Product Differentiation, Oligopoly, and Resource Allocation

Bruno Pellegrino

University of California Los Angeles

Job Market Paper

Please click here for the most recent version

January 23, 2020

Abstract

Industry concentration and profit rates have increased significantly in the United States over the past two decades. There is growing concern that oligopolies are coming to dominate the American economy. I investigate the welfare implications of the consolidation in U.S. industries, introducing a general equilibrium model with oligopolistic competition, differentiated products, and hedonic demand. I take the model to the data for every year between 1997 and 2017, using a data set of bilateral measures of product similarity that covers all publicly traded firms in the United States. The model yields a new metric of concentration—based on network centrality—that varies by firm. This measure strongly predicts markups, even after narrow industry controls are applied. I estimate that oligopolistic behavior causes a deadweight loss of about 13% of the surplus produced by publicly traded corporations. This loss has increased by over one-third since 1997, as has the share of surplus that accrues to producers. I also show that these trends can be accounted for by the secular decline of IPOs and the dramatic rise in the number of takeovers of venture-capital-backed startups. My findings provide empirical support for the hypothesis that increased consolidation in U.S. industries, particularly in innovative sectors, has resulted in sizable welfare losses to the consumer.

JEL Codes: D2, D4, D6, E2, L1, O4

Keywords: Competition, Concentration, Entrepreneurship, IPO, Market Power, Mergers, Misallocation, Monopoly, Networks, Oligopoly, Startups, Text Analysis, Venture Capital, Welfare

I gratefully acknowledge financial support from the Price Center for Entrepreneurship at UCLA Anderson. I am grateful to Jerry Hoberg, Gordon Phillips, and Nagpurnanand Prabhala for sharing their product-similarity data. I thank Andy Atkeson, David Baqaee, Ariel Burstein, Hugo Hopenhayn, Ed Leamer, William Mann, Stavros Panageas, Tommaso Valletti, Jonathan Vogel, Nico Voigtländer, and Romain Wacziarg for helpful feedback and discussions. I also wish to acknowledge thoughtful writing and presentation advice from John Asker, Alberto Bisin, Michela Giorcelli, Paola Giuliano, Thomas Groesbeck, Sebastian Ottinger, Nicola Persico, and Maria Lucia Yanguas, as well as outstanding data-collection assistance from Michele Schincaglia.

1. Introduction

Markups, profit rates, and industry concentration have all increased in the United States during the past two decades (De Loecker et al., 2018; Barkai, 2016; Grullon et al., 2018). These recent findings have spurred important public debates over whether these trends reflect a generalized oligopolization of U.S. industries and whether a revised antitrust policy is warranted (Khan, 2018; Werden, 2018). While standard price-theory arguments suggest that the welfare implications of these trends might be significant, interpreting these trends presents an imposing methodological challenge. The study of market power has traditionally resided within the domain of empirical industrial organization (EIO). Yet, there is a consensus that these trends are macroeconomic in nature: standard EIO methodologies are unfeasible, as they require data that is just not available for more than a handful of industries (Syverson, 2019).

This paper investigates the causes and welfare consequences of increasing industry concentration in the United States from 1997 to 2017. I address the existing methodological challenges by introducing a novel general-equilibrium model with oligopolistic competition, differentiated products, and hedonic demand. I estimate my model using an innovative data set by Hoberg and Phillips (2016), which is generated by a computational linguistic analysis of regulatory forms and that covers all publicly traded U.S. corporations. I compute several counterfactuals and policy experiments that address the following questions: (1) Are rising profits and markups a consequence of the increase in concentration? (2) How have consumer surplus and the welfare costs of concentration evolved as a consequence of increased industry consolidation during this period? (3) Why has industry concentration increased in the first place?

Using this new theoretical framework, I show that the increased concentration is consistent with the hypothesis that U.S. industries have indeed become more oligopolized. Specifically, I find that the deadweight losses induced by oligopolistic behavior have increased substantially over the last 20 years, and so has the share of total surplus that accrues to producers. My methodology also allows to me identify a proximate explanation for these trends: the dramatic rise in takeovers of venturecapital-backed startups that began in the mid 1990s, which coincided with the well-known secular decline in initial public offerings (IPOs) (Kahle and Stulz, 2017).

Economists have long been concerned with market power. Since the 1980s, the EIO literature has been developing a conceptual "toolkit" that researchers and antitrust enforcement practitioners have used to analyze market power across many industries (Einav and Levin, 2010).The EIO approach requires the researcher to first understand the structure of product rivalries in an industry: a firm's ability to price above marginal cost depends critically on the intensity of competition from firms that produce similar products. As a consequence, oligopoly power is inextricably linked to the notion of product differentiation: to measure market power in an industry with n firms, the economist effectively needs to first estimate n^2 cross-price demand elasticities—one for each pair of rivals. In industry studies, this is usually achieved by using a hedonic demand system (Berry, Levinsohn, and Pakes, 1995).

The recent interest in market power and antitrust has a distinctive macroeconomic angle. Because we do not observe output volume, prices, or product characteristics for a sufficiently large cross-section of industries, the EIO approach cannot be directly applied in a macroeconomic context. This challenge is compounded by the problem that, even at the macro level, product-market rivalry is not well approximated by industry classifications. The reason is that industry classifications (such as NAICS) tend to be based on similarities in the production process, not on the degree of product substitutability. In other words, they are appropriate for estimating production function, but they are unreliable when it comes to measuring the cross-price elasticity of substitution between products. In addition, the very concept of industry/sector is more fluid than economists tend to assume. While industry classifications are static, a significant percentage of the larger companies (those more likely to have market power) move with remarkable ease from one industry to another using R&D, mergers, spinoffs, and strategic alliances; some of them may be active in multiple industries and have been shown to strategically manipulate their industry classification—a phenomenon that has been dubbed *industry window dressing* (Chen et al., 2016).

Despite these challenges, the macroeconomics literature has made progress in incorporating market power into general equilibrium models: Baqaee and Farhi (2017, henceforth BF) have recently shown how to approximate the welfare costs of factor misallocation, under minimal assumptions, using the cross-sectional distribution of markups. This approach—by design—is agnostic about the origin of the observed variation in markups: its advantage is that it captures all observed variation in markups (and therefore all sources of inefficiency); the downside is that it does not model how the observed dispersion in markups originates in the first place. Therefore, a separate theory of markup formation is required to simulate specific policies.

This study breaks new ground by providing a theory of firm size and profitability that generalizes the Cournot oligopoly model to differentiated products and hedonic demand, and embeds it in a general equilibrium model. The objective of my model, rather than capturing all sources of misallocation, is to isolate the variation in firm size and markups that can be reliably attributed to product-market rivalry. Through this approach, I can quantify the contribution of each individual producer to aggregate welfare, and I can study the general equilibrium effects of events that are relevant to antitrust policy, such as mergers or the breaking up of an alleged monopoly.

To achieve this, my theoretical model dispenses with the notions of industry and sector altogether, building instead on the tradition of hedonic demand (Rosen, 1974). Thus, I can link the cross-price elasticity of demand between all firms in the economy to the fundamental attributes of each firm's product portfolio. Each firm's output is modeled as a bundle of characteristics that are individually valued by the representative consumer. The cross-price elasticity of demand between two firms depends on the characteristics embedded in their output. If the product portfolios of two companies contain similar characteristics, the cross-price elasticity of demand between their products is high. The result is a rather different picture of the product market: not a collection of sectors, but a network, in which the distance between nodes reflects product similarity and strategic interaction between firms.

I show that firms play a linear-quadratic game over a weighted network (Ballester et al., 2006; Ushchev and Zenou, 2018) if the following assumptions hold: (1) the representative consumer holds a linear-quadratic hedonic utility (Epple, 1987); (2) firms compete à la Cournot; and (3) the marginal-cost function is linear in output. Moreover, I embed the game in a full general equilibrium framework by replacing the numeraire good in the Epple (1987) utility specification with a linear disutility of labor. My model yields a prediction that can be directly tested using firm-level data. A firm that produces a distinct product (in the sense that it embeds characteristics that are scarce in the products offered by other firms) is going to earn significant monopoly rents.

To take the model to the data, I use a recently developed data set (Hoberg and Phillips, 2016, henceforth HP)that provides measures of product similarity for every pair of publicly traded firms in the United States. These product-similarity scores—which are based on a computational-linguistics analysis of the firms' regulatory 10-K forms—give rise to a continuous, high-dimensional representation of the product space. My model maps these bilateral similarity scores to an $n \times n$ matrix of cross-price elasticities of demand. Moreover, because HP's similarity scores are time-varying (yearly observations since 1997), my model is unique in that the degree of substitution between firms is allowed to change over time.

I find that—consistent with the model's prediction—firms that occupy a more peripheral position in the product network (as measured by HP's similarity scores) systematically charge higher markups. This is true even after controlling for narrow (6-digit) industry classifications.

I then use my model to compute the (static) deadweight loss from oligopoly and to simulate changes in total surplus and consumer surplus for a number of policy counterfactuals. I find sizable welfare costs of concentration. By moving to an allocation in which firms price at marginal cost (that is, in which they behave as if they were atomistic players in a perfectly competitive market), total surplus would rise by approximately 13.3 percentage points. At the same time, consumer surplus would more than double, as the entire surplus is reallocated from producers to consumers. By computing a separate counterfactual that only rectifies allocative distortions (markups are equalized, rather than eliminated, and input supply is assumed to be inelastic), I can determine that most of the welfare losses from market power—about 10 percentage points of the aforementioned 13.3—occur by way of factor misallocation. In other words, the deadweight losses are driven not by an underutilization of inputs but rather by a suboptimal mix of goods being produced. I also simulate a counterfactual in which all firms in the economy are owned by a single producer. Under this scenario, total surplus would drop by one-quarter, while consumer surplus would decrease by about 40%.

By mapping my model to the data for a period of 21 consecutive years, I can investigate the welfare consequences of the observed trends in concentration and markups between 1997 and 2017. I find that the share of surplus appropriated by companies in the form of oligopoly profits has increased from 33.2% (in 1997) to nearly 43.4% (in 2017). Concordantly, the welfare costs of oligopoly have increased over this period. In terms of total surplus, the gap between the competitive equilibrium and the first best has increased from 10% (in 1997) to 13.3% (in 2017); in terms of consumer surplus, the gap is much larger—increasing from slightly below 40% to nearly 51%. The resulting effect on the consumer could be best described as a double whammy: less surplus is produced overall (as a percentage of the surplus that *could* be produced), and less of the diminished surplus is allocated to the consumer in equilibrium. Thus, another unique feature of this model: the ability to analyze the distributional consequences of the increased industry consolidation across a number of industries.

Finally, I use the counterfactual-building capabilities of the model to explore the causes of rising concentration and oligopoly power. In particular, I study the effects of the dramatic secular shift in the type of venture capital (VC) exits observed in the past 20 years:[1] in the early 1990s, most VC-backed startups (80%–90%), if successful, would exit through IPOs. Today, the near entirety (~94%) of successful VC exits conclude with the startup being acquired by an incumbent. I find that this shift accounts not only for the secular decline in the number of public corporations in the United States (from about 7,500 in 1997 to about 3,500 in 2017) but also for the measured increase in the welfare costs of oligopoly, as well as the rising profit share of surplus. Moreover, it is consistent with the cross-section of these trends, as the welfare costs of oligopoly have seen the most dramatic increase in sectors that are most affected by VC activity (such as information technology and biotechnology). Overall, my results suggest that increased concentration

[1]In entrepreneurial finance lingo, an "exit" is the termination of a VC investment and should not be confused with a business termination. If the VC investor exits with an IPO, that event marks the opposite of an enterprise death.

and markups resulted in sizable welfare losses and affected how surplus is shared between producers and consumers.

This paper contributes to two growing branches of macroeconomics that study imperfect competition. First is the literature on misallocation (see Hopenhayn, 2014, for a review). This literature starts from the observation that, in a frictionless economy, marginal-revenue-cost gaps should be equalized across productive units. It therefore uses markups (Baqaee and Farhi, 2017) or a metric derived from markups (Hsieh and Klenow, 2009) as a sufficient statistic of competitive distortions. Second is the branch of macroeconomics (Autor et al., 2017; Gutiérrez and Philippon, 2017; Grullon et al., 2018) that is less concerned with generic frictions than with frictions that arise from concentration and product-market rivalry. Here, the agenda is to accurately collect evidence relating macroeconomic trends to changes in the extent of competition across firms. Because this literature is more specific in terms of what drives the inefficiencies, it leads more directly to actionable policies; yet, it can't speak about welfare in a general equilibrium setting[2].

This paper also incorporates key insights from the work of Edmond, Midrigan, and Xu (2018), specifically the role of demand convexity in modeling markup heterogeneity across firms. While my model maps differently to the data and relaxes the free-entry assumption to pose a different set of policy questions (their focus is fiscal policy, while my focus is competition policy), we both depart from CES preferences in favor of a demand aggregator that correlates markups to firm size[3].

This paper also speaks to the empirical corporate-finance literature on the secular decline in the number of public companies (Kahle and Stulz, 2017; Doidge et al., 2018) and on the declining rate of IPOs (Gao, Ritter, and Zhu, 2013; Bowen, Frésard, and Hoberg, 2018; Ewens and Farre-Mensa, 2018). Specifically, it offers a way to quantify the effects of these trends on competition and on public corporations' ability to appropriate surplus.

The rest of the paper is organized as follows. In Section 2, I present my theoretical model. In Section 3, I present the data used throughout the empirical part of the paper (including HP's data set) and show how it can be mapped to the model. In Section 4, I present empirical results. In Section 5, I perform welfare calculations and present a number counterfactual analyses. In Section 6, I investigate the causes of the increase in oligopoly power, focusing on the role of the decline of IPOs and the rise of startup takeovers. In Section 7, I discuss a number of extensions, robustness checks, and limitations. In Section 8, I present my conclusions and discuss how my findings can inform the current debate on market power and antitrust policy.

[2]These two literatures share a common precursor in Harberger (1954), who famously estimated the welfare costs of monopoly power using sector-level data from the 1920s: 0.1% of the industrial value added.

[3]They use the Kimball aggregator. I adopt a linear-quadratic aggregator, which has similar properties: namely, it can support a high degree of demand "super-elasticity."

2. A theory of Imperfect, Networked Competition

In this section, I present a general equilibrium model in which firms produce differentiated products and compete à la Cournot. My model has some points of contact with the canonical model of Atkeson and Burstein (2008), who also assume oligopolistic competition with differentiated products. A major difference lies in the functional specification of the representative agent's utility. I depart from (nested) CES preferences in favor of a hedonic demand model with quadratic utility. The advantage of this specification is that it does not require us to assume a certain nesting structure in the preferences. This will in turn allow, in the empirical application of the model, the 10-K text data to speak with regard to the degree of substitutability between products. Even more important, it offers a way to model how the matrix of cross-price elasticities evolves over time as the number of firms and their product portfolios change.

I start by laying out the basic model. After characterizing the equilibrium of this model economy, I study its properties and use them to define a new firm-level measure of product-market competition, which I will later use in the empirical section.

2.1. Notation

I use standard linear algebra notation throughout the paper: lowercase italic letters (s) denote scalars, lowercase bold letters ($\mathbf{v}$) denote vectors, and uppercase bold letters denote matrices ($\mathbf{M}$). The following convention is the only nonstandard notation in this paper: when the same letter is used simultaneously (without subscript) to denote both a vector and a matrix, it represents the same object arranged, respectively, as a column vector and as a diagonal matrix. Moreover, when the same letter is used as an uppercase italic, it represents the summation of the components of the vector. Formally:

$$\mathbf{v}=\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \qquad \Longrightarrow \qquad \mathbf{V} \overset{\text{def}}{=} \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_n \end{bmatrix} \qquad \text{and} \qquad V \overset{\text{def}}{=} \mathbf{1}'\mathbf{v} \equiv \sum_i v_i \tag{1}$$

2.2. Basic Setup

There are n firms that produce differentiated products. Following the tradition of hedonic demand in differentiated product markets (Rosen, 1974), I assume that consumers value each product $i \in \{1, 2, ..., n\}$ as a bundle of k characteristics: each product i can provide more or less of a certain characteristic $j \in \{0, 1, 2, ..., k\}$, and can therefore be described by a k-dimensional unit vector $\mathbf{a}_i$ The coordinate a_{ji} is the number of units of characteristic j embedded in product i. Formally, the unit-valuation restriction for $\mathbf{a}_i$ is:[4]

$$\sum_{i=1}^{k} a_{ji}^2 = 1 \quad \forall\, i \in \{1, 2, ..., n\} \tag{2}$$

The assumption that $\mathbf{a}_i$ is a unit vector amounts to a normalization assumption. For every product,

[4]I abstract from multiproduct firms in the basic exposition of the model. In Section 7, I show how the model can be extended to multiproduct firms with minimal modifications.

we need to pick an output volume metric (kilograms, pounds, gallons, etc.). The normalization consists in picking the volume unit so that each unit is geometrically represented by a point on the k-dimensional hypersphere.

We can stack all the vectors $\mathbf{a}_i$—that is, the coordinates of all firms in the product characteristics space $\mathbb{R}^k$—inside matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \tag{3}$$

The n-dimensional vector $\mathbf{q} = (q_1, q_2, ..., q_n)$ contains, for every coordinate $i = 1, 2, ..., n$, the number of units of good i that are purchased by the representative consumer. Because the welfare properties of the economy can be entirely described in terms of the quantity produced by each firm ($\mathbf{q}$), I shall refer to a generic $\mathbf{q}$ as an "allocation."

Definition 1 (Allocation)**.** A vector $\mathbf{q}$ that specifies, for every firm, the number of units produced and sold is called an *allocation.*

The matrix $\mathbf{A}$ transforms the vector of units of goods purchased $\mathbf{q}$ into units of characteristics $\mathbf{x}$, which is what the representative agent ultimately values:

$$\mathbf{x} = \mathbf{A}\mathbf{q} \tag{4}$$

Labor is the only factor of production. I denote by the vector $\mathbf{h} = (h_1, h_2, ..., h_n)$ the labor input acquired by every firm. Firms convert labor hours into units of differentiated goods using the following linear technology:

$$q_i = \omega_i h_i \tag{5}$$

where ω_i is firm i's productivity. Because I take labor hours to be the numeraire good of this economy, the unit cost of production is simply equal to the inverse of productivity; that is, $c_i = \omega_i^{-1}$.

The representative consumer's preferences are defined using a quasi-linear, hedonic utility specification $U(\cdot)$ that is quadratic in the vector of characteristics $\mathbf{x}$, and incorporates a linear disutility for the total number of hours of work supplied (H):

$$U(x_1, x_2, .., x_k, H) \stackrel{\text{def}}{=} \sum_{j=1}^{k} \left(b_j x_j - \frac{1}{2} x_j^2 \right) - H \tag{6}$$

where b_j is a characteristic-specific preference shifter. This function can be concisely written, in vector form, as:

$$U(\mathbf{x}, H) \stackrel{\text{def}}{=} \mathbf{b}'\mathbf{x} - \frac{1}{2}\mathbf{x}'\mathbf{x} - H \tag{7}$$

A similar utility specification was adopted by Epple (1987), who used it to study the econometric identification of hedonic demand systems. To close the model, I modify the utility function by simply replacing the numeraire good with a linear disutility for labor hours. The representative consumer buys goods vector $\mathbf{q}$ taking $\mathbf{p}$ (the vector of prices) as given. Moreover, I assume that the representative consumer is endowed with the shares of all the companies in the economy. As a consequence, the aggregate profits are paid back to her. Her consumption basket, defined in terms of the unit purchased $\mathbf{q}$, has to respect the following budget constraint:

$$H + \Pi \; = \; \sum_{i=1}^{k} p_i q_i \tag{8}$$

where Π is the sum of the profits earned by all firms in the economy. The aggregate resource constraint is:

$$H \stackrel{\text{def}}{=} \sum_i h_i \; = \; \sum_i c_i q_i \; = \; \sum_i \omega_i^{-1} q_i \tag{9}$$

Plugging equation (4) inside (7), we obtain the following Lagrangian for the representative consumer:

$$\mathscr{L}(\mathbf{q}, H) \; = \; \mathbf{q}'\mathbf{A}'\mathbf{b} - \frac{1}{2}\mathbf{q}'\mathbf{A}'\mathbf{A}\mathbf{q} - H - \Lambda\left(\mathbf{p}'\mathbf{q} - H - \Pi\right) \tag{10}$$

The choice of work hours as the numeraire immediately pins down the Lagrange multiplier $\Lambda = 1$. Then, the consumer chooses a demand function $\mathbf{q}(\mathbf{p})$ to maximize the following consumer-surplus function:

$$S(\mathbf{q}) = \mathbf{q}'\left(\mathbf{A}'\mathbf{b} - \mathbf{p}\right) - \frac{1}{2}(\mathbf{q}'\mathbf{A}'\mathbf{A}\mathbf{q}) \tag{11}$$

The matrix $\mathbf{A}'\mathbf{A}$, which I assume to be invertible, contains the *cosine similarity scores* between all pairs of firms in the characteristics space. More explicitly, the component $(\mathbf{A}'\mathbf{A})_{ij} = \mathbf{a}_i'\mathbf{a}_j$ measures the cosine of the angle between vectors $\mathbf{a}_i$ and $\mathbf{a}_j$ in the space of characteristics $\mathbb{R}^k$:[5] A higher cosine similarity score reflects a lower angular distance. In other words, if the cosine similarity between i and j ($\mathbf{a}_i'\mathbf{a}_j$) is high, the outputs of i and j contain a more similar set of characteristics. The intuition for the fact that the quadratic term contains this matrix is that if two products i and j contain a *similar* set of characteristics (that is, if the cosine between i and j is high), there is a high degree of substitution between these two products; as a consequence, an increase in the supply of product i will have a large negative impact on the marginal utility provided by one additional unit of product j.

Figure 1 helps visualize this setup for the simple case of two firms—1 and 2—competing in the space of two characteristics A and B. As can be seen in the figure, both firms exist as vectors on the unit circle (with more than three characteristics, it would be a hypersphere). The cosine similarity $\mathbf{a}_i'\mathbf{a}_j$ captures the tightness of the angle θ and, therefore, the similarity between firm 1 and firm 2. An increase in the cosine of the angle θ (a lower angular distance) reflects a more similar set of characteristics, and therefore a higher degree of substitution between firm 1 and firm 2.

The demand and inverse demand functions are given by:

$$\text{Aggregate demand}: \qquad \mathbf{q} = \left(\mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b} - \mathbf{p}\right) \tag{12}$$

$$\text{Inverse demand}: \qquad \mathbf{p} = \mathbf{A}'\mathbf{b} - \mathbf{A}'\mathbf{A}\mathbf{q} \tag{13}$$

Notice that the quantity sold by each firm may affect the price of the output sold by every other firm in the economy (unless the matrix $\mathbf{A}'\mathbf{A}$ equals the identity matrix), hence there is imperfect substitutability

[5]This is a consequence of the normalization assumption that all vectors $\mathbf{a}_i$ are unit vectors.

among the products. In particular, the derivative $\partial p_i/\partial q_j$ is equal to $\mathbf{a}_i'\mathbf{a}_j$: the product similarity between i and j. The closer two firms are in the product-characteristics space, the higher the cross-price elasticity between the two firms. Because $\mathbf{A}'\mathbf{A}$ is symmetric, we have $\partial q_i/\partial p_j = \partial q_j/\partial p_i$ by construction. As a consequence, the (inverse) cross-price elasticities of demand are:

$$\text{Inverse cross} - \text{price elasticity of demand}: \quad \frac{\partial \log p_i}{\partial \log q_j} = -\frac{q_j}{p_i} \cdot \mathbf{a}_i'\mathbf{a}_j \tag{14}$$

$$\text{Cross} - \text{price elasticity of demand}: \quad \frac{\partial \log q_i}{\partial \log p_j} = -\frac{p_j}{q_i} \cdot (\mathbf{A}'\mathbf{A})^{-1}_{ij} \tag{15}$$

My choice to use a linear demand system is motivated by a recent literature that has investigated the implications of different demand systems on allocative efficiency and market power.[6] Linear demand has *super-elasticity*—that is, the elasticity of demand decreases with firm size.[7] I discuss the implications of linear demand at length in Section 7.

I can now define the vector of firm profits π as follows:

$$\begin{aligned} \pi(\mathbf{q}) &\stackrel{\text{def}}{=} \mathbf{Q}[\mathbf{p}(\mathbf{q}) - \mathbf{c}] \\ &= \mathbf{Q}(\mathbf{A}'\mathbf{b} - \mathbf{c}) - \mathbf{Q}\mathbf{A}'\mathbf{A}\mathbf{q} \end{aligned} \tag{16}$$

each component π_i quantifies the profits of firm i . Firms compete à la Cournot. That is, each firm i strategically chooses its output volume q_i by taking as given the output of all other firms. By taking the profit vector as a payoff function and the vector of quantities produced $\mathbf{q}$ as a strategy profile, I have implicitly defined a linear-quadratic game over a weighted network.

This class of games, which has been analyzed by Ballester et al. (2006, henceforth BCZ), belongs to a larger class of games called "potential games"(Monderer and Shapley, 1996): their key feature is that they can be described by a scalar function $\Phi(\mathbf{q})$, which we call the game's *potential*. The potential function can be thought of, intuitively, as the objective function of the "pseudoplanner" problem that is solved by the Nash equilibrium allocation. The potential function is shown below, together with the aggregate profit function $\Pi(\mathbf{q})$ and the aggregate welfare function $W(\mathbf{q})$:

$$\text{Aggregate Profit}: \quad \Pi(\mathbf{q}) = \mathbf{q}'(\mathbf{A}'\mathbf{b} - \mathbf{c}) - [\mathbf{q}'\mathbf{q} + \mathbf{q}'(\mathbf{A}'\mathbf{A} - \mathbf{I})\mathbf{q}] \tag{17}$$

$$\text{Cournot Potential}: \quad \Phi(\mathbf{q}) = \mathbf{q}'(\mathbf{A}'\mathbf{b} - \mathbf{c}) - \left[\mathbf{q}'\mathbf{q} + \frac{1}{2}\mathbf{q}'(\mathbf{A}'\mathbf{A} - \mathbf{I})\mathbf{q}\right] \tag{18}$$

$$\text{Total Surplus}: \quad W(\mathbf{q}) = \mathbf{q}'(\mathbf{A}'\mathbf{b} - \mathbf{c}) - \left[\frac{1}{2}\mathbf{q}'\mathbf{q} + \frac{1}{2}\mathbf{q}'(\mathbf{A}'\mathbf{A} - \mathbf{I})\mathbf{q}\right]. \tag{19}$$

The three functions are visually similar to each other; they differ only by the scalar weight applied to the quadratic terms. In writing these functions, I separated, on purpose, the diagonal components of the quadratic term from the off-diagonal components. As can be seen from equations (17)-(19), the Cournot

[6] See Dhingra and Morrow (forthcoming); Edmond et al. (2018); Haltiwanger et al. (2018).

[7] While other utility specifications display super-elasticity (Edmond et al., 2018 for example use instead a Kimball aggregator), linear demand makes the model not only tractable but also empirically relevant as it enables one to relate the cross-derivatives of the demand system to available product-similarity data.

potential is somewhat of a hybrid: the diagonal elements of the quadratic terms are the same as the aggregate profit function, while the off-diagonal terms are the same as the aggregate surplus function. By maximizing the potential $\Phi(\mathbf{q})$, we find the Cournot-Nash equilibrium.

Proposition 1. *The Cournot-Nash equilibrium of the game described above is given by the maximizer of the potential function $\Phi(\cdot)$, which I label $\mathbf{q}^{\Phi}$:*

$$\mathbf{q}^{\Phi} \quad \stackrel{\text{def}}{=} \quad \arg\max_{\mathbf{q}} \Phi(\mathbf{q}) \quad = \quad (\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}(\mathbf{A}'\mathbf{b}-\mathbf{c}) \tag{20}$$

Proof. The derivation of the potential function, as well as the proof that its maximizer $\mathbf{q}^{\Phi}$ is the genuine Nash equilibrium, appear in Appendix A. □

An intuitive interpretation of (20), which characterizes the Cournot-Nash equilibrium, is that there are two ways for firms to be profitable (in this model): (1) produce at low costs and sell large volumes[8]; (2) produce a highly differentiated good that commands a high markup (Porter, 1991). The value-adding capabilities of firm i, which are captured by the term $(\mathbf{a}_i\mathbf{b}-c_i)$, are not the only determinant of a firm's profitability: the firm's position in the product characteristic space, captured by the matrix $(\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}$, also matters.

The discrepancy between the potential function and the total-surplus function implies that the network Cournot game delivers an equilibrium allocation that will generally differ from the social optimum. Benevolent social planners can theoretically improve on the market outcome because they can coordinate output choices across firms.

Under the current volumetric normalization, the unit profit margin is easily verified to be equal to the output volume:

$$\mathbf{p}-\mathbf{c} \;=\; \mathbf{q} \quad . \tag{21}$$

This is a consequence of my output volumetric unit normalization. It is easy to see that the Cournot-Nash equilibrium respects the markup rule. The Lerner index is equal to the inverse of the own price elasticity (remember that I've normalized the output measurement units in such a way that the own slope of the inverse demand function is one for all firms):

$$\frac{p_i-c_i}{p_i} \;=\; \frac{\partial p}{\partial q}\cdot\frac{q_i}{p_i} \;=\; \frac{q_i}{p_i} \quad . \tag{22}$$

2.3. The Symmetric Case: Comparative Statics, Endogenous Entry, and Identification in the Structure-Conduct-Performance Literature

In this subsection, I perform some comparative statics on the Cournot-Nash equilibrium of a special case of the model just described—specifically, that in which the network is symmetrical. Here, the matrix $\mathbf{A}'\mathbf{A}$ takes the form:

$$\mathbf{A}'\mathbf{A}_{ij} \;=\; \begin{cases} 1 & \text{if} \quad i=j \\ a & \text{if} \quad i\neq j \end{cases} \tag{23}$$

[8]In the strategic management literature, this would be called *cost leadership* (Porter, 1991).

and firms are identical. I focus on this special case because I wish to study how various equilibrium quantities react to changes in similarity between firms, which in this case can be summarized by a single parameter. In particular, assume that

$$\mathbf{a}_i'\mathbf{b} \;=\; 3, \qquad c_i = 1 \qquad \forall\, i \in \mathcal{I} \tag{24}$$

the parameter $a \in [0,1]$ controls the degree of similarity across firms. The equilibrium quantities and markup simplify to:

$$q_i \;=\; \mu_i - 1 \;=\; \frac{1}{1+\frac{1}{2}\left(n-1\right)a} \quad \forall i \tag{25}$$

where n is the number of firms.

Figure 2 plots the markup $\mu\left(a\right)$ for the case considered above, for different values of the similarity parameter a and different number of firms n. As the similarity (a) coefficient goes to zero, every firm effectively becomes a monopolist. This is also reflected in the potential function $\Phi\left(\mathbf{q}\right)$, which converges to the aggregate profit function $\Pi\left(\mathbf{q}\right)$. Conversely, as similarity coefficients go to one, the model reduces to the classical Cournot Oligopoly. The canonical Cournot model inherited by this model offers another feature: it yields the first-best allocation in the limit—that is, as the number of firms in the economy grows unboundedly (see subsection 5.3 for a convergence result that applies to the nonsymmetric case).

I use this simple symmetrical example to illustrate how modeling product substitution across firms explicitly is critical in avoiding a major identification problem that for decades has plagued the so-called structure-conduct-performance literature (Bresnahan and Schmalensee, 1987; Bresnahan, 1989). Let us start by rearranging the equilibrium condition in equation (25) to obtain the following:

$$\frac{4-2\mu_i}{\mu_i - 1} \;=\; \left(n-1\right)a \quad \forall i \tag{26}$$

Notice that we can write n as the inverse of the Herfindahl-Hirschmann Index (as firms are all identical), which I denote by the Greek capital letter Heta (Ͱ). Rearranging we obtain:

$$\log\frac{\mu_i - 1}{4-2\mu_i} \;=\; \log\frac{\text{Ͱ}}{1-\text{Ͱ}} - \log a \quad \forall i \tag{27}$$

The lefthand side is an increasing function of the markup, while the righthand side is the difference between an increasing function of the Herfindahl-Hirschmann Index (HHI) and the log of similarity (a). This relationship reflects that a higher concentration leads to a lower demand elasticity for individual firms. Motivated by the positive relationship implied by the Cournot model, the structure-conduct-performance literature has tried, for a long time, to uncover a positive association at the industry level between markups and concentration. One major problem, I argue, in correlating concentration with markups is that omitting a can fundamentally invalidate the HHI as a measure of concentration.

To see just how problematic it can be to ignore product similarity in interpreting HHIs, let me introduce, in this simple symmetrical model, endogenous entry. Firms now pay a fixed cost f^2 to enter, and the number of firms in the economy is determined by the free-entry condition:

$$\sqrt{\pi_i} \;=\; q_i \;=\; \mu_i - 1 \;=\; f \tag{28}$$

This effectively renders Ͱ endogenous. Moreover, in the free-entry equilibrium, the following equality holds:

$$\log a \;=\; \log\frac{\mathrm{H}}{1-\mathrm{H}}-\log\frac{f}{1-f} \tag{29}$$

by replacing this expression for $\log a$ inside equation (27), we can verify that the implied correlation between markups and concentration is zero. Hence, omitting product similarity in these types of regressions leads to nonidentificaion when entry is endogenous. The product similarity data of Hoberg and Phillips (2016) offers an avenue to measure this important (previously) unobserved variable.

2.4. Separability of Consumer Surplus

In this model, two forces drive cross-sectional differences in market power across firms. The more familiar one is the incomplete passthrough from marginal cost to prices, which allows larger firms to charge higher markups. The second force, a feature of hedonic demand models, is the varying degree of firms' product *uniqueness.* That is, each product might have few or many other products with similar characteristics that can act as substitutes. A firm's size and product uniqueness affect its ability to appropriate surplus. This begs the question, How do we measure surplus appropriation in a model like this, where patterns of substitution may differ across firm pairs?

To answer that question, I now discuss a key property of my utility specification. I show that the consumer surplus earned by the representative agent is *separable* by firm. This means we can attribute a certain share of the consumer surplus produced to each firm. This is possible, in this particular case, thanks to the linear quadratic utility specification. This separability property is the key to deriving a measure of competition that varies by firm, which I will later use in my empirical analysis.

Definition 2 (Separable Consumer Surplus)**.** Assume that the allocation $\mathbf{q}$ maximizes the consumer utility given the price vector $\mathbf{p}$. We say that the consumer surplus $S(\mathbf{q})$ is *separable* if it can be written as the sum across firms of some function $s(\mathbf{a}_i, q_i, p_i)$ that only depends on the triple $(\mathbf{a}_i, q_i, p_i)$. That is:

$$S(\mathbf{q}) \;=\; \sum_i s(\mathbf{a}_i, q_i, p_i)$$

Proposition 2. *In the model previously presented, the consumer-surplus function $S(\mathbf{q})$ is separable.*

Proof. Noting that the inverse demand function can be rewritten as $\mathbf{A}'\mathbf{A}\mathbf{q} = \mathbf{A}'\mathbf{b} - \mathbf{p}$, we have:

$$\begin{aligned} S(\mathbf{q}) \;&=\; \mathbf{q}'(\mathbf{A}'\mathbf{b}-\mathbf{p}) - \frac{1}{2}\mathbf{q}'\mathbf{A}'\mathbf{A}\mathbf{q} \\ &=\; \mathbf{q}'(\mathbf{A}'\mathbf{b}-\mathbf{p}) - \frac{1}{2}\mathbf{q}'(\mathbf{A}'\mathbf{b}-\mathbf{p}) \;=\; \frac{1}{2}\mathbf{q}'(\mathbf{A}'\mathbf{b}-\mathbf{p}) \end{aligned}$$

the last term can be rewritten in summation form as $\sum_i \frac{1}{2} q_i (\mathbf{a}_i'\mathbf{b} - p_i)$. □

The firm-level consumer surplus, which attributes to each firm i a certain share s_i of the consumer surplus $S(\mathbf{q})$, is then defined as:

$$\mathbf{s}(\mathbf{q}) \quad \stackrel{\text{def}}{=} \quad \frac{1}{2}\mathbf{Q}\mathbf{A}'\mathbf{A}\mathbf{q} \tag{30}$$

We can then also define a firm-level total surplus function, which specifies for every firm i a certain share w_i of the total surplus $W(\mathbf{q})$, is defined as follows:

$$\mathbf{w}\left(\mathbf{q}\right) \quad \stackrel{\text{def}}{=} \quad \pi\left(\mathbf{q}\right) + \mathbf{s}\left(\mathbf{q}\right) \quad = \quad \mathbf{Q}\left(\mathbf{A}'\mathbf{b} - \mathbf{c}\right) - \frac{1}{2}\mathbf{Q}\mathbf{A}'\mathbf{A}\mathbf{q} \tag{31}$$

2.5. Measuring Oligopoly Power at the Firm Level

The canonical Cournot oligopoly model establishes the HHI as a measure of market power. The reason for that is that the HHI relates the (market share-)weighted average firm-level inverse demand elasticity to the industrywide inverse demand elasticity. Let $Q = \sum_i q_i$. Then:

$$\begin{aligned} \frac{\partial \log p}{\partial \log q_i} &= \frac{q_i}{Q} \cdot \frac{\partial \log p}{\partial \log Q} \\ \Longrightarrow \frac{q_i}{Q} \cdot \frac{\partial \log p}{\partial \log q_i} &= \left(\frac{q_i}{Q}\right)^2 \cdot \frac{\partial \log p}{\partial \log Q} \\ \Longrightarrow \sum_i \frac{q_i}{Q} \cdot \frac{\partial \log p}{\partial \log q_i} &= \mathrm{H} \cdot \frac{\partial \log p}{\partial \log Q} \end{aligned} \tag{32}$$

The only reason the HHI is informative as a measure of market power is that the individual *market shares* are themselves informative about the demand elasticity of firm i—this fact is frequently overlooked or forogotten The first line of equation (32) evinces this: the ratio of the inverse demand elasticities for firm i and the industry as a whole is simply the market share of firm i. Hence, if we wanted to derive a firm-level counterpart of the HHI index, it would simply be the market share of firm i. Next, I return to the network Cournot model.

Definition 3 (Weighted Market Share)**.** I define the (similarity-)weighted market share σ_i of firm i as follows:

$$\sigma_i \quad \stackrel{\text{def}}{=} \quad \frac{q_i}{\sum_j \mathbf{a}_{ij} q_j} \tag{33}$$

It is possible to show that the following lemma holds.

Lemma 1. *In the Cournot-Nash equilibrium allocation, the ratio of firm profits π_i to consumer surplus s_i is equal to twice the (similarity-)weighted market share σ_i of firm i:*

$$\frac{\pi_i}{2s_i} \quad = \quad \sigma_i \quad \stackrel{\text{def}}{=} \quad \frac{q_i}{\sum_j \mathbf{a}_{ij} q_j} \tag{34}$$

Proof. The element-by-element ratio $(\mathbf{s}/\pi)$ is equal to $\frac{1}{2}\mathbf{Q}^{-2}\mathbf{Q}\left(\mathbf{A}'\mathbf{A}\right)\mathbf{q} \;=\; \frac{1}{2}\mathbf{Q}^{-1}\left(\mathbf{A}'\mathbf{A}\right)\mathbf{q}$. □

Notice that, under canonical Cournot oligopoly ($\mathbf{A}'\mathbf{A} = \mathbf{1}\mathbf{1}'$) this is simply the market share of firm i . Therefore, in the Network Cournot model, we can use the similarity-weighted market share σ_i as a firm-level alternative to the HHI that accounts for product differentiation.

Lemma 2. Equation (20) can be rewritten in terms of σ_i as follows:

$$q_i \;=\; \frac{\sigma_i}{1+\sigma_i}\left(\mathbf{a}_i'\mathbf{b} - c_i\right) \tag{35}$$

Proof. See Appendix R. □

By comparing equation (35) to equation (20), we can see that σ_i effectively allows us to replace, in the equilibrium allocation, the matrix $(\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}$ with a diagonal matrix:

$$\begin{aligned} \mathbf{q} &= (\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}(\mathbf{A}'\mathbf{b}-\mathbf{c}) \\ &= (\mathbf{I}+\Sigma)^{-1}\Sigma(\mathbf{A}'\mathbf{b}-\mathbf{c}) \end{aligned}$$

In other words, it allows me to summarize the network of strategic rivalries among the n firms in the model into a vector. This provides a second, intuitive reason for how it measures the intensity of competition for different firms.

Similarly to the HHI, the similarity-weighted market share is an equilibrium object—an endogenous outcome variable. To test whether a firm's position in the network predicts the firm's markup (one of the key testable predictions of the model), I want to derive a measure of oligopoly power at the firm level that uses no endogenous outcome variable (such as the firms' output vector $\mathbf{q}$). To do so, I simply compute the similarity-weighted market shares, for the special case where $(\mathbf{A}'\mathbf{b}-\mathbf{c})$ is constant across firms. I call this the *Inverse Centrality* index.

Definition 4 (Inverse Centrality)**.** I define χ_i, the inverse centrality index for firm i, as the solution to the following equation:

$$\frac{\chi_i}{1+\chi_i} = (\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}\mathbf{1}_i \tag{36}$$

I call this *Inverse Centrality* because this measure coincides with the Katz-Bonacich Centrality of firms in the network with negatively signed adjacency matrix $(\mathbf{I}-\mathbf{A}'\mathbf{A})$. Because the network weights are negative, the interpretation is opposite: it captures *peripherality* (or *eccentricity*), rather than centrality. The relationship between these measures of oligopoly power and known measures of network centrality is explained in Appendix B.

Two main differences exist between the *Weighted Market Share* and the *Inverse Centrality* measures. The first difference has to do with how they are computed: the first measure requires firm-level financial data to be computed, while the latter measure can be computed solely on the basis of the network links. The second difference has to do with the economic intuition behind each measure. The first measure captures two dimensions of concentration and market power: sheer size, and the firm's position in the network. The latter measure, on the other hand, captures only the network configuration (not firm size).

In the next section, I show how to compute both these measures using the data of Hoberg and Phillips (2016). I will then validate them and put them to use in the empirical section of this paper (Section 4).

3. Data and Calibration

In this section, I outline the data used to estimate the model in Section 2. The sources, as well as the mapping to the theory, are outlined in Table 1.

3.1. Text-Based Product Similarity

The key data input required to apply my model to the data is the matrix of product similarities $\mathbf{A}'\mathbf{A}$. The empirical counterpart to this object is provided by Hoberg and Phillips (2016, henceforth HP), who computed product cosine similarities for firms in Compustat by analyzing the text of their 10-K forms.

By law, every publicly traded firm in the United States must submit a 10-K form annually to the Securities and Exchange Commission (SEC);the form contains a *product description* section, which is the target of the algorithm devised by HP. They built a vocabulary of 61,146 words that firms use to describe their products,[9] and that identify product characteristics. Based on this vocabulary, HP produced, for each firm i, a vector of word occurrences $\mathbf{o}_i$.

$$\mathbf{o}_i = \begin{bmatrix} o_{i,1} \\ o_{i,2} \\ \vdots \\ o_{i,61146} \end{bmatrix} \tag{37}$$

This vector is normalized to be of length one. HP then used the dot product of the normalized vectors to compute a matrix of cosine similarities. To the extent that the vocabulary used by HP correctly identifies product characteristics, the resulting matrix is the exact empirical counterpart to $\mathbf{A}'\mathbf{A}$——the matrix of cross-price effects in my theoretical model. The fact that all publicly traded firms in the United States are required to file a 10-K form makes the HP data set unique: it is the only data set that covers the near entirety (97.8%) of the CRSP-Compustat universe.

One of HP's objectives in developing this data set was to remedy two well-known shortcomings of the traditional industry classifications: (1) the inability to capture imperfect substitutability between products, which is the most salient feature of my model; and (2) the fact that commonly used industry classifications, such as SIC and NAICS, are based on similarity in *production processes*, rather than in product characteristics—in other words, they are appropriate for estimating production function, but unsuitable for measuring the elasticity of substitution between different products.

Though other data sets have a network structure that one could potentially utilize to estimate $\mathbf{A}'\mathbf{A}$[10], they all have the following shortcomings; (a) they are all either directly or indirectly based on industry classifications; and (b) they fail to meet the data-coverage requirements for my empirical exercise. In terms of coverage, both across firms and across time, no other available data set comes close to HP's—it is the only data set that allows me to cover a meaningful share of the economic activity in the United States, for every year since 1997.

I use a second similarity score in this paper (Section 6). I source this data from Hoberg, Phillips, and Prabhala (2014); it is computed analogously to that described above. It measures product similarity between firms in Compustat and startups from the VenturExpert database. The authors extend HP's similarity scores

[9] I report here verbatim the description of the methodology from the original article by Hoberg and Phillips (2016):"[...] *In our main specification, we limit attention to nouns (defined by Webster.com) and proper nouns that appear in no more than 25 percent of all product descriptions in order to avoid common words. We define proper nouns as words that appear with the first letter capitalized at least 90 percent of the time in our sample of 10-Ks. We also omit common words that are used by more than 25 percent of all firms, and we omit geographical words including country and state names, as well as the names of the top 50 cities in the United States and in the world.* [...]"

[10] Bloom, Schankerman, and Van Reenen (2013), for example, use Compustat segment data. Another data set that contains rivalry relationships is CapitalIQ.

to VCbacked startups by using, for this universe of firms, their business descriptions from VenturExpert in place of the 10-K product description.[11]

3.2. Mapping the Model to Compustat Data

My data source for firms' financials and performance measures is the CRSP-Compustat merged database. From this database, I extract information on firm revenues, operating costs (costs of goods sold [COGS] + selling, general, and administrative costs [SG&A]), market capitalization, stock price, common shares outstanding, and book and redemption value of preferred stock and total assets. Using these balance-sheet and stock-market figures, I compute the enterprise value.

One challenge of mapping my model to firm-level data is that, because there is no uncertainty in demand or productivity, profits are always positive and markups are always above one. In this sense, my model should be thought of as a long-run equilibrium. This complicates applying the model to the data, since many firms in Compustat report negative income.

In what follows, I propose a mapping that preserves the average level of markups. The key is to obtain a strictly positive measure of *expected* profits. One way to achieve this is to use the market value of the company, which can be obtained from Compustat-CRSP and can be interpreted as the discounted present value of the company's future profits. Letting k_{it} be the market value of company i's assets at the beginning of period t and π_{it} the expected profits during period t we have, by no-arbitrage:

$$k_{it} = \frac{\pi_{it} - i_{it} + k_{it+1}}{1 + r_t} \tag{38}$$

where r is the required rate of return on capital and i_{it} is the required capital investment during period t. Let δ be the rate of depreciation. If the per-unit price of capital is constant, and the company's assets follow the law of motion

$$k_{it+1} = i_{it} + (1 - \delta)\, k_{it} \tag{39}$$

by plugging equation (39) inside (38) and rearranging, we obtain the following relationship:

$$\pi_{it} = (r_t + \delta)\, k_{it} \tag{40}$$

I can therefore use, subject to an estimate of $(r_t + \delta)$, the market value of each company from CRSP-Compustat to obtain a strictly positive measure of expected profits. My estimate of $(r_t + \delta)$ is:

$$r_t + \delta = \frac{\Pi_t}{K_t} \tag{41}$$

where Π_t is of aggregate *realized* operating profit, which I sum over all Compustat companies in year t, and K_t is the aggregate market value of all capital. In other words, I use the expected profits predicted by the cross-section of firm valuations to smooth out losses as well as short-term volatility. The implicit assumption that underlies my estimate of $(r_t + \delta)$ is that there is no aggregate risk on profits.

I adopt this approach for two reasons. First, by construction, given a firm-level measure of variable

[11] These scores cover 1996 to 2008. For the most recent years, I use the last-available score, where available. For a few firms that join the Compustat database after 2008, I use the VC-similarity score predicted by their NAICS industry membership.

costs[12] and letting $\tilde{p}_i q_i$ being firm i's realized revenues, this choice leaves unaffected the aggregate sum of profits as well as the cost-weighted average markup

$$\frac{\sum_{i=1}^{n} \tilde{p}_i q_i}{\sum_{i=1}^{n} c_i q_i} \quad \equiv \quad \frac{\sum_{i=1}^{n} p_i q_i}{\sum_{i=1}^{n} c_i q_i} \quad \equiv \quad \sum_{i=1}^{n} \frac{c_i q_i}{\mathbf{c}'\mathbf{q}} \mu_i \tag{42}$$

which is shown by Edmond et al. (2018) to be the welfare-relevant average markup at the aggregate level. In this basic version of the model, I have constant returns to scale; therefore, the markup is simply equal to the ratio of revenues over operating costs. In Section 7, I extend the model to incorporate nonconstant returns to scale and fixed costs.

Second, this approach hedges my measures of expected profits against aggregate fluctuations in the stock market: this is because, by construction, any change in total stock market value that does not anticipate a shift in next-period profits is exactly offset by $(r + \delta)$. Another benefit of this approach is that it is nearly model-free and does not require any econometric estimation, thus reducing arbitrariness.

In order to map my model to the data, I need firm-level estimates of the output volume q_i. As is the case in other macro models with differentiated products (see for example Hsieh and Klenow, 2009), in this model there is no single objective volume metric of firm-level output, since the firms in my data set come from widely different industries and their outputs are intrinsically not comparable: it is simply not possible to compare, say, laptops to cars.

Fortunately, the welfare properties of the economy do not depend on the firm-level volumetric units used. In other words, when it comes to computing surplus or profits, if the model is correctly written, it should not matter whether we measure the output of any particular firm in kilograms, liters, or number of items. The only thing that should matter from a welfare perspective are (1) profits, (2) markups, and (3) the own and cross-elasticity of demand. It is easy to show that all these measures are invariant to the output volume unit chosen. These facts are consistent with the theoretical results of Baqaee and Farhi (2017) and Hsieh and Klenow (2009), who show that the aggregate costs of misallocation can be written exclusively in terms of the vector of markups or revenue-based productivity, at least to a first-degree approximation.

Because of this invariance, for every cross-section of data, we can therefore pick an arbitrary volumetric unit for each good. The resulting value of q_i automatically pins down the corresponding p_i and c_i .

In this case, my choice for the volumetric system is preconditioned by my assumption that the vector of characteristics space coordinates $\mathbf{a}_i$ has norm one for every firm i. Again, this is a normalization assumption that has no effect on my aggregate measures of welfare. Under this normalization, the model-consistent measure of firm-level output is identified as the square root of profits, that is:

$$q_i \;=\; \sqrt{\pi_i} \;=\; \sqrt{(r + \delta)\, k_i} \tag{43}$$

This identity can be derived directly from equation (21): it is a consequence of the fact that, in the chosen volumetric space, the residual inverse demand function has slope -1 and the marginal cost curve is flat (I relax the latter assumption in Section 7).

[12] De Loecker, Eeckhout, and Unger (2018, henceforth DEU) suggest using COGS as a measure of variable costs, while Traina (2018) suggests operating costs (XOPR). In the basic version of the model, I use Traina's mapping and assume that operating costs are all variable. In Section 7, I propose an alternative version of the model with fixed costs and nonconstant returns to scale in which I map variable costs to COGS and overhead to SG&A, consistently with DEU.

Once $\mathbf{q}$ is obtained, the vector $(\mathbf{A}'\mathbf{b}-\mathbf{c})$ is identified by equation (20):

$$\mathbf{A}'\mathbf{b}-\mathbf{c} \;=\; (\mathbf{I}+\mathbf{A}'\mathbf{A})\,\mathbf{q} \tag{44}$$

It is important to stress that the level q_i is unrelated to actual output volume: it is simply a monotone transform of the firms' expected profits. By simply replacing $\mathbf{q}$ with $\sqrt{\pi}$, we can see that all welfare metrics in this model can be written in terms of the profit vector π and the similarity matrix $\mathbf{A}'\mathbf{A}$. This is also a consequence of volumetric invariance.

3.3. Model Calibration

The hedonic demand system introduced in Section 2 yields an equivalence between the firms' degree of similarity in the product-characteristic space and the matrix of cross-price elasticities of the demand system from equation (12), specifically:

$$\frac{\partial p_i}{\partial q_j} \;=\; \mathbf{a}_i'\mathbf{a}_j \tag{45}$$

The matrix of product similarity of Hoberg and Phillips (2016), which I adopt as the empirical counterpart to $\mathbf{A}'\mathbf{A}$, is interpreted by my model as a measure of substitutability, and is the key data ingredient to compute the cross-price elasticity of demand between all firms in the economy, as well as the aggregate metrics of consumer and producer surplus.

Based on the model presented in Section 2, one could, in theory, directly map HP's cosine similarity scores to the matrix of cross derivatives $\partial\mathbf{p}/\partial\mathbf{q}$. That is, if $\widehat{\mathbf{a}_i'\mathbf{a}_j}$ is HP's product-similarity score for (i,j)—our empirical estimate of $\mathbf{a}_i'\mathbf{a}_j$—we could assume:

$$\mathbf{a}_i'\mathbf{a}_j \;=\; \widehat{\mathbf{a}_i'\mathbf{a}_j} \tag{46}$$

While it is hard to dispute that HP's text-based product-similarity scores accurately identify competitor relationships, the assumption that $\widehat{\mathbf{A}'\mathbf{A}}$ is an unbiased estimate of $\mathbf{A}'\mathbf{A}$ seems nonetheless too strong: from an intuitive standpoint, it requires that the occurrence of the words used to construct HP's similarity scores maps exactly to the hedonic characteristics embedded in the firms' product portfolios. It is relatively easy to come up with scenarios under which $\widehat{\mathbf{a}_i'\mathbf{a}_j}$, as an estimate of $\mathbf{a}_i'\mathbf{a}_j$, could be upward- or downward-biased. For example, one driver of substitutability that is not captured by these similarity scores is geography. If firms compete in geographical segments that do not overlap perfectly, the cosine similarity matrix will provide an upward-biased estimate of the cross-price demand elasticity.

To account for these potential biases, I am going to impose the following, weaker assumption:

$$\mathbf{a}_i'\mathbf{a}_j \;=\; \widehat{\mathbf{a}_i'\mathbf{a}_j}^{\,\lambda} \tag{47}$$

That is, the "true" cosine similarity between firm i and firm j is log-linear in the one measured by Hoberg and Phillips. The newly introduced parameter λ governs the steepness of this relationship. The reason I use a log-linear relationship is that the cosine similarity score is mathematically restricted to the interval $[0,1]$. The log-linear mapping has the desirable property

$$\widehat{\mathbf{a}_i'\mathbf{a}_j}=1 \qquad \Longleftrightarrow \qquad \mathbf{a}_i'\mathbf{a}_j=1 \tag{48}$$

$$\widehat{\mathbf{a}_i'\mathbf{a}_j} = 0 \qquad \Longleftrightarrow \qquad \mathbf{a}_i'\mathbf{a}_j = 0 \tag{49}$$

That is, if two firms use exactly the same set of words to describe their products, then they must be producing products that are perfectly substitutable ($\mathbf{a}_i'\mathbf{a}_j = 1$); if two firms use a completely different set of words to describe their product, they must be producing unrelated goods.

Because all $\mathbf{a}_i'\mathbf{a}_j$ lie on the $[0,1]$ interval, a value of λ above one implies that HP's cosine similarity scores yield upward-biased estimates of the cross-price elasticity matrix. Conversely a value of λ below one implies that HP's cosine similarity scores yield downward-biased estimates of the cross-price elasticity matrix. By introducing and appropriately calibrating the parameter λ, I can account for and correct this bias, which might otherwise affect my aggregate metrics of welfare.

My strategy for calibrating λ is to target the most reliable estimates for the cross-price elasticity of demand. Those are microeconometric estimates from industrial organization (IO) studies. In order to calibrate λ, I obtain, for a number of product pairs, estimates of the cross-price demand elasticity from empirical IO studies that estimate the demand function econometrically. I match these estimates of the cross-price elasticity of demand to the corresponding firm pair in Compustat. Next, for different values of λ, I compare the microeconometric estimate to the corresponding model-based elasticity that is based on HP's cosine similarity data. I calibrate λ to the value that provides the closest fit between the text-based cross-price elasticity and the microeconometric estimates. This procedure yields an estimate of λ of 2.26. To offer the best possible reassurance that the combination of model and text data provides robust estimates of the cross-price elasticity of demand for the whole sample, I also validate the calibrated model using an econometric analysis of the firm's stock-market reactions to patent announcements. The full methodology that I used to obtain and validate my calibrated value of λ is presented in Appendix C.

4. Empirical Findings

4.1. Mapping the Product Space

In my first empirical exercise, I use the Hoberg and Phillips (2016) data set to produce a bidimensional visualization of the product characteristic space for all firms in the CRSP-Compustat universe. To do this, I have to drastically reduce the dimensionality of the data: each firm exists in a space of characteristics that has as many dimensions as there are words in the vocabulary that HP used to create their similarity data set ($\sim 61,000$).

To create a bidimensional visualization of the product space, I use the algorithm of Fruchterman and Reingold (1991, henceforth FR), which is widely used in network science to visualize weighted networks. The algorithm models the network edges as a gravitational field, letting the nodes dynamically arrange themselves on a bidimensional surface as if they were particles subject to attractive and repulsive forces.[13]

The result of this exercise is shown in Figure 3: every single dot in the graph is a publicly traded firm as of 2004. Firm pairs that have a high cosine similarity score appear closer; they are also joined by a thicker line. Conversely, firms that are more dissimilar are not joined, and tend to be more distant. The product space

[13] One known shortcoming of this algorithm is that it is sensitive to the initial configurations of the nodes, and it can have a hard time uncovering the cluster structure of large networks. To mitigate this problem, and to make sure that the cluster structure of the network is displayed correctly, before running FR I prearrange the nodes using a different algorithm, OpenOrd, (Martin et al., 2011) which was explicitly developed for this purpose.

is manifestly uneven: some areas are significantly more densely populated with firms than others. Also, the network displays a pronounced community structure: large groups of firms tend to cluster in certain areas of the network.

Because more densely populated areas are likely to be associated with lower markups, this network configuration can impact allocative efficiency. The theoretical model presented in Section 2 leverages this network structure to capture heterogeneity in competition and market power across firms.

One could argue that perhaps this visualization might be an artifact of dimensionality reduction and/or of measurement error, and as such, uninformative. In Appendix D, I show that this is not the case: notwithstanding the dimensionality reduction, a remarkable degree of overlap exists between the macro-clusters of this network and broad economic sectors. This allows me to independently validate the product similarity network data of Hoberg and Phillips (2016).

4.2. Concentration, Centrality and Markups

One of the key testable predictions of my model is that having few substitute products—as measured by the Inverse Centrality Index—allows a firm to charge a high markup. To see why this is so, consider the special case in which all firms are identical except for their network position. The firm-level markup simplifies to the following expression:

$$\mu \;=\; \mathbf{1} + \left(\mathbf{I} + \mathbf{A}'\mathbf{A}\right)^{-1} \mathbf{1} \tag{50}$$

this expression can be rewritten in scalar form in terms of the firm-level centrality score:

$$\mu_i \;=\; 1 + \frac{\chi_i}{1+\chi_i} \tag{51}$$

I have thus obtained a “predicted” markup for each firm—specifically, that which is implied by the firm’s centrality in the product network. Two major differences exist between my predicted markup and that computed by Traina (2018) and De Loecker et al. (2018). First, their measures of markups are “supply-side” markups, and are estimated using cost-minimization assumptions, while my predicted markup is a “demand-side” markup, in the sense that is micro-founded on my specific assumptions about the demand system faced by the firms. Second, while the supply-side markups are computed using firm financials, my predicted markup is computed using nothing but the 10-K data.

To empirically validate my model, I use regression analysis to verify whether inverse centrality (the demand-side markup) predicts supply-side markups computed from balance-sheet data using the techniques used in Traina (2018) and De Loecker et al. (2018). This is arguably the strongest, most direct test of the empirical validity of my theoretical model. The reason is that the distribution of markups is a key input in the measurement of deadweight losses and there is no overlap in the data used to compute the left-hand-side variable (the supply-side markup) and the right-hand-side variable (the inverse centrality).

I start by computing, for all firm-years in my sample (1997–2017), the two measures of supply-side markups. The first, revenues divided by operating costs, is consistent with the model’s definition of unit costs as well as Traina (2018)’s methodology. The second, revenues divided by the cost of goods sold, is consistent with the methodology of De Loecker et al. (2018). Both these measures are ex-post measures, in the sense that firms’ actual realized revenues are used in the computation.

I regress both of these measures of markups, in logs, on the Inverse Centrality Index, controlling for the

NAICS sector classification (at the 2-, 3-, 4-, and 6-digit level), for year fixed effects. The main explanatory variable—inverse centrality—is standardized to simplify the interpretation of the regression coefficients. Because the linear demand assumption implies that larger firms should charge higher markups, I add to this specification a firm size control: the log of the book value of the total assets; I use the book value as opposed to the market value because the latter is likely to incorporate future expectations of profitability, and is therefore likely to suffer from reverse causality.

My measures of markups differ from those of De Loecker (2011) and Traina (2018) by a scalar multiplier that is equal to the elasticity of output with respect to the variable input, and which varies by sector.[14] This scalar is usually estimated econometrically. However, because I run the regressions in logs, this sector-level scalar θ_s is absorbed by fixed effects:

$$\mu_i = \theta_s \frac{\text{Revenues}_i}{\text{Variable Costs}_i} \quad \Longrightarrow \quad \log \mu_i = \log \theta_s + \log \frac{\text{Revenues}_i}{\text{Variable Costs}_i} \tag{52}$$

Consequently, I run my regressions by simply using the log of revenues over costs as the left-hand-side variable. By construction, this specification yields identical estimates, as if I were using the De Loecker/Traina markups. At the same time, it is also transparent and easily replicable with available data.

The results of this regression analysis are shown in Table 2. Panel A displays regression results for markups based on operating costs, while Panel B displays regression results for markups based on COGS. Standard errors are clustered at the firm level. Each panel has six specifications, each characterized by an increasingly fine level of sector controls from left to right. The regression results show that both size and (more important) less centrality in the product-similarity network reliably predict higher markups. In all but one specification, the reaction coefficients for both variables are statistically different from zero at the 1% confidence level.

The economic magnitude of these coefficients is significant as well. In Panel A, the magnitude of the coefficient on the *Inverse Centrality* ranges from 0.04 and 0.12, implying that a standard-deviation increase in this variable is associated with an 4% to 12.5% higher markup; the coefficient on *log Assets* implies that a doubling of the book value of assets yields an effect of comparable magnitude: the predicted markup is 11% to 12% higher. More important, these coefficients remain remarkably stable as we add increasingly fine sectoral controls.

The coefficients from the regression in Panel B imply that a standard-deviation increase in *Inverse Centrality* is associated with 1% to 5% higher COGS-based markups; a doubling in the book value of assets results in a 6.5% higher COGS-based predicted markup. The most likely explanation for the divergence in the magnitudes of the coefficients estimated in the two panels is that COGS-based markups are significantly more volatile than XOPR-based markups, perhaps reflecting the fact that operating costs might be less amenable to accounting manipulations than COGS, as they are more tightly linked to the companies' cashflows. This view is also consistent with the lower R^2 in the Panel B regressions.

While these regression results are uninformative with respect to how firms ultimately attain a less-central position in the network, they do show that measures of centrality based on HP's product-similarity data (unlike the original HHI) do actually predict markups: this finding is consistent with the interpretation that firm-level variations in exposure to competition translate into measurable differences in pricing power,

[14] See De Loecker (2011) for the detailed methodology.

provided that we actually take into account product similarity.

5. Counterfactual Analyses and Policy Experiments

5.1. Basic Counterfactuals

In the next few paragraphs, I define four counterfactual allocations of interest. Later in the section, I present my computations of these counterfactuals based on firm-level data covering the universe of public firms in the United States.

The first counterfactual that I compute is the first-best allocation, or *social optimum*, which is presented in equation (53): it is the equivalent of an equilibrium in which all firms are acting as atomistic producers, and pricing all units sold at marginal cost.[15]

Definition 5 (First-best allocation)**.** The first-best allocation is defined as that which maximizes the aggregate total surplus function $W(\mathbf{q})$

$$\mathbf{q}^W \quad \stackrel{\text{def}}{=} \quad \arg\max_{\mathbf{q}} W(\mathbf{q}) \quad = \quad (\mathbf{A}'\mathbf{A})^{-1}(\mathbf{A}'\mathbf{b}-\mathbf{c}) \tag{53}$$

The second counterfactual that I consider is what I call the *Monopoly* allocation: the allocation that maximizes the aggregate profit function and represents a situation in which one agent has control over all the firms in the economy. The *Monopoly* counterfactual is interesting; it allows us to explore the welfare consequences of letting corporate consolidation continue until its natural limit: that is, the point where one monopolist owns all the firms.

Definition 6 (Monopoly allocation)**.** The *Monopoly* allocation is defined as that which maximizes the aggregate profit function $\Pi(\mathbf{q})$:

$$\mathbf{q}^\Pi \quad \stackrel{\text{def}}{=} \quad \arg\max_{\mathbf{q}} \Pi(\mathbf{q}) \quad = \quad \frac{1}{2}(\mathbf{A}'\mathbf{A})^{-1}(\mathbf{A}'\mathbf{b}-\mathbf{c}) \tag{54}$$

The Monopoly allocation is simply the first-best scaled by one-half. This is because, similarly to the social planner, and unlike the pseudo-planner (who delivers the Nash-Cournot equilibrium), the Monopolist can coordinate production across units. However, he does not internalize the consumer surplus; therefore, he has an incentive to restrict overall output.

Another highly policy-relevant allocation, one in line with the spirit of the original Harberger (1954) paper, is one in which resources are allocated *efficiently*—that is, in which the social planner maximizes the aggregate surplus function subject to the constraint of using no more input than in the observed equilibrium. Both Harberger and Hsieh and Klenow (2009, henceforth HK) consider a counterfactual in which some (indirect) measure of markups is equalized across productive units. The earlier paper focuses on differences in rates of return on capital, while the latter focuses on differences in revenue productivity (TFPR) across

[15] The "first-best" in my model is different from that of Baqaee and Farhi (2017), in the sense that I remove only variation in markups that is captured by firm size and position in the network. The reason is, my model is built to isolate market power related to oligopoly/concentration.

firms[16]; yet, the basic intuition is the same. The equalization of markups implies allocative efficiency because it occurs for allocations where the marginal total surplus added by every productive unit is the same, which is what a benevolent social planner would seek to attain in order to maximize welfare. The following result allows me to pin down the constant markup for all firms that leaves total input usage H unchanged.

Definition 7. I define the resource-efficient (or *misallocation-free*) counterfactual $\mathbf{q}^H$ as the solution to the following constrained maximization problem:

$$\mathbf{q}^H = \arg\max_{\mathbf{q}} W(\mathbf{q}) \quad \text{s.t.} \quad H(\mathbf{q}) = H \tag{55}$$

Proposition 3. *The resource-efficient counterfactual takes the form:*

$$\mathbf{q}^H = \left(\mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b} - \mu\mathbf{c}\right) \tag{56}$$

where μ is the markup charged by all firms:

$$\mu = \frac{\mathbf{c}'\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\mathbf{b} - H}{\mathbf{c}'\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{c}} \tag{57}$$

Proof. See Appendix R. □

Because this counterfactual uses the same number of inputs as the observed equilibrium, by comparing welfare in this allocation to the first-best we can effectively disentangle the welfare costs of monopoly into two components: misallocation and factor-suppression.

5.2. Merger Simulations

An important tool of antitrust policy is the review of proposed mergers. In the United States, merger review is carried out jointly by the Federal Trade Commission and the Department of Justice. The Hart-Scott-Rodino (HSR) Act mandates that merging companies (subject to a certain size threshold) must notify the relevant agencies before proceeding with the merger. The FTC and the DOJ then review proposed mergers that are considered at risk of violating antitrust provisions. When a merger is deemed to violate U.S. antitrust laws, the FTC or the DOJ can attempt to reach a settlement with the involved parties or they can sue them in a court of law to prevent the merger. *Merger simulation* is used by antitrust enforcement agencies to predict the effects of mergers on competition, to determine whether a certain merger is in violation of U.S. antitrust legislation.

The model presented in Section 2 allows me to simulate the aggregate welfare effects of mergers by appropriately modifying the potential function $\Phi(\cdot)$. The comparative advantage of this model when it comes to merger simulation, relative to the IO and antitrust literature, is that it allows me to analyze the general equilibrium effects of multiple mergers, both within and across industries. For specific transactions, my model would provide comparatively less accurate estimates relative to the best practices in that literature, which model more complex strategic interactions.

[16] Pellegrino and Zheng (2017) show that, in HK, TFPR is simply a noisy proxy for markups. The only source of interfirm variation in TFPR is markups, and the only sources of interfirm variation in markups are the unobservable policy distortions that are the focus of HK's analysis (by assumption).

We have already considered the extreme case of the *Monopoly* counterfactual, where all firms are merged at once. I now focus on mergers between specific firms, modeling mergers as *coordinated pricing* (similar to Bimpikis et al., 2019). That is, following the merger, the parent firms do not disappear; instead, a single player chooses the output of both firms to maximize the joint profits.

Proposition 4. *Consider a merger among a set of companies that are localized by the coordinate matrix* $\mathbf{A}_1$*, such that the matrix* $\mathbf{A}$ *is partitioned as follows:*

$$\mathbf{A} \;=\; \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \tag{58}$$

The new equilibrium allocation maximizes the following modified potential:

$$\Psi\left(\mathbf{q}\right) \;=\; \mathbf{q}'\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)-\mathbf{q}'\mathbf{q}-\frac{1}{2}\mathbf{q}'\begin{bmatrix} 2\left(\mathbf{A}_1'\mathbf{A}_1-\mathbf{I}\right) & \left(\mathbf{A}_2'\mathbf{A}_1-\mathbf{I}\right) \\ \left(\mathbf{A}_2'\mathbf{A}_1-\mathbf{I}\right) & \left(\mathbf{A}_2'\mathbf{A}_2-\mathbf{I}\right) \end{bmatrix}\mathbf{q} \tag{59}$$

Proof. See Appendix R. □

The maximizer of $\Psi\left(\mathbf{q}\right)$, which corresponds to the postmerger equilibrium allocation, is:

$$\mathbf{q}^{\Psi} = \left(2\mathbf{I}+\begin{bmatrix} 2\left(\mathbf{A}_1'\mathbf{A}_1-\mathbf{I}\right) & \left(\mathbf{A}_2'\mathbf{A}_1-\mathbf{I}\right) \\ \left(\mathbf{A}_2'\mathbf{A}_1-\mathbf{I}\right) & \left(\mathbf{A}_2'\mathbf{A}_2-\mathbf{I}\right) \end{bmatrix}\right)^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right) \tag{60}$$

That is, to simulate the new equilibrium following a merger between existing firms, I need only to amend the potential function by doubling the off-diagonal quadratic terms corresponding to the merging firms.

When all firms are merged, $\Psi\left(\mathbf{q}\right)$ simply becomes the aggregate profit function $\Pi\left(\mathbf{q}\right)$, and the equilibrium allocation simply converges to the Monopoly counterfactual (equation 54).

5.3. Breakup Simulations and the First Welfare Theorem

The ability to break up a company is another important tool of antitrust policy. Although it is seldom used nowadays, it was famously applied in the case of Standard Oil Company (in 1911) and the Bell System (in 1984). Similar to merger cases, for this to happen the Federal Trade Commission or the Department of Justice need to sue the company that is holding the alleged monopoly.

To simulate the breaking up of a trust, I need to make an assumption about the resulting companies' coordinates in the product characteristics space. Because this information is generally unavailable, a neutral assumption that the firm is broken up into N companies that are all identical to the initial one (same coordinate vector $\mathbf{a}$, same unit cost $\mathbf{c}$). Without loss of generality, consider breaking up the company corresponding to the first component of the column vector $\mathbf{q}$.

Proposition 5. *Following the breaking up of company 1 into* N *separate entities whose product offering is identical to that of the parent company, the new equilibrium allocation maximizes the following modified potential:*

$$\Upsilon\left(\mathbf{q}\right) \;=\; \mathbf{q}'\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)-\mathbf{q}'\begin{bmatrix} \frac{1+N}{2N} & 0 \\ 0 & \mathbf{I} \end{bmatrix}\mathbf{q}-\frac{1}{2}\mathbf{q}'\left(\mathbf{A}'\mathbf{A}-\mathbf{I}\right)\mathbf{q} \tag{61}$$

Proof. See Appendix R. $\square$

The maximizer of $\Upsilon(\mathbf{q})$, which corresponds to the postbreakup equilibrium allocation, is

$$\mathbf{q}^{\Upsilon}=\left(\begin{bmatrix}\frac{1}{N} & 0\\ 0 & \mathbf{I}\end{bmatrix}+\mathbf{A}'\mathbf{A}\right)^{-1}(\mathbf{A}'\mathbf{b}-\mathbf{c}) \tag{62}$$

That is, to simulate the new equilibrium following the breaking up of a trust, I only need to amend the potential function by multiplying by $(1+N)/2N$ the diagonal quadratic terms corresponding to the firms being split. By doing this for all firms and taking the limit as N goes to infinite, I can theoretically implement the first-best allocation (equation 53).

Theorem 1 (Asymptotic First Welfare Theorem)**.** *As all firms are broken up into an arbitrary number of independent competitors, the total surplus of the economy $W(\mathbf{q})$ converges to the first-best $W\left(\mathbf{q}^{W}\right)$.*

Proof. To break down every firm i into N_i separate entities, as shown above, I multiply each diagonal element of the quadratic term of the potential function by $(1+N_i)/2N_i$. I then take the limit for all $N_i\to\infty$. Thus, $\Upsilon(\mathbf{q})\to W(\mathbf{q})$ and $\mathbf{q}^{\Upsilon}\to\mathbf{q}^{W}$. $\square$

5.4. Counterfactual Welfare Calculations

I now report my model-based estimates of the aggregate total surplus and consumer surplus, together with a set of counterfactual calculations based on the scenarios described in Subsection 5.1. The estimates are all shown in Table 3.

The first column presents estimates for the "actual" surplus, which is based on the assumption that firms compete as posited in the model of Section 2. The (publicly traded) firms in my sample earn an aggregate operating profit of \$2.8 trillion and produce an estimated total surplus of \$6.5 trillion. For context, the gross operating surplus of U.S. corporations (the aggregate equivalent of operating profits) in the same year (2017) is \$4.7 trillion. Consumer surplus is therefore estimated to be about \$3.7 trillion. About half of the total surplus produced is appropriated by firms in the form of oligopoly profits. The first counterfactual I consider, the first-best allocation, appears in the second column, Again, this is an allocation in which all firms behave as if markets were perfectly competitive (firms price at marginal cost). In this allocation, aggregate surplus is significantly higher than in the Network Cournot equilibrium allocation: \$7.5 trillion. Because firms price at marginal cost, all of the surplus goes entirely to the consumer and firms make no profits. The percentage welfare increase with respect to the Network Cournot equilibrium is significant: 13.3%. Yet, the relative increase in consumer surplus is even higher: it more than doubles, as the consumer earns the entire surplus.

The next counterfactual I analyze, the *Monopoly* counterfactual, appears in the third column: it represents a scenario in which all firms are controlled by a single decision-maker. In this allocation, aggregate surplus is significantly lower than in the Network Cournot equilibrium allocation: \$5.7 trillion. Despite the decrease in aggregate welfare, profits are significantly higher: \$3.8 trillion. Consequently, consumer surplus is reduced to just \$1.9 trillion. In this scenario, the consumer earns the smallest share of total surplus she can hope to attain: 33%.

The last scenario I consider, the no-misallocation counterfactual, depicts a scenario in which the social planner maximizes total surplus subject to not changing overall labor usage. In this scenario, markups across firms have been equalized, but not eliminated. By removing all dispersion in markups, this counterfactual targets the malallocative effects of concentration.

The total surplus produced in this last scenario is \$7.3 trillion, about 10% higher than the observed Cournot-Nash equilibrium. Most of the surplus produced (\$5.8 trillion) goes to the consumer, and profits are reduced to \$896 billion. The main takeaway from the Harberger counterfactual is that most of the welfare-reducing effects of market power appear to come from resource misallocation. In other words, it is the dispersion of markups (rather than the level) that appears to have the most significant impact on aggregate welfare.

One important caveat of this analysis: these welfare calculations are relevant only in the short term, as they only capture the intensive margin of competition across firms. Results are likely to vary substantially if entry and exit are allowed to occur. To construct a long-run counterfactual, a dynamic model of the network $\mathbf{A}'\mathbf{A}$ is needed.

5.5. Time Trends in Total Surplus and Consumer Surplus

The data used for this paper, sourced from HP's data set and CRSP-Compustat, is available as far back as 1997. By mapping my model to firm-level data, year by year, I can produce annual estimates of aggregate surplus, its breakdown into profits and consumer surplus, and the welfare costs of oligopoly. This allows me to interpret secular trends in markups and concentration in a general equilibrium setting.

In Figure 5, I plot aggregate total surplus $W(\mathbf{q})$, consumer surplus $S(\mathbf{q})$, and profits $\Pi(\mathbf{q})$ for every year between 1997 and 2017. I also plot the profit share of surplus $\Pi(\mathbf{q})/W(\mathbf{q})$. The graph shows that the aggregate total surplus produced by American public corporations has increased from about \$4.6 trillion to about \$7.5 trillion. Profits have increased disproportionately more, from about \$1.3 trillion to about \$2.8 trillion. As a consequence, the profit share of surplus has increased from about 32.3% of total surplus to nearly 43.4%.

In Figure 6, I plot, over the same period, the percentage gain in total surplus and consumer surplus resulting from moving from the competitive equilibrium $\mathbf{q}^{\Phi}$ to the first best $\mathbf{q}^{W}$. These can be seen as a measure of the aggregate deadweight loss. Both series experienced upward trends that mimic that of profit share of surplus: the total surplus gains have increased from about 10% (in 1997) to about 13.3% (in 2017). The consumer surplus gains, which are generally much larger in magnitude, have increased from 39.8% to 50.9%.

These findings are consistent with the interpretation that the increasing concentration in U.S. industries over the past few decades reflect a generalized increase in oligopoly power that has negatively affected total surplus consumer welfare.

6. The Rise of M&A and the Decline of IPOs

6.1. Effects on Concentration and Aggregate Welfare

The empirical analysis of Section 4 suggests that—consistent with secular trends in profit rates and concentration—the deadweight loss from oligopoly and the profit share of surplus have indeed increased in the United States. In this section, I use the counterfactual-building capabilities of my model to offer an explanation for these secular trends.

My hypothesis is motivated by two known facts about U.S. public equity markets. First, the number of public companies in the United States has decreased dramatically in the past 20 years, from about 7,500 in 1997 to about 3,500 in 2017 (Kahle and Stulz, 2017). Second, while in 1995 the vast majority of successful VC exits were IPOs, today virtually all of them are acquisitions (Gao et al., 2013). In addition, the finance literature has recently started documenting so-called "killer acquisitions" (Cunningham et al., 2018)—where the acquirer purchases a startup with a clear intent of obstructing the development of new products that might pose a competitive threat to its existing product portfolio.

In what follows, I show that the secular decline in IPOs and the surge in acquisitions of VC-backed startups can quantitatively account for the dramatic decline in the number of public corporations in the United States, the increasing profit share of surplus, and the rising welfare costs of oligopoly measured in Section 4. While this is not the only plausible explanation for the observed trends, this effect can be investigated through the lens of the model of Section 2, it works quantitatively, and it can be rationalized by known changes in the regulatory and technological environment.

I start by collecting and visualizing data on VC exits. There are multiple data sources for VC transactions. They differ in terms of coverage across firms and time. Yet, the data shows, regardless of the source, that IPO exits have been declining while M&A exits have surged since the mid-'90s.

Figure 7 displays the number of successful VC exits in the United States by year and type for two different data sources. Panel A shows data from Dow Jones VentureSource for 1996 to 2017. While the shift is noticeable from this first graph, it becomes even more dramatic when looking back to the 1980s. The graph in Panel B, which displays data from the National Venture Capital Association (NVCA) for 1985 to 2017, shows that, before the 1990s, virtually all VC exits were IPOs.

The next question I ask is how much of the decline in the number of firms in the merged Compustat-HP data set observed in the past twenty years can be accounted for by the decline in IPOs. I answer this question by computing a counterfactual in which the rate of IPOs, as a percentage of successful exits, would have stayed constant after 1996. To compute this counterfactual, I need to make an assumption about the probability of each additional entrant "surviving" from one year to another. I assume that these additional entrants would have disappeared from Compustat at the same rate as the firms actually observed in Compustat.

The result of this exercise is shown in Figure 8. It can be seen from the picture that, under the counterfactual scenario, the number of public firms would have increased slightly over the period 1997–2017, rather declining by nearly half.

Finally, I ask to what extent this decrease in the number of public firms affected competition and welfare dynamics in the United States, and to what extent it can explain the uptrends in deadweight losses and the profit share of surplus I measured in Section 4. To this end, I cannot just add these firms back to the Compustat sample: because they were acquired, this would equate to effectively double-counting them. Instead, what I need to perform is a "break-up," using equation (59). Moreover, one would presume that an assumption needs to be made about where these "missing startups" would enter in the product space. In other words, I need to have an estimate of the cosine similarity between these startups and the firms observed in Compustat.

Fortunately, no assumption of this sort needs to be made, as this similarity score was already computed in Hoberg, Phillips, and Prabhala (2014), who kindly shared their data, allowing me to compute this specific counterfactual. The score $\mathbf{a}_i'\mathbf{a}_0$ provides, for every year, the cosine similarity score between all firms i from Compustat and the VC-backed firms from the VenturExpert data set that were financed in that year.

Let n_0 be the number of additional firms that would be active in the "constant IPO" scenario. The counterfactual I study here consists in breaking up the n firms in Compustat into $n + n_0$ firms. Each firm gets broken into $(1 + n_{i0}/n)$ firms, where n_{i0} is proportional to the similarity score $(\mathbf{a}_i'\mathbf{a}_0)$ between i and the VenturExpert firms. Formally:

$$N_i = \frac{n + n_{i0}}{n} \qquad \text{where} \qquad n_{i0} = \frac{\mathbf{a}_i'\mathbf{a}_0}{\frac{1}{n}\sum_i \mathbf{a}_i'\mathbf{a}_0} \cdot n_0 \tag{63}$$

this implies that

$$\sum_i N_i = n + n_0 \tag{64}$$

The potential function for the "constant IPO" scenario is therefore

$$\Upsilon(\mathbf{q}) \;=\; \mathbf{q}'(\mathbf{A}'\mathbf{b} - \mathbf{c}) - \mathbf{q}' \begin{bmatrix} \frac{1+N_1}{2N_1} & 0 & \cdots & 0 \\ 0 & \frac{1+N_2}{2N_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1+N_n}{2N_n} \end{bmatrix} \mathbf{q} - \frac{1}{2}\mathbf{q}'(\mathbf{A}'\mathbf{A} - \mathbf{I})\mathbf{q} \tag{65}$$

In other words, we are breaking up the n firms in Compustat into $n + n_0$ entities, yet more of the "breaking up" is loaded on firms that directly compete with VC-backed startups.

The counterfactual surplus functions from this scenario are shown in Figure 9: it shows that, under this alternative scenario, aggregate total surplus would be larger by about $410 billion. Even more important, a significantly larger share of surplus gained between 1997 and 2017 would have been accrued to the consumer: instead of increasing from 33.2% to 43.4%, the profit share would have increased to just 34.2%.

The deadweight losses due to oligopoly show a similar dynamic under the "constant IPO rate" scenario. Instead of increasing from 10% to 13.3%, the deadweight loss from oligopoly would have decreased to about 7.9%; instead of rising from 39.8% to 50.9%, the loss in aggregate consumer surplus would have decreased to about 39.4%.

One potential concern with this counterfactual exercise is that these trends in the aggregate welfare metrics might be biased by measurement error in the similarity score for VC-backed startups $(\mathbf{a}_i'\mathbf{a}_0)$. I rule out this possibility by computing an alternative version of the same scenario in which, instead of using the similarity score for VC-backed startups, I impose

$$n_{i0} = n_0 \qquad \forall i \tag{66}$$

Recomputing my estimates using this alternative methodology yields nearly identical estimates.

While, again, this set of counterfactual exercises is no hard proof that the decline in IPOs and the rise of takeovers of VC-backed startups are the root cause of the observed trends in market power and concentration, it establishes that this is indeed a possibility that researchers and policy analysts need to consider seriously, not least because it can *also* account for the secular decline in the number of public companies in the United States.

6.2. Startup Takeovers and Concentration: the "Big Five Tech"

As the public debate over concentration and market power has heated up, the five largest American technology companies—Alphabet (Google), Amazon, Apple, Facebook, and Microsoft—have been regularly mentioned as a case-in-point of the rise of oligopoly power.[17]

While obviously sheer size plays a role, these companies stand out, even among the largest corporations, for another reason: their notoriously active involvement in the market for corporate control of VC-backed startups. Data collected by the website TechTakeovers show that these five companies alone acquired over 600 startups from 1996 to 2017. This compares with 10,260 acquisition of startups recorded in the VenturExpert database, and a pool of about 5,000 companies per year in Compustat over the same period. A simple back-of-the-envelope calculation suggests that the Big Five Tech may account for as many as 6% of all the acquisitions of startups, and that they might be acquiring startups at a rate 60 times that of the average company in Compustat.

This observation suggests another, more qualitative test of the link between product-market competition and startup takeovers. If indeed the increased oligopolization of U.S. industries has been accelerated by the increased rate of startups being taken over, we should observe a noticeable increase in the Inverse Centrality score of the Big Five with respect to their peer group. In Figure 11, I plot the relative increase in the Inverse Centrality for three groups of companies—the Big Five Tech, the other Tech companies, and the rest of Compustat—for the period following the dot-com bubble. The plot shows a dramatic increase in product-market isolation in favor accruing to the Big Five—not just in comparison to the rest of Compustat, but also in comparison to the Tech peer group. The magnitude of the increase (0.4 points) is particularly striking, due to the fact that the normal range of the Inverse Centrality Index is 0 to 1.

While this additional analysis does not conclusively prove that the startup takeovers are the chief reason why U.S. industries have become more concentrated, it provides additional evidence that they might have been a significant contributor to these trends. It also shows that there might be a substantive, nonideological rationale for why the Big Five have undergone significant scrutiny in recent times. One potential objection to the findings thus far presented is that large tech firms might be engaging in acquisitions of startups with the objective of gaining ownership of valuable technology. In Appendix P, I show additional evidence from 10-K data that suggests that the surge in takeovers of startups was not associated with increased product market innovation.

6.3. Potential Causes of the Collapse of the IPO Market

The previous counterfactual implicitly treats the rate of IPOs as an exogenous variable. Yet, being acquired rather than going public is a choice; in other words, the IPO rate itself is an endogenous variable. In this subsection, I discuss some potential factors that might be driving the secular shift from IPOs to takeovers.

While this decline has an effect on the intensity of product-market competition, there is no reason to assume that the shift from IPOs to takeovers is itself the result of a friction at play. Indeed, the most simple explanation for this shift is that private equity markets have become more efficient. Two trends in IPOs seem to point to this explanation: (1) the average and median size of the IPOs has increased dramatically

[17] See, for example, the Economist (Jan 18th 2018); the Report of the Stigler Center Subcommittee for the Study of Digital Platforms; as well as Dolata (2017) and Moore and Tambini (2018).

since the early 1990s; and (2) the vast majority of firms going public today make losses; this the opposite of what was the case in the early 1990s. In Appendix I, I present these statistics and propose a simple Roy-style model that endogenizes the startup's decision to go public vis-à-vis getting acquired. I show that the observed IPO trends are consistent with this selection model.

To conclude this section, I discuss other proposed explanations for the plunge in the IPO rate. One such explanation, is provided by Gao et al. (2013), who suggest that the importance of bringing new products to the customer quickly has increased over this period: this, in turn, has reduced the incentives of new entrants to challenge existing competitors and increased the expected gains from selling to a strategic acquirer. A second, related explanation is offered by Bowen et al. (2018): they use natural-language processing of patent data to quantify the degree of disruptiveness and technological breadth of every patent. They show that both of these variables are associated, at the firm level, with a greater probability of going public vis-à-vis getting acquired, and that both variables have plummeted since the early nineties—this in turn is driving down the the rate of IPOs and contributing to the rise in M&A. Finally, a third explanation has been offered by Ewens and Farre-Mensa (2018), who provide evidence that the National Securities Markets Improvement Act of 1996 has increased the ability of firms to raise private equity capital, and this in turn has increased their incentive to delay going public.

7. Robustness, Extensions and Additional Remarks

7.1. Extending the Model to Fixed Costs and Nonconstant Returns to Scale

The model can be easily expanded to accommodate nonconstant returns to scale as well as fixed costs. As long as the total cost function is quadratic, firms still play a linear quadratic game over a weighted network with a closed-form equilibrium. In what follows, I illustrate this generalization of the model, and propose an alternative mapping to the data that is consistent with this modification, as well as with the empirical framework of De Loecker et al. (2018).

Suppose that the firms' *average* variable cost function is now:

$$\bar{\mathbf{c}}\left(\mathbf{q}\right) \ = \ \mathbf{c} + \beta\mathbf{q} \tag{67}$$

The coefficient β captures the scale elasticity. In turn, the *marginal* cost function is:

$$\tilde{\mathbf{c}}\left(\mathbf{q}\right) \ = \ \mathbf{c} + 2\beta\mathbf{q} \tag{68}$$

Also, suppose that firms that produce a positive output quantity pay an additional (firm-specific) fixed cost $\mathbf{f}$. By plugging the new cost function inside the potential $\Phi\left(\mathbf{q}\right)$, we obtain the following new equilibrium relationships:

$$\mathbf{p} - \tilde{\mathbf{c}} \ = \ \mathbf{q} \tag{69}$$

as well as:

$$\mathbf{p} - \bar{\mathbf{c}} \ = \ \left(1 + \beta\right)\mathbf{q} \tag{70}$$

Then, conditional to an estimate of β, the real output output vector $\mathbf{q}$ can be identified using the following amended formula:

$$q_i = \sqrt{\frac{\pi_i + f_i}{1+\beta}} \tag{71}$$

To take this extended model to the data, we need to calibrate the scale-elasticity parameter β. This parameter should ideally be allowed to vary through time—as is the case for the demand elasticities—to accommodate the superstar/mega firm effect (see next subsection). To obtain a time-varying estimate of the parameter β, I use the following relationship, which can be derived by combining equations (69) and (70):

$$1+\beta = \left(\frac{\bar{\mu}_i - 1}{\bar{\mu}_i}\right) \Big/ \left(\frac{\mu_i - 1}{\mu_i}\right) \qquad \forall i \tag{72}$$

where $\overline{\mu}_i$ is markup over the average cost and μ_i is the (properly-defined) markup over marginal cost:

$$\overline{\mu}_i \stackrel{\text{def}}{=} \frac{p_i}{\bar{c}_i}; \qquad \mu_i \stackrel{\text{def}}{=} \frac{p_i}{\tilde{c}_i} \tag{73}$$

that is, the parameter β, if different from zero, drives a wedge between these two markups. While marginal costs are not be observed, I can obtain a measure of the average marginal cost (weighted by revenues) from De Loecker, Eeckhout, and Unger (2018, henceforth DEU), who analyze the same universe of companies. The authors compute markups over the marginal cost for every firm, by estimating a production function at the sector level with a single variable output. By estimating the output-input elasticity for this input, they can recover the firm-level markup over the marginal cost.

Conditional on a value of β, my model produces a distribution of markups that will generally produce a different average markup from that of DEU: this difference, as noted above, is primarily driven by the scale elasticity. This allows me to calibrate, for every year between 1997 and 2017, a time-varying β that exactly matches DEU's sales-weighted markup . That is, for every t between 1997 and 2017, I define β_t as:

$$\beta_t \quad : \quad \sum_i \frac{p_{it} q_{it}}{\mathbf{p}_t' \mathbf{q}_t} \cdot \mu_{it}\left(\beta_t\right) = \mu_t^{\text{DEU}} \tag{74}$$

where the μ_t^{DEU} the average markup computed by DEU, and the tilde ($\tilde{\cdot}$) sign denotes that we are using the realized (accounting) revenues to weight.[18]

To implement this alternative model, I modify the mapping of the model to match that used by DEU, who exploit the breakdown of operating costs in Compustat into two buckets: COGS and SG&A. Following DEU, I map variable costs to COGS and fixed costs to SG&A. Revenues are mapped directly to the corresponding Compustat figure. This alternative mapping, as well as the welfare calculations from this alternative model are presented in Appendix J: both the levels as well as the trends mimic closely those found in Section 4: the increased concentration is leading to larger welfare losses (8.6% of total surplus in 1997, 12.5% in 2017) and a lower consumer share of total surplus (28% in 1997, 35% in 2017). As is the case in the baseline model, the fall in the rate of IPO appears to account for the entirety of these trends. Overall, the results appear to be robust to the modeling of fixed costs and nonconstant returns to scale.

[18]My cost function is different from that of DEU. In particular, they assume a Cobb-Douglas/Translog production function. My model can match DEU's average level of markups but not the firm-specific markup.

7.2. Consolidation is Unlikely to Be Driven by Increasing Product Market Overlap or Changing Returns to Scale

Industry consolidation is not per se indicative of a decrease in product-market competition. Recent research has put forward a different explanation for the rising concentration, namely the emergence, over the past two decades, of *superstar* or *mega*firms (Autor et al., 2017; Hall, 2018). The narrative is that a structural transformation has occurred in the macroeconomy that has lead to the reallocation of market share from less productive to more productive producers, more exits, and thus more concentration.

This channel is also present in the model presented in Section 2, and it is most easily seen at work in the endogenous entry/symmetric version previously analyzed in Subsection 2.3. With endogenous entry, the equilibrium number of firms is given by:

$$n = 1 + \frac{2}{a} \cdot \frac{1-f}{f} \tag{75}$$

An increase in product substitutability reduces the number of firms in equilibrium, resulting in an increase in concentration. Because the model is symmetric, there is no reallocation along the intensive margin—yet the basic mechanism is the same as that of Autor et al. (2017). It is therefore natural to ask whether industry consolidation is somehow a *result* of an increase in product-market overlap. The data set of Hoberg and Phillips (2016) allows me to test this channel directly.

To answer this question, I study the evolution of the product-similarity matrix $\mathbf{A}'\mathbf{A}$ over the period under consideration. To the extent that increased competition is responsible for industry consolidation, we should observe an increase in the average product similarity over the same period. The results of this analysis are shown in Appendix M. After controlling for survivor bias, I find that, far from increasing, product similarity among Compustat firms has actually slightly *decreased* (by a cumulative 2.4%) since 1996. This suggests that there has not been any measurable increase in product substitution that can justify the dramatic consolidation that we observed over the past twenty years.

Syverson (2019) notes that that a generalized increase over time in the degree of returns of scale might also have lead to a "superstar firm" effect and increased concentration (as well as higher markups). In Appendix M, I show that this effect is also unlikely to be responsible for the observed trends in concentration. I show that my calibrated values of β (the slope of the marginal cost function in the alternative model presented in Section 7) do not trend over this period. This finding is in line with previous estimates independently obtained by other researchers (Ho and Ruzic, 2018), and suggests that changing economies of scale are unlikely to be driving industry consolidation.

To conclude, while I cannot completely rule out the competing "superstar firms" explanation, these findings allow me to impose significant discipline on the potential channels through which the superstar effect could be at play. In particular, the superstar effect, if at all present, is unlikely to be driven by increased product substitutability or a change in returns to scale. A more promising channel for the "superstar" hypothesis might be offered by an increase in competition due to more extensive geographic overlap among firms. This effect would be akin to an opening up of trade in a spatial model with heterogeneous firms (see for example Melitz and Ottaviano, 2008). It would also rationalize some recent preliminary findings of diverging concentration trends at the national and local level (Rossi-Hansberg et al., 2018), though current data limitations render any conclusion along this dimension premature (Barnatchez et al., 2017).

7.3. Multiproduct Firms (Diversification vs. Differentiation)

The model I present in Section 2 assumes, for the sake of tractability, that every firm produces only one product. A legitimate concern is to what extent the model and the data can accommodate the presence of a vast number of multiproduct firms in Compustat. In particular, one might ask: (1) to what extent do the model's empirics hold after we relax the assumption that firms are undiversified? and (2) to what extent do HP's product-similarity scores conflate diversification with differentiation?

In what follows, I clarify the assumptions under which the model presented in Section 2 can be generalized to multiproduct firms. Conditional on the validity of such assumptions, it will become clear that, even if HP's similarity scores were to capture diversification, that would be a desirable feature of the data, rather than a bug.

Suppose that there are still n firms and k characteristics, but now the n firms produce a total of $m \geq n$ products. The same product might be produced by multiple firms and the same firm may produce more than one product. The vector of units produced for each good is now the m-dimensional vector $\mathbf{y}$. Similarly to matrix $\mathbf{A}$ in Section 2, matrix $\mathbf{A}_1$ projects units of products onto units of characteristics:

$$\mathbf{x} \quad = \quad \mathbf{A}_1\mathbf{y} \tag{76}$$

Because firms are diversified, each firm now produces a basket of goods: instead of representing the number of units produced of each product, the vector $\mathbf{q}$ now represents the number of baskets produced by each firm. The matrix $\mathbf{A}_2$ projects quantity indices for each basket/firm onto units of products supplied:

$$\mathbf{y} \quad = \quad \mathbf{A}_2\mathbf{q} \tag{77}$$

Now I put together the previous two equations. Letting $\mathbf{A} = \mathbf{A}_1\mathbf{A}_2$, I have

$$\mathbf{x} \quad = \quad \mathbf{A}_1\mathbf{y} \quad = \quad \mathbf{A}_1\mathbf{A}_2\mathbf{q} \quad = \quad \mathbf{A}\mathbf{q} \tag{78}$$

The relationship above demonstrates how the linear hedonic structure of the model makes the model immediately generalizable to multiproduct firms. The intuition is that, if the output of a certain firm i is not a single product, but rather a basket of products, one can equivalently project the basket quantity index q_i onto the characteristics space in two steps (by projecting it first onto goods and then onto characteristics), or in one single step (using the composite projection matrix $\mathbf{A}$).

It is clear that the linearity assumption buys a lot in this case: what are, then, the implications, in terms of modeling differentiation and diversification? The implication is that the two are undistinguishable from the point of view of the model. This is obviously less than ideal, but the assumption is required given the nature of the data (I do not observe goods, only firms). I and one that makes sufficient sense for the application of measuring the cross-price elasticity between firms. To see why this assumption makes sense for this application, consider a simplified characteristics space where goods can either be "spoon-like" or "fork-like"; then, compare the following two cases[19]:

Example 1. The quantity index of diversified firm i contains two products—forks and spoons—in equal

[19]Thanks to Simon Board for coming up with this nifty example.

amounts; a second firm j only produces spoons.

Example 2. A single-good firm i produces "sporks," which contain equal loadings on the spoon-like characteristic and the fork-like characteristic; a second firm j produces only spoons.

For the purpose of deriving the cross-price derivative $\partial p_i / \partial q_j$, only half of i's product offering is affected by the change in supply of the product(s) produced by j. Although in the first case both firms produce product B, it would be inappropriate to model the baskets i and j as perfectly substitutable (as one might be tempted to do).

The main limitation of the multi-product interpretation of the model is that it implicitly assumes away the reallocation of resources within the company (the consumer chooses $\mathbf{q}$, not $\mathbf{y}$). If the quantity supplied by a competitor decreases, a company producing spoons and forks might produce a larger quantity of both, but it won't be able vary the ratio between the two.

This clarifies why it is desirable for HP's similarity scores to capture diversification into overlapping products as well as differentiation into overlapping characteristics, as these produce identical effects in terms of the product substitutability under the linearity assumption.

7.4. Sample Selection

As is the case for other papers that use Compustat data to describe trends in market power, the data I use cover only public firms. As a consequence, while the model I developed could be estimated using economy-wide data, private firms are absent in the empirical implementation of the paper. As is the case with Baqaee and Farhi (2017) and Edmond et al. (2018), I argue that my empirical analyses are still informative, for three reasons. First, the overwhelming majority of private firms compete only at a local level and do not interact strategically with the firms covered in Compustat. This is shown by a large literature on small enterprises (see for example Hurst and Pugsley, 2011) that finds confirmation in some of the latest empirical evidence of strategic complementarity in pricing: Amiti et al. (2018) provide evidence in microdata of a two-speed market, in which small firms do not interact strategically with larger ones. Second, while Compustat covers a limited *number* of firms, these firms account for about 60% of the operating profits earned by U.S. corporations[20] and and 40% of nonfarm business GDP, and are the most likely to behave oligopolistically. Moreover, as shown by Gabaix (2011), these corporations make a disproportionately large contribution to aggregate fluctuations in economic activity. Third, the trends in markups and concentration that kickstarted the current debate on concentration and antitrust were measured on public firms in the first place; therefore, studying public firms is interesting in and on itself.

I investigate nonetheless the robustness of my analysis to the inclusion of private and foreign competitors. In Appendix K, I show how one can combine Compustat and sector-level data to construct weights that can be applied to the firms in my model to act as proxies for unobserved firms. I show alternative welfare calculations based on this method. While, as one would expect, the overall level of my chosen metrics of market power (profit share of surplus, deadweight losses as percentage of total potential surplus) is somewhat smaller than in the baseline analysis, all these metrics show an equally pronounced upward trend over the

[20]For this estimate, I compare Compustat operating profits to Gross Operating Surplus in the corporate sector from U.S. national accounts.

1997-2017 period. This additional check suggests that the finding that market power has increased as a consequence of industry consolidation is unlikely to be determined by the sample selection.

7.5. Linear Quadratic Utility Formulation

In addition to allowing the model to be extended to multiproduct firms, the linear demand assumption has a number of desirable properties. Yet, because it is the one assumption of my model that cannot be relaxed, it is important to think about the implications of this assumption. In particular, one potential objection is that such demand specification might be restrictive and that perhaps an alternative utility specification, such as Dixit-Stiglitz preferences (CES), should be used instead. In this subsection, I discuss why the linear demand assumption is the most appropriate choice for this setting.

Firstly, I note that my utility specification is simply the hedonic, discrete counterpart of a type of preference aggregator that is already widely used in macroeconomics, trade, and industrial organization (see for example Asplund and Nocke 2006; Foster et al. 2008; Melitz and Ottaviano 2008; Syverson 2019):

$$U(q) \; = \; b_0 q_0 + b \int_{i \in \mathcal{I}} q_i \, di - a \left(\int_{i \in \mathcal{I}} q_i \, di \right)^2 - \int_{i \in \mathcal{I}} q_i^2 \, di \tag{79}$$

This type of linear-quadratic aggregators are generally preferred in contexts where product substitutability and imperfect competition are important (as is the case in this study), as they are better able to capture the relationship between size and markups than CES. My hedonic demand specification simply adds, to this setup, asymmetry in the degree of substitutability between different firm pairs as well as the ability interpret the cross derivatives of the demand system in terms of similarity between products.

The reason why CES (or nested CES) are inadequate to my setting is that they yield a constant markup across firms within a sector and this in turn makes the economy tend toward allocative efficiency.[21] Dhingra and Morrow (forthcoming) show that this property of CES preferences does *not* generalize to any other demand specification.

Modeling product differentiation and heterogeneity in markups across firms in a realistic way requires that both own and cross-price elasticity be able to adjust to in response to changes in firm size and expenditure shares. By definition, neither CES nor nested CES allow for this flexibility. This intuition is consistent with the empirics of Edmond et al. (2018): using the Kimball aggregator as a demand specification, they show that a significant degree of demand super-elasticity (a feature of linear demand) is required to capture the observed heterogeneity in markups across firms. Concordantly, their calibration of the aggregate demand function is more consistent with a linear demand specification than CES.

While (nested) CES preferences have a number of desirable properties when it comes to modeling the relationship between aggregate productivity and the distribution of firm-level productivity, they become hard to rationalize in the context of this paper, not just empirically but also theoretically. If one tries, for example, to write the isoelastic counterpart to the demand system in equation (12)

$$\log \mathbf{p} \; = \; \mathbf{A}'\mathbf{b} - \mathbf{A}'\mathbf{A} \log \mathbf{q} \tag{80}$$

[21] This is what allows Hsieh and Klenow (2009) to identify policy distortions simply by observing variation in revenue productivity.

and derive the utility specification that generates it, she will quickly find that it requires the highly stringent integrability condition that $\mathbf{A}'\mathbf{A}$ be diagonal. In other words, no representative agent utility exists that rationalizes the demand system of equation (80) for values of $\mathbf{A}'\mathbf{A}$ other than the identity matrix, in which case it just collapses back to the canonical CES demand aggregator.

7.6. Factoring in Capital Costs

Thus far, my analysis has implicitly treated the consumption of fixed capital as a sunk cost. It is therefore possible that the increase in the (operating) profit share of surplus might be justified by an increase in capital costs, captured by the rate of depreciation. To verify whether this is the case, we need to obtain, from Compustat data, a second measure of aggregate profits from which capital consumption costs are netted. I define an alternative measure of aggregate profits $\bar{\Pi}$, which detracts interest and capital expenditures as well:the advantage of this measure, which in many ways resembles *free cash flow*, is that it is readily computed using Compustat data, thus avoiding the problem of having to cherrypick among many imperfect proxies for the cost of capital and the depreciation rate. It also allows in some sense to control for changes in investment opportunities, which could be relevant given the fact that during the relevant period productivity growth and investment slowed down.

In Appendix L, I replicate Figure 5 using this alternative measure of aggregate profits ($\bar{\Pi}$ is used in place of Π). The resulting profit share increases more dramatically over the 1997-2017 period: from about 23% to nearly 35%. This alternative computation suggests that fixed costs are not the driving factor behind this trend. The deadweight losses from oligopoly increase when we factor capital costs: 18.4% of total surplus in 2017. There are two reasons behind this finding: first, factoring in fixed costs decreases total surplus both in the observed equilibrium and in the competitive outcome, but it has a smaller impact in percentage terms on the latter because the denominator is larger; second, in the competitive outcome some firms shut down and a portion of the the fixed costs can be avoided. This increase in profits net of fixed costs do not seem to be explained by an increase in the rate of return on capital, since most available measures for the rate of return on capital decrease over this period. While the upward trend in the deadweight loss appears somewhat less dramatic under this definition of profits, as it starts from a level of 15.2% in 1997.

Overall, this robustness exercise tends to confirm the previous findings, suggests that increasing fixed costs are likely not to be blamed for the increasing profits.

7.7. Policy Distortions and Other Frictions

One natural extension of my model that can accommodate for additional sources of variation in markups is to include policy/financial wedges in the style of Hsieh and Klenow (2009). If shadow taxes on inputs/output are present, the expression for the Cournot-Nash equilibrium allocation can be amended as follows:

$$\mathbf{q}^{\Phi} = (\mathbf{I} + \mathbf{A}'\mathbf{A})^{-1} [\mathbf{A}'\mathbf{b} - \mathbf{C}(\mathbf{1} + \tau)] \tag{81}$$

where τ is a vector of shadow taxes. The vector $(\mathbf{A}'\mathbf{b} - \mathbf{c})$ thus needs to be replaced by $[\mathbf{A}'\mathbf{b} - \mathbf{C}(\mathbf{1} + \tau)]$. Fortunately, both these vectors are obtained from the data as $(\mathbf{I} + \mathbf{A}'\mathbf{A})\sqrt{\pi}$; therefore, this extension does not affect any of my metrics of welfare. In this sense, although my theoretical framework does not model these shadow costs explicitly, it is robust to their inclusion. The only caveat is that, if these shadow costs are present, the interpretation of the empirical counterpart of the vector $(\mathbf{A}'\mathbf{b} - \mathbf{c})$ changes (the marginal cost should be thought to include the shadow tax).

7.8. Limitations, Extensions, and Future Work

The model and the empirical results in this paper come with some caveats and limitations.

First, this paper does not aim to propose an all-encompassing theory of market power; rather, it seeks to formulate a theory that relates firm performance to industry concentration in a differentiated product setting. Many important industry-specific drivers of market power are not necessarily directly related to industry concentration. One example, particularly relevant to the healthcare sector, is search costs (Lin and Wildenbeest, 2019). Other important drivers of market power are geography (Rossi-Hansberg et al., 2018), brand equity (Goldfarb et al., 2009), and collusion (Asker et al., forthcoming).

The empirical industrial organization literature exists because it has been successful in modeling these industry-specific dynamics. Capturing this vast heterogeneity is neither feasible nor desirable for a macroeconomic model, and is in any case it's significantly beyond the scope of this paper. This paper aims to estimate the welfare effects of the consolidation that has occurred across most U.S. industries.

Another limitation that applies to the model presented in Section 2 is that it abstracts away from input-output linkages: all firms sell their output directly to the representative household (horizontal economy). Similarly, the baseline model does not incorporate market-size effects: that is, all firms are presumed to sell to the same population of consumer. While it's obviously not feasible to address every possible limitation of my new model in a general-interest paper like this one, in Appendix O, I explore some extensions of the model that incorporate these factors, and perform some sensitivity analysis.

The most significant impediment to incorporating these additional forces in my new model is the lack of data that simultaneously cover firm similarity, input-output linkages, and market size. If these data shortcomings can be overcome, studying the interactions between these different forces in a general equilibrium setting could provide an interesting avenue for future research.

8. Conclusions

In this study, I have presented a new general equilibrium model of oligopolistic competition with hedonic demand and differentiated products, with the objective of measuring the welfare consequences of rising oligopoly power in the United States from 1997 to 2017. I applied the model to the data using a data set (recently developed by Hoberg and Phillips, 2016) of bilateral product-similarity scores that covers all public firms in the United States on a year-by-year basis. Through the lens of my model, these similarity scores are used to retrieve the cross-price elasticity of demand for every pair of publicly traded firms.

My measurements suggest that industry concentration has a considerable and increasing effect on aggregate welfare. In particular, I estimate that, if all publicly traded firms were to behave as atomistic competitors, the total surplus produced by this set of companies would increase by 13.3 percentage points. Consumer welfare would increase even more dramatically—it would more than double—as surplus would be entirely reallocated from producers to consumers. I find that most of the deadweight loss caused by oligopoly (10 percentage points of the total surplus produced) can be attributed to resource misallocation—that is, most of the deadweight losses could theoretically be recovered by a benevolent social planner, even if we assume inelastic input supply. I also find that, at the other extreme, consolidating firm ownership in the hands of one producer would depress aggregate surplus by about 8.6 percentage points. Consumer surplus would suffer even more, with a projected decrease of about 49 percentage points. Overall, my analysis of firm-level data suggests that there is evidence of sizable welfare distortions and resource misallocation due to oligopoly power.

By mapping my model to firm-level data for every year between 1997 and 2017, I find that, while both the profits earned by U.S. public corporations and the corresponding consumer surplus have increased over this period, profits have increased at a significantly faster pace: consequently, the share of surplus appropriated by firms in the form of oligopoly profits has increased substantially (from 33% to 44%). Consistent with this finding, I estimate that the welfare costs of oligopoly, computed as the percentage increase in surplus that is obtained by moving to the competitive outcome, have increased (from 12% to 13.3%). Overall, my estimates are consistent with the hypothesis that the observed secular trends in markups and concentration have resulted in increased welfare losses, particularly at the expense of the consumer.

The model allows me to compute a number of novel antitrust-relevant counterfactuals, and to shed light on the possible causes of the oligopolization of U.S. industries. I have shown that a potential explanation might lie in the secular decline of U.S. public equity markets. In particular, I show that the secular decline in IPOs and the surge in takeovers of VC-backed startups can quantitatively account for the decline in the number of corporations listed on U.S. exchanges, as well as the observed increase in the deadweight loss from oligopoly and the larger share of surplus accruing to producers.

This paper contributes—both methodologically and empirically—to a growing literature in finance and macroeconomics that is devoted to incorporating heterogeneity and imperfect competition in general equilibrium models. In particular, it shows that combining firm financials with measures of similarity based on natural-language processing of regulatory filings offers a promising avenue to model product differentiation and imperfect substitutability at the macroeconomic level: it affords the opportunity to impose a less arbitrary structure on the degree of substitution across sectors and firms.

My quantitative findings in this paper inform a growing debate about the welfare implications of rising concentration and markups in the United States and potential policy responses. One potential policy implication of my findings is that, while antitrust agencies tend to focus most of their merger-review work on mergers between large incumbents, acquisitions of VC-backed startups might also have important implications for competition. Given there is also emerging microeconomic evidence on "killer acquisitions" (Cunningham et al., 2018), antitrust agencies may wish to consider increasing their scrutiny of these transactions, as they might provide an avenue for large corporations to restrict competition while circumventing antitrust regulations.

References

Amiti, M., O. Itskhoki, and J. Konings (2018): "International Shocks, Variable Markups, and Domestic Prices," *The Review of Economic Studies*.

Asker, J., A. Collard-Wexler, and J. De Loecker (forthcoming): "(Mis)Allocation, Market Power, and Global Oil Extraction," *The American Economic Review*.

Asplund, M. and V. Nocke (2006): "Firm turnover in imperfectly competitive markets," *The Review of Economic Studies*, 73, 295–327.

Atkeson, A. and A. Burstein (2008): "Pricing-to-market, trade costs, and international relative prices," *American Economic Review*, 98, 1998–2031.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, J. Van Reenen, et al. (2017): "The Fall of the Labor Share and the Rise of Superstar Firms," Tech. rep., National Bureau of Economic Research, Inc.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's who in networks. Wanted: The key player," *Econometrica*, 74, 1403–1417.

Baqaee, D. and E. Farhi (2019): "Networks, Barriers, and Trade," Tech. rep., National Bureau of Economic Research.

Baqaee, D. R. and E. Farhi (2017): "Productivity and Misallocation in General Equilibrium." Working paper, National Bureau of Economic Research.

——— (2018): "Macroeconomics with heterogeneous agents and input-output networks," Working paper, National Bureau of Economic Research.

Barkai, S. (2016): "Declining labor and capital shares," *Stigler Center for the Study of the Economy and the State New Working Paper Series*, 2.

Barnatchez, K., L. D. Crane, and R. Decker (2017): "An assessment of the national establishment time series (nets) database," .

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 841–890.

Bimpikis, K., S. Ehsani, and R. İlkiliç (2019): "Cournot competition in networked markets," *Management Science*, 65, 2467–2481.

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying technology spillovers and product market rivalry," *Econometrica*, 81, 1347–1393.

Bonacich, P. (1987): "Power and centrality: A family of measures," *American journal of sociology*, 92, 1170–1182.

Bowen, D. E. I., L. Frésard, and G. Hoberg (2018): "Technological disruptiveness and the evolution of IPOs and sell-outs," .

Bresnahan, T. F. (1989): "Empirical studies of industries with market power," *Handbook of industrial organization*, 2, 1011–1057.

Bresnahan, T. F. and R. Schmalensee (1987): "The empirical renaissance in industrial economics: an overview," *The Journal of Industrial Economics*, 371–378.

Chen, H., L. Cohen, and D. Lou (2016): "Industry window dressing," *The Review of Financial Studies*, 29, 3354–3393.

Chintagunta, P. K. (2002): "Investigating category pricing behavior at a retail chain," *Journal of Marketing Research*, 39, 141–154.

Correia, S., P. Guimarães, and T. Zylkin (2019): "PPMLHDFE: Fast poisson estimation with high-dimensional fixed effects," *arXiv preprint arXiv:1903.01690*.

Cunningham, C., F. Ederer, and S. Ma (2018): "Killer acquisitions," *Available at SSRN 3241707*.

De Loecker, J. (2011): "Recovering markups from production data," *International Journal of Industrial Organization*, 29, 350–355.

De Loecker, J., J. Eeckhout, and G. Unger (2018): "The rise of market power and the macroeconomic implications," *NBER Working Papers*.

De Loecker, J. and P. T. Scott (2016): "Estimating market power: Evidence from the US Brewing Industry," Working paper, National Bureau of Economic Research.

Dhingra, S. and J. Morrow (forthcoming): "Monopolistic Competition and Optimum Product Diversity Under Firm Heterogeneity," *Journal of Political Economy*.

Doidge, C., K. M. Kahle, G. A. Karolyi, and R. M. Stulz (2018): "Eclipse of the public corporation or eclipse of the public markets?" *Journal of Applied Corporate Finance*, 30, 8–16.

Dolata, U. (2017): "Apple, Amazon, Google, Facebook, Microsoft: Market concentration-competition-innovation strategies," Tech. rep., Stuttgarter Beiträge zur Organisations-und Innovationsforschung, SOI ....

Edmond, C., V. Midrigan, and D. Y. Xu (2018): "How costly are markups?" Working paper, National Bureau of Economic Research.

Einav, L. and J. Levin (2010): "Empirical industrial organization: A progress report," *Journal of Economic Perspectives*, 24, 145–62.

Epple, D. (1987): "Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products," *Journal of political economy*, 95, 59–80.

Ewens, M. and J. Farre-Mensa (2018): "The deregulation of the private equity markets and the decline in IPOs," *Available at SSRN 3017610*.

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?" *American Economic Review*, 98, 394–425.

Fruchterman, T. M. and E. M. Reingold (1991): "Graph drawing by force-directed placement," *Software: Practice and experience*, 21, 1129–1164.

Gabaix, X. (2011): "The granular origins of aggregate fluctuations," *Econometrica*, 79, 733–772.

Gao, X., J. R. Ritter, and Z. Zhu (2013): "Where have all the IPOs gone?" *Journal of Financial and Quantitative Analysis*, 48, 1663–1692.

Goeree, M. S. (2008): "Limited information and advertising in the US personal computer industry," *Econometrica*, 76, 1017–1074.

Goldfarb, A., Q. Lu, and S. Moorthy (2009): "Measuring brand value in an equilibrium framework," *Marketing Science*, 28, 69–86.

Grullon, G., Y. Larkin, and R. Michaely (2018): "Are US industries becoming more concentrated?" *Forthcoming, Review of Finance*.

Gutiérrez, G. and T. Philippon (2017): "Declining Competition and Investment in the US," Working paper, National Bureau of Economic Research.

Hall, R. E. (2018): "New evidence on the markup of prices over marginal costs and the role of mega-firms in the us economy," Tech. rep., National Bureau of Economic Research.

Haltiwanger, J., R. Kulick, and C. Syverson (2018): "Misallocation measures: The distortion that ate the residual," Working paper, National Bureau of Economic Research.

Harberger, A. C. (1954): "Monopoly and Resource Allocation," *The American Economic Review*, 77–87.

Ho, S.-J. and D. Ruzic (2018): "Returns to Scale, Productivity Measurement, and Trends in US Manufacturing Misallocation," in *2018 Meeting Papers*, Society for Economic Dynamics, 119.

Hoberg, G. and G. Phillips (2016): "Text-based network industries and endogenous product differentiation," *Journal of Political Economy*, 124, 1423–1465.

Hoberg, G., G. Phillips, and N. Prabhala (2014): "Product market threats, payouts, and financial flexibility," *The Journal of Finance*, 69, 293–324.

Hopenhayn, H. A. (2014): "Firms, misallocation, and aggregate productivity: A review," *Annu. Rev. Econ.*, 6, 735–770.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124, 1403–1448.

Hurst, E. and B. W. Pugsley (2011): "What Do Small Businesses Do?" *Brookings Papers on Economic Activity*, 73.

Kahle, K. M. and R. M. Stulz (2017): "Is the US public corporation in trouble?" *Journal of Economic Perspectives*, 31, 67–88.

Katz, L. (1953): "A new status index derived from sociometric analysis," *Psychometrika*, 18, 39–43.

Khan, L. (2018): "The new Brandeis movement: America's antimonopoly debate," .

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017): "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 132, 665–712.

Lin, H. and M. R. Wildenbeest (2019): "Non-Parametric Estimation of Search Costs for Differentiated Products: Evidence from Medigap," *Journal of Business & Economic Statistics*, 1–29.

Martin, S., W. M. Brown, R. Klavans, and K. W. Boyack (2011): "OpenOrd: an open-source toolbox for large graph layout," *Visualization and Data Analysis 2011*, 7868, 786806.

Melitz, M. J. and G. I. Ottaviano (2008): "Market size, trade, and productivity," *The review of economic studies*, 75, 295–316.

Monderer, D. and L. S. Shapley (1996): "Potential games," *Games and economic behavior*, 14, 124–143.

Moore, M. and D. Tambini (2018): *Digital dominance: the power of Google, Amazon, Facebook, and Apple*, Oxford University Press.

Nevo, A. (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342.

Pellegrino, B. and G. Zheng (2017): "Institutions and Resource Misallocation: Firm-level Evidence," *UCLA Anderson Working Paper*.

Porter, M. E. (1991): "Towards a dynamic theory of strategy," *Strategic management journal*, 12, 95–117.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *The Journal of Political Economy*, 82, 34–55.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2018): "Diverging Trends in National and Local Concentration," Working Paper 25066, National Bureau of Economic Research.

Silva, J. S. and S. Tenreyro (2010): "On the existence of the maximum likelihood estimates in Poisson regression," *Economics Letters*, 107, 310–312.

Syverson, C. (2019): "Macroeconomics and Market Power: Facts, Potential Explanations, and Open Questions," Working paper, Brookings.

Traina, J. (2018): "Is aggregate market power increasing? production trends using financial statements," Working paper, Chicago Booth Stigler Center.

Ushchev, P. and Y. Zenou (2018): "Price competition in product variety networks," *Games and Economic Behavior*, 110, 226–247.

Werden, G. J. (2018): "Back to School: What the Chicago School and New Brandeis School Get Right," *Available at SSRN 3247116*.

Figure 1: Example Product Space: Two Firms, Two Characteristics

The following diagram exemplifies the hedonic demand model presented in Section 2, for the simple case where there are only two product characteristics (A and B) and only two competitors (1 and 2). Each firm's output is valued, by the representative consumer, as a basket of characteristics, and each firm exists as a vector on the unit hypersphere of product characteristics (in this example, we have a circle). The tighter the angle θ, the higher the cosine similarity, and the higher the cross-price elasticity of demand between the two firms.

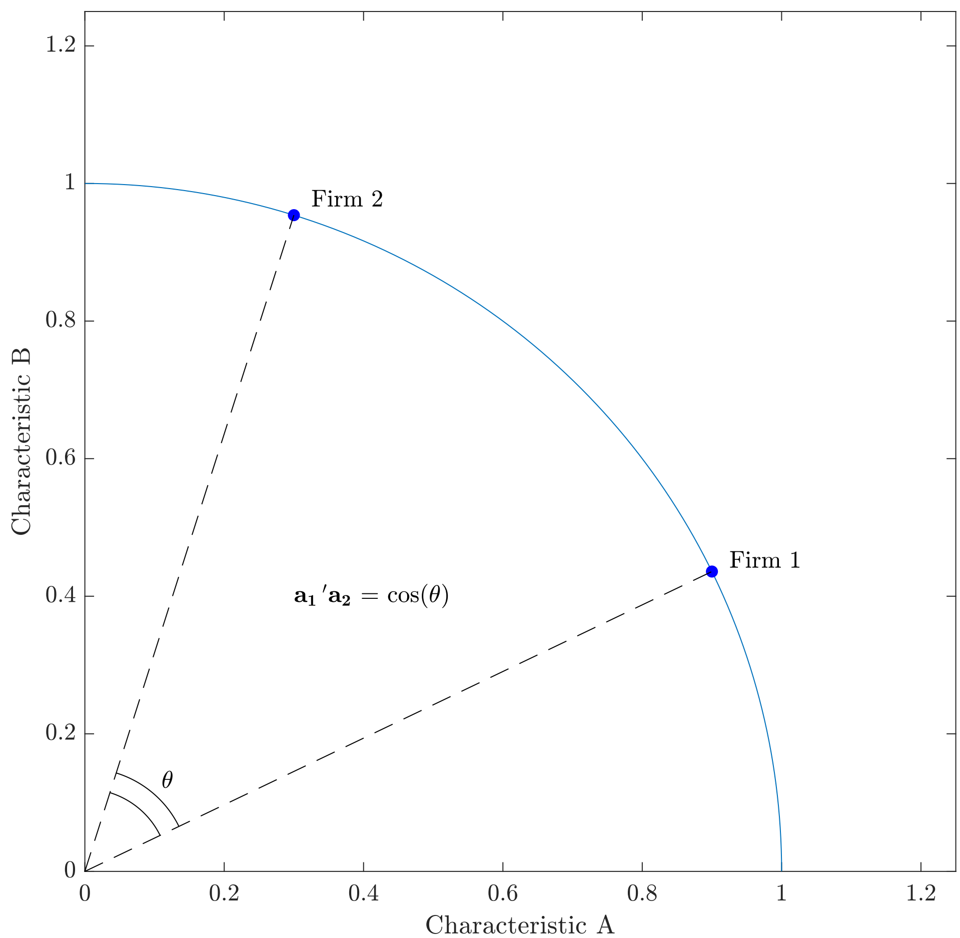


FIGURE 2: MARKUPS AND PRODUCT DIFFERENTIATION IN THE SYMMETRIC CASE

The following figure plots a comparative static for the firm-level markup in the symmetric case of the model described in Section 2: all firms are identical, and the firms' cosine similarity matrix in the product space given by $\mathbf{A}'\mathbf{A} = \mathbf{I} + a\left(\mathbf{11}' - \mathbf{I}\right)$. The parameter a controls the sparsity of the similarity matrix; n is the number of firms. When $a = 0$, all firms produce unrelated goods and individually behave as monopolists. When $a \to 1$, the firms produce identical goods and the game collapses to the canonical Cournot Oligopoly.

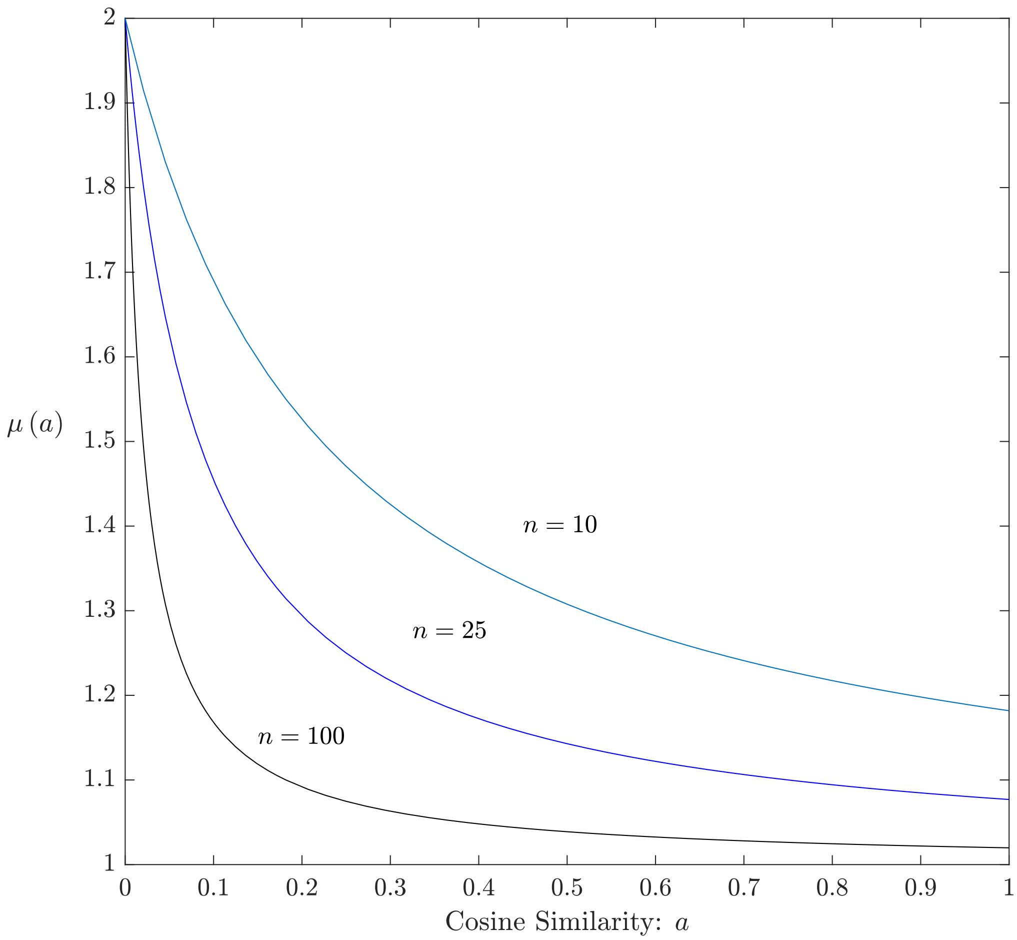


Figure 3: Visualization of the Product Space (CRSP-Compustat)

The following diagram is a two-dimensional representation of the network of product similarities computed by Hoberg and Phillips (2016), which is used in the estimation of the model presented in Section 2. The data covers the universe of Compustat firms in 2004. Firm pairs that have thicker links are closer in the product market space. These distances are computed in a space that has approximately 61,000 dimensions. To plot such a high-dimensional object over a plane, I applied the gravity algorithm of Fruchterman and Reingold (1991), which is standard in social network analysis.

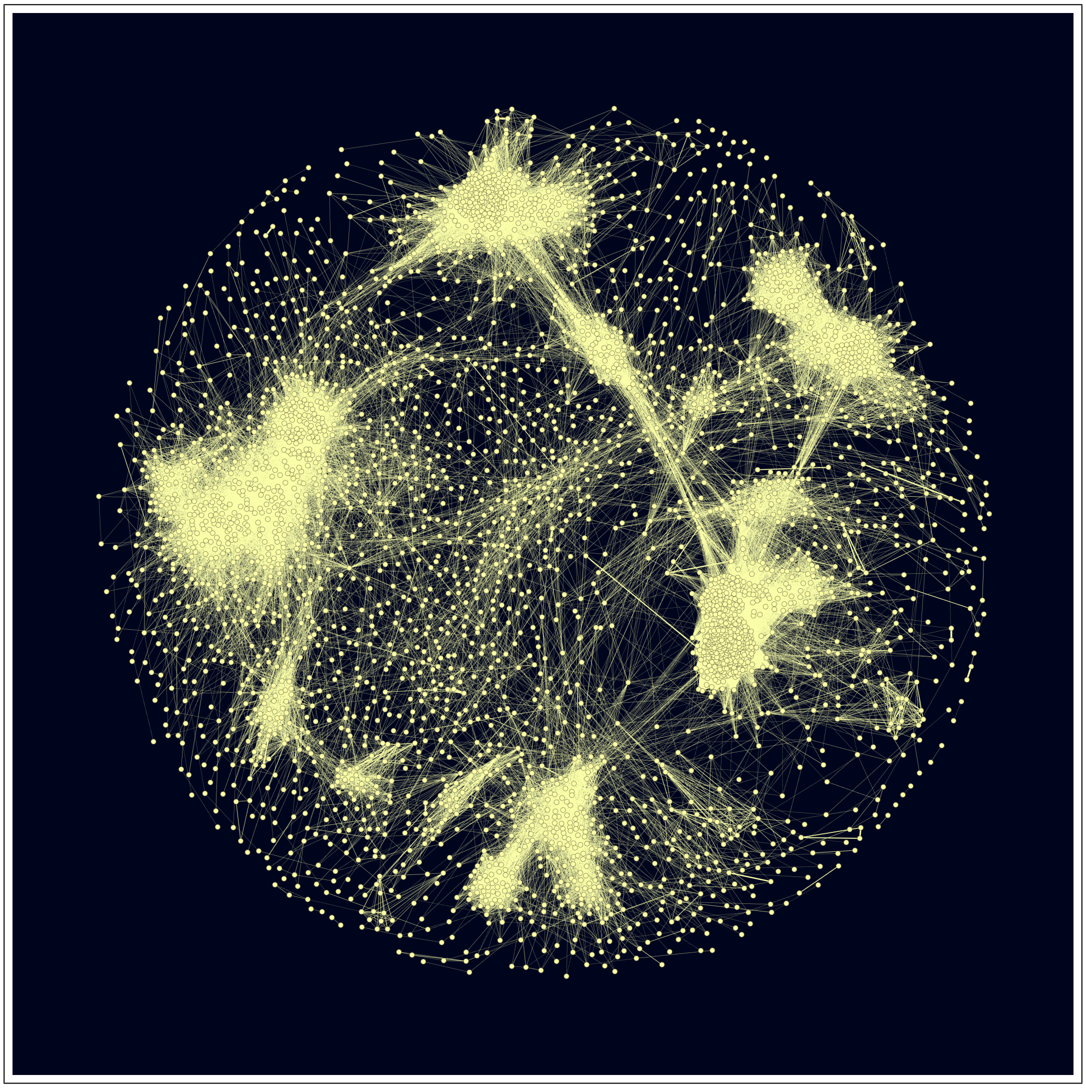


Figure 4: Weighted Market Share, Centrality, and surplus Appropriation

The following graph plots the two metrics of concentration presented in Section 2 (the Weighted Market Share and the Inverse Centrality Index) against each other. The main difference between these two measures is that the first (y-axis) captures both firm size and network configuration, while the latter (x-axis) only captures network configuration. Because the Weighted Market Share is exactly equal to half the ratio of profits to consumer surplus (π_i/s_i), the same variable can be read on the right axis scale as the producer/consumer surplus ratio. Note that some firms have a slightly negative Inverse Centrality Index: these negative values correspond to firms that, given their network position, would not survive in a counterfactual in which all firms are otherwise identical.

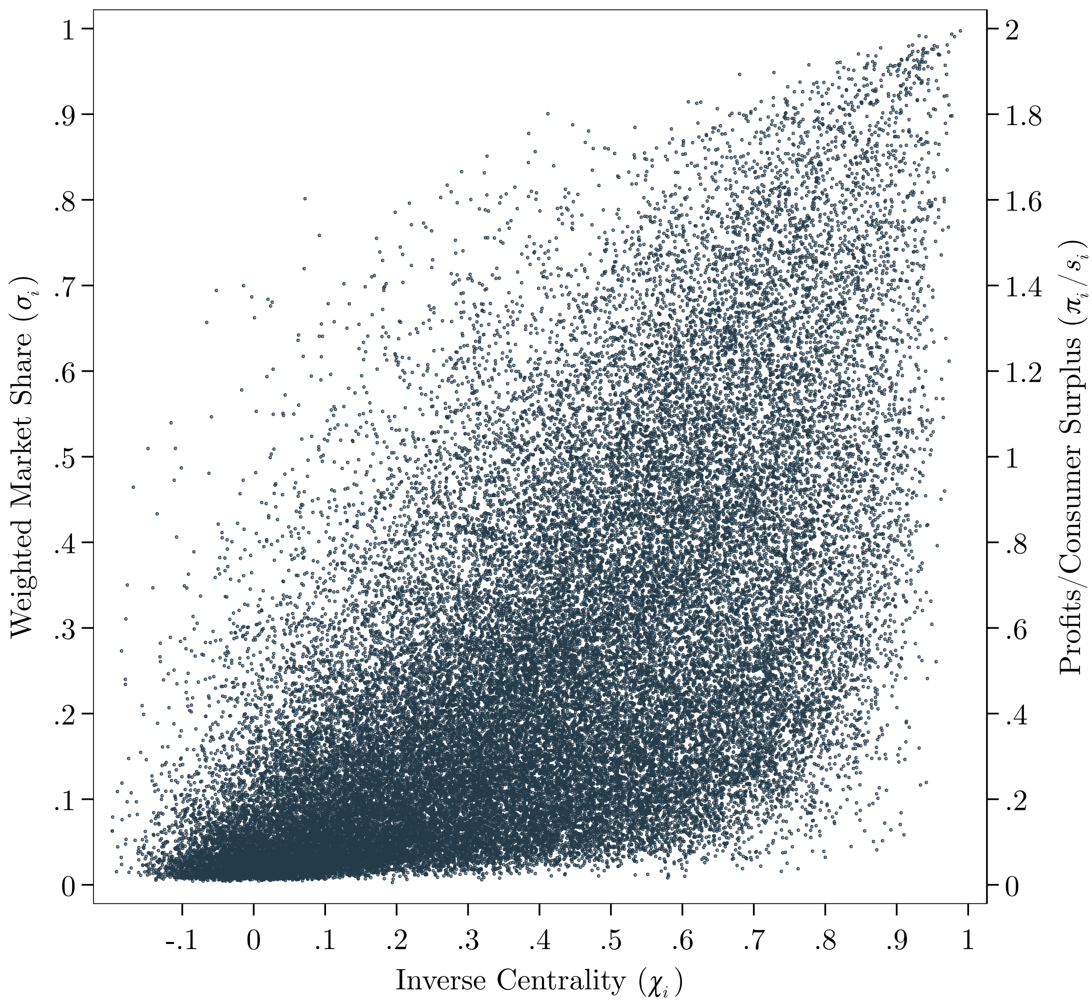


Figure 5: Evolution of Aggregate Profits and Consumer Surplus (1997-2017)

The following figure plots the evolution, between 1997 and 2017, of the aggregate profit function $\Pi(\mathbf{q})$, the aggregate consumer surplus $S(\mathbf{q})$ as well as the total surplus function $W(\mathbf{q})$ from the model described in Section 2. Profits as a percentage of total surplus (Π/W, black dotted line) are shown on the right axis. These functions are estimated for the Compustat-CRSP universe.

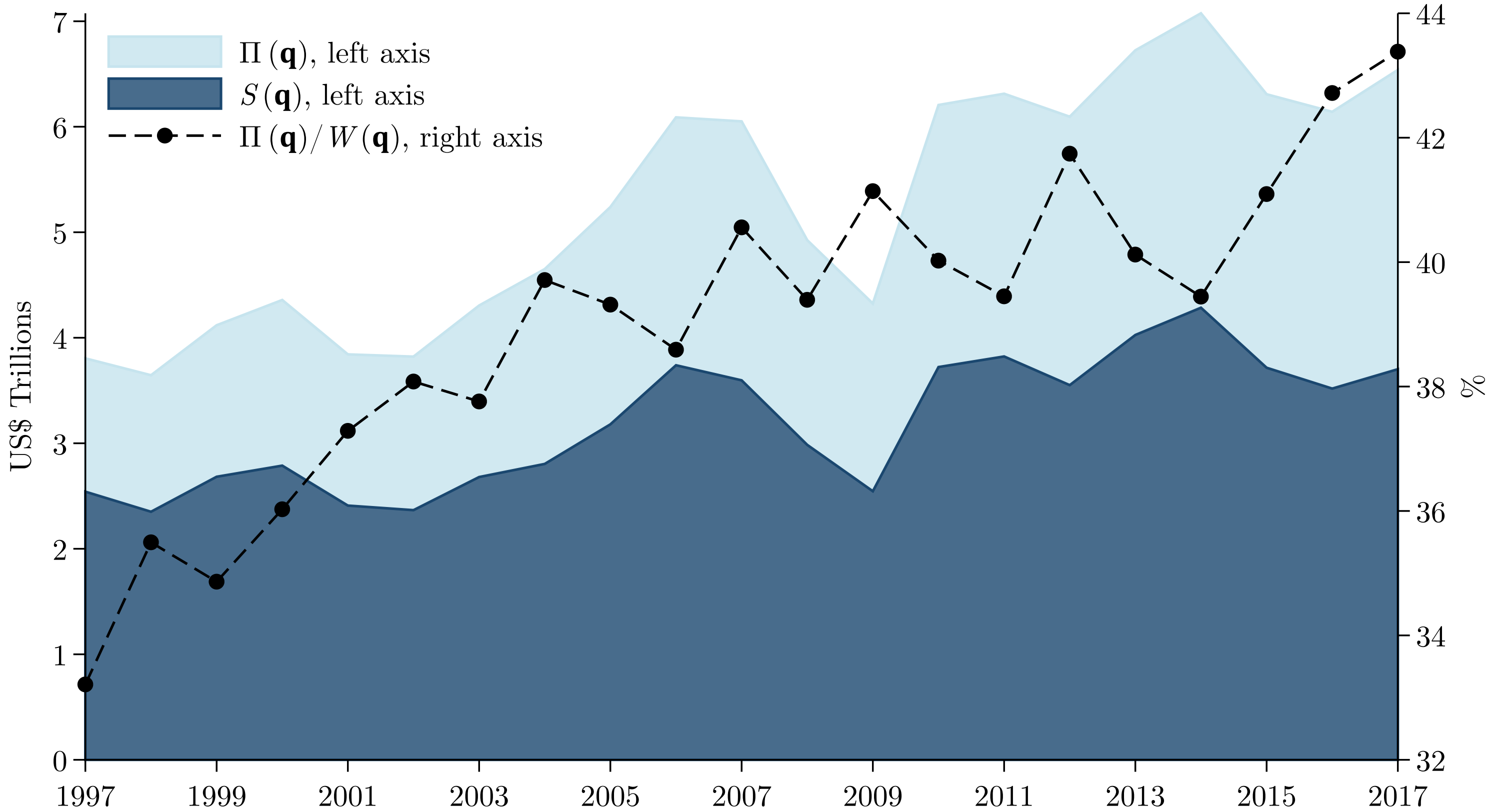


Figure 6: Evolution of the Welfare Costs of Oligopoly (1997-2017)

The following figure plots the evolution, between 1997 and 2017, of the deadweight loss from oligopoly, that is, the percentage gain in total surplus $\Delta\%W(\mathbf{q})$ from moving to the first-best allocation . The percentage gain in consumer surplus function $\Delta\%S(\mathbf{q})$ is shown on the right axis.

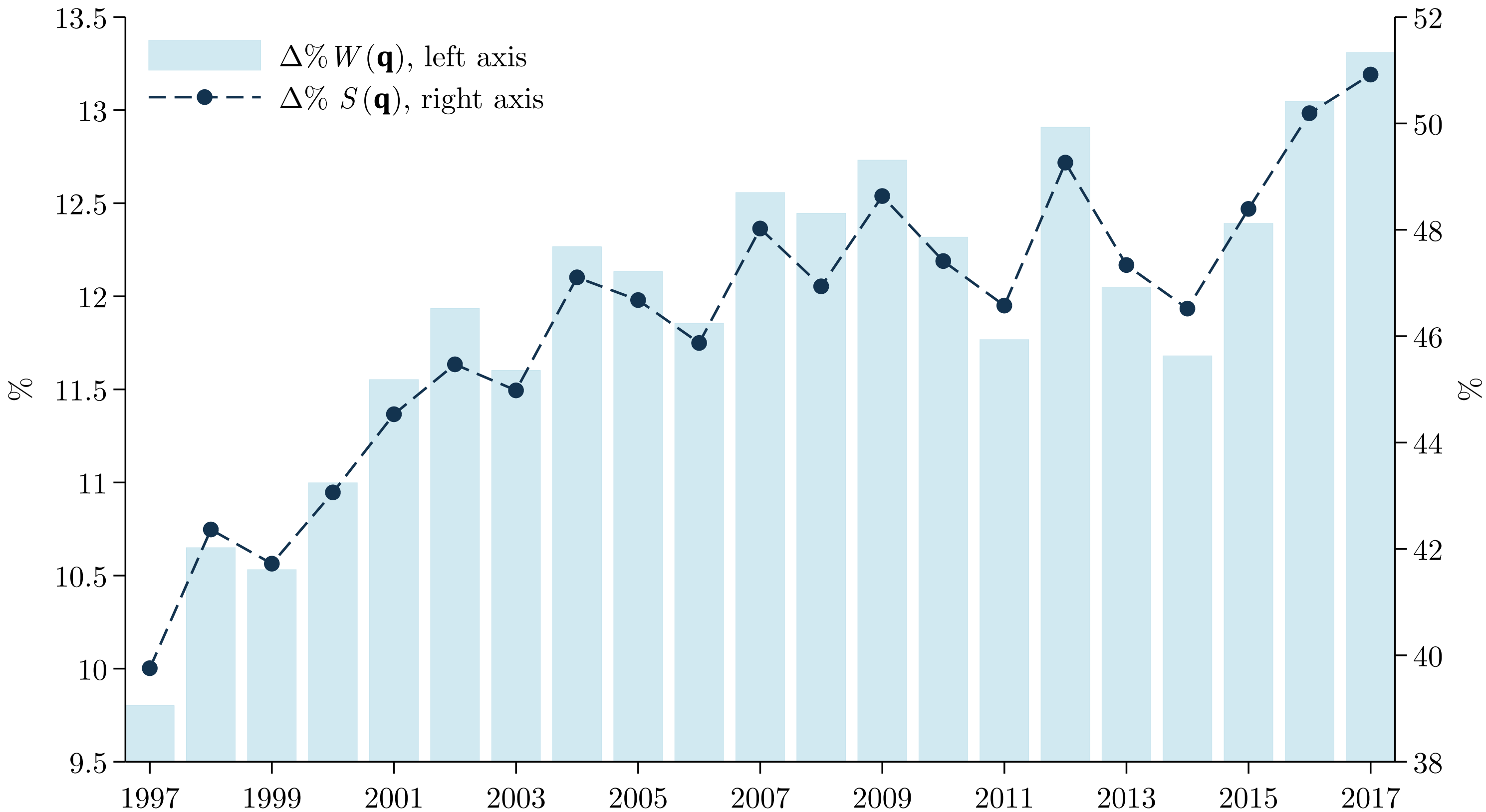
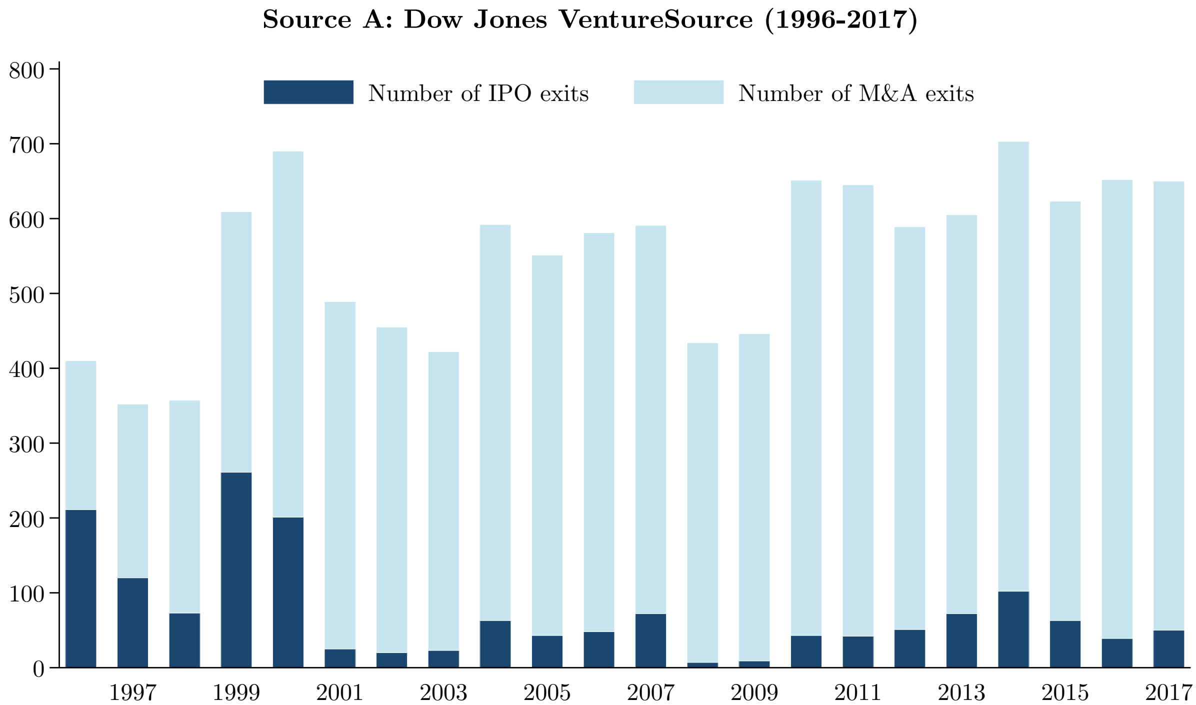


Figure 7: Successful Venture Capital Exits by Year and Type

The following figure plots the number of successful venture capital exits in the United States by year and type (Initial Public Offering v/s Acquisition) for two different data sources.



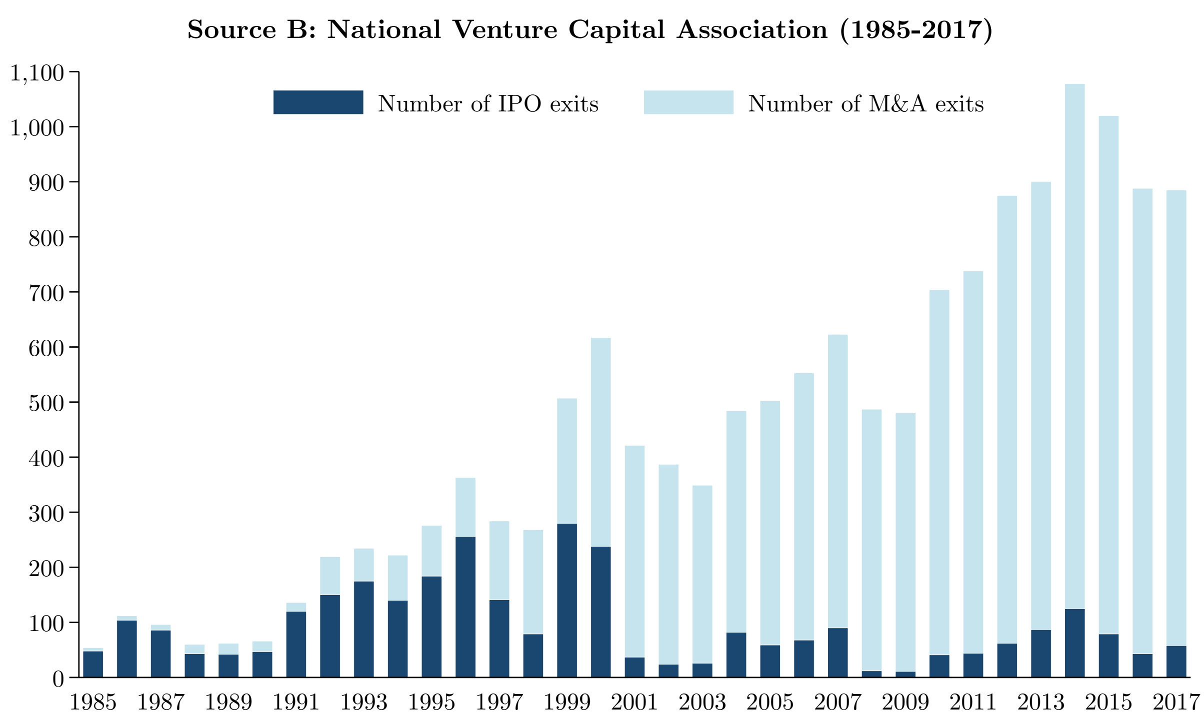


Figure 8: Number of Firms in Compustat vs. Counterfactual

The following figure plots the number of firms that are present in the merged Compustat-Hoberg and Phillips (2016) data set against a counterfactual in which the ratio of IPO VC exits to M&A exits would have stayed constant after 1996.

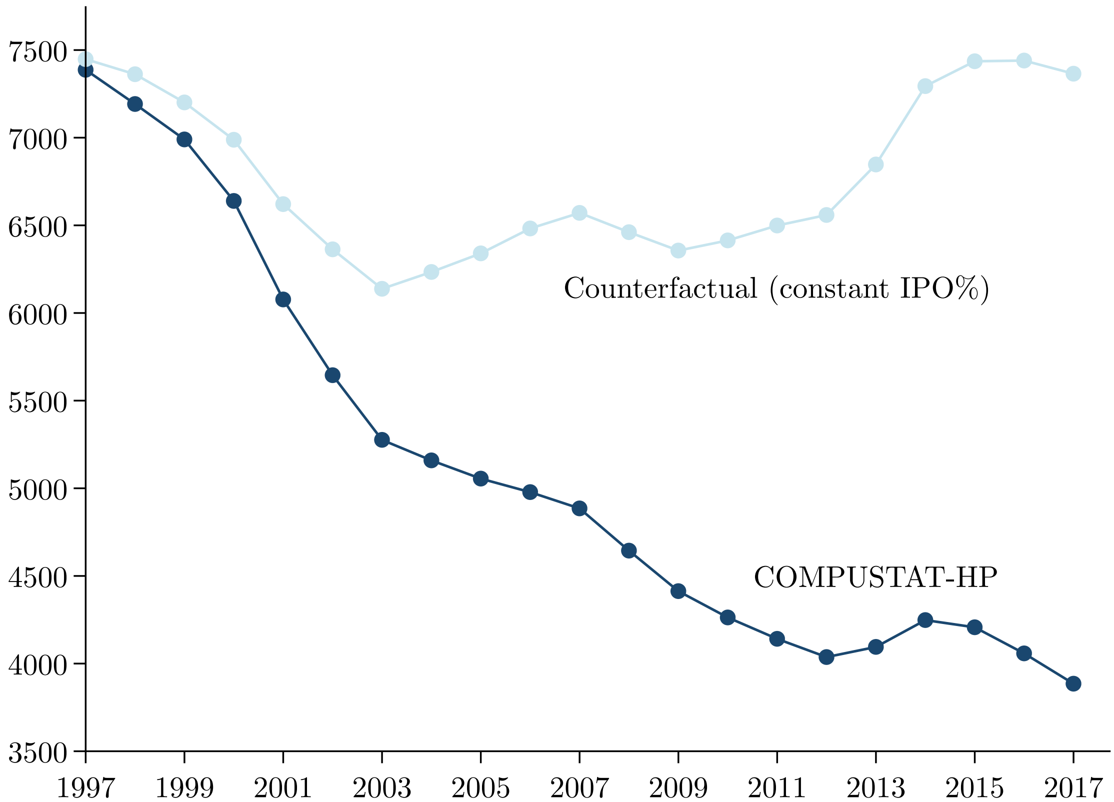


Figure 9: Aggregate Profits and Consumer Surplus in the "Constant IPO rate" Scenario (1997-2017)

The following figure plots the evolution, between 1997 and 2017, of the aggregate profit function $\Pi(\mathbf{q})$, the aggregate consumer surplus $S(\mathbf{q})$ as well as the total surplus function $W(\mathbf{q})$ in the counterfactual presented in Section 2, which assumes that the rate of IPOs as a percentage of VC exits would have stayed constant after 1996. Profits as a percentage of total surplus (Π/W, black dotted line) are shown on the right axis.

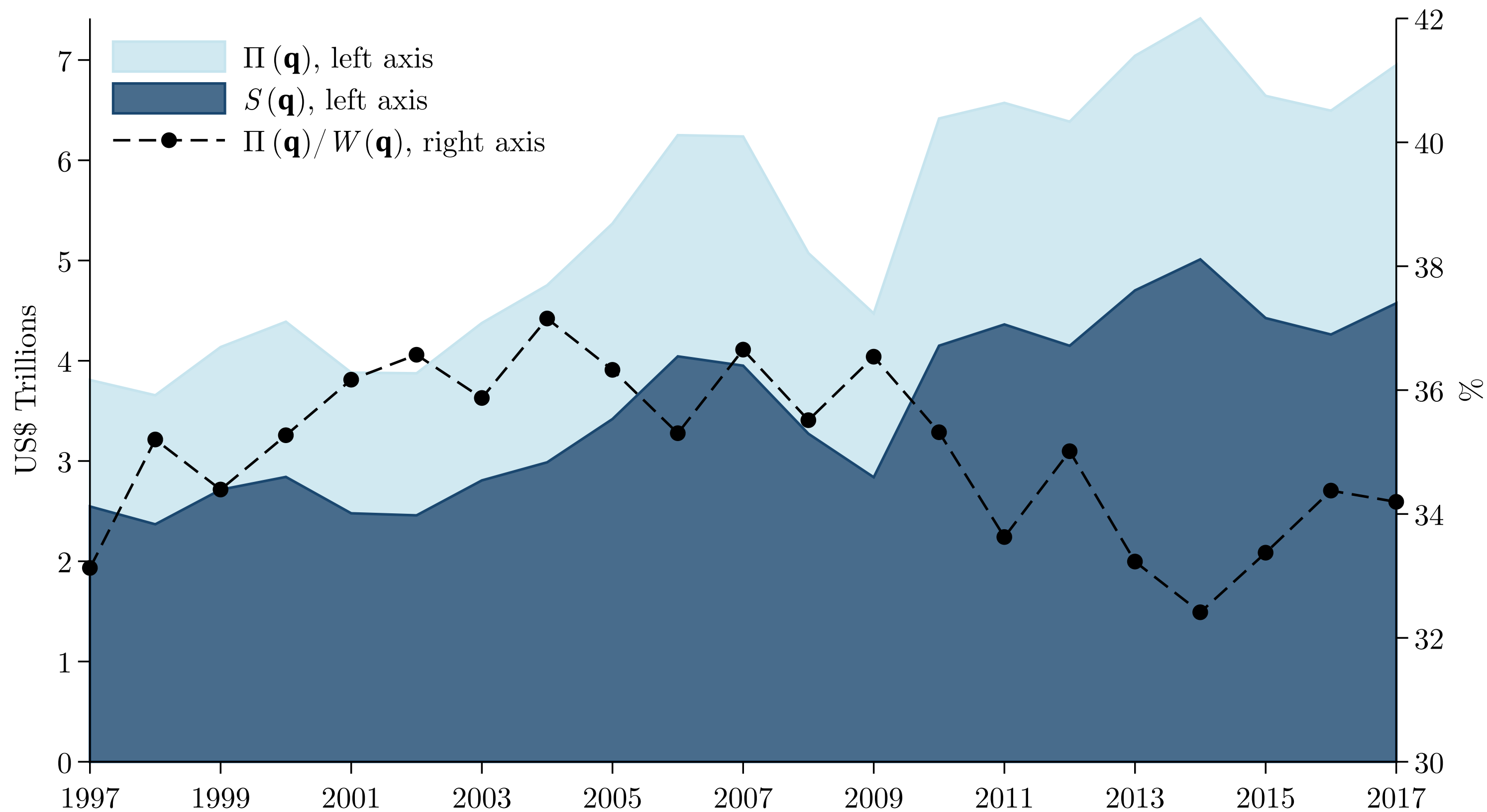


Figure 10: Welfare Costs of Oligopoly in the "Constant IPO rate" Scenario (1997-2017)

The following figure plots the evolution, between 1997 and 2017, of the deadweight loss from oligopoly—that is, the percentage gain in total surplus $\Delta\% W(\mathbf{q})$ from moving to the first-best allocation, in the counterfactual presented in Section 2, which assumes that the rate of IPOs as a percentage of VC exits would have stayed constant after 1996. The percentage gain in consumer surplus function $\Delta\% S(\mathbf{q})$ is shown on the right axis.

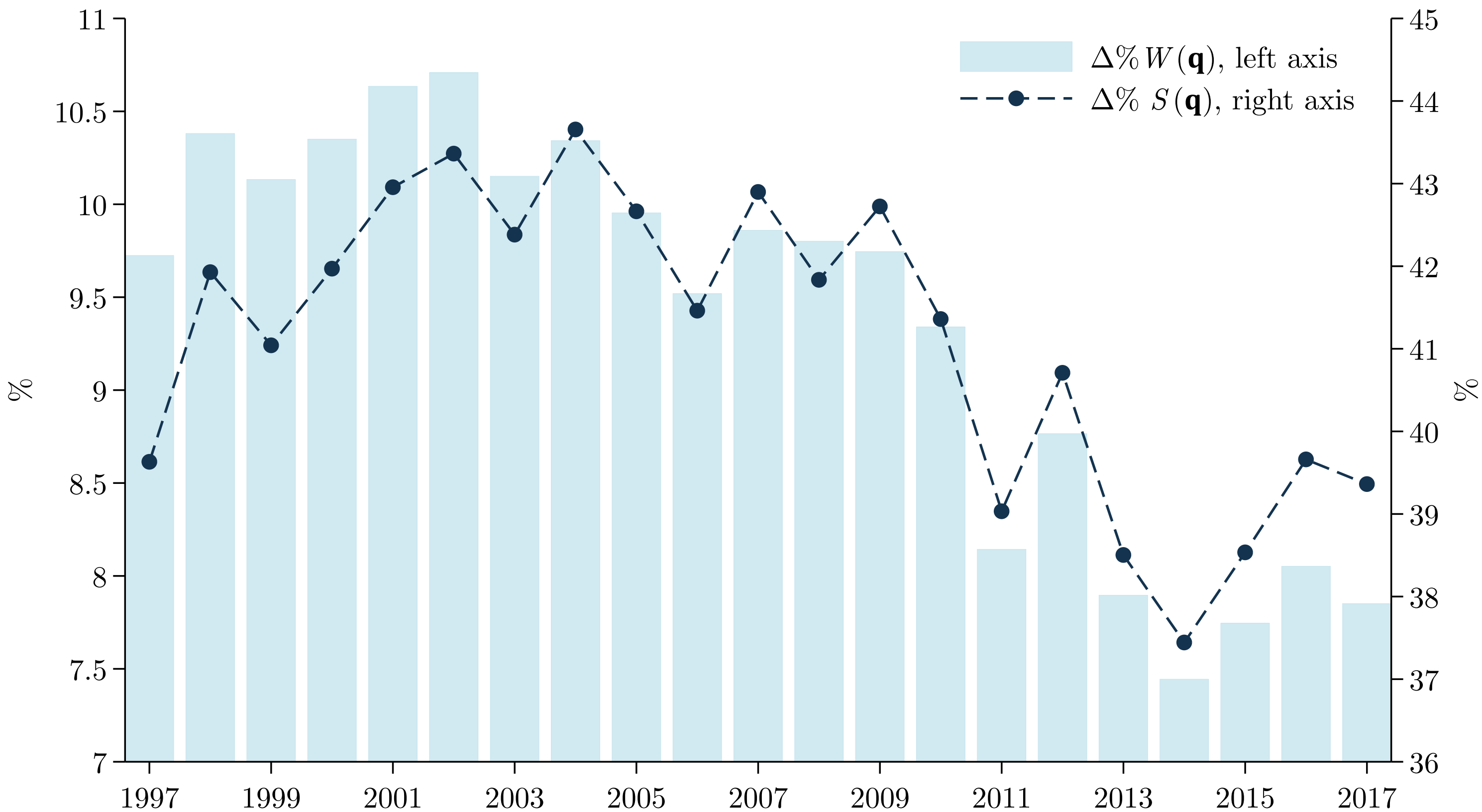


Figure 11: Cumulative Average Change in Inverse Centrality (2002-2017)

The following figure plots the evolution, between 2002 and 2017, of the Inverse Centrality Score χ_i, a firm-level metric of oligopoly power, for different groups of companies. This measure is distributed roughly uniformly over the interval $(0, 1)$. *Big 5 Tech* = {Alphabet, Amazon, Apple, Facebook, Microsoft}. *Other Tech* refers to GICs code 45. The computation of inverse centrality is based only on the firms' similarity in the product characteristics space (see model in section 2) and does not involve the use of any measure of firm size).

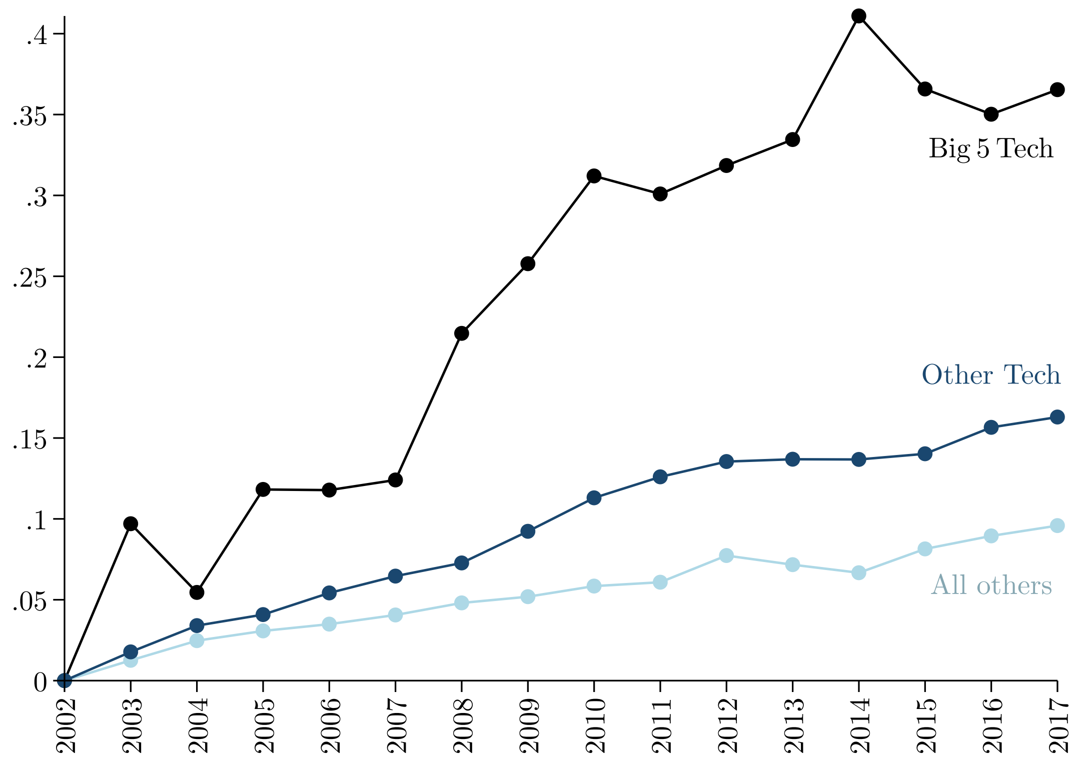


Table 1: Variable Definitions and Mapping to Data

The following table presents descriptions of the variables used in the estimation of the baseline model presented in section 2.

Panel A: Observed Variables

Notation	Observed Variable	Measurement
$\mathbf{k}$	*Enterprise Value*	Market value of Equity + Debt *source*: CRSP+Compustat
Π	*Aggregate profits*	Total sum of operating profits *source*: CRSP+Compustat
$\mathbf{A}'\mathbf{A}$	*Product similarity* for *Compustat* firms	Text similarity in 10-K product description *source*: Hoberg and Phillips (2016)
$\mathbf{A}'\mathbf{a}_0$	*Product similarity for VenturExpert firms*	Text similarity in product description *source*: Hoberg, Phillips, and Prabhala (2014)

Panel B: Model-Inferred Variables

Notation	Derived Variable	Computation/Identification
$r+\delta$	*Aggregate rate of return + Depreciation rate*	$=\Pi/K, \quad \text{where} \quad K \stackrel{\text{def}}{=} \sum_i k_i$
π	*(Expected) Profits*	$= (r+\delta)\,\mathbf{k}$
$\mathbf{q}$	*Output*	$= \sqrt{\pi}$
$\mathbf{Ab}-\mathbf{c}$	*Marginal Surplus at* $\mathbf{q}=\mathbf{0}$	$= (\mathbf{I}+\mathbf{A}'\mathbf{A})\,\mathbf{q}$

Table 2: Markups, Concentration, and Network Centrality: Regression Analysis

The following table presents OLS estimates for the following linear regression model:

$$\log \mu_{it} \;=\; \log\left(\theta_{s(i)} \frac{p_i q_i}{c_i q_i}\right) \;=\; \alpha_{s(i)} + \tau_t + \beta_1 \cdot \chi_{it} + \beta_2 \cdot \log k_{it} + \varepsilon_{it} \tag{82}$$

Where μ_{it} is the markup, χ_{it} is "Inverse Centrality" (computing using 10-K text data), k_{it} is the book value of assets, $\alpha_{s(i)}$ are sector fixed effects and τ_t are year fixed effects. Log markup is computed as the log difference between revenues and variable costs. The scale elasticity parameter is omitted from the markup calculation as it is absorbed by sector fixed effects. All specifications contain year fixed effects. Standard errors clustered at the firm level in parentheses: $^{*}p < .1; ^{**}p < .05; ^{***}p < .01$

Panel A: Variable Costs = Operating Costs

	(1)	(2)	(3)	(4)	(5)	(6)
Inverse Centrality (*standardized*)	0.086***	0.053***	0.038***	0.119***	0.093***	0.080***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
log Assets (*book value*)				0.116***	0.109***	0.112***
				(0.003)	(0.003)	(0.003)
R^2	0.202	0.252	0.299	0.289	0.322	0.365
Observations	102,426	102,423	102,395	102,426	102,423	102,395
Companies	12,413	12,410	12,382	12,413	12,410	12,382
Sector (NAICS) Fixed Effects	3-digit	4-digit	6-digit	3-digit	4-digit	6-digit

Panel B: Variable Costs = Costs of Goods Sold

	(1)	(2)	(3)	(4)	(5)	(6)
Inverse Centrality (*standardized*)	0.033***	0.025***	0.008	0.052***	0.048***	0.032***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
log Assets (*book value*)				0.064***	0.064***	0.066***
				(0.004)	(0.004)	(0.004)
R^2	0.171	0.200	0.252	0.190	0.217	0.268
Observations	102,426	102,423	102,395	102,426	102,423	102,395
Companies	12,413	12,410	12,382	12,413	12,410	12,382
Sector (NAICS) Fixed Effects	3-digit	4-digit	6-digit	3-digit	4-digit	6-digit

TABLE 3: WELFARE CALCULATIONS

The following table shows my estimates of aggregate profits, consumer surplus and total surplus of the counterfactuals presented in section 5. I use the short-hand W^* for the total surplus in the first-best allocation - that is, $W^* = W\left(\mathbf{q}^W\right)$.

	Scenario	*Cournot-Nash Eq.* (1)	*First-best* (2)	*Monopoly* (3)	*No-misallocation* (4)
Welfare Statistic	**Variable**	$\mathbf{q}^{\Phi}$	$\mathbf{q}^{W}$	$\mathbf{q}^{\Pi}$	$\mathbf{q}^{H}$
Total Surplus (US$ trillions)	$W(\mathbf{q})$	6.539	7.543	5.657	7.278
Aggregate Profits (US$ trillions)	$\Pi(\mathbf{q})$	2.837	0.000	3.771	0.896
Consumer Surplus (US$ trillions)	$S(\mathbf{q})$	3.702	7.543	1.886	6.382
Total Surplus / First Best	$\frac{W(\mathbf{q}^{\Phi})}{W^*}$	0.867	1.000	0.750	0.965
Aggregate Profit / Total Surplus	$\frac{\Pi(\mathbf{q}^{\Phi})}{W(\mathbf{q}^{\Phi})}$	0.434	0.000	0.667	0.123
Consumer Surplus / Total Surplus	$\frac{S(\mathbf{q}^{\Phi})}{W(\mathbf{q}^{\Phi})}$	0.566	1.000	0.333	0.877

Online Appendices for

Product Differentiation, Oligopoly and Resource Allocation

Bruno Pellegrino, UCLA

A. Derivation of the Cournot Potential

In the Network Cournot model, each firm i chooses its own output level to maximize its own profit by taking as given the output of every other firm:

$$q_i^* = \arg\max_{q_i} \pi\left(q_i; \bar{\mathbf{q}}_{-i}\right) \tag{83}$$

where $\overline{\mathbf{q}}_{-i}$ is the vector of output for every firm except i, treated as fixed by firm i. The first order condition for this problem is

$$0 = \mathbf{a}_i'\mathbf{b} - c_i - 2q_i - \sum_{j \neq i} \left(\mathbf{a}_i'\mathbf{a}_j\right) \bar{q}_j \tag{84}$$

which can be expressed, in vector form, as:

$$0 = \left(\mathbf{A}'\mathbf{b} - \mathbf{c}\right) - 2\mathbf{q} - \left(\mathbf{A}'\mathbf{A} - \mathbf{I}\right) \bar{\mathbf{q}} \tag{85}$$

This system of reaction functions defines a vector field $\mathbf{q}\left(\bar{\mathbf{q}}\right)$ which represents the firms' best response as a function of every other firms' strategy. To find the Cournot-Nash Equilibrium, we look for the fixed point $\mathbf{q}^*$ such that $\mathbf{q} = \bar{\mathbf{q}} = \mathbf{q}^*$. Plugging this inside the equation above yields the first order condition that is needed to maximize the potential function $\Phi\left(\mathbf{q}\right)$:

$$0 = \left(\mathbf{A}'\mathbf{b} - \mathbf{c}\right) - \left(\mathbf{I} + \mathbf{A}'\mathbf{A}\right) \mathbf{q}^* \tag{86}$$

which clarifies why the maximizer of the potential function solves the Network Cournot game. The potential $\Phi\left(\mathbf{q}\right)$ is then obtained as the solution to the following system of partial differential equations

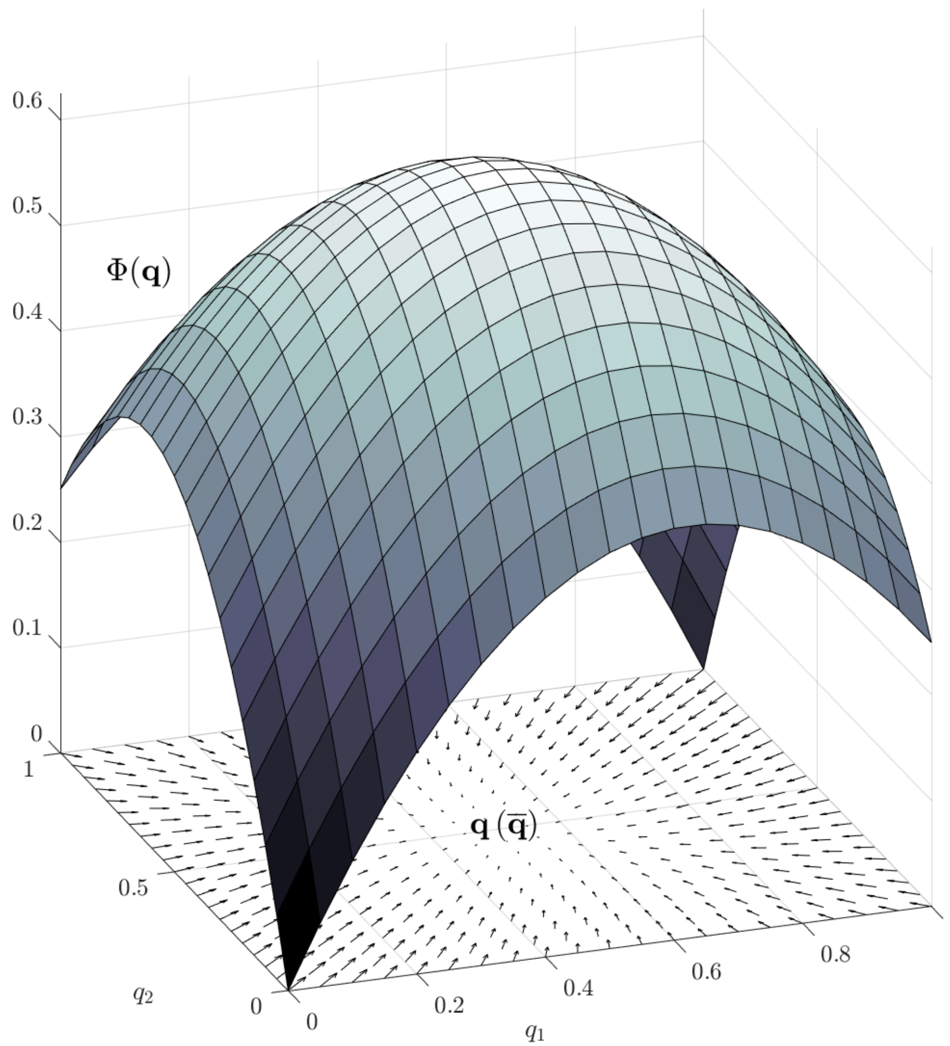
$$\nabla_{\mathbf{q}} \Phi\left(\mathbf{q}\right) \;=\; \left(\mathbf{A}'\mathbf{b} - \mathbf{c}\right) - \left(\mathbf{I} + \mathbf{A}'\mathbf{A}\right) \mathbf{q} \tag{87}$$

which equates the gradient of the potential function to the linear system of Cournot reaction functions.

The relationship between the potential function and the Cournot-Nash equilibrium is represented graphically, for the two-firm case, in Figure 12. The blue arrows represent the vector field defined by the firms' reaction functions. The potential function is defined to be the scalar-valued function whose gradient coincides with this vector field.

A game is a potential game if the vector field defined by the players' reaction functions is a conservative field - that is, if it is the gradient of some scalar function. We call that function the game's potential.

Figure 12: Graphing the Cournot Potential for the Two-firm Case



B. Nash-Cournot Equilibrium and Network Centrality

In this appendix, I explain in more detail the relationship between key metrics from the model described in Section 2 and the measures of network centrality developed by Katz (1953) and Bonacich (1987), which are widely used in the social networks literature.

The game played by the firms from Section 2 is a linear quadratic game played over a weighted network. Ballester, Calvó-Armengol, and Zenou (2006, henceforth BCZ) show that players' equilibrium actions and payoffs in this class of games depends on their centrality in the network.

In the game played the firms that populate by model, the adjacency matrix of the network over which the Cournot game is played, is given by the matrix $(\mathbf{I}-\mathbf{A}'\mathbf{A})$, which we find in the quadratic term of all the key welfare functions (equations 17,18,19).

Before discussing how the linkage extends to my model, I am going to formally define the metric of centrality.

Definition 8 (Katz-Bonacich Centrality)**.** For a weighted network with adjacency matrix $\mathbf{G}$, we define the centrality $\mathbf{f}$ with parameters $(\alpha, \mathbf{z})$ to be:

$$\begin{aligned} \mathbf{f}\left(\mathbf{G};\alpha,\mathbf{z}\right): \qquad \mathbf{f} &= \alpha\,\mathbf{G}\,\mathbf{f}+\mathbf{z} \\ &= \left(\mathbf{I}-\alpha\,\mathbf{G}\right)^{-1}\mathbf{z} \end{aligned} \tag{88}$$

Recursivity is the key property of this class of centrality indices: a node receives a higher centrality score the higher is the centrality of the nodes it is connected to.

The Nash-Bonacich linkage extends to my model: the Cournot-Nash equilibrium allocation of the model presented in Section 2 (equation 20) can be easily verified to coincide with the Katz-Bonacich centrality of the nodes in the network with adjacency matrix $(\mathbf{I}-\mathbf{A}'\mathbf{A})$, with parametrization $\left(\frac{1}{2},\frac{1}{2}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)\right)$:

$$\mathbf{q}^{\boldsymbol{\Phi}} \;\equiv\; \mathbf{f}\left(\mathbf{I}-\mathbf{A}'\mathbf{A};\,\frac{1}{2},\,\frac{1}{2}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)\right) \qquad . \tag{89}$$

The peculiarity of the Cournot game played by the firms in my model is that it is played over a *negatively-weighted* network. The consequence is that the interpretation of the centrality index is flipped (a higher centrality score actually reflects a more peripheral position—which is the reason I rename it *inverse* centrality) and the effect of centrality on the firm's strategy choice ($\mathbf{q}$) is reversed: firms grow larger and more profitable if they have a more *peripheral* position in the network.

Finally, the predicted markup from equation (51) is also a monotonic increasing function of the (inverse) centrality in the network of product similarities.

C. Calibrating λ and Testing the Model Using Patent Spillovers

C.1. Initial Calibration of λ

The parameter λ governs the relationship between the measures of product similarity of Hoberg and Phillips (2016) and the cross-price elasticity of demand in the model presented in Section 2. To calibrate λ, I target the microeconometric estimates of the cross-price demand elasticity from five empirical industrial organization studies. Table 4 lists the industry studies from which I sourced the cross-price elasticity data used for the calibration of λ.

For each of these studies (one per industry) I perform the following workflow: 1) I extract the matrix of cross-price elasticities; 2) I invert it to obtain the inverse cross price elasticities ($\partial \log \mathbf{p} / \partial \log \mathbf{q}$); 3) I match each of the off-diagonal elements to specific pairs of firms in Compustat; 4) for different values of λ, I compute the corresponding model estimate of the inverse cross-price elasticity based on HP's cosine similarity scores, and compare it to the microeconometric estimate obtained from the literature. I calibrate λ to the value that provides the best fit (in mean square error terms) between these two series.

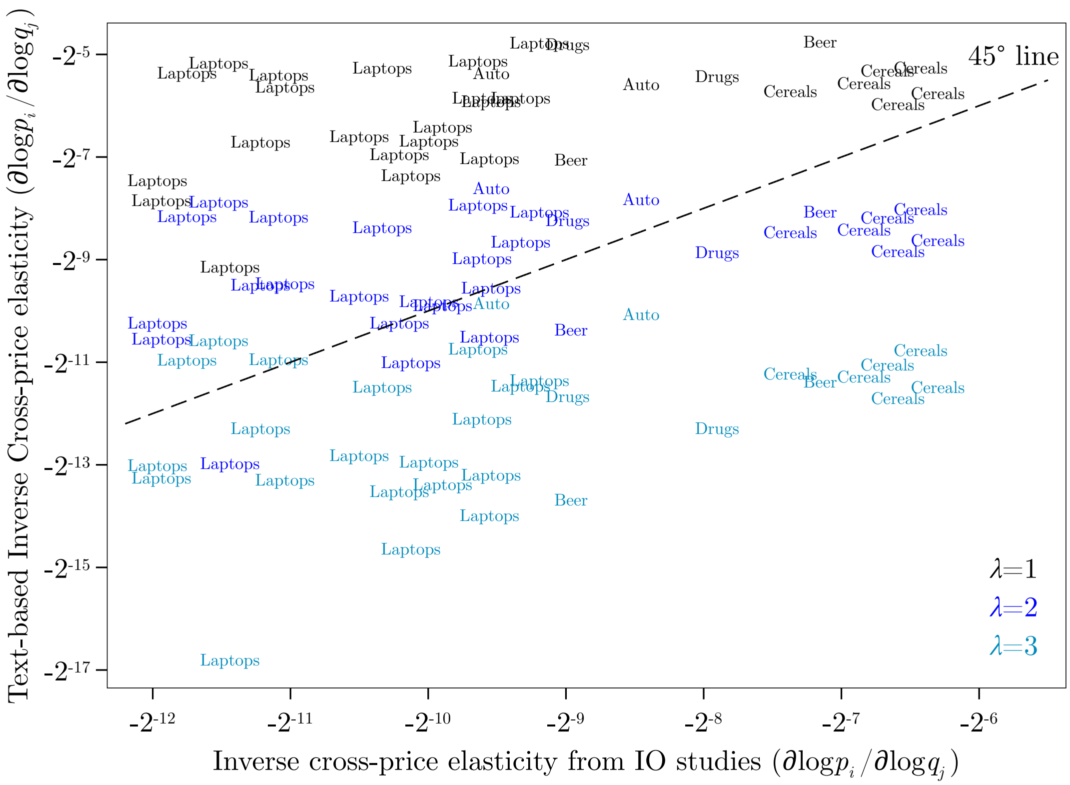
In Figure 13, I demonstrate graphically the calibration exercise: I plot the relationship between the model-based cross-price elasticity of demand (derived from HP's text-based similarity scores)—for different values of λ—against the respective microeconometric estimate. Different values of λ are represented by different colors. Visual inspection of the graph suggests that a reasonable value of λ should fall somewhere around 2. Using a distance-minimization algorithm, I obtain a value a value of 2.06. By repeating the same exercise for the model with nonconstant returns to scale (see section 7) I obtain a value of 2.46. I settle for the intermediate value 2.26.

Table 4: List of Industry Studies Used for the Calibration Exercise

Industry	Citation	Journal/Series
Auto	Berry, Levinsohn, and Pakes (1995)	Econometrica
Beer	De Loecker and Scott (2016)	NBER Working Papers
Cereals	Nevo (2001)	Econometrica
Drugs	Chintagunta (2002)	Journal of Marketing Research
Laptops	Goeree (2008)	Econometrica

While aligning my estimates of the demand elasticity to match microeconometric estimates is an important first step, given the importance of the parameter λ, we want an extra degree of confidence in this calibration. For this reason, I test the robustness of this estimate using patents and stock market data.

Figure 13: Model Calibration using cross-price elasticities from Empirical IO



C.2. Patents market value data

To empirically validate my model, I use a measure of the cumulative dollar value of the patents issued, in a year, to each firm in Compustat. The dollar value of each individual patent is measured by gauging the stock market reaction to the issuance of the patent.

This database was introduced in a recent paper by Kogan, Papanikolaou, Seru, and Stoffman (2017, henceforth KPSS), who have carried out this measurement for all patents from Google's patent database that they were able to match to firms from the CRSP data set. Like that of Hoberg and Phillips (2016), this data set is also made publicly-available by the authors: it contains, for the period 1996–2010, a total of 1,560,032 patents that can be matched to CRSP-Compustat companies.

Let k_{it} be the total value of company i at time t, and let Δk_{it} be the 1-year change in the company's value. Now, consider breaking down the 1-year change in company i's enterprise valuation into continuous increments:

$$\Delta k_{it} = \int_{t-1}^{t} dk_{iu} \qquad u \in [t-1, t] \tag{90}$$

Let ϑ_{it} be an indicator of firm i being issued a patent at instant u. We can then obtain θ_{it}, the cumulative impact of all patents issued to i in the interval $[t-1, t]$ on the company's market value by integrating over the jumps:

$$\theta_{it} = \int_{t-1}^{t} \vartheta_{iu} \, dk_{iu} \qquad u \in [t-1, t] \tag{91}$$

In practice, there are two issues that need to be accounted for in implementing this measurement. First, the instantaneous return may be either unobservable, or the stock market's reaction to the patent issuance may be less-than-instantaneous; as a consequence KPSS look at idiosyncratic 3-day returns around the patent issuance date. Second, the market is likely to have prior knowledge of the patent at the time the patent is granted. KPSS assume that, given an unconditional probability $\overline{\pi} = 56\%$ of the patent being issued, only a share $(1 - \overline{\pi}) = 44\%$ of the value of the patent is realized at issuance date. I follow this assumption, and make the additional assumption that the remaining 56% of the value of the patent is realized on the application date (the only other date that is consistently available in their data set). Because I do not want my assumptions on the exact timing of the patent value realization to be a source of concern, I perform the empirical exercises that follow at the year level.

Using KPSS' data, I compute θ_{it} for 65,415 firm-year observations. Next, I describe how to use this variable to construct shifters for individual companies' output and test the model's quantitative predictions.

C.3. Testing the Model using patent spillovers

Given the importance of the parameter λ as an input in my aggregate measures of market power, it is desirable to test whether the model — thus calibrated — yields accurate predictions.

To be more explicit, consider the equilibrium equation for the firms' output vector $\mathbf{q}$:

$$\mathbf{q} \; = \; \left(\mathbf{I} + \mathbf{A}'\mathbf{A}\right)^{-1} \left(\mathbf{A}'\mathbf{b} - \mathbf{c}\right) \tag{92}$$

the vector $(\mathbf{A}'\mathbf{b} - \mathbf{c})$ captures idiosyncratic shifts in each firm's residual demand and supply. Because firms interact strategically, a shift in (say) firm i's marginal cost c_i will shift firm j's residual demand curve. The magnitude of the shift will depend on the degree of strategic interaction between all firms, as captured by the matrix $\mathbf{A}'\mathbf{A}$. For simplicity, let us define:

$$\Xi \; = \; [\xi_{ij}] \; = \; \left(\mathbf{I} + \mathbf{A}'\mathbf{A}\right)^{-1} \tag{93}$$

If c_i shifts downwards by one unit, i's equilibrium output increases by ξ_{ii} and j's equilibrium output shifts by ξ_{ij} . Because ξ_{ij} depends on the calibrated value of λ, we can verify that this calibration is sensible by estimating a regression of output quantity on an appropriate cost shifter, and checking whether the regression yields the expected coefficient.

Because my metric of output q is derived by the firm's market value k, we can use KPSS's patent valuations as shifters. I impose the following identifying assumption: a patent announcement for firm i only shifts i's marginal cost c_i, and not i's similarity to other firms $\mathbf{a}_i$. The reason why this assumption makes sense given the data, is that there is nearly no variation from year to year in similarity scores. The reason is that 10-K forms change slowly over time, and are highly unlikely to incorporate variation from individual patents.

To see how we can use KPSS' patent data to construct shifters, recall that the value θ_{it} of the patents granted to i during period t is simply the variation in k_i induced by the patents, and that there is a bivariate mapping linking q_{it} to k_{it} . Keeping everything else constant, we then have the following change in q_{it} as a consequence of all the patents granted to i and to its competitors at time t:

$$dq_{it} \;=\; \frac{\partial q_{it}}{\partial k_{it}} \cdot \theta_{it} \;+\; \sum_{j \neq i} \frac{\xi_{ij}}{\xi_{ii}} \cdot \frac{\partial q_{jt}}{\partial k_{jt}} \cdot \theta_{jt} \tag{94}$$

Because $q_{it} = \sqrt{(r+\delta)\, k_{it}}$, we can rewrite the equation above by taking the derivative of q with respect to k :

$$\frac{\partial q_{it}}{\partial k_{it}} \;=\; \frac{q_{it}}{2k_{it}} \tag{95}$$

We can now define the following variables. The first, which I call *Patents*, is simply the change in q_i due to all patents issued to i during period t:

$$\text{Patents}_{it} \overset{\text{def}}{=} \frac{q_{it}}{2k_{it}} \cdot \theta_{it} \tag{96}$$

The second, which I call *Rivalry*, is the change in q_i due to all patents issued to i's competitors $j \neq i$ during period t:

$$\text{Rivalry}_{it} \overset{\text{def}}{=} \sum_{j \neq i} \frac{\xi_{ij}}{\xi_{ii}} \cdot \frac{q_{it}}{2k_{it}} \cdot \theta_{it} \tag{97}$$

Because different values of λ result in different values of Σ, the variable *Rivalry* is a function of the calibrated value of λ. Hence, we can run the following regression:

$$dq_i \;=\; \alpha_i + \tau_t + \delta \cdot \text{Rivalry}_{it} + \gamma \cdot \text{Patents}_{it} + \varepsilon_{it} \tag{98}$$

where α_i is a firm fixed effect and τ_t is a year fixed effect, and look at the regression coefficients for different values of λ. If the model is correctly specified and calibrated, we should expect this regression to yield:

$$\gamma \;=\; \delta \;=\; 1 \tag{99}$$

One potential concern of regressing firm size on the variable *Rivalry* (thus defined) is that the resulting regression coefficient might conflate the effect of rivalry with that of technology spillovers. To be more specific, suppose that firm j is granted a patent that increases its market value. This in turn might have two directionally opposite effects on the size of firm i: 1) a negative one, due to the fact that an increase in the supply of j's product(s) decreases the demand for i's output if the two firms produce substitute products; 2) a positive one, if the two firms use the same technology and i can benefit from R&D spillovers. In other words, the regression might suffer from an omitted variable problem if there are R&D spillovers.

Bloom, Schankerman, and Van Reenen (2013, henceforth BSV) proposed a method to disentangle these effects. The idea is to compute a variable, *TechSpillovers*, that is identical to *Rivalry* except for the fact we use the cosine similarity in the technology space $\mathbf{t}_i'\mathbf{t}_j$ to average across competitors:

$$\text{TechSpill}_{it} \overset{\text{def}}{=} \sum_{j \neq i} \left(\mathbf{t}_i'\mathbf{t}_j\right) \frac{q_j}{2k_j} \theta_{jt} \tag{100}$$

The vector $\mathbf{t}_j$ counts, for each company j, the patents obtained in different patent classifications. Like $\mathbf{a}_i$, $\mathbf{t}_i$ is also normalized to be unit-valued and is used to construct a measure of similarity in the technology space.

One final concern about this specification is that firms could be engaging in R&D to attain product innovation—that is, in order change their position in the product space. The coefficient on *Rivalry* could be biased if the cosine similarity scores are themselves affected by product innovation. To account for this possibility, I add one further control variable (which I obtain from the data set of Hoberg, Phillips, and Prabhala, 2014) to the regression above: the variable is called *AngularVelocity*, and is defined as the arccosine of the similarity between $\mathbf{a}_{it}$ (firm i's coordinate vector in year t) and $\mathbf{a}_{it-1}$ (i's coordinate vector in the previous year) – formally:

$$\text{AngularVelocity}_{it} \;=\; \text{acos}\left(\mathbf{a}_{it}'\mathbf{a}_{it-1}\right) \tag{101}$$

Armed with these four variables, we can then run the following fixed-effects panel regression

$$\Delta q_{it} = \alpha_i + \tau_t + \delta \cdot \text{Rivalry}_{it} + \rho \cdot \text{Patents}_{it} + \gamma \cdot \text{TechSpill}_{it} + \varsigma \cdot \text{AngularVelocity}_{it} + \varepsilon_{it} \tag{102}$$

To account for the obvious scale effects in the specification, and to correct for the resulting heteroskedasticity, we want to run this regression in percentage changes rather than level changes. I do so, without changing the variable definitions, by performing a weighted least squares estimation (WLS) and weighting each observation by the inverse of the lagged value of q_{it}, so that the left-hand side variable effectively becomes $\Delta q_{it}/q_{it-1}$.

The estimation results for this regression are displayed in Table 5. Columns 1–3 show the estimation of equation (98) for calibrated values of $\lambda = 1,2,3$ respectively. Columns 4–5 show regression results for $\lambda = 2$, with the added controls *TechSpillovers* and *AngularVelocity*; in column 5, *TechSpillovers* is lagged. The reason I include this specification is to account for the potential ambiguity in the timing of the stock market reaction to technology spillovers. If, at the moment the patent is issued, the stock market is able to foresee a technology spillover to a rival firm, then the contemporaneous value of the variable *TechSpillovers* should be included. Otherwise, it would make sense to use the first lag (as do BSV).

Across all specifications, the regression coefficient for *Patents* is statistically significant and close to one, in line with the predicted value. In all specifications, I fail to reject the null hypothesis that the coefficient is one. In column 1, where a blue of $\lambda = 1$ is used to compute the variable *Rivalry*, the regression coefficient for this variable is statistically significant at the 10% level but notably lower than one (0.003), suggesting that calibrating λ to one would lead to an over-estimation of the magnitude of the cross price elasticities. In column 2 and 3, the corresponding estimate is, respectively 0.78 and 1.47: this is in line with our previously-calibrated value of λ (2.26). Reassuringly, across all specifications (even when λ is calibrated to a low value of 1), the effect of *Rivalry* remains positive and statistically significant (at the 1% level, with the exception of column 1), as the model would predict. The estimated coefficient for *TechSpillovers,* in column 4, is -0.002. This estimate is statistically significant, but the sign is opposite to that predicted; the corresponding estimate for *Rivalry* is 0.547, somewhat lower than the value obtained in column 2. When, in column 5, we lag the control *TechSpillovers,* the sign of the coefficient for this variable flips, and the corresponding estimate for *Rivalry* approaches one. The estimated coefficient for *AngularVelocity* is equal to 0.27 and statistically significant, at the 1% level.

The results of this regression analysis support calibrating λ to a value slightly above 2, which is (reassuringly) in line with the value of 2.26 previously obtained.

TABLE 5: TESTING THE MODEL USING PATENT SPILLOVERS

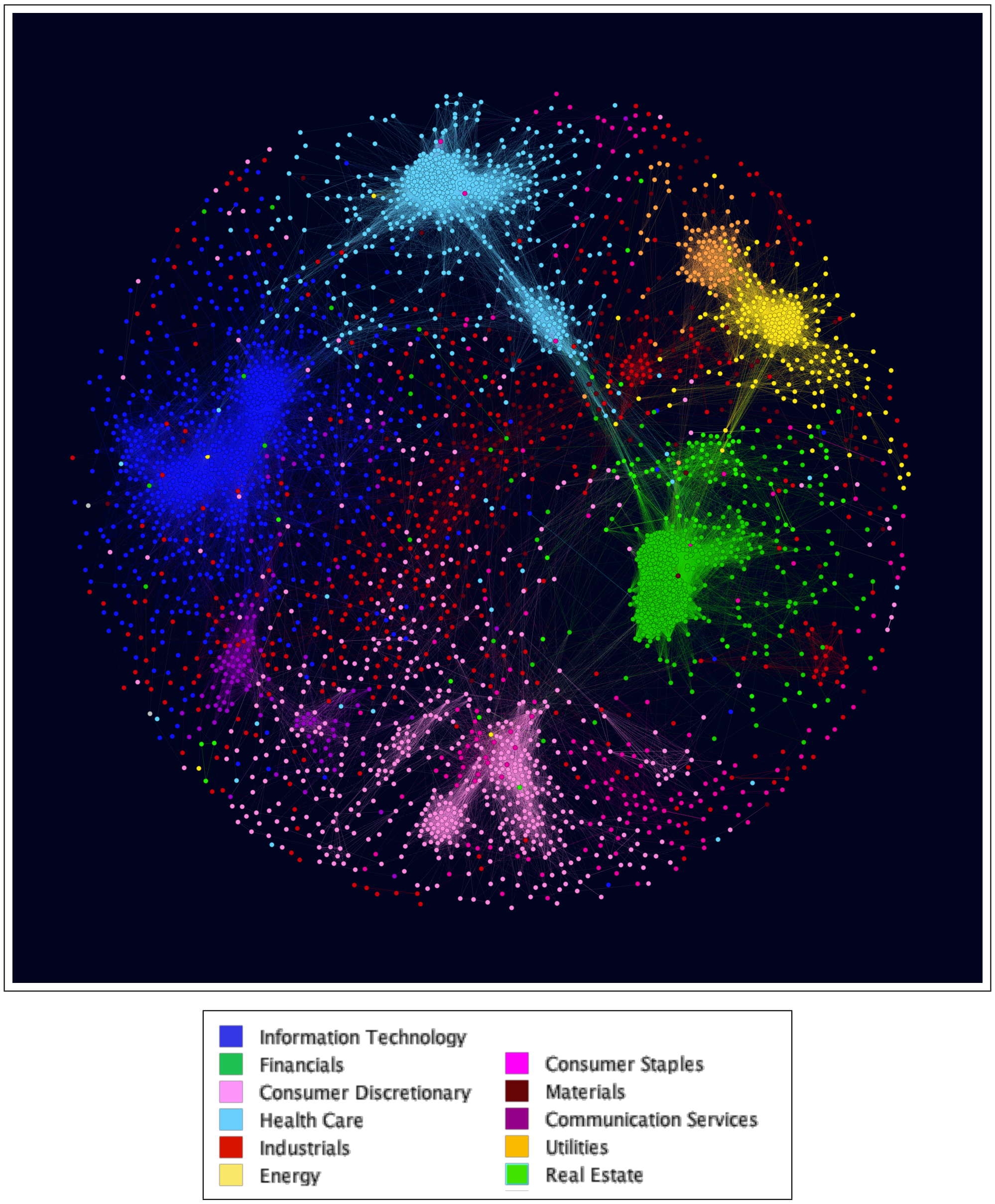
Weighted Least Squares Regression (WLS)		(1)	(2)	(3)	(4)	(5)
Dependent variable: Δq_{it} ; **weight**: $\frac{1}{q_{it-1}}$		$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 2$	$\lambda = 2$
Rivalry	(*predicted value* $= 1$)	0.002*	0.898***	2.144***	0.716***	1.195***
		(0.001)	(0.138)	(0.387)	(0.146)	(0.143)
Patents	(*predicted value* $= 1$)	1.083***	1.135***	1.133***	1.173***	1.115***
		(0.128)	(0.130)	(0.130)	(0.133)	(0.131)
TechSpillovers	(*predicted sign*: ≥ 0)				-0.002***	0.002***
					(0.000)	(0.000)
AngularVelocity	(*predicted sign*: ?)				0.269***	0.266***
					(0.018)	(0.018)
R^2		0.232	0.233	0.233	0.239	0.239
Observations		63,650	63,650	63,650	62,675	62,675
Companies (Clusters)		8,596	8,596	8,596	8,535	8,535
TechSpillovers is lagged		–	–	–	No	Yes
Company Fixed Effects		Yes	Yes	Yes	Yes	Yes
Year Fixed Effects		Yes	Yes	Yes	Yes	Yes

Cluster-robust standard errors in parentheses: $^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

D. Independent validation of text-based product similarity scores

In this appendix, I validate independently the text-based product similarity measures of Hoberg and Phillips (2016). In the figure below I produce a graph similar to that of Figure 3, while coloring different nodes according to the respective firm's GIC economic sector. The figure shows that there is significant overlap between the macro clusters of the network of product similarity and the broad GIC sectors. To produce this visualization, the dimensionality of data has been reduced from 61,000 to 2; yet, the overlap is nonetheless very clearly visible.

Figure 14: Visualization of the product space (alternate coloring)



E. Active fiscal policy

I now wish to move to the question of what can a social planner do to rectify the allocative distortions induced by imperfect competition. Unsurprisingly, the level of total surplus that can be attained depends on what constraints are placed on the social planner. In what follows, I assume that the planner can impose a vector of per-unit taxes (or subsidies, if negative) $\mathbf{t}$, so that the net marginal cost imposed on the producer is now $(\mathbf{c}+\mathbf{t})$ and the vector of firm profits is now given by

$$\pi(\mathbf{q}) = \mathbf{Q}(\mathbf{A}'\mathbf{b}-\mathbf{c}-\mathbf{t}) - \frac{1}{2}\mathbf{Q}\mathbf{A}'\mathbf{A}\mathbf{q} \quad . \tag{103}$$

Because, after the wedges are applied, the Cournot-Nash equilibrium moves to:

$$\mathbf{q}^{\mathbf{t}} = (\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}(\mathbf{A}'\mathbf{b}-\mathbf{c}-\mathbf{t}) \tag{104}$$

This implies that, if the social planner desires to implement allocation $\hat{\mathbf{q}}$, it has to impose the following wedge vector:

$$\mathbf{t}(\hat{\mathbf{q}}) = (\mathbf{A}'\mathbf{b}-\mathbf{c}) - (\mathbf{I}+\mathbf{A}'\mathbf{A})\hat{\mathbf{q}} \tag{105}$$

Another metric of welfare that is useful to keep track of is the government budget surplus/deficit that is generated by the unit taxes and subsidies imposed by the social planner to enforce a certain allocation.

Definition 9 (Government budget function)**.** We define the government budget, as a function of allocation $\mathbf{q}$, as the dot product of the allocation and of the corresponding wedge vector required to implement it:

$$G(\mathbf{q}) = \mathbf{q}'\mathbf{t}(\mathbf{q}) \tag{106}$$

The budget deficit to implement the first best—that is, $G\left(\mathbf{q}^W\right)$ — is negative. In other words, to implement the social optimum, the planner has to run a deficit — unless, of course, she can impose flat transfers on individual firms. For this reason, it is interesting to look more broadly at a set of allocations that maximize aggregate welfare subject to an upper cap to the government budget $G(\cdot)$: we call these allocations *budget-efficient.*

Definition 10 (Budget-efficient allocations)**.** We say an allocation $\hat{\mathbf{q}}$ is budget-efficient if it solves the following constrained social planner problem, for some $\bar{G}$:

$$\hat{\mathbf{q}} = \arg\max_{\mathbf{q}} W(\mathbf{q}) \qquad \text{s.t.} \qquad G(\mathbf{q}) \geq \overline{G} \qquad . \tag{107}$$

Proposition 6. *The set of budget-efficient allocations takes the following form:*

$$\mathbf{q} = \left(1-\frac{\nu}{2}\right)(\nu\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}(\mathbf{A}'\mathbf{b}-\mathbf{c}) \tag{108}$$

where ν is an increasing function of the budget cap $\bar{G}$.

Proof. See Appendix R. $\square$

Although λ depends nonlinearly on the budget cap $\overline{G}$ and does not appear to have a closed formula, it can be solved for numerically. In addition, in 5.4, I am going to provide a geometric characterization of the set of budget-efficient allocations.

One of the mechanisms that drive factor misallocation in this model is that large firms have more market power (in the sense that they face a less elastic residual demand). As a consequence, they are able to charge higher markups. Therefore, a fiscal policy counterfactual that is interesting to analyze is to let the social planner impose a size-dependent tax, so that:

$$\pi(\mathbf{q}) \;=\; \mathbf{Q}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}-\tau\mathbf{q}\right)-\frac{1}{2}\mathbf{Q}\mathbf{A}'\mathbf{A}\mathbf{q} \tag{109}$$

where τ is a scalar. The Cournot-Nash equilibrium then moves to:

$$\mathbf{q}^{\tau} \;=\; \left[(1+\tau)\,\mathbf{I}+\mathbf{A}'\mathbf{A}\right]^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right) \tag{110}$$

It is easy to see that a tax that impacts disproportionately large firms exacerbates allocative distortions rather than alleviating them.

F. Plotting the welfare function

I now want to consider how the aggregate surplus functions $W(\cdot)$ and $S(\cdot)$ vary over a set of reasonable counterfactual allocations. The problem with trying to visualize the welfare function $W(\cdot)$ is that $\mathbf{q}$ exists in a space of several thousand dimensions (one for each firm). Some degree of dimensionality reduction is necessary to study how the the aggregate welfare changes in response to the allocation $\mathbf{q}$.

To plot the welfare function, I restrict my attention to a bidimensional set of counterfactuals that can be obtained as linear combinations of the allocations $\mathbf{q}^{\Phi}$, $\mathbf{q}^{W}$ and $\mathbf{q}^{\Pi}$. This set of allocations is indexed by two parameters, $\alpha, \beta \in [0,1]$:

$$\mathbf{q} = \left(1 - \frac{\alpha}{2}\right)\left(\mathbf{I}\beta + \mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b} - \mathbf{c}\right) \tag{111}$$

When $\alpha = 0$ and $\beta = 0$, the equation above yields the first-best allocation $\mathbf{q}^{W}$; when $\alpha = 0$ and $\beta = 1$, it yields to the Cournot-Nash equilibrium which we observe in the data; finally, when $\alpha = 1$ and $\beta = 0$, the it yields the *Monopoly* counterfactual.

Eyeballing equation (108), we can see that one property of the sub-space of allocations indexed by (α, β), is that the set of budget-efficient allocations, which we previously characterized as a function of the Lagrange multiplier λ, is arranged on the 45° line—that is, it is the set of allocations such that:

$$\lambda \;=\; \alpha \;=\; \beta, \tag{112}$$

where the specific value of λ depends positively on the budget cap.

Figures 15 and 16 display, respectively, the total surplus function $W(\mathbf{q})$ and the consumer surplus function $S(\mathbf{q})$ over the space of allocations indexed by (α, β). Both total welfare and consumer welfare decrease as we move away from the first-best allocation, although they do so faster as we move along the horizontal axis. The main difference between the total surplus $W(\cdot)$ and the consumer surplus $S(\cdot)$ is that the latter is significantly steeper in the neighborhood around the first-best: this reflects the fact that, as we gradually eliminate markups, not only total surplus increases, but the consumer's share also increases.

Figure 15: Aggregate total surplus for counterfactual allocations

The following diagram plots the aggregate total surplus function $W(\mathbf{q})$, as a heat map, over a space of counterfactual allocations that are indexed by the parameters α and β

$$\mathbf{q} = \left(1 - \frac{\alpha}{2}\right)\left(\mathbf{I}\beta + \mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b} - \mathbf{c}\right)$$

These allocations can be computed as linear combinations of the three fundamental output vectors $\mathbf{q}^{\Phi}$ (the observed Cournot-Nash equilibrium), $\mathbf{q}^{\Pi}$ (the aggregate profit maximizer) and $\mathbf{q}^{W}$ (the social optimum). The aggregate total surplus $W(\mathbf{q})$ is measured in trillions of U.S. dollars.

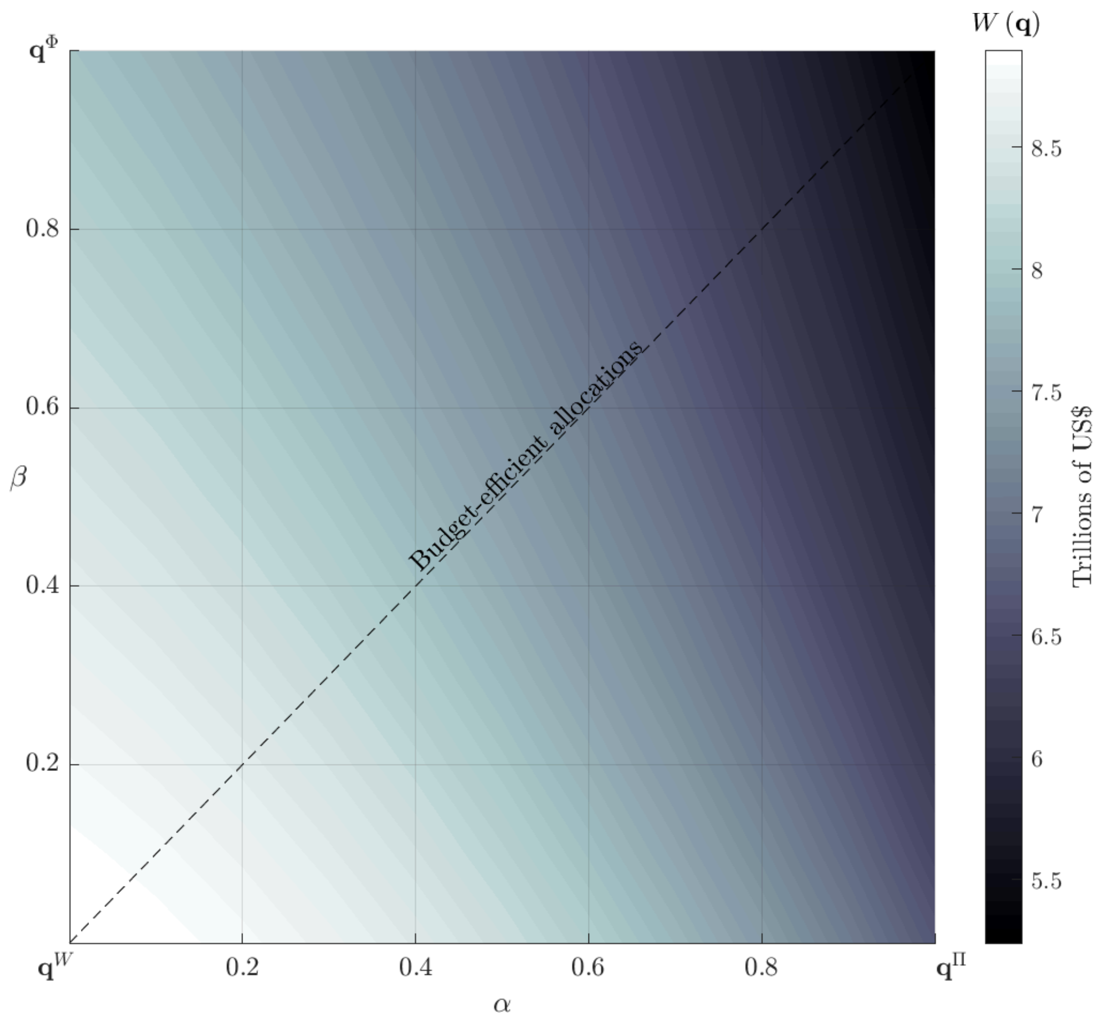
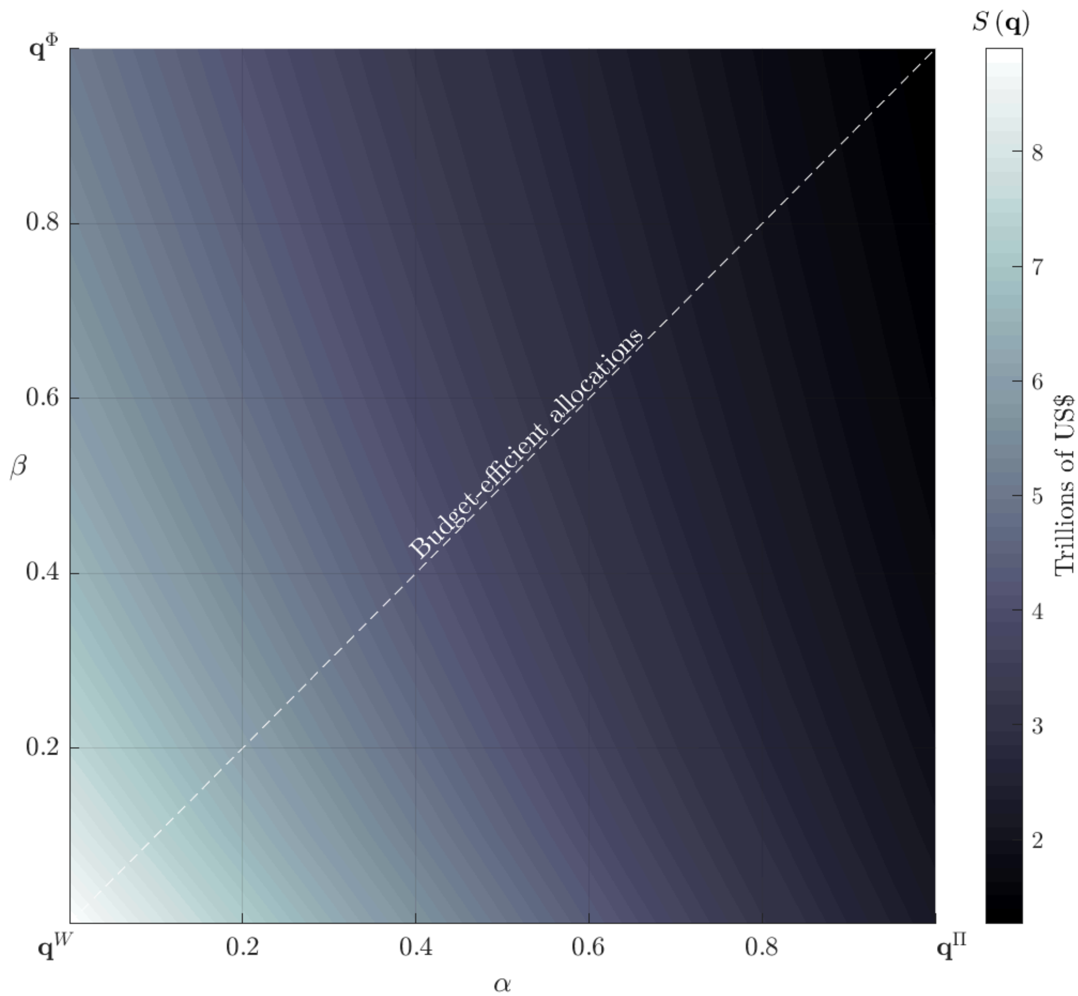


Figure 16: Aggregate consumer surplus for counterfactual allocations

The following diagram plots the aggregate consumer surplus function $S(\mathbf{q})$, as a heat map, over a space of counterfactual allocations that are indexed by the parameters α and β

$$\mathbf{q} = \left(1 - \frac{\alpha}{2}\right)\left(\mathbf{I}\beta + \mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b} - \mathbf{c}\right)$$

These allocations can be computed as linear combinations of the three fundamental output vectors $\mathbf{q}^{\Phi}$ (the observed Cournot-Nash equilibrium), $\mathbf{q}^{\Pi}$ (the aggregate profit maximizer) and $\mathbf{q}^{W}$ (the social optimum). The aggregate consumer surplus $S(\mathbf{q})$ is measured in trillions of U.S. dollars.



G. Why is most of the deadweight loss linked to misallocation?

In this Appendix I further discuss my finding that most of the deadweight losses (10 out of 13 percent of total potential surplus) can be recovered without changing the labor supply. This finding may look surprising without further context, given that there is a linear disutility for labor (high labor supply elasticity). To rationalize this finding I rely on the theoretical work of Baqaee and Farhi (2019, henceforth BF).

The first thing to clarify is that what is measured are changes in welfare — not in GDP. To see why the supply of labor supply changes significantly when we move from $\mathbf{q}^W$ to $\mathbf{q}^H$ while welfare doesn't, define the change in GDP as the revenue-weighted change in real output:

$$\Delta \log Y \stackrel{\text{def}}{=} \sum_i \frac{p_i q_i}{\mathbf{p}'\mathbf{q}} \Delta \log q_i \tag{113}$$

Define also the change in aggregate Total Factor Productivity (TFP) as:

$$\Delta \log \text{TFP} \stackrel{\text{def}}{=} \Delta \log Y - \Delta \log H \tag{114}$$

BF show that a change in TFP around any allocation with uniform markups (including the first-best and the misallocation-free counterfactual) can be approximated nonparametrically — to a second order — by:

$$\Delta \log\text{TFP} \approx \frac{1}{2} \sum_i \frac{p_i q_i}{Y} \Delta \log q_i \, \Delta \log \mu_i \tag{115}$$

They also show that a change in Total Surplus around the first best allocation $\mathbf{q}^W$ can be approximated (again, to a second order) by:

$$\Delta \log W \approx \frac{1}{2} \sum_i \frac{p_i q_i}{W} \Delta \log q_i \, \Delta \log \mu_i \tag{116}$$

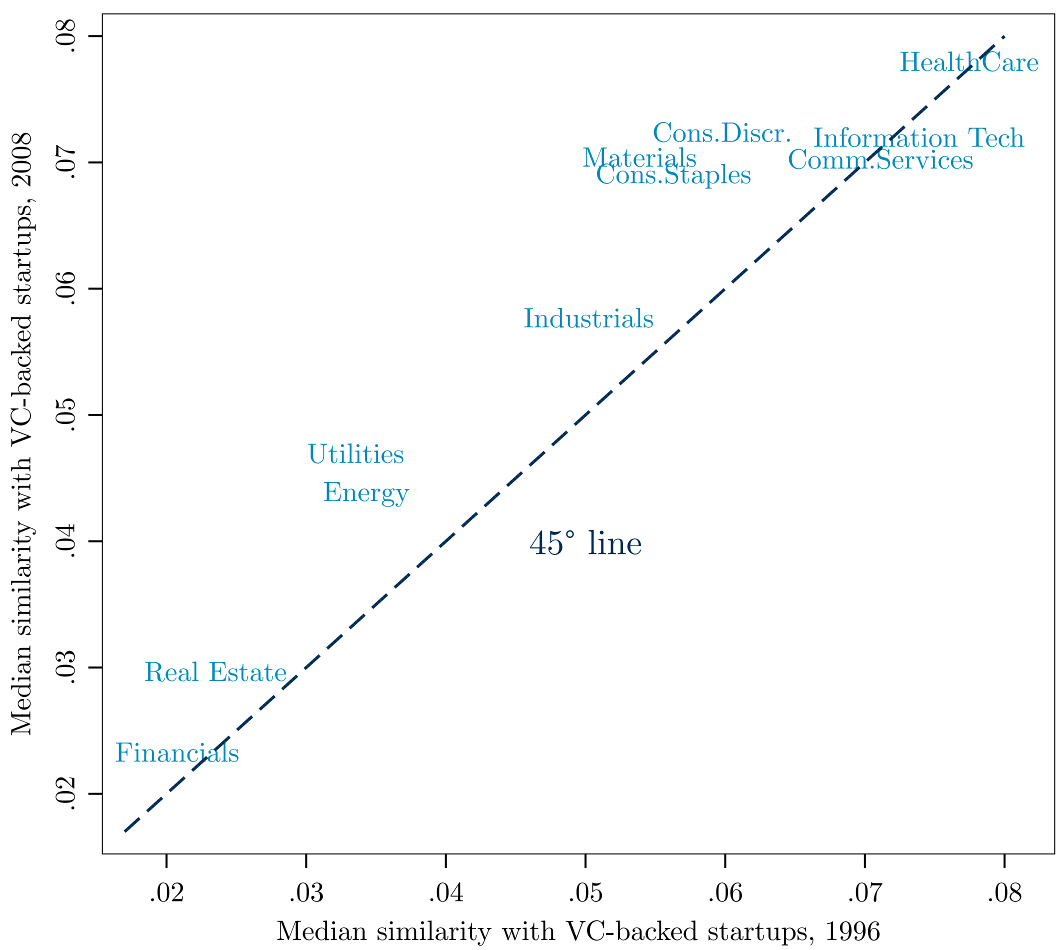
Equations (115) and (116), as well as the lack of first-order effects are all a direct consequence of two envelope conditions: the first-best allocation maximizes welfare, while any allocation with uniform markups maximizes TFP. The two equations are identical except for the denominator. BF show that this approximation holds *regardless* of the functional form of the representative agent's utility and regardless of the elasticity of labor supply. Now, consider moving from the first best $\mathbf{q}^W$ to the no-misallocation counterfactual $\mathbf{q}^H$. Because we moved from one constant-markup allocation to another (we have raised markups symmetrically), the change in (115) is zero. But this implies that the change in welfare—equation (116) — must *also* be approximately zero. This is true not just to a *first* degree approximation but also to a *second* order approximation, and is due to the fact that we are moving from the first best to another allocation with uniform markups.

The intuition behind this finding is that, at the first-best allocation, the optimal consumption bundle is (locally) a perfect substitute for leisure. Hence, when we introduce a uniform markup (for example, when we move from the first-best to the no-misallocation counterfactual), the consumer is unaffected because she can mantain approximately the same level of welfare by substituting its overall consumption bundle with leisure while leaving the composition of the consumption bundle unaffected. This is not the case when we introduce a non-uniform markup – which causes misallocation – because in this instance the average consumption bundle is no longer a perfect substitute for leisure.

H. No changes in the sectoral composition of VC activity

In this appendix I present evidence that the change in the number of IPOs as a percentage of successful venture capital exits was not driven by a change in the sectoral distribution of VC-backed startups: there has hardly been any in the years in which the IPO decline occurred. For the years 1996 and 2008 and for every GIC sector level, I take the median score measuring the 10-K similarity among Compustat companies in that sector to the set of VC-backed startups from the VenturExpert database that have been financed in that year. Hoberg, Phillips, and Prabhala (2014, henceforth HPP) extend the similarity scores to private firms using the VenturExpert product descriptions in place of the 10-K product description. The graph shows that the sectoral composition of VC startups has barely changed between 1996 and 2008 (the first and last years covered by HPP).

FIGURE 17: MEDIAN SIMILARITY TO VC STARTUPS, BY SECTOR AND YEAR



I. Private equity market efficiency and the choice to go public

In this appendix I endogenize the startups' choice to go public using a simple model. My objective is to show how the "efficiency explanation" (that the secular shift from IPO to acquisitions is due to an increase in the efficiency of private equity markets) can be reconciled with aggregate trends in IPOs.

Consider the problem of startup i, who needs to decide whether to go public or sell to an incumbent $j \neq i$, and knows it can obtain the following valuation by going public:

$$V\left(\pi_i^t; r\right) \tag{117}$$

where π_i^t is the stream of future expected profits and $V(\cdot; r)$ is the discounted present value function with discount rate r. Public firm j is willing to pay, for startup i, the difference between Π_{ij}^t — the aggregate stream of profits that the i-j combined entity can make — and π_j^t — the expected stream of profits it would make if it didn't buy i — discounted at a subjective rate r_j:

$$V\left(\Pi_{ij}^t - \pi_j^t; r_j\right) \tag{118}$$

Company i chooses to go public if the price it can get on the stock market is larger than the highest valuation it can get in the private market:

$$V\left(\pi_i^t; r\right) \;>\; \max_j V\left(\Pi_{ij}^t - \pi_j^t; r_j\right) \tag{119}$$

Let ι be the company j with the largest valuation for i — that is: $\iota = \arg\max_j V\left(\Pi_{ij}^t - \pi_j^t; r_j\right)$. Then i will go public if:

$$V\left(\Pi_{i\iota}^t - \pi_i^t - \pi_\iota^t; r_j\right) \;<\; V\left(\pi_i^t; r\right) - V\left(\pi_i^t; r_\iota\right) \tag{120}$$

here I have used the fact that, for a fixed discount rate, the DPV function is additive. Also, because V is monotonic in r, we can write (under the assumption of the present value being positive) the right hand side as an increasing function f of the differences in discount rates $r_\iota - r$. Inverting function f we have:

$$f^{-1}\left(V\left(\Pi_{i\iota}^t - \pi_\iota^t - \pi_i^t; r_j\right); \pi_i^t\right) \;<\; r_\iota - r \tag{121}$$

Clearly, ι's discount rate for the transaction (r_ι) is crucial. I assume that it takes the following expression:

$$r_\iota \;=\; r + \tau_\iota + g_\iota\left(\underset{-}{\pi_{it}}, \underset{+}{V\left(\pi_i^t; r\right)}\right) \tag{122}$$

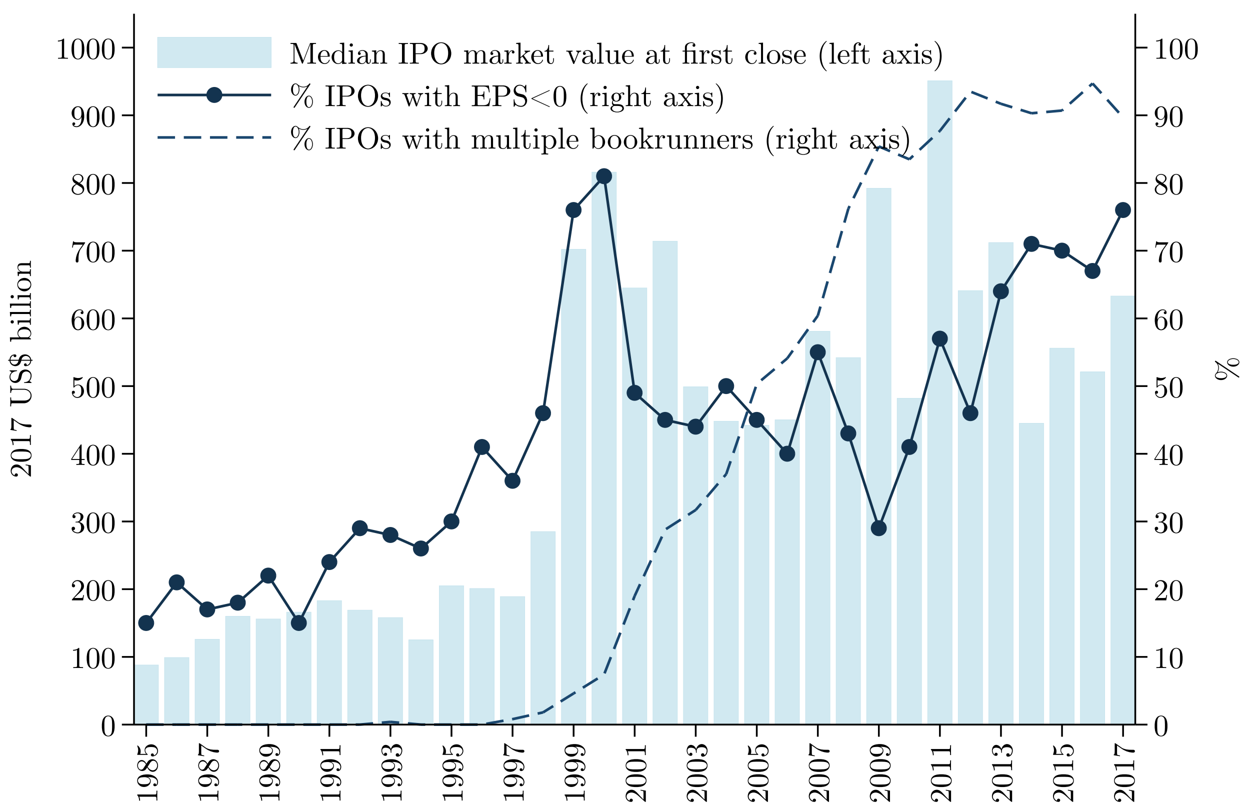
where τ is an idiosyncratic friction (we can think of it as the cost of learning about the acquisition target or the cost of performing a due diligence) and the function $g(\cdot)$ is assumed to be positive, decreasing in current profits π_{it} and increasing in the market value of firm i. The function $g(\cdot)$ captures the fact that large, unprofitable startups are likely to attract a smaller pool of buyers: this is because they make for riskier acquisition targets and because they require larger liquidity reserves to be acquired. We can finally re-write the optimality condition as follows:

$$f^{-1}\left(V\left(\Pi_{i\iota}-\pi_{\iota}-\pi_{i};r_{j}\right);\pi_{i}\right)-\tau_{\iota}\ <\ g\left(\underset{-}{\pi_{it}},\underset{+}{V\left(\pi_{i}^{t};r\right)}\right) \tag{123}$$

Suppose now that private equity markets become more efficient—that is, keeping everything else unchanged, the distribution of τ shifts to the left. Then, conditional on i getting acquired, g will increase in expectation. This implies that the firms that choose to IPO will be, in expectation, larger, less profitable and riskier. At the extreme, all but the largest and riskiest startups get acquired.

Figure 18 shows some aggregate IPO statistics, which I obtain from Jay Ritter's database, from which three trends clearly emerge: 1) the size of the median IPO (adjusted for inflation) has nearly quadrupled since the early 1990s; 2) at the beginning of the 1990s most companies going public (80%) had positive earnings per share (EPS), while today nearly 80% of the companies going public are making losses; 3) up to the mid 90s all IPOs had one have multiple book runners while today over 90% do. Because one of the main reasons why underwriters choose to syndicate an IPO is to share the risk, the increase in this latter figure is consistent with the hypothesis that more recent IPOs have been larger and riskier.

Figure 18: IPO Statistics (1985-2017; source: Jay Ritter)



J. Nonconstant returns to scale and fixed costs

TABLE 6: VARIABLE DEFINITIONS AND MAPPING TO DATA - ALTERNATIVE MODEL

The following table presents descriptions of the variables used in the estimation of the alternative model with nonconstant returns to scale and overhead presented in section 7.

Panel A: Observed variables

Notation	Observed Variable	Measurement
$\bar{\mathbf{C}}\mathbf{q}$	*Total Variable Costs*	Costs of Goods Sold (CoGS)
		source: CRSP+Compustat
$\mathbf{Pq}$	*Revenues*	max (Operating Revenues, CoGS)
		source: CRSP+Compustat
$\mathbf{f}$	*Fixed Costs (Overhead)*	Selling, General and Administrative Costs (SG&A)
		source: CRSP+Compustat
Π	*Aggregate profits*	Total sum of operating profits
		source: CRSP+Compustat
$\mathbf{A}'\mathbf{A}$	*Product similarity* for *Compustat* firms	Text similarity in 10-K product description
		source: Hoberg and Phillips (2016)
$\mathbf{A}'\mathbf{a}_0$	*Product similarity for VenturExpert firms*	Text similarity in product description
		source: Hoberg, Phillips, and Prabhala (2014)

Panel B: Model-inferred variables

Notation	Derived Variable	Computation/Identification
π	*Profits*	$= \mathbf{Pq} - \mathbf{Cq}$
$\mathbf{q}$	*Output*	$= \sqrt{\frac{1}{1+\beta}(\pi + \mathbf{f})}$
$\mathbf{Ab} - \mathbf{c}$	*Marginal Surplus at* $\mathbf{q} = \mathbf{0}$	$= [(1+2\beta)\mathbf{I} + \mathbf{A}'\mathbf{A}]\mathbf{q}$

Figure 19: Aggregate profits and consumer surplus in the alternative model with nonconstant returns to scale

The following figure replicates the calculations of Figure 5 for the alternative model with fixed costs and nonconstant returns to scale described in Section. 7

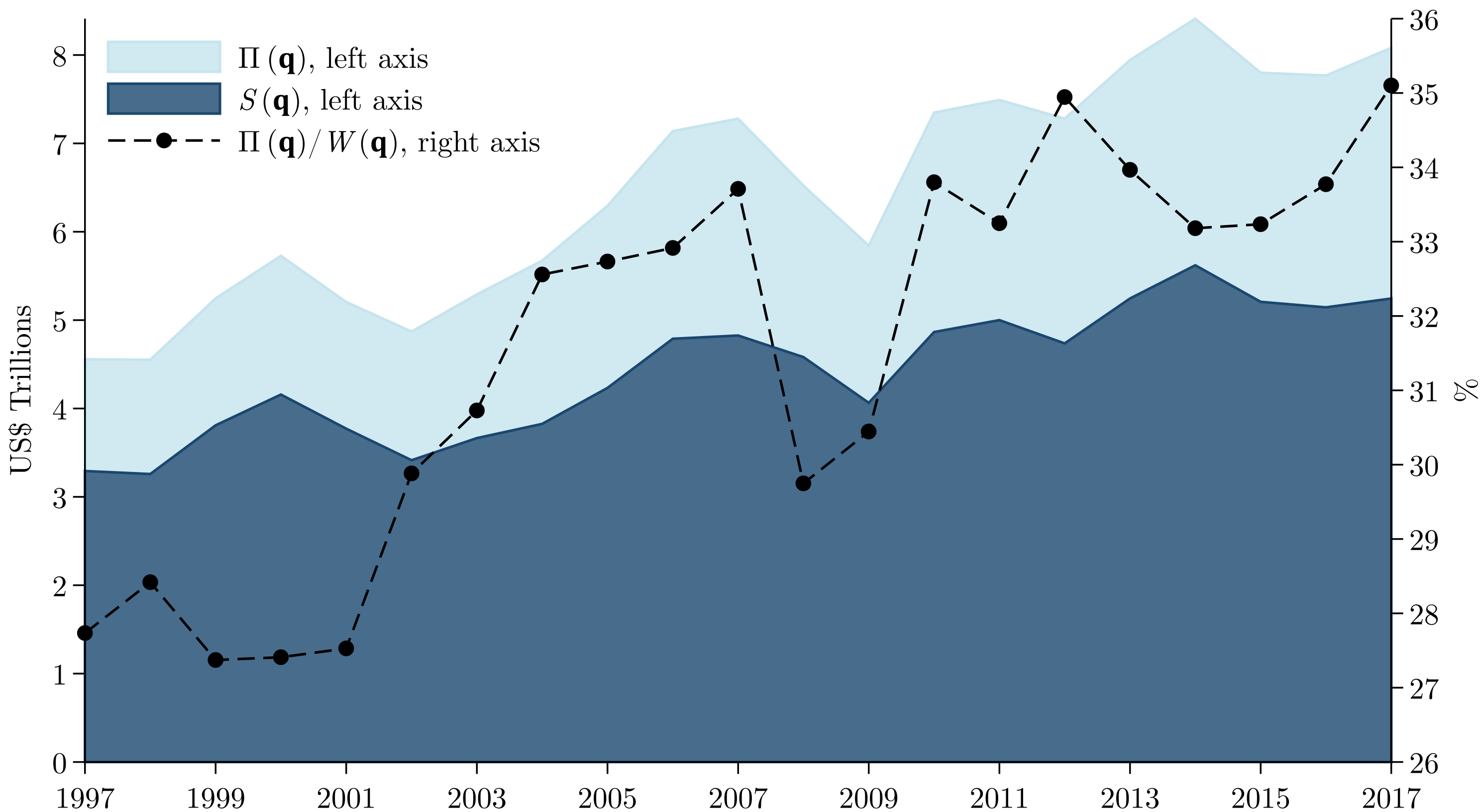
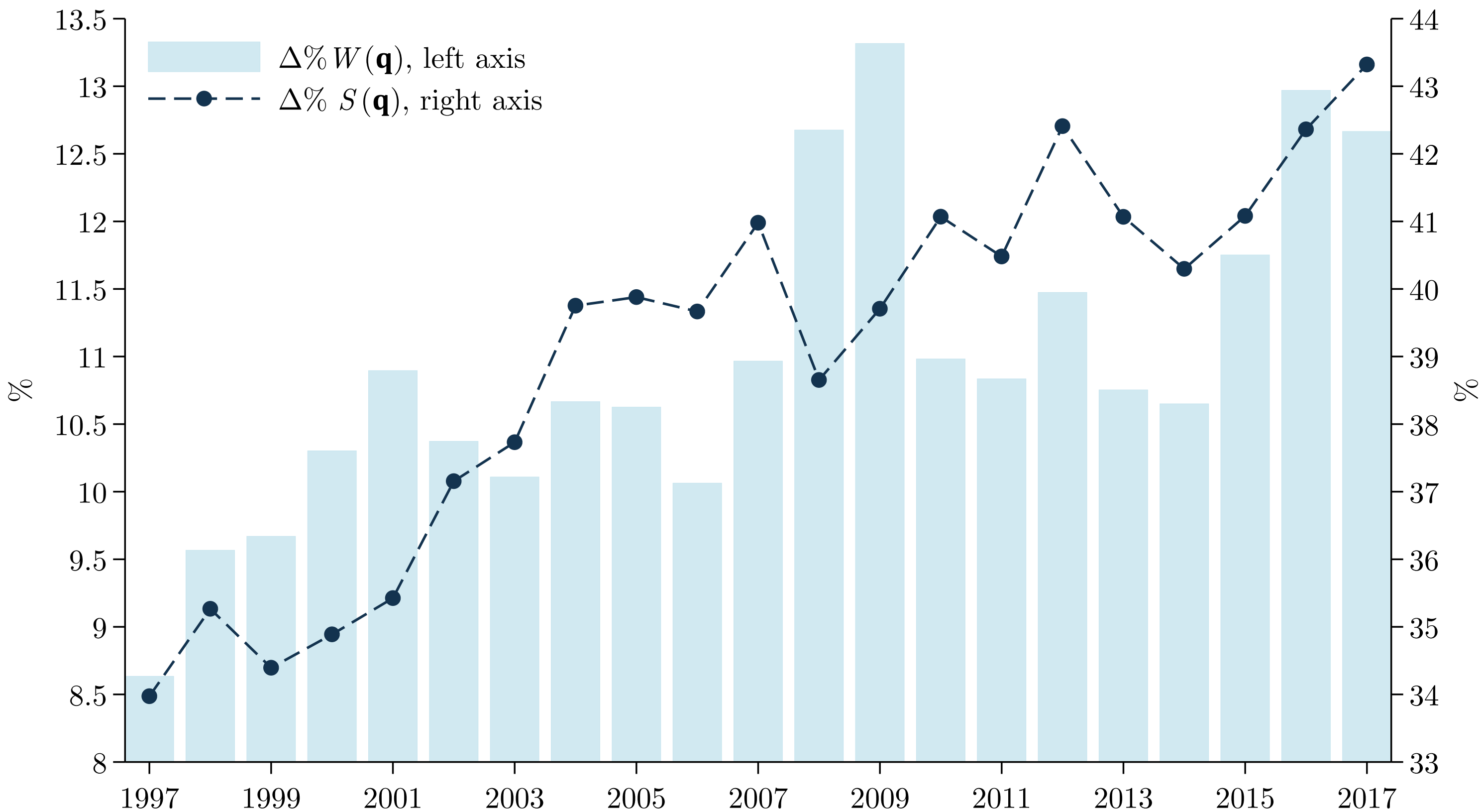


Figure 20: Welfare costs of oligopoly in the alternative model with nonconstant returns to scale

The following figure replicates the calculations of Figure 6 for the alternative model with fixed costs and nonconstant returns to scale described in Section. 7



K. Proxying for Private and Foreign firms

In this Appendix, I address one of the main shortcomings of my data set– sample selection. Compustat does not cover private and foreign firms: this could lead to incorrect inferences about welfare trends if the preponderance of these unobserved firms with respect to public U.S. firms varies substantially over time.

While firm-level data for these companies (particularly, similarity scores) is unavailable, I show here how one can use aggregate data to derive some error bounds for the observed welfare measures. Let us start from the Nash-Cournot best response function, which allows us to identify the vector $(\mathbf{A}'\mathbf{b} - \mathbf{c})$:

$$\mathbf{a}_i'\mathbf{b} - c_i \;=\; q_i + \sum_{j \in \mathcal{I}} \mathbf{a}_i'\mathbf{a}_j q_j \tag{124}$$

As a matter of fact, we can only observe the Compustat subset $\mathcal{C}$ of the firm population $\mathcal{I}$. Let:

$$\zeta_i = \frac{\sum_{j \in \mathcal{I}} \mathbf{a}_i'\mathbf{a}_j q_j}{\sum_{j \in \mathcal{C}} \mathbf{a}_i'\mathbf{a}_j q_j} \tag{125}$$

We can then rewrite:

$$\mathbf{a}_i'\mathbf{b} - c_i \;=\; q_i + \zeta_i \sum_{j \in \mathcal{C}} \mathbf{a}_i'\mathbf{a}_j q_j \tag{126}$$

ζ_i can be thought of as a measure of the bias caused by the omission of private and foreign firms. If $\zeta_i > 1$, omitting private and foreign firms (which equates to assuming $\zeta_i = 0$) leads to downward biased estimate of $(\mathbf{a}_i'\mathbf{b} - c_i)$: the intuition is that we underestimate how much competition Compustat firms face. As a consequence, we end up finding that observed profits are a larger share of surplus than they actually are. One way to correct my estimates for the unobserved private and foreign firms is to try and use aggregate data to bound ζ. I argue that the following expression provides a heuristic upward bound to ζ:

$$\zeta^{\max} \;=\; \frac{\text{GDP} + \text{Foreign Value Added}}{\text{COMPUSTAT Value Added}} \;>1 \tag{127}$$

To compute this value I obtain the value added by foreign firms (at the broad sector level) from the OECD Trade in Value Added (TiVA) data set . This upward bound $\zeta^{\max}$ implicitly assumes that Compustat firms are in some sense representative of private and foreign firms, and is therefore unrealistically conservative. A reasonable value ζ should realistically fall somewhere between 1 and $\zeta^{\max}$. How close ζ is to either of these two values depends on the degree of product similarity between the firms which we observe (Compustat) and those which we do not observe (private and foreign). In the limit case where these firms do not interact strategically with the Compustat sample, no correction is needed.

Nonetheless, I next present empirical estimates from these "conservatively-corrected" estimates, applying the upper bound $\zeta^{\max}$ to all firms in order compute my welfare measures. Figures 21 and 22 present alternative welfare calculations in which these weights are used to proxy for unobserved private and foreign firms. The levels of both the profit share of surplus as well as the deadweight loss from oligopoly are (as expected) lower than the corresponding baseline figures (21% and 2.6%, respectively). However, they continue to show a pronounced upward trend. This suggests that, while the baseline model is likely to overstate somewhat the overall extent of market power, the finding that market power has increased substantially over the past two decades appears to be robust to the inclusion of private and foreign firms.

Figure 21: Aggregate profits and consumer surplus – proxying for private and foreign firms

The following figure replicates the calculations of Figure 5 for the alternative model in which population weights are used to proxy for unobserved (private and foreign) firms.7

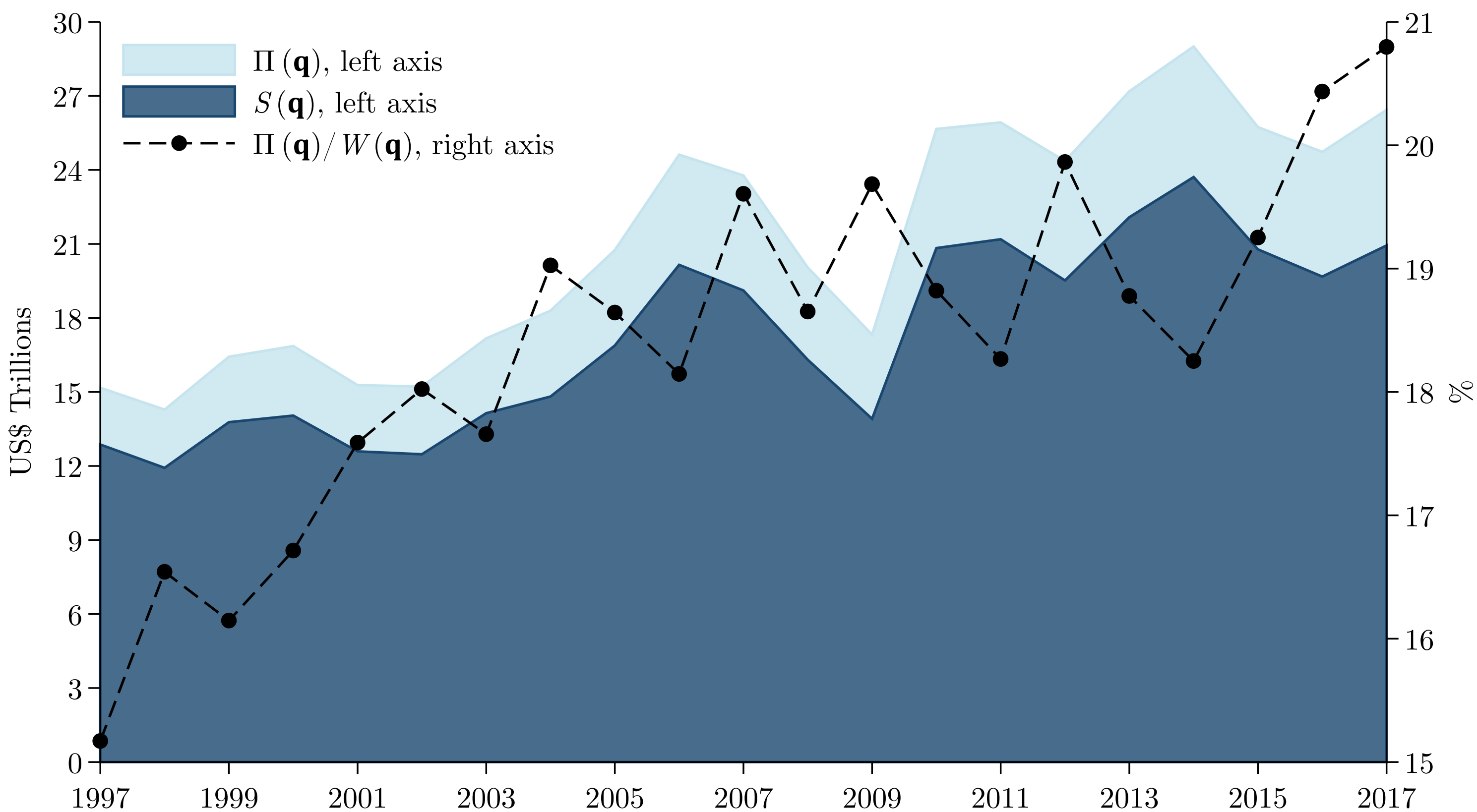
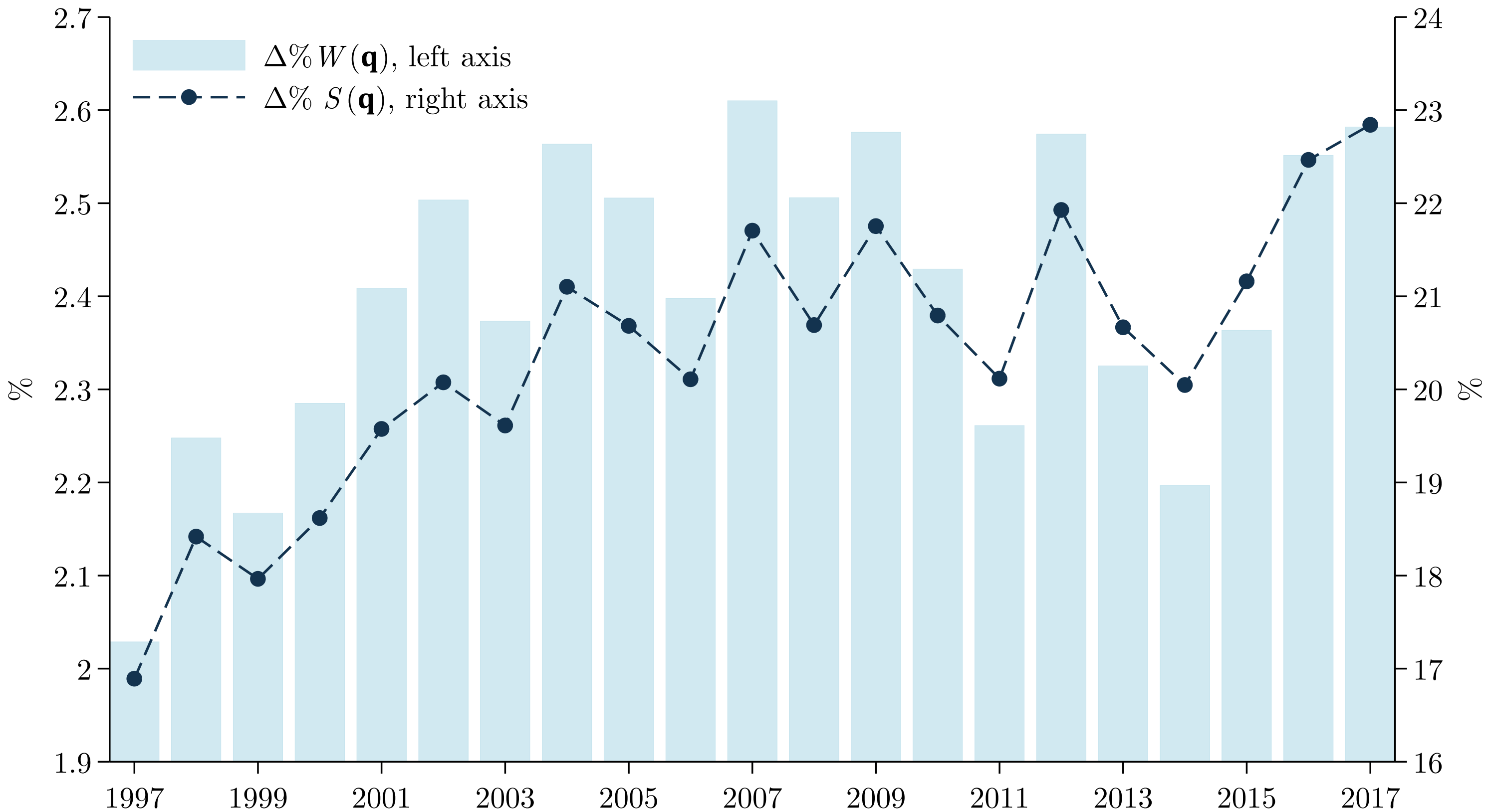


Figure 22: Welfare costs of oligopoly in the alternative model – proxying for private and foreign firms

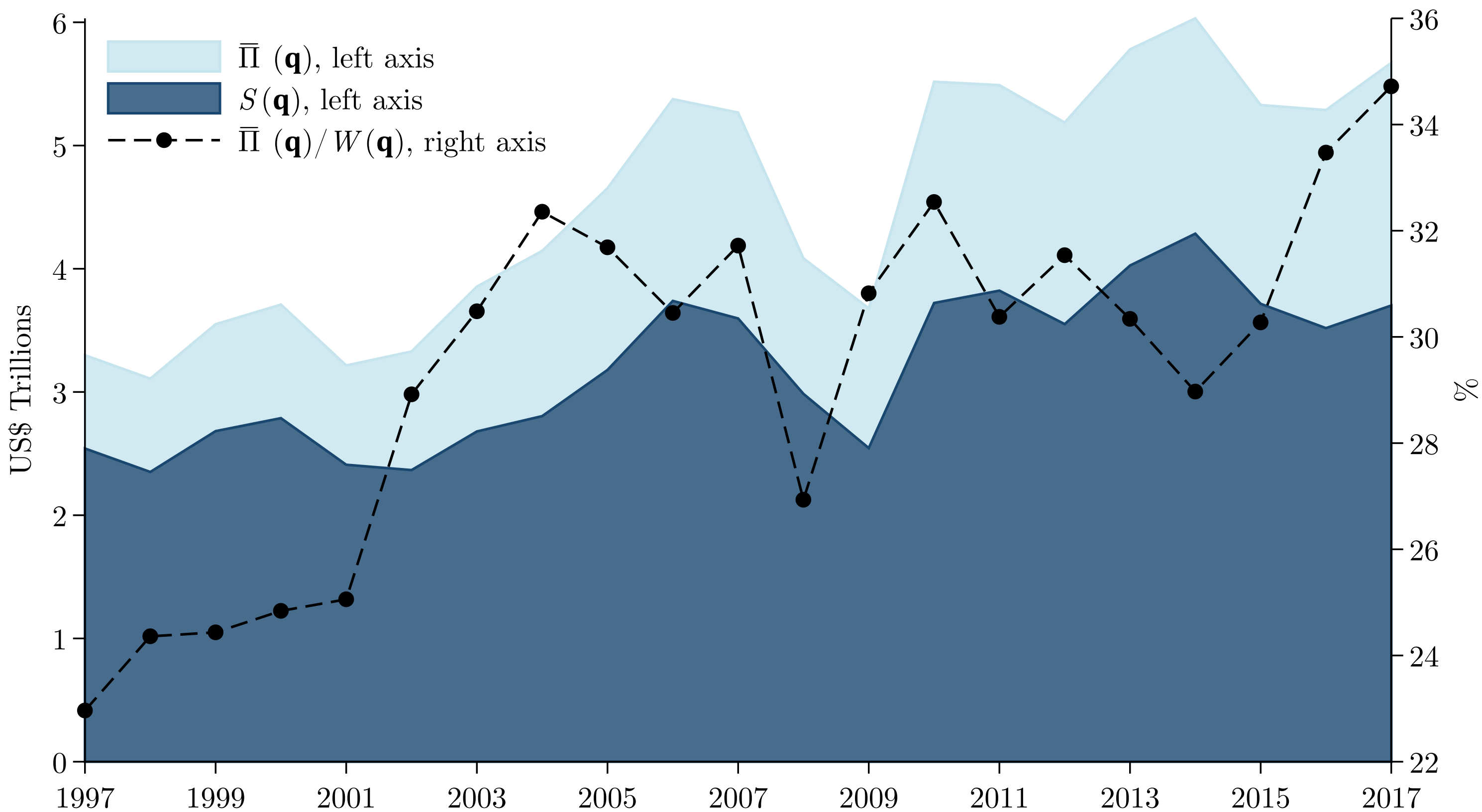
The following figure replicates the calculations of Figure 6 for the alternative model in which population weights are used to proxy for unobserved (private and foreign) firms.7



L. Factoring in capital costs

The diagram below replicates the analysis of Figure 5, by using an alternative measure of aggregate profits, $\overline{\Pi}$, that accounts for capital consumption costs and the expected growth rate by deducting capex (see subsection 7.6). The figure shows that, after making this adjustment, the rise in the profit share of surplus is even more dramatic. This measure of profit accounts for all costs except the direct cash dividends to providers of capital. This suggests that—unless the opportunity cost of capital has increased sensibly—fixed costs do not seem to be able to account for the rising profits.

Figure 23: Aggregate profits (alternative measure) and consumer surplus (1997-2017)



M. Consolidation unlikely to be a result of increasing competition or changing returns of scale

In this Appendix I address the question of whether industry consolidation can be explained by an increase in competition or changing economies of scale. I do so by first asking whether the average product similarity has increased over the period 1996-2015. I perform a Pseudo-Poisson Maximum Likelihood regression (PPML - Silva and Tenreyro, 2010; Correia et al., 2019) of bilateral similarity scores on year fixed effects.[22] The beta coefficients on the year fixed effects can be read as the cumulative percentage increase in similarity with respect to 1996.

However, endogenous entry and exit add a level of complication to the analysis, because entry and exit themselves have a mechanical effect on the average level of similarity across firms. Suppose, for example, that there are n firms and that none of them changes their product offering between times t_1 and t_2 (hence there is no change in bilateral product similarity), yet n_0 of these firms exit over the same period—then the average similarity between the remaining firms might still change simply as a result of survivor bias. To account for this survivor bias, I also perform the same regression with the addition of firm pair fixed effects: this way I can isolate the change in product similarity *within* firm pair.

Figure plots the cumulative percentage change in average product similarity among Compustat firms over the period 1996-2015 using these different measurements.[23] The the darker line with circular markers plots the year coefficients from a regression without firm pair fixed effects: it shows that average product similarity has increased steadily by a cumulative 14% between 1996 and 2015. The lighter line with square markers plots the year coefficients effects from a regression that includes firm pair fixed effects: after accounting for entry and exit, we can see that average similarity *within* firm pairs has actually decreased by a cumulative 2.4%; the vertical grey bars represent the difference between the two lines, and can be interpreted as the survivor bias.

To investigate whether upward movement in returns to scale might be responsible for the measured trends in consumer and producer welfare, I consider the alternative model with fixed costs and nonconstant returns scale presented in Section 7. For this model, I calibrate the scale elasticity parameter (β) year-by-year by picking the level that allows me to match exactly the revenue-weighted markup of De Loecker et al. (2018). The time series for the calibrated β are shown in Figure 25. The value of β oscillates, over the last twenty years, between .16 and .25 but with no clear trend.[24]

This data analysis suggests that concentration is unlikely to be driven by intensifying product market competition or changing returns to scale. Once entry and exit are controlled for, it appears there is no evidence of a measurable increase in substitutability among the firms' outputs, and the survivor bias runs opposite to what would be implied by the "increasing competition" hypothesis. Moreover, there appears to be no measurable trend, over this period, in the degree of returns to scale, in line with existing macro evidence (Ho and Ruzic, 2018).

[22] Using a panel regression and dividing the year fixed effects by the baseline value (1996) yields nearly identical results.

[23] For this computation, the "truncated" version of HP's data set was used, since the complete data set is too large for regression analysis in STATA. The truncated data set was available up until late 2016 on Hoberg and Phillips' data repository.

[24] A linear regression on a time trend reveals no measurable trend in the series.

Figure 24: Cumulative % change in average product similarity

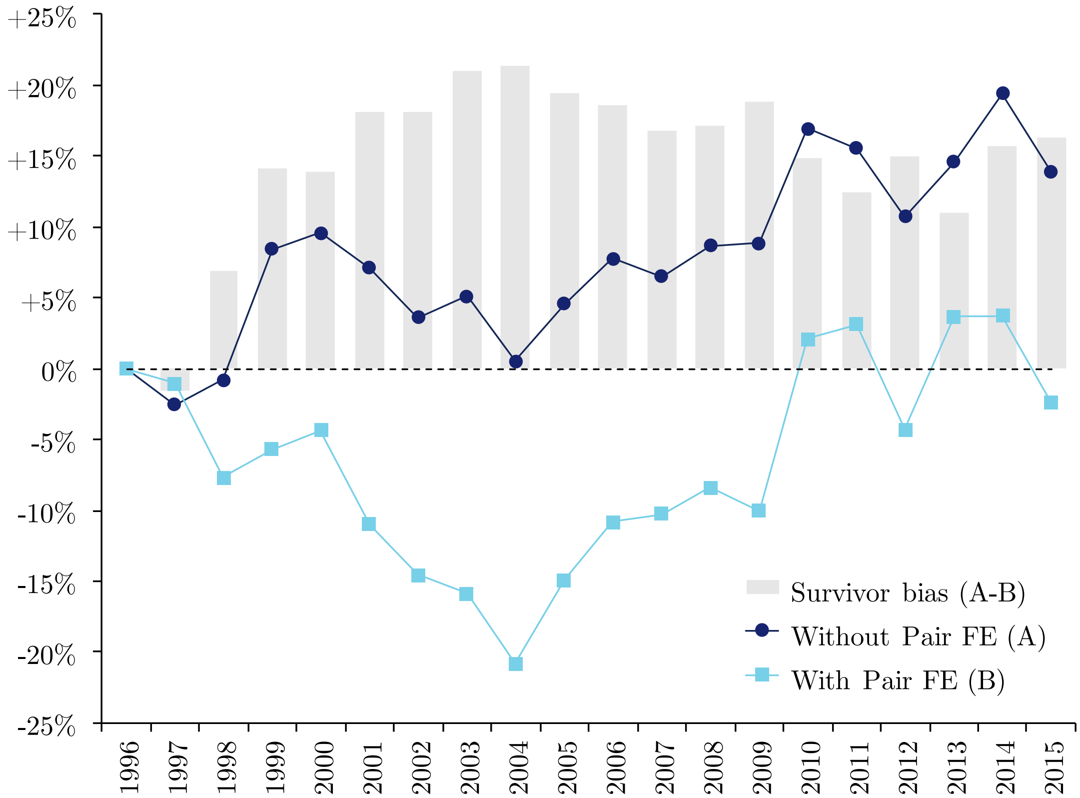
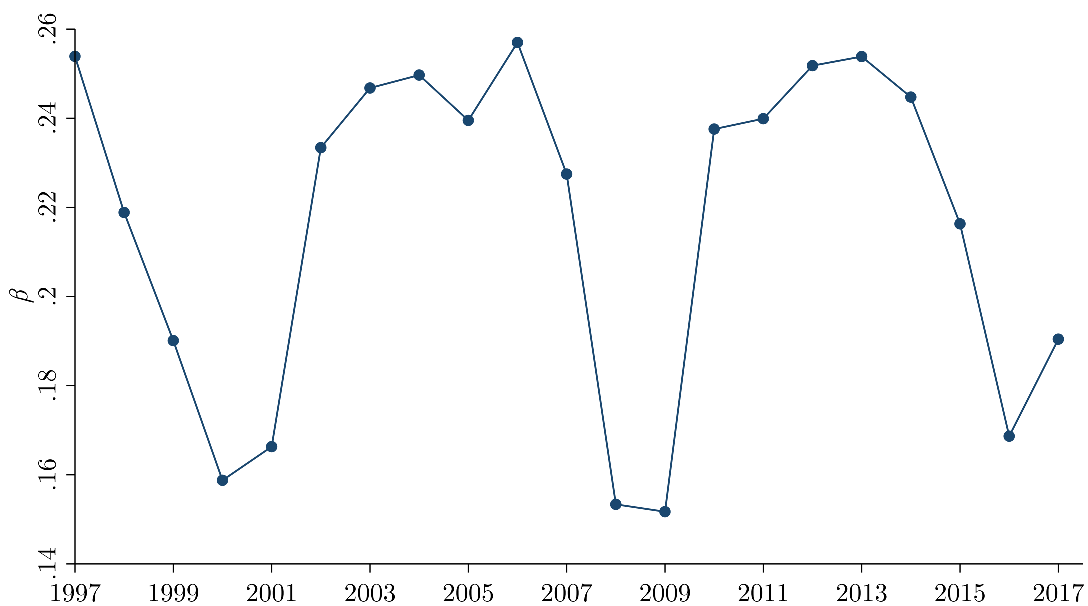


Figure 25: Calibrated scale elasticity parameter (β) over time - Alternative Model



N. Incorporating market size effects

The basic model presented in Section 2 abstracts away from market size effects. In other words, it is assumed that all firms cater to the same consumer population. In this appendix, I study an extension of the model where firms have different market sizes, in the sense that there is a force in the model that allows the firms' real output to grow independently of the firms' markups. I start from the demand function. Let $\mathbf{m}$ be a (strictly positive) n-dimensional strictly vector that contains, in every i^{th} component, the "market size" for company i. Then the demand function in a model with market size would be:

$$\mathbf{q} \;=\; \mathbf{M}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{p}\right) \tag{128}$$

This demand system yields the following equilibrium unit margin and quantity:

$$\mathbf{p}-\mathbf{c} \;=\; \left(\mathbf{I}+\mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right) \tag{129}$$

$$\mathbf{q} \;=\; \mathbf{M}\left(\mathbf{I}+\mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right) \tag{130}$$

Hence a shift in market sizes affects quantity vector $\mathbf{q}$ but not the price vector $\mathbf{p}$. The representative agent utility that would yield such demand system is:

$$U\left(\mathbf{x},H\right) \;=\; \mathbf{x}'\mathbf{b}-\frac{1}{2}\mathbf{x}'\hat{\mathbf{M}}^{-1}\mathbf{x}-H \tag{131}$$

where $\hat{\mathbf{M}}$ is a square $k\times k$ matrix that satisfies:

$$\mathbf{A}'\hat{\mathbf{M}}^{-1}\mathbf{A} \;=\; \mathbf{A}'\mathbf{A}\mathbf{M}^{-1} \tag{132}$$

If $\mathbf{AA}'$ is invertible (this condition will be satisfied if there are as many firms as characteristics and no two firms are the same) it is possible to find some $\left(\mathbf{M},\hat{\mathbf{M}}\right)$ such that this relationship holds.

Next, I study what are the implications, for unobserved welfare measures such as consumer surplus, of assuming no market size effects (as in the baseline model) when the market size model is the correct one—that is, when we erroneously assume $\mathbf{M}=\mathbf{I}$. In this case, we erroneously map:

$$\hat{\mathbf{q}} \;=\; \hat{\mathbf{p}}-\hat{\mathbf{c}} \;=\; \mathbf{M}^{1/2}\left(\mathbf{I}+\mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)$$

where the hat symbol indicates that these are the measured quantity, price and cost statistics as opposed to the true ones. that is, the market size is equally loaded on the quantity vector and on the unit margin vector when it should instead be loaded entirely on the quantity vector. The correct surplus in the "market size" model is:

$$\begin{aligned} S \;&=\; \frac{1}{2}\mathbf{q}\mathbf{A}'\mathbf{A}\mathbf{M}^{-1}\mathbf{q} \\ &=\; \frac{1}{2}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)'\left(\mathbf{I}+\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{M}\mathbf{A}'\mathbf{A}(\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right) \end{aligned}$$

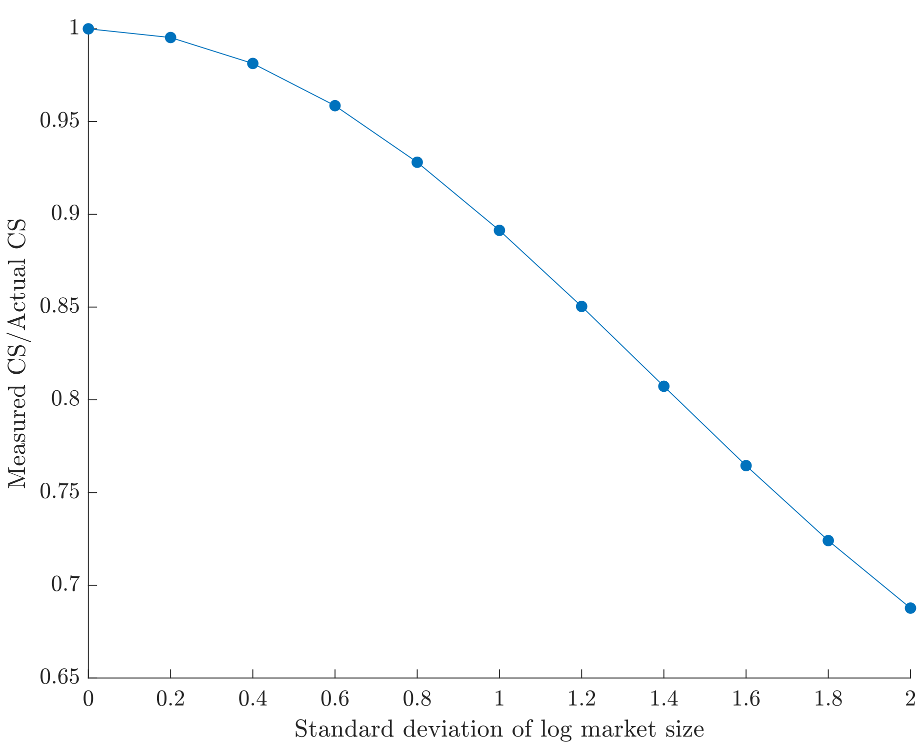
The incorrectly measured surplus is, instead:

$$\begin{aligned}\hat{S} &= \frac{1}{2}\hat{\mathbf{q}}'\mathbf{A}'\mathbf{A}\hat{\mathbf{q}} \\ &= \frac{1}{2}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)'\left(\mathbf{I}+\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{M}^{1/2}\mathbf{A}'\mathbf{A}\mathbf{M}^{1/2}(\mathbf{I}+\mathbf{A}'\mathbf{A})^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)\end{aligned}$$

Hence, a difference would arise between actual and measured consumer surplus. This difference would depends crucially on the dispersion of the vector $\mathbf{m}$ (if all market sizes are the same, then the benchmark and the market size model are equivalent). While this market size is unobserved, we can use the other (observed) data to perform a Monte Carlo sensitivity analysis. The idea is to draw random $\mathbf{m}$ vectors with increasing degrees of dispersion and study how the measurement error of S changes. To perform this exercise, I assume that m is lognormally-distributed.

Figure 26 shows the result of this analysis. The horizontal axis represents the standard deviation of the log of m_i, while the vertical axis represents the ratio of the measured to actual consumer surplus. As can be seen from the graph, unless market size is irrelevant, the measured consumer surplus tends to undershoot the actual consumer surplus. However, the error is not large for reasonable values of the standard deviation of log market size. For the measured surplus to undershoot the actual surplus by significantly more than 10%, the standard deviation of the log market size should exceed one (meaning that at the 75^{th} percentile has a market size nearly 4 times that of a firm at the 25^{th} percentile). In other words, for the error to be large the market size would have to be very dispersed.

Figure 26: Sensitivity of consumer surplus to market size effects



O. Incorporating input-output linkages

In this appendix, I explore an extension of the model with input-output linkages, and show under what assumption we can still write the welfare metrics in the model using the matrix of product similarities $\mathbf{A}'\mathbf{A}$.

Firms use both labor as well as an intermediate input to produce. Firms $i = \{1, 2, ..., n\}$ sell their output to both the representative agent as well as to the intermediate goods firm, and they demand an amount y_i of the input. We can normalize the price of this input to be 1 (same as labor). We assume that the total demand for the goods $i = \{1, 2, ..., n\}$, quantified by the vector $\mathbf{q}$, is the sum of the demand from the household $\bar{\mathbf{q}}$ and the demand from the intermediate good firm $\tilde{\mathbf{q}}$.

Figure 27 shows the input-output structure of the economy after introducing the intermediate good firm (firm 0): the economy has a "roundabout" structure Baqaee and Farhi (2018). L represents the labor supply, and HH represent the household sector.

In order for us to be able to write the welfare functions in terms of the similarity matrix $\mathbf{A}'\mathbf{A}$, we need to specify a technology for firm 0 that results in a linear demand function $\tilde{\mathbf{q}}$ that can be aggregated with that of the household $\bar{\mathbf{q}}$. Thus, I assume that the representative household has the following preferences

$$U(\mathbf{x}, H) \; = \; \mathbf{b}'\mathbf{x} - \frac{1}{2\alpha}\mathbf{x}'\mathbf{x} - H \tag{133}$$

implying the following consumer surplus function:

$$S(\bar{\mathbf{q}}) \; = \; \bar{\mathbf{q}}'\mathbf{A}'\mathbf{b} - \frac{1}{2\alpha}\bar{\mathbf{q}}'\mathbf{A}'\mathbf{A}\bar{\mathbf{q}} \tag{134}$$

I also assume that the intermediate good firm (firm 0) has the following production function:

$$Y(\tilde{\mathbf{q}}) \; = \; \tilde{\mathbf{q}}'\mathbf{A}'\mathbf{b} - \frac{1}{2(1-\alpha)}\tilde{\mathbf{q}}'\mathbf{A}'\mathbf{A}\tilde{\mathbf{q}} \tag{135}$$

The parameter α determines the relative size of the business-to-business market and that of the retail market. This yields the same inverse aggregate demand function as in the baseline model:

$$\mathbf{p} \; = \; \mathbf{A}'\mathbf{b} - \mathbf{A}'\mathbf{A}\mathbf{q} \tag{136}$$

We can verify that the balance of payments in this economy is zero. The consumer expenditure $\mathbf{p}'\bar{\mathbf{q}}$ is equal to the budget, which is itself the sum of the profits of firms $i > 0$,

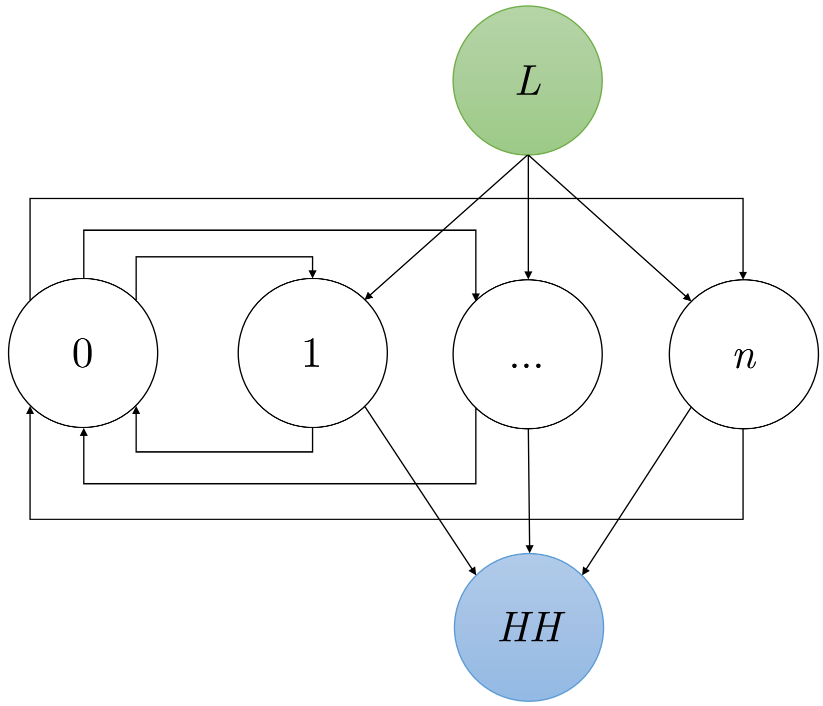
$$\bar{\Pi} \; = \; \mathbf{p}'\mathbf{q} - Y - H \tag{137}$$

the profits earned by firm 0

$$\tilde{\Pi} \; = \; Y - \mathbf{p}'\tilde{\mathbf{q}} \tag{138}$$

and the labor share H. What changes in the input-output model is the computation of consumer surplus. H is no longer equal to total cost: we need to subtract the intermediate inputs Y, and we need to compute the consumer surplus function using only the final good demand from the household $\bar{\mathbf{q}}$.

FIGURE 27: INPUT-OUTPUT STRUCTURE OF THE ECONOMY



P. Startup Acquisitions and (the lack of) Product Innovation

One potential objection to the results of Section 6 is that one reason why firms, particularly in Tech, acquire startups is to gain ownership of valuable technology and human capital. By preventing startups from being acquired, we would be limiting the ability of larger firms to innovate.

While I cannot completely exclude the innovative motive, in this Appendix I provide evidence that the increased rate of startups acquisitions was not associated with more product innovation, but less. In particular, I show that innovation (measured using 10-K text data), decreased particularly sharply for the Big 5 tech companies (Alphabet, Amazon, Facebook, Google, Microsoft). I use again the Angular Velocity score

$$\text{AngularVelocity}_{it} \; = \; \text{acos}\left(\mathbf{a}_{it}'\mathbf{a}_{it-1}\right) \tag{139}$$

which measures year-to-year variation in 10-K product description, as an objective gauge of product market dynamism at the firm-level. I measure cumulative percentage change in this variable using a PPML regression, which allows to control for survival bias (as I did for similarity scores in Appendix M).

In Figure 28, I show that, far from increasing, average velocity for Compustat firms decreased by nearly 20% since 1997. This change was *not* driven by survival bias, since a similar trend is obtained by controlling for firm fixed effects. In Figure 29, I zoom in on the post Tech "bubble" period, and in particular on the Big 5 tech. For these firms, an even more dramatic decrease in velocity (nearly 45%) occurred. These trends cast doubt on the hypothesis that the intense startup acquisition activity of large tech companies was driven by a product innovation motive.

Figure 28: Cumulative % change in angular velocity

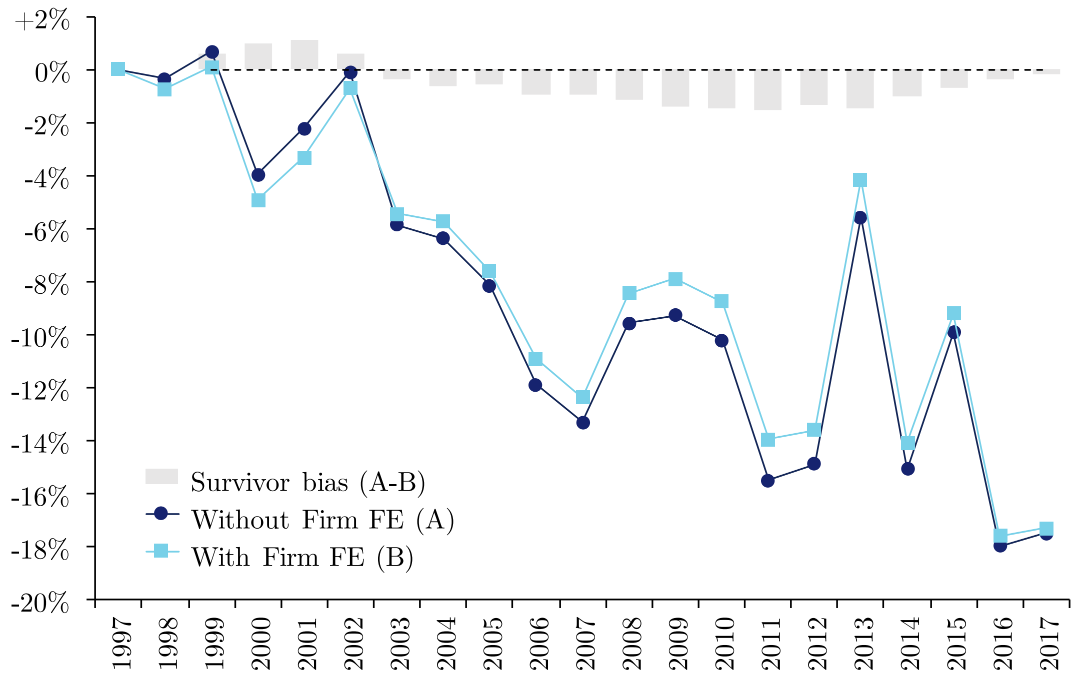
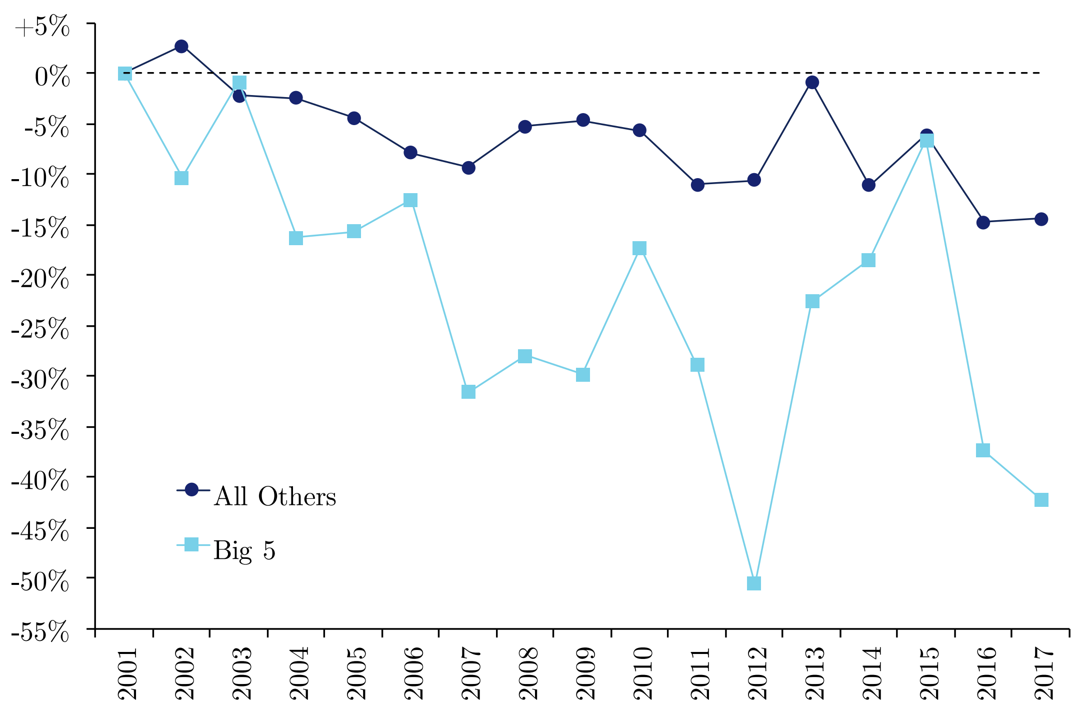


Figure 29: Cumulative % change in angular velocity (with firm FE)



Q. Bertrand Competition

In this Appendix I derive the equilibrium of the Bertrand game corresponding to the Cournot game played by the firms in the model presented in Section 2. I start from writing the profit function in terms of the price vector $\mathbf{p}$:

$$\pi = (\mathbf{P}-\mathbf{C})\left(\mathbf{A}'\mathbf{A}\right)^{-\mathbf{1}}(\mathbf{A}'\mathbf{b}-\mathbf{p}) \tag{140}$$

Now let $\mathbb{D}$ and $\mathbb{O}$ be, respectively, the matrices containing the diagonal and off-diagonal elements of $\left(\mathbf{A}'\mathbf{A}\right)^{-1}$ so that:

$$\left(\mathbf{A}'\mathbf{A}\right)^{-1} = \mathbb{D}+\mathbb{O} \tag{141}$$

Then we can write:

$$\pi = (\mathbf{P}-\mathbf{C})\left[\mathbb{D}(\mathbf{A}'\mathbf{b}-\mathbf{p})+\mathbb{O}(\mathbf{A}'\mathbf{b}-\bar{\mathbf{p}})\right] \tag{142}$$

and take the first order condition firm-by-firm by taking the price vector of other firms (denoted by the upper bar as in Appendix A) as given:

$$0 = \mathbb{D}(\mathbf{A}'\mathbf{b}-2\mathbf{p})+\mathbb{O}(\mathbf{A}'\mathbf{b}-\mathbf{p})-\mathbb{D}\mathbf{c} \tag{143}$$

which we can re-write in terms of $\mathbf{q}$ as:

$$0 = \mathbf{A}'\mathbf{b}-\mathbf{c}-\left(\mathbb{D}^{-\mathbf{1}}+\mathbf{A}'\mathbf{A}\right)\mathbf{q} \tag{144}$$

the corresponding Bertrand potential is

$$\Phi^B = \mathbf{q}'(\mathbf{A}'\mathbf{b}-\mathbf{c})-\left[\frac{1}{2}\mathbf{q}'\left(\mathbf{I}+\mathbb{D}^{-1}\right)\mathbf{q}+\frac{1}{2}\mathbf{q}'(\mathbf{A}'\mathbf{A}-\mathbf{I})\mathbf{q}\right] \tag{145}$$

and the Bertrand equibrium is:

$$\mathbf{q}^B = \left(\mathbb{D}^{-1}+\mathbf{A}'\mathbf{A}\right)^{-1}(\mathbf{A}'\mathbf{b}-\mathbf{c}) \tag{146}$$

because $\mathbb{D}^{-1}$ is a diagonal matrix whose diagonal components are (heuristically) between zero and one, from the above equations we can be seen how Bertrand is a more “intense” form of competition than Cournot.

R. Proofs and Derivations

Proof to Lemma 2. Rearranging equation (20) we have

$$\mathbf{Q}^{-1}\mathbf{A}'\mathbf{A}\mathbf{q} \;=\; \mathbf{Q}^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)-1 \tag{147}$$

which can be re-written as

$$\frac{1}{\sigma_i} \;=\; \frac{\mathbf{a}_i'\mathbf{b}-c_i}{q_i}-1 \tag{148}$$

which, rearranged, leads to equation 35. □

Proof to Proposition 3. We write the Lagrangian of this problem, conveniently picking $(1-\mu)$ as the Lagrangian multiplier:

$$\mathscr{L}\left(\mathbf{q};\hat{\mathbf{q}}\right) \;=\; \mathbf{q}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right)-\frac{1}{2}\mathbf{q}\mathbf{A}'\mathbf{A}\mathbf{q}+\left(1-\mu\right)\left(\mathbf{c}'\mathbf{q}-\mathbf{c}'\hat{\mathbf{q}}\right) \tag{149}$$

Then the first order condition yields the following form for the solution, as in equation 56

$$\mathbf{q}=\left(\mathbf{A}'\mathbf{A}\right)^{-1}\left[\mathbf{A}'\mathbf{b}-\mu\left(\hat{\mathbf{q}}\right)\mathbf{c}\right] \tag{150}$$

to find the markup μ we use the budget constraint:

$$\mathbf{c}'\mathbf{q} \;=\; \mathbf{c}'\left(\mathbf{A}'\mathbf{A}\right)^{-1}\left[\mathbf{A}'\mathbf{b}-\mu\left(\hat{\mathbf{q}}\right)\mathbf{c}\right] \;=\; \mathbf{c}'\hat{\mathbf{q}} \tag{151}$$

which leads directly to equation (57) for μ. □

Proof to Proposition 4. To simulate a merger, we sum the first rows of the profit function that correspond to the merging firms:

$$\begin{bmatrix}\Pi_1\\ \pi_2\end{bmatrix}=\begin{bmatrix}\mathbf{q}_1'\\ \mathbf{Q}_2\end{bmatrix}\begin{bmatrix}\mathbf{A}_1'\mathbf{b}-\mathbf{c}_1\\ \mathbf{A}_2'\mathbf{b}-\mathbf{c}_2\end{bmatrix}-\begin{bmatrix}\mathbf{q}_1\\ \mathbf{Q}_2\end{bmatrix}'\begin{bmatrix}\mathbf{q}_1\\ \mathbf{q}_2\end{bmatrix}-\begin{bmatrix}\mathbf{q}_1\\ \mathbf{Q}_2\end{bmatrix}'\begin{bmatrix}\mathbf{A}_1\mathbf{A}_1-\mathbf{I} & \mathbf{A}_1'\mathbf{A}_2\\ \mathbf{A}_2'\mathbf{A}_1 & \mathbf{A}_2\mathbf{A}_2-\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{q}_1\\ \mathbf{q}_2\end{bmatrix} \tag{152}$$

I have partitioned the profits vector into a scalar Π_1, which collects the joint profits of the new entity, and vector π_2, in which I stack the profits of all the other companies that are not included in the merger. If there are n firms and two of them are merging, this is a $(n-1)$ dimensional column vector. The system of first order condition solved by the surviving firms is:

$$\begin{bmatrix}0\\ 0\end{bmatrix}=\begin{bmatrix}\mathbf{A}_1'\mathbf{b}-\mathbf{c}_1\\ \mathbf{A}_2'\mathbf{b}-\mathbf{c}_2\end{bmatrix}-2\begin{bmatrix}\mathbf{q}_1\\ \mathbf{q}_2\end{bmatrix}-\begin{bmatrix}2\mathbf{A}_1\mathbf{A}_1-\mathbf{I} & \mathbf{A}_1'\mathbf{A}_2\\ \mathbf{A}_2'\mathbf{A}_1 & \mathbf{A}_2\mathbf{A}_2-\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{q}_1\\ \bar{\mathbf{q}}_2\end{bmatrix} \tag{153}$$

Proof to Proposition 5. To simulate the breakup of firm 1 into N firms, we add $N-1$ rows to the profit vector and partition it into π_1, the profit vector of the resulting entities, and π_2, the profit vector of firms that were not broken up. Similarly, we partition the new matrix of coordinates $\mathbf{A}$ into $\mathbf{A}_1$ $\mathbf{A}_2$. To form the profit function π_1, we use the assumption that all child firms are identical to their parent and therefore $\mathbf{A}_1'\mathbf{A}_1 = \mathbf{1}\mathbf{1}'$:

$$\begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1\left(\mathbf{A}_1'\mathbf{b}-\mathbf{c}_1\right) \\ \mathbf{Q}_2\left(\mathbf{A}_2'\mathbf{b}-\mathbf{c}_2\right) \end{bmatrix} - \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{1}\mathbf{1}'-\mathbf{I} & \mathbf{1}\mathbf{a}_1'\mathbf{A}_2 \\ \mathbf{A}_2'\mathbf{a}_1\mathbf{1}' & \mathbf{A}_2\mathbf{A}_2-\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} \tag{154}$$

This new set of firms solves the following system of first order conditions:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1'\mathbf{b}-\mathbf{c}_1 \\ \mathbf{A}_2'\mathbf{b}-\mathbf{c}_2 \end{bmatrix} - 2\begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{1}\mathbf{1}'-\mathbf{I} & \mathbf{1}\mathbf{a}_1'\mathbf{A}_2 \\ \mathbf{A}_2'\mathbf{a}_1\mathbf{1}' & \mathbf{A}_2\mathbf{A}_2-\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} \tag{155}$$

because the child companies are all identical, in equilibrium they must produce a single quantity $\frac{1}{N}Q_1$. Then we can rewrite this system by collapsing the first N rows and re-writing them in terms of the scalar Q_1 rather than the vector $\mathbf{q}_1$:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1'\mathbf{b}-\mathbf{c}_1 \\ \mathbf{A}_2'\mathbf{b}-\mathbf{c}_2 \end{bmatrix} - 2\begin{bmatrix} \frac{1}{N}Q_1 \\ \mathbf{q}_2 \end{bmatrix} - \begin{bmatrix} N-1 & \mathbf{a}_1'\mathbf{A}_2 \\ N\mathbf{A}_2'\mathbf{a}_1 & \mathbf{A}_2\mathbf{A}_2-\mathbf{I} \end{bmatrix} \begin{bmatrix} \frac{1}{N}Q_1 \\ \mathbf{q}_2 \end{bmatrix} \tag{156}$$

rearranging and re-defining

$$\mathbf{q} = \begin{bmatrix} Q_1 \\ \mathbf{q}_2 \end{bmatrix} \tag{157}$$

where the first row is no longer the output of the parent company but the joint output of all child companies, we finally obtain the new set of first order conditions

$$0 \;=\; \mathbf{A}'\mathbf{b}-\mathbf{c}-2\begin{bmatrix} \frac{1+N}{2N} & 0 \\ 0 & \mathbf{I} \end{bmatrix}\mathbf{q}-\left(\mathbf{A}'\mathbf{A}-\mathbf{I}\right)\mathbf{q} \tag{158}$$

which can trivially be seen to admit a potential in the function $\Psi\left(\mathbf{q}\right)$. ☐

☐

Proof to Proposition 6. To characterize the set of solutions - that is, the set of budget-efficient allocations, let us write the Lagrangian for this problem:

$$\mathscr{L}\left(\mathbf{q}\right) = W\left(\mathbf{q}\right) + \ell\left\{\mathbf{q}'\left[\mathbf{A}'\mathbf{b}-\mathbf{c}-\left(\mathbf{I}+\mathbf{A}'\mathbf{A}\right)\mathbf{q}\right]-\overline{G}\right\} \tag{159}$$

This yields the following solution as a function of the Lagrange multiplier ℓ

$$\mathbf{q} = \left(1-\frac{\ell}{1+2\ell}\right)\left(\frac{2\ell}{1+2\ell}\mathbf{I}+\mathbf{A}'\mathbf{A}\right)^{-1}\left(\mathbf{A}'\mathbf{b}-\mathbf{c}\right) \tag{160}$$

by making the following, convenient change of variable:

$$\nu \stackrel{\text{def}}{=} \frac{2\ell}{1+2\ell} \tag{161}$$

we obtain equation 108. $\square$
