Scaling Up: How Technology and Policy Shape Firm Dynamics

Thomas Hasenzagl*

University of Minnesota

Click here for the most recent version.

This version: November 5, 2024.

First version: April 24, 2024.

Abstract

This paper develops a quantitative model of entrepreneurial choice to analyze the aggregate and welfare impacts of mergers in the presence of scale efficiencies. The model incorporates varying markups to reflect different levels of market power and varying returns to scale to reflect lower selling, general, and administrative (SG&A) costs. This framework generates a trade-off where mergers simultaneously increase market power and lower operational costs. Key model parameters are estimated using firm-level microdata from Compustat and the US Census Bureau. The model is then applied to evaluate the aggregate effects of mergers and the potential impacts of different merger policies. The analysis reveals that although mergers typically lead to higher markups, the scale efficiencies achieved through optimized SG&A functions can lower the price increases due to mergers by one-third. Furthermore, I measure that markups and the returns to scale in SG&A have increased over the last 20 years. However, stricter merger enforcement is not warranted due to the mitigating benefits of scale efficiencies.

*Email: hasen019@umn.edu. Website: thomashasenzagl.com. I am greatly indebted to my advisors Kyle Herkenhoff, Loukas Karabarbounis, Ellen McGrattan, and Mike Waugh for their invaluable advice and support. I also wish to thank Fil Babalievsky, Anmol Bhandari, Jason Hall, Martin Holm, Carlos Esquivel, William Jungerman, Tobey Kass, Moritz Lenel, Wendy Morrison, Filippo Pellegrino, Luis Perez, and Ludwig Straub for helpful comments and suggestions. All errors are my own.

1. Introduction

What are the aggregate effects of mergers in the presence of scale efficiencies? To answer this question, I develop a quantitative model of mergers with varying markups and varying returns to scale. With varying returns to scale mergers lead to cost reductions and efficiency improvements, potentially lowering prices for consumers. However, with varying markups they also increase firms' market power, resulting in higher prices that may harm consumers. The net effect of a merger thus depends on this trade-off.

The aggregate effects of mergers are particularly relevant in the context of the recent wave of consolidation across industries. Worldwide, the value of merger and acquisition transactions has more than doubled between 1995 and 2016, with many firms citing cost efficiencies as a primary motivation for these transactions.[1] This wave of mergers has also raised concerns about the potential negative effects on competition and consumer welfare. In the United States, the Federal Trade Commission (FTC) and the Department of Justice (DOJ) have been increasingly scrutinizing mergers for potential antitrust violations.[2]

In my theoretical model, I integrate SG&A functions into the broader framework of firm dynamics. The model allows for varying returns to scale in SG&A functions and varying markups to reflect different levels of market power. It also features endogenous mergers, and entry and exit of firms. The model is estimated using firm-level data from Compustat and the US Census Bureau. First, I confirm the prediction that firms have increasing returns to scale in SG&A functions. Second, I find that the returns to scale in SG&A functions have increased over time, suggesting that firms have become more efficient in their SG&A operations, potentially due to technological advancements.

I then use the model to conduct two quantitative exercises to assess the aggregate implications of mergers. First, I study the effects of mergers on the aggregate economy, assessing the implications for prices, output, and wages. When two firms merge the acquirer captures the demand, or customer base, for the target's product, increases its market share, and scales up its operations. The increase in the acquirer's market share leads to higher markups, while the scale efficiencies from the merger lead to lower costs for the acquirer. The net effect of the merger on prices and output depends on the relative magnitudes of these two effects. In equilibrium, there are two types of mergers. In the first type, scale efficiencies dominate, typically when the productivity differential between the acquirer and the target is large and the demand for the target's products is small. In the

[1]See Institute for Mergers, Acquisitions and Alliances (IMAA) (2024).
[2]See DOJ and FTC (2023).

second type, the demand for the target's product becomes so large that the acquirer can increase markups to a level that offsets the cost savings from scale efficiencies.

In equilibrium, the second type of mergers, which leads to higher prices and markups and lower output, occurs more frequently than the first type. The reason for this is that acquirers base their purchasing decisions on the target firms' demand states and not on the target's productivity. This leads to a selection effect where the mergers that are most profitable to acquirers are those that increase markups. The scale efficiencies from these mergers are not sufficient to offset the negative effects of higher markups, resulting in lower output and higher prices.

However, this does not imply that cost savings through scale efficiencies, should be ignored when evaluating the effects of mergers on aggregate outcomes. In the second exercise, I simulate a specific form of technical change that increases the returns to scale in SG&A, based on increases observed in the data from the periods 1976 to 1995 and 1995 to 2016. This change leads to cost savings through a reduction in the average number of SG&A workers per firm and shifts the distribution of firm sizes towards larger firms, which charge higher markups. Initially, observing this rise in markups might suggest the need for more aggressive antitrust measures in the later period. However, the increase in scale efficiencies from SG&A functions in the later period implies that more aggressive policy measures are not necessary, as the cost savings from SG&A functions offset the negative effects of higher markups. In that case, if antitrust measures are too aggressive, they will block mergers that generate gains from scale economies, leading to lower output and higher prices.

The paper begins with a review of the relevant literature (Section 1.1), followed by a detailed discussion of a production technology with varying returns to scale via SG&A workers (Section 2). The quantitative model is then introduced (Section 3), with subsequent sections describing the data (Section 4), presenting preliminary results (Section 5), and concluding with a discussion of the broader implications of the findings (Section 6).

1.1. Literature Review

Firm Scales Over Time. This paper relates to the empirical work of De Loecker et al. (2020), who estimated production functions for publicly traded U.S. firms, finding that the average scale elasticity of firms has increased since the 1970s. Subsequently, Hasenzagl and Perez (2023) expanded on these findings, also emphasizing the growing importance of SG&A activities within firms. Building on their work, this paper develops a theoretical model that incorporates auxiliary activities to study how they influence the welfare and aggregate implications of mergers. Importantly, if firms operate at higher returns to scale, the potential efficiency gains from mergers could be larger, given that higher returns to scale imply that producers can increase output by increasing inputs less than proportionately. This paper also connects to a broad literature on recent changes in the firm size distribution in the U.S. (Decker et al.; 2016; Hsieh and Rossi-Hansberg; 2023).

Efficiency gains of mergers. A large empirical literature has been dedicated to estimating the effects of competition on productivity. Holmes and Schmitz Jr (2010) argue that increased competition leads to higher productivity. Blonigen and Pierce (2016) find that mergers in the U.S. manufacturing sector do not generate productivity gains. Ashenfelter et al. (2015) and Blonigen and Pierce (2016) find merger induced efficiency gain of 2% in the US beer industry and the French dairy dessert market, respectively. This paper contributes to this literature by providing a theoretical framework to study the efficiency gains of mergers. It also analyzes how technological change can affect the efficiency gains of mergers.

Mergers and the Macroeconomy. This paper examines the macroeconomic implications of mergers, focusing on their impact on the aggregate economy. This approach contrasts with the bulk of existing literature that primarily addresses the microeconomic effects of mergers. An important exception is David (2020), who developed a model that includes complementarities and merger-induced efficiency gains, arguing that mergers can increase aggregate output by 14%. Similarly, Guntin and Kochen (2024) explore the interplay between financial frictions and mergers using an entrepreneurial choice model akin to the one in this paper. They conclude that mergers, in the presence of financial frictions, can lead to allocative efficiency. However, unlike their model, this paper also incorporates market power, which is a critical aspect of our analysis. Additionally, Berger et al. (2023) use a structural model to study the effects of mergers on workers in the presence of monopsony power and focus on developing a set of merger guidelines that can be used to evaluate the effects of mergers on workers.

Intangible Capital. This paper is also related to the literature on intangible capital, especially the work of De Ridder (2019). They argue that intangible capital, which they define as capital that is used in production but is not physically embodied, has become increasingly important in the production process. They argue that intangible capital reduces a firm's marginal costs and increases its fixed costs, which then leads to lower productivity growth and lower business dynamism. While SG&A tasks are not typically considered intangible capital, they are related in that some SG&A tasks, such as marketing and branding, can be seen as investments into future intangible capital. The relationship between intangible capital and mergers has been studied by Bhandari et al. (2022).

2. Scale Efficiencies and SG&A Employees

2.1. Understanding SG&A Expenses

In the United States, companies with publicly traded stock are mandated by the Securities and Exchange Commission (SEC) to report their Selling, General, and Administrative (SG&A) expenses on their income statements as part of their annual 10-K filings. The Generally Accepted Accounting Principles (GAAP) in the U.S. do not provide a strict definition for what constitutes SG&A expenses, allowing companies to report these costs according to their internal accounting practices. This leads to variability in how SG&A expenses are categorized across different companies and industries.

Despite this variability, SG&A expenses typically include costs associated with the sales, marketing, and advertising departments, which are instrumental in driving the firm's revenue. They also cover overhead costs involved in managing the business, such as those related to human resources, legal, and accounting departments. These costs are categorized as operating expenses because they stem from the core operational activities of the company, rather than from production or direct manufacturing.

This paper borrows the term "SG&A workers" to include all workers primarily working on the day-to-day operations of a business and are not directly tied to the production of goods or services. In this section I argue that these tasks play a role in the determination of a firm's optimal scale of operation.

2.2. The Role of SG&A Workers in Firm Operations

I now introduce a production function based on the idea that firms produce final goods and services y by combining outputs from three inputs: capital k, production labor n_p and SG&A labor n_s, expressed as:

$$y = F(k, n_p, n_s) = z^{1-\nu} f\left(k, n_p, n_s^{\eta}\right)^{\nu}. \tag{1}$$

The variable z represents the firm manager's talent, which determines the firm's productivity. The parameter ν is the span-of-control parameter, capturing the managerial complexities that arise as firms expand, as in Lucas (1978). The parameter η governs the returns to scale for SG&A employees, which reflects how the firm's output responds to proportional changes in SG&A employees.

When $\eta > 1$, the production function exhibits increasing returns to scale with respect to SG&A employees, such as accountants. This means, that as firms grow larger they higher fewer accountants relative to their other production inputs. To see why this might be the case, consider a scenario where two equally sized firms, each with their own accounting departments, merge. After the merger, the combined firm typically consolidates these departments, employing fewer accountants than the total previously employed by the two separate firms. This reduction is feasible because certain accounting tasks—such as financial reporting and audit compliance—are required only once for the entire entity, regardless of its size. Since the output of the merged firm has effectively doubled, while the input of accounting resources has increased by less than double, this results in cost savings. These savings represent a key benefit of firm expansion.

While increasing returns to scale in SG&A employees suggest that a firm could grow indefinitely, expanding firms face increasing management challenges. These complexities are captured by the span-of-control parameter ν, typically estimated to be between 0 and 1. As the firm grows, the coordination of a larger number of operations and workers becomes more difficult. Eventually, the management complexities outweigh the benefits of scale, defining the firm's optimal size. This balance between cost efficiencies from expansion and rising managerial challenges determines the firm's scale.

The function f combines the inputs of capital, production labor, and scaled SG&A labor to produce output. In the next section, I analyze the properties of the properties if this function, focusing on the returns to scale of the firm.

2.3. Analysis of the Production Function

Proposition 1. *Let the production function $F : \mathbb{R}^3_+ \to \mathbb{R}_+$ be defined by equation (1), where f is quasi-concave on its domain and exhibits constant returns to scale in k, n_p, n_s^η. The parameters are such that $\eta > 1$ and $0 < \nu < 1$. For F to exhibit varying returns to scale, f must have non-constant elasticities of substitution. The production function F has the following properties:*

- ***Standard Properties:*** *F is non-negative, monotonic in k, n_p, n_s, and continuous across its domain.*

- ***Returns to Scale:*** *F exhibits varying returns to scale due to the interaction between ν and η, provided f has non-constant elasticities of substitution.*

- ***Quasi-Concavity:*** *F is quasi-concave because f is quasi-concave, and exponentiation by ν preserves quasi-concavity.*

Proposition 2. *Let the production function $F : \mathbb{R}^3_+ \to \mathbb{R}_+$ be defined by equation (1), where f is quasi-concave on its domain, exhibits constant returns to scale in k, n_p, and n_s^η, and allows for non-constant elasticities of substitution. The parameters are such that $\eta > 1$ and $0 < \nu < 1$. The production function F has the following properties:*

- ***Standard Properties:*** *F is non-negative, monotonic, and continuous across its domain.*
- ***Returns to Scale:*** *F exhibits varying returns to scale across different points in its domain.*
- ***Quasi-Concavity:*** *F is quasi-concave on its domain.*

Proof. See Appendix A.1.1. □

The non-negativity, monotonicity, and continuity properties of the production function stem directly from the inherent characteristics of the constituent functions and the methods used to combine them, without adding much economic meaning beyond that. However, the quasi-concavity property is essential for ensuring that the first-order conditions of the cost minimization problem are not only necessary but also sufficient for the cost function to achieve a global minimum.

The most economically meaningful property from Proposition 2 is the Returns to Scale. It implies that the production function can exhibit increasing, constant, or decreasing returns to scale, depending on the firm's size, i.e., the quantity of output the firm produces. This property also implies that the production function is non-homogeneous. The key assumption for this property to hold is that f has non-constant elasticities of substitution. This rules out the Cobb-Douglas form for f, but is consistent with f being a CES or translog function.

Next, I will provide a more detailed analysis of the returns to scale of the production function.

2.4. Local Returns to Scale

The returns to scale of a production function measure how the output of the function responds to a proportional increase in all inputs. The local returns to scale function provides a more nuanced view of this relationship by examining how the output responds to an infinitesimal proportional increase in all inputs around a specific point. This local view allows me to determine whether the production function exhibits increasing, constant, or decreasing returns to scale at a particular point. I first define the local returns to scale function for a generic production function with n-dimensional input vector $\mathbf{x}$ and then apply it to the production function defined in equation 1. [3]

Definition 1 (Local Returns to Scale)**.** *Let $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ be a differentiable production function. The* local returns to scale *function at a point $\mathbf{x} \in \mathbb{R}^n_+$, denoted by $R(\mathbf{x})$, is defined in two equivalent ways:*

1. *As the limit of the logarithmic change rate:*

$$R(\mathbf{x}) = \lim_{t \to 1^+} \frac{\log f(t\mathbf{x}) - \log f(\mathbf{x})}{\log t}, \tag{2}$$

   where t is a positive scalar approaching 1 from above.

2. *Using the gradient of f as:*

$$R(\mathbf{x}) = \frac{\mathbf{x} \cdot \nabla f(\mathbf{x})}{f(\mathbf{x})}, \tag{3}$$

   where $\nabla f(\mathbf{x})$ is the gradient of f at $\mathbf{x}$, and the dot product $\mathbf{x} \cdot \nabla f(\mathbf{x})$ represents the directional derivative of f in the direction of $\mathbf{x}$.

The function f exhibits:

- ***Constant local returns to scale** at $\mathbf{x}$ if $R(\mathbf{x}) = 1$,*
- ***Increasing local returns to scale** at $\mathbf{x}$ if $R(\mathbf{x}) > 1$,*
- ***Decreasing local returns to scale** at $\mathbf{x}$ if $R(\mathbf{x}) < 1$.*

This local returns to scale function $R(\mathbf{x})$ measures how the output of the production function f responds proportionally to an infinitesimal proportional increase in all inputs around the point $\mathbf{x}$.[4] Section **??** shows that a simple application of L'Hopital's rule establishes the equivalence between the two definitions of local returns to scale. The advantages

[3]This idea to define returns to scale for non-homogeneous production locally goes back to Frisch and Christophersen (1964).

[4]Consider the production function $f(x) = x^\nu$. Using the local returns to scale definition, we find:

$$\log f(x) = \nu \log x, \quad \log f(tx) = \nu \log(tx) = \nu(\log t + \log x), \quad \log f(tx) - \log f(x) = \nu \log t.$$

Then,

$$R(x) = \lim_{t \to 1^+} \frac{\nu \log t}{\log t} = \nu,$$

indicating that returns to scale are constant and equal to ν. This implies increasing, constant, or decreasing returns for $\nu > 1$, $\nu = 1$, and $\nu < 1$, respectively.

of the derivative-based definition are that it provides a more intuitive interpretation of the local returns to scale function as the sum of the output elasticizes of the production function and that it is easier to apply to production functions in practice.

$$R(k, n_p, n_s) = \nu \left[\frac{\partial \log f}{\partial \log k} + \frac{\partial \log f}{\partial \log n_p} + \eta \frac{\partial \log f}{\partial \log n_s^\eta} \right]. \tag{4}$$

Since the production function f is assumed to be constant returns to scale in k, n_p, n_s^η, the partial derivatives of $\log f$ with respect to $\log k, \log n_p$, and $\log n_s^\eta$ sum to one. Therefore, the assumption of constant returns to scale in the production function implies that the non-homotheticity of the production function is driven the parameters ν and η. This is a key feature of the model that allows for endogenously varying returns to scale due to cost efficiencies from scale.

In the next section, I will describe the quantitative model of mergers in order to study the interaction between market power and cost efficiencies. The model is based on the production function defined in equation and incorporates the local returns to scale properties of the production function.

3. Quantitative Model

I now introduce an entrepreneurial choice model with varying markups and varying returns to scale. The purpose of this model is to study the trade-offs associated with allowing firms to merge.

3.1. Environment, Preferences, and Production

There exists a continuum of agents indexed by $i \in [0, 1]$ who live for an infinite number of periods. Each agent is endowed with one unit of time per period, which can be allocated to firm ownership or wage labor. If agent i chooses to be a firm owner, they operate a monopolistically-competitive firm that produces variety i. In addition to the firm owner's talent z^i, each firm has a demand state ψ^i, which summarizes the firm's customer base. Managers can trade access to customer bases in the market for firms.

Throughout the paper, I use the terms "manager" and "worker" to refer to agents who choose to operate firms and work as wage laborers, respectively. The terms "firm" and "variety" are distinguished where "firm" refers to the business entities managed by managers, and "variety" refers to the specific goods or services these firms produce. The terms "SG&A labor" and "production labor" are used to denote the workers employed by managers to perform auxiliary tasks and directly produce goods and services, respectively. Additionally, I use superscript indices agents, while subscript indices are employed for varieties or firms, in order to distinguish between the roles of individuals and the products or entities they are associated with.

Preferences. Managers and workers both choose to maximize their lifetime utility, which is the sum of discounted utility from consumption. Consumer preferences are described by an indirect translog utility function. Consumer i's indirect utility function is given by:

$$\begin{aligned}\log u(c^i, P, \mathbf{p}, \boldsymbol{\psi}) = \; & \underbrace{\log(Pc^i)}_{\text{Total Expenditure}} \\ & \underbrace{- \int_0^1 \log\left(\frac{p_j}{\psi_j}\right) dj}_{\text{Mean Log Price}} \\ & + \underbrace{\frac{\gamma}{2}\left(\int_0^1 \log^2\left(\frac{p_j}{\psi_j}\right) dj - \left(\int_0^1 \log\left(\frac{p_j}{\psi_j}\right) dj\right)^2\right)}_{\text{Variance in Log Price}},\end{aligned} \tag{5}$$

where c^i is a consumption index of consumer i and P is an aggregate price index. The vector $\mathbf{p}$ is infinitely dimensional and contains the price p_j for each variety j. The vector $\boldsymbol{\psi}$ is also infinitely dimensional and contains demand states ψ_j for each variety.

The demand state ψ_j is a measure of the variety's customer base, which is determined by the variety's quality, brand, or product appeal. The demand state ψ_j is a key determinant of the variety's market power, as it captures the variety's ability to capture demand from consumers. The demand state ψ_j is also a key determinant of the variety's price, as a higher ψ_j will allow managers charge higher prices to consumers. Going forward, I will refer to ψ_j as either the variety's "demand state" or its "customer base".

I choose this utility function for four reasons. First, the utility function leads to a demand system with non-constant price elasticities, which allow the demand sensitivity to price changes to vary depending on price levels and consumption patterns. This variability is essential for the model as it enables the possibility of endogenously varying markups, crucial for modeling market power and competitive pricing strategies.

Second, the utility function is symmetric, meaning it treats all goods equivalently, ensuring utility depends only on the quantities consumed, not on the labeling or ordering of these quantities. It is also homogeneous of degree 0 in prices and expenditure, meaning that utility remains constant under proportional changes to all prices and total expenditure.

Third, preferences over the varieties are constant irrespective of the number of goods available. Specifically, there exists a choke price for each variety, above which consumers will not purchase the good. This feature allows for entry and exit of varieties without affecting the utility function's properties.

Fourth, the translog form of the utility function is highly flexible second order approximation of any symmetric and homogeneous utility function.

An application of Roy's identity to the indirect utility function 5 implies that agent i's demand for variety j is given by

$$c_j^i(p_j, \psi_j, \mathbf{p}, \boldsymbol{\psi}) = \frac{Pc^i}{p_j} s_j(p_j, \psi_j, \mathbf{p}, \boldsymbol{\psi}), \tag{6}$$

where s_j is the expenditure share of variety j. The expenditure share of variety j is given by

$$s_j(p_j, \psi_j, \mathbf{p}, \boldsymbol{\psi}) = 1 + \gamma \left(\int_0^1 \log\left(\frac{p_k}{\psi_k}\right) dk - \log\left(\frac{p_j}{\psi_j}\right) \right). \tag{7}$$

Note that the expenditure share of variety j is invariant across consumers, which is a consequence of the homogeneity in the utility function. This uniformity also extends to the aggregate level and implies that the individual expenditure share of each variety is equal to the aggregate expenditure share, which is the fraction of total market expenditure PY that is spent on variety j. The aggregate demand for variety j is given by

This allows me to compute the elasticity of total demand for variety j with respect to its own price from the expenditure share s_j:

$$\varepsilon(p_j, \psi_j, \mathbf{p}, \boldsymbol{\psi}) = 1 - \frac{\partial \log s_j(p_j, \psi_j, \mathbf{p}, \boldsymbol{\psi})}{\partial \log p_j} = 1 + \frac{\gamma}{s_j(p_j, \psi_j, \mathbf{p}, \boldsymbol{\psi})}. \tag{8}$$

Note that the elasticity of demand for variety j is decreasing in the expenditure share of the variety. This implies that the elasticity of demand is decreasing in the price of the variety. This is a key feature of the model that allows for endogenously varying markups.

Production. Each agent is endowed with one unit of time per period, which can be allocated to two occupations: management or wage labor. Each agent i has managerial productivity z^i, which experiences random fluctuations. This variability is modeled through a log-linear stochastic process:

$$\log z_t^i = \bar{z} + \rho_z \log z_{t-1}^i + \epsilon_t^i, \tag{9}$$

where ϵ_t^i represents an independently and identically distributed random shock with a mean of zero and a variance of σ_ϵ^2. This process captures the uncertainty and dynamics of individual productivity over time, and provides a basis for each consumer's choice between management and wage labor and manager's decisions to trade firms.

Manager i operates a monopolistically competitive firm and produces a distinct variety denoted by i. This manager manages a firm that serves a specific customer base ψ_i. The manager's returns, denoted by $\pi(z^i, \psi_i, \mathbf{p}, \boldsymbol{\psi})$, are derived from the following profit maximization problem:

$$\pi(z^i, \psi_i, \mathbf{p}, \boldsymbol{\psi}) = \max_{p_i, y_i, n_{p,i}, n_{s,i}, k_i} p_i y_i - w_p n_{p,i} - w_s n_{s,i} - (r+\delta) k_i, \tag{10}$$

This objective function is subject to two constraints. The first, described by equation 1, describes the firm's production capabilities:

$$y(z^i) = (z^i)^{1-\nu} f\left(k_i, n_{p,i}, (n_{s,i})^\eta\right)^\nu, \tag{11}$$

This production function shows the relationship between a firm's output y_i and the inputs capital k_i, production labor $n_{p,i}$, and SG&A labor $n_{s,i}$, with efficiency governed by the manager's talent z^i.

The second constraint is the aggregate demand for the firm's variety, which is equal to the sum of the demands of all consumers for the variety:

$$y_i(p_i, \psi_i, \mathbf{p}, \boldsymbol{\psi}) = \int_0^1 c_i^k(p_i, \psi_i, \mathbf{p}, \boldsymbol{\psi})\, dk = \frac{PY}{p_i} s_i(p_i, \psi_i, \mathbf{p}, \boldsymbol{\psi}). \tag{12}$$

In equilibrium, the firm's decisions—concerning output, pricing, and input allocations—are functions of both the technological state (z^i), the demand state (ψ_i) and the market structure ($\boldsymbol{p}, \boldsymbol{\psi}$). Importantly, since the input allocations depend on the agent's demands state, the firm's cost functions-the total cost TC, the marginal cost MC, and the average cost AC-are functions of the demand state ψ_i. This is a direct consequence of the non-homotheticity of the production function, which implies that the cost functions vary with firm size.

The first order conditions from the profit maximization problem imply that the optimal pricing function $p(z^i, \psi_i, \mathbf{p}, \boldsymbol{\psi})$ solves the following fixed point equation:

$$p_i(z^i, \psi_i, \mathbf{p}, \boldsymbol{\psi}) = \widetilde{\mu}(p_i(z^i, \psi_i, \mathbf{p}, \boldsymbol{\psi}), \psi_i, \mathbf{p}, \boldsymbol{\psi}) MC(z^i, \psi_i, \mathbf{p}, \boldsymbol{\psi}), \tag{13}$$

where $\widetilde{\mu}$ is the markup function given by

$$\widetilde{\mu}(p_i, \psi_i, \mathbf{p}, \boldsymbol{\psi}) = \frac{\varepsilon(p_i, \psi_i, \mathbf{p}, \boldsymbol{\psi})}{\varepsilon(p_i, \psi_i, \mathbf{p}, \boldsymbol{\psi}) - 1} = 1 + \frac{s(p_i, \psi_i, \mathbf{p}, \boldsymbol{\psi})}{\gamma}. \tag{14}$$

Optimal profits can be written in terms of markups and the returns to scale:

$$\pi(z_i, \psi_i, \mathbf{p}, \boldsymbol{\psi}) = \left(1 - \frac{R(z_i, \psi_i, \mathbf{p}, \boldsymbol{\psi})}{\mu(z_i, \psi_i, \mathbf{p}, \boldsymbol{\psi})}\right) p(z_i, \psi_i, \mathbf{p}, \boldsymbol{\psi}) y(z_i, \psi_i, \mathbf{p}, \boldsymbol{\psi}). \tag{15}$$

Note that profits are decreasing in the returns to scale and increasing in markups. If you assume that the markup is equal to one, as in the case of perfect competition, managers will only operate at decreasing returns to scale. If instead the returns to scale are fixed at one firms will only operate if they can charge a markup greater than one. In the case where R and μ are both varying, managers will only choose to operate whenever markups are greater than the returns to scale. When two firms decide to merge, the returns to scale will typically fall, the markups will increase, and the income of the manager will increase. Therefore, this relationship between markups and the returns to scale is crucial for understanding the incentives of managers to merge their firms.

3.2. Timing and Dynamic Programs

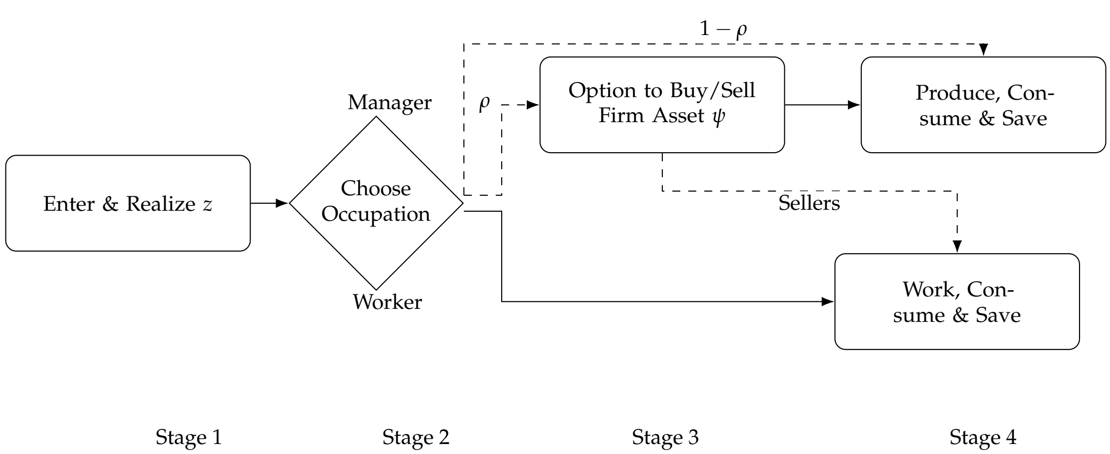

Figure 1: Stages within a period.

The timing within a period is shown in Figure 1. In the first stage of the period, consumer i enters the economy and learns their productivity z^i. In the second stage, the consumer choose their occupation for the current period. That means, they choose whether to be a manager or a worker, and if they choose to be a worker they also choose whether to be a production or SG&A worker. Workers then immediately move to stage four, where they work, consume, and save for the next period. With probability ρ manager i can move to stage three, where they randomly get matched with another firm j and can buy the firm's asset ψ_j or sell their own asset ψ_i to firm j. If manager i decides to sell, they become an SG&A worker in stage four. In the fourth stage, managers produce, consume, and save for the next period. I will now describe the value functions of the agents in each stage, starting in stage four and moving back to stage one.

Stage 4: Work, Produce, Consume & Save. I now describe the value functions of the agents in each stage in stage four. Since I am focused on stationary equilibria, I will ignore the vectors $\mathbf{p}$ and $\boldsymbol{\psi}$ in the value functions.

The state variables for the manager i are assets a^i, productivity z^i, and ψ_i. The value function of a manager in stage four is denoted by V_4 and is given by

$$
\begin{aligned}
V_4(a^i, z^i, \psi_i) = \max_{c^i, a'^i} & \left\{ u(c^i) + \beta \mathbb{E}_{z'^i} \widetilde{V}_1(a'^i, z'^i, \psi_i) \right\} \\
\text{s.t.} \quad & Pc^i + a'^i \leq \pi(z^i, \psi_i) + (1+r)a^i, \\
& c^i \geq 0, \quad a'^i \geq 0.
\end{aligned}
$$

Here, the managerial income $\pi(z^i, \psi_i)$ is given by equation 15. Note that I do not separate the firm's manager from the firm's owner and the manager receives the firm's monopoly profits as income.

The state variables for the worker i are assets a^i, productivity z^i, and occupation o^i. The value function of a worker in stage four is denoted by W_4 and is given by:

$$
\begin{aligned}
W_4(a^i, z^i, o^i) = \max_{c^i, a'^i} & \left\{ u(c^i) + \beta \mathbb{E}_{z'^i} \widetilde{V}_1(a'^i, z'^i, 1) \right\} \\
\text{s.t.} \quad & Pc^i + a'^i \leq w(o^i) + (1+r)a^i, \\
& c^i \geq 0, \quad a'^i \geq 0.
\end{aligned}
$$

The state o^i can take on two values: P for production worker and S for SG&A workers. The occupation o^i determines the wage that the worker i earns, that is $w(\mathrm{P}) = w_p$ and

$w(S) = w_s$. However, since z^i is managerial talent and not worker talent, it does not affect the wage. It still affects the worker's value function since it affects the probability of becoming a manager in the next period. Note that I normalize ψ_i to one for workers since they do not own firms and therefore do not have a customer base.

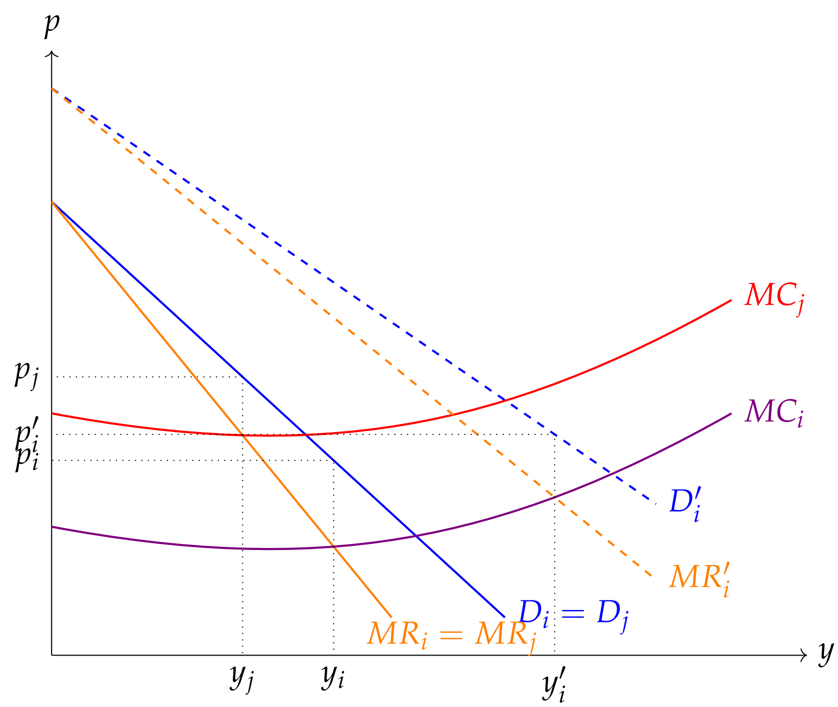

Figure 2: Graphical illustration of firm i buying firm j.

Stage 3: The Market for Firms. In stage three, the market for firm managers operates under a probabilistic matching mechanism. In the market for firms, managers trade claims to customer base ψ. With a probability ρ, each manager is randomly matched with another manager, leading to one of two potential outcomes: a merger or no merger. Upon a successful merger, the buyer, manager i, acquires the customer base ψ_j of the seller, manager j. After the acquisition, the buyer produces a new variety i', which is a composite of the buyer's and seller's varieties. This new variety has a demand state $\psi_{i'}$ that is a weighted average of the demand states of the original varieties. The demand state of the new variety is set such that the consumers' demand for the new variety reflects the individual contributions from each of the original varieties:

$$y_{i'}(p_{i'}, \psi_{i'}, \mathbf{p}, \boldsymbol{\psi}) = \frac{p_i}{p_{i'}} y_i(p_i, \psi_i, \mathbf{p}, \boldsymbol{\psi}) + \frac{p_j}{p_{i'}} y_j(p_j, \psi_j, \mathbf{p}, \boldsymbol{\psi}). \tag{16}$$

Since the demand curve is non-linear in prices, there is no demand state $\psi_{i'}$ that ensures that equation 16 holds for all prices. Instead, I pick the demand state $\psi_{i'}$ that ensures that the demand curve for the new variety is equal to the weighted average of the demand curves of the original varieties at the prices p_i and p_j that would prevail if buyer would produce both original varieties separately using the buyer's managerial talent z^i. That is, $p_i = p_j = p_i(z^i, \psi_i, \mathbf{p}, \boldsymbol{\psi})$.

Figure 3.2 illustrates the case in which firm i buys firm j. Before the merger, both firms face the same demand curve $D_i = D_j$, but different marginal cost curves MC_i and MC_j. Firm i is assumed to be more productive than firm j and therefore the marginal cost curve MC_i is below the marginal cost curve MC_j. The optimal pre-merger output levels for firms i and j are y_i and y_j, respectively. The optimal pre-merger prices for firms i and j are p_i and p_j, respectively. Suppose that post merger, firm i produces both varieties separately. Since the demand is the same for the two varieties, firm i earns revenues $2p_iy_i$. However, the merger generates a new variety i' with a new demand curve $D_{i'}$. This new demand curve $D_{i'}$ is set such that the revenue from producing the new variety is equal to the revenue from producing the two original varieties separately, that is, $p_{i'}y_{i'} = 2p_iy_i$.

Following the merger, the seller transitions into an SG&A worker role. The transfer price, θ^{ij}, paid by the buyer to the seller is determined through Nash bargaining.

The value function of a manager i in stage three is given by:

$$\begin{aligned} V_3(a^i, z^i, \psi_i) = \int \Big[& \mathbb{1}_{i \text{ buys } j} V_4(a^i - \theta^{ij}, z^i, \psi_{i'}) \\ & + \mathbb{1}_{j \text{ buys } i} W_4(a^i + \theta^{ji}, z^i, \mathrm{S}) \\ & + \mathbb{1}_{\text{no merger}} V_4(a^i, z^i, \psi_i) \Big] d\lambda_{\mathrm{M}}(a^j, z^j, \psi_j) \\ & + (1-\rho) V_4(a^i, z^i, \psi_i), \end{aligned} \tag{17}$$

where $\lambda_{\mathrm{M}}(a^j, z^j, \psi_j)$ is the distribution of managers in the economy.

This value function is an expectation over the possible outcomes of the merger. There are three possible outcomes: manager i buys firm j, manager j buys firm i, or no merger.

In the case in which manager i buys firm j, the buyer i pays the price θ^{ij} to the seller j and the buyer's assets decrease by this amount. The buyer then produces the new variety i' and adopts the demand state $\psi_{i'}$. The buyer's value function in stage four is then

$V_4(a^i - \theta^{ij}, z^i, \psi_{i'})$. The buyer's value function is then multiplied by the probability of this outcome occurring, which depends on the probability of manager i entering the market for firms ρ and the probability of manager i being matched with manager j who is randomly drawn from the distribution $\lambda(a^j, z^j, \psi_j)$.

In the second case, in which manager j buys firm i, the buyer j pays the price θ^{ji} to the seller i and the buyer's assets increase by this amount. The buyer then transitions to be a SG&A worker for the remainder of the period. The seller's value function in stage four is then $W_4(a^i + \theta^{ji}, z^i, \text{S})$. As above, the seller's value function is then multiplied by the probability of this outcome occurring.

In the third case, in which there is no merger, the manager's value function in stage four is $V_4(a^i, z^i, \psi_i)$, that is, the manager continues to operate their firm without a change to their states. Similarly, with probability $1 - \rho$, the manager does not enter the market for firms and the manager's value function in stage four is equal to $V_4(a^i, z^i, \psi_i)$.

Nash Bargaining The value of a merger may be different for the buyer and the seller. Therefore, the price θ^{ij} paid by buyer i to seller j is determined through a bargaining process. The surplus to buyer i when they buys seller j is given by:

$$S^i(\theta^{ij}) = V_4(a^i - \theta^{ij}, z^i, \psi_{i'}) - V_4(a^i, z^i, \psi_i)$$

This expression captures the difference in manager i's value function when they spend θ^{ij} to acquire firm j, compared to its outside value, which is equal to the value function of the manager when they do not buy or sell a firm. Similarly, the seller's surplus, $S^j(\theta^{ij})$, for firm j is given by:

$$S^j(\theta^{ij}) = W_4(a^j + \theta^{ij}, z^j, \text{S}) - V_4(a^j, z^j, \psi^j).$$

This is the increase in value that seller j realizes upon selling their firm and receiving the payment θ^{ij} from firm i.

The Nash bargaining solution then seeks to maximize the Nash product,

$$NP(\theta^{ij}) = S^i(\theta^{ij})^{\chi} \times S^j(\theta^{ij})^{1-\chi}$$

where χ is the bargaining power of buyer i. The maximization is subject to the constraints that both surpluses must be non-negative, ensuring that the deal is mutually beneficial.

In cases where the Nash bargaining solution is not feasible, the price θ^{ij} is set to zero, and the buyer and seller continue to operate their firms independently. In cases where it is feasible for manger i to buy firm j and for firm j to buy i, the trade that generates the highest Nash product is chosen.

Probabilities of buying and selling The probability that manager i buys firm j is given by:

$$\mathcal{P}_{\text{sell}}(a^i, z^i, \psi_i) = \rho \times \int \mathbb{1}_{i \text{ buys } j} \, d\lambda_{\text{M}}(a^j, z^j, \psi_j).$$

The transition probability for ψ from the beginning of the period to stage four of the period is given by:

$$\begin{aligned} \mathcal{P}(\psi' | a^i, z^i, \psi_i) = \rho \int & \mathbb{1}_{\psi_{i'} = \psi'} \mathcal{P}_{\text{buy}}(a^i, z^i, \psi_i; a^j, z^j, \psi_j) d\lambda_{\text{M}}(a^j, z^j, \psi_j) \\ & + (1 - \rho) \mathbb{1}_{\psi' = \psi} \\ & + \rho \mathbb{1}_{\psi' = 1} \mathcal{P}_{\text{sell}}(a^i, z^i, \psi_i) \\ & + \rho \mathbb{1}_{\psi' = \psi} \mathcal{P}_{\text{no change}}(a^i, z^i, \psi_i) \end{aligned}$$

Similarly, the probability that a manager will be at asset state $\tilde{a}$ in stage four given the states at the beginning of the period depends on the probability of buying or selling a firm in stage three:

$$\begin{aligned} \mathcal{P}(\tilde{a}^i | a^i, z^i, \psi_i) = \rho \int & \Big[\mathbb{1}_{\tilde{a}^i = a^i - \theta^{ij}} \mathcal{P}_{\text{buy}}(a^i, z^i, \psi_i; a^j, z^j, \psi_j) \\ & + \mathbb{1}_{\tilde{a}^i = a^i + \theta^{ji}} \mathcal{P}_{\text{sell}}(a^i, z^i, \psi_i; a^j, z^j, \psi_j) \Big] \, d\lambda_{\text{M}}(a^j, z^j, \psi_j) \\ & + (1 - \rho) \mathbb{1}_{\tilde{a}^i = a^i} \\ & + \rho \mathbb{1}_{\tilde{a}^i = a^i} \mathcal{P}_{\text{no change}}(a^i, z^i, \psi_i) \end{aligned}$$

Stage 2: Occupation Choice In Stage 2, agents choose their occupation for the current period. The value function at this stage is given by:

In this model, I incorporate an exogenous probability ω that influences the decision-making process at Stage 2. The agent chooses between continuing as a manager or transitioning to a worker, with an inherent risk of an involuntary change due to external factors.

The value function is defined as follows:

$$V_2(a^i, z^i, \psi_i) = \omega \max \left\{ V_3(a^i, z^i, \psi_i), \widetilde{W}_4(a^i, z^i) \right\} + (1-\omega)\widetilde{W}_4(a^i, z^i),$$

where ω is the probability that forces the manager to shut down their firm and become a worker, even if the value from being a manager, $V_3(a^i, z^i, \psi_i)$, is greater than the value from being a worker, $\widetilde{W}_4(a^i, z^i)$.

In scenarios where $V_3(a^i, z^i, \psi_i)$ exceeds $\widetilde{W}_4(a^i, z^i)$, the agent would typically prefer to continue as a manager. However, due to ω, there's a risk that the agent will be forced to transition to being a worker and will lose the customer base of their variety. Conversely, if the utility of being a worker is higher, the agent straightforwardly opts to be a worker, and the probability ω has no effect since the choice aligns with the lower-risk option. This ensures that the dynamic problem remains stationary.

This also implies that the probability of becoming a manager in stage 2 is given by:

$$\mathcal{P}_2(\mathrm{M}|a^i, z^i, \psi_i) = \begin{cases} 1-\omega & \text{if } V_3(a^i, z^i, \psi_i) > \widetilde{W}_4(a^i, z^i), \\ 0 & \text{otherwise,} \end{cases}$$

and the probability of becoming a worker at stage 2 is given by:

$$\mathcal{P}_2(\mathrm{W}|a^i, z^i, \psi_i) = \begin{cases} \omega & \text{if } V_3(a^i, z^i, \psi_i) > \widetilde{W}_4(a^i, z^i), \\ 1 & \text{otherwise.} \end{cases}$$

Here, the subscript 2 in $\mathcal{P}_2$ denotes the that this is the occupation choice probability at the end of stage two.

Agents who become workers then face a further decision: whether to become production workers (P) or SG&A workers (S). Their utility from each occupation is influenced by a taste shock $\varepsilon(o)$, which is a random factor affecting their preference for each occupation. The value function for a worker is given by:

$$\widetilde{W}_4(a^i, z^i) = \max_{o \in \{\mathrm{P,S}\}} \left\{ W_4(a^i, z^i, o) + \varepsilon(o) \right\},$$

where $o \in \{\mathrm{P,S}\}$ represents the two occupational choices. Note that the future value function for the worker is the stage four value function for the worker, W_4, since workers do not have the option to buy or sell firms and therefore skip stage three.

The taste shock $\varepsilon(o^i)$ follows a Type 1 Extreme Value distribution (also known as a Gumbel distribution), characterized by the following cumulative distribution function (CDF):

$$F(\varepsilon) = \exp(-\exp(-\varepsilon)).$$

This distribution leads to softmax choice probabilities (also called logit choice probabilities), which describe the likelihood that an agent chooses one occupation over the other. These probabilities are given by:

$$\mathcal{P}_2(o \mid a^i, z^i) \equiv \frac{\exp\left(\frac{W_3(a^i, z^i, o)}{\sigma}\right)}{\exp\left(\frac{W_3(a^i, z^i, \mathrm{P})}{\sigma}\right) + \exp\left(\frac{W_3(a^i, z^i, \mathrm{S})}{\sigma}\right)}, \tag{18}$$

where σ is the scale parameter of the Type 1 Extreme Value distribution, which governs the magnitude of the taste shocks. The subscript 2 in $\mathcal{P}_2$ denotes the that this is the occupation choice probability at the end of stage two.

In summary, the agent's decision-making process at Stage 2 involves first choosing between becoming a manager or a worker, and then, if becoming a worker, deciding between being a production worker or an SG&A worker. The agent's choice probabilities are shaped by both the expected utility from each occupation and the random taste shocks that follow the Type 1 Extreme Value distribution.

The total probability of an agent becoming a specific type of worker, either a production worker (P) or an SG&A worker (S), is computed by considering both the general probability of becoming a worker and the conditional probability of choosing a specific worker role:

$$\mathcal{P}_2(o|a^i, z^i, \psi_i) = \mathcal{P}_2(\mathrm{W}|a^i, z^i, \psi_i) \times \mathcal{P}_2(o|a^i, z^i).$$

Stage 1: Productivity Shock In Stage one, agents take expectations over their future productivity shock z'. The productivity shock z follows a stochastic process, and the transition probability of the productivity shock is given by $P(z' \mid z)$, as implied by the stochastic process in equation 9.

The agent's expected value function in Stage 1 is given by:

$$\mathbb{E}_z V_1(a^i, z^i, \psi_i) = \int V_2(a^i, z^{i'}, \psi_i)\, dP(z^{i'} \mid z^i).$$

3.3. Stationary Distribution

Before defining the model's equilibrium, I establish the stationary distributions over all states. To do so, I first need to find the probability of each agent being a manager or a worker at the end of the period in stage four. The distribution at the end of a period is the stationary distribution at stage four of that period. The law of motion for the distribution describes the transition from stage four in one period to stage four in the next period. The transition probabilities that I define in this section are the probabilities of the agent being in stage four of the current period, given their states at the beginning of the current period.

Probabilities at Stage Four. The probability that an agent is an SG&A worker at the end of stage four is defined as follows. It is the sum of the probability that the agent is an SG&A worker at the end of stage two and the probability that an agent chooses to be a manager in stage two but decides to sell their firm in stage three:

$$\mathcal{P}_4(\mathrm{S}|a^i, z^i, \psi_i) = \mathcal{P}_2(\mathrm{S}|a^i, z^i, \psi_i) + \mathcal{P}_2(\mathrm{M}|a^i, z^i, \psi_i) \times \mathcal{P}_{\text{sell}}(a^i, z^i, \psi_i),$$

where $\mathcal{P}_{\text{sell}}$ is the probability of selling the firm, as defined above. The probability that an agent remains a manager at the end of stage three is the product of the probability that the agent is a manager at the end of stage two and the complement of the probability of selling their firm:

$$\mathcal{P}_4(\mathrm{M}|a^i, z^i, \psi_i) = \mathcal{P}_2(\mathrm{M}|a^i, z^i, \psi_i) \times (1 - \mathcal{P}_{\text{sell}}(a^i, z^i, \psi_i)).$$

Distribution of Managers. The distribution of managers across states (a, z, ψ) evolves according to:

$$\lambda_{\mathrm{M}}(a', z', \psi') = \int \lambda_{\mathrm{M}}(a, z, \psi) \mathbb{1}\{a' = a'(\tilde{a}, z, \psi)\} \mathcal{P}(\tilde{a}|a, z, \psi) P(z'|z) \mathcal{P}(\psi'|a, z, \psi) \, da \, dz \, d\psi,$$

where $\mathcal{P}(\psi'|a, z, \psi)$ if the transition probability of the customer base and $\mathcal{P}(\tilde{a}|a, z, \psi)$ is the transition probability of the asset state between stages three and four, as defined above.

Distribution of Workers. The distribution of workers, λ_{W}, over states (a, z, o) follows a similar evolutionary pattern:

$$\lambda_{\mathrm{W}}(a', z', o') = \int \lambda_{\mathrm{W}}(a, z, o) \mathbb{1}\{a' = a'(a, z, o)\} P(z'|z) \mathcal{P}_3(o'|a, z, o) \, da \, dz \, do.$$

This equation reflects the changes in occupational choices based on asset policy functions, productivity shocks, and the decisions made at the end of stage three.

Mass of Managers. The mass of managers, denoted as Λ_{M}, is computed by integrating the probability of being a manager across all states:

$$\Lambda_{\mathrm{M}} = \int\int\int \mathcal{P}_3(\mathrm{M}|a,z,\psi)\,\lambda(a,z,\psi)\,da\,dz\,d\psi,$$

where $\lambda(a,z,\psi)$ is the distribution of agents across states (a,z,ψ).

Mass of Workers. Similarly, the mass of workers, denoted as Λ_{W}, is calculated by integrating the complement of the manager's probability (i.e., the probability of being a worker) across all states:

$$\Lambda_{\mathrm{W}} = \int\int\int (1-\mathcal{P}_3(\mathrm{M}|a,z,\psi))\,\lambda(a,z,\psi)\,da\,dz\,d\psi.$$

Full Distribution of Agents. The full distribution of agents, denoted by λ, is a weighted combination of the distributions of managers and workers over the state space (a,z,ψ,o), where o represents the occupational role (manager or worker). It is defined as:

$$\lambda(a,z,\psi,o) = \Lambda_{\mathrm{M}}\cdot\lambda_{\mathrm{M}}(a,z,\psi) + \Lambda_{\mathrm{W}}\cdot\lambda_{\mathrm{W}}(a,z,o),$$

3.4. Aggregation

Given the distribution over agents λ, I can compute the aggregate variables in the economy. The aggregate output, capital, and labor demands can be computed using the distribution over managers λ_m and are given by:

$$\begin{aligned}
Y &= \int \left(\frac{p(z,\psi,\mathbf{p},\boldsymbol{\psi})}{P} y(z,\psi,\mathbf{p},\boldsymbol{\psi})\right)\lambda_{\mathrm{M}}(a,z,\psi)\,da\,dz\,d\psi,\\
K &= \int k(z,\psi,\mathbf{p},\boldsymbol{\psi})\lambda_{\mathrm{M}}(a,z,\psi)\,da\,dz\,d\psi,\\
N_p^d &= \int n_p(z,\psi,\mathbf{p},\boldsymbol{\psi})\lambda_{\mathrm{M}}(a,z,\psi)\,da\,dz\,d\psi,\\
N_p^s &= \int n_s(z,\psi,\mathbf{p},\boldsymbol{\psi})\lambda_{\mathrm{M}}(a,z,\psi)\,da\,dz\,d\psi.
\end{aligned} \tag{19}$$

Labor supply is given by the mass of workers in each occupation:

$$\begin{aligned}
N_p^s &= \int \lambda_{\mathrm{W}}(a,z,\mathrm{P})\,da\,dz,\\
N_s^s &= \int \lambda_{\mathrm{W}}(a,z,\mathrm{S})\,da\,dz.
\end{aligned}. \tag{20}$$

In the stationary equilibrium, the aggregate assets in the economy are given by:

$$A = \int a'(a, z, \psi, o)\lambda(a, z, \psi, o)\, da\, dz\, d\psi\, do. \tag{21}$$

3.5. Equilibrium

I now define a stationary equilibrium for the model.

Definition 2 (Stationary Equilibrium)**.** *A stationary equilibrium is prices $\{r, w_p, w_s, \mathbf{p}, \boldsymbol{\theta}\}$, policy functions c and a' for workers and managers, choice probabilities over the discrete choices, and distributions over agents λ such that:*

1. *The prices $\boldsymbol{\theta}$ satisfy the Nash bargaining solution.*
2. *The policy functions solve the dynamic programs.*
3. *Managers choose $\mathbf{p}$, y, k, n_p, and n_s to maximize profits.*
4. *The distribution over agents is stationary.*
5. *Markets clear. That is,*
   (a) *Asset market: $A = K$.*
   (b) *Labor markets: $N_p^d = N_p^s$ and $N_s^d = N_s^s$.*
   (c) *Goods market: $PY = PC + \delta K$.*

3.6. Computational Approach

Solving the model requires solving the dynamic programs for managers and workers, finding the Nash bargaining solution, and computing the stationary distribution over agents, and solving for the aggregate prices $\{r, w_p, w_s, P\}$. I begin by discretizing the state space for the dynamic programs and by normalizing the aggregate price index P to 1.

The problem requires a guess for the stationary distribution over agents λ.

3.6.1. Step 1: Pricing of Varieties

Given a guess for the stationary distribution, I can solve for equilibrium prices using equation 13. This equation is a fixed point equation for the price p and the markup μ. The fixed point equation is solved using a fixed point algorithm.

Note that the fixed point equation is also a function of the marginal cost of each variety. The marginal cost is given by the marginal cost function $MC(z, \psi)$. Due to the complexity of the non-homothetic production function, the marginal cost function does not have a closed-form solution. Instead, I solve for it using numerical root finding methods.

After computing the prices p, I can compute firm level output y from the demand curve, and factor inputs k, n_p, and n_s from the first order conditions of the profit maximization problem. Together with the guess of the stationary distribution λ, I can now solve for the aggregate output, capital, and labor demands using equations 19. I can also solve for the labor supply using equation 20 and, using the grid for assets, the aggregate assets using equation 21. This allows me to impose the market clearing conditions and compute the aggregate prices w_p, w_s, and r.

Together with the stationary distribution, the prices are also used to compute the mean and variance of the log price index, which are used in the indirect utility function in equation 5. This is used to compute the indirect utility function for each agent when solving the dynamic programs.

3.6.2. Step 2: Dynamic Programs

Given the incomes of manager and workers, the aggregate mean and variance of the log price index and the aggregate prices I can solve the dynamic programs using value function iteration. For managers this involves solving the Nash bargaining problem for every possible merger, so that I can compute the expected value in equation 17. The solution to the dynamic programs gives the policy functions for consumption and savings for workers and managers, and the occupation choice probabilities for workers, the transition probabilities for ψ and the Nash bargaining prices.

3.6.3. Step 3: Stationary Distribution

Given the policy functions and the choice probabilities, I can update the stationary distribution using the laws of motion for the distribution. The distribution are updated until they converge.

4. Parametrization

I parameterize the model using data from the U.S. Census Bureau and the Compustat database. First, I use the data to estimate the parameters of the utility function, the production function, and the stochastic process for productivity. Conditional on these estimates, I then calibrate the parameters that govern the merger decisions to data on the frequency of mergers and the parameters that govern the occupation choice to data on the relative fraction of production and SG&A worker in the US economy.

I run the production function and markup estimation both Census data and Compustat data. The reason is that both datasets have their advantages and disadvantages. The Census data contains detailed information on production inputs and outputs, but is limited to the manufacturing sector. The Compustat data contains detailed financial information on firms, including a measure of SG&A costs, but is limited to publicly traded firms.[5]

4.1. Data

I use two sets of data sources for this paper: (1) the Longitudinal Business Database (LBD), the Census of Manufacturing and Annual Survey of Manufactures, and the Census of Auxiliary Establishments (AUX) from the U.S. Census Bureau, and (2) the Compustat database.

Longitudinal Business Database (LBD) The LBD is a longitudinal database that contains the universe of employer firms in the United States. The LBD is constructed from the Business Register, which is a census of all employer firms in the United States and contains information on firm characteristics, such as employment, payroll, industry classification, and ownership changes.

The LBD covers all business in the US except for non-employer firms, such as self-employed individuals, and firms that are not legally operating.

Census of Manufacturing (CMF) and Annual Survey of Manufactures (ASM) The CMF and ASM are annual surveys of manufacturing establishments in the United States. The CMF is a census of all manufacturing establishments in the United States and takes place every five years, in years that end in 2 and 7. The ASM is a sample survey of manufacturing establishments in the United States and takes place every year. The CMF and ASM contain detailed information on establishment characteristics, such as employment, payroll, and industry classification.

Compustat The U.S. Compustat database is a comprehensive database of financial information on publicly traded companies in the United States. The database contains

[5] The results using Census data are in the process of being disclosed and will be added to this paper once they are available.

information on firm characteristics, including revenues and expenditure on SG&A. The advantage of Compustat is that it contains detailed financial information on firms, which is not available in the LBD or LEHD. The disadvantage of Compustat is that it only contains information on publicly traded firms.

4.2. Census Variables

I now describe the variables that I will use in the production function and markup estimation on Census data. All nominal variables are deflated using the NBER-CES deflators.

Output The output variable is total value of shipments, adjusted for inventories.

Production Labor Production labor is measured as the total number of production hours worked at the establishment.

SG&A Labor SG&A labor is constructed as the sum of three primary components:

1. **Plant-Level SG&A Workers:** I calculate SG&A labor at the plant level by determining the total number of non-production hours worked at each establishment. This is done by subtracting the total hours worked by production workers from the total hours recorded at the establishment.

2. **Outsourced SG&A Labor:** To account for potentially outsourced SG&A activities, I use a measure of outsourced hours based on the total expenditure on services such as legal, accounting, IT, and advertising services. I convert these expenditures into labor hours using the average wage and average annual hours typical for SG&A workers.

3. **Auxiliary Establishments:** Employment at auxiliary establishments is quantified using data from the Longitudinal Business Database (LBD). First, I use the AUX dataset to identify the NAICS codes of establishments that likely serve as auxiliary establishments. Most of these establishments fall under two-digit NAICS codes 48 (Transportation), 49 (Warehousing), 51 (Information), 54 (Professional), 55 (Management), 56 (Administrative), and 81 (Repair) as auxiliary establishments. For each firm with auxiliary establishments, I compute the ratio of auxiliary workers to total workers and apply this ratio to the total number of workers at the firm's manufacturing establishments.

Capital Capital stock is measured as the sum of equipment and structures, both computed via the perpetual inventory method.

4.3. Production Function Estimation

I estimate the production function using a translog specification and the Olley and Pakes (1996) method. The general translog production function is given by:

$$\begin{aligned}\log y &= (1-\nu)\log(z) + \nu \log f(k, n_p, n_s^{\eta}) \\ &= \nu\alpha_k \log(k) + \nu\alpha_p \log(n_p) + \nu\eta\alpha_s \log(n_s) \\ &\quad + \frac{1}{2}\nu\alpha_{pp}(\log(n_p))^2 + \frac{1}{2}\nu\alpha_{ss}\eta^2(\log(n_s))^2 + \frac{1}{2}\nu\alpha_{kk}(\log(k))^2 \\ &\quad + \nu\alpha_{pk}\log(n_p)\log(k) + \nu\eta\alpha_{ps}\log(n_p)\log(n_s) + \nu\eta\alpha_{ks}\log(k)\log(n_s).\end{aligned}$$

Let j refer to a firm in the Compustat data or a manufacturing establishment in Census data and let t refer to a year. The goal is to estimate the parameters ν, η, and the α coefficients in the following setting:

$$\log y_{it} = \nu \log f(k_{it}, n_{p,it}, n_{s,it}^{\eta}) + (1-\nu) z_{it} + \xi_{it},$$

where ξ_{it} is a measurement error term that satisfies $\mathbb{E}_t(\xi_{it}|k_{it}, n_{p,it}, n_{s,it}) = 0$. Productivity shocks z_{it} are assumed to follow a Markov process with transition probabilities $\mathcal{P}(z_{it+1}|z_{it})$. Shocks to productivity are denoted by ζ_{it} and therefore the productivity process can be written as:

$$z_{it} = \mathbb{E}_t(z_{it}|z_{it-1}) + \zeta_{it} = g(z_{it-1}) + \zeta_{it}.$$

In the Olley and Pakes (1996) method firm investment i_{it} is used as a proxy variable to estimate unobserved productivity (z_{it}). The key assumption is that there exists an invertible relationship between investment and productivity, which is captured by the following equation:

$$z_{it} = h(i_{it}, k_{it}).$$

To estimate the production function, estimation procedure then involves two steps. In the first step, I estimate regression of y_{it} on the labor inputs $n_{p,it}$, $n_{s,it}$ and third order polynomial terms for capital k_{it} and investment i_{it} in order to capture the potentially non-linear relationships described by the h function. In the second step, I construct estimates of the productivity shocks $\hat{z}_{it}$ using the estimated coefficients from the first step and run a third-order polynomial regression of $\hat{z}_{it}$ on $\hat{z}_{it-1}$ to obtain estimates of the transition function g and the productivity shocks ζ_{it}. Then, I can use the estimated productivity shocks to build a GMM estimator that exploits the orthogonality of the productivity shocks to the lagged inputs $n_{p,it-1}$, $n_{s,it-1}$ and i_{it-1}, following De Loecker and Warzynski (2012).

The estimated production function parameters are combinations of ν, η, and the α coefficients. In order to separate these parameters, I impose two additional conditions:

$$\alpha_p + \alpha_s + \alpha_k = 1 \quad \text{and} \quad \alpha_{pp} + \alpha_{ss} + \alpha_{kk} + 2\alpha_{pk} + 2\alpha_{ps} + 2\alpha_{ks} = 0.$$

These conditions are necessary conditions for the function f to exhibit constant returns to scale while maintaining flexible input substitution. They also allow me to identify the parameters ν, η, and the α coefficients.

The production function estimation allows me to compute the output elasticities with respect to each input, which are used to compute the markups in the next step.

4.4. Markup Estimation

The markups are estimated using the Hall (1988) method. Hall's derived an expession for the markup from the first order conditions of a cost minimization problem with respect to a variable input. Formally, he showed that the markup can be computed as

$$\mu = \frac{\phi_j}{\kappa_j}, \tag{22}$$

where $\phi_j \equiv (\partial y / \partial x_j)$ is the elasticity of output with respect to variable input j, and $\kappa_j \equiv p_j x_j / (py)$ is the revenue share of input j.

I choose materials as the variable input and obtain output elasticities from the production function estimation. Then I compute establishment-level markups using equation 22 and aggregate them up for each year using a harmonic sales weighted average as in Hasenzagl and Perez (2023) to find the aggregate markup in the economy, μ^{data}.

In the model, markups are given by equation 14, which can be aggregated over firms to yield an aggregate markup in the model economy that is given by $\mu^{model} = 1 + \frac{1}{\gamma}$. The

Table 1: Parameter Estimations from Production Function Analysis

Parameter	1976-1995	1996-2016
Production function estimation with Compustat data		
η Returns to scale in SG&A	1.30	1.45
ν Span of control parameter	0.96	0.96
γ Price dispersion sensitivity	5.88	4.34
Production function estimation with Census data		
η Returns to scale in SG&A	–	–
ν Span of control parameter	–	–
γ Price dispersion sensitivity	–	–

parameter γ is then exactly pinned down by the estimated aggregate markup:

$$\gamma = \frac{1}{\mu^{data} - 1}.$$

Table 4.4 shows the parameter estimates from the production function analysis for two periods 1976-1995 and 1996-2016. The table shows the returns to scale in SG&A, the span of control parameter, and the price dispersion sensitivity. Detailed results on all the estimated parameters are available in appendix 4. Note that, in the Compustat sample the returns to scale in SG&A are estimated to be 1.3 for the period 1976-1995 and 1.45 for the period 1996-2016, indicating increasing returns to scale in SG&A in both periods. In the later period the returns to scale are higher, which is consistent with the evidence from Hasenzagl and Perez (2023). The span of control parameter is estimated to be 0.96 in both periods, indicating that managers have a high degree of control over their firms. This is larger than typical estimates for the span of control in the literature, which are typically around 0.8. These values are usually estimated using the simulated method of moments in models without markups or scale efficiencies.

The price dispersion sensitivity is estimated to be 5.88 in the period 1976-1995 which corresponds to an aggregate markup of 1.17 in that period. It is estimated to be 4.34 in the period 1996-2016 indicating a higher aggregate markup of 1.23 in that period.

4.5. Assigned Parameters

I set the discount factor β to 0.96, which implies an annual discount rate of 4%. I set the capital depreciation rate γ to 0.1 and the bargaining power of the buyer firm χ to 0.5.

I use the data on the distribution of firm sizes to construct the grid for the demand states ψ. The grid is constructed such that the demand states are evenly spaced across the distribution of firm sizes. The grid is then used to compute the demand states ψ for each firm in the model.

4.6. Other calibrated parameters

The main parameters governing the merger decisions is ρ, the probability of a manager entering the market for firms, which I calibrate to match the frequency of mergers in the data. The parameters governing the merger decisions are ω, the exogenous probability that a manager becomes a worker, and σ, the scale parameter of the Type 1 Extreme Value distribution. I calibrate these parameters to match the fraction of entrepreneurs in the economy and the fraction of workers in the US economy that are SG&A workers.

5. Results

5.1. Returns to scale and cost functions

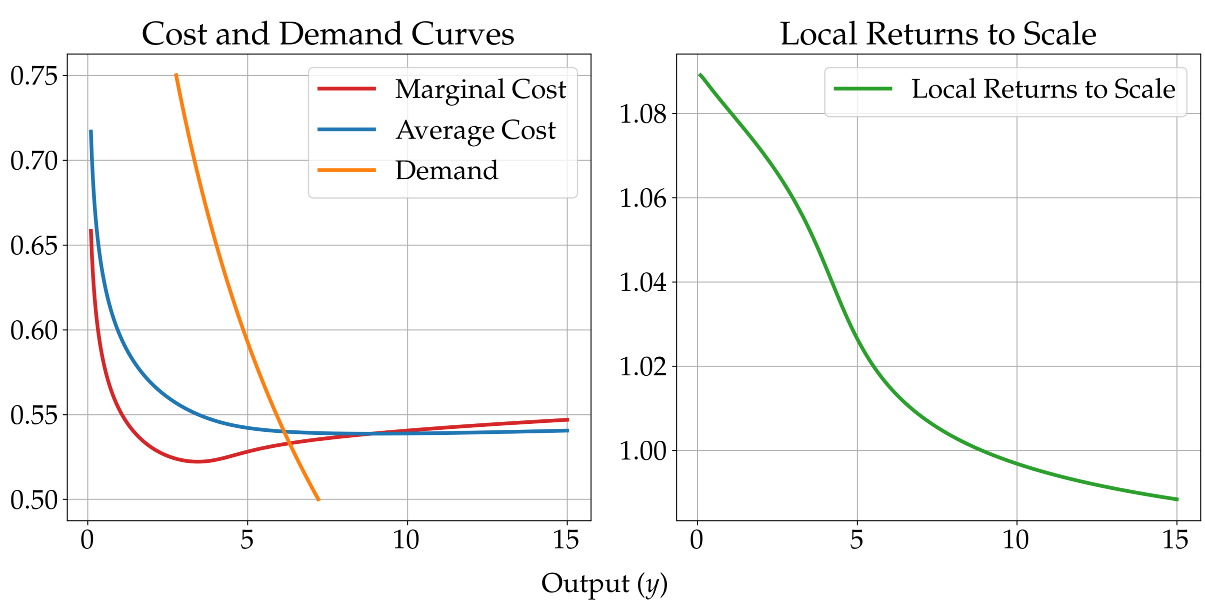

Figure 3: Cost and demand curves on the left showing Marginal Cost, Average Cost, and Demand. Local Returns to Scale on the right as a function of output.

Figure 3 shows the cost and demand curves for a typical firm in the model economy. The left panel displays the marginal cost, average cost, and demand curves. Both the marginal and average cost curves exhibit a U-shaped pattern. The firm experiences increasing returns to scale when the average cost curve lies above the marginal cost curve, and decreasing returns to scale when the average cost curve is below the marginal cost curve. The point where the marginal cost curve intersects the minimum of the average cost curve represents the point of constant returns to scale. The right panel plots the these changes in local returns to scale as a function of output.

Importantly, the demand curve intersects the cost curves in the region with increasing returns to scale. Given that profits diminish with increasing returns to scale, the firm would require sufficient market power to compensate for these losses in order to operate. This is consistent with the evidence from Hasenzagl and Perez (2023), who find that aggregate measures of returns to scale have increased to levels that are consistent with increasing returns to scale.

Importantly, with increasing returns to scale, mergers can lead to cost efficiencies that can be passed on to consumers in the form of lower prices. The reason is that the merged firm can produce at a lower average cost than the sum of the costs of the individual firms, which allows it to lower prices while maintaining profitability.

5.2. The aggregate effects of mergers

Now I turn to the aggregate effects of mergers in the model economy. I compare the effects of mergers on firm level and aggregate variables across two distinct equilibria to understand the importance of scale efficiencies. In the first scenario, no scale efficiencies are present ($\eta = 1$). In the second scenarioη is equal to 1.3, as I estimate for the period from 1976 to 1995 and in the third scenario η is equal to 1.45, which I estimate for the period from 1996 to 2015.

Table 2: Impact of mergers on firm level variables and economic aggregates.

Metric	$\eta = 1$	$\eta = 1.3$	$\eta = 1.45$
I. Firm level statistics relative to no merger case (%)			
Avg. Number of Production Workers Per Firm	3.21	4.37	5.87
Avg. Number of SG&A Workers Per Firm	7.62	4.78	4.02
Avg. Amount of Output Produced Per Firm	3.44	4.87	5.45
II. Aggregate statistics relative to no merger case (%)			
Aggregate Output	-3.09	-1.97	-1.86
Aggregate Markup	3.10	5.21	6.12
Aggregate Prices	3.41	2.93	3.04

Panel I of Table 5.2 shows the impact of mergers on average, firm level variables. The results show that, with and without scale efficiencies, mergers lead to an increase in the average firm size: allowing for mergers increases the average number of production workers per firm by 3.21% in the case without scale efficiencies, by 5.37% in the case with $\eta = 1.3$, and by 5.87% in the case with $\eta = 1.45$. The average number of SG&A workers per firm increases by 7.62% in the case without scale efficiencies, by 4.78% in the case with $\eta = 1.3$, and by 4.02% in the case with $\eta = 1.45$. Note that the average number of SG&A workers per firm decreases as η increases since firms can produce more output with the same number of SG&A workers. Mergers also imply that the average amount of output produced per firm increases. In the case without scale efficiencies output per firm increases by 3.44%, in the case with $\eta = 1.3$ it increases by 4.87%, and in the case with $\eta = 1.45$ it increases by 5.45%.

Panel II of Table 5.2 shows the impact of mergers on aggregate variables. In all cases, mergers lead to a decrease in aggregate output. This is true since the mass of firms in the economy falls, even though the average firm size increases. The decrease in aggregate output is larger in the case without scale efficiencies. Markups increase in all cases, but the increase is larger when η is greater. This is true, since the mass of active managers falls in the case with scale efficiencies, so merging firms can increase their markups more easily. Aggregate prices also increase in all cases, but the increase is larger when η is smallest. With $\eta = 1$, prices increase by 3.4% while markups increase by 3.1%. The difference between the increase in prices and markups is due to the changing firm size distribution and the fact that with $\eta = 1$, firms operate with decreasing returns to scale, so marginal costs increase as firms merge and become larger. With $\eta = 1.3$, prices increase by 2.9% while markups increase by 5.2%. With $\eta = 1.45$, prices increase by 3.0% while markups increase by 6.1%. This shows that the presence of scale efficiencies can mitigate some of the negative effects of mergers on aggregate output and prices.

5.3. The effects of a simple merger policy

What do these results imply for merger policy? This section evaluates the economic impacts of allowing the FTC/DOJ to block mergers based on firm size, reporting the results as a percentage change from the baseline "all mergers allowed" equilibrium. Specifically, I explore the consequences of a policy that prevents the largest 10% and the largest 50% of firms from engaging in acquisitions.

I do this for two scenarios: one where η is equal to 1.3 and γ is equal to 5.8 , as I estimate for the period from 1976 to 1995 and one where η is equal to 1.45 an d*gamma* is equal to

4.8, which I estimate for the period from 1996 to 2015.

Table 3: The effects of different merger policies.

Policy	P	μ	Y
I. Effects of policy on aggregates with $\eta = 1.3$(%)			
Block largest 10% of firms from purchasing	-1.41	-1.18	1.32
Block largest 50% of firms from purchasing	1.20	-0.81	-0.91
II. Effects of policy on aggregates with $\eta = 1.45$ (%)			
Block largest 10% of firms from purchasing	-1.35	-1.21	1.28
Block largest 50% of firms from purchasing	1.21	-0.81	-0.87

When we block the largest 10% of firms from engaging in acquisitions, prices decrease by 1.41%, markups decrease by 1.18%, and output increases by 1.32%. This is because the largest firms have the highest markups and are least likely to generate costs savings from scale efficiencies. However, when we block the largest 50% of firms from engaging in acquisitions, average prices increase by 1.2%, even though markups decrease by 0.8%. This is because we are now blocking firms that are more likely to generate cost savings from scale efficiencies that are larger than the increase in markups.

When comparing panel I and II of Table 3, we see that the effects of the merger policy are similar across the two scenarios. This is because the increase in markups is due to a decrease in the mass of firms in the economy due to increasing returns to scale. The increase in returns to scale also increases the cost savings from merger, which cancels out the market power effect of mergers. This implies that even though aggregate markups increase as η increases, more stringent merger policies could have the effect of harming consumers by blocking mergers that generate cost savings from scale efficiencies.

6. Conclusion

This paper presents a dynamic model of firm dynamics, mergers, and occupational choice. The model features a production function with non-homotheticity and scale efficiencies, and incorporates a dynamic decision-making process for managers and workers. The model is calibrated to match key features of the US economy, including the distribution of firm sizes, the frequency of mergers, and the relative fraction of production and SG&A workers.

The results of the model suggest that mergers tend to increase prices and markups, but that this effect is mitigated in the presence of scale efficiencies. The model also shows that blocking the largest firms from engaging in acquisition activities can lead to lower prices and markups, higher output, and a slight reduction in wages. These findings highlight the importance of considering scale efficiencies and firm size when evaluating the economic impacts of mergers.

References

Ashenfelter, O. C., Hosken, D. S. and Weinberg, M. C. (2015). Efficiencies brewed: pricing and consolidation in the us beer industry, *The RAND Journal of Economics* **46**(2): 328–361.

Berger, D. W., Hasenzagl, T., Herkenhoff, K. F., Mongey, S. and Posner, E. A. (2023). Merger guidelines for the labor market, (31147).
URL: *http://www.nber.org/papers/w31147*

Bhandari, A., Martellini, P. and McGrattan, E. (2022). A theory of business transfers. Working Paper.

Blonigen, B. A. and Pierce, J. R. (2016). Evidence for the effects of mergers on market power and efficiency, *Technical report*, National Bureau of Economic Research.

David, J. M. (2020). The Aggregate Implications of Mergers and Acquisitions, *The Review of Economic Studies* **88**(4): 1796–1830.
URL: *https://doi.org/10.1093/restud/rdaa077*

De Loecker, J., Eeckhout, J. and Unger, G. (2020). The rise of market power and the macroeconomic implications, *The Quarterly Journal of Economics* **135**(2): 561–644.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status, *American Economic Review* **102**(6): 2437–71.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.102.6.2437*

De Ridder, M. (2019). Market power and innovation in the intangible economy.

Decker, R. A., Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2016). Declining business dynamism: What we know and the way forward, *American Economic Review* **106**(5): 203–207.

DOJ and FTC (2023). 2023 draft merger guidelines, *Technical report*, United States Federal Government.

Frisch, R. and Christophersen, R. (1964). *Theory of Production*, Springer Netherlands.

Guntin, R. and Kochen, F. (2024). Financial frictions and the market for firms.

Hall, R. E. (1988). The relation between price and marginal cost in us industry, *Journal of Political Economy* **96**(5): 921–947.

Hasenzagl, T. and Perez, L. (2023). The Micro-Aggregated Profit Share, *arXiv.org* .

Holmes, T. J. and Schmitz Jr, J. A. (2010). Competition and productivity: a review of evidence, *Annu. Rev. Econ.* **2**(1): 619–642.

Hsieh, C.-T. and Rossi-Hansberg, E. (2023). The industrial revolution in services, *Journal of Political Economy Macroeconomics* **1**(1): 3–42.

Institute for Mergers, Acquisitions and Alliances (IMAA) (2024). Value of mergers and acquisition (m&a) transactions worldwide from 1985 to may 2024 (in billion u.s. dollars) [graph], In Statista. Retrieved October 28, 2024, from https://www.statista.com/statistics/267369/volume-of-mergers-and-acquisitions-worldwide/.

Lucas, R. (1978). On the size distribution of business firms, *Bell Journal of Economics* **9**(2): 508–523.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry, *Econometrica* **64**(6): 1263–1297.

Appendices

"Scaling Up: How Technology and Policy Shape Firm Dynamics"

Thomas Hasenzagl

November 5, 2024

List of Appendices

A. Additional Material

A.1. The Production Problem

A.1.1. Quasi-concavity of the Production function

Consider the production function:

$$y = z^{1-\nu} f(x_p, x_s^{\eta})^{\nu},$$

- **Homothetic Function:** A function $f(x_p, x_s^{\eta})$ is homothetic if it can be represented as $h(g(x_p, x_s^{\eta}))$ where g is homogeneous and h is a strictly increasing function.
- **Quasi-Concavity:** A function f is quasi-concave if for all x, y in its domain and for any $\lambda \in [0, 1]$, the function satisfies

$$f(\lambda x + (1-\lambda)y) \geq \min(f(x), f(y)).$$

- **Given Parameters:**
  - $\nu \in (0, 1)$: This parameter affects the exponent in the power function applied to f.
  - $\eta \geq 0$: This parameter determines the power to which x_s is raised within the function f.

Proof of Quasi-Concavity

Step 1: Analyzing x_s^{η} For $\eta \geq 0$, the function x_s^{η}:

- is constant and equals 1 if $\eta = 0$,
- is an identity transformation if $\eta = 1$,
- is quasi-concave for $\eta \geq 1$ since it is convex and increasing for $x_s > 0$,
- retains quasi-concavity for $0 < \eta < 1$ as it is concave and increasing for $x_s > 0$.

Step 2: Homothetic and Quasi-Concave Nature of f Given f is homothetic and quasi-concave, and it can be represented as $f(x_p, x_s^\eta) = h(g(x_p, x_s^\eta))$, the quasi-concavity of f follows from the quasi-concavity of g and the increasing nature of h. This implies f itself is quasi-concave since the composition of a quasi-concave function with an increasing function preserves quasi-concavity.

Step 3: Power Function Quasi-Concavity $f(x_p, x_s^\eta)^\nu$ Raising f to the power ν where $\nu \in (0, 1)$ preserves quasi-concavity. This is because the operation involves raising each value of the function to a power within this range, which compresses the upper tails of the function's distribution, thus preserving the convexity of its level sets.

Step 4: Multiplication by $z^{1-\nu}$ The function $z^{1-\nu}$ is quasi-concave in z as it is a power function with $z > 0$ and $1 - \nu > 0$. The product of two quasi-concave functions that are non-decreasing in their arguments preserves quasi-concavity.

B. Indirect utility function

There is a perfectly competitive final good producer who produces the final good using the intermediate goods as inputs. The quantity of variety j at time t is denoted by y_{jt}. Let Y_t be the quantity of the final good at time t, which is a function of the infinitely dimensional price vector $\mathbf{p}_t = \{p_{jt}\}_{j\in\Omega_t}$, the infinitely dimensional vector of demand shifters $\boldsymbol{\psi}_t = \{\psi_{jt}\}_{j\in\Omega_t}$, and the scalar expenditure E_t:

$$Y_t\left(\mathbf{p}_t, \boldsymbol{\psi}_t, E_t\right) = \max_{\mathbf{y}_t} Q_t\left(\mathbf{y}_t, \boldsymbol{\psi}_t\right) \quad \text{s.t.} \quad \int_0^1 p_{jt} y_{jt}\, dj = E_t,$$

where $\mathbf{y}_t$ represents the infinitely dimensional vector of varieties, and $Q_t(\mathbf{y}_t)$ represents the final good index aggregated over all varieties at time t.

Suppose that the final output index $Y_t(\mathbf{p}_t, \boldsymbol{\psi}_t, E_{it})$ is given by the translog function:

$$\log Y_t = \log E_t + \int_\Omega \theta(j) \log\left(\frac{p_{jt}}{\psi_{jt}}\right) dj + \frac{1}{2}\int_\Omega \int_{\Omega_t} \phi(j,k) \log\left(\frac{p_{jt}}{\psi_{jt}}\right) \log\left(\frac{p_{kt}}{\psi_{kt}}\right) dj\, dk$$

Here, $\theta(j)$ and $\phi(j,k)$ are kernels. The $\theta(j)$ coefficients capture the linear effect of the log of the price of good j on the log of the consumption index. The $\phi(j,k)$ terms are the second-order coefficients, which capture the interaction effects between the logarithms of the prices p_{jt} and p_{kt}. This term accounts for how the indirect utility changes when there are simultaneous changes in the prices of two different goods, j and k. If $j = k$, $\phi(j,j)$ represents the curvature of the utility function with respect to the price of good j. If $j \neq k$, $\phi(j,k)$ represents the cross-price effect on the utility.

I impose two further assumptions on the kernels $\theta(j)$ and $\phi(j,k)$ to ensure that the problem is tractable:

1. The kernel $\theta(j) = -1$ for all j.

2. The kernel $\phi(j,j) = \gamma$ and $\phi(j,k) = -\gamma$ for $j \neq k$, with $\gamma > 0$.

These two assumptions imply that the indirect consumption index satisfies the following properties:

1. **Symmetry**: For symmetry, the cross-partial derivatives must be equal, i.e., $\phi(j,k) = \phi(k,j)$.

2. **Homogeneity of degree zero in prices and income**: This requires that if all inputs are scaled by a positive constant t, the utility function is also scaled by t. For this property to hold, we must have:

$$\int \theta(j)\, dj = -1$$
$$\int \phi(j,k)\, dj = 0 \quad \text{for all } k$$
$$\int \phi(j,k)\, dk = 0 \quad \text{for all } j$$

Then, the indirect consumption index becomes:

$$\begin{aligned}\log U_t &= \log E_t + \int \theta(j) \log\left(\frac{p_{jt}}{\psi_{jt}}\right) dj + \frac{1}{2}\int\int \phi(j,k) \log\left(\frac{p_{jt}}{\psi_{jt}}\right) \log\left(\frac{p_{kt}}{\psi_{kt}}\right) dj\, dk, \\ &= \log E_t - \int \log\left(\frac{p_{jt}}{\psi_{jt}}\right) dj + \frac{\gamma}{2}\left[\int \log^2\left(\frac{p_{jt}}{\psi_{jt}}\right) dj - \left(\int \log\left(\frac{p_{jt}}{\psi_{jt}}\right) dj\right)^2\right].\end{aligned}$$

The second order term captures potential gains from variety or differentiation. Intuitively, if some prices are much lower relative to others (high variance), firms might be able to substitute more towards those cheaper inputs, leading to a more efficient allocation of resources, which increases the overall output.

Denote the aggregate price index by P_t. The homogeneity of degree one property implies that $P_t Y_t = E_t$. Then, the log of the aggregate price index is given by:

$$\log P_t = \int_\Omega \log\left(\frac{p_{jt}}{\psi_{jt}}\right) dj - \frac{\gamma}{2}\mathrm{Var}\left(\log\left(\frac{p_{jt}}{\psi_{jt}}\right)\right).$$

Next, I find the demand function for variety j by apply Roy's identity. Note that,

$$\frac{\partial}{\partial \log p_{jt}} \left(-\int_{\Omega} \log \left(\frac{p_{jt}}{\psi_{jt}} \right) dj \right) = -1,$$

$$\frac{\partial}{\partial \log p_{jt}} \left(\frac{\gamma}{2} \int_{\Omega} \left(\log \left(\frac{p_{jt}}{\psi_{jt}} \right) \right)^2 dj \right) = \gamma \log \left(\frac{p_{jt}}{\psi_{jt}} \right),$$

$$\frac{\partial}{\partial \log p_{jt}} \left(-\frac{\gamma}{2} \int_{\Omega} \int_{\Omega} \log \left(\frac{p_{jt}}{\psi_{jt}} \right) \log \left(\frac{p_{kt}}{\psi_{kt}} \right) dj\, dk \right) = -\gamma \int_{\Omega} \log \left(\frac{p_{kt}}{\psi_{kt}} \right) dk.$$

$$y_{jt} = \frac{E_t}{p_{jt}} \left[1 + \gamma \left(\int_{\Omega} \log \left(\frac{p_{kt}}{\psi_{kt}} \right) - \log \left(\frac{p_{jt}}{\psi_{jt}} \right) dk \right) \right].$$

The demand elasticity of good j with respect to its own price is given by:

$$\varepsilon_{jt} = 1 + \frac{\gamma}{1 + \gamma \left(\int_{\Omega} \log \left(\frac{p_{kt}}{\psi_{kt}} \right) - \log \left(\frac{p_{jt}}{\psi_{jt}} \right) dk \right)}.$$

The markup μ_{jt} for variety j is given by:

$$\mu_{jt} = 1 + \frac{1 + \gamma \left(\int_{\Omega} \log \left(\frac{p_{kt}}{\psi_{kt}} \right) - \log \left(\frac{p_{jt}}{\psi_{jt}} \right) dk \right)}{\gamma}.$$

Suppose that the marginal cost of good j is given by mc_{jt}. Then, the price of good j solves:

$$p_{jt} = \left[1 + \frac{1 + \gamma \left(\int_{\Omega} \log \left(\frac{p_{kt}}{\psi_{kt}} \right) - \log \left(\frac{p_{jt}}{\psi_{jt}} \right) dk \right)}{\gamma} \right] mc_{jt}.$$

B.1. Choke Price

The choke price is the price at which the expenditure share is zero, the demand elasticity is infinite, and the markup is 1. I denote the choke price by p^c_{jt}, and it is given by:

$$p^c_{jt} = \psi_{jt} \exp \left(\frac{1}{\gamma} + \int_{\Omega_t} \log \left(\frac{p_{kt}}{\psi_{kt}} \right) dk \right).$$

- **Demand Shifter** ψ_{jt}: The choke price scales with the demand shifter ψ_{jt}, meaning that if ψ_{jt} increases, the choke price for variety j increases as well.

- **Price Distribution** $\int_\Omega \log\left(\frac{p_{kt}}{\psi_{kt}}\right) dk$: The choke price also depends on the distribution of prices and demand shifters across all varieties. If the average log price (adjusted by ψ_{kt}) is high, the choke price for variety j will be higher.

- **Substitution Parameter** γ: The parameter γ affects the choke price by determining how sensitive the demand for variety j is to changes in its price relative to others. A higher γ implies a lower choke price, indicating greater price sensitivity.

C. Market Clearing Conditions

Recall that the mass of agents is normalized to one and indexed by $i \in [0, 1]$. I now re-index the agents such that agents on $[0, \omega_m)$ are managers, agents on $[\omega_m, \omega_p)$ are production workers, and agents on $[\omega_p, 1]$ are SG&A workers.

The labor market clearing conditions are given by

$$N_t^p \equiv \underbrace{\int_0^{\omega_m} n_{jt}^p \, dj}_{\text{Demand}} = \underbrace{\int_{\omega_m}^{\omega_p} d\,i}_{\text{Supply}}, \qquad N_t^s \equiv \underbrace{\int_0^{\omega_m} n_{jt}^s \, dj}_{\text{Demand}} = \underbrace{\int_{\omega_p}^{1} d\,i}_{\text{Supply}}.$$

Aggregate capital demand is given by

$$K_t \equiv \int_0^{\omega_m} k_{jt} \, dj.$$

Aggregate asset supply is given by

$$A_t \equiv \int_0^1 a_{it} \, di.$$

The capital market clearing condition is given by

$$K_t = A_t.$$

The total income of all agents is given by

$$
\begin{aligned}
M_t &= \int_0^{\omega_m} \left[p_{jt}y_{jt} - (r_t+\delta)k_{jt} - w_t^p n_{jt}^p - w_t^s n_{jt}^s \right] dj + w_t^p \int_{\omega_m}^{\omega_p} di + w_t^s \int_{\omega_p}^{1} di, \\
&= \int_0^{\omega_m} p_{jt}y_{jt}\, dj - (r_t+\delta)\int_0^{\omega_m} k_{jt}\, dj - w_t^p \int_0^{\omega_m} n_{jt}^p\, dj - w_t^s \int_0^{\omega_m} n_{jt}^s\, dj + w_t^p N_t^p + w_t^s N_t^s, \\
&= \int_0^{\omega_m} p_{jt}y_{jt}\, dj - (r_t+\delta)K_t.
\end{aligned}
$$

Now, summing the budget constraints of all agents, I obtain the aggregate budget constraint:

$$
\begin{aligned}
\int_0^1 \int_0^{\omega_m} p_{jt}c_{ijt}\, dj\, di + \left(\int_0^1 a_{i,t+1}\, di - \int_0^1 a_{it}\, di \right) &= \int_0^1 m_{it}\, di + \int_0^1 r_t a_{it}\, di, \\
\int_0^1 \int_0^{\omega_m} p_{jt}c_{ijt}\, dj\, di + A_{t+1} - A_t &= \int_0^{\omega_m} p_{jt}y_{jt}\, dj - (r_t+\delta)K_t + r_t A_t, \\
\int_0^1 \int_0^{\omega_m} p_{jt}c_{ijt}\, dj\, di + \delta K &= \int_0^{\omega_m} p_{jt}y_{jt}\, dj, \\
\int_0^{\omega_m} p_{jt} \int_0^1 c_{ijt}\, di\, dj + \delta K &= \int_0^{\omega_m} p_{jt}y_{jt}\, dj, \\
\int_0^{\omega_m} p_{jt}c_{jt}\, dj + \delta K &= \int_0^{\omega_m} p_{jt}y_{jt}\, dj.
\end{aligned}
$$

D. Production function parameters

Table 4: Parameter Estimates for Two Periods

Parameter	**1976-1995**	**1996-2016**
ν	0.96	0.96
η	1.30	1.45
α_p	0.52	0.59
α_s	0.39	0.31
α_k	0.09	0.10
α_{pp}	0.22	0.19
α_{ps}	-0.14	-0.11
α_{ss}	0.10	0.08
α_{pk}	-0.05	-0.05
α_{sk}	0.00	-0.00
α_{kk}	0.05	0.05