Intangible Capital and the Technology of Scale : Disentangling Scalability from Market Power

Yusuf Ozkara*

April 21, 2026

Abstract

This paper shows that intangible capital fundamentally reshapes firm-level scale economies and the measurement of market power. I develop a structural nonparametric production framework that incorporates intangible capital as a factor input and external R&D as a source of productivity spillovers, and estimate it using U.S. public firm data from 1975 to 2024. I find that returns to scale increase sharply with firm size only when intangible capital is included; absent intangibles, returns to scale are approximately constant across firms. Markups rise with intangible intensity but follow an inverted-U shape in firm size, indicating that pricing power reflects asset composition rather than scale. Incorporating intangible capital lowers aggregate markup estimates by up to 20 percent, implying that traditional models overstate market power by attributing technological scale effects to price-cost margins. Large firms exhibit lower relative productivity within industries, suggesting they grow through scalability rather than efficiency. Consistent with this, exposure to external R&D spillovers declines with firm size as large firms internalize innovation through intangible accumulation. These findings suggest that intangible capital is a central source of modern scale economies and that failing to account for it leads to a systematic conflation of technological scalability with market power.

JEL Codes: D22, D24, E22, L11, O30

Keywords: Intangible Capital, Returns to Scale, Markups, Knowledge Spillovers, Production Functions, Superstar Firms

*Ph.D. in Economics, Boston College. Email: yusufozkara.usa@gmail.com.

1 Introduction

Over the last several decades, advanced economies have undergone a fundamental transformation. Industry concentration has risen (Autor et al., 2020; Van Reenen, 2018), markups have increased in many sectors (De Loecker et al., 2020), and productivity dispersion between frontier and laggard firms has widened (Akcigit and Ates, 2023). At the same time, aggregate productivity growth has slowed and knowledge diffusion across firms has weakened (Goldin et al., 2024). While existing explanations typically attribute these trends to either superior firm selection or rising market power, a deeper question remains: has the underlying production technology itself changed in ways that structurally favor scale?

This paper argues that the rise of the modern "superstar" firm reflects a shift in production technology driven by intangible capital — and that standard production-function estimates of markups, by omitting this shift, systematically conflate technological scalability with markups. While prior work emphasizes the macroeconomic importance of intangible investment and its role in firm valuation and growth (Corrado et al., 2009; Crouzet and Eberly, 2020), less is known about how intangible capital reshapes the production function itself. Are increasing returns a mechanical byproduct of firm size, or do they arise from differences in asset composition? Do large firms grow because they are more productive, or because intangible assets enable scalable expansion? And how does the accumulation of intangible capital affect knowledge diffusion across firms?

To address these questions, it is important to understand how intangible capital differs from physical capital. Intangible assets — such as patents, software, proprietary data, and organizational know-how — possess two defining economic characteristics that jointly determine their structural consequences: non-rivalry in use, which generates increasing returns within the firm, and limited excludability, which generates knowledge spillovers across firms (Crouzet et al., 2022).

First, intangible capital is largely non-rival within the firm. Because it represents codified or tacit information, it can be deployed simultaneously across multiple production processes without being depleted. A software algorithm, for example, can coordinate thousands of machines at once, whereas each physical machine performs only a single task at a time. This feature implies that the marginal cost of scaling production declines with firm size, generating increasing returns for firms able to deploy intangible assets at scale.

Second, intangible capital — particularly knowledge capital — exhibits limited excludability.[1] Although patents and intellectual property rights provide some protection, knowl-

[1]Unlike knowledge capital, organizational capital (such as firm-specific management practices, internal processes, workforce expertise, and distribution systems) is difficult to codify and transfer (Lev and Radhakrishnan, 2003; Lev et al., 2016). It is therefore largely firm-specific and does not generate comparable

edge can often be imitated, reverse-engineered, or used to inspire related innovations. This partial non-excludability generates knowledge spillovers across firms (Bloom et al., 2013).

These two properties also have direct implications for how we measure market power. Production-based markups are recovered as the ratio of a flexible input's output elasticity to its revenue share (Loecker and Warzynski, 2012). When intangible capital is excluded, its contribution to output is misattributed to observable inputs, inflating their estimated elasticities and therefore measured markups. As intangible capital becomes more central to production, this bias grows, implying that part of the rise in measured market power reflects changes in production technology rather than pricing behavior.

To quantify and correct these mechanisms, I develop a structural nonparametric production framework that incorporates intangible capital as a distinct factor input and external R&D as a productivity shifter. I estimate nonparametric production functions following the methodology of Demirer (2025), allowing output elasticities, returns to scale, markups, and spillovers to vary flexibly across firms while incorporating both labor-augmenting and Hicks-neutral productivity.[2] I measure intangible capital as the sum of accumulated R&D and organizational capital following Peters and Taylor (2017) and Ewens et al. (2025), and estimate the model using panel data on U.S. public firms from 1975 to 2024 — a time span that captures the full transition from a predominantly physical-capital-based production regime to one increasingly governed by intangible assets.

The first and central finding is that returns to scale increase sharply with firm size only when intangible capital is included in the production function. When intangibles are excluded, estimated returns to scale remain approximately constant across the firm-size distribution. Once intangible capital is incorporated, returns to scale rise monotonically with both firm size and intangible intensity, reaching values above 1.5 among large firms. This scalability is driven by a systematic reallocation of factor elasticities: labor elasticity declines with firm size but rises with intangible intensity; physical capital elasticity exhibits the opposite pattern but remains small in magnitude; and materials elasticity remains largely flat. Because the intangible capital elasticity far exceeds that of physical capital among large firms, the steep rise in returns to scale is driven primarily by intangible assets.

This shift has deepened over time: comparing the pre- and post-2000 periods reveals that scale advantages have strengthened among large firms while weakening among small firms. I also document a secular decline in both labor and physical capital elasticities over time (Autor et al., 2020), accompanied by a substantial increase in intangible elasticity —

spillovers.

[2]Distinguishing between Hicks-neutral and labor-augmenting productivity is important because a large literature argues that technological change — particularly skill-biased technological change — is closely linked to the diffusion of intangible assets and knowledge capital (Krusell et al., 2000; Acemoglu, 2002).

suggesting that the growing dominance of large firms reflects not only changes in market structure but a deeper transformation in the technology of production itself.

Second, incorporating intangible capital into the production function substantially lowers aggregate markup estimates throughout the sample period — by approximately 20 percent at the peak of the measured markup surge, as documented in Section 6. Because the omitted output contribution of intangible capital is absorbed by flexible inputs, standard production-function estimates overstate market power by misattributing technological scale effects to price-cost margins (Bond et al., 2021; Koh et al., 2020). This bias is externally validated: incorporating intangible capital improves the explanatory power of estimated markups for cross-sectional variation in Tobin's Q by approximately 10 percent after 2000, confirming that corrected markups better reflect the genuine price-cost wedge that investors price into firm value.

Beyond the aggregate bias, the cross-sectional pattern of markups reveals that pricing power is driven more by asset composition than by scale per se. Markups increase monotonically with intangible intensity but exhibit an inverted-U relationship with firm size.[3] The largest firms do not charge the highest margins — instead they appear to pursue a volume-margin trade-off, expanding output at low marginal cost rather than extracting maximum price-cost wedges, consistent with models of strategic market-share expansion (Milgrom and Roberts, 1982; Atkeson and Burstein, 2008) and with evidence that firms with cost advantages pass them through to expand sales rather than raise margins (Edmond et al., 2023; Baqaee and Farhi, 2020).

Third, R&D spillover exposure declines with firm size but increases with intangible intensity. Firms with greater intangible intensity benefit more from external knowledge, reflecting complementarities between internal capabilities and external ideas (Bloom et al., 2013; Akcigit and Ates, 2023). At the same time, larger firms exhibit substantially lower exposure to spillovers from other firms, and as economic activity shifts toward these firms, sales-weighted spillover exposure declines over time. This pattern is consistent with increased internalization of innovation among dominant firms: as scalable intangible assets enable firms to expand internally, the marginal contribution of external knowledge to productivity declines. The result connects to the literature on declining knowledge diffusion (Akcigit and Ates, 2023) and provides a production-based mechanism for it.

Finally, relative to their industry-year peers, Hicks-neutral productivity declines with firm size, and among the largest firms, declines further with intangible intensity. Firms may expand even as their relative productivity declines because intangible capital generates scale economies that allow output to grow disproportionately relative to inputs. Intangible

[3]Intangible intensity is defined as the share of intangible capital in the firm's total capital stock.

capital thus operates through different channels across the size distribution — enhancing productivity among small firms but enabling scalability among large ones. This finding qualifies selection-based interpretations of firm growth (Syverson, 2011; Hsieh and Klenow, 2009) and points to scalability rather than efficiency as the primary source of the rise of superstar firms.

Together, these findings suggest that a substantial share of what the literature attributes to market power instead reflects technological scalability rooted in intangible capital. Rising concentration and declining knowledge diffusion are better understood as consequences of a structural shift in production technology than as symptoms of anti-competitive behavior. This distinction matters both for measurement, since standard markup estimates are increasingly biased upward, and for policy, since interventions targeting price-cost margins may be poorly suited to an environment in which the underlying driver is intangible scalability.

This paper contributes to four strands of the literature. First, it contributes to the markup measurement literature by identifying intangible capital omission as a source of systematic upward bias in production-based markup estimates. De Loecker et al. (2020) document a dramatic rise in markups using the Loecker and Warzynski (2012) framework; Bond et al. (2021) and Koh et al. (2020) raise concerns about input omission and the labor share. This paper provides a structural correction that directly incorporates intangible capital into the production function, quantifies the resulting bias over a fifty-year sample, and validates the correction against external market valuations. Second, it contributes to the intangible capital literature (Corrado et al., 2009; Crouzet and Eberly, 2020; Peters and Taylor, 2017) by estimating the production function implications of intangible accumulation — the elasticities, returns to scale, and markups— rather than focusing on valuation or investment. Third, it provides a production-based mechanism for the rise of superstar firms (Autor et al., 2020) by showing that increasing returns emerge from intangible asset composition rather than physical scale or productivity advantage. Fourth, it contributes to the knowledge diffusion literature (Bloom et al., 2013; Akcigit and Ates, 2023) by embedding R&D spillovers in a structural production framework and documenting their systematic decline as large firms internalize innovation.

The remainder of the paper is organized as follows. Section 2 presents the model, beginning with a simple motivating example. Section 3 explains the estimation strategy. Section 4 describes the identification of output elasticities, markups, and knowledge spillovers. Section 5 introduces the data and empirical implementation. Section 6 presents the main results. Section 7 reports robustness checks. Section 8 concludes.

2 Model

This section develops a nonparametric production function framework that accommodates both intangible capital and knowledge spillovers. I extend the framework of Demirer (2025) by introducing intangible capital as a distinct factor input and allowing firm productivity to depend on external knowledge through R&D spillovers. I begin with a simple parametric illustration of the markup bias from omitting intangible capital, then present the full nonparametric model, the structural assumptions, and the control variables required to identify factor elasticities, markups, and spillover effects.

A Simple Illustration of the Bias

Before developing the full nonparametric framework, I illustrate the core mechanism in a simple parametric setting. Suppose the true production function is Cobb-Douglas with intangible capital:

$$Y_{it} = K_{it}^{\alpha}(K_{it}^{int})^{\beta} L_{it}^{\gamma} M_{it}^{\theta} \exp(\omega_{it}^{H}) \exp(\epsilon_{it})$$

Under cost minimization with competitive input markets, the firm-level markup is:

$$\mu_{it} = \frac{\theta}{\alpha_{it}^{M}}$$

where α_{it}^{M} is the revenue share of materials. Now suppose one omits intangible capital and estimates instead:

$$Y_{it} = K_{it}^{\tilde{\alpha}} L_{it}^{\tilde{\gamma}} M_{it}^{\tilde{\theta}} \exp(\tilde{\omega}_{it}^{H}) \exp(\epsilon_{it})$$

Because intangible capital is correlated with output through $(K_{it}^{int})^{\beta}$, and this variation must be absorbed by the remaining inputs, the estimated materials elasticity satisfies $\tilde{\theta} > \theta$. The implied markup from the misspecified model is:

$$\tilde{\mu}_{it} = \frac{\tilde{\theta}}{\alpha_{it}^{M}} > \frac{\theta}{\alpha_{it}^{M}} = \mu_{it}$$

The bias $\tilde{\mu}_{it} - \mu_{it}$ is increasing in β — the more important intangible capital is in production, the larger the overstatement of markups. Two implications follow. First, the bias is not constant across firms: firms with larger intangible stocks generate more omitted variation, leading to greater inflation of their estimated flexible-input elasticities. Second, the bias is not constant over time: as the economy becomes more intangible-intensive (Figure 1), the aggregate overstatement of markups grows — precisely the pattern documented in Section 6.

This example clarifies why standard markup estimates may confuse scalability with mar-

ket power. The output contribution of intangible capital is real, but when the production function omits it, this contribution is misattributed to flexible inputs, inflating the ratio $\tilde{\theta}/\alpha^M$ that defines the markup. At the same time, because the intangible elasticity β is absent from the sum of output elasticities, estimated returns to scale are biased downward — understating the degree of increasing returns among intangible-intensive firms. The nonparametric framework developed below corrects both biases by incorporating intangible capital directly into the production function, allowing the correction to vary flexibly across firms, industries, and time.

2.1 Production Function

In each period t, firm i produces output Y_{it} according to the following production technology:

$$Y_{it} = F_t(K_{it}, K_{it}^{int}, \omega_{it}^L L_{it}, M_{it}) exp(\omega_{it}^H) exp(\epsilon_{it}) \tag{1}$$

where K_{it} denotes physical capital, K_{it}^{int} is intangible capital, M_{it} represents intermediate materials, and L_{it} is labor. The model accommodates multi-dimensional unobserved heterogeneity. Specifically, $\omega_{it}^L \in \mathbb{R}$ captures labor-augmenting productivity, such that $\omega_{it}^L L_{it}$ represents effective labor[4]. Furthermore, $\omega_{it}^H \in \mathbb{R}$ captures Hicks-neutral productivity, which symmetrically scales output for any given input bundle. Finally, $\epsilon_{it} \in \mathbb{R}$ represents a random idiosyncratic shock to production.

The firm's inputs are partitioned into two groups according to their dynamic properties. Physical and intangible capital are *predetermined*: firms choose their investment in period $t-1$, so capital stocks are fixed before productivity is realized. Labor and materials are *flexible* inputs: firms choose them in period t to minimize production costs, conditional on their information set, $\mathcal{I}_{it}$, which includes both productivity components $(\omega_{it}^L, \omega_{it}^H)$, current capital stocks, and the firm's past history. The idiosyncratic shock is orthogonal to the firm's information set $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}] = 0$, so it is unobserved by the firm at the time of input decisions.

Input markets are assumed to be perfectly competitive, implying that firms take input prices as given. In contrast, the output market is characterized by imperfect competition, allowing firms to exert market power and charge heterogeneous markups over marginal costs.

[4]Because the nonparametric production function cannot separately identify multiple unobserved heterogeneity terms, I follow Demirer (2025) in specifying a single flexible productivity component as labor-augmenting. Three considerations motivate this choice. First, Doraszelski and Jaumandreu (2018) show that technological change is predominantly labor-augmenting. Second, cross-sectional variation in labor costs substantially exceeds that of other inputs, suggesting considerable unobserved heterogeneity in workforce composition and efficiency (Abowd et al., 1999). Third, because workers frequently internalize and carry knowledge across firms, the spillover effects of intangible assets manifest most naturally through labor productivity (Haskel and Westlake, 2017).

A distinctive feature of the model is the incorporation of intangible capital into the production function. Because both the intensity and impact of intangible assets vary substantially across industries and over time (Crouzet et al., 2022; Crouzet and Eberly, 2019), and because large firms tend to be more capital-intensive while smaller firms rely more heavily on labor (Holmes and Schmitz, 2010), I allow the production function $F_t(\cdot)$ to be industry-specific and time-varying, with output elasticities that vary across firms. This flexibility also accommodates broader structural shifts such as the decline in the labor share (Autor et al., 2020).

2.2 Assumptions

This section details the structural assumptions necessary to identify and estimate the firm's production function. The foundational assumption is weak homothetic separability, which permits the expression of labor productivity as a function of observed input choices. Building on Levinsohn and Petrin (2003), I then establish conditions to invert Hicks-neutral productivity from intermediate input demand. Finally, I define the evolution of both productivity processes, explicitly formalizing the spillover effects of external knowledge capital.

Assumption 2.1: Weak Homothetic Separability

I assume the production function takes the functional form:

$$Y_{it} = F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)exp(\epsilon_{it}) \tag{2}$$

where $h_t(\cdot)$ is a homogeneous function. This implies that the production function is weakly separable, partitioning the predetermined inputs (physical and intangible capital) from a homogeneous function of flexible inputs (materials and effective labor). While this imposes structure on the production technology, it nests widely used functional forms. For instance, the Cobb-Douglas production function:

$$Y_{it} = K_{it}^{\alpha}(K_{it}^{int})^{\beta}(\omega_{it}^L L_{it})^{\gamma} M_{it}^{\theta} exp(\omega_{it}^H)exp(\epsilon_{it})$$

satisfies this condition, as $h(\cdot) = (\omega_{it}^L L_{it})^{\gamma} M_{it}^{\theta}$ is homogeneous and separable from the capital inputs. However, Cobb-Douglas rigidly imposes a constant unit elasticity of substitution and fixed output elasticities across firms.

Similarly, the Constant Elasticity of Substitution (CES) functional form:

$$Y_{it} = \left(\beta_k K_{it}^{\sigma} + \beta_I (K_{it}^{int})^{\sigma} + \beta_l (\omega_{it}^L L_{it})^{\sigma} + (1 - \beta_k - \beta_I - \beta_l) M_{it}^{\sigma}\right)^{v/\sigma} exp(\omega_{it}^H)exp(\epsilon_{it})$$

satisfies Assumption 2.1, with $h(\cdot) = \beta_l(\omega_{it}^L L_{it})^\sigma + (1 - \beta_k - \beta_I - \beta_l)M_{it}^\sigma$. While CES is less restrictive than Cobb-Douglas, it still imposes a constant elasticity of substitution, σ, across all firms. By relying solely on weak homothetic separability, the framework maintains a highly flexible form that accommodates heterogeneous output elasticities and non-constant substitution patterns across the firm distribution.

Assumption 2.2: Firms minimize their costs

Firms choose flexible inputs (labor and materials) in period t to minimize short-run costs, conditional on their predetermined capital stocks. This static optimization framework also facilitates the recovery of firm-level markups from the first-order conditions, following the methodology of Loecker and Warzynski (2012).

Firms solve the following cost minimization problem given competitive input prices and output demand, $\tilde{Y}_{it}$:

$$\min_{L_{it}, M_{it}} \left\{ p_t^L L_{it} + p_t^M M_{it} : F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)exp(\epsilon_{it}) \geq \tilde{Y}_{it} \right\}$$

where p_t^L and p_t^M are the respective prices of labor and materials at time t. Because input markets are perfectly competitive, these prices vary over time but are common across firms.

By Shepherd's lemma and Assumption 2.1, this cost minimization yields the optimal material-to-labor ratio (proof in Appendix):

$$\tilde{M}_{it} = \frac{M_{it}}{L_{it}} = r_t(\omega_{it}^L) \tag{3}$$

where $r_t(\cdot)$ is an unknown function. Crucially, Equation 3 establishes that the optimal flexible input ratio is determined entirely by labor-augmenting productivity and is independent of Hicks-neutral productivity.

Assumption 2.3: Elasticity of substitution between effective labor and material is less than 1 or greater than 1.

To invert $r_t(\cdot)$ and recover unobserved labor-augmenting productivity, the material-to-labor ratio must be strictly monotonic in ω_{it}^L. The elasticity of $\tilde{M}_{it}$ with respect to ω_{it}^L is strictly given by $1 - \sigma_t(\omega_{it}^L L_{it}, M_{it})$, where $\sigma_t(\cdot)$ is the elasticity of substitution between effective labor and materials. I assume $\sigma_t(\cdot) \neq 1$. Consequently, $\tilde{M}_{it}$ is strictly increasing or strictly decreasing in ω_{it}^L. Under this condition, ω_{it}^L can be expressed as an inverse function of the flexible input ratio:

$$\omega_{it}^L = r_t^{-1}(\tilde{M}_{it}) \tag{4}$$

Unobserved labor productivity can now be characterized as an unknown function of the

material-to-input ratio.

Assumption 2.4: Material demand is monotone in Hicks-neutral productivity Following Levinsohn and Petrin (2003), I assume material demand is strictly increasing in Hicks-neutral productivity, conditional on capital and labor-augmenting productivity:[5]

$$M_{it} = s_t(K_{it}, K_{it}^{int}, \omega_{it}^L, \omega_{it}^H) \tag{5}$$

where $s_t(\cdot)$ is a strictly increasing function in ω_{it}^H : more productive firms demand higher materials. This framework extends the standard Levinsohn-Petrin approach in two dimensions: it incorporates labor-augmenting productivity alongside Hicks-neutral productivity, and it includes both physical and intangible capital in the state space.

By strict monotonicity, ω_{it}^H can be inverted as a function of factor inputs. Substituting Equation 4 into this inversion isolates Hicks-neutral productivity from ω_{it}^L:

$$\omega_{it}^H = s_t^{-1}(K_{it}, K_{it}^{int}, r_t^{-1}(\tilde{M}_{it}), M_{it}) = \bar{s}_t(K_{it}, K_{it}^{int}, \tilde{M}_{it}, M_{it}) \tag{6}$$

where $\bar{s}_t(\cdot)$ is an unknown function. Hicks-neutral productivity is now expressed entirely in terms of observables: physical capital, intangible capital, the material-to-labor ratio, and materials

Assumption 2.5: Physical and intangible capital are predetermined variables The physical and intangible capital are determined by investments in the preceding period, capturing a time-to-build dynamic (Kydland and Prescott (1982)):

$$K_{it} = \kappa_{1t}(K_{it-1}, I_{it-1})$$
$$K_{it}^{int} = \kappa_{2t}(K_{it-1}^{int}, I_{it-1}^{int})$$

where I_{it} and I_{it}^{int} are physical and intangible investment, respectively. The unknown functions $\kappa_{1t}(\cdot)$ and $\kappa_{2t}(\cdot)$ accommodate unobserved fixed and convex adjustment costs. This timing implies that investment becomes fully productive in the following period[6]

Assumption 2.6: Productivity shocks have a first-order Markov process

[5] I follow the material demand approach of Levinsohn and Petrin (2003) rather than the investment-based proxy of Olley and Pakes (1996) because investment is frequently lumpy and zero at the firm level, violating the strict monotonicity required for inversion. Material expenditures, by contrast, are positive and smoothly varying for nearly all firm-year observations in the sample.

[6] While the productive lag period of intangible investments is highly heterogeneous — ranging from rapidly deployable software to multi-year R&D projects— I adopt a standard one-period (annual) lag. This aligns with the annual frequency of the Compustat data and the perpetual inventory methods used to construct intangible capital stocks (Ewens et al., 2025; Corrado et al., 2009). The nonparametric specification of $\kappa_{2t}(\cdot)$ further absorbs unobserved adjustment and integration costs that may arise during this time-to-build phase.

The unobserved productivity shocks follow a joint first-order Markov process, augmented to incorporate external knowledge spillovers :

$$P(\omega_{it}^L, \omega_{it}^H \mid \mathcal{I}_{t-1}) = P(\omega_{it}^L, \omega_{it}^H \mid \omega_{it-1}^L, \omega_{it-1}^H, \sum_{j \in I, j \neq i} R_{jt-1})$$

where R_{jt} is the knowledge capital stock of firm j at time t. This specification allows firm i's productivity to depend not only on its own past productivity but also on the accumulated knowledge capital of other firms, following the endogenous growth insight that the aggregate stock of knowledge serves as a productive externality (Romer, 1990). Because knowledge capital is characterized by limited excludability, innovations generated by one firm create positive externalities for others operating in similar technological or industry environments ((Bernstein and Nadiri, 1988; Bloom et al., 2013)).

Assuming the productivity processes are continuous, I use the Skorohod representation of random variables to define labor-augmenting productivity as:

$$\omega_{it}^L = g_1 \left(\omega_{it-1}^L, \omega_{it-1}^H, \sum_{j \in I, j \neq i} R_{jt-1}, u_{it}^1 \right), \quad u_{it}^1 \mid \omega_{it-1}^L, \omega_{it-1}^H, \sum_{j \in I, j \neq i} R_{jt-1} \sim \text{Uniform}(0,1)$$

u_{it}^1 represents an unobserved innovation term. Unlike standard approaches, this disturbance is non-separable. It captures a firm's idiosyncratic labor productivity rank, conditional on past productivity and external R&D spillovers. Hicks-neutral productivity is defined analogously:

$$\omega_{it}^H = g_2 \left(\omega_{it-1}^L, \omega_{it-1}^H, \sum_{j \in I, j \neq i} R_{jt-1}, u_{it}^1, u_{it}^2 \right), \quad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H, \sum_{j \in I, j \neq i} R_{jt-1}, u_{it}^1 \sim \text{Uniform}(0,1)$$

where u_{it}^2 represents the conditional Hicks-neutral productivity rank after accounting for the past productivity and the external R&D spillovers.

3 Estimation Strategy: A Control Variable Approach

To resolve the endogeneity of flexible inputs, I use the control function approach of Imbens and Newey (2009). This methodology accommodates non-separable, multidimensional disturbances, whereas standard proxy variable techniques (Olley and Pakes (1996)) assume a single, additively separable scalar unobservable.

A control variable is characterized by the conditional distribution function of the endogenous variable given the instruments. It must satisfy two properties: strict monotonicity in

the endogenous variable and independence from the instruments. I construct distinct control variables for labor-augmenting and Hicks-neutral productivity.

Let $W_{it-1} = (K_{it-1}, K^{int}_{it-1}, \tilde{M}_{it-1}, M_{it-1})$ denote the vector of lagged inputs. By substituting the labor productivity process into the flexible input ratio derived in Equation 3, $\tilde{M}_{it}$ can be written as:

$$\begin{aligned}\tilde{M}_{it} &= r_t\big(\omega^L_{it}\big) = r_t\Big(g_1\big(\omega^L_{it-1}, \omega^H_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u^1_{it}\big)\Big) \\ &= \tilde{r}_t\Big(W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u^1_{it}\Big)\end{aligned}$$

where $\tilde{r}_t(\cdot)$ is an unknown function. Because ω^L_{it} is strictly monotonic in u^1_{it} (by the Skorohod representation) and $\tilde{M}_{it}$ is strictly monotonic in ω^L_{it} (Assumption 2.3), $\tilde{M}_{it}$ is strictly monotonic in u^1_{it}. Furthermore, the timing assumptions ensure that u^1_{it} is independent of the predetermined inputs W_{it-1} and aggregate knowledge spillovers (proof in Appendix).

Because u^1_{it} satisfies both the monotonicity and independence criteria and is uniformly distributed by definition, it can be recovered as the conditional cumulative distribution function (CDF) of the flexible input ratio:

$$u^1_{it} = F_{\tilde{M}_{it}|W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}}\Big(\tilde{M}_{it} \mid W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}\Big) \tag{7}$$

where $F_{\tilde{M}_{it}|W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}}(.)$ is the cumulative distribution function of $\tilde{M}_{it}$ conditional on $W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}$. The intuition is straightforward: among firms with identical input histories and spillover exposure, a higher material-to-labor ratio implies a higher productivity rank u^1_{it}. With u^1_{it} recovered, labor-augmenting productivity is fully characterized by observables and the control variable:

$$\begin{aligned}\omega^L_{it} &= g_1\Big(\omega^H_{it-1}, \omega^L_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u^1_{it}\Big) \\ &= c_{1t}\Big(W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u^1_{it}\Big)\end{aligned} \tag{8}$$

where $c_{1t}(\cdot)$ is an unknown function. The control variable for Hicks-neutral productivity is constructed analogously, using the monotonicity of material demand

$$\begin{aligned}M_{it} &= s_t(K_{it}, K^{int}_{it}, \omega^L_{it}, \omega^H_{it}) \\ &= s_t\Big(K_{it}, K^{int}_{it}, g_1\big(\omega^L_{it-1}, \omega^H_{it}, \sum_{j\in I, j\neq i} R_{jt-1}, u^1_{it}\big), g_2\big(\omega^L_{it-1}, \omega^H_{it}, \sum_{j\in I, j\neq i} R_{jt-1}, u^1_{it}, u^2_{it}\big)\Big)\end{aligned}$$

Substituting equation 4 and 6 into ω_{it}^L and ω_{it}^H in the equation above, material demand becomes

$$M_{it} = \tilde{s}_t\big(K_{it}, K_{it}^{int}, W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u_{it}^1, u_{it}^2\big)$$

where $\tilde{s}_t()$ is an unknown function. The timing assumptions guarantee that u_{it}^2 is independent of predetermined capital, the lagged state variables W_{it-1}, and the labor-augmenting innovation u_{it}^1. Furthermore, material demand is strictly increasing in u_{it}^2. Therefore, the control variable for Hicks-neutral productivity is identified via the conditional CDF of material demand:

$$u_{it}^2 = F_{M_{it}|K_{it},K_{it}^{int},W_{it-1},\sum_{j\in I,j\neq i} R_{jt-1}}\Big(M_{it} \mid K_{it}, K_{it}^{int}, W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}\Big) \tag{9}$$

where $F(\cdot)$ is the CDF of M_{it} conditional on $K_{it}, K_{it}^{int}, W_{it-1}, \sum_{j\in I,j\neq i} R_{jt-1}$. This implies that conditional on capital stocks, historical inputs, and knowledge spillovers, variation in material demand therefore identifies the Hicks-neutral productivity rank. Finally, ω_{it}^H can be expressed as:

$$\omega_{it}^H = c_{2t}\big(W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u_{it}^1, u_{it}^2\big) \tag{10}$$

where $c_{2t}(\cdot)$ is an unknown function. This indicates that, conditional on the two control variables, external knowledge spillovers and lagged factor inputs, Hicks-neutral productivity is fully determined — there is no residual variation.

4 Identification

This section outlines the identification of the output elasticities of factor inputs, firm-level markups, and the productivity effects of knowledge spillovers.

The cost minimization problem for the flexible inputs (materials and labor) yields the following optimality condition (proof in Appendix):

$$\frac{\theta_{it}^M}{\theta_{it}^L} = \frac{\alpha_{it}^M}{\alpha_{it}^L} \tag{11}$$

where θ_{it}^M and θ_{it}^L denote the output elasticities of materials and labor, and α_{it}^M and α_{it}^L their revenue shares. The ratio of flexible input elasticities equals the ratio of their observed revenue shares. I exploit this proportionality to identify the individual elasticities.

Let lower-case letters denote logarithms. The logarithmic production function is given

by:

$$y_{it} = f_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it})) + \omega_{it}^H + \epsilon_{it}$$

where $f_t(\cdot)$ is the logarithm of the $F_t(\cdot)$ function. Assuming $h_t(\cdot)$ is homogeneous of degree one, effective labor can be factored out: $h_t(\omega_{it}^L L_{it}, M_{it}) = L_{it} h_t(\omega_{it}^L, \tilde{M}_{it})$. Substituting the inverted labor productivity from Equation 4, this aggregate flexible input function becomes $L_{it} h_t(r_t(\tilde{M}_{it}), \tilde{M}_{it}) \equiv L_{it} \bar{h}_t(\tilde{M}_{it})$, where $\bar{h}_t(\cdot)$ is an unknown function.

Because both arguments of the original $h_t(\cdot)$ function depend on $\tilde{M}_{it}$, its internal structure is not separately identified. However, identifying the composite function $\bar{h}_t(\cdot)$ is sufficient to recover the output elasticities. The production function can thus be rewritten as:

$$y_{it} = f_t(K_{it}, K_{it}^{int}, L_{it} \bar{h}_t(\tilde{M}_{it})) + \omega_{it}^H + \epsilon_{it}$$

The following proposition establishes the identification of the flexible input elasticities.

Proposition 1. *The sum of labor and material output elasticities is equal to:*

$$\theta_{it}^M + \theta_{it}^L = f_{t3}(K_{it}, K_{it}^{int}, L_{it} \bar{h}_t(\tilde{M}_{it})) L_{it} \bar{h}_t(\tilde{M}_{it})$$

where $f_{t3}(\cdot)$ is the derivative of $f_t(\cdot)$ with respect to the third argument.

The proof is provided in the appendix. The proposition demonstrates that derivative of $f_t(\cdot)$ with respect to its third argument is equivalent to the sum of flexible input elasticities. Importantly, the sum of elasticities do not depend on the derivative of $h_t()$ - identifying $\bar{h}_t(\cdot)$ is sufficient to determine the sum of elasticities.

Substituting Equation 11 into this sum of elasticities isolates the individual labor and material elasticities

$$\begin{aligned} \theta_{it}^L &= \frac{\alpha_{it}^L}{\alpha_{it}^L + \alpha_{it}^M} f_{t3}(\cdot) L_{it} \bar{h}_t(\tilde{M}_{it}) \\ \theta_{it}^M &= \frac{\alpha_{it}^M}{\alpha_{it}^L + \alpha_{it}^M} f_{t3}(\cdot) L_{it} \bar{h}_t(\tilde{M}_{it}) \end{aligned} \tag{12}$$

Because the input revenue shares, α_{it}, are directly observed in the data, no additional estimation is needed. Once the sum of elasticities is identified, the individual elasticities follow immediately.

Following Loecker and Warzynski (2012) the first order condition of the cost minimization problem yields the expression for the firm-level markup, μ_{it}:

$$\mu_{it} = \frac{\theta_{it}^M}{\alpha_{it}^M} = \frac{\theta_{it}^L}{\alpha_{it}^L} \tag{13}$$

The markup equals the ratio of a flexible input's output elasticity to its revenue share, and can be computed equivalently using either labor or materials. I use the elasticities from Equation 12 to construct firm-level markups.

The output elasticities of physical and intangible capital are identified analogously.

Proposition 2. *The output elasticity of physical and intangible capital is*

$$\begin{aligned} \theta_{it}^{int} &= f_{t2}(K_{it}, K_{it}^{int}, L_{it}\bar{h}_t(\tilde{M}_{it}))K_{it}^{int} \\ \theta_{it}^{K} &= f_{t1}(K_{it}, K_{it}^{int}, L_{it}\bar{h}_t(\tilde{M}_{it}))K_{it} \end{aligned} \tag{14}$$

where $f_{t2}(\cdot)$ and $f_{t1}(\cdot)$ is the derivative of $f_t(\cdot)$ with respect to the second and first argument, respectively.

As with the flexible input elasticities, identifying the composite function $\bar{h}_t(\cdot)$ is sufficient; the internal structure of $h_t(\cdot)$ does not need to be characterized.

Finally, the following proposition identifies knowledge spillover elasticity - the effect of external knowledge on firm productivity.

Proposition 3. *The productivity elasticity of knowledge spillover is*

$$\theta_{it}^{spillover} = \frac{\partial c_{2t}\left(W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u_{it}^1, u_{it}^2\right)}{\partial \sum_{j\in I, j\neq i} R_{jt-1}} \frac{\sum_{j\in I, j\neq i} R_{jt-1}}{c_{2t}(\cdot)} \tag{15}$$

The spillover elasticity is the derivative of the recovered Hicks-neutral productivity function with respect to aggregate external knowledge capital, expressed in elasticity form.

5 Data and Empirical Model

This section describes the data sources, the construction of key variables, and the empirical strategy used to estimate the model.

5.1 Data

I use firm-level financial statement data from U.S. Compustat to measure intangible capital and estimate the production functions. The sample covers all public U.S. firms from 1975 to 2024. The analysis is restricted to public firms because the structural measurement of intangible capital requires granular, consistent reporting of R&D and SG&A expenses, which are systematically unavailable for private companies. Following standard sampling procedures in the literature, I exclude utilities (SIC codes 4900–4999), financial firms (SIC

Table 1: Summary Statistics

	Obs.	Mean	Std. Dev.	Min	p25	Median	p75	Max
Sales	139,475	2,270.9	12,878.0	0.25	40.7	189.8	923.0	680,985.0
Employees (thousands)	139,475	8.8	40.7	0.01	0.3	1.2	5.1	2,300.0
Physical Capital	139,475	1,569.7	10,242.1	0.46	17.3	93.2	501.1	477,185.0
Intangible Capital	139,475	571.1	3,445.4	0.51	13.7	50.5	210.8	265,123.8
Materials	139,475	1,790.5	11,769.4	0.29	20.7	118.3	618.3	526,207.6

Notes: This table reports summary statistics for the primary variables used in the production function estimation. All variables are deflated using industry-year price deflators from the BEA. Sales, physical capital, intangible capital, and materials are in millions of dollars. Employees are in thousands.

codes 6000–6999), and government entities (SIC codes 9000 and above). I further restrict the sample to firms with at least 10 employees and remove observations with missing or negative sales, capital, employment, Selling, General, and Administrative (SG&A), or Cost of Goods Sold (COGS). All variables are winsorized at 1 % to minimize the impact of extreme outliers.

Table 1 reports summary statistics. The sample comprises 139,475 firm-year observations, averaging approximately 2,800 firms per year. The median firm has $190 million in sales, 1,200 employees, and an intangible capital stock of $50.5 million — roughly half the size of its physical capital stock ($93.2 million). The large gap between medians and means across all variables reflects substantial right-skewness in the firm size distribution, with the largest firms orders of magnitude above the median. This skewness validates the nonparametric approach, which allows the production function to vary flexibly across the firm distribution rather than imposing a common parametric structure.

Physical capital (K) is measured as the gross book value of property, plant, and equipment (item PPEGT). This gross measure captures the total physical capacity available to the firm and avoids the idiosyncratic measurement error inherent in firm-specific accounting depreciation schedules. Labor (L) is defined as the total number of employees (item EMP), and output (Y) is measured as deflated net sales (item SALE).

I construct material inputs (M) following the residual approach of Keller and Yeaple (2009), with two structural modifications to accommodate the intangible capital framework. First, consistent with the use of gross physical capital, I do not deduct accounting depreciation from total operating expenses. This ensures that the material flow is measured gross of non-cash capital utilization. Second, I subtract both the total wage bill and the industry-specific fraction of Selling, General, and Administrative (SG&A) expenditures allocated to organizational capital accumulation. This adjustment is structurally necessary to prevent the double-counting of intangible investments: it ensures that resources used for the long-term accumulation of organizational capital are treated as a predetermined state variable

(K^{int}) rather than a flexible intermediate input (M).

All variables constructed are deflated using industry-year price deflators from the Bureau of Economic Analysis (BEA). Annual, industry-level wage data are sourced from the Quarterly Census of Employment and Wages (QCEW) provided by the Bureau of Labor Statistics (BLS).

Measurement of Intangible Capital. I construct the intangible capital at the firm level as outlined in Ewens et al. (2025) (along with insights from other studies such as Peters and Taylor (2017)). Intangible capital consists of two components: knowledge capital and organizational capital. I measure knowledge capital using reported Research and Development (R&D) expenses. Because R&D investments are recorded as flow variables in Compustat, I accumulate them into a stock variable using the perpetual inventory method:

$$R_{it} = (1 - \delta_{R\&D})R_{it-1} + R\&D_{it} \tag{16}$$

where R_{it} is the knowledge capital of firm i in year t, $\text{R\&D}_{it}$ is the firm's R&D investment, and $\delta_{R\&D}$ is the industry-specific R&D depreciation rate estimated by Ewens et al. (2025). The initial knowledge stock, R_{i0}, is normalized to zero.

Organizational capital is constructed using Selling, General, and Administrative Expenses (SG&A). Because SG&A encompasses a broad range of overhead and operating expenses, it requires allocating a specific fraction of total SG&A as investment in organizational capital Lev and Radhakrishnan (2003). Applying the perpetual inventory method :

$$O_{it} = (1 - \delta_{SG\&A})O_{it-1} + \gamma SG\&A_{it} \tag{17}$$

where O_{it} is the organizational capital stock of firm i in year t, and γ is the industry-specific proportion of SG&A allocated to organizational investment. I set $\delta_{SG\&A} = 0$ following Ewens et al. (2025) and use their industry-specific estimates of the capitalization rate γ.[7]. The initial organizational stock, O_{i0}, is set to zero.

Total intangible capital stock is therefore the sum of these two components:

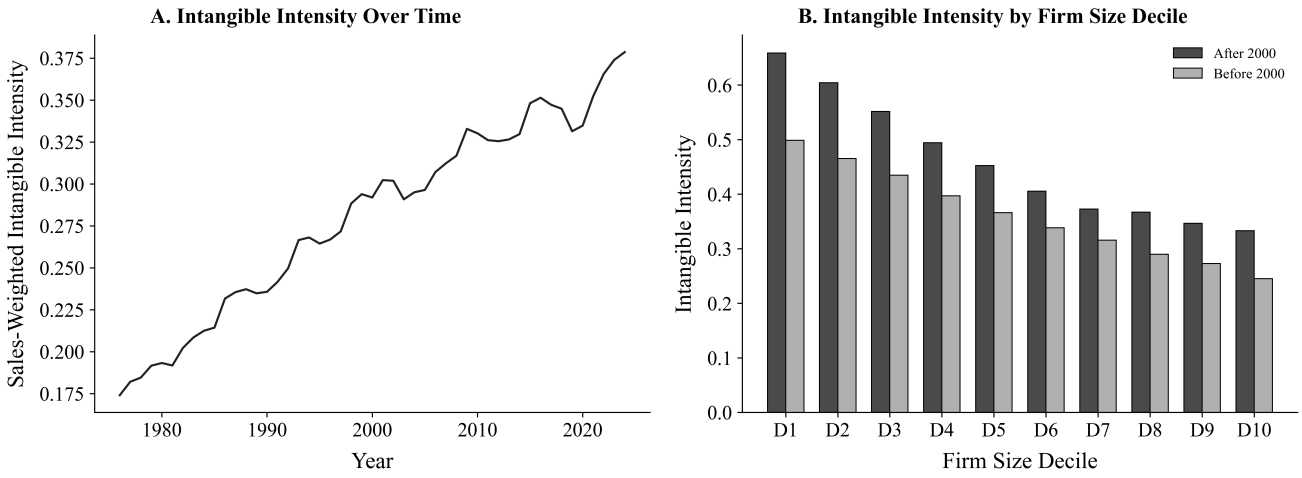
$$K_{it}^{int} = R_{it} + O_{it} \tag{18}$$

where K_{it}^{int} is the firm i's intangible capital stock in year t.

Panel A of Figure 1 plots the sales-weighted average of intangible intensity — defined as the ratio of intangible capital to total capital (intangible plus physical) — over the sample

[7]In the Robustness Check Section, I demonstrate that the core empirical results are robust to alternative parameterizations, including a uniform 30% capitalization rate.

Figure 1: Intangible Intensity Over Time and Across Firm Size



Notes: Panel A reports sales-weighted average intangible intensity over time. Panel B reports average intangible intensity by firm size decile before and after 2000. Intangible intensity is defined as $K_{it}^{int}/(K_{it}^{int} + K_{it})$. Firms are assigned to size deciles within each year based on sales.

period. Intangible intensity roughly doubles, rising from approximately 0.18 in the mid-1970s to 0.38 by 2024. This secular rise underscores the growing importance of intangible capital in production function estimation.

While Panel A documents the aggregate trend, the cross-sectional distribution of intangible intensity across firm size reveals important heterogeneity. Panel B reports average intangible intensity by firm size decile before and after 2000. Intangible intensity declines monotonically with firm size in both periods: smaller firms hold a larger share of intangible capital in their total capital stock. The post-2000 shift raises intangible intensity across all deciles, but the magnitude of this increase is highly asymmetric: the smallest firms experience a higher increase in intensity than the largest firms. This shift preserves the overall negative size gradient — a pattern whose implications for production are examined in Section 6.

5.2 Estimation Strategy

I estimate the production function separately across industries and over time. To construct the industry definitions, I broadly categorize firms at the 2-digit NAICS level, aggregating closely related sectors such as Retail Trade (44–45) and Transportation and Warehousing (48–49). However, to capture nuanced technological and production differences, I explicitly isolate key 3-digit subsectors from their parent categories, breaking out highly specialized industries such as Chemical Manufacturing (325), Computer and Electronic Product Manufacturing (334), and Non-store Retailers (454). Finally, to ensure sufficient statistical power

and maintain adequate sample sizes within each industry-year cell, I estimate the model using 8-year rolling windows.

As established in the identification section, the logarithmic production function is given by:

$$y_{it} = f_t(K_{it}, K_{it}^{int}, L_{it}\bar{h}_t(\tilde{M}_{it})) + \omega_{it}^H + \epsilon_{it}$$

Substituting the Hicks-neutral productivity term recovered via the control function approach (Equation 10) yields:

$$y_{it} = f_t(K_{it}, K_{it}^{int}, L_{it}\bar{h}_t(\tilde{M}_{it})) + c_{2t}\big(W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u_{it}^1, u_{it}^2\big) + \epsilon_{it} \tag{19}$$

Because the control variables resolve the endogeneity of the flexible inputs, the ex-post shock is strictly orthogonal to the firm's information set: $\mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, \sum_{j\neq i} R_{jt-1}] = 0$. While this condition permits estimation by minimizing the sum of squared residuals, exploiting the predetermined nature of capital provides additional identifying moments.

To form these moment conditions, I project the productivity process into a separable framework widely utilized in the proxy variable literature (Olley and Pakes (1996)). Specifically, I model Hicks-neutral productivity as:

$$\omega_{it}^H = c_{3t}(\omega_{it-1}^H, \omega_{it-1}^L, \sum_{j\in I, j\neq i} R_{jt-1}) + \upsilon_{it} \tag{20}$$

where $c_{3t}(\cdot)$ is an unknown function, and υ_{it} is separable from $c_{3t}(.)$ with $\mathbb{E}[\upsilon_{it}|I_{t-1}] = 0$. Unlike the innovation terms u_{it}^1 and u_{it}^2 in the previous section, which are non-separable and fully independent, υ_{it} is separable and mean-independent.

Using the lagged state variables, this projection can be rewritten as $\omega_{it}^H = \bar{c}_{3t}(W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}) + \upsilon_{it}$ for an unknown function $\bar{c}_{3t}(\cdot)$. Substituting this into the production function yields the final structural equation:

$$y_{it} = f_t(K_{it}, K_{it}^{int}, L_{it}\bar{h}_t(\tilde{M}_{it})) + \bar{c}_{3t}\big(W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}\big) + \upsilon_{it} + \epsilon_{it} \tag{21}$$

By construction, $\mathbb{E}[\upsilon_{it} + \epsilon_{it} \mid \mathcal{I}_{it-1}] = 0$. Because physical and intangible capital are predetermined, they must be orthogonal to this composite error term, providing the moment conditions $\mathbb{E}[\upsilon_{it} + \epsilon_{it} \mid K_{it}] = 0$, $\mathbb{E}[\upsilon_{it} + \epsilon_{it} \mid K_{it}^{int}] = 0$ and $\mathbb{E}[\upsilon_{it} + \epsilon_{it} \mid \sum_{j\in I, j\neq i} R_{jt-1}] = 0$.

Combining these conditions with the residual moment from Equation 19, the empirical

objective function to be minimized over the industry-year rolling window is:

$$\frac{1}{NT}\sum_{i,t}\epsilon_{it}^2 + \sum_k \left(\frac{1}{NT}\sum_{i,t}(\epsilon_{it}+\upsilon_{it})Z_{it}^k\right)^2 \tag{22}$$

where NT is the total number of observations in the estimation cell, and Z_{it}^k represents a vector of polynomial basis functions of physical capital, intangible capital, and external knowledge capital with k indexing its elements.

Minimizing this objective function requires estimating the control variables and approximating the functions $f_t(\cdot), \bar{h}_t(\cdot), c_{2t}(\cdot)$, and $\bar{c}_{3t}(\cdot)$. The estimation proceeds in three steps:

1. **Recovering Control Variables:** I estimate the control variables u_{it}^1 and u_{it}^2 (Equations 7 and 9) using logistic regression. Because the outcome variables are continuous in both cases, I partition their support into a grid of up to 500 points. I estimate the conditional CDFs at these points with second-order polynomials using logistic regression, and then linearly interpolate between them.

2. **Function Approximation:** I approximate the composite flexible input function $\bar{h}_t(\cdot)$ using second-order polynomials. Conditional on $\bar{h}_t(\cdot)$, I approximate the production function $f_t(\cdot)$ and the productivity functions $c_{2t}(\cdot)$ and $\bar{c}_{3t}(\cdot)$ using second-order polynomials.

3. **Optimization and Inference:** I minimize the empirical criterion function (Equation 22) to recover the structural parameters. Output elasticities, markups, and spillover effects are then calculated analytically according to the propositions derived in Section 3. I perform 100 bootstrap replications to estimate standard errors for the identified outcomes, treating firms as independent clusters and resampling them with replacement.

6 Results

This section presents the empirical findings from the structural estimation. I begin by examining returns to scale (RTS) and the elasticity decomposition across factor inputs. I then analyze how markups vary with firm size and intangible intensity, and compare aggregate markup dynamics over time with and without intangible capital in the model. Finally, I discuss productivity levels and dispersion by firm size and intangible assets, and document the knowledge spillover results across time and firms.

6.1 Returns to Scale and Factor Input Elasticities

To examine how intangible capital affects production, I estimate the production function under two specifications: one that includes intangible capital and knowledge spillovers, and one that excludes both. The specification with intangible capital is described in the Model section. The second re-estimates the production function after removing the intangible capital input and the spillover term.

For each specification, I construct firm-level returns to scale as the sum of estimated output elasticities. Because elasticities are recovered nonparametrically at the firm–year level, RTS vary flexibly across firms, industries, and time. I assign firms to size deciles within each year based on sales and average RTS within each decile–year cell to isolate systematic variation with scale.

Figure 2 presents the RTS results. When intangible capital is included as an input in the production function, returns to scale increase monotonically with firm size, from approximately 1.1 in the smallest decile to 1.5 in the largest. In contrast, when intangible capital is excluded, estimated RTS exhibit little variation across the size distribution. Increasing returns do not arise mechanically from physical scale; they emerge only when intangible capital enters the production technology.

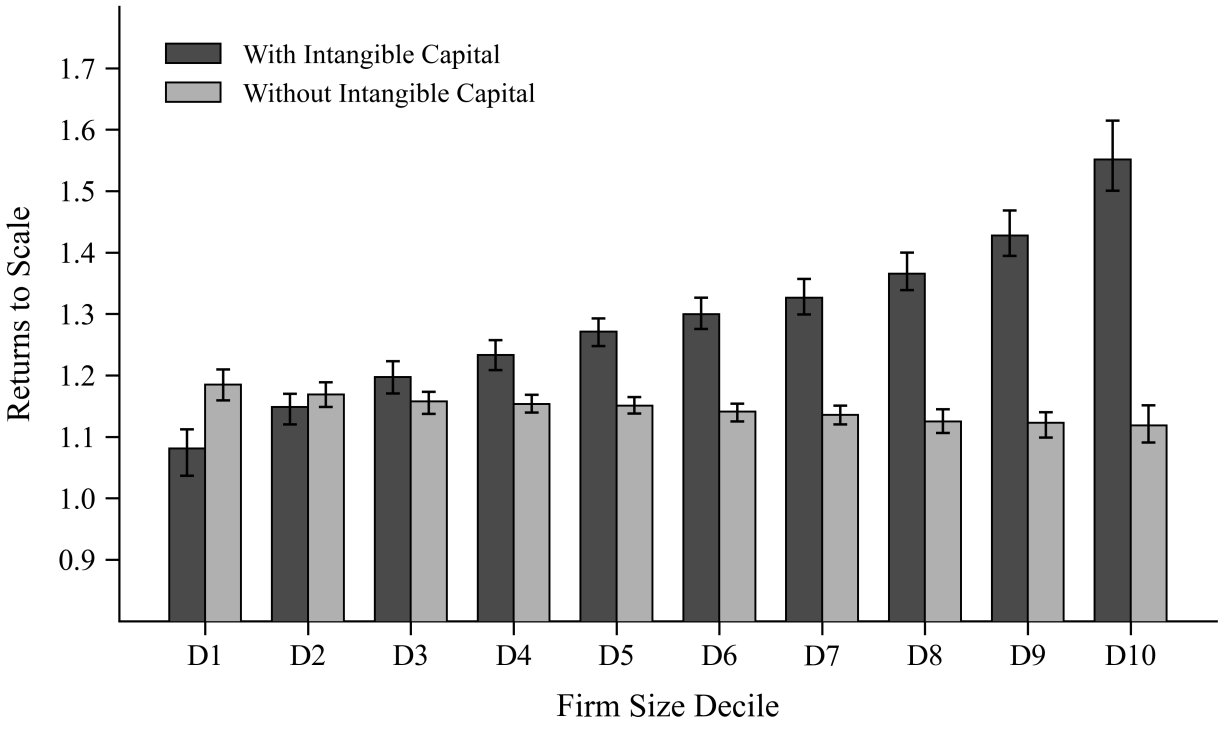
Figure 3 examines the time dimension by estimating returns to scale separately for the pre- and post-2000 periods. In both periods, RTS increase with firm size. However, the gradient steepens substantially after 2000. Small firms experience a decline in returns to scale, moving closer to constant returns, while large firms exhibit a pronounced rise. The post-2000 pattern is therefore not a uniform upward shift in scale economies but a divergence: the growing importance of intangible capital has amplified the technological advantages of firms able to deploy intangible assets at scale while offering limited scale benefits to smaller firms.

Figure 4 decomposes the size–RTS gradient by reporting output elasticities for each factor input across size deciles, separately before and after 2000.

Panel A (Intangible Capital). The elasticity of intangible capital increases with firm size in both periods. Before 2000, the elasticity is near zero for the smallest firms and rises gradually to approximately 0.3–0.4 in the top deciles. After 2000, the gradient steepens dramatically: the smallest decile exhibits a negative elasticity (approximately -0.3), while the largest decile reaches 0.6. The post-2000 divergence between small and large firms is striking — the range of intangible elasticities more than doubles relative to the earlier period.

The negative or near-zero elasticity among the smallest firms does not imply that intangible investment reduces production. Rather, conditional on other inputs and productivity,

Figure 2: Returns to Scale by Firm Size Decile



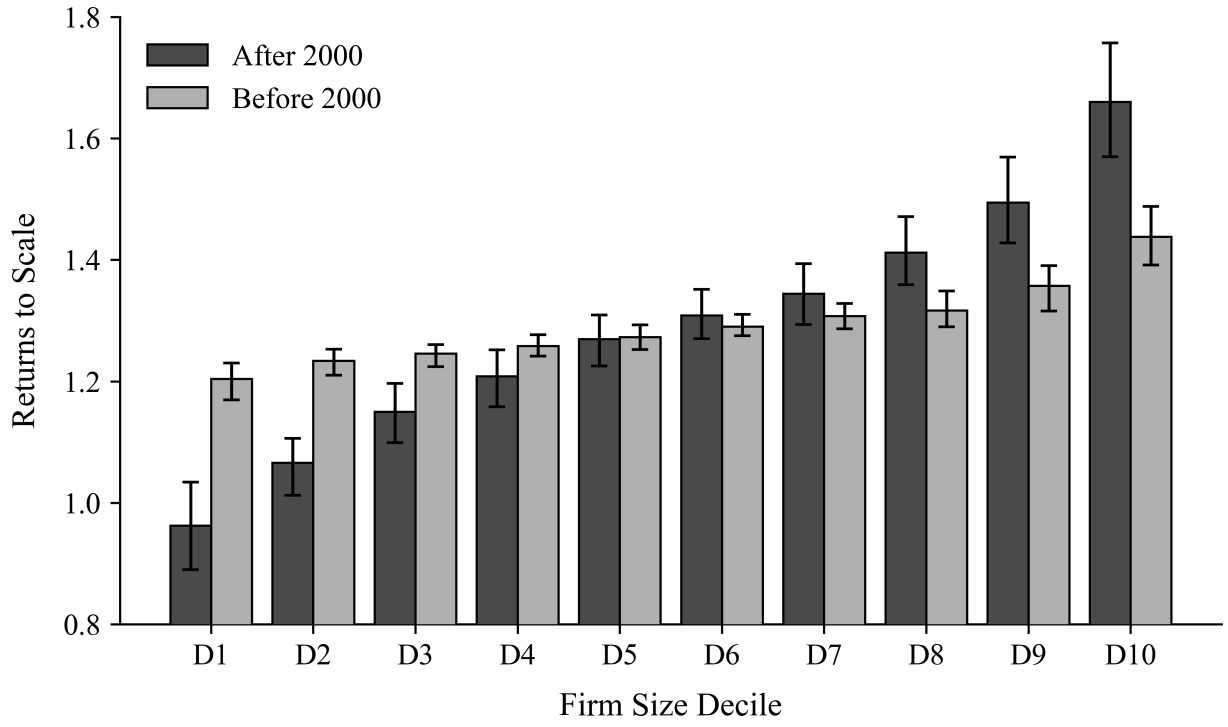
Note: This figure reports returns to scale by firm size decile. Firms are assigned to size deciles within each year based on sales. Returns to scale are averaged within each decile-year cell and then across years. Error bars denote 95% bootstrap confidence intervals (100 replications).

small firms are unable to translate intangible capital into proportional output gains.[8] This is especially striking given that these firms exhibit the highest intangible intensity in both periods (Figure 1, Panel B), consistent with high fixed costs (De Ridder, 2024) and scale thresholds (Haskel and Westlake, 2017) required to effectively exploit intangible assets. The decline after 2000 suggests that modern intangible assets entail substantially higher integration costs, leaving smaller firms in a deeper productivity J-curve (Brynjolfsson et al., 2021). At the other end of the distribution, the large and rising elasticity among top-decile firms confirms that scalable intangible capital is the primary driver of the increasing returns documented in Figures 2 and 3.

Panel B (Physical Capital). The elasticity of physical capital increases modestly with firm size in both periods, but the magnitude is small relative to that of intangible capital. While there is a slight post-2000 increase among large firms and a mild decline among small firms, these changes are quantitatively limited. Physical capital therefore does not drive the

[8]Because elasticities are estimated conditional on control variables that control for labor-augmenting and Hicks-neutral productivity, a negative elasticity reflects a low marginal contribution of intangible capital to contemporaneous output.

Figure 3: Returns to Scale by Firm Size Decile Before & After 2000



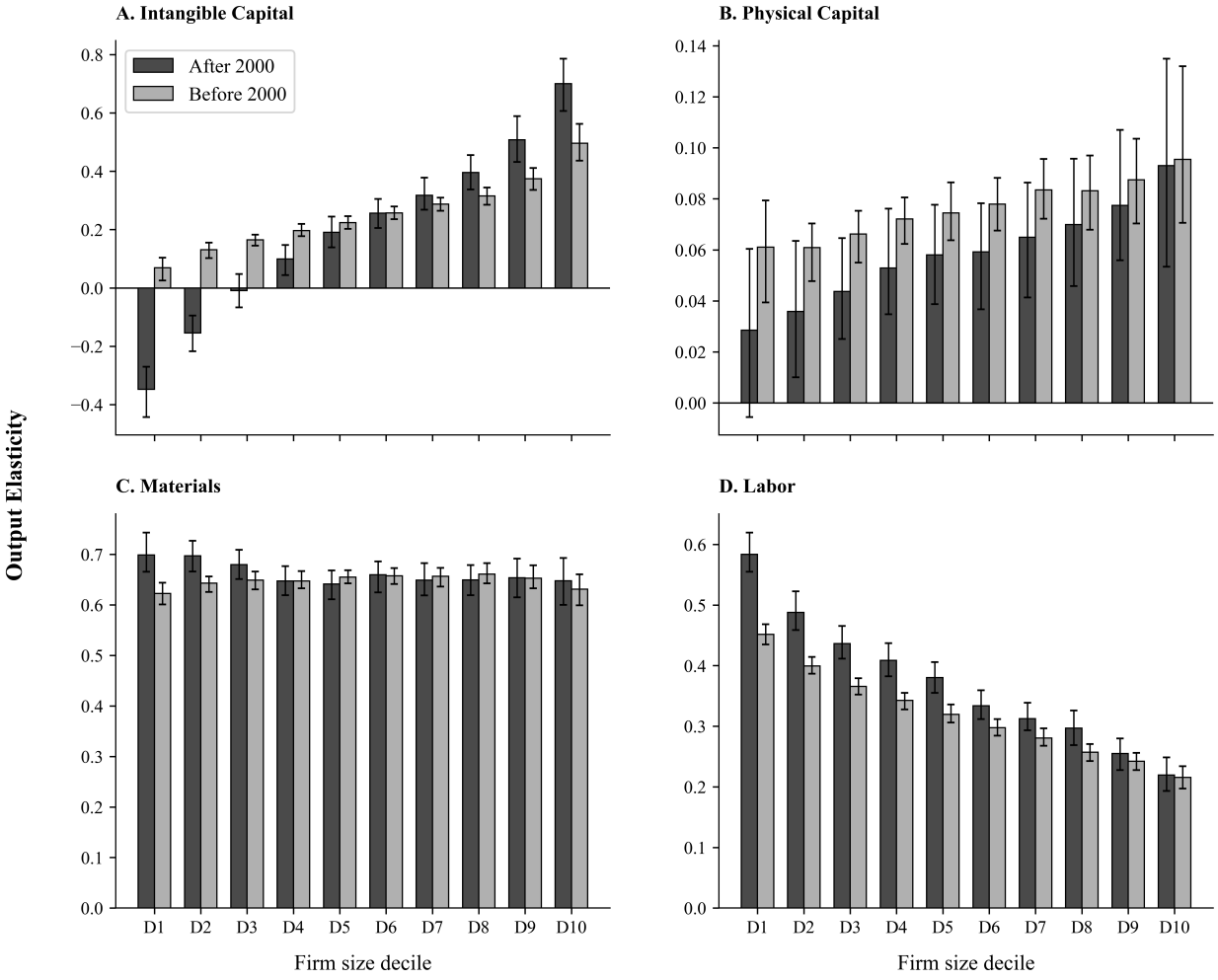
Note: This figure reports returns to scale by firm size decile. Firms are assigned to size deciles within each year based on sales. Returns to scale are averaged within each decile-year cell and then across years. Error bars denote 95% bootstrap confidence intervals (100 replications).

steepening of the RTS gradient.

Panel C (Materials). The materials elasticity is largely flat across the size distribution and stable over time. Because materials are strictly variable inputs tied directly to the quantity of output produced, this stability confirms that large firms are not generating increasing returns simply by extracting more final output from a given unit of raw materials.

Panel D (Labor). The labor elasticity declines systematically with firm size in both periods, falling from approximately 0.5 in the smallest decile to around 0.2 in the largest. The downward slope steepens after 2000, with labor elasticities declining across nearly the entire size distribution relative to the earlier period — consistent with the broader decline in the labor share documented by Autor et al. (2020). Mirroring the rising intangible elasticity in Panel A, this pattern reveals a reallocation of output elasticities at the top of the size distribution: as firms scale, production depends less on labor and more on intangible capital.

Figure 4: Factor Input Elasticities by Firm Size Decile



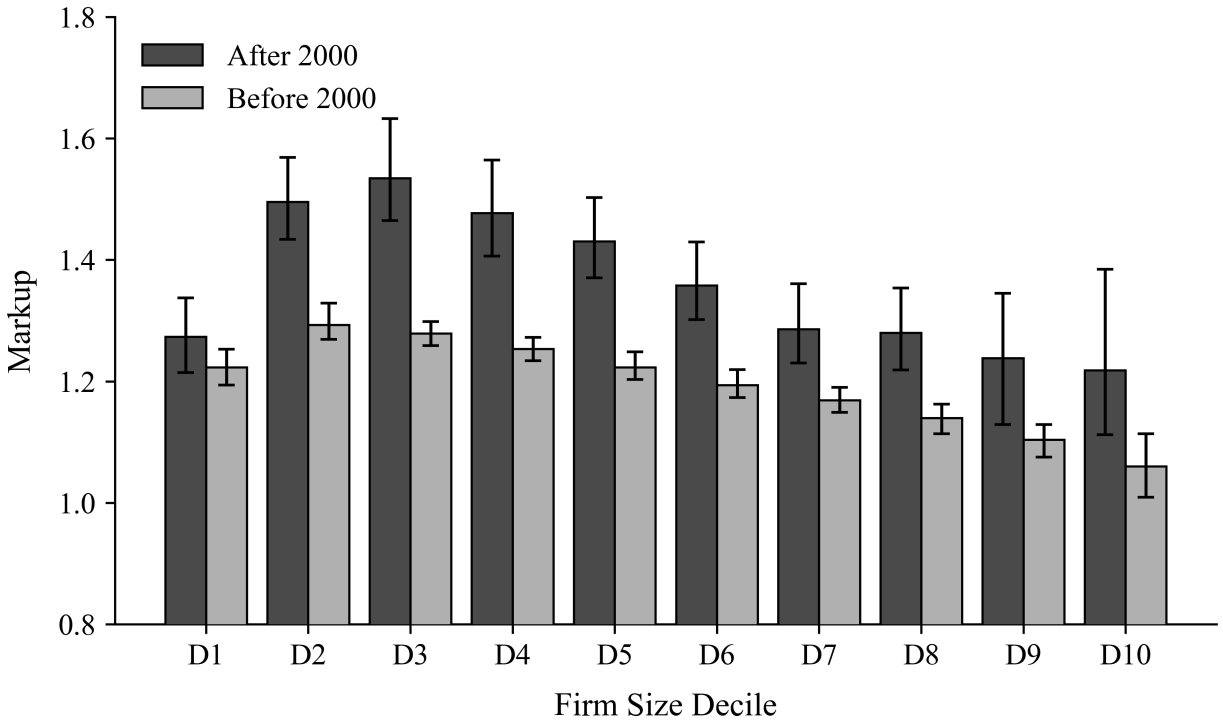
Note: This figure reports the output elasticities of intangible, physical capital, labor, and material by firm size decile. Firms are assigned to size deciles within each year based on sales. Elasticities are averaged within each decile-year cell and then across years. Error bars denote 95% bootstrap confidence intervals (100 replications).
Note: Y-axis scales differ across panels.

6.2 Markups

Having established that returns to scale vary systematically with firm size and over time, I now turn to markups. As described in the Identification section, firm-level markups are recovered from the first-order conditions of cost minimization following Loecker and Warzynski (2012) : the markup equals the ratio of a flexible input's output elasticity to its revenue share.

Figure 5 reports average markups by firm size decile before and after 2000. Markups exhibit an inverted-U relationship with firm size in both periods: they rise from the smallest

Figure 5: Markup by Firm Size Decile Before & After 2000



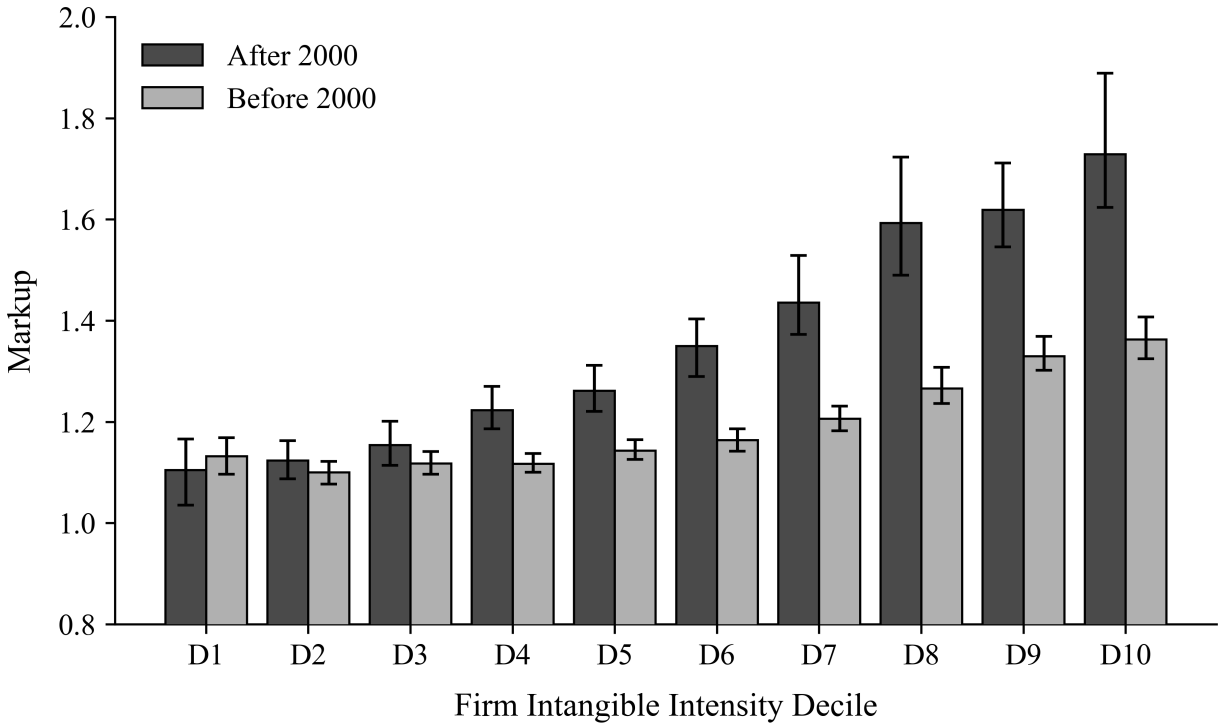
Note: Firms are assigned to size deciles within each year based on sales. Markups are averaged within each decile-year cell and then across years. Error bars denote 95% bootstrap confidence intervals (100 replications).

deciles to the middle of the distribution and then decline for the largest firms. Pricing power is therefore not monotonically increasing in size. Markups also shift upward after 2000 across nearly all size deciles, with the increase most pronounced among medium-sized firms, while the inverted-U shape remains intact. Despite exhibiting the strongest scale economies (Figure 3), the largest firms do not charge proportionally higher markups than mid-sized firms.

Figure 6 sharpens this distinction by reporting markups across deciles of intangible intensity, separately for the pre- and post-2000 periods. In contrast to the inverted-U pattern by firm size, markups increase monotonically with intangible intensity. This positive gradient is present in both periods and becomes noticeably steeper after 2000.

Taken together, Figures 5 and 6 reveal that intangible capital reshapes both production technology and competitive outcomes, but along different margins: it generates increasing returns among large firms while supporting higher markups among intangible-intensive firms. Scale and market power are therefore related but not equivalent consequences of intangible accumulation. The fact that the smallest firms fail to capture the highest markups despite

Figure 6: Markup by Firm Intangible Intensity Decile Before & After 2000



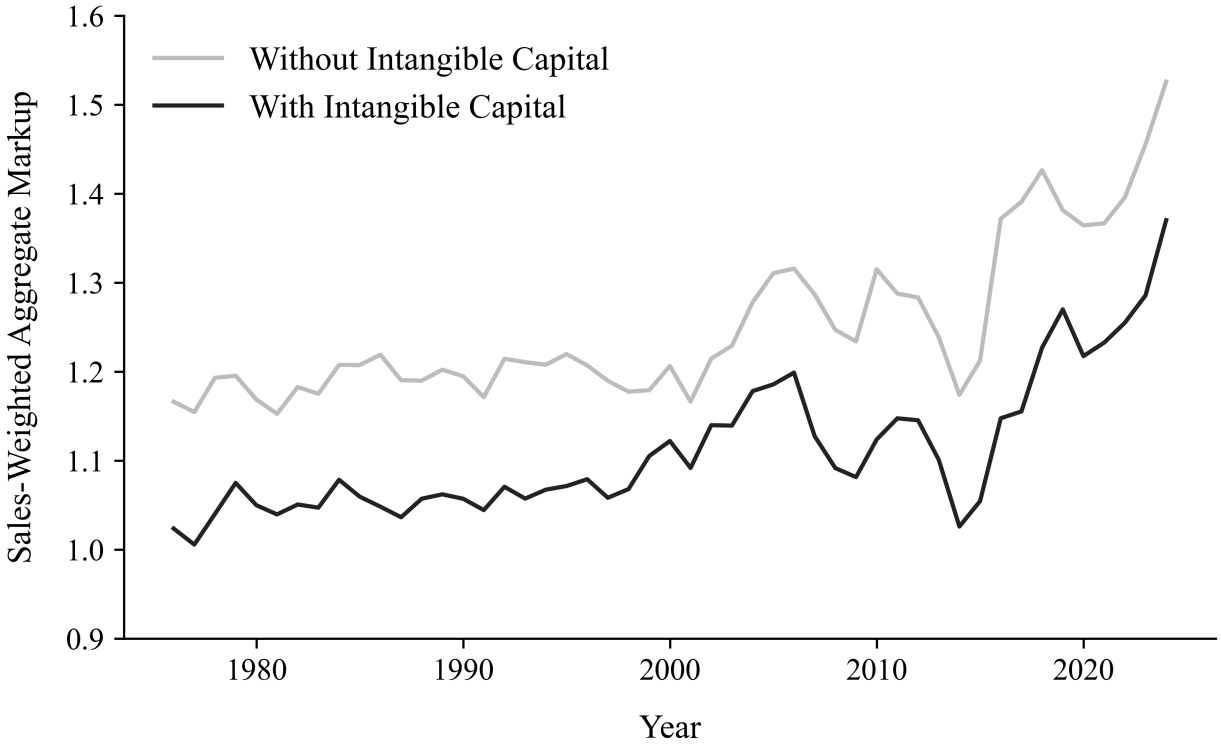
Note: This figure reports markup by firm intangible intensity decile. Firms are assigned to intangible intensity deciles within each year. Markups are averaged within each decile-year cell and then across years. Error bars denote 95% bootstrap confidence intervals (100 replications).

their high intangible intensity (Figure 1, Panel B) suggests that substantial scale thresholds are required to translate intangible capital into either returns to scale or pricing power.

Figure 7 examines the time-series evolution of sales-weighted average markups under two alternative production specifications: one that includes intangible capital as a factor input and one that excludes it. In both specifications, average markups trend upward over the sample period, particularly after 2000, consistent with the broader empirical literature (De Loecker et al., 2020).

Markups estimated from the model that excludes intangible capital are systematically higher throughout the sample, on average overstating the aggregate markup by 0.132 before 2000 and 0.145 after 2000. The gap reaches a peak of 0.235 in 2017 — an overstatement of more than 20 percent. The corrected markup estimates — approximately $1.05 - 1.10$ before 2000 — are more closely aligned with aggregate markup estimates from the macroeconomic literature (Basu and Fernald, 1997) than with the substantially higher levels reported by De Loecker et al. (2020). These higher unadjusted levels have been criticized as implausibly large (Basu, 2019); the corrected estimates largely address this concern.

Figure 7: Average Markup Over Time With and Without Intangible Capital



Note:This figure reports sales-weighted average markups over time, with and without intangible capital in the production function.

The gap widens after 2000, precisely when intangible capital intensifies in the economy. Standard production-function estimates therefore overstate market power by misattributing technological scale effects to price-cost margins (Bond et al., 2021; Koh et al., 2020). Table 2 reports the sales-weighted average output elasticities across both specifications. Omitting intangible capital inflates the estimated elasticities of all inputs: physical capital by 76%, materials by 12.5%, and labor by 12.1%. Because markups are structurally recovered as the ratio of a flexible input's elasticity to its revenue share, the 12.5% average bias in materials elasticity translates directly into an upward bias in markups of comparable magnitude, consistent with the average gap documented in Figure 7.

6.3 Productivity and Knowledge Diffusion

I next examine the diffusion margin embedded in the model. The spillover elasticity is identified as the derivative of the recovered Hicks-neutral productivity function with respect to the aggregate knowledge stock of other firms, measuring how sensitive a firm's productivity is to external R&D conditional on its own inputs. This approach connects directly to

Table 2: Sales-Weighted Average Output Elasticities: With and Without Intangible Capital

	With Intangible	Without Intangible	Difference	% Change
Intangible Capital	0.660	—	—	—
Physical Capital	0.094	0.165	0.071	76.1%
Materials	0.651	0.733	0.082	12.5%
Labor	0.199	0.223	0.024	12.1%
RTS	1.604	1.121	-0.484	-30.1%

Notes: This table reports sales-weighted average output elasticities across the two production function specifications. Difference = Without − With. % Change = (Without − With) / With × 100. RTS is the sum of all output elasticities.

the innovation-spillover literature (Bloom et al., 2013; Akcigit and Ates, 2023) but embeds spillovers inside a structural production framework.

Figure 8 shows that prior to 2000, spillover exposure is relatively flat across the size distribution, consistent with earlier evidence of widespread inter-firm knowledge diffusion (Bernstein and Nadiri, 1988; Bloom et al., 2013). After 2000, this pattern shifts sharply: spillover exposure declines monotonically with firm size, with small firms exhibiting substantially higher external spillover elasticities and large firms displaying little to no exposure. This shift is consistent with rising fixed costs of intangible capital (De Ridder, 2024). Because intangible investment increasingly requires large upfront fixed costs, large firms internalize knowledge creation through intensive R&D and organizational investment, while smaller firms rely more heavily on external sources.[9]

These findings complement the earlier results on scale economies and markups: the non-rivalry and scale thresholds of intangible capital give large firms a common advantage across all three dimensions — production, pricing, and knowledge diffusion. Their ability to deploy intangible assets at scale generates increasing returns, allows them to pursue volume rather than high markups, and enables them to internalize innovation rather than rely on external sources.

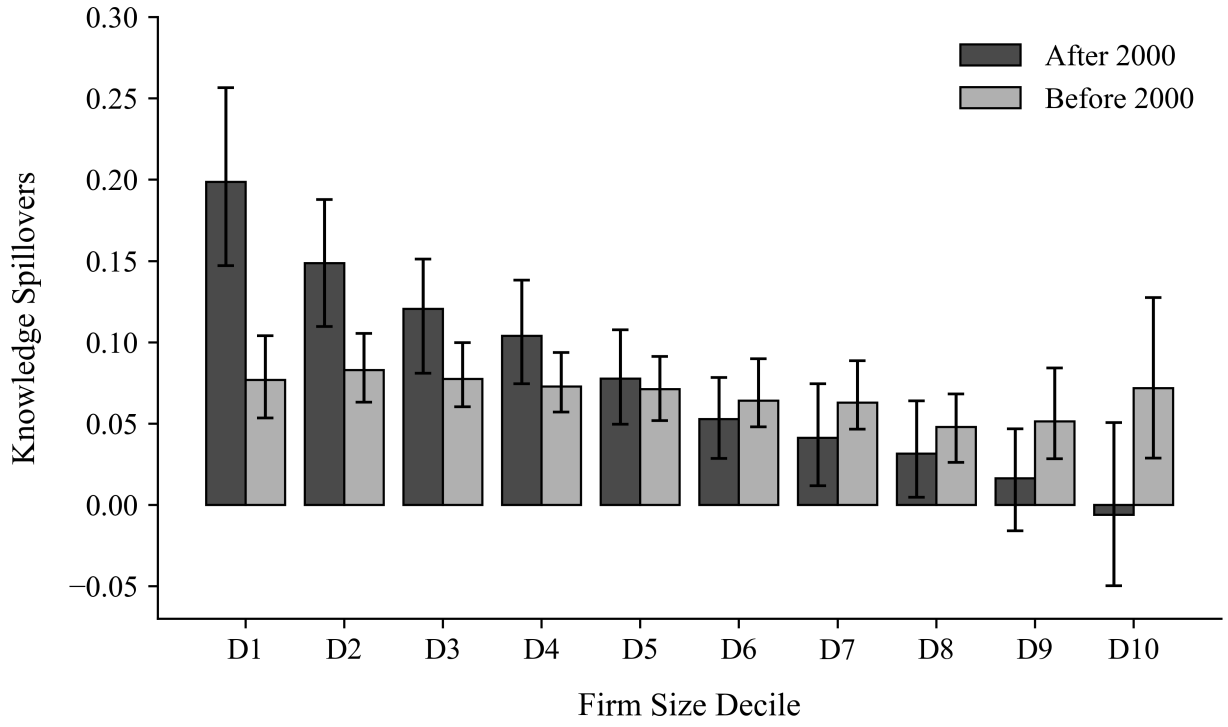
Figure 9 provides a complementary perspective by reporting industry–year demeaned productivity across joint deciles of firm size and intangible intensity.[10] Productivity refers to the Hicks-neutral productivity demeaned by industry–year. Deciles are constructed within industry–year cells using sales and intangible intensity, respectively.

First, relative productivity declines systematically with firm size. Moving from the smallest to the largest output deciles, demeaned productivity falls sharply — from 0.24 to −0.91

[9]Examples include large-scale software platforms, proprietary datasets, and complex organizational systems, which require substantial upfront investment and can only be amortized at scale.

[10]The untrimmed version yields nearly identical qualitative patterns and is reported in Appendix Figure A4.

Figure 8: Average Knowledge Spillover Before and After 2000 by Firm Size



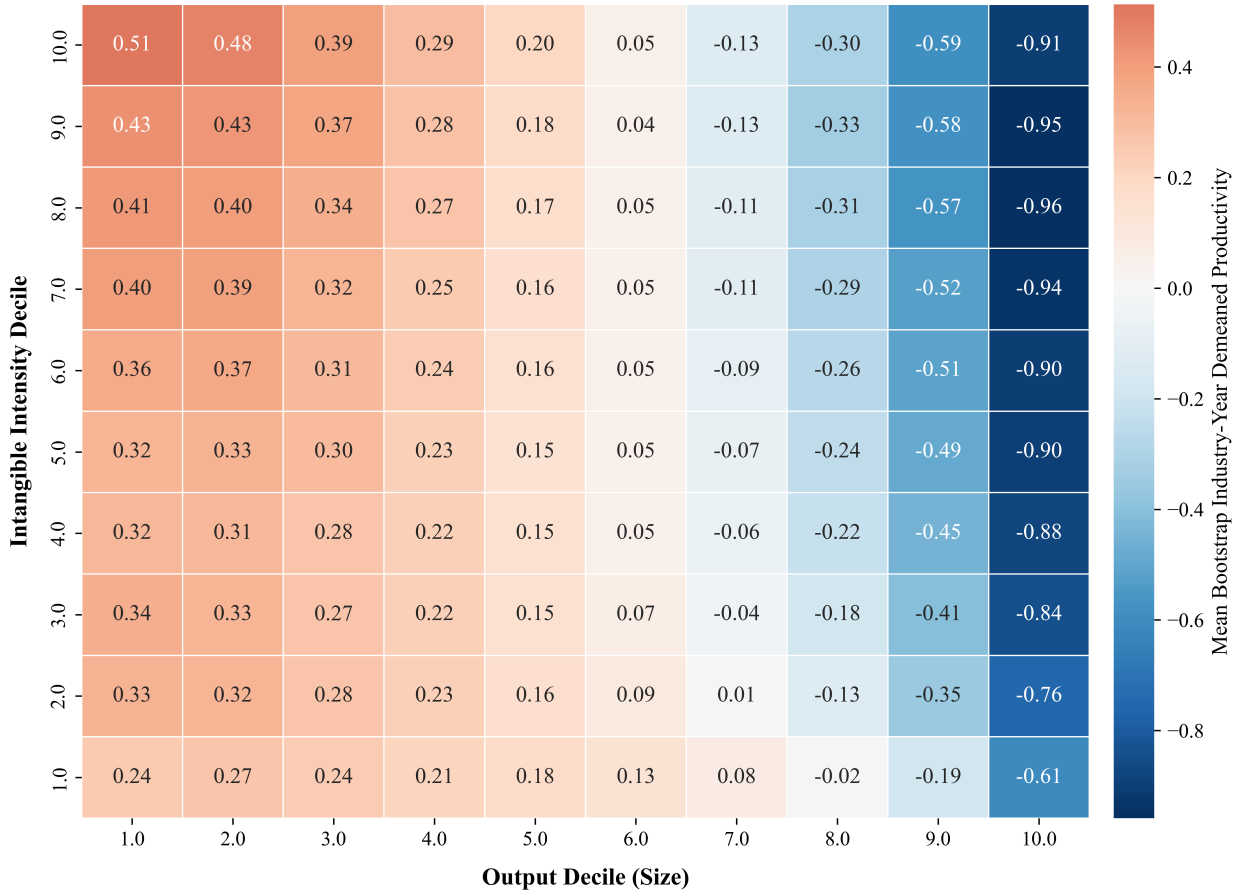
Notes: The figure reports average external R&D spillover exposure by firm size decile before and after 2000. Firms are assigned to size deciles within each year based on sales. Spillover elasticity is averaged within each decile-year cell and then across years. Error bars denote 95% bootstrap confidence intervals (100 replications).

at the extremes. Large firms are therefore not more productive in the Hicks-neutral sense. This finding qualifies selection-based interpretations of firm growth (Syverson, 2011; Hsieh and Klenow, 2009) and contrasts with models in which larger firms dominate purely because they are more efficient.

Second, the relationship between productivity and intangible intensity varies across the size distribution. Among small and medium-sized firms, higher intangible intensity is associated with higher relative productivity — consistent with evidence that intangible capital enhances firm performance (Doraszelski and Jaumandreu, 2018; Corrado et al., 2009). Among the largest firms, however, this gradient reverses: the most intangible-intensive large firms exhibit the *lowest* relative productivity. This pattern is consistent with the paper's central mechanism. Large intangible-intensive firms do not require high residual productivity to grow because their intangible capital generates increasing returns to scale — their output is explained by the scalability of their inputs, not by residual productivity.

Combining the productivity and returns to scale results reveals that intangible capital

Figure 9: Industry-Year Relative Productivity by Firm Size and Intangible Intensity



Notes: The figure reports industry-year demeaned productivity (2.5% trimmed) by firm size and intangible intensity deciles. Deciles are formed within each industry-year using log output and intangible intensity, respectively. Each cell reports the unweighted mean productivity across firms in that decile.

operates through fundamentally different channels across the firm size distribution. For small firms, intangibles enhance productivity but do not translate into increasing returns to scale. For large firms, the pattern reverses: intangibles generate strong scale economies while productivity declines. Intangible capital thus shapes firm performance across the size distribution — through productivity among small firms and through scalability among large ones.

6.4 External Validation

To formally address the potential for measurement bias, I validate the structural markup estimates against Tobin's Q, a market-based measure of firm value that is entirely external to the production function estimation. Because markups are recovered as the ratio of a

flexible input's output elasticity to its revenue share, including intangible capital as an additional factor input could mechanically reduce the estimated elasticities of the flexible inputs, thereby lowering measured markups by construction. If this mechanical adjustment were the sole driver of the corrected estimates, one would expect the model to exhibit a weaker or unchanged correlation with external market valuations — a lower markup that reflects nothing more than an accounting reallocation should not produce a stronger relationship with forward-looking investor assessments of firm value.

Table 3 reveals a pattern that is difficult to reconcile with a pure measurement artifact. Before 2000, when production was predominantly physical-capital-intensive and intangible stocks varied modestly across the size distribution, the two specifications yield closer results. The corrected model provides a marginal improvement of 3.74% in R^2. This near-equivalence serves as a falsification benchmark: in an environment where omitting intangibles generates little bias, the correction produces little change.

Table 3: External Validation: Markups and Market Valuation (Tobin's Q)

	Before 2000		**After 2000**	
	(1)	(2)	(3)	(4)
Variable: log(Tobin's Q)	Without Intangible	With Intangible	Without Intangible	With Intangible
log(Markup)	0.627***	0.659***	0.402***	0.417***
	(0.037)	(0.034)	(0.028)	(0.023)
Industry & Year FE	Yes	Yes	Yes	Yes
R^2	0.1561	0.1620	0.1101	0.1206
% Improvement in R^2	**+3.74%**		**+9.59%**	
Observations	69,947	69,947	60,218	60,218

Notes: Dependent variable is log(Tobin's Q). Without Intangible uses markups without intangibles; With Intangible uses the main structural model. Standard errors clustered at firm level.

A sharper divergence emerges after 2000. In the modern intangible-intensive regime, the specification incorporating intangible capital yields a 9.59% improvement in R^2 relative to the traditional baseline — nearly three times the pre-2000 gain. This is the opposite of what a mechanical artifact would produce: a purely accounting-driven reduction in elasticities should weaken, not strengthen, the connection between estimated markups and external valuations. The result instead suggests that traditional estimators conflate technological scale effects with pricing power, and that once intangible capital is properly accounted for, the corrected markups more accurately reflect the price-cost wedge that investors price into firm value.

7 Robustness Checks

7.1 Controlling for Product Demand

One concern about Compustat data is that variables are measured in revenue rather than quantities. Output in the regression is PQ where P is the product price and Q is quantity. Similarly, $K_{it}^{int} = P_t^{int} Q_{it}^{int}$, $K_{it} = P_t^k Q_{it}^k$, and $M_{it} = P_t^m Q_{it}^m$. BBecause input markets are perfectly competitive, input prices are common across firms within each industry-year and are absorbed by deflation. However, since the output market is imperfectly competitive, deflating sales does not eliminate cross-firm output price variation. After deflating, the estimation becomes

$$\frac{R_{it}}{P_t} = \frac{P_{it} Q_{it}}{P_t} = F_t(Q_{it}^k, Q_{it}^{int}, h_t(\omega_{it}^L L_{it}), Q_{it}^m) \exp(\omega_{it}^H) \exp(\epsilon_{it}) \tag{23}$$

where R_{it} is the revenue of firm i in year t.

I follow Klette and Griliches (1996) to control for cross-firm price variation. Under a CES demand aggregator, demand for each product is

$$Q_{it} = \left(\frac{P_{it}}{P_t}\right)^{-\sigma} Q_t \exp(\epsilon_{it}^d) \tag{24}$$

where σ is the elasticity of substitution, Q_t is aggregate sectoral demand, and ϵ_{it}^d is an idiosyncratic demand shock. Rearranging, the deflated output price is

$$\frac{P_{it}}{P_t} = \left(\frac{Q_{it}}{Q_t}\right)^{-1/\sigma} \exp((\epsilon_{it}^d)^{1/\sigma}) \tag{25}$$

Substituting into Equation 23 and taking logarithms yields

$$r_{it} = \left(1 - \frac{1}{\sigma}\right) f_t(Q_{it}^k, Q_{it}^{int}, h_t(\omega_{it}^L L_{it}), Q_{it}^m) + \left(1 - \frac{1}{\sigma}\right) \omega_{it}^H + \frac{1}{\sigma} q_t + \frac{1}{\sigma} \epsilon_{it}^d + \left(1 - \frac{1}{\sigma}\right) \epsilon_{it} \tag{26}$$

where r_{it} is log deflated revenue and q_t is log aggregate sectoral output. Revenue-based estimation biases the coefficients because the production function is scaled by $\left(1 - \frac{1}{\sigma}\right)$ and total industry demand enters as an additional regressor.

To isolate the interaction between demand and production coefficients, I re-estimate the structural model including q_t as an additional regressor to identify σ, then recover quantity-based elasticities by scaling the revenue-based estimates by $(1 - 1/\sigma)^{-1}$. As shown in Table 4, the results remain remarkably stable. The RTS gradient and the markup divergence persist after purging demand-side contamination, suggesting that the scale economies documented

in this paper are a supply-side phenomenon that survives after controlling for demand-side variation. Detailed estimation results are reported in the Appendix C.

Table 4: Robustness of Scale Economies, Knowledge Spillovers, and Productivity Dispersion

	Returns to Scale		Knowledge Spillovers		Productivity Dispersion	
Specification	Δ_{50}	Δ_{10}	Δ_{50}	Δ_{10}	Δ_{50}	Δ_{10}
Baseline Model	0.205	0.455	−0.070	−0.127	−0.697	−1.594
Demand Control (CES)	0.188	0.384	−0.017	−0.025	−0.572	−1.147
Organizational Capital ($\gamma = 0.3$)	0.241	0.527	−0.043	−0.066	−0.659	−1.498
Measurement Error Simulation	0.038	0.086	−0.065	−0.116	−0.434	−0.898
External Intangible Capital	0.199	0.450	−0.046	−0.085	−0.639	−1.475
Third-Order Polynomial	0.206	0.444	−0.066	−0.120	−0.670	−1.516
TNIC Spillover (normalized)	0.150	0.361	−0.006	−0.012	−0.858	−1.987

Notes: Each row reports a separate robustness specification. Δ_{50} = top quintile minus bottom quintile (deciles 6–10 vs. 1–5); Δ_{10} = decile 10 minus decile 1. Returns to scale and knowledge spillovers are sales-share-weighted averages including intangible capital and inter-firm R&D spillovers. Productivity dispersion is the demeaned log productivity gap across quantiles.

7.2 Measurement Error in Intangible Capital

A fundamental empirical challenge in this framework is the potential mismeasurement of intangible capital. To ensure that the core findings are not driven by specific capitalization assumptions or idiosyncratic measurement noise, I conduct three distinct robustness exercises. First, I test the sensitivity of the model to the organizational capital parameters. Rather than applying the industry-varying capitalization rates estimated by Ewens et al. (2025), I impose a uniform capitalization rate of 30% ($\gamma = 0.3$) for SG&A expenditures across all sectors. This fixed 30% parameterization serves as a strict and widely utilized benchmark in the macroeconomic and finance literature to measure organizational capital (Eisfeldt and Papanikolaou, 2014) .

Second, I explicitly simulate the impact of measurement error on the baseline intangible capital stock via a Monte Carlo approach. For each observation, I add a normally distributed noise term with a mean of zero and a standard deviation calibrated to 10% of the cross-sectional standard deviation of intangible capital within that specific industry-year cell. I generate 100 independent, perturbed datasets, re-estimate the structural production function for each sample, and compute the average of the resulting estimates.

Third, I account for the inorganic accumulation of intangible assets through mergers and acquisitions (M&A). When firms acquire targets, they inherently absorb the target's

existing intangible capital stock. To prevent these discrete, externally acquired assets from confounding the firm's endogenous input choices, I incorporate acquired intangible capital as a structural shifter within the Markov productivity process. This explicitly controls for M&A-driven jumps in the state space, which are particularly prevalent among large incumbent firms.

As shown in Table 4, all three exercises yield results consistent with the baseline model. The core patterns — increasing returns to scale with firm size, declining spillover exposure among large firms, and productivity dispersion across the size distribution — are preserved across all specifications. The Monte Carlo simulation produces slightly attenuated parameters, a predictable consequence of classical measurement error, but does not alter the qualitative findings. Detailed results are reported in Appendix C.

7.3 Functional Form of the Approximations

To verify that the structural estimates are not driven by the choice of polynomial degree, I re-estimate the production function using third-order polynomial approximation. The baseline uses second-order polynomials following Demirer (2025), but higher-order terms could capture additional curvature in the production function and productivity processes at the expense of overfitting.

Table 4 highlights that the results are quantitatively and qualitatively consistent with the baseline. The estimated output elasticities, markups, and spillover effects are all preserved under the more flexible specification. Detailed estimates are reported in Appendix C.

7.4 TNIC Product-Market Peer Spillovers

A potential concern with the knowledge spillover measure used in the main specification is that aggregating R&D capital stocks at the industry level may capture broad sectoral trends rather than firm-level knowledge diffusion. Although this specification is theoretically motivated by the non-rivalry of knowledge following Romer (1990) — wherein the total stock of sectoral knowledge is available to all firms regardless of their specific technological proximity — one might argue that the industry aggregate conflates knowledge spillovers with common shocks that drive both R&D investment and productivity across all firms in a sector simultaneously.

To address this concern, I construct an alternative spillover measure using the Text-based Network Industry Classification (TNIC) database of Hoberg and Phillips (2016). Rather than aggregating R&D stocks within broad industry cells, this measure weights each peer firm's R&D capital stock by its pairwise product-market similarity score, which is derived from the

textual overlap in firms' 10-K product descriptions. Specifically, for each focal firm i in year t, the TNIC-weighted spillover is constructed as

$$S_{it}^{TNIC} = \sum_{j \neq i} \tilde{\omega}_{ij} R_{jt-1} \tag{27}$$

where $\tilde{\omega}_{ij} = \omega_{ij} / \sum_{k \neq i} \omega_{ik}$ are similarity scores normalized to sum to one within each focal firm-year, and ω_{ij} is the raw TNIC-3 pairwise similarity score between firms i and j. I apply the standard similarity threshold of 0.01 following Hoberg and Phillips (2016), retaining only firm pairs with meaningful product market overlap. Normalizing scores to sum to one ensures that the measure captures the average intensity of spillover exposure from product-market peers rather than the total quantity, controlling for mechanical differences in peer network size across firms of different sizes.

This measure differs from the main specification along two important dimensions. First, it defines knowledge communities by product-market proximity rather than industry classification boundaries, providing a more granular and data-driven delineation of the relevant peer group. Second, because TNIC similarity is based on 10-K product descriptions, it covers all public firms regardless of patenting activity — unlike patent-based proximity measures (Bloom et al., 2013), which are limited to firms with active patent portfolios.

The TNIC data is available from 1989 onward, yielding a robustness sample of 72,088 firm-year observations covering 8,435 firms through 2023 with a mean of 26.4 product-market peers per firm. An important feature of this sample is that TNIC coverage is somewhat lower among smaller firms, ranging from approximately 55% in the smallest size decile to 83% in the largest. This size gradient in coverage makes the robustness check conservative for my purposes: since small firms exhibit the highest spillover exposure in the main specification, their underrepresentation in the TNIC sample mechanically compresses the estimated size gradient in spillover exposure. Finding a qualitatively consistent declining gradient despite this compositional bias therefore provides stronger evidence for the robustness of the result.

Table 4 reports the results. The RTS gradient, productivity dispersion, and knowledge spillover patterns are all preserved under the TNIC specification. All gradients are attenuated in magnitude relative to the baseline, consistent with the underrepresentation of small firms in the TNIC sample. Detailed results are reported in Appendix E.

8 Conclusion

The modern economy's shift toward intangible-intensive production has profound implications for how we measure and interpret firm behavior. This paper develops a flexible non-

parametric production framework that incorporates intangible capital as a distinct factor input and external R&D as a structural productivity shifter, and uses it to revisit three central questions: what drives increasing returns among large firms, what determines markups, and what shapes knowledge diffusion across firms.

When intangible capital is incorporated into the production function, returns to scale rise sharply with firm size — a pattern that disappears when intangibles are excluded. Yet the same large firms do not charge the highest markups; pricing power follows intangible intensity rather than size, and the largest firms appear to pursue a volume-margin trade-off consistent with strategic market-share expansion. Meanwhile, Hicks-neutral productivity declines with firm size, and among the largest firms, declines further with intangible intensity. Large firms grow not because they are more efficient, but because their intangible assets — non-rival, scalable, and increasingly internalized — allow output to expand disproportionately relative to inputs.

This has direct implications for markup measurement. Omitting intangibles biases markups upward — increasingly so after 2000, when intangible elasticities rise most sharply among large firms. Corrected estimates not only lower the aggregate level but also better predict cross-sectional variation in market valuations. Part of what the literature has documented as rising market power is more accurately characterized as a transformation in production technology.

The knowledge diffusion findings reinforce this interpretation. Before 2000, spillover exposure was broadly shared across the firm size distribution. After 2000, it declines sharply with size, as large firms increasingly rely on internally generated knowledge rather than external R&D flows. The concentration of economic activity among these firms weakens the aggregate spillover measures — not because knowledge has become less valuable, but because the firms that dominate the economy have reduced their dependence on others.

Together, these results suggest that the standard policy framework for evaluating concentration and market power requires recalibration. If rising concentration reflects scalable intangible assets rather than anti-competitive pricing, then markup-based measures will systematically overstate the welfare costs of concentration, and remedies targeting price-cost margins may be poorly targeted.

Appendix

A Proofs

A.1 Proof of Equation 3

The firms minimize their flexible inputs as physical and intangible capital are predetermined variables. They optimize their level to produce at least planned output, $\bar{Y}_{it}$. The cost minimization problem, then, becomes:

$$\min_{L_{it},M_{it}} \left\{ p_t^L L_{it} + p_t^M M_{it} : \mathbb{E}[F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)exp(\epsilon_{it})|\mathcal{I}_{it}] \geq \bar{Y}_{it} \right\} \tag{28}$$

By the timing assumption, as capitals and ω_{it} and ω_{it}^L are known, the cost problem turns out

$$\min_{L_{it},M_{it}} \left\{ p_t^L L_{it} + p_t^M M_{it} : F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)\mathbb{E}[exp(\epsilon_{it})\mathcal{I}_{it}] \geq \bar{Y}_{it} \right\}$$

This problem can be written as the choice of effective labor rather than labor. $\bar{L}_{it} = L_{it}\omega_{it}^L$ is defined as the effective labor. Defining quality adjusted wages $p_{it}^L = \frac{p_t^L}{\omega_{it}^L}$ I can now rewrite the problem in terms of effective labor, $\bar{L}_{it}$

$$\min_{\bar{L}_{it},M_{it}} \left\{ \bar{p}_{it}^L \bar{L}_{it} + p_t^M M_{it} : F_t(K_{it}, K_{it}^{int}, h_t(\bar{L}_{it}, M_{it}))exp(\omega_{it}^H)\mathbb{E}[exp(\epsilon_{it})] \geq \bar{Y}_{it} \right\} \tag{29}$$

These two problems are equivalent as I just redefine the optimization problem with effective labor and labor productivity ω_{it}^L is known by the firms at time t. For easy notation, defining $\tilde{Y}_{it} = \frac{\bar{Y}_{it}}{\mathbb{E}[exp(\epsilon_{it}\mathcal{I}_{it}]exp(\omega_{it}^H)}$ the cost problem becomes

$$C_t(K_{it}, K_{it}^{int}, \bar{Y}_{it}, p_{it}^L, p_t^M, \omega_{it}^H) = \min_{\bar{L}_{it},M_{it}} \left\{ \bar{p}_{it}^L \bar{L}_{it} + p_t^M M_{it} : F_t(K_{it}, K_{it}^{int}, h_t(\bar{L}_{it}, M_{it})) \geq \tilde{Y}_{it} \right\}$$

As $F_t()$ is a homogeneous function in $h_t()$, I can invert $F_t()$ function in $h_t()$. As the objective function is linear I can reconstruct the problem as

$$C_t(K_{it}, K_{it}^{int}, \bar{Y}_{it}, p_{it}^L, p_t^M, \omega_{it}^H) = \min_{\bar{L}_{it},M_{it}} \left\{ \bar{p}_{it}^L \bar{L}_{it} + p_t^M M_{it} : h_t(\bar{L}_{it}, M_{it}) \geq F_t^{-1}(K_{it}, K_{it}^{int}, \tilde{Y}_{it}) \right\}$$

As $K_{it}, K_{it}^{int}, \omega_{it}^H$ and $\bar{Y}_{it}$ is given at time t, the problem can be rewritten as

$$C_t(K_{it}, K_{it}^{int}, \bar{Y}_{it}, p_{it}^L, p_t^M, \omega_{it}^H) = F_t^{-1}(K_{it}, K_{it}^{int}, \tilde{Y}_{it}) \min_{\bar{L}_{it},M_{it}} \left\{ \bar{p}_{it}^L \bar{L}_{it} + p_t^M M_{it} : h_t(\bar{L}_{it}, M_{it}) \geq 1 \right\}$$

Therefore, the cost minimization problem pins down to

$$C_t(K_{it}, K_{it}^{int}, \bar{Y}_{it}, p_{it}^L, p_t^M, \omega_{it}^H) = C_{1t}(K_{it}, K_{it}^{int}, \bar{Y}_{it}, \omega_{it}^H) C_{2t}(p_{it}^L, p_t^M)$$

where I redefine $F_t^{-1}()$, $C_{1t}(K_{it}, K_{it}^{int}, \bar{Y}_{it}, \omega_{it}^H) = F_t^{-1}(K_{it}, K_{it}^{int}, \tilde{Y}_{it}) \min_{\bar{L}_{it}, M_{it}}$. C_{2t} is the results of minimization problem. Since p_{it}^L and p_t^M are the only state variables left under the cost minimization problem, the minimization problem will only be a function of them. In other words, minimization part will be $C_{2t}(p_{it}^L, p_t^M)$.

Using Shephard's Lemma, the firms' optimal material and labor demand will be a derivative of the minimized cost function with respect to the corresponding prices, i.e.

$$M_{it} = \frac{\partial C_t(K_{it}, K_{it}^{int}, \bar{Y}_{it}, p_{it}^L, p_t^M, \omega_{it}^H)}{\partial p_t^M} = C_{1t}(K_{it}, K_{it}^{int}, \bar{Y}_{it}, \omega_{it}^H) \frac{\partial C_{2t}(p_{it}^L, p_t^M)}{\partial p_t^M}$$
$$\bar{L}_{it} = \frac{\partial C_t(K_{it}, K_{it}^{int}, \bar{Y}_{it}, p_{it}^L, p_t^M, \omega_{it}^H)}{\partial p_{it}^L} = C_{1t}(K_{it}, K_{it}^{int}, \bar{Y}_{it}, \omega_{it}^H) \frac{\partial C_{2t}(p_{it}^L, p_t^M)}{\partial p_{it}^L}$$

As the first part of the cost minimization problem, $C_{1t}(K_{it}, K_{it}^{int}, \bar{Y}_{it}, \omega_{it}^H)$ does not depend on flexible input prices, I factor out when taking derivatives. I will only need to take the derivative of $C_{2t}(p_{it}^L, p_t^M)$ to find the flexible input demand. Taking the ratio between material and labor demand, the equations above becomes

$$\frac{M_{it}}{\bar{L}_{it}} = \frac{\partial C_{2t}(p_{it}^L, p_t^M)/\partial p_t^M}{\partial C_{2t}(p_{it}^L, p_t^M)/\partial p_{it}^L}$$

The material-to-effective labor ratio only depends on the factor input prices. Replacing the effective labor $\bar{L}_{it} = L_{it}\omega_{it}^L$ into the equation, material-to-labor ratio becomes

$$\frac{M_{it}}{L_{it}} = \frac{\partial C_{2t}(p_{it}^L, p_t^M)/\partial p_t^M}{\partial C_{2t}(p_{it}^L, p_t^M)/\partial p_{it}^L} \omega_{it}^L$$

Note that this function only depends on the input prices and labor productivity, ω_{it}^L. Thus, the flexible input ratio becomes

$$\frac{M_{it}}{L_{it}} = r_t(p_t^L, p_t^M, \omega_{it}^L) = r_t(\omega_{it}^L)$$

where $r_t()$ is an unknown function. As $r_t()$ function is time-varying, and input prices are constant across firms due to the competitive input market, these prices will be a constant in this function. Therefore, the material-to-labor ratio becomes a function of labor augmenting productivity ω_{it}^L.

A.2 Proof of Equation 4

I will now show that the material-to-labor ratio is monotonous in ω_{it}^L in order to write labor productivity as a function of the material-to-labor ratio.

Since the cost function is a homogeneous of degree with one, the derivative of the cost function with respect to the input prices would be homogeneous of degree zero. Using this property, dividing by p_t^M will not change the partial derivative of the cost function with respect to flexible inputs. Therefore, I can rewrite it as follows

$$\frac{M_{it}}{L_{it}} = \frac{\partial C_{2t}(\frac{p_{it}^L}{p_t^M}, \frac{p_t^M}{p_t^M})/\partial p_t^M}{\partial C_{2t}(\frac{p_{it}^L}{p_t^M}, \frac{p_t^M}{p_t^M})/\partial p_{it}^L}\omega_{it}^L = \frac{\partial \tilde{C}_{2t}(\frac{p_{it}^L}{p_t^M})/\partial p_t^M}{\partial \tilde{C}_{2t}(\frac{p_{it}^L}{p_t^M})/\partial p_{it}^L}\omega_{it}^L$$

where $\tilde{C}_{2t}(\frac{p_{it}^L}{p_t^M}) = C_{2t}(\frac{p_{it}^L}{p_t^M}, 1)$. For the sake of easy notation, let $\tilde{C}_{2t}(\frac{p_{it}^L}{p_t^M})/\partial p_t^M = \tilde{C}_{2m}$, $\tilde{C}_{2t}(\frac{p_{it}^L}{p_{it}^L})/\partial p_t^M = \tilde{C}_{2l}$, and $\tilde{M}_{it} = \frac{M_{it}}{L_{it}}$. Taking the logarithm of the equation I can rewrite it as

$$log(\tilde{M}_{it}) = log\left(\frac{\tilde{C}_{2m}}{\tilde{C}_{2l}}\right) + log(\omega_{it}^L)$$

Taking the derivative of the equation with respect to $log(\omega_{it}^L)$

$$\frac{\partial log(\tilde{M}_{it})}{\partial log(\omega_{it}^L)} = \frac{\partial \tilde{C}_{2m}/\tilde{C}_{2l}}{\partial log(p_{it}^L/p_t^M)}\frac{log(\partial p_{it}^L/p_t^M)}{\partial log(\omega_{it}^L)} + 1$$

Since $\frac{log(\partial p_{it}^L/p_t^M)}{\partial log(\omega_{it}^L)} = 1$ and $\frac{\partial \tilde{C}_{2m}/\tilde{C}_{2l}}{\partial log(p_{it}^L/p_t^M)} = -\sigma(\omega_{it}^L L_{it}, M_{it})$ is the elasticity of substitution between effective labor and material, the equation becomes

$$\frac{\partial log(\tilde{M}_{it})}{\partial log(\omega_{it}^L)} = -\sigma(\omega_{it}^L L_{it}, M_{it}) + 1$$

From the elasticity substitution assumption $\sigma(\omega_{it}^L L_{it}, M_{it}) > 1$ or $\sigma(\omega_{it}^L L_{it}, M_{it}) < 1$, the derivative of $\tilde{M}$ with respect to ω_{it}^L is either positive or negative. IT will never become zero. This implies that $\tilde{M}_{it}$ is strictly increasing or decreasing function of ω_{it}^L. Thus, $\tilde{M}_{it}$ is monotone in ω_{it}^L. Then, we can invert the $r_t(\omega_{it}^L)$ function in ω_{it}^L, then labor productivity, ω_{it}^L becomes

$$\omega_{it}^L = r_t^{-1}\left(\frac{M_{it}}{L_{it}}\right) = r_t^{-1}(\tilde{M}_{it})$$

Thus, this shows that labor productivity is a function of the material-to-labor ratio.

A.3 Proof of Equation 11

The cost minimization problem of the firm i at time t is

$$\min_{L_{it},M_{it}} \left\{ p_t^L L_{it} + p_t^M M_{it} : F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)\mathbb{E}[exp(\epsilon_{it})\mathcal{I}_{it}] \geq \bar{Y}_{it} \right\}$$

The first order condition for material and labor is

$$p_t^M = \lambda_{it} \frac{\partial F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)\mathbb{E}[exp(\epsilon_{it})\mathcal{I}_{it}]}{\partial M_{it}}$$
$$p_t^L = \lambda_{it} \frac{\partial F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)\mathbb{E}[exp(\epsilon_{it})\mathcal{I}_{it}]}{\partial L_{it}}$$

where λ_{it} is the Lagrange multiplier on the constraint. λ_{it} will capture the firms' marginal cost. Multiplying the first equation by $\frac{M_{it}}{p_{it}Y_{it}}$ and the second equation by $\frac{M_{it}}{p_{it}Y_{it}}$

$$\frac{p_t^M M_{it}}{p_{it}Y_{it}} = \lambda_{it} \frac{\partial F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)\mathbb{E}[exp(\epsilon_{it})\mathcal{I}_{it}]}{\partial M_{it}} \frac{M_{it}}{p_{it}Y_{it}}$$
$$\frac{p_t^L L_{it}}{p_{it}Y_{it}} = \lambda_{it} \frac{\partial F_t(K_{it}, K_{it}^{int}, h_t(\omega_{it}^L L_{it}, M_{it}))exp(\omega_{it}^H)\mathbb{E}[exp(\epsilon_{it})\mathcal{I}_{it}]}{\partial L_{it}} \frac{L_{it}}{p_{it}Y_{it}}$$

The right-hand side of the first equation becomes the material share in revenue. Similarly, the right-hand side of the second equation becomes labor share in revenue. Let's denote them as α_{it}^M and α_{it}^L as the material and labor share in revenue, respectively. On the left-hand side of the equations, $\frac{\partial F_t()}{\partial M_{it}} \frac{M_{it}}{Y_{it}}$ is the output elasticity of material, and similarly $\frac{\partial F_t()}{\partial L_{it}} \frac{L_{it}}{Y_{it}}$ is the output elasticity of labor. Let's denote ϵ_{it}^L, and ϵ_{it}^M as the labor and material elasticity, respectively. Thus, we have

$$\alpha_{it}^M = \frac{\lambda_{it}}{p_{it}} \epsilon_{it}^M$$
$$\alpha_{it}^L = \frac{\lambda_{it}}{p_{it}} \epsilon_{it}^L \tag{30}$$

Taking their ratios we will have

$$\frac{\alpha_{it}^M}{\alpha_{it}^L} = \frac{\epsilon_{it}^M}{\epsilon_{it}^L} \tag{31}$$

This shows that the ratio of revenue shares of flexible inputs is equal to their ratio of output elasticities.

A.4 Proof of Equation 13

In the equation 32, $\frac{\lambda_{it}}{p_{it}}$ is equal to the inverse markup because the markup is defined as price over marginal cost. Since λ_{it} is the marginal cost of the firm markup becomes $\mu_{it} = \frac{p_{it}}{\lambda_{it}}$ by definition. Therefore, we can rewrite equation 32 as

$$\begin{aligned} \alpha_{it}^{M} &= \frac{\epsilon_{it}^{M}}{\mu_{it}} \\ \alpha_{it}^{L} &= \frac{\epsilon_{it}^{L}}{\mu_{it}} \end{aligned} \tag{32}$$

Rearranging the equations markups can be rewritten as

$$\mu_{it} = \frac{\epsilon_{it}^{M}}{\alpha_{it}^{M}} = \frac{\epsilon_{it}^{L}}{\alpha_{it}^{L}}$$

The markup charged by the firms equals the ratio of their flexible input elasticity to the flexible input's share in revenue.

A.5 Proof of Proposition 1

For the production function,

$$y_{it} = f_t(K_{it}, K_{it}^{int}, L_{it} h_t(\omega_{it}^{L}, \tilde{M}_{it})) + \omega_{it}^{H} + \epsilon_{it}$$

The material and labor elasticity are

$$\begin{aligned} \theta_{it}^{M} &= f_{t3} h_{t2} M_{it} \\ \theta_{it}^{L} &= f_{t3}(h_t - h_{t2}\tilde{M}_{it}) L_{it} \end{aligned}$$

where f_{t3} is the derivative of f_t function with respect to the third arguments and h_{t2} is the derivative of h_t function with respect to the second argument. The sum of labor and material elasticity is

$$\theta_{it}^{M} + \theta_{it}^{L} = f_{t3} h_{t2} M_{it} + f_{t3}(h_t - h_{t2}\tilde{M}_{it}) L_{it} = f_{t3} h_t L_{it} = f_{t3} \bar{h}_t L_{it}$$

Since the sum of elasticities do not depend on the derivative of h_t function I can replace $h_t()$ with $\bar{h}_t()$. The last equality comes from that fact.

B Data and Estimation

B.1 Estimation Strategy

This section provides a detailed estimation strategy employed. I first construct the control variables, u^1_{it} and u^2_{it} to control for Hicks-neutral and labor augmenting productivity. u^1_{it} is defined in equation 7 as

$$u^1_{it} = F_{\tilde{M}_{it}|W_{it-1},\sum_{j\in I,j\neq i} R_{jt-1}}\Big(\tilde{M}_{it} \mid W_{it-1}, \sum_{j\in I,j\neq i} R_{jt-1}\Big)$$

u^1_{it} is computed as a conditional CDF of $\tilde{M}$ given $W_{it-1}, \sum_{j\in I,j\neq i} R_{jt-1}$. However, the dependent variable, $\tilde{M}_{it}$ is a continuous variable so I can't directly apply the logistic regression. To do that, I first split the data into year and sector subsets since I do production function estimation at the sector year level. I then partition the sub-sample data (within a year and sector) into 500 parts. In cases where 500 partitioning is not feasible I partition them into the one tenth of the sample size[11]. At the boundary points of each discretized data points (named as $q \in Q$) I construct the logistic regression as follow

$$P(\tilde{M}_{it} \leq q \mid W_{it-1} = w, \sum_{j\in I,j\neq i} R_{jt-1} = r) = s_{1t}(q, w, r)$$

The dependent variable becomes a discrete variable depending on whether $\tilde{M}_{it}$ is less than q. I approximate the s_{2t} function with the second-order polynomials[12]. I, then, estimate the predicted CDF for each q points using logistic regression. I compute the remaining points using a linear interpolation.

To estimate the u^2_{it} I follow the similar strategy as u^1_{it}. The equation 9 defines u^2_{it} as follow

$$u^2_{it} = F_{M_{it}|K_{it},K^{int}_{it},W_{it-1},\sum_{j\in I,j\neq i} R_{jt-1}}\Big(M_{it} \mid K_{it}, K^{int}_{it}, W_{it-1}, \sum_{j\in I,j\neq i} R_{jt-1}\Big)$$

As in u^1_{it}, the dependent variable M_{it} is not a discrete variable, it is a continuous variable. I partition the data 500 parts at the sector year level. If the data is not feasible I split it into the one-tenth of the data. At each q level, I do

$$P(M_{it} \leq q \mid K_{it} = k, K^{int}_{it} = k^{int}, W_{it-1} = w, \sum_{j\in I,j\neq i} R_{jt-1} = r, u^1_{it} = u^1) = s_{2t}(q, k, k^{int}, w, r, u^1)$$

[11] I use different partition size such as 5. They don't change the results

[12] Third-order approximation gives the same results.

The dependent variable becomes a discrete variable depending on whether M_{it} is less than q. I similarly approximate s_{2t} function with second order polynomials. I then find the CDF at the points where $M_{it} = q$. For the remaining points, I compute the linear interpolation so that I have an estimator for u_{it}^2 at each point in the data.

After estimating u_{it}^1 and u_{it}^2, I can now estimate the productions functions. I estimate the productions at each sector over 8-year rolling windows in the equation 19

$$y_{it} = f_t(K_{it}, K_{it}^{int}, L_{it}\bar{h}_t(\tilde{M}_{it})) + c_{2t}\big(W_{it-1}, \sum_{j\in I, j\neq i} R_{jt-1}, u_{it}^1, u_{it}^2\big) + \epsilon_{it}$$

If I know $\bar{h}_t(.)$ function, I can estimate the regression using polynomial approximations for $f_t()$ and $c_{2t}()$ functions. I first approximate the $h_t()$ function with third-order polynomials.

$$h_t(\tilde{m}_{it}) = \alpha_{1t}\tilde{m}_{it}^2 + \alpha_{3t}\tilde{m}_{it}^3$$

I approximate the $f_t()$ function with second-order polynomials. For the sake of easy notation, let's define $v_{it} = l_{it} + h(m_{it})$. With the approximated $h_t()$ given, I can write $f_t()$ as follow

$$\begin{aligned} f_t() = \beta_{0t} + \beta_{1t}k_{it} \quad &+ \quad \beta_{2t}k_{it}^{int} + \beta_{3t}v_{it} + \beta_{4t}k_{it}^2 + \beta_{5t}k_{it}k_{it}^{int} + \beta_{6t}k_{it}v_{it} \\ &+ \quad \beta_{7t}(k_{it}^{int})^2 + \beta_{8t}k_{it}^{int}v_{it} + \beta_{9t}v_{it}^2 \end{aligned}$$

where β's represent the coefficients of the polynomial approximations. I have time subscripts because I estimate them over 8 year rolling windows within each sector. These coefficients will be sector-specific and time varying. I similarly approximate c_{2t} and c_{3t} with second-order polynomials. I do then minimize the moment conditions with these approximated functions in equation 22.

I estimate the production function parameters β's and $c_2()$ and $c_3()$ coefficients given α's. I first guess α values and then estimate the moment conditions using the least squares. I iterate the loops to minimize the moment conditions in equation 22. This will give me the β coefficients and productivity function coefficients. This is not a very hard loop because given *alphas* it is straightforward to find other parameters. Therefore, in the loops we are minimizing the OLS estimates to find 3 α parameters to minimize the residuals.

After estimating the parameters of $f_2()$ and c_2 functions, I can compute the elasticities as defined in the identification section. The output elasticity of intangible and tangible capital is the derivative of estimated f_t function with respect to the capital and intangible capital, respectively. Since I know that the approximated the f_t function the capital and intangible

elasticity are

$$\theta_{it}^{k} = \beta_{1t} + 2\beta_{4t}k_{it} + \beta_{5t}k_{it}^{int} + \beta_{6t}v_{it}$$
$$\theta_{it}^{int} = \beta_{2t} + \beta_{5t}k_{it}^{int} + 2\beta_{7t}k_{it}^{int} + \beta_{8t}v_{it}$$

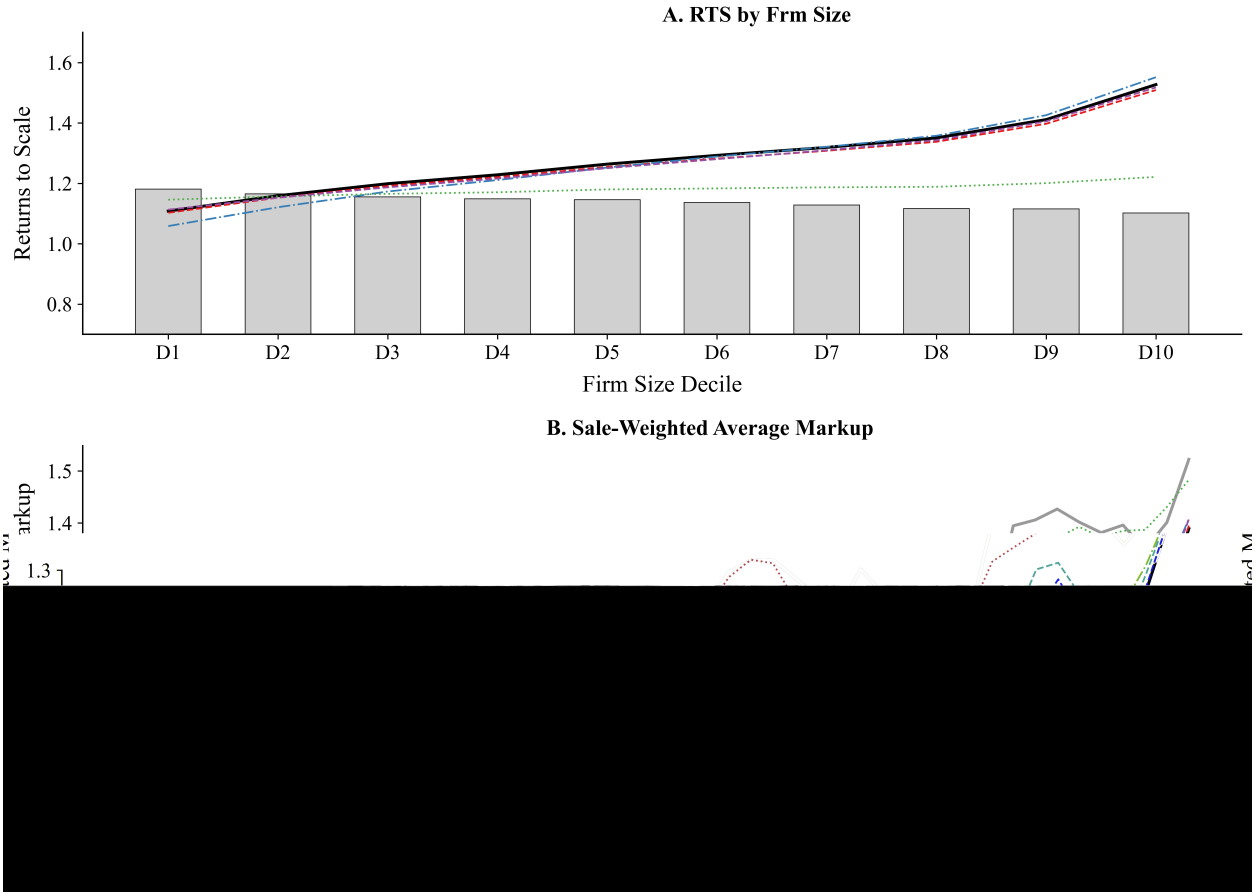
Due to the input heterogeneity this generates elasticity heterogeneity across firms. For the identification of flexible inputs I similarly take the derivative of f_t function with respect to v, this would give me the sum of flexible input elasticities. Multiplying with the flexible share in revenue I compute the labor and material elasticities. The productivity elasticity of the external knowledge is the derivative of c_2 function with respect to external knowledge capital. I construct them similar to the output elasticities.

C Robustness Results

Figure A1 summarizes the robustness of the two central findings — the RTS gradient and the markup bias — across all alternative specifications described in Section 7. Panel A reports returns to scale by firm size decile. The gray bars show the specification without intangible capital, which exhibits approximately constant returns across the size distribution. All specifications that include intangible capital produce a monotonically increasing RTS gradient, clustering tightly around the baseline. The demand-corrected specification is slightly attenuated, consistent with purging output price variation, and the measurement error simulation yields a flatter gradient, as expected from classical attenuation bias. The third-order polynomial and alternative organizational capital ($\gamma = 0.3$) specifications are nearly indistinguishable from the baseline.

Panel B reports sales-weighted average markups over time. The without-intangibles specification (gray) produces systematically higher markups throughout the sample, with the gap widening after 2000. All robustness specifications that include intangible capital track the baseline closely, confirming that the markup correction is not driven by specific parameterization choices, functional form assumptions, or demand-side contamination.

Figure A1: Robustness of Returns to Scale and Markup Estimates Across Specifications



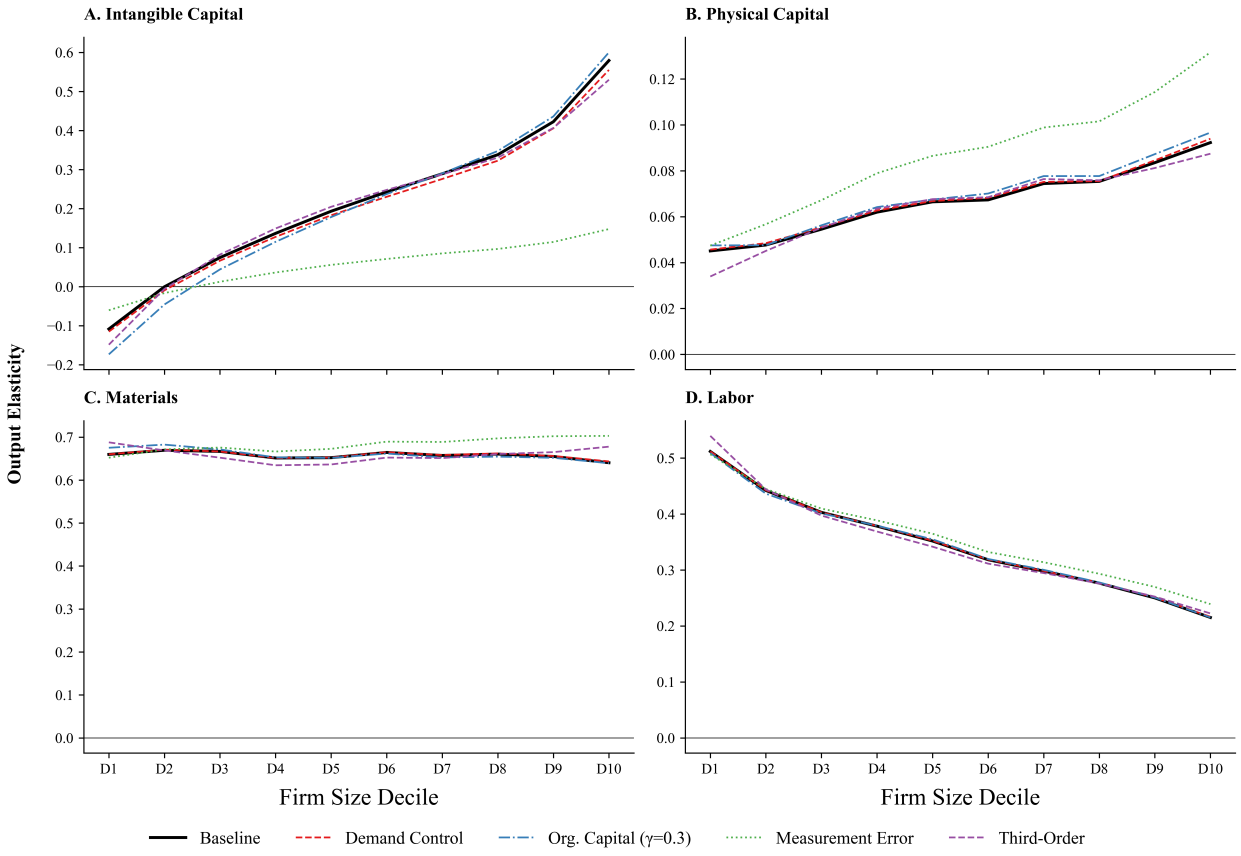
Note: Panel A reports returns to scale by firm size decile. Gray bars denote the specification without intangible capital. Lines denote the baseline and robustness specifications with intangible capital. Panel B reports sales-weighted average markups over time across all specifications. See Section 7 for details on each robustness exercise. Confidence intervals are omitted for brevity.

C.1 Robustness of Factor Input Elasticities

Figure A2 reports factor input elasticities by firm size decile across all robustness specifications. The core patterns are preserved throughout. Intangible capital elasticity (Panel A) rises monotonically with firm size across all specifications. Physical capital elasticity (Panel B) is modest and stable. Materials elasticity (Panel C) is flat across the size distribution. Labor elasticity (Panel D) declines systematically with firm size. The measurement error specification produces attenuated intangible elasticities and elevated physical capital and labor elasticities, consistent with classical attenuation bias redistributing the omitted variation to other inputs. The remaining specifications cluster tightly around the baseline, confirming that the elasticity decomposition is not driven by parameterization or functional

form choices.

Figure A2: Factor Input Elasticities by Firm Size Decile Across Robustness Specifications



Note: Each panel reports output elasticities by firm size decile across all robustness specifications described in Section 7. Confidence intervals are omitted for brevity.

Note: Y-axis scales differ across panels.

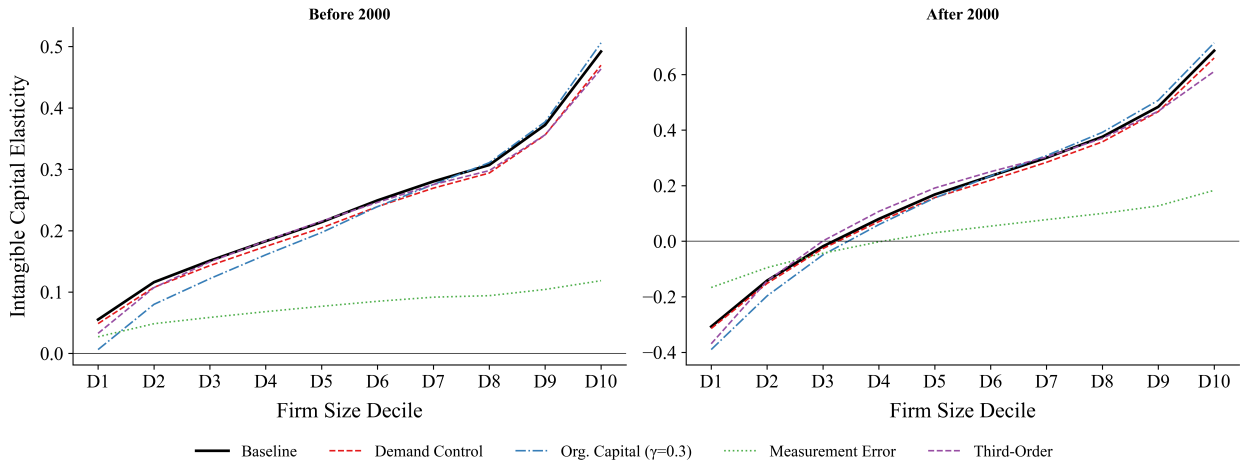
C.2 Intangible Capital Elasticity: Pre- and Post-2000 Across Specifications

A key finding in Section 6 is that the intangible capital elasticity gradient steepens after 2000, with the smallest firms exhibiting negative elasticities in the post-2000 period. Figure A3 examines the robustness of this time pattern. Before 2000 (Panel A), all specifications produce similar gradients: intangible elasticity is near zero for the smallest firms and rises gradually to approximately 0.4–0.5 for the largest. The specifications cluster tightly, confirming that the pre-2000 pattern is robust.

After 2000 (Panel B), the gradient steepens substantially across all specifications. The negative D1 elasticity is present in the baseline, demand control, organizational capital, and

third-order specifications, confirming it is a feature of the data rather than a functional form artifact. The measurement error specification produces a flatter gradient with a less negative D1, consistent with classical attenuation. The monotonically increasing gradient — the paper's central finding — is robust across all specifications and both periods.

Figure A3: Intangible Capital Elasticity by Firm Size Decile: Pre- and Post-2000 Across Specifications



Note: Panel A reports intangible capital elasticity by firm size decile before 2000. Panel B reports the same after 2000. Each line corresponds to a robustness specification described in Section 7. Confidence intervals are omitted for brevity.

D Untrimmed Productivity Heatmap

Figure A4 replicates Figure 9 without trimming the productivity distribution. The qualitative patterns are unchanged: productivity declines with firm size and, among the largest firms, declines with intangible intensity. Magnitudes are more extreme at the tails, reflecting the influence of outlier observations.

Figure A4: Industry-Year Relative Productivity by Firm Size and Intangible Intensity (Untrimmed)



Notes: The figure reports industry-year demeaned productivity (untrimmed) by firm size and intangible intensity deciles. Deciles are formed within each industry-year using log output and intangible intensity, respectively. Each cell reports the unweighted mean productivity across firms in that decile.

E TNIC Spillover Robustness

Construction

For each focal firm i in year t, the TNIC-weighted spillover is constructed as

$$S_{it}^{TNIC} = \sum_{j \neq i} \tilde{\omega}_{ij} R_{jt-1}$$

where $\tilde{\omega}_{ij} = \omega_{ij} / \sum_{k \neq i} \omega_{ik}$ are TNIC-3 similarity scores normalized to sum to one within each focal firm-year, and ω_{ij} is the raw pairwise similarity score between firms i and j from Hoberg

and Phillips (2016). I apply the standard similarity threshold of 0.01, retaining only firm pairs with meaningful product market overlap. Normalizing scores to sum to one ensures that the measure captures the average quality of spillover exposure from product-market peers rather than the total quantity.

Sample Properties

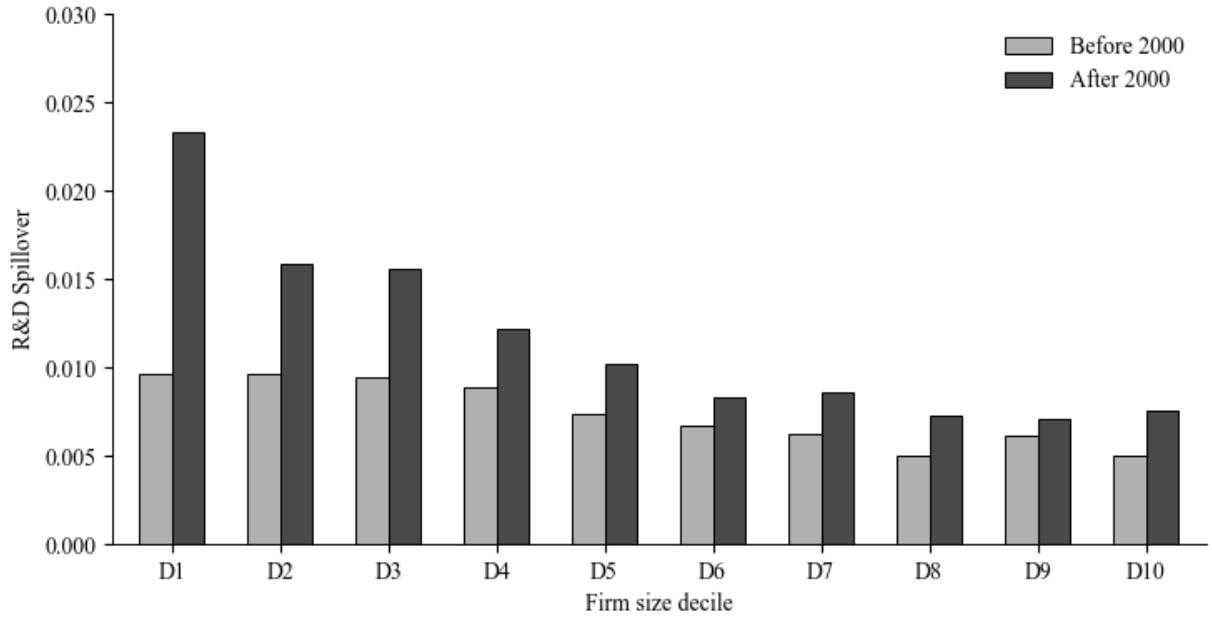
The TNIC data is available from 1989 onward. Post-1988 coverage reaches 72.7% of firm-year observations, yielding a robustness sample of 72,088 observations covering 8,435 firms through 2023 with a mean of 26.4 product-market peers per firm. TNIC coverage is somewhat lower among smaller firms, ranging from approximately 55% in the smallest size decile to 83% in the largest. This size gradient in coverage makes the robustness check conservative: since small firms exhibit the highest spillover exposure in the main specification, their underrepresentation mechanically compresses the estimated size gradient.

Results

The knowledge spillover gradient is negative as in the main specification — spillover exposure declines with firm size — but is attenuated in magnitude, with Δ_{50} of -0.006 and Δ_{10} of -0.012 compared to -0.070 and -0.127 in the baseline. This attenuation has a straightforward interpretation. The TNIC measure captures narrow product-market peer spillovers, while the main industry-aggregate measure captures broad sectoral knowledge flows. The declining size gradient in the main specification reflects large firms reducing their exposure to broad sectoral knowledge as they internalize R&D. The weaker gradient under TNIC suggests that large firms maintain relatively higher engagement with immediate product-market competitors — consistent with competitive pressure forcing monitoring of close rivals — even as they reduce reliance on broader industry knowledge flows.

Figure A5 reports TNIC-weighted spillover exposure by firm size decile before and after 2000. The post-2000 pattern shows a clear declining gradient consistent with Figure 8, while the pre-2000 pattern is relatively flat across the size distribution.

Figure A5: Average Knowledge Spillover Before and After 2000 by Firm Size (TNIC Weighted)



Notes: The figure reports average TNIC-weighted external R&D spillover exposure by firm size decile before and after 2000. Spillover exposure is constructed as the normalized similarity-weighted sum of peer R&D capital stocks following Hoberg and Phillips (2016), applying the standard similarity threshold of 0.01.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Acemoglu, D. (2002). Directed technical change. *The review of economic studies 69*(4), 781–809.

Akcigit, U. and S. T. Ates (2023). What happened to us business dynamism? *Journal of Political Economy 131*(8), 2059–2124.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The quarterly journal of economics 135*(1), 105–163.

Basu, S. (2019). Are price-cost markups rising in the united states? a discussion of the evidence. *Journal of Economic Perspectives 33*(3), 3–22.

Basu, S. and J. G. Fernald (1997). Returns to scale in us production: Estimates and implications. *Journal of political economy 105*(2), 249–283.

Bernstein, J. and M. I. Nadiri (1988). Interindustry r&d spillovers, rates of return, and production in high-tech industries. NBER Working Papers 2554, National Bureau of Economic Research, Inc.

Bloom, N., M. Schankerman, and J. Van Reenen (2013). Identifying technology spillovers and product market rivalry. *Econometrica 81*(4), 1347–1393.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of monetary economics 121*, 1–14.

Brynjolfsson, E., D. Rock, and C. Syverson (2021). The productivity j-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics 13*(1), 333–372.

Corrado, C., C. Hulten, and D. Sichel (2009). Intangible capital and us economic growth. *Review of income and wealth 55*(3), 661–685.

Crouzet, N. and J. Eberly (2020). Rents and intangible capital: A q+ framework. *Unpublished manuscript, Northwestern University*.

Crouzet, N. and J. C. Eberly (2019). Understanding weak capital investment: The role of market concentration and intangibles. Technical report, National Bureau of Economic Research.

Crouzet, N., J. C. Eberly, A. L. Eisfeldt, and D. Papanikolaou (2022). The economics of intangible capital. *Journal of Economic Perspectives 36*(3), 29–52.

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

De Ridder, M. (2024). Market power and innovation in the intangible economy. *American Economic Review 114*(1), 199–251.

Demirer, M. (2025). Production function estimation with factor-augmenting technology: An application to markups. *Econometrica*. Conditionally accepted.

Doraszelski, U. and J. Jaumandreu (2018). Measuring the bias of technological change. *Journal of Political Economy 126*(3), 1027–1084.

Edmond, C., V. Midrigan, and D. Y. Xu (2023). How costly are markups? *Journal of Political Economy 131*(7), 1619–1675.

Eisfeldt, A. L. and D. Papanikolaou (2014). The value and ownership of intangible capital. *American Economic Review 104*(5), 189–94.

Ewens, M., R. H. Peters, and S. Wang (2025). Measuring intangible capital with market prices. *Management Science 71*(1), 407–427.

Goldin, I., P. Koutroumpis, F. Lafond, and J. Winkler (2024). Why is productivity slowing down? *Journal of Economic Literature 62*(1), 196–268.

Haskel, J. and S. Westlake (2017). Capitalism without capital: The rise of the intangible economy.

Hoberg, G. and G. Phillips (2016). Text-based network industries and endogenous product differentiation. *Journal of political economy 124*(5), 1423–1465.

Holmes, T. J. and J. Schmitz (2010). Competition and productivity: A review of evidence. *Annual Review of Economics 2*(1), 619–642.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics 124*(4), 1403–1448.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

Keller, W. and S. R. Yeaple (2009). Multinational enterprises, international trade, and productivity growth: firm-level evidence from the united states. *The review of economics and statistics 91*(4), 821–831.

Klette, T. J. and Z. Griliches (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of applied econometrics 11*(4), 343–361.

Koh, D., R. Santaeulàlia-Llopis, and Y. Zheng (2020). Labor share decline and intellectual property products capital. *Econometrica 88*(6), 2609–2628.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1053.

Kydland, F. E. and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, 1345–1370.

Lev, B. and S. Radhakrishnan (2003). The measurement of firm-specific organization capital.

Lev, B., S. Radhakrishnan, and P. Evans (2016). Organizational capital: A ceo's guide to measuring and managing enterprise intangibles. *The Center for Global Enterprise*.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies 70*(2), 317–341.

Loecker, J. D. and F. Warzynski (2012). Markups and firm-level export status. *American economic review 102*(6), 2437–2471.

Milgrom, P. and J. Roberts (1982). Limit pricing and entry under incomplete information: An equilibrium analysis. *Econometrica: Journal of the Econometric Society*, 443–459.

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

Peters, R. H. and L. A. Taylor (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics 123*(2), 251–272.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy 98*(5, Part 2), S71–S102.

Syverson, C. (2011). What determines productivity? *Journal of Economic literature 49*(2), 326–365.

Van Reenen, J. (2018). Increasing differences between firms: Market power and the macro-economy.