Firm Specialization vs. Diversification

Anmol Bhandari[1], Sanhitha Jugulum[2], Laura Veldkamp[3], Venky Venkateswaran[4]

[1]University of Minnesota and NBER
[2]Columbia Business School
[3]Columbia Business School and NBER
[4]New York University and NBER

March 18, 2026

PRELIMINARY AND INCOMPLETE*

Abstract

This paper investigates the fundamental trade-off between firm specialization and diversification. While management theories emphasize core competency to maximize efficiency, finance favors diversification to mitigate cash flow volatility. We develop a heterogeneous-firm general equilibrium model to quantify these opposing forces, introducing new empirical measures for firm-sector productivity and firm scope. To quantify the model, we extract granular revenue shares from 10-K filings, using a large language model. We define economic distance between products as the inverse of their revenue covariance. Our analysis reveals that firms expand non-randomly into related sectors to leverage expertise. However, they also diversify into distant sectors to smooth risk, which endogenously lowers their markups. This diversification imposes a significant productivity penalty. Our calibration shows that broadening scope to reduce firm risk compromises 19.31% of U.S. aggregate output. This framework unifies finance, management, and macro-IO perspectives to provide a novel understanding of firm scope.

Expanding the number of products a firm produces is central to modern theories of growth and firm dynamics. In models such as Klette and Kortum (2004) and their many descendants, firms grow by innovating and introducing new goods, driving creative destruction and aggregate productivity. Yet these models are largely silent on what kinds of goods choose to produce. Should a firm deepen its expertise by producing related products, or diversify into new and distinct lines? Other fields offer guidance, but they disagree. Management theory emphasizes core competency: firms should specialize in related activities to exploit shared technologies, brands, and know-how (Rumelt, 1974; Prahalad and Hamel, 1990). This view

*This is a rough draft. Results are subject to change. Please do not cite or distribute.

elevates Adam Smith's idea that specialization is the source of efficiency, to the level of the firm. Finance, in contrast, stresses diversification: firms should expand into different businesses to reduce cash flow volatility and risk (Lewellen, 1971; Stein, 1997; Maksimovic and Phillips, 2002). Both logics are compelling, yet they push in opposite directions. How firms choose to specialize or diversify and the aggregate consequence of those decisions remains an open question.

In this paper, we develop a heterogeneous-firm GE model to measure both specialization and diversification in firm expansion. While the model uses standard components from macro and finance, the framework guides our development of new empirical measures of firm-sector productivity and firm scope. This allows us to assess the magnitude of the trade-off. We find that product choice and risk significantly affects aggregate output. Risk mitigation by firms – specifically, expanding their operations in sectors with lower productivity for diversification – lowers output by 19.31% relative to an otherwise identical economy but without risk. These results offer a new perspective on the cost of risk: Risk is costly for economies because it induces firms to take risk-reduction measures, like broadening scope, that compromise output.

In our model (Section 1), firms choose a portfolio of products that differ in both productivity potential and revenue covariance. Producing goods that are similar allows firms to leverage existing expertise, while producing goods that are dissimilar reduces exposure to shocks. The model implies that the distance between products—negatively related to the correlation of their revenue streams—captures the key economic dimension of relatedness. Within this structure, we estimate a parameter governing the rate of decay in productivity as firms move away from their core competencies. This parameter quantifies the gains to specialization: it tells us whether firms are more productive in activities close to their expertise, and by how much.

The model's ingredients together offer a new interpretation of firm scope that reflects how differently a firm's products perform, not just how many it makes or how many industries it operates in. The model also reveals that scope is directly linked to a firm's markups. Because markups compensate firms for bearing risk, and because broader scope reduces risk through diversification, firms with greater scope should charge lower markups in equilibrium. These predictions follow from firms' productivity–risk trade-offs, offering a novel explanation for markup and scope patterns across firms and industries.

It is worth emphasizing the difference between scope and firm size. Larger firms typically charge *higher* markups, reflecting scale economies or market power. In contrast, firms with greater scope—that is, firms that produce a broader and more diversified set of goods—charge *lower* markups, consistent with diversification lowering the cost of risk-bearing. This

distinction uncovers a new margin of heterogeneity in markups and risk exposure.

After describing the data, Section 2 constructs three new measures: 1) Firm-sector productivities, 2) firm scope and 3) a price of risk, based on firms' revealed preference. While previous work (Hoberg and Phillips, 2025) significantly advanced the measurement of firm scope by capturing the number of sectors a firm operates in, this is not sufficient to assess diversification. It would be like judging stock portfolio risk according to the number of stocks held. What would be missing is the covariance of the stocks and how much one holds of each. Similarly, we advance scope measurement by adding sector covariances and revenue shares in each sector. This allows scope to fit our purpose – to measure firms' trade-off between diversification and specialization.

Our measure of each firm's productivity in each sector of activity is also a novel empirical contribution, of potentially broader interest to the firm dynamics literature. This exercise uncovers a detailed productivity profile across sectors, which we summarize in a low-dimensional representation consisting of a firm's "home-sector" productivity and a rate of decay in product space. This representation is both interpretable and portable, providing a set of sufficient statistics that others can use to study firm productivity in each of its activities.

Deriving these sector-level productivities requires estimating two key objects: (i) each firm's revenue shares across sectors and (ii) the covariance structure of revenues between sectors, which our model interprets as economic distances. The firm-sector revenue share is a new data set that we create using textual analysis of firms' 10-K reports. The covariances come from standard data sources, but provide the empirical foundation for our measure of relatedness between industries. These measures allow us to connect firm behavior to its multi-sector productivity, its cash flow risk and its scope.

Using this new firm-sector level data, Section 3 documents three key facts:

Fact 1: Scope reduces risk. Broader scope is associated with a lower probability of exit, consistent with the diversification hypothesis.

Fact 2: Markups decline with scope. Changes in firm scope are associated with higher marksups, consistent with firms demanding higher compensation for operating in riskier sectors. This pattern also helps us pin down a key parameter in the model: the price of risk.

Fact 3: Productivity declines with distance from the home sector. This is the gains to specialization. The set of sectors firms operate in is not random: firms are more likely to be in sectors/products that are closer to their existing (or 'home') sector.

Firms also have higher TFP in these nearby sectors compared to ones that are further away.

Fact 2 allows us to measure the price of risk in a new way. While risk price measures abound, most are derived from equity markets, reflecting the preferences and constraints of equity traders. Our measure reflects the revealed preference of firms trading off their own return for risk reduction through scope expansion.[1]

Finally, Section 4 uses the productivity, scope and risk measures for each firm to estimate the aggregate cost of firm risk. If each firm faced no uncertainty, its production decisions would be driven solely by productivity. We estimate that output in such an environment would be about 19.31% higher. In other words, the cost of firms' actions to mitigate risk is about 19.31% of aggregate income. We show that almost a third of this loss can be attributed to the covariance between risk and productivity: in particular, the fact that the most productive firms demand the most compensation for risk. This pattern amplifies the output losses from risk (relative to a counterfactual world in which risk compensation and productivity were uncorrelated).

Contribution to the existing literature. Our paper unifies three historically distinct perspectives on firm scope through a measurable construct of product distance and revenue covariance. By doing so, we bridge the gap between the finance view of diversification, the management view of core competencies, and the macro-IO view of multi-product dynamics.

The Finance View: Diversification and Risk Management In corporate finance, the study of firm scope is primarily framed through the lens of risk management and internal capital markets. Early work by Lewellen (1971) and more recently Stein (1997) suggests that firms expand across activities to mitigate financing frictions and smooth idiosyncratic shocks. Maksimovic and Phillips (2002) and Matvos and Seru (2014) further explore how diversified firms allocate resources across segments, often finding that internal capital markets can provide stability even if they lead to a valuation discount. We complement with work with a production economy that also features gains to specialization and estimating how the two forces trade off.

The Management View: Core Competencies and Specialization In contrast to the finance view, management and organizational scholars emphasize the costs of straying

[1]While firms' choices maximize equity values in ideal settings, Jensen and Meckling (1976); Myers and Majluf (1984) and Stein (2003) have shown that incentive misalignment, constraints and frictions can break this link, causing the firm's internal price of risk to differ from the equity holders'.

too far from a firm's core. Rumelt (1974) and the seminal work by Prahalad and Hamel (1990) argue that firms should focus on their core competencies to maintain organizational coherence and efficiency. Teece et al. (1994) formalized the notion that firms are constrained by their existing knowledge and routines, a finding echoed by empirical evidence from Bresnahan et al. (2011), who show that firms expand most successfully into technologically or organizationally proximate areas. Our model translates these qualitative notions of "coherence" and "relatedness" into a quantitative distance parameter (θ), allowing us to quantify the trade-off between the gains from specialization and the gains from diversification.

The Macro-IO View: Multi-Product Firms and Markups The macroeconomics and industrial organization literatures typically treat scope as the number of products produced or sectors served, sometimes linked through firm-level constraints. Foundational models such as Bernard et al. (2010) and Mayer et al. (2016) focus on product switching and how firms sort across products based on productivity. However, many models follow a Klette and Kortum (2004) or Eaton and Kortum (2002) structure where products arrive randomly, making the models silent on the composition of firm's products. We extend this literature by introducing an explicit product space and showing why the distance of firms' products from each other matters for aggregate output.

Our model builds on the multi-attribute product space pioneered by Pellegrino (2025), to explain various industrial organization pricing and elasticity facts. Our key departure is the introduction of firms that price revenue risk. This is essential for studying the costs and benefits of product diversification. Consistent with our theory, Boar and Midrigan (2023) demonstrate that much of the cross-sectional markup dispersion stems from compensation for risk.

Our ideas are most closely related to recent theories of innovation. Berlingieri et al. (2025) characterizes innovation as happening in "bursts," where a single discovery leads a firm to introduce multiple new products at once, to capture market power. Furthermore, De Ridder (2021) argues that information technology increases the fixed costs of industry entry while decreasing marginal costs, which incentivizes firms to specialize and narrow their production scope. Ma et al. (2024) show that business concentration has risen, in a way consistent with the growing importance of intangible assets. Our approach estimates these forces as gains to specialization and complements this work by examining how these types of gains compete with diversification.

Finally, we speak to the macro literature on markups. Edmond et al. (2023) document the rise of markup inequality. In contrast, our structural counterfactual provides a new decomposition of markups, identifying how much of this dispersion is driven by heterogeneity

in a firms' scope across product space.

Scope Measurement Methodologically, our work contributes to the burgeoning field of text-based analysis in economics. We build on the pioneering work of Hoberg and Phillips (2016), who used textual similarity in 10-K filings to define product differentiation and competitors. However, whereas Hoberg and Phillips (2025) and Kim et al. (2023) often rely on cosine similarity or transformer-based embeddings to measure linguistic proximity, we innovate in three ways: (1) we map firms to NAICS sectors, to better align with macroeconomic data. (2) More conceptually, we extract firm-sector revenue shares. This allows us to describe how much of each sector's risk a firm is exposed to. (3) We estimate firm-sector productivity, which is an essential input into our estimate of the cost of firm uncertainty and could be useful for other applications. (4) Using revenue covariance, we measure the distance of sectors from each other. Instead of just counting up the number of sector codes assigned to the firm's establishments, we distinguish between operating in nearby industries, or branching out into business lines that are less related.

Eberly et al. (2023) define scope in terms of input applicability across tasks, while our framework focuses on outputs and revenue. One could also study how firms diversify input risk and the costs of doing so, with a similar methodology.

Finally, Autor et al. (2020) document a recent decline in measured firm scope at the 4-digit industry level; our distance-based measure provides a richer perspective on these trends, emphasizing that reductions in nominal scope need not imply lower diversification when products move within nearby industries.

Measuring distance has deep roots in the patent literature: Akcigit et al. (2013) and Cai and Li (2019) use patent linkages to trace technological proximity.[2] We extend these insights from patents to products, capturing market-based rather than invention-based proximity. This distinction is important: Most patents are never brought to market and most products are not patented.

1 A Model of Firm Scope

Our environment is a canonical heterogeneous-firm equilbrium production economy. with two key departures: 1) the existence of a product space, and 2) the assumption that firms care about revenue variability, rather than just expected profits. Both are essential to connect to diversification motives. Both are also realistic. Risk management practices are central to

[2]Related work values these intangible assets (Ayyagari et al., 2023; Eisfeldt and Papanikolaou, 2014) and their macro roles (Acemoglu et al., 2018).

the way business schools train corporate decision makers.

1.1 Model setup

The economy has a continuum of sectors, indexed by $j \in \mathcal{J}$ (e.g., $\mathcal{J} = [0,1]$ with Lebesgue measure). Within each sector j, a continuum of firms, indexed by $i \in \mathcal{I}_j$ (e.g., $\mathcal{I}_j = [0,1]$), produces differentiated varieties. The measure of operating firms in sector j is N_j, taken as exogenous unless otherwise specified. Sectoral output is a composite of different varieties, using a CES aggregator:

$$Q_j = \left(\int_{i \in \mathcal{I}_|} Q_{ij}^{\frac{\eta-1}{\eta}} di \right)^{\frac{\eta}{\eta-1}} . \tag{1}$$

The single final good, denoted Q, is a CES-composite of the sectoral output:

$$Q = \left(\int_{j \in \mathcal{J}} B_j^{\frac{1}{\theta}} Q_j^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}} . \tag{2}$$

where $B_j > 0$ is a preference shock, the source of risk. The shocks $\{B_j\}_{j \in \mathcal{J}}$ are normalized so that $\mathbb{E}[B_j^{\frac{1}{\theta}}] = 1$. The joint second moments exist, with $\text{Var}(\ln B_j^{1/\theta}) = \sigma^2$ and $\text{Cov}(\ln B_j^{1/\theta}, \ln B_k^{1/\theta}) = \rho_{jk}\sigma^2$, where $\rho_{jk} = \rho_{kj}$, $\rho_{jj} = 1$.

The assumption of sectoral demand shocks as the only source of risk is made for simplicity. It is a stand-in for the many types of uncertainty firms face.

Preferences, numeraire, and household budget. The representative household has preferences over consumption and leisure, with a utility function $u(C, L) = \ln C - L$.

Normalizing the price of the final good to 1, the household budget constraint is

$$C = WL + \Pi ,$$

where Π denotes aggregate firm profits.

Technology, inputs, and costs. Production by firm i in sector j is given by $Q_{ij} = A_{ij}L_{ij}$. Labor markets are competitive. The marginal cost is thus $MC_{ij} = \frac{W}{A_{ij}}$ where W denotes the wage.

Timing of decisions. Firms choose production $\{Q_{ij}\}$ observing $\{A_{ij}\}$ (and factor prices) but before the realization of the sectoral demand shocks $\{B_j\}$. The household makes its

consumption decisions after $\{B_j\}$ is observed, so the realized prices are given by

$$P_{ij} = \left(\frac{Q_{ij}}{Q_j}\right)^{-\frac{1}{\eta}} P_j \qquad \text{where} \qquad P_j = B_j^{\frac{1}{\theta}} \left(\frac{Q_j}{Q}\right)^{-\frac{1}{\theta}} . \tag{3}$$

Equation (3) thus defines the inverse demand for each variety, as a function of the choices of other firms (and sectors) and the demand shock.

Risk and the firm's problem. We assume that firms price risk. Otherwise, they would have no motive to diversify. Specifically, they discount cash flows linearly in the variance of revenues (a mean–variance objective). Firms make production decisions under uncertainty to maximize risk-adjusted profits:

$$\max_{\{Q_{ij}\}} \qquad \mathbb{E} \sum_j \left[P_{ij} Q_{ij} - \frac{W}{A_{ij}} Q_{ij} \right] - \gamma_i \operatorname{Var}\left(\sum_j P_{ij} Q_{ij} \right), \tag{4}$$

where $\gamma_i \geq 0$ is the firm's risk price. The variance is taken with respect to demand shocks $\{B_j\}$.[3]

Equilibrium: An equilibrium in this economy consists of (i) production choices $\{Q_{ij}\}$ that are optimal for firms, i.e. solve (4), (ii) sectoral and economy-wide quantities, $\{Q_j\}$ and Q respectively, that are consistent with the firm-level choices, i.e. satisfy (1)-(2), and (iii) wages W that clear labor markets, i.e. satisfy the household optimality condition $Q = W$.

1.2 Equilibrium: Risk, Scope and Markups

We begin by approximating and then solving the firm's problem. Then we express the solution in terms of markups, risk expose and a risk price. Finally, we map these objects into our firm scope measure. This forms the basis for our measurement.

To derive revenue variability, $\operatorname{Var}\left(\sum_j P_{ij} Q_{ij}\right)$, we use a first-order approximation of revenues (around $B_j^{\frac{1}{\theta}} = \mathbb{E}[B_j^{\frac{1}{\theta}}]$:

$$\sum_j P_{ij} Q_{ij} \approx \sum_j R_{ij} + \sum_j R_{ij} \left(\ln B_j^{\frac{1}{\theta}} - \ln \mathbb{E}[B_j^{\frac{1}{\theta}}] \right) . \tag{5}$$

[3]This mean-variance objective is consistent with empirical corporate finance evidence on firms' decisions (Eckbo, ed, 2008) and with macro evidence that firms increase investment and sales in response to a randomized data treatment (Kumar et al., 2023). It could represent a convex cost, firm bankruptcy risk or the risk-aversion of a manager with equity compensation.

where $R_{ij} \equiv \mathbb{E}[B_j^{\frac{1}{\theta}}] Q_j^{\frac{1}{\eta}-\frac{1}{\theta}} Q^{\frac{1}{\theta}} Q_{ij}^{1-\frac{1}{\eta}}$ is the expected revenue of firm i from sector j. Then,

$$\text{Var}\left(\sum_j P_{ij} Q_{ij}\right) \approx \sum_j \sum_k R_{ij} R_{ik} \sigma_{jk} \ , \tag{6}$$

where $\sigma_{jk} \equiv \text{Cov}\left(\ln B_j^{\frac{1}{\theta}}, \ln B_k^{\frac{1}{\theta}}\right)$.

The solution to the firm's problem (4) can be expressed in terms of markups over marginal costs. In particular, the optimal expected markup for firm i in sector j is

$$\frac{\mathbb{E}[P_{ij}]}{MC_{ij}} = \frac{1}{1-\tau_{ij}} \cdot \frac{\eta}{\eta-1} \ , \tag{7}$$

where the factor $\frac{1}{1-\tau_{ij}}$ is a risk wedge. It scales up the standard markup, $\frac{\eta}{\eta-1}$ and indexes the compensation required by the firm for bearing the risk from its sector j activities. The term τ_{ij} is the product of a (firm-specific) price of risk and (firm-sector-specific) exposure:

$$\tau_{ij} \ := \ RiskPr_i \cdot RiskExp_{i,j} \tag{8}$$

$$\text{where} \qquad RiskPr_i \ := \ 2\gamma_i R_i \sigma^2 \tag{9}$$

$$RiskExp_{i,j} \ := \ \sum_k \frac{R_{ik}}{R_i} \rho_{jk}. \tag{10}$$

The price of risk increases with the firm's sensitivity to revenue risk, γ_i, size (expected total revenue R_i) and the variance of sectoral demand shocks, σ^2. The risk exposure is the revenue-weighted average correlation between sector j and the other sectors in the firm's portfolio. It measures how much additional risk sector j contributes to the firm's overall revenue risk. If sector j is highly correlated with the rest of the firm's activities (high ρ_{jk}), then $RiskExp_{i,j}$ is high. This raises τ_{ij} and therefore increases the optimal markup. Intuitively, selling more in a highly correlated sector amplifies firm-level risk, so the firm requires higher markups as compensation. Conversely, sectors that provide diversification benefits reduce effective risk exposure and see lower markups.

The firm-level markup is:

$$\frac{\sum_j \mathbb{E}[P_{ij}] Q_{ij}}{\sum_j MC_{ij} Q_{ij}} = \frac{1}{1 - RiskPr_i \cdot RiskExp_i} \ , \tag{11}$$

$$\text{where} \qquad RiskExp_i = \sum_j \frac{R_{ij}}{R_i} \ RiskExp_{ij} \tag{12}$$

is the overall risk exposure of the firm. It is a weighted average of $\{RiskExp_{ij}\}$ with expected revenue shares $\frac{R_{ij}}{R_i} = \frac{\mathbb{E}[P_{ij}Q_{ij}]}{\sum_k \mathbb{E}[P_{ik}Q_{ik}]}$ as weights. A more diversified portfolio of activities lowers the average risk exposure and therefore, the firm finds it optimal to charge a lower markup.

Thus, measured markups reflect not only demand conditions but also the riskiness of the firm's portfolio of activities. The model thus predicts systematic cross-sector and cross-firm variation in markups driven by risk considerations and diversification motives, even when elasticities are constant.

This insight motivates the following measure of firm i's scope (from a risk perspective):

$$Scope_i \quad = \quad 1 - RiskExp_i \quad = \quad 1 - \sum_j \sum_k \frac{R_{ij}}{R_i} \frac{R_{ik}}{R_i} \rho_{jk} \, . \tag{13}$$

A firm which operates in only 1 sector (or in a set of perfectly correlated sectors) has $Scope_i = 0$. At the other extreme, a perfectly diversified firm has no exposure to sectoral demand shocks and has $Scope_i = 1$. Thus, $Scope_i$ indexes the breadth of the firm's operations from a risk perspective.

1.3 Aggregation

We now aggregate firm-level decisions to obtain implications for sectoral and economy-wide objects. This will highlight how risk compensation distorts aggregate production.

Sectoral aggregation. Since the sectors are symmetric, in equilibrium, aggregate output Q is the same as sectoral output Q_j. In Appendix A, we derive the following expression for equilibrium Q_j:

$$Q_j = \frac{\eta - 1}{\eta} \left(N_j^{\varpi} \int (1 - \tau_{ij})^{\eta - 1} A_{ij}^{\eta - 1} \, di \right)^{\frac{1}{\eta - 1}} . \tag{14}$$

In the absence of risk (e.g. $\gamma_i = 0 \Rightarrow \tau_{ij} = 0$ for all i), this expression collapses to the standard formula in a CES economy. Risk compensation enters as a wedge that reduces effective productivity. Intuitively, risk induces firms to contract production, relative to the level implied by productivity alone.

To quantify the output losses from risk considerations, define the no-risk benchmark Q_j^*, which denotes sectoral (or aggregate) output when $\gamma_i = 0$ (so $\tau_{ij} = 0$):

$$Q_j^* = \frac{\eta - 1}{\eta} \left(N_j^{\varpi} \int A_{ij}^{\eta - 1} \, di \right)^{\frac{1}{\eta - 1}} . \tag{15}$$

In the next subsection, we outline how we take the model – specifically, equations (14)-(26) – to the data and quantify the effects of risk on markups and aggregate output.

1.4 From Model to Data

Sufficient statistics. Firm–sector productivities $\{A_{ij}\}$ and risk compensation, $\{\tau_{ij}\}$ are not directly observable. Using equations (14)-(26), we can recover them from a small set of sufficient statistics: (i) expected revenues by firm-sector $\{R_{ij}\}$, (ii) sectoral demand-shock correlations ρ_{jk}, and (iii) firm-level markups. The model equations directly link these variables to the structural parameters that drive risk-adjusted markups and aggregate distortions.

Measuring risk exposure. Given expected revenue shares and the correlations, risk exposures (both at the firm-sector level and at the firm level) can be directly pinned down from equations (10) and (12).

Identifying the price of risk. Equations (11)-(13) imply that, as a firm expands or contracts its sectoral scope, its risk exposure changes. The price of risk $RiskPr_i$ then governs the change in markups. Therefore, the variation in firm-level markups as risk exposure (or equivalently scope) changes is informative about how much risk matters to the firm. In other words, observed joint movements in firm scope and markups identifies the price of risk.

Recovering productivities. Here, we make use of the model equations which link firm–sector revenues $\{R_{ij}\}$ to the corresponding productivities $\{A_{ij}\}$ and markups. The latter can be explicitly accounted for (using the risk exposures and price of risk), allowing us to back out productivities $\{A_{ij}\}$. Specifically, using the firm's optimality conditions and the demand functions, we can derive the following relationship:

$$\ln A_{ij} = \frac{1}{\eta - 1} \ln s_{ij} + \ln\left(\frac{1}{1 - \tau_{ij}}\right) + \ln\left(\frac{\eta}{\eta - 1}\right) \qquad \text{where} \tag{16}$$

$$s_{ij} := \frac{P_{ij} Q_{ij}}{P_j Q_j} \tag{17}$$

denotes firm i's market share in sector j.

In summary, the approach proceeds in three steps: (i) combine revenue shares with sectoral correlations to measure risk exposure, and (ii) use the relationship between scope and markups to estimate the price of risk and (iii) use revenues to infer market shares and

effective productivities. These objects are sufficient to quantify how scope and risk shape markups and matter for aggregate output. In the following section, we detail our sources and measurement approach.

2 Data and Measures

2.1 Data sources and definitions

This subsection describes our measurement strategy, including the data sources, sector definitions, firm filings, and markup estimates.

Sector definitions and correlations. We define sectors at the 4-digit North American Industry Classification System (NAICS) industry codes. This level of aggregation balances sufficient granularity with adequate data coverage across our sources.

A measure of correlation between demand shocks hitting each sector (ρ_{jk}) is essential for our exercise. We use the correlation of sectoral employment from the Business Dynamics Statistics (BDS), published by the U.S. Census Bureau, as our preferred measure of correlation in demand. The BDS provides total employment at the 4-digit NAICS level for every year starting in 1978. For each sector-pair j, k, we compute the correlation between total sectoral employment over 1978–2022. This gives us a measure of $\rho_{jk} = \rho_{kj}$. We end up with a total of 288 sectors and therefore a 288×288 matrix of sectoral correlations. We assume that the correlation matrix is time-invariant.

Firm sample. Our sample consists of U.S. publicly traded firms that file annual reports (10-Ks). The sample period runs from 2000 to 2023, comprising approximately 32,687 observations. To capture meaningful changes in firm scope and exposure, we analyze filings at a horizon of $h = 5$ years, pairing 10-K reports five years apart for the same firm. This horizon smooths short-run noise and focuses on medium-run adjustments in business composition.

Markups and risk measures. Firm-level markups are computed using the methodology of De Loecker et al. (2020), implemented on Compustat financial statement data. This approach recovers markups from production data by combining cost shares with estimated output elasticities and is widely used in the empirical industrial organization and macro-productivity literature.[4]

[4]We use an open-source Python implementation published by Basile Grassi to ensure consistency and replicability.

2.2 Measuring Revenue Shares

Measuring a firm's diversification requires knowing i) the sectors in which a firm operates (what previous measures captured), but also ii) the revenue shares of each of these sectors in the firm's portfolio, and iii) the covariance of these sectors. Sector covariance (iii) is straightforward to measure. We return to that later.

Traditional data sources do not reveal i) or ii). For instance, Compustat segment data is too broad. These are parsed from management disclosures. Two-thirds of firms report only one segment. One could, in principle, use the Census establishment data to get a detailed segment level data at the firm level. However, it is restricted to domestic operations and does not attribute headquarter services to sectors. A third approach uses textual analysis of regulatory filings. Prior work (e.g., Hoberg and Phillips (2025)) measures industry exposure as the fraction of words in a firm's 10-K filing that match industry-specific word dictionaries. While innovative, it is not obvious how to map industry exposure into revenue *shares* at the firm level. That is, the objective of the exercise was to determine the extensive margin of operation in an industry. It does not provide an obvious way to transform these exposure measures into revenue shares at the firm level. At the very least, it requires strong assumptions, and small changes in language can generate noisy inferences about economic activity.

Therefore, we develop new techniques to identify sectors and measure of firm-sector revenue shares (π_{ijt}). We use a large language model (LLM) to parse the business, risk factors, and management's discussion and analysis sections of firm 10-K filings. We collect all the 10-K filings from the U.S. Securities and Exchange Commission using the SEC-API between 2000 and 2025. The approach relies on narrative disclosures in firms' annual 10-K filings. Specifically, we use ChatGPT's GPT-5.2 model to translate qualitative business descriptions into quantitative sectoral revenue shares. The procedure is designed to recover a standardized NAICS-based revenue allocation that can be linked directly to BEA sectoral data and to our model-based measures of firm scope and risk exposure. Figure 1 summarizes the two-stage pipeline.

To ensure replicability and minimize simulation noise, we use a two-stage procedure. The first stage summarizes the relevant business-line descriptions from Sections 1, 1A, 1B, 7 and 8. The model reads selected sections of the 10-K and produces a summary that isolates information relevant for industry classification. The prompt for stage 1 input reads as follows:

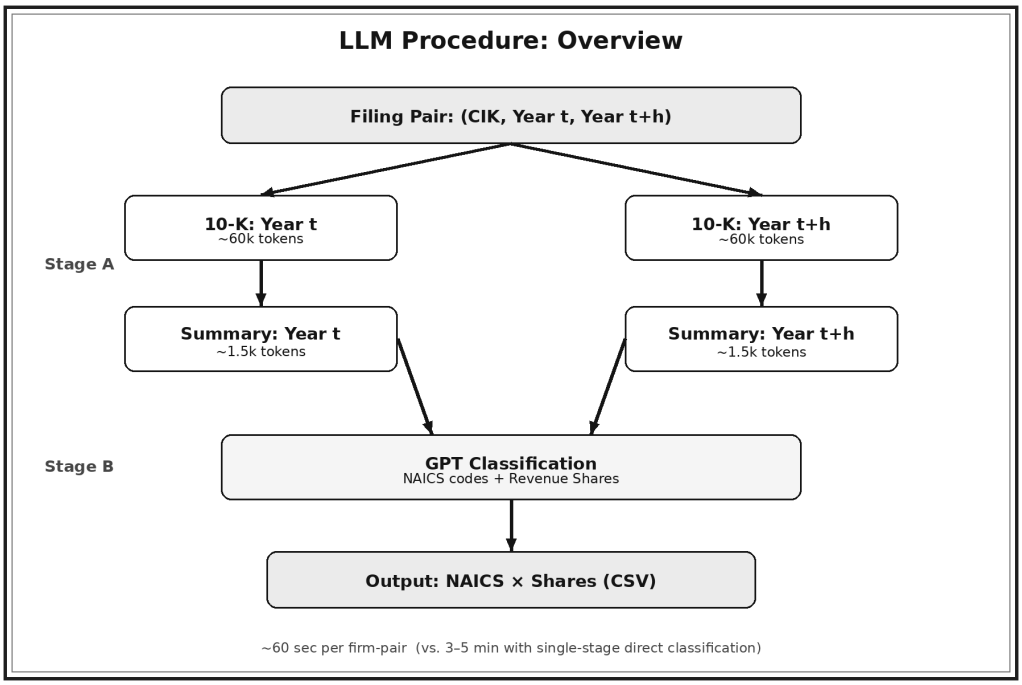


Figure 1: Two-stage LLM pipeline used to construct firm-sector revenue shares from 10-K filings

Stage A Prompt

Analyze this 10-K filing and extract ONLY revenue-relevant information for NAICS classification.

OUTPUT FORMAT - Provide a structured summary with these sections:

1. BUSINESS SEGMENTS: List each business segment/division with:

- Segment name; Products/services offered
- Estimated revenue share (% of total) if mentioned

2. REVENUE SOURCES: List primary revenue streams:

- Product sales vs services; Geographic breakdown if available
- Customer types (B2B, B2C, government, etc.)

3. INDUSTRY INDICATORS: Key phrases indicating industry:

- Manufacturing activities; Retail/wholesale operations
- Technology/software services; Financial services; Healthcare activities

4. CHANGES NOTED: Any mentions of:

- New business lines; Discontinued operations; Acquisitions or divestitures

Keep the summary concise (under 2000 words). Focus on facts that help determine 4-digit NAICS codes.

The sections used are those most likely to contain economically meaningful descriptions of the firm's operating activities: Section 1 (Business) describes the firm's main products and markets, Section 1A (Risk Factors) discusses key risks facing the firm; Section 1B (Unresolved Staff Comments) contains regulatory commentary when relevant; and Section 7 (Management's Discussion and Analysis) provides management's perspective on operations and performance. Together, these sections provide rich descriptions of the firm's economic activities across sectors. The output is a concise structured summary, that preserves the parts of the filing most informative for mapping the firm into 4-digit NAICS industries. Because we can store the descriptions, this stage allows others to use these summaries to replicate the analysis, despite changing LLM models and outputs. The first stage also reduces noise by focusing the model's attention on text that contains specific information about revenue-generating activities and business segments, avoiding the inclusion of unrelated industrial codes.

In the second stage, we pass these summaries to GPT-5.2, to obtain approximate revenue shares. We pass two filing summaries for the same firm at a five-year horizon ($h = 5$) and ask the model to produce a structured CSV output containing all active 4-digit NAICS codes and the corresponding revenue shares in the base and final years. The prompt requires the model to list all NAICS codes appearing in either year, assign a revenue share to each code in both periods, using zeros when a code is absent in one year. Stage B prompt reads as follows:

Stage B Prompt

Based on these revenue summaries from two years, identify the 4-digit NAICS codes the company operates in. REQUIREMENTS:

1. List all NAICS codes from both years combined
2. For each code, provide revenue share (%) for Year 1 (Base) and Year 2 (Final)
3. If a code wasn't present in a year, use 0 for that year's share
4. CRITICAL: Base Share column MUST sum to exactly 100
5. CRITICAL: Final Share column MUST sum to exactly 100

Output format (CSV):
NAICS Code, Base Share, Final Share
3361, 50, 10
5413, 45, 40
2389, 0, 30
3399, 5, 20
TOTAL, 100, 100
Include the TOTAL row to verify your sums equal 100.

By passing both years jointly, we force the model to reason about the evolution of the firm's business mix across filings. This paired-year setup helps the algorithm distinguish true business-line expansion, contraction, entry, and exit from random variation in wording across annual reports. We also ask the model to ensure that the shares in each year sum to exactly 100. This serves as a simple validation exercise. The output is accepted only if the revenue shares for each year sum to 100 up to a small tolerance of 5%. Failed instances are reprocessed with stronger instructions. The final output has the following form: NAICS code, base share, final share. The procedure yields a mapping from business line descriptions in firm 10-Ks directly to 4-digit NAICS sectors. The resulting revenue shares can be integrated with Bureau of Economic Analysis (BEA) data that reveals sectoral distances.

We also conduct a few human audits for selected firms by comparing assigned NAICS codes and shares to validating textual evidence in the underlying 10-Ks, including segment disclosures when available. These audits are not intended to claim that the LLM recovers true revenue shares without error. Rather, they verify that the estimated shares are grounded in recognizable business descriptions and that the inferred changes over time align with well-known shifts in firms' product portfolios. Table 1 shows Amazon's business line over

Table 1: Amazon NAICS Assignment and Revenue Share Validation

Panel A. Amazon 2000		
NAICS sector	**Share**	**Validating Evidence (Human Audit)**
Online retail (NAICS 4541)	98%	**Sem:** "leading online retailer"; "online retail stores ... books, music, DVD/video". **Int:** Retail is presented as the core business, with the longest and richest product discussion, implying a dominant share.
Online marketplaces (NAICS 5191)	2%	**Sem:** "Web-based marketplaces ... transactions"; "Auctions, zShops". **Int:** Positioned as a new add-on with limited detail, implying a small share.
Panel B. Amazon 2020		
NAICS sector	**Share**	**Validating Evidence (Human Audit)**
Online retail (NAICS 4541)	57%	**Sem:** "online and physical stores"; "hundreds of millions of unique products". **Int:** Retail and fulfillment are framed as the core consumer business, implying the largest share.
Marketplace / nonstore retail (NAICS 4543)	20%	**Sem:** "sellers... sell their products in our stores"; "not the seller of record... earn fees". **Int:** Seller programs are sizeable, but clearly secondary to first-party retail.
Data processing / cloud (NAICS 5182)	12%	**Sem:** "AWS segment... global compute, storage, database". **Int:** Segment with repeated emphasis, implying a large services share.
Grocery / physical stores (NAICS 4451)	6%	**Sem:** "Whole Foods Market... included... from acquisition"; "physical stores". **Int:** Physical grocery is present with moderate discussion, implying a modest share.

Panel A note: The 2000 classification is primarily retail and is supported by actual segment revenue disclosure in the 10-K: U.S. retail contributes $1.3 billion out of $1.6 billion in overall revenue.
Panel B note: The 2020 classification is supported by disclosed segment information: AWS accounts for about 12% of net sales, and Whole Foods enters the filing after the 2017 acquisition.

the last two decades as captured by the LLM. In 2000, its 10-K describes a firm that is overwhelmingly an online retailer, and the model correspondingly assigns almost all revenue to online retail. By 2020, the filing emphasizes not only retail, but also third-party seller services, AWS cloud computing, physical grocery operations, and advertising. The model translates these changes in narrative prominence into a more diversified NAICS revenue vector, with retail remaining largest but substantial shares emerging in cloud, marketplace services, grocery, and advertising.

2.3 Measuring Firm Scope

Our measure of scope is grounded in the model's notion of risk exposure. Recall that firm i's risk from its operations in sector j depends on how revenues in that sector comove with the firm's total revenues. This logic naturally delivers a risk-based measure of scope: firms are broader in scope when their revenue composition provides diversification benefits and narrower in scope when their activities are concentrated in correlated sectors.

Empirically, we combine two ingredients: (i) firm-sector revenue shares we obtain from our LLM procedure and (ii) sectoral correlations. In the model, (13) described scope as a function of expected revenue shares R. What we measure are realized revenues. Let $\frac{Rev_{ikt}}{Rev_{it}}$ denote the relized share of firm i's revenues in sector k in period t. Adapting the function for realized revenue, the scope of firm i at time t is:

$$Scope_{it} = \sum_{j,k} \frac{Rev_{ijt}}{Rev_{it}} \frac{Rev_{ikt}}{Rev_{it}} (1 - \rho_{jk}). \tag{18}$$

Recall from equation (12) that $RiskExp_i = 1 - Scope_i$. While risk exposure is the object that enters the markup equation (11), we find scope more useful for building intuition and for connecting our measure to the broader literature on diversification.

The scope measure has a clean interpretation as a form of *Rao's Quadratic Entropy*. Quadratic entropy is defined as $\sum_{j,k} \frac{Rev_{ijt}}{Rev_{it}} \frac{Rev_{ikt}}{Rev_{it}} d_{jk}$, where d_{jk} is a measure distance between sectors j and k. It measures the expected distance between two randomly drawn revenue dollars from the firm's portfolio. When those two dollars come from distant sectors, the firm is diverse; when they come from nearby sectors, it is concentrated. Our theory pins down the relevant distance metric: to measure exposure to sectoral risk, the appropriate distance is $d_{jk} = 1 - \rho_{jk}$, the complement of the demand-shock correlation between sectors j and k. Two sectors that experience identical shocks ($\rho_{jk} = 1$) are at zero distance from a risk perspective, while two sectors whose shocks are uncorrelated ($\rho_{jk} = 0$) are maximally distant.

This formulation nests several familiar measures as special cases under alternative distance metrics. If one replaces d_{jk} with a discrete metric ($d_{jk} = 1$ for $j \neq k$, $d_{jj} = 0$), the expression reduces to $1 - \text{HHI}$, the standard revenue-based Herfindahl index complement. A Hamming-style metric based on the hierarchical NAICS coding structure—where distance increases as sectors diverge at higher digits—would instead capture product-space proximity in the spirit of the classification system itself. Our correlation-based distance is the one that the theory selects as relevant for risk: it measures the degree to which operating across sectors actually provides hedging value.

Connection to other scope measures Hoberg and Phillips (2025) use word counts from 10-K filings to index a firm's industry exposure as the fraction of an industry dictionary in the firm's filing. While word counts may capture features of a firm's operations, it is difficult to reliably map them to revenue shares or revenue diversification, the key object of interest in this paper. Input-output measures characterize a production structure rather than revenue shares. Our measure captures diversification, which requires going beyond counting the number of business lines. Two firms with the same number of sectors can have very different scope if one operates in tightly related industries and the other in unrelated ones. A firm active in many sectors that move together with the same shocks has limited effective scope, while a firm active in sectors that respond to different shocks has broad effective scope.

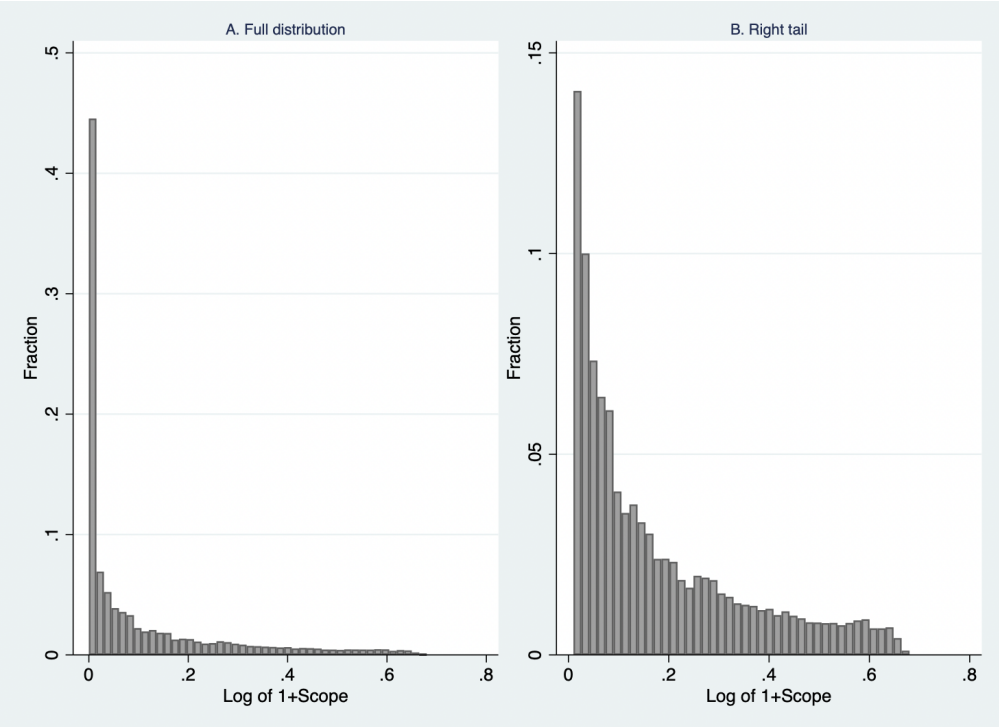


Figure 2: Histogram of Firm Scope Measure

Notes: Panel A shows the full distribution of $\log(1 + Scope)$. *Panel B excludes observations with* $\log(1 + Scope) \leq 0.01$ *in order to zoom in on the right tail of the distribution.*

Figure 2 depicts a histogram of our scope measure.[5] It shows a significant mass at zero (single-sector firms) and a long right tail of highly diversified conglomerates. Table 2 summarizes the distribution of $\ln(1 + Scope)$ in our sample.

[5]For our empirical analysis, we work with a transformation, $\ln(1 + Scope)$.

Table 2: Summary Statistics (Firm-Level)

Variable	N	Mean	Std. Dev.	P25	P75
Scope	32,689	0.13	0.20	0	0.16
Log Sales	32,693	5.563	2.24	4.06	7.18
Markup	22,470	1.86	1.42	1.13	1.95

Notes: Number of observations: $33,962$ firm-years.

Furthermore, Table 3 sheds light on the distribution of activities within an average firm in our sample. An average firm operates in about 2.5 sectors with revenues mostly concentrated in the top sector. The top sector contributes 83% of total revenues. Conditioning on multi-sector firms, an average multi-sector firm remains heavily concentrated, with the top sector contributing 75% of its revenues.

Table 3: Summary Statistics (Firm-level Revenue Shares)

	Mean	SD	P25	P75
$Scope_{ij}$	0.25	0.31	0.00	0.45
Top-1 sector share	0.83	0.20	0.70	1.00
Top-2 sector share	0.18	0.14	0.06	0.29
Top-3 sector share	0.09	0.07	0.03	0.13
No. of sectors	2.41	1.61	1.00	3.00

Finally, we also report statistics on how firms are distributed within a sector in Table 4. An average sector in our sample has about 19 firms. The revenues in the sector are concentrated, with an HHI of 0.42.

Table 4: Summary Statistics (Sectoral)

	Mean	SD	P25	P75
No. firms	18.6	30.1	5.0	18.0
Top firm share	0.54	0.22	0.37	0.72
HHI	0.42	0.23	0.22	0.56

Notes: Number of obs: $4,385$ sector-years. A sector-year with < 3 firms is mapped to the closest dense sector in distance and the sectoral average on the combined sector (with firms in both the sectors) are assigned to both the sectors.

3 Three Facts about Firm Scope

In this section, we document three novel facts based on our measure of firm scope. First, we show that our risk-based scope predicts firm exit, providing a degree of external validation. Next, we quantify the gains from specialization by estimating how productivity decays as firms stray from their core competency, or home sector. Third, we demonstrate a systematic relationship between firm-level markups and scope. This reveals the risk price: how much profit firms are willing to forgo to mitigate risk.

Fact 1: Scope Reduces Risk

One might reasonably doubt that applying a language model to firm 10-Ks would deliver sensible revenue shares and scope measures. So, our first fact also serves as a form of validation. In Table 5, we run a probit regression where the dependent variable equals 1 if the firm exits within the next 5 years (columns 1–2) or 10 years (columns 3–4) and 0 otherwise. Exit is defined as a firm permanently leaving the Compustat sample. A firm that last appears in Compustat at t is said to exit at t. The independent variables are standardized values of $\ln(1 + Scope)$ and controls. The results show that the scope has a statistically and economically meaningful association with exit. A 1-standard deviation change in $\ln(1 + Scope)$ is associated with a 5% reduction in the probability of exit over the next 5-10 years. This pattern is consistent with the hypothesis at the core of this paper: expanding scope reduces risk.

Table 5: Scope and Exit.

	P(Exit in 5yrs)	P(Exit in 5yrs)	P(Exit in 10yrs)	P(Exit in 10yrs)
log(1+scope)	-0.0508***	-0.0505***	-0.0589***	-0.0625***
	(-6.36)	(-5.84)	(-6.90)	(-6.72)
log inf adj. Rev	-0.2476***	-0.2559***	-0.2904***	-0.3168***
	(-30.07)	(-27.55)	(-32.03)	(-28.74)
Observations	26758	19976	23826	17838
Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

t statistics in parentheses

Notes: The dependent variable is an indicator equal to one if the firm exits within the next 5 years (columns 1–2) or 10 years (columns 3–4). All independent variables are standardized as $(x - \bar{x})/\sigma_x$. Additional controls (included where indicated) are R&D intensity, operating profits, and net leverage. All specifications include year fixed effects. Standard errors are robust.

Fact 2: Markups Decline with Scope

Next, we demonstrate a key empirical pattern: markups at the firm-level decline as the the firm expands its scope. This validates a key implication of the theory: that firms charge higher markups as compensation for risk. This exercise also allows us to pin down a crucial parameter: the price of risk.

We exploit the structural relationship between firm-level markups and our risk-based measure of scope: equations (11)-(12). In principle, with a sufficiently long panel, one could estimate this relationship at the firm-level. Given the data limitations, however, we have to impose more structure. We assume the price of risk is scale-dependent, so that $\gamma_{it} R_{it} = \hat{\gamma}$ is constant across firms and time. To estimate $\hat{\gamma}$, we run the following regression:

$$MKP_{it} = \beta_0 + \beta_1 \ln(1 + Scope_{it}) + \beta_2 \ln Rev_{it} + \varphi_i + \epsilon_{it} \ , \tag{19}$$

where MKP_{it} is the markup estimated from the data (as described above), Rev_{it} is firm's sales and φ_i is a fixed effect.[6] Table 6 presents the results. The coefficient of interest, β_1, is negative and significant. In other words, firm-level markups decline with scope, consistent with the theory.

Table 6: Scope and Markups.

	(1)	(2)
log(1+scope)	-0.3283***	-0.3063***
	(-6.01)	(-5.66)
Log of inflation adj. revenue	0.0395***	0.0013
	(4.62)	(0.14)
Observations	21603	21470
R^2	0.808	0.813
Firm FE	Yes	Yes
Controls	No	Yes

t-statistics in parentheses

Notes: The dependent variable is firm level measured mark-up calculated using the production function approach as in De Locker et al. (2020). Key regressors include log(1 + scope) and log inflation-adjusted revenues. Additional controls (included where indicated) are R&D intensity, and net leverage. All specifications include firm fixed effects. Standard errors are robust.

To recover the price of risk, $RiskPr_i$, we note that (11) implies

$$\frac{R_i}{\sum_j MC_{ij} Q_{ij}} \approx \frac{\eta}{\eta - 1}(1 + RiskPr_i) - \frac{\eta}{\eta - 1} \ RiskPr_i \ Scope_i \ . \tag{20}$$

[6] We include $\ln Rev_{it}$ as a control to account for any size-dependence in markups.

Approximating $\ln(1 + Scope_i) \approx \frac{1}{1+\overline{Scope}}(Scope_i - \overline{Scope})$ around the average scope, $\overline{Scope}$, we obtain

$$RiskPr_i = \frac{-\beta_1}{1 + \overline{Scope}} \frac{\eta - 1}{\eta} . \tag{21}$$

We follow the firm dynamics literature and set $\eta = 5$ – see, for example, Atkeson and Burstein (2010). Plugging in the estimates from Table 6 (along with an average scope of 0.13 and $\eta = 5$) yields $RiskPr_i = 0.23$. This implies that a single sector firm (with $Scope_i = 0$) charges a markup that is $\frac{1}{1-0.23}$ times the benchmark without risk (or equivalently, the markup of a fully diversified firm). Under our calibration with $\eta = 5$, this translates to a markup of 1.62 (the benchmark without risk is 1.25). At our estimated average scope of 0.13, the markup is $\frac{1}{1-(0.23)(1-0.13)} \frac{\eta}{\eta-1} = 1.56$.

Fact 3: Productivity Declines with Distance

To quantify the gains to specialization, we measure the change in a firm's total factor productivity as it moves away from its core competency, defined as the 4-digit NAICS industry in which the firm achieves its maximum measured productivity (A_{ijt}). This home sector, denoted j_i^*, serves as the benchmark: the distance of a given sector k from firm i's core competency is defined as $1 - \rho_{kj_i^*}$.

The estimation proceeds in two steps. First, we non-parametrically estimate a hazard function to determine the probability that a firm operates in a given sector as a function of the distance, $1 - \rho_{kj_i^*}$. Second, we recover firm-sector productivity (A_{ijt}) from sectoral market shares, price of risk (estimated above) and $RiskExp_{ij}$ (measured as revenue weighted correlation of sector j to all sectors firm i operates in), leveraging the insight in Equation (16). Sectoral market shares are computed within the sample as firm's revenue in a sector relative to the average revenue of firms operating in that sector.[7]

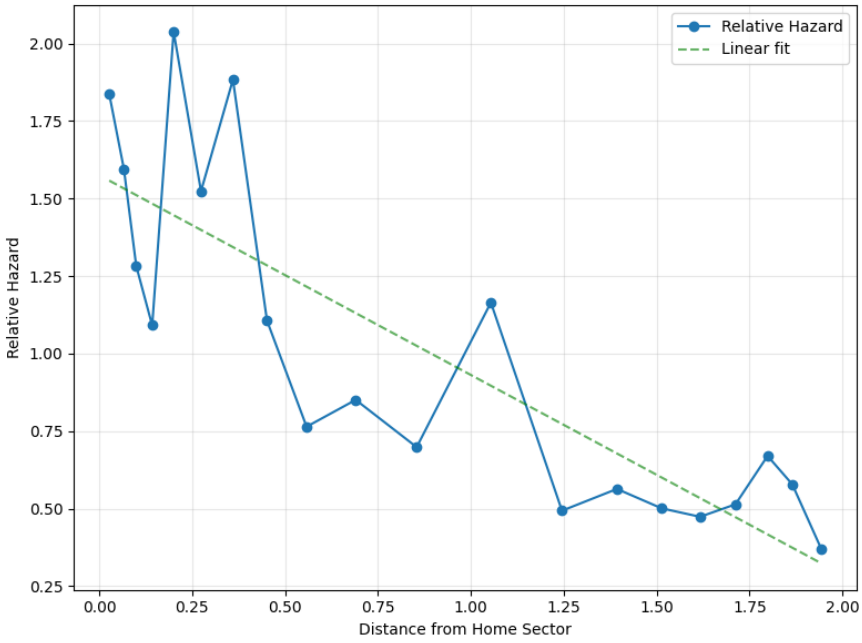
Figure 3 displays the estimated hazard: it shows the average probability of a firm operating in a sector for 20 quantile bins of the distance $1 - \rho_{kj_i^*}$, along with a fitted line:

$$h(\rho_{kj^*}) = 1.59 - 0.66(1 - \rho_{kj^*}) = 0.93 + 0.66\ \rho_{kj^*} , \tag{22}$$

where $h(\rho_{jk})$ denotes the likelihood of operating in sector $k \neq j^*$. The estimates imply that the probability declines sharply as the correlation with the home sector falls: a firm is 4.5 times more likely to operate in a sector whose correlation with the home sector is 0.9

[7]The model assumes symmetry across sectors. In our empirical implementation, we remove a sector-by-year fixed effect from the estimated A_{ijt} to control for heterogeneity across sectors.

Figure 3: Firms Less Likely to be in Sectors Far from Home



Notes: The figure plots the fraction of firm-years operating in a sector at a given distance from the home sector, using quantile bins of sectoral distance. Relative hazard rescales this fraction by its average across bins.

compared to a sector with correlation -0.9. This extensive margin pattern is inconsistent with theories of firm scope where new products arrive randomly.

Figure 4 depicts the gains to specialization: it plots the average productivity of non-home sector (relative to home sector TFP) against distance $1 - \rho_{kj_i^*}$ (in quantile bins). The estimted line of best fit implies that the TFP of firm i in a sector $k \neq j_i^*$

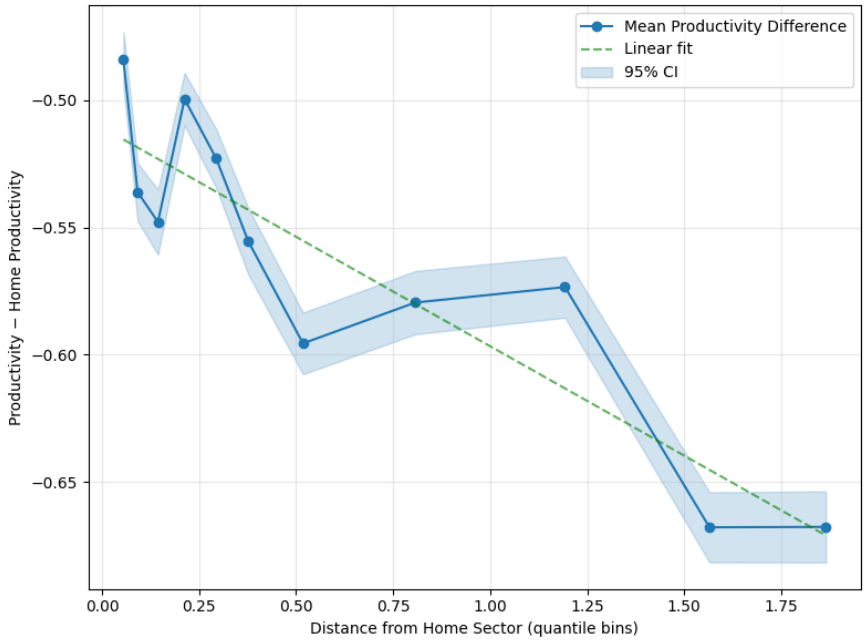
$$\ln A_{ik} = -0.52 - 0.1\ (1 - \rho_{kj_i^*}) = -0.62 + 0.1\ \rho_{kj_i^*} \quad . \tag{23}$$

Productivity declines rapidly as a firm moves away from its core competency: the penalty for a sector close to the home sector is about 50 log points, which increases to 70 log points for sectors that are furthest away. This TFP decay implies that diversifying into distant sectors may provide risk-reduction benefits, but at a significant efficiency cost.

4 Quantification: The Cost of Uncertainty

Does the diversification vs. specialization trade-off matter for the macroeconomy? We use the estimates so far and the model structure to quantify the output cost of firms' diversifi-

Figure 4: Gains to Specialization: Distance from Home Sector Reduces Productivity



Notes: The figure plots the log difference in productivity relative to the firm-year home sector against distance from the home sector, using quantile bins of sectoral distance. Productivity is residualized to remove sector-year fixed effects. The home sector is the sector with the highest implied productivity within the firm-year.

cation. This is not a welfare cost: Firms optimally balance foregone output and the benefits of risk reduction. To quantify aggregate output losses from risk, we conduct a simple a Lucas (1987) style exercise: we compare the aggregate output of our baseline economy to a counterfactual economy where the price of risk parameter is set to zero ($\gamma = 0$).

Using the estimates of productivity and risk wedges from the previous section we construct the following objects for each sector. The loss due to risk, denoted $\mathcal{L}^{Risk}$, is defined as the

$$\mathcal{L}^{Risk} = \ln Q_j^* - \ln Q_j = \frac{1}{\eta - 1}\left[\ln\left(\int (1-\tau_{ij})^{\eta-1} A_{ij}^{\eta-1}\, di\right) - \ln\left(\int A_{ij}^{\eta-1}\, di\right)\right] \tag{24}$$

where the second equality comes from substituting in (15)

The estimated loss due to risk, $\mathcal{L}^{Risk}$, (averaged across sectors, weighted by the number of firms in the sector) is substantial, at 19.31%. Of this, the loss due to the average tax, $\mathcal{L}^{Avg}$, accounts for about two-thirds (12.95%), while the rest can be attributed to the covariance between the risk wedge and productivity.

Figure 5: Aggregate Output Loss from Firm Scope

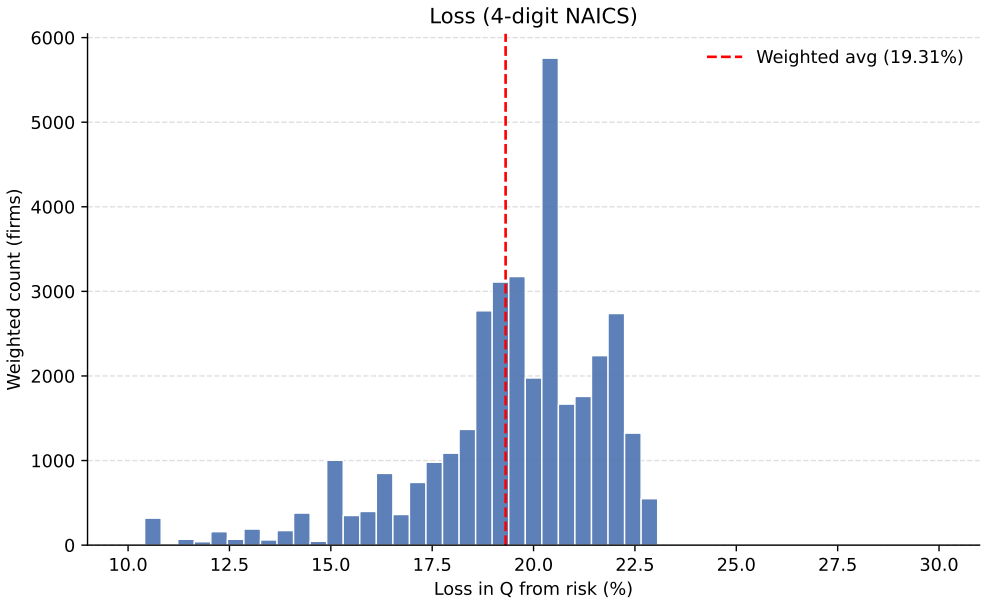


Figure 5 presents a weighted histogram of the estimated output losses by sector (the weights being the number of firms in the sector). It shows large losses across the board, ranging from 10% at the low end to 22%. Figure

To understand where this result comes from, we define another useful benchmark: what output would be if the wedge $1-\tau_{ij}$ was independent of productivity, A_{ij}:

$$Q_j^{Ind} = \frac{\eta-1}{\eta}\left(N_j^{\varpi}\int(1-\tau_{ij})^{\eta-1}di \quad N_j^{\varpi}\int A_{ij}^{\eta-1}\,di\right)^{\frac{1}{\eta-1}}. \tag{25}$$

We can then decompose the total loss into two components: one stemming from the average wedge and the other attributable to the covariance between the wedge and productivity:

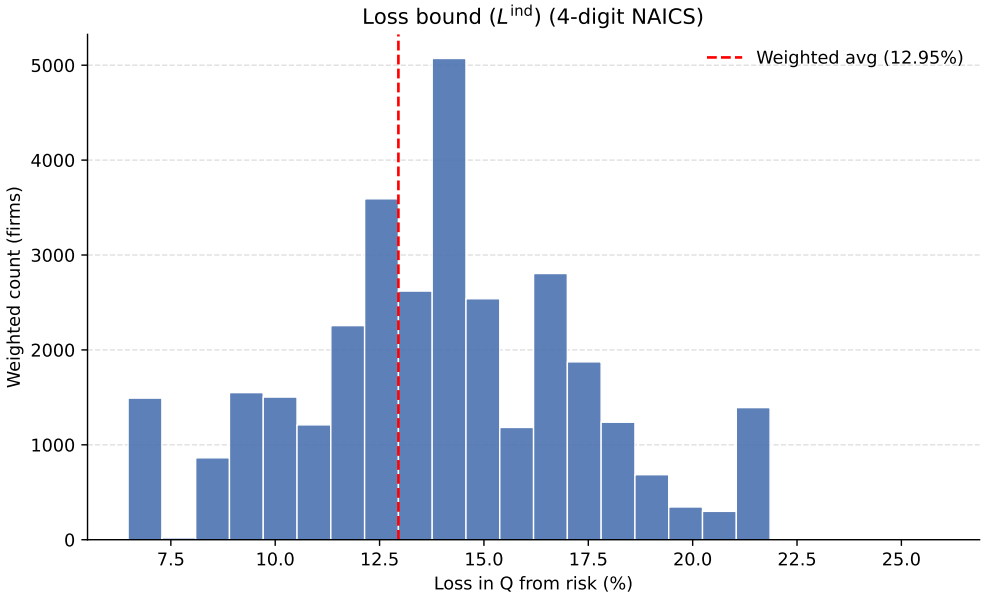
$$\mathcal{L}^{Risk} = \underbrace{\ln Q_j^* - \ln Q_j^{Ind}}_{\text{Average Wedge Loss } \equiv\ \mathcal{L}^{Avg}} + \underbrace{\ln Q_j^{Ind} - \ln Q_j}_{\text{Covariance Loss } \equiv\ \mathcal{L}^{Cov}} \tag{26}$$

The first term, $\mathcal{L}^{Avg}$, is equal to $\frac{1}{\eta-1}\ln\int(1-\tau_{ij})^{\eta-1}di$: the reduction in output due to the average level of risk compensation among firms operating in the sector. The second term arises when the risk wedge varies systematically with productivity. When more productive firms require higher risk compensation, this covariance is positive and further depresses output. In contrast, if highly productive firms are better diversified and face lower τ_{ij}, the covariance term can mitigate aggregate losses. Thus, output depends not only on the distri-

bution of productivities but also on the distribution of risk compensation and its correlation with productivity.[8]

Figure 6 depicts a histogram of the loss attributable to the average wedge, $\mathcal{L}^{Avg}$. One could think of this as a conservative estimate of losses, given that the covariance is likely to be estimated with noise.

Figure 6: Bound on Aggregate Output Loss from Firm Scope



The difference between the two histograms also reveals the covariance of firm-level productivities with the losses from risk. The lower levels of loss on the x-axis of Figure 6 relative to Figure 5 imply that covariance adds substantially to the loss. Specifically, productivity decreases with the potential for diversification. This means that the high productivity firms in a sector will typically demand a higher risk compensation, driving up the output losses from risk.

In summary, these results suggest that the 'distance technology'—the ease with which a firm can span disparate industries—explains a substantial share of the markup dispersion and firm-size inequality that traditional macro models often attribute purely to productivity shocks.

[8]Another way to see this decomposition is to approximate (14) around the no-risk benchmark:

$$\ln Q_j \approx \ln Q_j^* - \left[\overline{\tau_{ij}} + \text{Cov}\left(\tau_{ij}, \frac{A_{ij}^{\eta-1}}{\int A_{ij}^{\eta-1} di}\right)\right] \qquad \text{where} \qquad \overline{\tau_{ij}} = \int \tau_{ij} di.$$

4.1 Understanding the Cost of Risk: An Example

Finally, we use a four-sector example to demonstrate how the productivity advantage of a firm's home sector influences the aggregate cost of risk. We show that when the productivity penalty for operating in distant sectors is higher, firms are less willing to diversify. While firms are most productive in this home sector, the risk concentration poses makes the produce less aggressively, leading to aggregate output losses.

To see the connection between productivity patterns and the effect of risk, consider the following special case. There are four sectors, whose demand shocks are correlated as follows:

$$\Omega = \begin{bmatrix} r_0 & r_1 & r_2 & r_3 \\ r_1 & r_0 & r_3 & r_2 \\ r_2 & r_3 & r_0 & r_1 \\ r_3 & r_2 & r_1 & r_0 \end{bmatrix} ,$$

where, without loss of generality, the sectors are ordered so that $1 = r_0 > r_1 > r_2 > r_3$. A continuum of firms operate in all 4 sectors. The within-sector elasticity of substitution across firms, η, is set to 2. Each firm has a (randomly drawn)'home' sector, where its productivity is highest. Firm i, in its home sector (denoted j_i^*), has a productivity normalized to 1, i.e. $\ln A_{ij_i^*} = 0$. Its productivity in the other sectors is:

$$\ln A_{ik} = -\xi_0 - \xi_1(1 - \rho_{kj_i^*}) \qquad \forall k \neq j_i^* ,$$

where the parameters $\xi_0, \xi_1 > 0$ index the intercept and slope of the productivity penalty, i.e. the decline in productivity as a function of the distance $(1 - \rho_{kj_i^*})$ from the home sector ($\rho_{kj_i^*}$ is the correlation of sector k with the home sector of the firm). Since $\xi_1 > 0$, the penalty is higher for less correlated sectors.

Given the symmetry, the set of risk wedges is the same for every firm. With a slight abuse of notation, we denote them as $\{\tau_h\}, h = 0, 1, 2, 3$, where $\tau_{ij} = \tau_h$ if $\rho_{jj_i^*} = r_h$. In matrix form,

$$\boldsymbol{\tau} = RiskPr \cdot \boldsymbol{\Omega}\ \boldsymbol{\omega} . \tag{27}$$

where $\boldsymbol{\tau} := [\tau_0 \ \ \tau_1 \ \ \tau_2 \ \ \tau_3]'$, $\boldsymbol{\omega} := [\omega_0 \ \ \omega_1 \ \ \omega_2 \ \ \omega_3]'$ with ω_h denoting the revenue share of the sector whose correlation with the home sector is r_h. For analytical tractability, we approximate the weights $\{\omega_h\}$ using the revenue shares at the no-risk limit. When $\gamma = 0$,

the revenue shares of the sector are given by

$$\omega_0 = \frac{R_{ij_i^*}}{\sum_k R_{ik}} = \frac{A_{ij_i^*}}{\sum_k A_{ik}} = \frac{1}{1+\sum_k e^{-\xi_0-\xi_1(1-r_k)}} \tag{28}$$

$$\omega_h = \frac{e^{-\xi_0-\xi_1(1-r_h)}}{1+\sum_k e^{-\xi_0-\xi_1(1-r_k)}} \qquad h=1,2,3\ . \tag{29}$$

The loss from risk – the difference between output without risk and equilibrium output, as defined in (24) – is:

$$\mathcal{L}^{Risk} = \ln(1-\boldsymbol{\omega}'\boldsymbol{\tau})\ . \tag{30}$$

The next result shows that the losses from risk are more severe when the productivity advantage of the firm's home sector is greater.

Lemma 1. (Gains to Specialization and the Output Cost of Risk)

$$\frac{d}{d\xi_0}\exp(\mathcal{L}^{Risk}) = 2\omega_0(\boldsymbol{\omega}'\boldsymbol{\tau}-\tau_0) \quad < \quad 0\ , \tag{31}$$

$$\frac{d}{d\xi_1}\exp(\mathcal{L}^{Risk}) = 2Cov(1-\rho_j,\tau_j) \quad < \quad 0\ . \tag{32}$$

Equation (31) says that the effect of the intercept parameter ξ_0 depends on the difference between the (weighted) average τ_j and that of the home sector. Intuitively, a larger penalty shifts the firm's portfolio away from non-home sectors and towards the riskiest home sector, increasing the losses. Analogously, from equation (32), the effect of the slope parameter ξ_1 depends on the covariance between the distance from the home sector $1-\rho_j$ and the wedge τ. When productivity declines more sharply with distance, incentives to diversify fall, shifting the composition of the firm's portfolio towards sectors closer to home (i.e. riskier). This force exerts a more negative effect on aggregate output when this shift is associated with a larger increase in τ.

5 Conclusion

This paper unifies three historically disparate perspectives on firm scope by modeling the trade-off between specialization and diversification in firm product choice. While the finance view emphasizes diversification as a tool for risk management and the management literature champions core competency for efficiency, the macro-IO perspective traditionally focuses on a count of products or sectors. By introducing a measurable construct of product distance

and revenue covariance, we unite these perspectives and provide a theory-guided framework to measure the cost of uncertainty and the resulting scope expansion. Our calibrated model reveals that risk induces firms to broaden their scope to mitigate volatility, compromising their gains to specialization, resulting in a loss of roughly 19.31% of aggregate output.

Looking forward, several extensions could be fruitful. While our current framework is static, incorporating dynamic models would allow for the study of how scope evolves over the firm lifecycle, from the rapid product switching of young startups to the entrenched breadth of superstar firms. Future research might also integrate fixed costs of expansion, particularly as information technology shifts the barriers to industry entry, to explore how these costs influence the wave of specialization seen in R&D-intensive firms (Ma, 2023). Finally, explicitly modeling the relationship between diversification and exit through bankruptcy could offer a more granular view of how risk-mitigation strategies impact firm survival and the resulting distribution of firm size and markups in the aggregate economy.

References

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr, "Innovation, Reallocation, and Growth," *American Economic Review*, November 2018, *108* (11), 3450–91.

Akcigit, Ufuk, Murat Alp Celik, and Jeremy Greenwood, "Buy, Keep or Sell: Economic Growth and the Market for Ideas," Working Paper 19763, National Bureau of Economic Research December 2013.

Atkeson, Andrew and Ariel Tomás Burstein, "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, June 2010, *118* (3), 433–484.

Autor, David, David Dorn, Lawrence F. Katz, and Christina Patterson, "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 2020, *135* (4), 1482–1541.

Ayyagari, Meghana, Asli Demirgüç-Kunt, and Vojislav Maksimovic, "The Rise of Star Firms: Intangible Capital and Competition," *The Review of Financial Studies*, 08 2023, p. hhad065.

Berlingieri, Giuseppe, Maarten De Ridder, Danial Lashkari, and Davide Rigo, "Creative Destruction through Innovation Bursts," Discussion Paper dp2095, Centre for Economic Performance, London School of Economics 2025. Revise-Resubmit, Econometrica.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, "Multiple-Product Firms and Product Switching," *American Economic Review*, March 2010, *100* (1), 70–97.

Boar, Corina and Virgiliu Midrigan, "Markups and Misallocation," *Working Paper*, 2023. Cited for the finding that the majority of cross-sectional markup dispersion stems from risk.

Bresnahan, Timothy, Shane Greenstein, and Rebecca Henderson, "Idiosyncratic knowledge and the diffusion of managed care," *Strategic Management Journal*, 2011. Cited as evidence that firms expand successfully into proximate areas.

Cai, Jie and Nan Li, "Growth Through Inter-sectoral Knowledge Linkages," *The Review of Economic Studies*, 2019, *86* (5), 1827–1866.

De Ridder, Maarten, "Market Power and Innovation in the Intangible Economy," Working Paper 1931, Cambridge 2021.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Eberly, Janice C., Nicolas Crouzet, Andrea Eisfeldt, and Dimitris Papanikolaou, "Intangible Capital, Non-rivalry, and Growth," 2023. Working Paper.

Eckbo, B. Espen, ed., *Handbook of Empirical Corporate Finance Set*, Amsterdam: Elsevier, 2008.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "How Much Are We Losing from Productivity Dispersion?," *Journal of Political Economy*, 2023. Cited to compare the contribution of scope to markup dispersion against productivity differences.

Eisfeldt, Andrea L and Dimitris Papanikolaou, "The value and ownership of intangible capital," *American Economic Review*, 2014, *104* (5), 189–194.

Hoberg, Gerard and Gordon M Phillips, "Scope, Scale and Competition: The 21st Century Firm," *The Journal of Finance*, 2025, *forthcoming* (n/a).

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

Jensen, Michael C. and William H. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 1976, *3* (4), 305–360.

Kim, J, K Lee, and H Lim, "Firm Technological Proximity and the Direction of Net Knowledge Spillovers," *Working Paper*, 2023. Cited for the use of transformer-based embeddings to study technological proximity.

Klette, Tor Jakob and Samuel Kortum, "Innovating firms and aggregate innovation," *Journal of political economy*, 2004, *112* (5), 986–1018.

Kumar, Saten, Yuriy Gorodnichenko, and Olivier Coibion, "The Effect of Macroeconomic Uncertainty on Firm Decisions," *Econometrica*, 2023, *91(4).*

Lewellen, Wilbur G., "A Pure Financial Rationale for the Conglomerate Merger," *Journal of Finance*, 1971, *26* (2), 521–537.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 2020. Benchmark for markup dispersion trends and aggregate concentration.

Ma, Yueran, Mengdi Zhang, and Kaspar Zimmermann, "Business Concentration around the World: 1900–2020," 2024. Working Paper.

Ma, Yueyuan, "Specialization in a Knowledge Economy," 2023. Working paper.

Maksimovic, Vojislav and Gordon Phillips, "Do Conglomerate Firms Allocate Resources Inefficiently Across Industries? Theory and Evidence," *Journal of Finance*, 2002, *57* (2), 721–767.

Matvos, Gregor and Amit Seru, "Resource Allocation within Firms and Financial Constraints," *The Review of Financial Studies*, 2014. Discussed in the context of internal capital markets and diversification performance.

Mayer, Thierry, Marc J Melitz, and Gianmarco IP Ottaviano, "Product Mix and Firm Productivity Responses to Trade Competition," *Review of Economic Studies*, 2016. Cited for the macro/IO view of product scope and markups.

Myers, Stewart C. and Nicholas S. Majluf, "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have," *Journal of Financial Economics*, 1984, *13* (2), 187–221.

Pellegrino, Bruno, "Product Differentiation and Oligopoly: A Network Approach," *American Economic Review*, 2025, *115* (4), 1170–1225.

Prahalad, Coimbatore Krishnarao and Gary Hamel, "The Core Competence of the Corporation," *Harvard Business Review*, 1990, *68*, 275–292.

Rumelt, Richard P., *Strategy, Structure, and Economic Performance*, Cambridge, MA: Harvard University Press, 1974.

Stein, Jeremy C., "Internal Capital Markets and the Competition for Corporate Resources," *Journal of Finance*, 1997, *52* (1), 111–133.

_ , "Agency, Information and Corporate Investment," in George M. Constantinides, Milton Harris, and René M. Stulz, eds., *Handbook of the Economics of Finance*, Vol. 1 of *Handbook of the Economics of Finance*, Elsevier, 2003, pp. 111–165.

Teece, David J, Richard Rumelt, Giovanni Dosi, and Sidney Winter, "Understanding Corporate Coherence: Theory and Evidence," *Journal of Economic Behavior & Organization*, 1994. Foundational work for the management view on core competencies and organizational coherence.

A Equilibrium

Since the sectors are symmetric and $\int B_j^{\frac{1}{\theta}} dj = \mathbb{E}[B_j^{\frac{1}{\theta}}] = 1$, we have

$$Q_j = Q \qquad \mathbb{E}[P_j] = 1 \qquad \forall j$$

Given the utility function, $u(C, L) = \ln C - L$, household optimality implies:

$$Q = W \tag{33}$$

where W is the wage. Using the firm's optimal price,

$$Q_{ij} = \left(\frac{\mathbb{E}[P_{ij}]}{\mathbb{E}[P_j]}\right)^{-\eta} Q_j = \left[\frac{\eta}{\eta - 1}\left(\frac{1}{1-\tau_{ij}}\frac{W}{A_{ij}}\right)\right]^{-\eta} Q \tag{34}$$

$$= \left(\frac{\eta - 1}{\eta}\right)^{\eta} (1-\tau_{ij})^{\eta} A_{ij}^{\eta} Q^{1-\eta} \tag{35}$$

where we use the general equilibrium condition $Q = W$. Aggregating, we can solve for Q:

$$Q = \left(N_j^{\varpi} \int Q_{ij}^{\frac{\eta-1}{\eta}} di \right)^{\frac{\eta}{\eta-1}} \tag{36}$$

$$= \left(N_j^{\varpi} \int \left(\frac{\eta-1}{\eta} \right)^{\eta-1} (1-\tau_{ij})^{\eta-1} A_{ij}^{\eta-1} \, di \right)^{\frac{\eta}{\eta-1}} Q^{1-\eta} \tag{37}$$

$$= \frac{\eta-1}{\eta} \left(N_j^{\varpi} \int (1-\tau_{ij})^{\eta-1} A_{ij}^{\eta-1} \, di \right)^{\frac{1}{\eta-1}} . \tag{38}$$