Essays in Asset Pricing and Labor Economics

by

Khalil Esmkhani

M.Sc., Arizona State University, 2016

M.Sc., Sharif University of Technology, 2014

B.Sc., Sharif University of Technology, 2011

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Business Administration)

The University of British Columbia
(Vancouver)

October 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Essays in Asset Pricing and Labor Economics

submitted by **Khalil Esmkhani** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Business Administration**.

Examining Committee:

Jack Favilukis, Associate Professor, Sauder School of Business, UBC
Supervisor

Lorenzo Garlappi, Professor, Sauder School of Business, UBC
Supervisory Committee Member

Guillermo Marshall, Associate Professor, Sauder School of Business, UBC
University Examiner

Jesse Perla, Associate Professor, Vancouver School of Economics, UBC
University Examiner

Adolfo de Motta, Associate Professor, Desautels Faculty of Management, McGill University
External Examiner

Additional Supervisory Committee Members:

Giovanni Gallipoli, Professor, Vancouver School of Economics, UBC
Supervisory Committee Member

Abstract

This thesis is a collection of four essays. In the first chapter, I build a model of a production network where each firm's choice of debt is endogenous. For a realistic network calibrated to the US economy, the model predicts that more central firms are safer, pay less to borrow, and choose higher debt. I confirm these predictions empirically. The model also produces an important externality: due to bankruptcy costs and network connections, there is a wedge between the socially optimal capital structure and the decentralized equilibrium. This suggests that government policies could reduce the externality, for example by making the tax shield a function of the network position.

In chapter 2, I study a standard CARA-normal asset pricing model with arbitrary information-sharing social networks. In the benchmark equilibrium, I show more central investors have access to more information transmitted via the network; and a more connected network always improves price informativeness. However, when there is uncertainty about the source of information and the quality of information varies across different sources, more information sharing could lead to a less informative market price. In addition more central investors' signals could become distorted, lowering their overall access to information via the network.

In chapter 3, we quantify firm heterogeneity in skill returns and present direct evidence of worker–firm complementarities. Using population data linked to cognitive and noncognitive skill measures, we estimate a model of firm-specific returns to these attributes. We find evidence of significant return heterogeneity, sorting, and earnings convexification.

In chapter 4, we study the introduction of Universal Basic Income (UBI) with a particular focus on how it affects real estate and the urban environment. In the baseline calibration with $5,000 UBI, about 38% of households see large welfare gains, but the remainder see smaller welfare losses. Prices, rents, and the ownership rate all fall. The wage rises and the makeup of the city's inner core versus outer suburbs also changes, although these changes depend on exactly how UBI is financed. The more progressive the financing scheme, the more likely high income households are to leave the city center.

Lay Summary

In the first chapter of the thesis, I study the implications of the complex network of inter-industry trades of intermediate goods on firms' financing decisions. In chapter two, I study the significance of information source uncertainty in investors' social networks for market price informativeness. In chapter three, we study the heterogeneity of firms in rewarding different skill attributes and its implications for the matching of employers and employees in the labor market. Finally, in chapter four, we quantitatively analyze the introduction of Universal Basic Income with a particular focus on real estate and the urban environment.

Preface

The research projects in chapters 1 and 2 were identified and performed solely by the author. Chapter 3 is based on an unpublished project with Giovanni Gallipoli (Vancouver School of Economics, University of British Columbia) and Michael Böhm (Department of Economics, University of Bonn). In this project, all authors worked on all aspects of the paper, including identifying the research question, theoretical analysis, numerical implementation, empirical work, and manuscript writing. Chapter 4 is based on an unpublished project with Jack Favilukis (Sauder School of Business, University of British Columbia) and Stijn Van Nieuwerburgh (Graduate School of Business, Columbia University). In this project, the author was primarily responsible for the theoretical extension of an existing model and numerical implementation of the extended model. In all projects, the author received valuable advice and guidance from his dissertation committee members and faculty at Sauder School of Business.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my extreme gratitude to the members of my supervisory committee: Jack Favilukis, Giovanni Gallipoli, and Lorenzo Garlappi, for their continuous support and mentorship. I would like to also thank the rest of the UBC Finance faculty, and Michael Böhm for their insightful and generous comments as well as for the tough questions that helped me develop my economic intuition. I am also thankful to Elaine Cho for her extraordinary dedication to helping students from the first step of the Ph.D. program to the last.

Last but not least, I wish to express my immense gratitude to my beloved, Hamideh, and my parents. Without their love and sacrifice none of this would have been possible.

Chapter 1

Capital Structure and Production Networks

1.1 Introduction

A firm is a production entity that converts a set of inputs to output(s). Naturally, the details of this transformation process are consequential for firms' outcomes. One significant aspect is the input-output relations. Production of goods and services is never done in isolation and firms rely heavily on various inputs from one another used as intermediate goods. Defining industries as a set of firms who produce the same good, in 2019 the intra-industry intermediate goods trade summed up to 76 percent of the US GDP. Studying the significance of these input-output dependencies, also known as production networks, for various macroeconomic phenomena is an active line of research, however, the cross-sectional implications for individual firms are less explored. Notably, these input-output connections are not symmetric and each industry plays a unique role in the economy producing a unique good. In other words, in the overall complex network of input-output relations, each industry has a unique "position" relative to the network. The main objective of this paper is to study the effects of these input-output relations (at the industry level), on firms' capital structure and bond prices.

To study the network implications, first I extend a standard trade-off model of capital structure to accommodate an arbitrary linear network. The model is an endowment economy in which individual firms' endowments are linked to others through a network. The model additionally features bankruptcy costs and an endogenous choice of capital structure. The model predicts that firms' position in the network determines their lower tail risk and hence their capital structure through two related but distinct channels. First, firms' cash flow distributions depend on their position in the network. Hence, by making explicit the cash flow distribution, the shape of the network has first-order effects on firms' choice of capital structure. Second, bankruptcies might have cascading effects and further influence the neighboring firms (Barrot and Sauvagnat, 2016). Taking into account these propagating effects could alter firms' financing incentives. This channel could also create strategic incentives for firms to alter their leverage to change other firms' behavior for their benefit.

I then extend the model to allow for production where I build a general equilibrium model of a production network and endogenous debt choice. I characterize how the overall production network of the economy and individual industries' unique position in this network affect their cash flow risk and optimal capital structure.

The model predicts that more central industries may be safer or riskier depending on the network structure and the ability of firms to substitute different inputs in their production. An industry's risk is determined by both its inputs and its outputs. All else equal, if a focal industry's output is used heavily as an input in other industries, that focal industry faces less risk in its output market and will have a higher leverage ratio. On the input side, relying on more goods as input in production could make industries less or more risky, depending on their ability to substitute different inputs with one another. If the production process allows an industry to easily substitute inputs, then relying on more goods lowers the risk, and increases its debt. On the other hand, if an industry is not able to substitute input goods, then relying on more goods increases the industry's risk and lowers its leverage. For the US economy, the observed production network is complex and both these channels coexist and interact in determining industries' risk and financing choices.

With reasonable parameter values calibrated to the US input-output accounts in 2019, the model predicts that more central industries issue more debt and sell their bond for a higher price. The natural measure of centrality that emerges from my model is the eigenvector centrality. In the calibrated network and absent any non-network heterogeneity, the eigenvector centrality explains more than 93 percent of the variation in the leverage ratios and bond prices in the model. Intuitively based on this measure, an industry is more central if the good it produces is utilized in the production of other central industries with high weights.

Moreover, in both models, I show that network connections create a source of externality that firms impose upon each other via their debt choice. Firms only take their individual probability of bankruptcy into account when deciding about how much debt to issue, however, their bankruptcy has implications for their neighboring firms as mentioned. More firms in a highly levered industry are likely to become bankrupt which in turn would affect the neighboring industries. This creates a wedge between the individual firms' debt choice and what is socially optimal. I show that government could reduce this externality by imposing industry-specific debt incentive schemes such as industry-specific tax shields.

In the last section of the paper, I find empirical evidence supporting these findings. I construct a series of detailed Social Accounting Matrices (SAM) for the US economy using the Input-Output and National Income and Product Accounts (NIPA) from the online database of the Bureau of

Economic Analysis (BEA). I show more central industries in the SAM matrix tend to issue more debt. Moreover, using the Mergent Fixed Income Securities Database (FISD) I show more central industries tend to issue bonds with higher credit ratings and lower spread in the market, however, the centrality effect on credit spread becomes statistically insignificant once firm variables are controlled for in a lower size sample for which the firm level variables are observed.

Related literature. This paper contributes to several strands of existing work. First, it contributes to a recent and growing literature on the implications of production networks in economics and finance. A key insight in this literature is that the shape of the network affects several macroeconomic phenomena. For instance, how asymmetric production networks can translate idiosyncratic shocks into aggregate fluctuations (Acemoglu et al., 2012, Atalay, 2017, Baqaee and Farhi, 2019, Gabaix, 2011, Horvath, 2000). Beyond the second moment, Acemoglu et al. (2017) analyzes the role of production networks in creating substantial macroeconomic downturns. Another related topic that is analyzed in the literature is the propagation of shocks and activities along the production network connections. Using natural disasters, Barrot and Sauvagnat (2016) find that idiosyncratic firm-level negative shocks propagate in production networks to their costumers. Baqaee (2018) show that the interaction of input-output networks with industry-level market structure could generate cascades of firm entry and exit across the economy. Ahern and Harford (2014) finds that mergers propagate in waves across the network through customer-supplier links. This paper contributes to this literature by endogenizing the bankruptcy decisions of firms in a production network which affects the production and real aggregate outcomes. In addition, the corporate bond market added to this class of models creates another channel through which the real and financial shocks can be originated, propagated, amplified, or dampened in the economy.

Second, this paper contributes to the literature focusing on externalities and systemic risk in network economies. Herskovic (2018) builds a production-based asset pricing model with production networks and approximates the network effect on the aggregate consumption asset using two sufficient statistics, network concentration and sparsity. In the cross-section, Ahern (2013) finds that more central firms earn higher stock returns. In addition, Herskovic et al. (2020) use a network model to explain the joint evolution of firm size and volatility distribution. Cossin and Schellhorn (2007) given a fixed network of borrowing and lending among firms, develop a structural model to price credit risk taking into account the entire network risk. In a somewhat similar setting but with a different objective, again with an exogenous borrowing and lending network among banks, Acemoglu et al. (2015) show that the extent of financial contagion exhibit a phase transition, that is densely connected financial networks stabilize the economy for small shocks and destabilize for big shocks at the same time. In a similar framework, Bernard et al. (2022) endogenizes intervention in financial crises as the strategic negotiation between a regulator and creditors of distressed banks. More broadly, in the literature studying externalities in financial markets, Lorenzoni (2008) argue limited commitment in financial contracts and determination of asset prices in spot markets lead to competitive contracts with a pecuniary externality and inefficient financing. Bianchi (2011) studies the systemic credit externality in a model with credit constraints linking debt to market-determined prices. They show private agents fail to internalize the financial amplification effects of carrying a large amount of debt when credit constraints bind. Theoretically, I extend the asset pricing network models from the literature to include a corporate bond market with endogenous debt choice to analyze the cross-sectional implications of production networks on firms' equilibrium debt and bond price. I show input-output linkages could lead to negative externality and inefficient borrowing. Further, I empirically document that more central firms in the production network tend to have lower credit risk and bond spread.

Finally, this paper also contributes to the broad literature on capital structure determinants. I use the standard trade-off theory of capital structure where firms balance benefits and costs of debt financing at the margin (Kraus and Litzenberger, 1973, Robichek and Myers, 1966, Scott, 1976, Stiglitz, 1972). Besides using the trade-off methodology, this paper is also related to the stakeholder theory of capital structure. Titman (1984) argues that the possibility of liquidation can impose costs on a firm's non-financial stakeholders such as its workers, customers, and suppliers. Titman and Wessels (1988) argue that a firm's customers face switching costs upon liquidation and these costs are higher when the firm's product is unique or there is a supplier-specific investment. Banerjee et al. (2008) show that firms with these bilateral relationships are likely to produce unique products and maintain lower leverage. Hennessy and Livdan (2009) propose a bargaining theory of leverage based on upstream-downstream relationships between firms. In this paper, I show that firms' position in the production network could affect their financial decisions beyond their immediate customers and suppliers. The entire network directly affects firms' cash-flow distribution and hence has a first-order effect on their choice of capital structure. Furthermore, through bankruptcy deadweight loss, firms' debt choices also feed back into the real economy and have propagating real and financial effects on other firms. As a result, firms' optimal leverage and debt price are jointly determined in equilibrium in which the shape of the production network plays a key role. Empirically, I confirm a similar finding of Prostakova (2018) showing more central firms in the production network tend to have higher leverage ratios even after controlling for known variables from the literature.

Model

I start this section by introducing a stylized trade-off model with exogenous cash flows and no production (section 1.2). Apart from the extension of adding a network feature here, this first endowment model is extensively used in the trade-off theory of capital structure literature. Later in section 1.3, I endogenize the cash flows by introducing production to the model in partial and general equilibrium settings. While the endowment model transparently highlights the main channels through which networks affect firms' financing decisions and the origins of debt externality in the network, the latter model with the production network is more structurally interpretable and easier to calibrate to available economic data.

1.2 Endowment Model

Model Environment. There are two periods $t = 0, 1$, a set of identical risk-neutral investors (or a representative investor), n industries, and each industry is comprised of a continuum of firms indexed by a uniform measure $f \in (0, 1)$. Each firm can issue debt at $t = 0$ and owns a project which produces a random cash flow in period $t = 1$. Let d_{fi} and x_{fi} denote the face value of debt and total cash flow of firm f in industry i respectively, and $D_i = \int_0^1 d_{fi} \mathrm{d}f$ and $X_i = \int_0^1 x_{fi} \mathrm{d}f$ are the industry aggregates of the corresponding variables. Moreover, bold letters without subscript denote the corresponding variables of industries stacked together in a column vector, for instance $\mathbf{D} = [D_i]_{n\times 1}$ and $\mathbf{X} = [X_i]_{n\times 1}$.

Cash-flow Network. Industries are connected in a network that ties their cash flows to one another. In particular, if firm f in industry i is not bankrupt (nb) then its total cash flow is equal to

$$x_{fi}^{\mathrm{nb}} = \mu_i + \sum_{j=1}^{n} w_{ij} X_j + \varepsilon_{fi}. \tag{1.1}$$

x_{fi}^{nb} can also be considered as fi's unlevered cash flow where μ_i is an industry-specific constant, w_{ij} is the direct marginal effect of industry j's aggregate cash flow on i's firms, and ε_{fi} is an idiosyncratic firm-specific random shock uniformly distributed on $(0, \varphi_i)$. w_{ij} represents the extent to which i is directly affected by j in the network and captures the structural relationships between industries' production relations modeled here as a reduced form linear link.[1] It is important to distinguish between the direct effects and equilibrium effects in this setting. In equilibrium, industry j's cash flow could indirectly affect i's even when $w_{ij} = 0$ through other industries. w_{ij} only captures the structural and direct effect of j on i.

An informative representation of $\mathbf{W}$ involves graph theory. In this representation, each industry corresponds to a node of a directed graph, and w_{ij} is the weight of the edge connecting node j to i. Figure 1.1 plots such a network for a generic economy.

At $t = 1$ if the realized x_{fi}^{nb} is smaller than fi's outstanding debt obligation d_{fi}, the firms is forced into bankruptcy and incur a bankruptcy cost proportional to the liabilities excess of the unlevered

[1] In section 1.3 I explicitly model the production functions and the exact nature of these connections in the context of production networks.

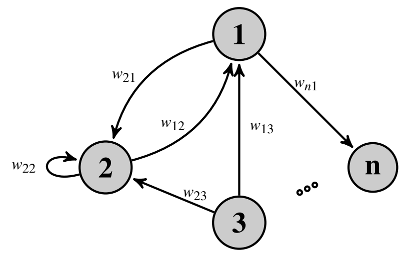


Figure 1.1: Graphical representation of cash flow dependencies.

cash flow. The following equation pins down the total cash flow of firm fi (net of bankruptcy costs) as a function of all industries' aggregate debt vector $\mathbf{D}$ (the aggregate state), its own debt d_{fi}, and the realized idiosyncratic shock ε_{fi}.

$$x_{fi}(d_{fi},\varepsilon_{fi};\mathbf{D}) = x^{\text{nb}}_{fi}(\varepsilon_{fi};\mathbf{D}) - \overbrace{\delta\ \max\left\{d_{fi} - x^{\text{nb}}_{fi}(\varepsilon_{fi};\mathbf{D}), 0\right\}}^{\text{bancruptcy cost}} \tag{1.2}$$

$$= \mu_i + \sum_{j=1}^{n} w_{ij}X_j(\mathbf{D}) + \varepsilon_{fi} - \delta\ \max\left\{d_{fi} - \left(\mu_i + \sum_{j=1}^{n} w_{ij}X_j(\mathbf{D}) + \varepsilon_{fi}\right), 0\right\}$$

$$X_j(\mathbf{D}) = \int_0^1 x_{fj}(d_{fj},\varepsilon_{fj};\mathbf{D})\,\mathrm{d}f \tag{1.3}$$

X_j is industry j's aggregate cash flow and $\delta > 0$ is the marginal cost of bankruptcy. The only source of uncertainty in the model is ε_{fi}, and since there is a continuum of firms within each industry, there is no aggregate uncertainty in the model. Thus the aggregate industry cash flows only depend on the aggregate debt vector $\mathbf{D}$, and individual firm fi's cash flow is a function of d_{fi}, ε_{fi}, as well as the aggregate state $\mathbf{D}$.

Debt Benefits. Despite the deadweight loss due to bankruptcy, debt could increase firm value through tax benefits, reduction in information frictions and agency costs, enhanced bargaining power, etc. In any trade-off model of capital structure, firms balance these costs and benefits at the margin to maximize their value. To keep the model tractable and focus on the network aspect, I abstract from modeling the exact source of these benefits and assume debt has a constant marginal benefit for firms' equity holders which is realized at $t = 0$, that is firm fi's equity holders receive a benefit of τd_{fi} at $t = 0$. The constant $\tau > 0$ represents the marginal benefits of debt in the model.

Cash Flow to Stakeholders. Depending on the bankruptcy state of firm fi, its cash flow at $t = 1$ is split between the creditors and shareholders of fi. If there is enough cash flow at $t = 1$ for full repayment of d_{fi}, equity holders fulfill the debt obligation and receive the residual cash flow

$x_{fi} - d_{fi}$. Otherwise, the firm is forced into bankruptcy, creditors take over the firm, and receive any residual cash flow net of bankruptcy costs.

$$\mathrm{CF}^{\mathrm{equity1}}_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D}) = \begin{cases} x_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D}) - d_{fi} & \text{if not bankrupt} \\ 0 & \text{if bankrupt} \end{cases} \tag{1.4}$$

$$\mathrm{CF}^{\mathrm{debt1}}_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D}) = \begin{cases} d_{fi} & \text{if not bankrupt} \\ x_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D}) & \text{if bankrupt} \end{cases} \tag{1.5}$$

$$x_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D}) = \mathrm{CF}^{\mathrm{equity1}}_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D}) + \mathrm{CF}^{\mathrm{debt1}}_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D}) \tag{1.6}$$

Equilibrium. At $t = 0$ the equilibrium debt d^*_{fi} is chosen to maximize fi's total value which is equal to the debt benefit τd_{fi} plus the time zero price of claims over the firm's cash flow to equity and debt at $t = 1$. With risk-neutral investors, the prices of these claims are equal to their expected cash flow and

$$\begin{aligned} d^*_{fi} &\in \arg\max_{d_{fi}} \quad \tau d_{fi} + \overbrace{\mathbb{E}^0[\mathrm{CF}^{\mathrm{equity1}}_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D})]}^{\text{equity value at } t=0} + \overbrace{\mathbb{E}^0[\mathrm{CF}^{\mathrm{debt1}}_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D})]}^{\text{debt value at } t=0} \\ &= \arg\max_{d_{fi}} \quad \tau d_{fi} + \mathbb{E}^0\left[x_{fi}(d_{fi}, \varepsilon_{fi}; \mathbf{D})\right] && (1.7) \\ &= \arg\max_{d_{fi}} \quad \underbrace{\tau d_{fi}}_{\text{debt benefit}} + \underbrace{\mathbb{E}^0\left[x^{\mathrm{nb}}_{fi}(\varepsilon_{fi}; \mathbf{D})\right]}_{\text{unlevered asset value}} - \underbrace{\delta \mathbb{E}^0\left[\max\left\{d_{fi} - x^{\mathrm{nb}}_{fi}(\varepsilon_{fi}; \mathbf{D}), 0\right\}\right]}_{\text{bancruptcy cost}} && (1.8) \end{aligned}$$

Some Notation Simplification. The dependence of functions $\mathrm{CF}^{\mathrm{equity1}}_{fi}(.)$, $\mathrm{CF}^{\mathrm{debt1}}_{fi}(.)$, and $x_{fi}(.)$ to a specific firm f are only through their arguments d_{fi} and ε_{fi}, hence I drop the f index for these functions hereinafter without loss of generality.

1.2.1 Debt Market Structure and Equilibrium Characterization

Throughout the paper, I only focus on symmetric equilibria in which all the firms within each industry choose the same level of debt $d_{fi} = D_i$ (for all fi). However, this shared debt D_i could arise in widely different market structures. I study equilibria in two extreme cases of markets with atomistic/competitive or strategic firms. If firms behave atomistically, then each individual firm does not internalize the equilibrium condition $d_{fi} = D_i$. On the other extreme, in a market with strategic behavior, firms within each industry fully internalize the equilibrium condition $d_{fi} = D_i$ in their decision-making and behave as if there is a representative firm within the industry. In both cases, rational investors price firms' assets consistent with the market structure. Let $\Theta \in \{0,1\}$ be an indicator parameter that is equal to zero for the competitive market structure and one for the strategic case. Definition 1 provides a Nash equilibrium notion for both $\Theta = 0$ and 1.

Definition 1. $\mathbf{D}^* = [D_i^*]_{n\times 1}$ *is an equilibrium if and only if for all* $i = 1, ..., n$

$$D_i^* \in \arg\max_{d_{fi}} \quad \tau d_{fi} + (1-\Theta)\mathbb{E}^0\left[x_i(d_{fi}, \varepsilon_{fi}; \mathbf{D}^*)\right] + \Theta\mathbb{E}^0\left[x_i(d_{fi}, \varepsilon_{fi}; (d_{fi}, \mathbf{D}^*_{-i})\right]$$

where $\Theta \in \{0,1\}$.

If $\Theta = 0$ firm fi's objective function reduces to $\tau d_{fi} + \mathbb{E}^0\left[x_i(d_{fi}, \varepsilon_{fi}; \mathbf{D}^*)\right]$. In this case, firm fi takes the entire aggregate state of the economy $\mathbf{D}^*$ as given (including D_i^*) when choosing their optimal debt. On the other hand, in the strategic case or $\Theta = 1$, the objective function is $\tau d_{fi} + \mathbb{E}^0\left[x_i(d_{fi}, \varepsilon_{fi}; (d_{fi}, \mathbf{D}^*_{-i})\right]$, firms in the industry i take only $\mathbf{D}^*_{-i}$ as given, and regard $D_i = d_{fi}$ as a decision variable.

Definition 1 implies the following optimality condition:

$$\tau + \frac{\partial \mathbb{E}[x_i(d_{fi}, \varepsilon_{fi}; \mathbf{D}^*)]}{\partial d_{fi}}\Big|_{d_{fi}=D_i^*} + \Theta \frac{\partial \mathbb{E}[x_{fi}(d_{fi}, \varepsilon_{fi}; (D_i, \mathbf{D}^*_{-i}))]}{\partial D_i}\Big|_{d_{fi}=D_i=D_i^*} \leq 0 \;\; \text{(with equality if } D_i^* > 0\text{)} \qquad (1.9)$$

The first term in the left-hand side of (1.9) is the marginal benefit of an extra unit of debt which is assumed to be constant τ. The second term is the marginal effect of an increase in d_{fi} on fi's expected cash flow if it behaves atomistically in a competitive market. The last term is the marginal effect of an increase in industry i's aggregate debt on fi's expected future cash flow. This last term is only present when firms act strategically ($\Theta = 1$). Proposition 1 further characterizes the equilibrium in this economy based on the underlying industry network.

Proposition 1. *Industries' equilibrium debt vector* $\mathbf{D}^*$ *solves the following system of equations:*

$$\delta \text{ Prob}\left(\mathbb{1}_i^{\mathsf{b}}(\mathbf{D}^*) = 1|\mathbf{D}^*\right) + \Theta\delta\left[\widetilde{\mathbf{W}}(\mathbf{D}^*) + \widetilde{\mathbf{W}}(\mathbf{D}^*)^2 + \dots\right]_{ii} \mathbb{P}\text{rob}\left(\mathbb{1}_i^{\mathsf{b}}(\mathbf{D}^*) = 1\Big|\mathbf{D}^*\right) = \tau \qquad (1.10)$$

for all $i = 1, ..., n$ *where*

$$\widetilde{\mathbf{W}}(\mathbf{D}^*) = \left(\mathbb{I} + \delta \mathbb{1}^{\mathsf{b}}(\mathbf{D}^*)\right)\mathbf{W}, \qquad (1.11)$$

$$\mathbb{1}^{\mathsf{b}}(\mathbf{D}^*) = \text{diag}\left[\mathbb{P}\text{rob}\left(\mathbb{1}_i^{\mathsf{b}}(\mathbf{D}^*) = 1\Big|\mathbf{D}^*\right)\right], \qquad (1.12)$$

$\mathbf{W} = [w_{ij}]_{n\times n}$ *is the network adjacency matrix, and* $\mathbb{1}^{\mathsf{b}}(\mathbf{D}^*)$ *is a diagonal matrix of bankruptcy probabilities.*

The right-hand side of the equation (1.10) is the marginal benefit of debt for firms in industry i, whereas the left-hand side is the expected marginal cost. The first term δ Prob$(\mathbb{1}_i^{\mathsf{b}}(\mathbf{D}^*) = 1|\mathbf{D}^*)$ is a baseline marginal cost due to the direct effect that more debt has on the probability of bankruptcy. More debt results in a higher probability of default and greater expected deadweight loss. This cost is present even when there are no network linkages between firms, in fact if $\mathbf{W} = [0]_{n\times n}$ then (1.10) reduces to $\text{Prob}(\mathbb{1}_i^{\text{b}}(\mathbf{D}^*) = 1|\mathbf{D}^*) = \tau/\delta$ and the default probability is the only determinant of the equilibrium debt. This term is the only perceived cost for firms and investors in an economy with atomistic firms i.e. $\Theta = 0$.

The second term in the right hand side of (1.10) captures the equilibrium effects of the network which itself is a series of different order terms weighted by $\delta\ \text{Prob}(\mathbb{1}_i^{\text{b}}(\mathbf{D}^*) = 1|\mathbf{D}^*)$. For a given $\mathbf{D}$, each firm in industry i will become bankrupt at $t = 1$ with probability $\text{Prob}(\mathbb{1}_i^{\text{b}}(\mathbf{D}) = 1|\mathbf{D})$ depending on its realized idiosyncratic cash flow shock. Since there is a continuum of firms in each industry, the Law of Large Numbers implies that exactly fraction $\text{Prob}(\mathbb{1}_i^{\text{b}}(\mathbf{D}) = 1|\mathbf{D})$ of industry i's firms will become bankrupt at $t = 1$. For solvent firms, a marginal increase of D_i does not affect their cash flow. On the other hand, a marginal increase in D_i lowers insolvent firms' cash flow at the rate δ due to bankruptcy costs. Hence the effect of a marginal increase in D_i on the industry aggregate is $\delta\ \text{Prob}(\mathbb{1}_i^{\text{b}}(\mathbf{D}) = 1|\mathbf{D})$.

This Marginal reduction in industry i's aggregate cash flow, further changes its firms' cash flow via the network at rate w_{ii} and $(1+\delta)w_{ii}$ for solvent and bankrupt firms respectively. The industry-wide aggregate is $\widetilde{w}_{ii} = (1+\delta\ \text{Prob}(\mathbb{1}_i^{\text{b}}(\mathbf{D}) = 1|\mathbf{D}))w_{ii}$.

In addition, the previous change in turn has a first-order effect on the cash flow of i's immediate neighbors in the network due to the direct network connections by rate $\widetilde{w}_{ji}$ for firms in industry j. Similar to $\widetilde{w}_{ii}$, $\widetilde{w}_{ji}$ depends on the fraction of bankrupt firms in industry j. Consequently, a change in j's aggregate cash flow will feed back into i and change its cash flow by rate $\widetilde{w}_{ij}$. Adding these effects for all j will result in $\sum_{j=1}^{n} \widetilde{w}_{ij}\widetilde{w}_{ji}$ which is exactly ii-th element of $\widetilde{\mathbf{W}}^2$. Essentially this term is the marginal reduction in i's aggregate cash flow due to first order change in i's immediate neighbors' cash flow caused by a reduction in i's cash flow itself.

The network effects are not confined to only first and second-order effects, and a change in i's aggregate cash flow will change that of neighbors of i's neighbors; and their neighbors, etc. Accordingly, the ii-th element of $\widetilde{\mathbf{W}}^3$ is the marginal effect in i's cash flow due to the change in immediate neighbors' of i's immediate neighbors cash flow caused by a reduction in i's cash flow itself. Adding all these higher order effects results in the total network effect in equation (1.10).

Proposition 1 essentially divides the effect of the network on firms' expected profit into a baseline and feedback of propagating terms. The baseline default probability is directly affected by the network since $\mathbf{W}$ has a first-order effect on the cash flow distribution of firms. The propagating terms capture the spreading effects that a change in D_i could cause. As discussed next, these propagating effects could be a source of externality in the network when not accounted for properly.

1.2.2 Debt Externality

Externalities pose fundamental economic policy problems when agents do not internalize the indirect costs or benefits of their decisions on other agents. In the context of this paper, an industry's debt choice affects the other industries via the network. If an industry issues too much debt relative to what is socially optimal, a larger fraction of its firms become bankrupt. These inefficient bankruptcies affect the cash flow of the neighboring industries and beyond through the network connections. However, firms do not internalize these effects when deciding about their own debt. This leads to a wedge between the socially optimal level of debt and what is chosen by the firms in a market equilibrium. First, in the aggregate level, there might be more or less than optimal issued debt and thus bankruptcy. Second, the distribution of the debt may be inefficient from a planner's perspective. For instance, a focal industry who is heavily connected to other industries might issue larger than optimal debt since its effect on other industries does not factor in its leverage decision. In principle, government interventions could ensure that all the benefits and costs are fully assumed and socially optimal outcomes are obtained.

To find the socially optimal distribution of debt, I maintain the assumption that there is a constant marginal benefit of raising debt and a socially optimal debt distribution maximizes an equally weighted sum of industries' expected future cash flows.

Definition 2. $\mathbf{D}^{\text{sp}} = [D_i^{\text{sp}}]_{n\times 1}$ *is the social planner's optimal debt choice if and only if*

$$\mathbf{D}^{\text{sp}} \in \arg\max_{\mathbf{D}} \sum_{j=1}^{n} \left(\tau D_j + \mathbb{E}[x_j(D_j, \varepsilon_{fj}; \mathbf{D})]\right) \tag{1.13}$$

The optimality conditions for the planner's problem are:

$$\tau + \frac{\partial \mathbb{E}[x_i(D_i, \varepsilon_{fi}; \mathbf{D}^{\text{sp}})]}{\partial D_i}\Big|_{D_i = D_i^{\text{sp}}} + \sum_{j=1}^{n} \frac{\partial \mathbb{E}[x_j(D_j^{\text{sp}}, \varepsilon_{fj}; \mathbf{D})]}{\partial D_i}\Big|_{\mathbf{D}=\mathbf{D}^{\text{sp}}} \leq 0 \quad (\text{with equality if } D_i^{\text{sp}} > 0) \tag{1.14}$$

The left-hand sides of (1.14) has the following extra terms compared to the market equilibrium condition (1.9):

$$(1-\Theta)\frac{\partial \mathbb{E}[x_i(D_i^{\text{sp}}, \varepsilon_{fi}; (D_i, \mathbf{D}_{-i}^{\text{sp}}))]}{\partial D_i}\Big|_{D_i = D_i^{\text{sp}}} + \sum_{j\neq i} \frac{\partial \mathbb{E}[x_j(D_j^{\text{sp}}, \varepsilon_{fj}; (D_i, \mathbf{D}_{-i}^{\text{sp}}))]}{\partial D_i}\Big|_{D_i = D_i^{\text{sp}}} \tag{1.15}$$

which reveals a source of externality. The first term is the marginal effect of an increase in D_i on i's firms, which is not accounted for by atomistic firms in a competitive market. The second term is the effect of an increase in D_i on all the other industries which only appear in the planner's problem as individual firms are only concerned with their own value. This term represents the externality imposed by industry i on all the other industries. Similar to proposition 1, the following proposition characterizes the planner's choice of debt in the economy as a function of the underlying network structure.

Proposition 2. *The social planner's debt vector* $\mathbf{D}^{\,\mathrm{sp}}$ *solves the following system of equations for all* $i = 1, ..., n$*:*

$$\delta \sum_{j=1}^{n} \left[\mathbb{I} + \widetilde{\mathbf{W}}(\mathbf{D}^{\,\mathrm{sp}}) + \widetilde{\mathbf{W}}(\mathbf{D}^{\,\mathrm{sp}})^2 + \dots\right]_{ji} \mathrm{Prob}\left(\mathbb{1}_i^{\mathrm{b}}(\mathbf{D}^{\,\mathrm{sp}}) = 1 \middle| \mathbf{D}^{\,\mathrm{sp}}\right) = \tau \tag{1.16}$$

The intuition behind equation (1.16) is similar to (1.10) once one notes that the planner always takes the effect of D_i on i's firms into account even if $\Theta = 0$, and the planner adds the marginal effects of change in D_i on all the other industries.

The above analysis suggests that a decentralized debt decision by industries could potentially lead to a socially inefficient distribution of debt because of network externalities. However the planner could move the economy towards its socially optimal debt distribution by changing firms' incentives to raise more or less debt based on their industry. One such solution is via industry-specific tax incentives. If an industry issues less than socially optimal debt, the government could increase its debt by providing a larger tax shield for the firms within that industry. Similarly lowering the tax benefits of debt would reduce the debt of an over-leveraged industry. An immediate corollary to propositions 1 and 2 is that the following industry-specific changes in the debt benefits will eliminate the externality and achieve socially optimal debt distribution in the decentralized equilibrium.

Corollary 1. *The planner could achieve the socially optimal debt distribution via a decentralized equilibrium by altering the marginal debt benefit for industry i to*

$$\begin{aligned}\tau_i = \tau - (1-\Theta)\delta \left[\widetilde{\mathbf{W}}(\mathbf{D}^{\,\mathrm{sp}}) + \widetilde{\mathbf{W}}(\mathbf{D}^{\,\mathrm{sp}})^2 + \dots\right]_{ii} \mathrm{Prob}\left(\mathbb{1}_i^{\mathrm{b}}(\mathbf{D}^{\,\mathrm{sp}}) = 1 \middle| \mathbf{D}^{\,\mathrm{sp}}\right) \\ - \delta \sum_{j \neq i} \left[\mathbb{I} + \widetilde{\mathbf{W}}(\mathbf{D}^{\,\mathrm{sp}}) + \widetilde{\mathbf{W}}(\mathbf{D}^{\,\mathrm{sp}})^2 + \dots\right]_{ji} \mathrm{Prob}\left(\mathbb{1}_i^{\mathrm{b}}(\mathbf{D}^{\,\mathrm{sp}}) = 1 \middle| \mathbf{D}^{\,\mathrm{sp}}\right)\end{aligned} \tag{1.17}$$

for all $i = 1, ..., n$*.*

Corollary 2. *If* $\mathbf{W} \geq 0$ *then* $D_i^* \geq D_i^{\mathrm{sp}}$*; and If* $\mathbf{W} \leq 0$ *then* $D_i^* \leq D_i^{\mathrm{sp}}$ *(for all i).*

The change in τ is exactly the marginal effects of D_i on industries that are not accounted for by the individual firms as mentioned above. Moreover, if all the network parameters are positive (negative), then the externality is always negative (positive) and the social planner chooses lower (higher) debt for all industries relative to the market equilibrium.

Takeaways. The endowment model suggests that the network affects the leverage choice via (i) determining the joint distribution of cash flows (baseline effect) and (ii) determining the propagating effects of bankruptcies across industries (propagation effect). Moreover, the network connections create a source of externality in the economy, which brings about inefficient aggregate level and distribution of debt and bankruptcies. The example in section 1.2.3 illustrates all of these points in a simple economy with three industries. Nonetheless, the linear and reduced-form nature of the endowment model falls short of capturing the actual form of these effects in the context of input-output networks. In section 1.3, I address this issue by introducing production to the model and explicitly modeling the production network that links industries.

1.2.3 Illustrative Example 1

In figure 1.2, the top panel summarizes the cash flow dependencies graph in this example economy. All the network weights are fixed except $w_{12} = w_{13} = w$ which varies in $[0, 0.3]$ and I find the equilibrium and planner's choice for all w in this interval. Increasing w makes industry 1 more central relative to 2 and 3, because as w rises, the cash flows of 1 depend more evenly on the entire network of industries. For all $i = 1, 2, 3$, I further assume $\mu_i = 0$; $\varepsilon_{fi} \sim \text{Uniform}(0, 1)$ are iid; $\tau = 0.1$; and $\delta = 0.5$. The bottom left panel in figure 1.2 plots the equilibrium debt D_1^* (for both cases of $\Theta = 0, 1$) and the planner's choice D_1^{sp} divided by firm's total value in industry 1. The bottom right panel plots the same graphs for industries 2 and 3 (plots for industry 3 are identical to 2 because of symmetry).

Figure 1.2 shows that as industry 1 becomes more central, its leverage ratio rises, and this is true for the atomistic, monopolistic, and social planner cases. This is mainly because of the baseline effect that w has on the joint distribution of cash flows in all industries.[2] The second result is that the propagating effects lower the strategic firms' leverage ($\Theta = 1$) relative to the competitive case ($\Theta = 0$). The third result is that when either atomistic or monopolistic firms choose leverage, their choice is above the social planner's. This is the aforementioned externality: first, monopolistic firms do not internalize the effects of their own leverage on the bankruptcy of other industries leading to too much debt relative to the social planner; in addition, atomistic firms also do not internalize the effects of their own leverage on the bankruptcy within their own industry leading to even more debt

1.3 Production Model

As mentioned before, the fixed linear structure of the endowment model is unable to capture the exact input-output relations of industries. Crucially, the ability of firms to substitute different inputs in response to equilibrium price shocks is absent in the endowment model. Moreover, the non-linear aspects of the production technologies (represented by Constant Elasticity of Substitution (CES) production functions) has been shown to be a source of shock amplification in production

[2] In this example w shifts the cash flow distribution of all firms to the right and makes them safer for each level of debt.

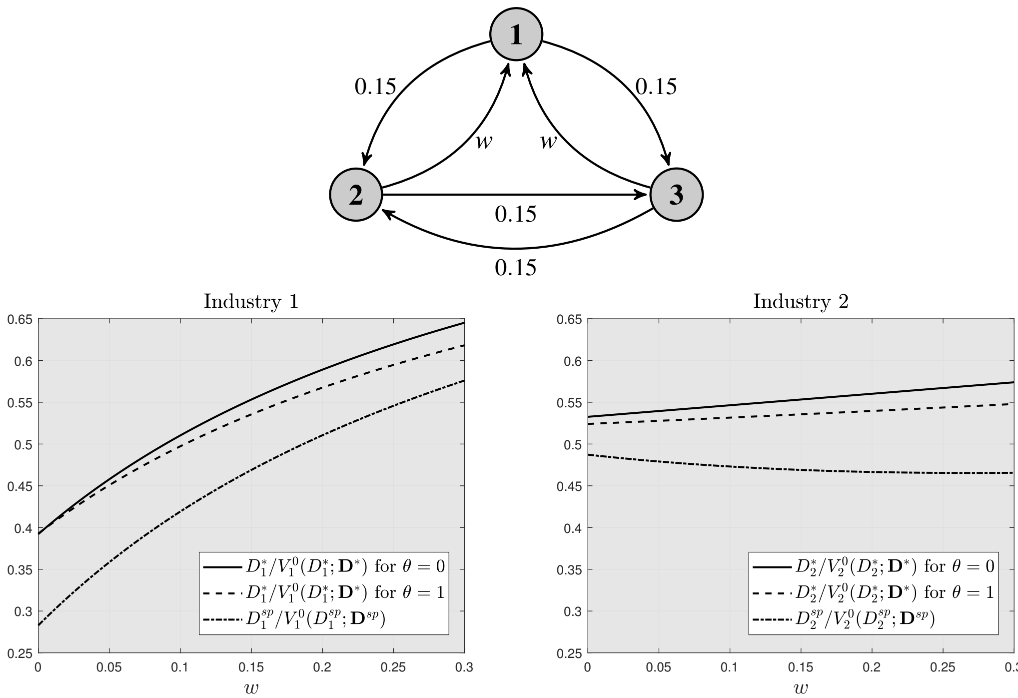


Figure 1.2: Example Economy

networks, which is also absent in the endowment model.

Following the literature on production networks, in this section I present a general equilibrium model of capital structure decision with multiple sectors connected through a production network. Contrary to the reduced-form representation of the industry linkages in the endowment model, here I explicitly model the network connections as weights of input goods from other industries that appear in each industry's production function. Similar to the endowment model, here the model also features a trade-off between the benefits of having higher debt and bankruptcy costs. On the production side, the model builds on the multisector economy of Long and Plosser (1983) which is extensively used in the recent literature (e.g. Acemoglu et al., 2012, 2017, Baqaee, 2018, Baqaee and Farhi, 2019, Herskovic, 2018). Crucially, in these type of models, the joint distribution of goods' prices is determined in a general equilibrium. Thus the unlevered baseline cash flow distribution of industries is a function of the underlying network through general equilibrium forces governing the goods' prices.

1.3.1 Environment

Time is discrete and there are two periods, $t = 0, 1$. There are n distinct goods produced by n different sectors or industries in the economy.[3] Each sector is comprised of a continuum of firms

[3] I use the words *sector* and *industry* interchangeably throughout the paper.

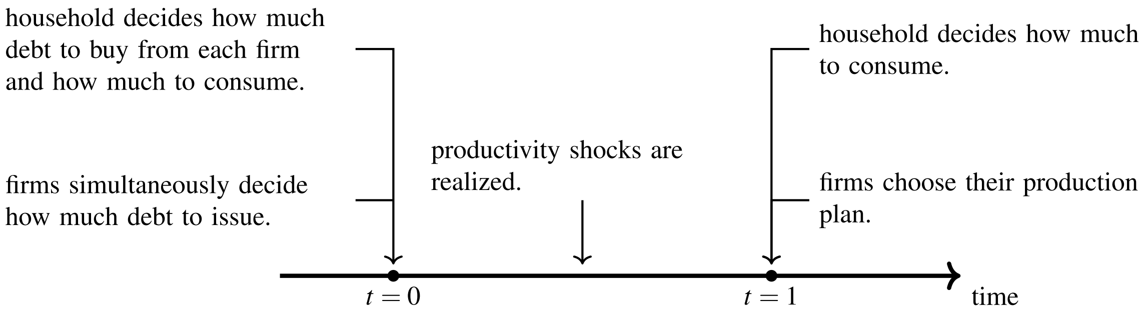


Figure 1.3: Time line of the general equilibrium model.

and production only takes place at $t = 1$. Firm f in industry i is exposed to a sector-specific random productivity shock Z_i, as well as an idiosyncratic shock z_{fi}, and uses goods built by other sectors as input to produce its own output. Firms issue defaultable debt and sell it to the household at $t = 0$, and repay (or possibly default on) their debt at $t = 1$. A representative household with a constant relative risk aversion (CRRA) utility function owns and lends to the firms. Figure 1.3 summarizes the timeline of the model.

I maintain the assumption that debt could increase firm value and again abstract from modeling the exact source of these benefits. I assume debt has a constant marginal benefit for firms' equity holders when solvent, namely firm fi's equity holders, if not bankrupt, receive a cash flow τd_{fi} at $t = 1$. The constant $\tau > 0$ represents the benefits of raising debt. In the following sections, I formally present the problems of the firms and household at each point in time; and next define and characterize equilibria in this economy.

1.3.2 Firms

There are two sets of choices that firms need to make (i) financial decisions at $t = 0$ and (ii) real decisions regarding production at $t = 1$. At $t = 0$ firms choose the amount of debt to issue, and at $t = 1$ they determine the amount of output to produce and the input combination to utilize in production. In what follows, I present firms' problems throughout the timeline of the model.

Firms at $t = 1$

Each industry is comprised of a continuum of firms indexed by a uniform measure between zero and one.[4] Given a debt level $d_{fi} \geq 0$ chosen at time zero, firm $f \in [0,1]$ in industry $i \in \{1,...,n\}$ produces its output y_{fi} according to the production function

$$y_{fi} = Z_i\,(1-\delta \mathbb{1}^{\mathrm{b}}_{fi}) z_{fi}\, I^{\eta}_{fi}. \tag{1.18}$$

[4]The measure of firms in all industries is normalized to one. This is without loss of generality and for simplicity of notation.

Z_i is a random industry productivity shock and z_{fi} ~uniform$(0, \varphi_i)$ is an idiosyncratic firm productivity shock. I_{fi} is an intermediate good produced from inputs, and $\eta \in (0,1)$ is the elasticity of the final output with respect to the intermediate good. $\mathbb{1}^{\mathrm{b}}_{fi}$ is an indicator that takes the value of one if the firm becomes bankrupt and otherwise is zero. A Bankrupt firm continues to produce output with reduced productivity $(1-\delta)z_{fi}$ where $\delta \in [0,1]$ represents real bankruptcy costs. If $\delta = 0$ then bankruptcy does not affect firms' real operations, and on the other extreme if $\delta = 1$ a bankrupt firm exits the market without any production. In the event of bankruptcy, creditors take over the firm and incur financial restructuring costs equal to a fixed value of Δ_0 plus fraction Δ of the operating profit. This financial cost represents the money that creditors pay to lawyers, bankers, etc. to restructure the firm. In the model, the household inelastically provides these services and collects this financial cost from the bankrupt firms.

The real cost δ is the origin of the propagating effects of bankruptcies in the model. A higher bankruptcy rate in an industry leads to lower average productivity of the industry and alters the equilibrium prices of goods. These altered prices will affect other industries in the network. The financial cost parameters Δ and Δ_0 create the trade-off between debt benefits and bankruptcy costs without any effect on production.

Production Network. The intermediate good I_{fi} is produced according to the following CES production function

$$I_{fi}(x_{fi1}, ..., x_{fin}) = \left[\sum_{j=1}^{n} w_{ij}^{\frac{1}{\sigma}} x_{fij}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{1.19}$$

where x_{fij} is the input of good j in the production of firm f in industry i, σ is the elasticity of substitution, and $w_{ij} \geq 0$ is the relative weight of input j in production of the intermediate good i. These weights are normalized such that

$$\sum_{j=1}^{n} w_{ij} = 1 \qquad \text{for all } i \in \{1, ..., n\}.$$

$\mathbf{W} = [w_{ij}]_{n \times n}$ drives the interdependence of firms' cash flows across industries in the entire economy. If $w_{ij} > 0$ then firms in industry i are customers of the good produced in industry j, exposing them to any uncertainty in the price of good j. On the other hand, firms in industry j are also exposed to uncertainties in industry i that could affect their demand for good j. Notably, $\mathbf{W}$ determines the flow of goods in the economy and the exposure of industries to demand and supply market risks. Similar to before, $\mathbf{W}$ represents the adjacency matrix of a directed graph where industries are nodes and w_{ij} is the weight of the edge between i and j. Contrary to the endowment model, the direction of edge w_{ij} is from node i to j. This change of direction in the production model compared to the endowment model is to ensure that the directions of edges consistently show the flow of money across industries. The flows of goods in the production model are in the opposite direction of the ones represented by the directed graph.

Elasticity of Substitution. The parameter σ determines the firms' ability to substitute different inputs with one another in production. As discussed in detail in section 1.3.6, σ plays a significant role in determining the risks that firms face in their input market. With $\sigma < 1$ the input goods are complements in production and requiring more inputs makes firms riskier. If $\sigma > 1$ firms can substitute away from the goods with higher prices and being able to utilize more inputs in production makes the firm safer.

At $t = 1$, after realization of the productivity shocks and given prices of the goods, firms maximize their equity value V_{fi}^{equity1} where

$$V_{fi}^{\text{equity1}}(d_{fi}, z_{fi}; \mathbf{D}, \mathbf{Z}) = [\pi_{fi} - d_{fi} + \tau d_{fi}]\,(1 - \mathbb{1}_{fi}^{\text{b}}). \tag{1.20}$$

The equity value is the operating profit π_{fi} minus the financial liabilities d_{fi} plus the debt benefit τd_{fi} if the firm is solvent ($\mathbb{1}_{fi}^{\text{b}} = 0$), and otherwise is zero. In the aggregate level, V_{fi}^{equity1} is a function of the aggregate state of the economy $(\mathbf{D}, \mathbf{Z})$ where $\mathbf{D} = [D_i]_{n\times 1}$ is the vector of aggregate debt issued in each industry and $\mathbf{Z} = [Z_i]_{n\times 1}$ is the vector of realized industry shocks. V_{fi}^{equity1} also depends on fi's individual debt obligation d_{fi} and its realized idiosyncratic productivity shock z_{fi}.

On the other hand, the value to the creditors at $t = 1$ is d_{fi} or π_{fi} depending on whether the firm is solvent or bankrupt respectively.

$$V_{fi}^{\text{debt1}}(d_{fi}, z_{fi}; \mathbf{D}, \mathbf{Z}) = \pi_{fi}\mathbb{1}_{fi}^{\text{b}} + d_{fi}(1 - \mathbb{1}_{fi}^{\text{b}}) \tag{1.21}$$

Since all the debt decisions are made at $t = 0$, maximizing value at $t = 1$ (regardless of the potential change of ownership due to restructuring) is equivalent to maximizing profit. In this regard firm fi chooses $\{x_{fij}\}_{j=1}^{n}$ and I_{fi} to maximize its profit

$$\pi_{fi}(d_{fi}, z_{fi}; \mathbf{D}, \mathbf{Z}) = \max_{\substack{\{x_{fij}\}_{j=1}^{n} \\ I_{fi}}} \quad \max \left\{ \begin{array}{l} p_i Z_i (1 - \delta \mathbb{1}_{fi}^{\text{b}}) z_{fi} I_{fi}^{\eta} - \sum\limits_{j=1}^{n} p_j x_{fij} \\ - \left\{ \Delta_0 + \Delta \left(p_i Z_i (1-\delta) z_{fi} I_{fi}^{\eta} - \sum\limits_{j=1}^{n} p_j x_{fij} \right) \right\} \mathbb{1}_{fi}^{\text{b}} \end{array}, 0 \right\} \tag{1.22}$$

subject to (1.19) and

$$\mathbb{1}_{fi}^{\text{b}} = 1 \quad \text{if and only if} \quad p_i Z_i z_{fi} I_{fi}^{\eta} - \sum_{j=1}^{n} p_j x_{fij} < d_{fi} \tag{1.23}$$

where p_j is the price of good j. The first line of the objective function in (1.22) is the revenue minus the cost of the input goods used in production and the second line corresponds to the financial costs of bankruptcy and restructuring. The restructuring cost is comprised of a fixed cost Δ_0 and fraction Δ of the firm's revenue minus input costs. As mentioned before, if the firm becomes bankrupt then the creditors take over the firm and operate it to maximize its profit with reduced productivity. Finally, $\max\{.,0\}$ ensures that the operating profit is always non-negative. Moreover, the bankruptcy constraint (1.23) implies that firms are forced into bankruptcy if they are unable to repay their debt without becoming bankrupt.[5]

Firms at $t = 0$

At $t = 0$ firms decide how much debt to issue. The amount of issued debt affects the price of debt, as well as firms' equity value in the market. Firms have rational expectations about how their debt affects these quantities. Let $q_{fi}^{\text{equity0}}(d_{fi}, \mathbf{D})$ and $q_{fi}^{\text{debt0}}(d_{fi}, \mathbf{D})$ be firm fi's belief about market value of their equity and debt price at $t = 0$ if they issues d_{fi} in debt and the aggregate state is $\mathbf{D}$. Firms choose debt to maximize their total market value V_{fi}^0 where

$$V_{fi}^0(d_{fi}, \mathbf{D}) = q_{fi}^{\text{equity0}}(d_{fi}, \mathbf{D}) + q_{fi}^{\text{debt0}}(d_{fi}, \mathbf{D}) \times d_{fi} \tag{1.24}$$

$q_{fi}^{\text{equity0}}(d_{fi}, \mathbf{D})$ is the market value of equity,[6] $q_{fi}^{\text{debt0}}(d_{fi}, \mathbf{D}) \times d_{fi}$ is the market value of debt, and firm's total value is the sum of the two. In addition to d_{fi}, the firm's total value and asset prices also depend on the aggregate debt vector issued by the industries $\mathbf{D}$. The collected debt value from the creditors is paid to the equity holders as a dividend at $t = 0$.

1.3.3 Household

All the firms are owned and operated by a representative household with a time separable CRRA utility function

$$u(C) = \frac{C^{1-\gamma} - 1}{1 - \gamma} \tag{1.25}$$

where C is a consumption aggregator and γ is the relative risk aversion parameter. Following the literature, I further assume that the consumption aggregator C is

$$C(c_1, ..., c_n) = \prod_{i=1}^{n} c_i^{\alpha_i} \tag{1.26}$$

[5] With this constraint I abstract from strategic incentives for bankruptcy and assume that firms declare bankruptcy only if they are unable to fully repay their debt without bankruptcy.

[6] The total number of shares is normalized to 1.

where c_i is the household's consumption of good i and $\alpha_i \geq 0$ is the elasticity of consumption aggregator with respect to c_i. α is normalized such that

$$\sum_{i=1}^{n} \alpha_i = 1$$

The household is endowed with ω^0 units of the aggregate consumption good at time zero and owns the firms. In the next section, I present the household's decisions at $t = 0$ and 1.

Household at $t = 1$

At $t = 1$ the only decision the household needs to make is how much to consume from each good given their budget set. The household's wealth is comprised of his asset holdings in the equity and debt market in addition to a lump sum transfer of T due to the restructuring services that they provide bankrupt firms with. Hence conditional on the aggregate state of the economy $(\mathbf{D}, \mathbf{Z})$, and given the household's portfolio in the securities market $\{\{s_{fi}, d_{fi}\}_{f\in[0,1]}\}_{i=1}^{n}$, they solve the following optimization problem.

$$\begin{aligned} U^1(\{\{s_{fi}, d_{fi}\}_{f\in[0,1]}\}_{i=1}^{n}; \mathbf{D}, \mathbf{Z}) = \quad & \max_{C^1, \{c_i\}_{i=1}^{n}} \quad u(C^1) \\ & \text{s.t.} \ \sum_{i=1}^{n} p_i c_i \leq \sum_{i=1}^{n} \int \left(V_{fi}^{\text{equity1}} s_{fi} + V_{fi}^{\text{debt1}} \frac{d_{fi}}{D_i} \right) \mathrm{d}f \ \ + \mathrm{T} \\ & \qquad C^1 = \prod_{i=1}^{n} c_i^{\alpha_i} \end{aligned} \tag{1.27}$$

$p_i c_i$ is the household's spending on the consumption good produced by industry i. With some abuse of notation s_{fi} and d_{fi}/D_i respectively denote the number of fi's shares and the fraction of fi's total debt purchased by the household at $t = 0$. In equilibrium $s_{fi} = d_{fi}/D_i = 1$ since all the assets are held by the representative household.

Household at $t = 0$

After observing each firm's issued debt, in this period in addition to consumption, the household decides how much debt and equity to buy from each firm.

$$\begin{aligned} U^0(\mathbf{D}) = \quad & \max_{C^0, \{\{s_{fi}, d_{fi}\}_{f\in[0,1]}\}_{i=1}^{n}} \quad u(C^0) + \beta \mathbb{E}^0 \left[U^1(\{\{s_{fi}, d_{fi}\}_{f\in[0,1]}\}_{i=1}^{n}; \mathbf{D}, \mathbf{Z}) \right] \\ & \text{s.t.} \qquad C^0 + \sum_{i=1}^{n} \int \left(q_{fi}^{\text{equity0}}(D_i, \mathbf{D}) s_{fi} + q_{fi}^{\text{debt0}}(D_i, \mathbf{D}) d_{fi} \right) \mathrm{d}f \\ & \qquad \leq \omega^0 + \sum_{i=1}^{n} \int V_{fi}^{0}(D_i, \mathbf{D}) \mathrm{d}f \end{aligned} \tag{1.28}$$

Household's utility is the sum of utility from consumption C^0 at $t = 0$ and the expected discounted (by factor $\beta \in [0,1]$) utility from $t = 1$. The left-hand side of the budget constraint is the household's spending on the aggregate consumption C^0 (with the normalized price of one) and their asset purchases from each firm. On the other side, the household is endowed with ω^0 units of the aggregate consumption good and owns all fi firms in the economy with a total value of V_{fi}^0.

1.3.4 Equilibrium Characterization

The definition of equilibria in this economy is mostly standard. In equilibrium, at each point in time, all the firms maximize their value; the representative household maximizes its utility; all the goods and securities markets clear; and all the beliefs are rational. I continue to focus only on symmetric equilibria in which all firms within an industry choose the same level of debt, or $d_{fi} = D_i$ for all fi.

Firm fi's total value V_{fi}^0 is the expected discounted cash flow that fi generates at $t = 1$, and the equilibrium debt D_i^* maximizes this time zero value. To characterize equilibrium debt vector $\mathbf{D}^*$, one needs to first determine firms' cash flow at $t = 1$ conditional on their debt choices, and the realized productivity shocks.

As previously argued, at $t = 1$ firm fi's stakeholders maximize their operating profits according to (1.22). For any outstanding debt, goods price vector $\mathbf{p}$, and realized industry productivity Z_i, this profit is increasing in fi's idiosyncratic productivity. As a result, the bankruptcy state admits a cutoff rule as follows.

Lemma 1. *At $t = 1$ firm fi becomes bankrupt if its idiosyncratic productivity shock is lower than z_{fi}^* where*

$$z_{fi}^{*\,\frac{1}{1-\eta}} = \frac{d_{fi}}{(1-\eta)\left[\eta^{\eta} p_i Z_i \left(\sum_{j=1}^{n} \frac{w_{ij}}{p_j^{\sigma-1}}\right)^{\frac{\eta}{\sigma-1}}\right]^{\frac{1}{1-\eta}}}. \tag{1.29}$$

The cutoff value z_{fi}^* is such that $\pi_{fi}(d_{fi}, z_{fi}^*; \mathbf{D}, \mathbf{Z}) = d_{fi}$, below which the firm is unable to repay the par value of its debt to creditors. z_{fi}^* is increasing in d_{fi} consistent with the intuition that firms with larger d_{fi} require a larger productivity shock to fulfill their debt obligation in full and avoid bankruptcy. The denominator of (1.29) is equal to a baseline operating profit of the firm when $z_{fi} = 1$ and the firm is solvent. Any z_{fi} scales this baseline profit (with power $1/1-\eta$), hence the bankruptcy cutoff z_{fi}^* is decreasing in this profit.

Similarly, there is a lower bound $\underline{z}_{fi} = \sup\{z | \pi_{fi}(d_{fi}, z; \mathbf{D}, \mathbf{Z}) = 0\}$ below which firm's value is zero. For any productivity lower than $\underline{z}_{fi}$ firm's profit from production fails to meet the fixed and variable costs of restructuring, resulting in nil value for the firm. $\underline{z}_{fi}$ and z_{fi}^* partition the support of z_{fi} into three intervals: (i) $[0, \underline{z}_{fi}]$ where the firm is bankrupt with zero residual value, (ii) $(\underline{z}_{fi}, z_{fi}^*)$

where the firm is bankrupt with positive value, and (iii) $[z^*_{fi}, \varphi_i]$ where the firm is solvent.

From (1.20) and (1.21) the total value of fi at $t = 1$ is

$$V^1_{fi} = V^{\text{equity1}}_{fi} + V^{\text{debt1}}_{fi} = \pi_{fi} + \tau d_{fi}(1 - \mathbb{1}^{\text{b}}_{fi}). \tag{1.30}$$

Let $EV^1_{fi}(d_{fi}; \mathbf{D}, \mathbf{Z})$ denote the expected value of V^1_{fi} with respect to z_{fi}. From the above partitioning of z_{fi}'s support and the uniform distribution of z_{fi} on $(0, \varphi_i)$

$$\begin{aligned} EV^1_{fi}(d_{fi}; \mathbf{D}, \mathbf{Z}) &= \int_0^{\varphi_i} V^1_{fi}(d_{fi}, z_{fi}; \mathbf{D}, \mathbf{Z}) \frac{1}{\varphi_i} \mathrm{d}z_{fi} \\ &= \int_{\underline{z}_{fi}}^{z^*_{fi}} \pi_{fi}(d_{fi}, z_{fi}; \mathbf{D}, \mathbf{Z}) \frac{1}{\varphi_i} \mathrm{d}z_{fi} + \int_{z^*_{fi}}^{\varphi_i} \left[\pi_{fi}(d_{fi}, z_{fi}; \mathbf{D}, \mathbf{Z}) + \tau d_{fi}\right] \frac{1}{\varphi_i} \mathrm{d}z_{fi}. \end{aligned} \tag{1.31}$$

The first integral corresponds to the generated cash flow in the bankruptcy region and is equal to the production profits minus the fixed and variable restructuring costs. The second integral is the expected value of production profits plus the debt benefits τd_{fi} in the solvency region. As mentioned above, the total value of fi at $t = 0$ or V^0_{fi} is equal to the expected discounted value of EV^1_{fi} and firms choose their debt to maximize it.

Market Structure. Similar to the endowment model, I study equilibria in competitive and monopolistic debt market structures. With a competitive debt market, firms behave atomistically and do not internalize their effect on the aggregate debt of their industry. On the other hand, in the market with strategic firms, all firms within each industry take the equilibrium condition $d_{fi} = D_i$ into account in their decision-making. In both cases the investors price assets accordingly and consistent with the market structure. Furthermore, in both cases, I assume that the goods' markets are competitive and individual firms are all price takers in their input and output markets.[7] The prevailing market price vector $\mathbf{p}$ clears the goods markets and depends on the aggregate state $(\mathbf{D}, \mathbf{Z})$.

Definition 3. *$\mathbf{D}^* = [D^*_i]_{n\times 1}$ is an equilibrium if and only if for all $i = 1, \ldots, n$*

$$D^*_i \in \arg\max_{d_{fi}} \, (1 - \Theta)\mathbb{E}^0_{\mathbb{Q}}\left[EV^1_{fi}(d_{fi}, \mathbf{D}^*, \mathbf{Z})\right] + \Theta\mathbb{E}^0_{\mathbb{Q}}\left[EV^1_{fi}(d_{fi}; (d_{fi}, \mathbf{D}^*_{-i}), \mathbf{Z})\right]$$

where $\Theta \in \{0, 1\}$ and the expectations are taken with respect to $\mathbf{Z}$ and under the pricing measure $\mathbb{Q}$.

Similar to the endowment model, $\Theta = 0$ represents the competitive debt market and $\Theta = 1$ corresponds to the market with strategic firms. The pricing measure $\mathbb{Q}$ determines how investors discount future random cash flows in different states of the economy. This measure crucially depends on investors' attitudes towards risk and is determined by the household's utility function.

[7] This is a common assumption in the literature.

Before analyzing the general case, it is informative to analyze debt choices under a special case of a partial equilibrium setting and risk-neutral investors with exogenous price and productivity distributions.

Proposition 3. *if investors are risk neutral and the joint distribution of* $\mathbf{p}$ *and* Z_i *is exogenous and independent of* d^*_{fi}*, then firm* fi *chooses its debt* d^*_{fi} *such that:*

$$\mathbb{E}^0\left[\frac{\tau}{\tau+(1-\eta)\left[1+\tau+\frac{\Delta_0}{d^*_{fi}}-(1-\Delta)(1-\delta)^{\frac{1}{1-\eta}}\right]}-\frac{z^*_{fi}}{\varphi_i}\right]=0 \qquad (1.32)$$

where the expectation is taken with respect to Z_i *and* $\mathbf{p}$*.*

If investors are risk neutral then they do not discount future cash flows for risk. In addition if the joint distribution of $(\mathbf{p}, Z_i)$ does not depend on d_{fi} then V^0_{fi} is equal to the expected value of equation (1.31) with respect to $(\mathbf{p}, Z_i)$.[8] Setting the expected partial derivative of (1.31) results in the optimality condition (1.32). z^*_{fi}/φ is the conditional probability of bankruptcy for any realized price vector $\mathbf{p}$ and industry productivity shock Z_i. To maximize value, fi increases debt up to a point where the expected conditional probability of default is equal to the first term within the expectation. This term is increasing in the marginal benefit of debt τ and decreasing in the cost parameters δ, Δ_0, and Δ. A higher marginal benefit or lower costs of bankruptcy leads to a higher expected probability of debt or higher d^*_{fi}.

Pricing Measure and Stochastic Discount Factor

The household's utility function determines the stochastic discount factor (SDF) in this economy that can be used to price any asset at $t = 0$. With CRRA household

$$\text{SDF}(\mathbf{D}, \mathbf{Z}) = \beta\left(\frac{C^1(\mathbf{D}, \mathbf{Z})}{C^0}\right)^{-\gamma} \qquad (1.33)$$

where C^0 and C^1 are household's consumption from the aggregate consumption good at $t = 0$ and 1 respectively. In equilibrium, $C^0 = \omega^0$ and C^1 is a function of the aggregate state and outcome of the equilibrium in the goods market. The equilibrium at $t = 1$ does not have a closed form expression, however, it can be characterized in terms of a system of equations. Proposition 4 fully characterizes equilibria in the model.

Proposition 4. *The following system of equations characterizes the industries equilibrium values*

[8] In the general equilibrium model if $\gamma = 0$ and $\Theta = 0$ these conditions are all satisfied.

at time $t = 1$ when the aggregate state is $(\mathbf{D}, \mathbf{Z})$.

$$\alpha^\top \log(\mathbf{p}) = \alpha^\top \log(\alpha) \tag{1.34}$$

$$M = \sum_{i=1}^{n} p_i y_i \tag{1.35}$$

$$[p_i y_i]_{n\times 1} = (1-\eta) M \left[\mathbb{I} - \eta \widetilde{\mathbf{W}}\right]^{-1} \alpha \tag{1.36}$$

$$\widetilde{\mathbf{W}} = [\widetilde{w}_{ki}]_{n\times n} \text{ where } \widetilde{w}_{ki} = \frac{\frac{w_{ik}}{p_k^{\sigma-1}}}{\sum_{j=1}^{n} \frac{w_{ji}}{p_j^{\sigma-1}}} \tag{1.37}$$

$$y_i = \left[\eta^\eta p_i^\eta \overline{Z}_i \left(\sum_{j=1}^{n} w_{ij} p_j^{1-\sigma}\right)^{\frac{\eta}{\sigma-1}}\right]^{\frac{1}{1-\eta}} \tag{1.38}$$

$$\overline{Z}_i = Z_i \left[\frac{1-\eta}{(2-\eta)\varphi_i}\left(\varphi_i^{\frac{2-\eta}{1-\eta}} - z_i^{*\frac{2-\eta}{1-\eta}}\right)\right]^{1-\eta} \tag{1.39}$$

$$z_i^* = \frac{D_i^{1-\eta}}{(1-\eta)\left[\eta^\eta p_i Z_i \left(\sum_{j=1}^{n} \frac{w_{ij}}{p_j^{\sigma-1}}\right)^{\frac{\eta}{\sigma-1}}\right]} \tag{1.40}$$

$$\text{SDF} = \beta \left(\frac{(1-\eta)M}{\omega^0}\right)^{-\gamma} \tag{1.41}$$

In addition, the equilibrium $\mathbf{D}^*$ *solves*

$$\frac{\partial}{\partial d_{fi}} \mathbb{E}^0 \left[\text{SDF}(\mathbf{D}^*, \mathbf{Z})\ EV_{fi}^1(d_{fi}, \mathbf{D}^*, \mathbf{Z})\right]\Big|_{d_{fi}=D_i^*} + \tag{1.42}$$

$$\Theta \frac{\partial}{\partial D_i} \mathbb{E}^0 \left[\text{SDF}((D_i, \mathbf{D}_{-i}^*), \mathbf{Z})\ EV_{fi}^1(d_{fi}; (D_i, \mathbf{D}_{-i}^*), \mathbf{Z})\right]\Big|_{d_{fi}=D_i=D_i^*} \leq 0 \ \ (\textit{with equality if } D_i^* > 0)$$

where the expectation is taken over the distribution of $\mathbf{Z}$.

The set of equations from (1.34) to (1.41) characterize the equilibrium at $t = 1$ for any given state $(\mathbf{D}, \mathbf{Z})$. They summarize a price normalization, optimality conditions for the firms and household, and market clearing conditions.

Equation (1.34) is a price normalization in the goods market that sets the aggregate consumption good's price to one. (1.35) defines M as the aggregate production value of all industries combined. (1.36) determines each industries output as shares $(1-\eta)[\mathbb{I} - \eta\widetilde{\mathbf{W}}]^{-1}\alpha$ of the aggregate production M, where $\widetilde{\mathbf{W}}$ is the production network matrix $\mathbf{W}$ weighted by prices to account for substitution in inputs. (1.40) is the cutoff value for the idiosyncratic productivity shock below which firms become bankrupt. $\overline{Z}_i$ in equation (1.39) is the average productivity of firms in industry i after accounting for the reduction $1-\delta$ in the productivity of the bankrupt firms. Equation (1.38) determines industry i's aggregate output y_i as a function of the production network and industry's average productivity $\overline{Z}_i$. Finally (1.41) pins down the stochastic discount factor which depends on the time discount parameter β, aggregate consumption $C^1 = (1-\eta)M$ and $C^0 = \omega^0$, and the household's risk aversion parameter γ.

Equation (1.42) is the first order condition of firms' value maximization problem at $t = 0$, where the pricing measure is explicitly substituted for using the SDF. The effect of a marginal increase in d_{fi} on fi's value can be split into two terms. The first term corresponds to the partial equilibrium effect where the equilibrium condition $d^*_{fi} = D^*_i$ is not accounted for. In other words, the first term is the marginal increase in fi's value if the aggregate state is fixed to its equilibrium value $(\mathbf{D}^*, \mathbf{Z})$. With a fixed aggregate state, the firm's cash flow does not depend on d_{fi} beyond the costs associated with bankruptcy. In particular, d_{fi} affects the firm neither in its input and output market via a change in the price vector $\mathbf{p}$, nor the discount factor $\text{SDF}(\mathbf{D}^*, \mathbf{Z})$. The standard trade-off between the costs of bankruptcy and benefits of debt applies. With atomistic firms, this term alone determines the equilibrium value of debt for each industry. In fact if $\Theta = 1$ then the optimality condition (1.42) reduces to

$$\mathbb{E}^0 \left[\text{SDF} \times \left(\frac{\tau}{\tau + (1-\eta)\left[1 + \tau + \frac{\Delta_0}{d^*_{fi}} - (1-\Delta)(1-\delta)^{\frac{1}{1-\eta}}\right]} - \frac{z^*_{fi}}{\varphi_i} \right) \right] = 0 \qquad (1.43)$$

which is equivalent to (1.32) when investors' attitudes towards risk are accounted for.

The second term in (1.42) represents the case where firms internalize the effect of their debt choice on their own industry. As seen in the endowment model, a marginal increase in d_{fi} also increases D_i by the same amount. An increase in D_i increases the fraction of bankrupt firms in industry i, which in turn brings about a reduction in the average productivity of industry i. Depending on i's position in the production network, lower productivity in industry i changes the prevailing price vector $\mathbf{p}$. These altered prices feed back into fi's value by changing its profit from production, as well as altering the discount factor of the investors. Similar to the endowment model, this term is only present when all the firms in each industry behave strategically.

The role of production network $\mathbf{W}$ in shaping the distribution of D^*_i across industries is again twofold. First, $\mathbf{W}$ determines the distribution of $\mathbf{p}$ which in turn drives the distribution of cash flows in each industry (baseline effect). Industries with larger probabilities of observing low cash flows (larger negative tail risk) due to their position in the production network are more prone to bankruptcy and issue less debt. The second effect is due to the propagation of bankruptcy shocks in the network (propagation effect). If firms within an industry are adversely affected by bankruptcies in other industries to a larger extent due to their position in the entire network, they issue less debt. Defining the position of industries in the entire network is not a straightforward task, as $\mathbf{D}^*$ depends on the entire $\mathbf{W}$ in principle. I revisit this issue later in the empirical section of the paper by introducing a well-known proxy variable for nodes' position in a network, known as eigenvector centrality.

1.3.5 Debt Externality

As seen before in the endowment model, the decentralized market distribution of debt is not necessarily optimal from the social planner's perspective. This is also the case here in the market equilibrium. Firms do not recognize their effect on other industries in their decision-making. Definition 4 defines the optimal debt distribution from the planner's point of view as the distribution that maximizes the economy's aggregate value at $t = 0$.

Definition 4. $\mathbf{D}^{\text{sp}} = [D_i^{\text{sp}}]_{n\times 1}$ *is the social planner's optimal debt choice if and only if*

$$\mathbf{D}^{\text{sp}} \in \arg\max_{\mathbf{D}} \sum_{j=1}^{n} \mathbb{E}^0 \left[\text{SDF}(\mathbf{D},\mathbf{Z}) \times EV_{fj}^1(D_j;\mathbf{D},\mathbf{Z})\right] \tag{1.44}$$

Taking the first order condition with respect to D_i results in the following optimality condition:

$$\begin{aligned}
&\frac{\partial}{\partial d_{fi}}\mathbb{E}^0\left[\text{SDF}(\mathbf{D}^*,\mathbf{Z})\; EV_{fi}^1(d_{fi},\mathbf{D}^*,\mathbf{Z})\right]\Bigg|_{d_{fi}=D_i^{\text{sp}}} + \\
&\quad \frac{\partial}{\partial D_i}\mathbb{E}^0\left[\text{SDF}((D_i,\mathbf{D}_{-i}^{\text{sp}}),\mathbf{Z})\; EV_{fi}^1(d_{fi};(D_i,\mathbf{D}_{-i}^{\text{sp}}),\mathbf{Z})\right]\Bigg|_{d_{fi}=D_i=D_i^{\text{sp}}} + \\
&\quad \sum_{j\neq i}\frac{\partial}{\partial D_i}\mathbb{E}^0\left[\text{SDF}((D_i,\mathbf{D}_{-i}^{\text{sp}}),\mathbf{Z})\; EV_{fj}^1(D_j;(D_i,\mathbf{D}_{-i}^{\text{sp}}),\mathbf{Z})\right]\Bigg|_{D_i=D_i^{\text{sp}}} \leq 0 \;\; (\text{with equality if } D_i^{\text{sp}} > 0)
\end{aligned} \tag{1.45}$$

The first term in (1.45) is identical to (1.42) which is the partial equilibrium effect of D_i on i's firms' value. The second term is the general equilibrium effect of D_i on i's firms which is present in (1.42) when $\Theta = 1$. Finally, the last term corresponds to the equilibrium effect of D_i on the value of the firms outside industry i which only appears in the planner's optimality condition. The second and last lines are the source of externality in the model.

In the next section, I numerically solve an illustrative example with three industries to further discuss the implications of the model.

1.3.6 Illustrative Example 2

The top panels in figure 1.4 plot the production network in two economies with three sectors. The edge connecting industry i to j is the weight of input good j in production of i and the directions

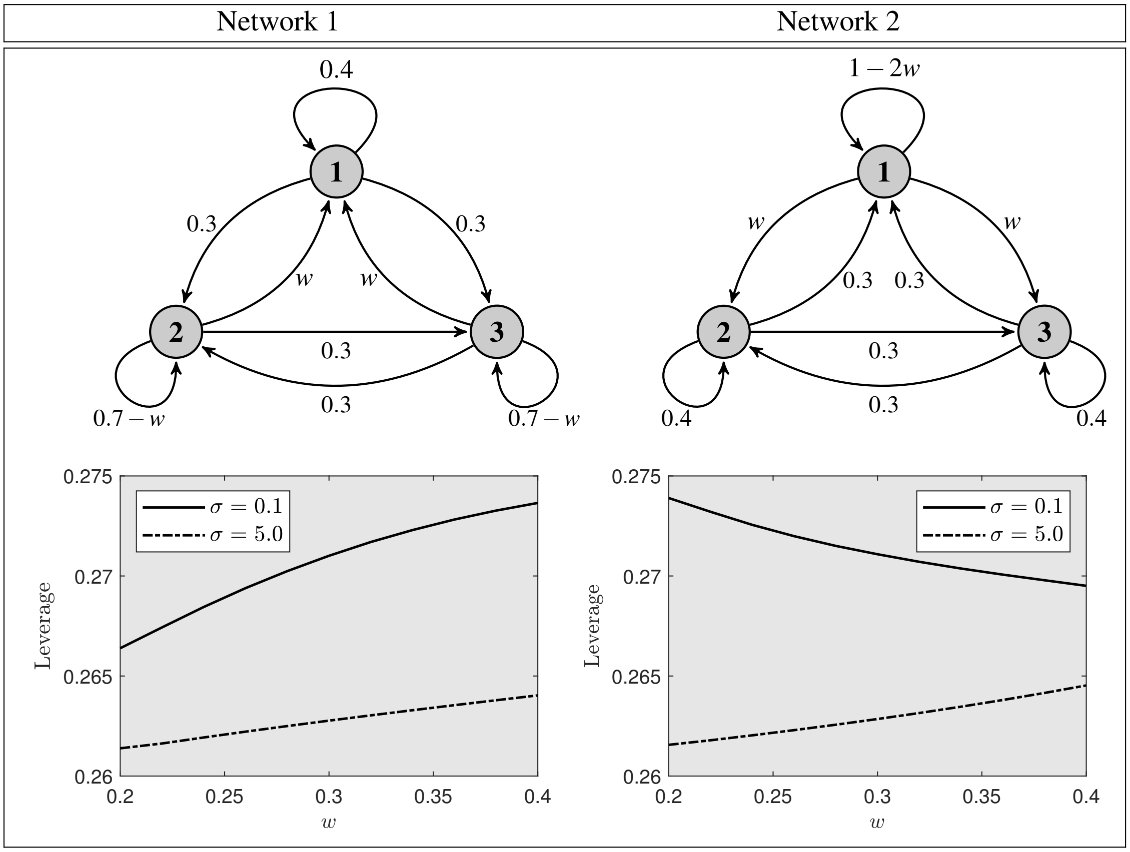


Figure 1.4: An illustrative example in the production model, part 1.

of arrows are from customers to suppliers (the direction of cash flows). In network 1, the production weights of industry 1 are fixed to $(0.4, 0.3, 0.3)$ from goods 1,2, and 3 respectively. The production weights for industry 2 and 3 are $(w, 0.7-w, 0.3)$ and $(w, 0.3, 0.7-w)$ respectively where $w \in [0.2, 0.4]$. In this network, industry 1's position in the input market is fixed and as w increases from 0.2 to 0.4, industry 2 and 3 become more reliant on 1's good in their production. Intuitively 1 becomes more central in the cash flow network as w increases.[9]

In network 2, industry 1's position in the output market is fixed (1's good appears in 2 and 3's production with fixed weights of 0.3). However, as w increases, 1 becomes more reliant on inputs from 2 and 3 in its production and becomes less central in the cash flow network. For each network in the first part of the example, I solve for equilibrium with atomistic firms $\Theta = 0$ with two values of $\sigma = 0.1$ and 5 separately. The rest of the model parameters are identical for both networks and are listed in the top left panel of figure 1.5.

The bottom panels in figure 1.4 plot the resulting equilibrium leverage choice of firms in indus-

[9] The exact notion of node centrality used here is defined later in the paper. Intuitively, more central in this context is equivalent to playing a more significant role in the production of inputs in the economy.

try 1 as a function of w for the corresponding networks. In network 1, as w increases, industry 1 plays an increasingly larger role as a supplier of 2 and 3 and becomes less risky and prone to the productivity shocks in either of its customers. Thus its leverage is increasing in w. This is the case for both $\sigma = 0.1$ and 5. However, if $\sigma = 0.1$ then 2 and 3 are even more reliant on 1's good because of their inability to substitute 1's good for other inputs. This is reflected in the larger slope of the leverage graph for $\sigma = 0.1$ relative to $\sigma = 5$.

In network 2, as w increases, firms in industry 1 become more safe or risky depending on the elasticity of substitution σ. If $\sigma < 1$ they face more risk as they require inputs from 2 and 3 with larger weight and cannot substitute for these inputs when their prices are high. As a result, with $\sigma = 0.1$ the firms' leverage decreases with w. On the other hand when $\sigma > 1$, being able to use more inputs that are gross substitutes in production lowers the firms' risk and increases their leverage in equilibrium.

To sum up, being a more central producer always makes an industry safer. On the other hand, being a more central consumer makes the industry safer (riskier) only if it is relatively easy (difficult) to substitute across inputs. Therefore, firms that are more central may be safer or riskier, depending on which of these is quantitatively more important.

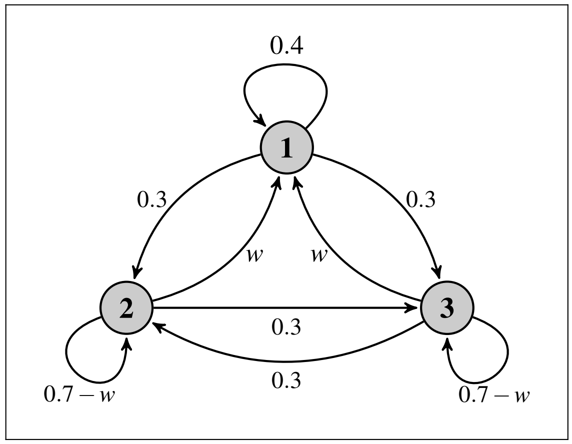
Figure 1.5 summarizes the second part of this section in which I further focus on network 1 and its resulting equilibrium outcomes. This network is symmetric for industry 2 and 3 and w determines the relative position and riskiness of 1 over 2 and 3 in the input market. Recall that for this network, a higher w makes industry 1 more central because industries 2 and 3 become more dependent on it. This makes industry 1 safer, and with $\Theta = 0$, industry 1's leverage and debt price are increasing with w, while the opposite occurs for 2 and 3.

When $\Theta = 1$ strategic behavior is more pronounced and dominates firms' financial behavior. With strategic players, firms internalize the effect of their financing on the aggregate state and hence product prices and operating profits. This provides an incentive for firms to affect the market price of their goods through their debt choice. All else equal, a higher debt leads to a higher share of bankrupt firms and lower average productivity of an industry due to productivity reduction in the bankrupt firms. As a result, all firms have a significantly higher leverage ratio compared to the competitive case $\Theta = 0$. Furthermore, larger w undermines 1's strategic position in the network and leads to higher equilibrium leverage for its firms. Since leverages is partially used as a strategic tool to affect market outcomes, and there is strategic complementary in industries' choice of debt, in this case, the price of debt moves opposite to the leverage ratio by w, namely the debt price is increasing in w for industry 1 and decreasing for 2 and 3. In addition, the price is always lower for all firms compared to the case $\Theta = 0$.

Finally, due to the aforementioned externalities, the social planner chooses the lowest leverage for all firms among the three equilibrium types. Moreover, the planer does not fall into the debt race due to the strategic complementary observed in $\Theta = 1$ and the leverage profile is similar to $\Theta = 0$.

Parameter	Value
γ	3
β	1
r_f	0
τ	0.6
η	0.7
σ	0.1
α	$[1/3, 1/3, 1/3]^\top$
φ	$[3, 3, 3]^\top$
δ	0.3
Δ_0	0.002
Δ	0.3
Z_i	uniform($\{0.8, 0.9, 1.0, 1.1, 1.2\}$)

Production Network



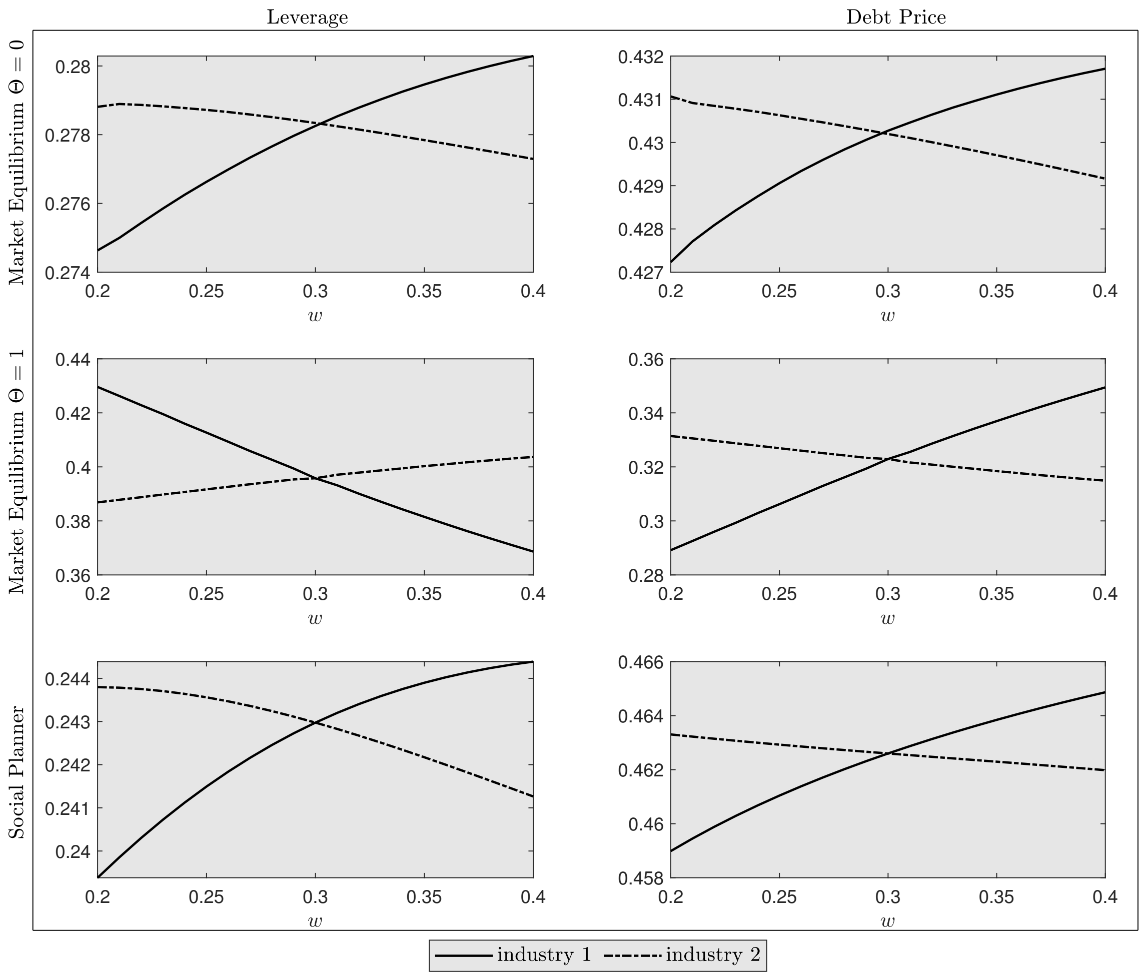


Figure 1.5: An illustrative example in the production model, part 2.

1.4 Calibration and Empirical Evidence

In this section, I provide empirical evidence around the role of industries' position in the production network in determining their firms' capital structure and cost of debt. To this end, first I calibrate the production network parameters to match the observed input-output accounts in 2019 for the US economy with 15 aggregated industries. The market equilibrium in this calibrated economy further shows that firms' leverage choice and bond prices depend on their position in the production network. In particular, the model predicts that firms in more central industries issue more debt in equilibrium and pay less to borrow. In addition, in the absence of any non-network heterogeneity, the eigenvector centrality, defined as the eigenvector corresponding to the largest eigenvalue of $\mathbf{W}$, explains a significant portion of the cross-sectional variation in the leverage ratios and bond prices. Then I empirically document these theoretical findings using microdata with more granular industry definitions.

1.4.1 Data

Following Ahern (2013) I construct the Social Accounting Matrix (SAM) for the US economy using the input-output and National Income and Product Accounts (NIPA) tables from the online database of the Bureau of Economic Analysis (BEA). A SAM represents the circular flow of transactions between all economic agents in the economy including the production activities (industries), factors of production (capital and labor), and institutions (household, foreign sector, and government). A SAM is a square matrix in which each row and column represent an agent, and every element of this matrix provides the receipts of the row agent from the column, or equivalently the expenditures of the column agent on the row.

BEA updates the detailed IO tables once in 5 years and the aggregated tables annually. Due to computational limitations, I use the aggregated IO tables in 2019 with 15 industries in the calibration for which I numerically solve the model only for the competitive case of $\Theta = 0$. Then I use the detailed tables with more than 400 industries in the empirical investigation. I construct one SAM matrix for every 5 years from 1980 to 2015. I also use Compustat for firms' financial data and Mergent Fixed Income Securities Database (FISD) for corporate bonds and follow Leary and Roberts (2014) to construct the other financial variables from the Compustat variables.

1.4.2 Calibration

Table 1.4.2 is the industry portion of the SAM constructed for 2019 that I use in this section. The calibration exercise is done in two steps:

1. I set $\sigma = 0.1$ and $\gamma = 3$ and for any set of other parameters in the model, namely $\mathbf{W}, \alpha$, and η, I solve a more flexible version of the model at $t = 1$ where the return to scale parameter η is industry-specific, however, there is no bankruptcy and uncertainty in this step. I iterate

over these parameters until the resulting SAM from the model matches the observed SAM constructed from the data. The objective of this step is to estimate the production network that is consistent with the observed SAM. Figure 1.6 plots the resulting production network from this step.

2. I eliminate all non-network heterogeneity in the parameters by setting $\alpha_i = 1/n$ and homogeneous $\eta = 0.7$ for all industries. Moreover I set $\delta = 0.1$, $\Delta_0 = 0.0001$, $\Delta = 0.3$, $\tau = 0.6$, $\varphi_i = 3$, and $Z_i \sim$uniform$(\{0.9, 1, 1.1\})$. For this set of parameters, I solve the full model and report the resulting equilibrium, however, the qualitative results are robust to alternative parametrization.

Figure 1.7 plots the resulting equilibrium leverage ratio and bond price for all the industries against their eigenvector centrality in the network. The only heterogeneity across industries must be due to their differential position in the input-output network. Industry position in the network, captured by the eigenvector centrality here, plays a key role in determining firms' equilibrium leverage and bond prices. More central firms have higher leverage ratios and sell their bond for a higher price. Moreover, the eigenvector centrality alone explains a remarkable portion of the variation in the leverage ratio and bond price (with R^2 of 0.93 and 0.94 respectively). In the next section, I empirically investigate these model predictions with more granular data from BEA.

1.4.3 Empirical Evidence

As a proxy for firms' position in the production networks, consistent with the literature and suggested by the calibration exercise, I use the eigenvector centrality of the industry that the firm belongs to in the network. First I normalize columns of SAM to sum to one, hence each entry represents the share of the column agent's expenditures on the row agents. Then the centrality of each industry is defined as the corresponding component of the eigenvector corresponding to the largest eigenvalue of the normalized SAM.[10] I use the natural logarithm of this eigenvector as the position variable in the empirical analysis.

The eigenvector centrality of each industry is defined recursively as a weighted sum of the centralities of its customers, and the weights are the normalized SAM elements. Intuitively, industry i is more central if it has customers that have high centrality themselves. In addition, if industry j buys a larger share of its inputs from i, then j's centrality will have a larger impact on that of i.

Table 1.1 provides the regression results of book and market leverage ratios onto industry centrality. I control for firm size, tangibility, profitability, and market-to-book ratio as standard control variables for leverage regressions in the literature. As evident from the table, more central firms

[10] This normalization guarantees that the spectral radius (the largest eigenvalue) of the normalized SAM is equal to one and the corresponding eigenvector is real-valued.

Industry Code	11	21	22	23	31G	42	44RT	48TW	51	FIRE	PROF	6	7	81	G
11	104.46	2.2	4.1	1.67	76.89	44.57	1.46	13.44	0.83	36.81	3.98	0.02	0.62	0.8	3.31
21	0.15	49.95	8.45	4.35	72.22	16.25	0.96	18.34	5.96	48.45	61.89	0	0.95	1.17	5.63
22	0	31.3	26.16	8.18	26.34	7.59	2.23	30.55	6.14	17.79	41.83	0.24	3.33	0.64	17.6
23	1.92	19.48	4.27	0.22	383.58	86.26	84.48	31.34	13.31	54.79	83.14	0.01	1.09	9.11	5.34
31G	263.71	331.84	46.78	14.23	1911	382	20.91	201.16	38.33	82.85	283.02	0.09	11.25	21.91	49.63
42	0.03	0.29	12.42	1.9	62.11	53.98	1.08	73.97	37.63	169.1	263.43	1.92	9.25	28.63	25.71
44RT	1.85	0.23	28.42	4.31	50.73	29.84	8.01	76.22	38.32	213.59	207.22	6.26	9.05	19.38	23.42
48TW	0.16	0.7	12.82	6.81	112.49	28.35	13.39	150.07	20.32	140.85	101.85	0.38	20.05	23.52	46.05
51	0.06	1.04	5.71	2.66	93.67	24.75	1.04	25.11	259.11	87.51	243.81	0.28	49.02	10.43	22.22
FIRE	0.11	0.4	95	158.72	82.23	28.49	10.29	43.1	101.41	1391.13	575.95	0.03	84.71	41.25	101.72
PROF	2.58	2.07	13.74	1.77	148.77	36.01	6.94	62.16	146.79	313.02	668.71	1.41	76.58	33.99	29.51
6	0.26	1.23	15.86	1.47	195.96	60.51	1.72	28.36	57.69	338.31	296.75	38.2	60.98	29.72	24.99
7	6.78	1.51	25.4	3.13	102.32	28.88	16.78	14.71	27.85	142.46	174.77	2.07	39.39	17.34	19.26
81	0.19	0.84	3.16	2.87	61.39	13.41	12.4	6.49	18.54	74.08	51.18	3.39	7.28	9.1	8.06
G	5.31	27.95	23.85	102.45	422.53	78.86	3.48	79.65	123.52	160.33	292.35	33.16	27.01	36.9	37.63

The values are in billions of US dollars.

Industry Code	Industry Name
11	Agriculture, forestry, fishing, and hunting
21	Mining
22	Utilities
23	Construction
31G	Manufacturing
42	Wholesale trade
44RT	Retail trade
48TW	Transportation and warehousing
51	Information
FIRE	Finance, insurance, real estate, rental, and leasing
PROF	Professional and business services
6	Educational services, health care, and social assistance
7	Arts, entertainment, recreation, accommodation, and food services
81	Other services, except government
G	Government

Table 1.1: Social Accounting Matrix (SAM) for 15 aggregated industries in 2019.

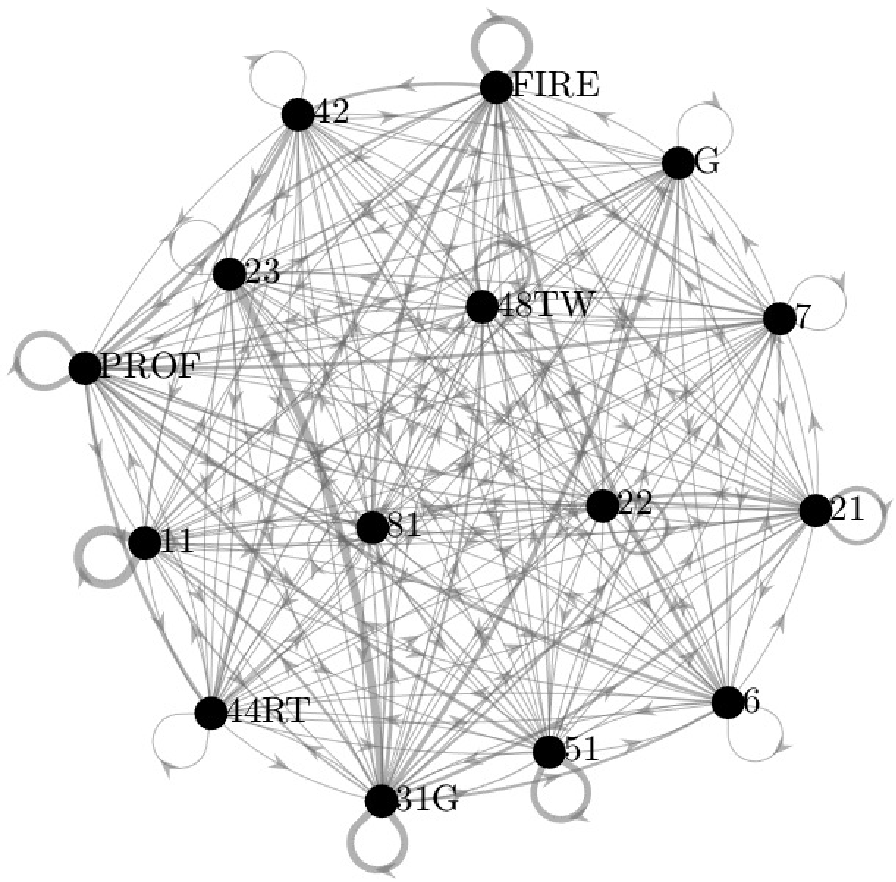


Figure 1.6: Calibrated production network for the US economy in 2019. Each node represents an aggregated sector and edges represent the production weights of input goods. The directions of edges represent the directions of the flows of goods and the thicknesses are proportional to the weights. For the exact description of the sectors and the underlying SAM refer to table 1.4.2.

tend to have higher market and book leverage ratios. The positive association is statistically significant and a one percent increase in a firm's centrality corresponds to a 0.01 percent increase in its leverage ratio.

The next result is related to the bond prices. In figure 1.8, the left panel plots the average centrality for different credit classes of corporate bonds from their S&P ratings. The figure suggests that firms' centrality is positively associated with the credit quality of their bond. On average higher quality debt is issued by more central firms. Similar patterns arise from Moody's and Fitch ratings as well. In the right panel, the cumulative density function of corporate bond spread is plotted for two classes of firms, namely the firms with log centrality smaller than the sample mean and the firms with larger than average centrality. The spread distribution for low centrality firms first-order

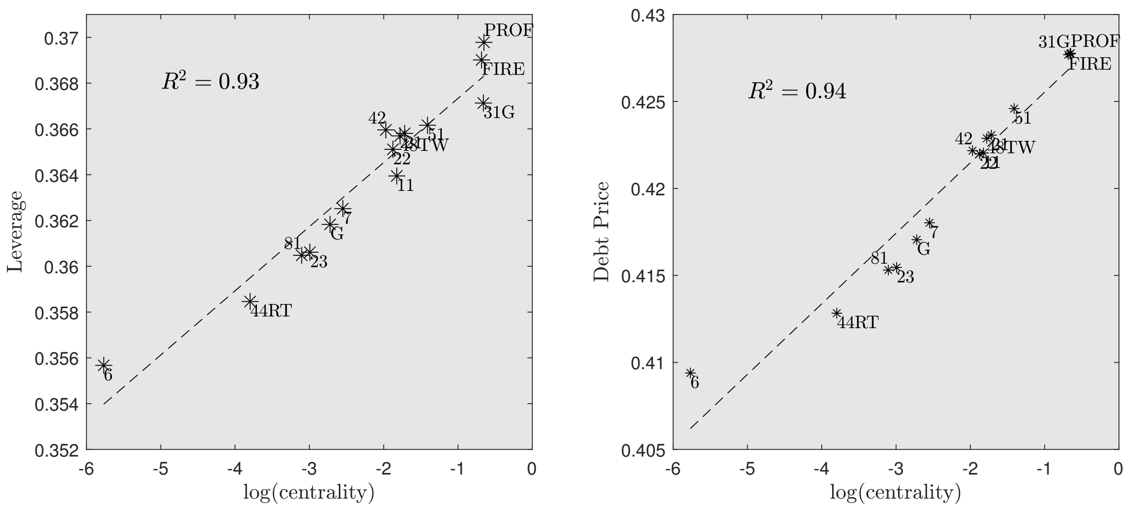


Figure 1.7: Equilibrium leverage and bond price for the calibrated network.

	BookLev	BookLev	BookLev	MarketLev	MarketLev	MarketLev
logCentrality	0.0118		0.00968	0.0234		0.0149
	(4.52)		(3.94)	(7.94)		(6.40)
Size		0.00231	0.00169		0.00800	0.00705
		(2.15)	(1.48)		(7.26)	(6.24)
Tangibility		0.220	0.215		0.209	0.203
		(13.16)	(12.45)		(11.73)	(11.26)
Profitability		-0.177	-0.175		-0.104	-0.101
		(-19.56)	(-19.24)		(-12.28)	(-12.20)
Market2Book		0.00117	0.00149		-0.0495	-0.0490
		(0.71)	(0.90)		(-24.97)	(-25.63)
N	112137	112137	112137	112137	112137	112137
R^2	0.007	0.070	0.072	0.035	0.202	0.209

Table 1.2: Regression results of leverage on log eigenvector centrality. All specifications contain year fixed effects and all the standard errors are clustered at industry-period levels.

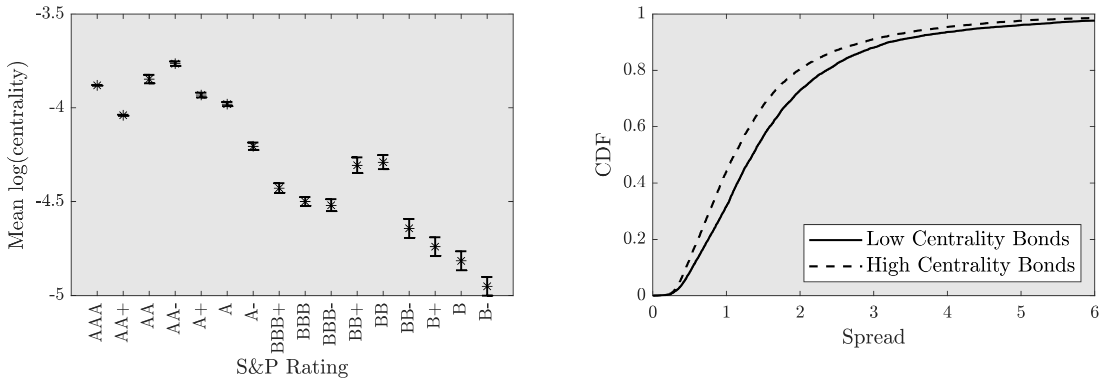


Figure 1.8: Left panel: Average log eigenvector centrality for different corporate bond credit ratings. The vertical bars show the 95% confidence intervals around the means.
Right panel: Cumulative distribution function of bond spreads for high/low centrality bonds.

	spread
logEigCent	-0.0757
	(-2.05)
N	17835
R^2	0.004

Table 1.3: Regression results of corporate bond spread on log eigenvector centrality.

stochastically dominates the high centrality distribution, suggesting that low centrality firms are more likely to issue bonds with a spread larger than any value compared to high centrality firms. The positive association of corporate bond with centrality is also present in a univariate regression of bond spread on log eigenvector centrality. With a negative coefficient of -0.07 and the t-stat of -2.05, the relation is economically and statistically significant. One percent increase in the centrality is associated with a 0.07 basis points reduction of the bond spread. However, this coefficient is not significantly different than zero when the firm level variables from table 1.1 are controlled for in the regression.

1.5 Conclusion

In this chapter, I showed that the complex interdependence of industries to input goods from one another, puts each industry in a unique position in the input-output production network. These positions interact with the ability of industries to substitute different inputs for one another and determine the risks each industry faces in its input and output markets. As a result, industries' leverage choices and bond prices depend on their overall position in the network. Being more central makes industries safer by diversifying their revenue sources and lowering their bankruptcy risks for each level of debt. Requiring more input goods on the other hand, could increase or lower an industry's risk depending on its elasticity of substitution. Overall, in a calibrated model and empirically I show more central industries tend to have higher leverage and pay less to borrow. Moreover, the network dependence is a source of externality that induces industries to issue sub-optimal levels of debt (both in the aggregate level and the distribution across industries), and consequently sub-optimal bankruptcy levels and distribution. This suggests a government intervention that could move the economy towards its socially optimal point by making the debt tax shields position-specific and/or by rescue policies for distressed firms that take their position in the network into account.

Chapter 2

Investors' Social Network with Uncertain Source of Information

2.1 Introduction

The idea that financial markets aggregate and disseminate investors' diverse information through prices has been the subject of an extensive literature in financial economics. There is also substantial evidence that investors exchange information about the stocks they trade,[1] and more recently, a growing literature focuses on the implications of direct information sharing social networks across agents for financial markets. In this paper, I analyze the effect of information sharing networks among investors on information aggregation and market price informativeness. I extend the standard CARA-normal large economy with private signals to allow for arbitrary information-sharing networks where agents share their portfolio holdings with their neighbors. I study two types of equilibria: (i) benchmark equilibrium with rational agents and no uncertainty about the source of information, and (ii) rational investors with uncertainty about the information source.

I present three main results. First, I show that, in the benchmark equilibrium, the position of an investor in the information-sharing network plays a key role in the quality of the information he receives. More central investors have access to better information via their neighbors in the network. This is in line with empirical findings of Ozsoylev et al. (2014) where, using an account-level dataset of traders in the Istanbul Stock Exchange, the authors show more central investors in

[1] Shiller and Pound (1989) in a survey of 131 institutional investors found that for a majority of them, conversation with their peers promoted their most recent stock purchase. Consistent with the word of mouth theory, Hong et al. (2005) show that a mutual fund manager is more likely to buy (or sell) a particular stock if other managers in the same city are buying (or selling) that same stock. Ivković and Weisbenner (2007) found a similar result for households (see also Feng and Seasholes (2004), Grinblatt and Keloharju (2001)). Sometimes sharing of information might be involuntary through common agents, for example, using detailed trade-level data from institutional investors and their brokers, Maggio et al. (2019) show central brokers gather information by executing informed trades and later leak the gained information to their other clients.

the empirical investor network earn higher returns. Maggio et al. (2019) also show clients of more central brokers earn higher returns by exploiting the information leakage from the brokers to their clients.

Second, in the benchmark equilibrium, a more connected investor network always leads to more informative prices, and any strongly connected social network is maximally efficient. However, if there is uncertainty about the source of information, more information sharing could lower price informativeness. This is because, in the benchmark equilibrium, more sharing always brings new information to some investors which in turn leads to better-informed agents and more informative prices. Furthermore, when the investor graph is strongly connected,[2] all of the available information reaches all of the investors and the price becomes maximally informative. When there is uncertainty about the source of information and the accuracy of information from different sources is heterogeneous, there is an extra layer of processing for investors, that is they need to infer the quality of information they receive from their neighbors. A more connected network could impede the ability of investors to uncover the information source and hence its accuracy. In other words, when the source of information on the network is uncertain, there is a trade-off between more information shared and the ability of investors to better identify the accuracy of the signals they receive. I show that the second channel could be strong enough to lower price efficiency for more connected networks.

Finally, I show, unlike the benchmark equilibrium, when there is uncertainty about the source of information, being more central does not always lead to access to more information via the network. There is an inherent trade-off between having access to more sources of information and the ability to infer the quality of signals received. Depending on the network structure and the quality of each investor's signal, I show the second channel could dominate and lower the overall access to information for more central investors. This is the case when there is a large variation in the quality of the signals received by more central investors. If they are not able to infer the quality of their signals, the high-quality signals get mixed with low-quality ones and lower the overall access to information for more central investors.

These results are closely related to the finding of Han and Yang (2013). Following the literature on costly information acquisition started by Grossman and Stiglitz (1980) and Verrecchia (1982),[3] Han and Yang (2013) show that when the information is exogenous, increase in the investors network connectedness improves price informativeness. Colla and Mele (2010) and Ozsoylev and Walden (2011) also show social communication improves market efficiency with exogenous information acquisition. I confirm the same finding in the benchmark equilibrium with a more general

[2]A directed graph is strongly connected if there is a directed path between any two nodes in the graph.

[3]See also Barlevy and Veronesi (2000), Dow et al. (2017), Garcia and Strobl (2011), Garcia and Vanden (2009), Kacperczyk et al. (2009), Mele and Sangiorgi (2015), Nieuwerburgh and Veldkamp (2010), Peress (2004, 2010), Van Nieuwerburgh and Veldkamp (2009).

form of information network. Han and Yang (2013) also show that when information acquisition is endogenous and costly, a more connected network could lead to less informative prices due to investors' incentive to free ride on the information provided to them from the network. As a result, less information is generated in equilibrium which leads to a loss of price efficiency. I complement these findings by identifying a new channel through which a more connected network could lead to a loss of price informativeness. In particular, I show that even if the information acquisition is exogenous, a more connected network could lower efficiency due to information source uncertainty. In addition, by using a general form of information-sharing networks, I show an investor's position in the network plays a crucial role in determining the amount of information they receive.

This paper also belongs to the broader literature on networks in finance and economics. For detailed surveys about networks in economics and finance see Allen and Babus (2009), Carvalho and Tahbaz-Salehi (2019), Easley et al. (2010), Jackson (2010, 2011). This paper contributes to the literature by studying the joint interaction of information networks and information source uncertainty and their effect on market outcomes.

2.2 Model

2.2.1 Setup

There are two periods $t = 0,1$ and a single risky asset that is traded at $t = 0$ which generates a random payoff $x \sim \mathrm{N}(\mu, \Sigma_x)$ at $t = 1$ where the mean of the distribution $\mu \sim \mathrm{N}(\bar{\mu}, \Sigma_\mu)$ is also random. The number of shares of the risky asset is normalized to 1, and there is a risk-free asset with a net zero return. There are also n investors in the economy, indexed by $i = 1,...,n$. Investor i owns an initial wealth of w_i and has CARA utility function with risk aversion parameter γ_i. Each investor decides how much to invest in each asset at $t = 0$ to maximize his expected utility at $t = 1$.

Information Structure. Each investor has a set of neighbor investors with whom he shares his portfolio holdings information.[4] Let a_{ij} denote an indicator that is equal to 1 if investor j shares his portfolio holding with i and otherwise is zero.[5] Using the graph theory terminology $A = [a_{ij}]_{n \times n}$ is the adjacency matrix of a graph in which investors are nodes and edges are the information-sharing relationship among neighbor investors about their portfolio holdings. I assume the structure of the investor network is common-knowledge among investors. I further assume in addition to their neighbors portfolio holdings, at $t = 0$ every investor i also receives a private signal $z_i = \mu + \varepsilon_i$ about the mean of the asset payoff where $\varepsilon_i \sim \mathrm{N}(0, \Sigma_i)$. Furthermore, a randomly chosen investor

[4] The model is static in the sense that all investors demand the asset at the same time. However, what I mean here by observing their neighbor's demand is investor i can condition his demand on his neighbors' demand, and in equilibrium, all these conditions must hold.

[5] In principle it is not required to have $a_{ij} = a_{ji}$ as investor i could share his demand with j without j sharing it with i.

k receives an additional signal $z_k^* = \mu + \varepsilon_k^*$ about the mean asset payoff where $\varepsilon_k^* \sim \mathrm{N}(0, \Sigma_k^*)$ and $\mathbb{P}\mathrm{rob}[i \text{ is doubly informed}] = \mathbb{P}\mathrm{rob}[i = k] = \pi_i$.

In the next section, I focus on a benchmark case where the identity of the investor k who has received the additional signal is common-knowledge among all investors and part of their information set. Later I will change this assumption to a case where all investors can only infer the probability of such an event from their information set using the Bayes Rule, except for investor k who knows with certainty that he is doubly informed. In other words, in the benchmark case, after k is chosen, there is no uncertainty about the source of the information floating in the investor network whereas, with information source uncertainty, most of the investors do not know with certainty where some of the information in the network is coming from and only able to indirectly infer the source of information as best as they could using the information available to them.

2.2.2 Benchmark Equilibrium

In this section, I study a model in which the identity of the doubly informed agent k is public knowledge. Assuming price taker investors and conditional on investor k being the doubly informed investor, investor i's demand for the risky asset is

$$d_i(k) = \frac{\mathbb{E}\left[\mu | \Omega_i(k)\right] - p}{\gamma_i \left(\Sigma_x + \mathbb{V}\mathrm{ar}\left[\mu | \Omega_i(k)\right]\right)} \tag{2.1}$$

where p is the price of the asset and $\Omega_i(k)$ is the information set of investor i conditional on k being doubly informed. $\Omega_i(k)$ consists of z_i as well as $d_j(k)$ for all $j \in \{m : a_{im} = 1\}$ that is the set of investors who share their portfolio information with i. In addition, $\Omega_k(k)$ also contains z_k^*.

In the model, I assume the only sources of information for the investors are their private signals and their neighbors' portfolios. In particular, I restrict investors from gaining information from the price of the asset. It is known in these settings that if the investors use the price in their information set, then the No-trade Theorem applies and the market collapses. As an alternative approach, one could introduce noise traders to the model to avoid the mentioned problem. With noise traders the model becomes more complicated, however all the intuition and results of the paper remains similar. I maintain the mentioned assumption throughout the paper for simplicity and tractability of the model.

Given the normality assumptions, in equilibrium, each investor's demand is linear in his signals and the price of the asset, that is:

$$d_i(k) = c_i(k) + \sum_{j=1}^{n} a_{ij}\alpha_{ij}(k)\, d_j(k) + \beta_i(k)\, z_i + \beta(k)\, z_k^*\, \mathbf{1}[i = k] - \lambda_i(k)\, p \tag{2.2}$$

The product $a_{ij}\alpha_{ij}$ is the effect of investor j's demand on that of i. This product is zero if j is not a sharing neighbor of i ($a_{ij} = 0$) and thus d_j is not in the information set of i, and is equal to

α_{ij} if j is a neighbor of i $(a_{ij}=1)$. β_i is the effect of i's private signal z_i on his asset demand, β is the effect of the second signal of investor k in his demand which appears in d_i only when $k=i$, λ_i is the price effect of the asset on i's demand, and c_i is just a constant.

An equilibrium is a set of matrix functions $[c_i(k)]_{n\times 1}$, $[a_{ij}\alpha_{ij}(k)]_{n\times n}$, $[\beta_i(k)]_{n\times 1}$, $[\beta(k)]_{1\times 1}$, and $[\lambda_i(k)]_{n\times 1}$ such that (2.1) and (2.2) hold; and the risky asset market clears:

$$\sum_{i=1}^{n} d_i(k) = 1 \tag{2.3}$$

In equilibrium, the market-clearing price aggregates and reflects the information available to the investors about the final asset payoff. This crucially depends on the amount of information available to each investor via his private signals, as well as the extent to which this information is shared with others. The following definition gives a quantitative measure of price efficiency in information aggregation.

Definition 5. *Given an information sharing network $A=[a_{ij}]$, define* ***Information Efficiency (IE)*** *of the network as:*

$$\mathrm{IE}(A) = 1 - \frac{\mathbb{V}\mathrm{ar}[x|p,A]}{\mathbb{V}\mathrm{ar}[x]} \tag{2.4}$$

$\mathrm{IE}\in[0,1]$ is the fraction of the total variation in the underlying asset payoff that can be explained by the market price. If the market price perfectly correlates with the underlying asset value, then $\mathrm{IE}=1$ and knowing the asset price eliminates any uncertainty about its final payoff. On the other extreme if the price has no information about the asset payoff, then $\mathrm{IE}=0$. An equivalent interpretation of IE in this linear economy is the r-squared of regression of asset payoff on market price.

The following set of results summarizes the implications of different information sharing networks on information aggregation and informativeness of the market price.

Lemma 2. *$A=[1]_{n\times n}$ is a most efficient information network:*

$$\mathrm{IE}([1]_{n\times n}) \geq \mathrm{IE}(B) \;\; \textit{for all} \;\; B\in\{0,1\}^{n\times n} \tag{2.5}$$

$A=[1]_{n\times n}$ corresponds to an economy with full information sharing in which all the investors observe everyone's demand. Intuitively, by observing all the demands, each investor could fully recover all the private signals in the economy. Hence the total available information about the fundamentals via private signals is fully utilized by all the investors. The market price aggregates the individual investors' information which leads to the most informative price of the asset.

It is worth noting that A being a most efficient network does not necessarily mean $\mathrm{IE}(A)=1$, rather it means among all the possible information-sharing networks, A achieves the highest IE possible. The level of IE depends on both how much information is available to the investors and how it is shared among them. For instance if $\Sigma_x = 0$ i.e. the only source of uncertainty about the asset payoff is through the uncertainty about its mean μ; $\Sigma_i = 0$ for some i, that is investor i directly observes μ without any noise; and finally $A = [1]_{n \times n}$ then $\mathrm{IE}(A) = 1$. In this example, there is enough information for one investor to fully learn the asset payoff and his information is shared with all the other investors in the market.

It is not surprising that full information sharing is the most efficient network, however, one also could achieve the maximum efficiency via a network with much fewer direct links between investors. Consider a star network in which there is a central investor who observes everyone else's demand and all the other investors only observe the central investor's demand. In this network, the central investor could recover all the peripheral investors' private signals from their demands and aggregate the information in his demand. At the same time, the peripheral investors could recover the aggregated information from the demand of the central investor and the market achieves the maximum efficiency. In other words, all the investors share their private information with the central investor and the central investor reveals the aggregated signal to everyone via his demand. Proposition 5 characterizes the structure of all most efficient networks.

Proposition 5. *If $\Sigma_i < \infty$ for all i, then network A is most efficient if and only if it is strongly connected, that is if there is a directed path between any two investors in A.*

Intuitively each investor's demand aggregates all the information available to him and is a sufficient statistic for anyone who wants to utilize this information in his demand. Hence the existence of a directed path between any two investors suffices to achieve full information sharing. It is also necessary because otherwise, one could "increase" information available to investors who are not connected via a directed path by having them share their demands. For detailed and formal proof refer to the appendix. Proposition 6 relates the effect of increased information sharing on IE.

Proposition 6. *if $A = [a_{ij}]_{n \times n}$ and $B = [b_{ij}]_{n \times n}$ where $a_{ij} \geq b_{ij}$ for all i and j, then network A is at least as efficient as B, that is* $\mathrm{IE}(A) \geq \mathrm{IE}(B)$.

Proposition 6 simply means if we pick any network B and have some new investors share their information (add some new edges to the network) then the efficiency of the network does not decrease.

The next result links the accuracy of each investor's signal to its relative position in the information-sharing network. Each investor's relative position in the network is a function of the entire network in principle but several measures, known as centrality measures, have been used in the literature to quantitatively summarize the position of nodes in a network e.g. betweenness, node, eigenvector, Katz, etc. I define $C_A(i)$, the centrality of node i in the network $A = [a_{ij}]_{n \times n}$ as the fraction of investors who are connected to i via an at least one directed path.

Proposition 7. *Suppose* $\Sigma_i = \overline{\Sigma}$*, and* $\Sigma_k^* = \infty$ *for all i and k. For a given information network* $A = [a_{ij}]_{n\times n}$ *if node i is more central than j i.e.* $C_A(i) \geq C_A(j)$ *then* $\mathbb{V}\text{ar}[x|\Omega_i] \leq \mathbb{V}\text{ar}[x|\Omega_j]$.

Condition $\Sigma_i = \overline{\Sigma}$ ensures that all the investors have access to private signals with the same precision, and $\Sigma_k^* = \infty$ simply means that none of the investors is doubly informed.[6] In other words, the only difference between investors is due to their position in the network and not due to their access to a better private source of information. Then a more central investor is better informed, that is he perceives a lower variance of the asset payoff conditional on his information set.

The key insight behind all the above results is that, ceteris paribus, more information sharing will result in more price informativeness and better-informed investors. This is consistent with the common belief that the market aggregates information available to investors and translates it to prices, and the more investors share their information the more prices become efficient in predicting the asset payoff. In the next section, I present a variation of the benchmark case in which this key insight fails to hold, that is more sharing does not necessarily translate to more informative prices. Namely, if the source of the information in a network is uncertain, then more sharing could lower the information content of prices.

2.2.3 Equilibrium with Information Source Uncertainty

In the benchmark equilibrium, I assumed that the structure of the information-sharing network, as well as the identity of the doubly informed investor k, is common knowledge. In this section, I maintain the assumption that all of the investors know the network structure, however, the identity of the doubly informed investor is not directly revealed to the investors (except the doubly informed agent himself). However, the investors try to learn the identity of the doubly informed agent from their signals in a Bayesian manner.

Let $\pi_k^{(i)} = \mathbb{P}\text{rob}[k \text{ is doubly informed}|\Omega_i]$ denote the posterior probability of k being doubly informed from investor i's point of view (conditional on his information set). With the CARA utility function, investor i's demand is determined by

$$d_i \in \arg\max_y - \sum_{k=1}^{n} \pi_k^{(i)} \mathbb{E}\left[e^{-\gamma_i y(x-p)} \middle| k \text{ is doubly informed}\right] \tag{2.6}$$

If all of the investors' signals fully reveal to them the state of the economy i.e. the identity of the doubly informed investor, then equation (2.6) reduces to the benchmark case in equation (2.1). However, this is not the case for many network structures. In other words, although investors' signals provide some information about the state of the economy beyond the prior distribution π_k, it is far from revealing the true state in many cases. In fact, as the network becomes more connected, inferring the state of the economy might get increasingly difficult for the investors and

[6] Note if $\Sigma_k^* = \infty$ then no investor is doubly informed and the information sets do not vary with k.

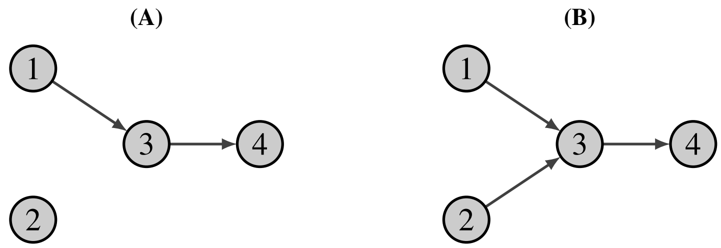


Figure 2.1: Illustrative Example: Investors' Information Sharing Network

Parameter	Value
$\overline{\mu}$	0
Σ_x, Σ_μ	15
Σ_4	3
Σ_1^*	1
Σ_2^*	20
γ_1, γ_2, γ_3	1
γ_4	0.25

Table 2.1: Parameter values in the example (all the signal variances not mentioned in this table are set to infinity).

lower the overall efficiency of prices, even though there are more opportunities to share information. I illustrate this phenomenon via the following simple example.

Example 1

Consider the following networks. Each node corresponds to an investor and each edge shows the flow of information between nodes. In the network (A), investor 1 shares his portfolio holdings with 3, and investor 3 with 4. The Network (B) is identical to (A) except there is an extra link from the investor 2 to 3. In both cases, I assume $\pi_1 = \pi_2 = 0.5$ and $\pi_3 = \pi_4 = 0$ that is with equal probabilities either investor 1 or 2 is doubly informed.

For simplicity of the example, I reduce the total number of informative signals available to the investors to three sources: (i) $z_4 = \mu + \varepsilon_1$ available to investor 4, (ii) $z_1^* = \mu + \varepsilon_1^*$, and (iii) $z_2^* = \mu + \varepsilon_2^*$ available to investors 1 and 2 respectively only when doubly informed. In any realization of the economy, only one of the (ii) or (iii) is available to the corresponding investor along with (i) which is always observed by investor 4.[7] The numerical values of the model parameters are summarized in table 2.1. Crucially, the signal strength of the two investors who may become doubly informed is very different. Investor 2, who is disconnected in case A but becomes connected in case B has a far less precise signal.

[7] This is equivalent to setting variance of the noise terms in all of the other signals equal to ∞.

Case 1: network (A)

Benchmark equilibrium. There are two states of the world i.e. when 1 is doubly informed (state 1) or 2 is doubly informed (state 2). Table 2.2 summarizes the state-contingent equilibrium demand of each investor. In both states of the economy and for every investor i, d_i is obtained from equation (2.1). In state 1, investor 1 observes one signal z_1^* about the average asset payoff μ and his demand is a linear combination of the prior mean of the μ distribution, the signal he observes, and the market price. Investor 2 observes no signal and his demand is only a function of $\overline{\mu}$ and p. Having observed d_1 and the market price p, and knowing 1's demand function, investor 3 could fully recover 1's private signal. Let $z_1^{*(3)}$ denote the signal z_1^* perceived by investor 3 from observing d_1. In this case d_1 is enough to fully recover the value of z_1^*, hence $z_1^{*(3)} = z_1^*$. Similarly, investor 4 could recover $z_1^{*(4)} = z_1^{*(3)} = z_1^*$ from d_3 and use it in his demand along with his own private signal z_4.

In state 2, investor 1 receives no signal and bases his demand on the prior distribution of μ. Investor 2 receives one signal z_2^* and forms his demand based on his private signal and the prior distribution of μ. The information set of investor 3 is identical to that of 1, and observing d_1 has no information value for 3. Similarly, investor 1 gains no information from d_3 and bases his demand solely on his private signal z_4 and the prior distribution of μ.

In summary, when investor 1 is doubly informed, his information propagates across the network from 1 to 3 and then from 3 to 4. On the other hand, when 2 is doubly informed, his information is not shared with any investor. Investors 3 and 4 also do not gain any new information from their neighbors because there is no new information available to investor 1 in the first place that could be shared.

Equilibrium with information source uncertainty. In the network (A) the equilibrium with source uncertainty is identical to the benchmark equilibrium. This is because even if the investors are not directly informed whether state 1 or 2 happened, they can infer the exact state of the economy from their signals. Investor 1 is either doubly informed (state 1) or not which corresponds to state 2. Similarly, investor 2 can infer the state of the economy without any uncertainty. If 1 is doubly informed, investor 3 could almost surely infer the state of the economy from observing d_1 alone. If state 2 happens, then investor 1's demand must be $d_1 = \frac{\overline{\mu}-p}{\gamma_1\left[\Sigma_x+\Sigma_\mu\right]}$. Any deviation from this demand means that state is 1 and investor 1 is doubly informed. Similar reasoning reveals the state of the economy to investor 4 using d_3.

Case 2: network (B) benchmark equilibrium

Table 2.3 summarizes the demand functions in this case. The only difference between this case and case 1 is that in state 2, the signal z_2^* also propagates across the network from investor 2 to 3 and

state 1:	investor 1 is doubly informed
	$d_1 = \frac{\frac{\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\overline{\mu} + \frac{\Sigma_\mu}{\Sigma_\mu+\Sigma_1^*} z_1^* - p}{\gamma_1 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\right]}$ $d_2 = \frac{\overline{\mu}-p}{\gamma_2 \left[\Sigma_x+\Sigma_\mu\right]}$ $d_3 = \frac{\frac{\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\overline{\mu} + \frac{\Sigma_\mu}{\Sigma_\mu+\Sigma_1^*} z_1^{*(3)} - p}{\gamma_3 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\right]}$ where $z_1^{*(3)} = \frac{\Sigma_\mu+\Sigma_1^*}{\Sigma_\mu}\left\{\gamma_1 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\right] d_1 - \frac{\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\overline{\mu} + p\right\} = z_1^*$ $d_4 = \frac{\frac{\Sigma_4 \Sigma_1^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_1^*+\Sigma_4\Sigma_1^*}\overline{\mu} + \frac{\Sigma_\mu \Sigma_1^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_1^*+\Sigma_4\Sigma_1^*} z_4 + \frac{\Sigma_\mu \Sigma_4}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_1^*+\Sigma_4\Sigma_1^*} z_1^{*(4)} - p}{\gamma_4 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_4 \Sigma_1^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_1^*+\Sigma_4\Sigma_1^*}\right]}$ where $z_1^{*(4)} = \frac{\Sigma_\mu+\Sigma_1^*}{\Sigma_\mu}\left\{\gamma_3 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\right] d_3 - \frac{\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\overline{\mu} + p\right\} = z_1^{*(3)} = z_1^*$
state 2:	investor 2 is doubly informed
	$d_1 = \frac{\overline{\mu}-p}{\gamma_1 \left[\Sigma_x+\Sigma_\mu\right]}$ $d_2 = \frac{\frac{\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\overline{\mu} + \frac{\Sigma_\mu}{\Sigma_\mu+\Sigma_2^*} z_2^* - p}{\gamma_1 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\right]}$ $d_3 = \frac{\overline{\mu}-p}{\gamma_3 \left[\Sigma_x+\Sigma_\mu\right]}$ $d_4 = \frac{\frac{\Sigma_4}{\Sigma_\mu+\Sigma_4}\overline{\mu} + \frac{\Sigma_\mu}{\Sigma_\mu+\Sigma_4} z_4 - p}{\gamma_4 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_4}{\Sigma_\mu+\Sigma_4}\right]}$

Table 2.2: demand functions in case 1: network (A)

then from 3 to 4, which they utilize in their demand.

Case 3: network (B) equilibrium with information source uncertainty

With a similar argument presented above, in this case, investors 1, 2, and 3 could infer the true state of the economy from their information set. Hence d_1, d_2, and d_3 are identical to their value in case 2. The only difference is that investor 4 is no longer able to eliminate uncertainty about the state of the world based on his signals. Although investor 4 is not able to infer the true state of the economy, he can update his prior belief about the probability of each state given his information set.[8] Let $\pi_1^{(4)} = \mathbb{P}\text{rob}[\text{state} = 1|\Omega_4]$ and $\pi_2^{(4)} = \mathbb{P}\text{rob}[\text{state} = 2|\Omega_4]$ denote the posterior belief of

[8]This is essentially what investors 1, 2, and 3 also do. In their cases, the posterior distribution over the states becomes degenerate i.e. they know the true state with probability 1.

state 1: investor 1 is doubly informed

$$d_1 = \frac{\frac{\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\overline{\mu}+\frac{\Sigma_\mu}{\Sigma_\mu+\Sigma_1^*}z_1^*-p}{\gamma_1\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\right]}$$

$$d_2 = \frac{\overline{\mu}-p}{\gamma_2\left[\Sigma_x+\Sigma_\mu\right]}$$

$$d_3 = \frac{\frac{\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\overline{\mu}+\frac{\Sigma_\mu}{\Sigma_\mu+\Sigma_1^*}z_1^{*(3)}-p}{\gamma_3\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\right]}$$

where $z_1^{*(3)} = \frac{\Sigma_\mu+\Sigma_1^*}{\Sigma_\mu}\left\{\gamma_1\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\right]d_1-\frac{\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\overline{\mu}+p\right\} = z_1^*$

$$d_4 = \frac{\frac{\Sigma_4\Sigma_1^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_1^*+\Sigma_4\Sigma_1^*}\overline{\mu}+\frac{\Sigma_\mu\Sigma_1^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_1^*+\Sigma_4\Sigma_1^*}z_4+\frac{\Sigma_\mu\Sigma_4}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_1^*+\Sigma_4\Sigma_1^*}z_1^{*(4)}-p}{\gamma_4\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_4\Sigma_1^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_1^*+\Sigma_4\Sigma_1^*}\right]}$$

where $z_1^{*(4)} = \frac{\Sigma_\mu+\Sigma_1^*}{\Sigma_\mu}\left\{\gamma_3\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\right]d_3-\frac{\Sigma_1^*}{\Sigma_\mu+\Sigma_1^*}\overline{\mu}+p\right\} = z_1^{*(3)} = z_1^*$

state 2: investor 2 is doubly informed

$$d_1 = \frac{\overline{\mu}-p}{\gamma_1\left[\Sigma_x+\Sigma_\mu\right]}$$

$$d_2 = \frac{\frac{\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\overline{\mu}+\frac{\Sigma_\mu}{\Sigma_\mu+\Sigma_2^*}z_2^*-p}{\gamma_2\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\right]}$$

$$d_3 = \frac{\frac{\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\overline{\mu}+\frac{\Sigma_\mu}{\Sigma_\mu+\Sigma_2^*}z_2^{*(3)}-p}{\gamma_3\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\right]}$$

where $z_2^{*(3)} = \frac{\Sigma_\mu+\Sigma_2^*}{\Sigma_\mu}\left\{\gamma_2\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\right]d_2-\frac{\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\overline{\mu}+p\right\} = z_2^*$

$$d_4 = \frac{\frac{\Sigma_4\Sigma_2^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_2^*+\Sigma_4\Sigma_2^*}\overline{\mu}+\frac{\Sigma_\mu\Sigma_2^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_2^*+\Sigma_4\Sigma_2^*}z_4+\frac{\Sigma_\mu\Sigma_4}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_2^*+\Sigma_4\Sigma_2^*}z_2^{*(4)}-p}{\gamma_4\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_4\Sigma_2^*}{\Sigma_\mu\Sigma_4+\Sigma_\mu\Sigma_2^*+\Sigma_4\Sigma_2^*}\right]}$$

where $z_2^{*(4)} = \frac{\Sigma_\mu+\Sigma_2^*}{\Sigma_\mu}\left\{\gamma_3\left[\Sigma_x+\frac{\Sigma_\mu\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\right]d_3-\frac{\Sigma_2^*}{\Sigma_\mu+\Sigma_2^*}\overline{\mu}+p\right\} = z_2^{*(3)} = z_2^*$

Table 2.3: demand functions in case 2: network (B) with benchmark equilibrium

investor 4 about the state of the economy after observing his signals. Then d_4 is

$$d_4 \in \arg\max_y \begin{array}{l} -\pi_1^{(4)} e^{-\gamma_4 \left[\frac{\Sigma_\mu \Sigma_4 \Sigma_1^*}{\Sigma_\mu \Sigma_4 + \Sigma_\mu \Sigma_1^* + \Sigma_4 \Sigma_1^*} \left(\frac{\overline{\mu}}{\Sigma_\mu} + \frac{z_4}{\Sigma_4} + \frac{z_1^{*(4)}}{\Sigma_1^*}\right) - p\right] y + \frac{1}{2}\gamma_4^2 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_4 \Sigma_1^*}{\Sigma_\mu \Sigma_4 + \Sigma_\mu \Sigma_1^* + \Sigma_4 \Sigma_1^*}\right] y^2} \\ -\pi_2^{(4)} e^{-\gamma_4 \left[\frac{\Sigma_\mu \Sigma_4 \Sigma_2^*}{\Sigma_\mu \Sigma_4 + \Sigma_\mu \Sigma_2^* + \Sigma_4 \Sigma_2^*} \left(\frac{\overline{\mu}}{\Sigma_\mu} + \frac{z_4}{\Sigma_4} + \frac{z_2^{*(4)}}{\Sigma_2^*}\right) - p\right] y + \frac{1}{2}\gamma_4^2 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_4 \Sigma_2^*}{\Sigma_\mu \Sigma_4 + \Sigma_\mu \Sigma_2^* + \Sigma_4 \Sigma_2^*}\right] y^2} \end{array} \tag{2.7}$$

where

$$\begin{aligned}
z_1^{*(4)} &= \frac{\Sigma_\mu + \Sigma_1^*}{\Sigma_\mu} \left\{ \gamma_3 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_1^*}{\Sigma_\mu + \Sigma_1^*}\right] d_3 - \frac{\Sigma_1^*}{\Sigma_\mu + \Sigma_1^*} \overline{\mu} + p \right\} \\
z_2^{*(4)} &= \frac{\Sigma_\mu + \Sigma_2^*}{\Sigma_\mu} \left\{ \gamma_3 \left[\Sigma_x + \frac{\Sigma_\mu \Sigma_2^*}{\Sigma_\mu + \Sigma_2^*}\right] d_3 - \frac{\Sigma_2^*}{\Sigma_\mu + \Sigma_2^*} \overline{\mu} + p \right\} \\
\pi_1^{(4)} &= \mathbb{P}\text{rob}[\text{state} = 1 | d_3, z_4] \\
&= \frac{\pi_1 \text{ pdf}[d_3 | \text{state} = 1, z_4]}{\pi_1 \text{ pdf}[d_3 | \text{state} = 1, z_4] + \pi_2 \text{ pdf}[d_3 | \text{state} = 2, z_4]} \\
\pi_2^{(4)} &= \mathbb{P}\text{rob}[\text{state} = 2 | d_3, z_4] \\
&= 1 - \pi_1^{(4)}
\end{aligned}$$

$z_1^{*(4)}$ is investor 4's inferred value of z_1^* from d_3 if the state is 1. Similarly, $z_2^{*(4)}$ is 4's inference about z_2^* if the state is 2. Furthermore, $\pi_1^{(4)}$ and $\pi_2^{(4)}$ are 4's posterior probabilities of states 1 and 2 respectively after observing d_3 and z_4.

In all of the above cases, the market price is determined by clearing the market for the risky asset, that is $d_1 + ... + d_4 = 1$.

Summary of Results

$$\text{IE}(B)_{\text{uncertain source}} < \text{IE}(A)_{\text{benchmark}} = \text{IE}(A)_{\text{uncertain source}} < \text{IE}(B)_{\text{benchmark}} \tag{2.8}$$

1. Network A's structure is such that the signals of investors are all fully revealing of the state of the economy. Hence the equilibrium with source uncertainty coincides with the benchmark equilibrium with identical information efficiency.

2. Adding the extra link from the investor 2 to 3 improves the price efficiency under the benchmark equilibrium. This is due to the extra piece of information shared (directly or indirectly) with investors 3 and 4, even though the extra signal has extremely lower quality relative to the private signal of investor 1 ($\Sigma_2^*/\Sigma_1^* = 20$).

3. When the source of information is not explicitly revealed to the investors, contrary to the

benchmark equilibrium, adding the extra sharing link from investor 2 to 3 lowers the price informativeness. This extra link has two opposing effects: (i) similar to the benchmark equilibrium, it provides investors 3 and 4 with an extra piece of information about the asset payoff and improves the efficiency; (ii) it also prevents the investor 4 from fully inferring the state of the economy from his signals. As a result, the relatively precise signal of investor 1 gets mixed with the extremely imprecise signal of investor 2 before reaching investor 4. Although investor 4 can partially update his belief about the state of the economy using the Bayes rule, this update is not enough and investor 4 is not able to eliminate the uncertainty about the state of the economy for himself. This leads to two types of inefficiencies in 4's demand namely putting more than optimal weight on the signal received from investor 3's demand when the state is 2; and putting less than optimal weight on this signal when the state is 1. In this example, the adverse effect of (ii) on efficiency is stronger than (i) causing the fall in IE of network A compared to B.

4. The central investor 4 could gain more information from different sources due to his position in the network, however being more central also makes it more difficult for him to infer the source of information he receives. When the quality of information from different sources varies substantially, the second channel could dominate and more centrality could lower the overall access to information.

2.3 Conclusion

Contrary to a widely accepted statement that with rational agents more information sharing is always favorable in terms of market outcomes, in this paper I show when information is shared among investors in a social network, more sharing could lead to opposite effects. Although I only focused on the case in which the source of information is uncertain, it is not the only case that more sharing could lead to adverse market outcomes. Han and Yang (2013) argue a similar result when information acquisition is endogenous. Banerjee (1992), Bikhchandani et al. (1992), Eyster and Rabin (2010), Han et al. (2018) show with certain documented behavioral biases (namely cursedness and overconfidence) more sharing could also lead to inefficiencies. This issue is particularly important from the policy perspective and information design. Maximizing price efficiency does not always require more information sharing, and the context in which information is shared is of great importance.

Chapter 3

Firm Heterogeneity in Skill Returns

This chapter is a joint work with Michael J. Böhm[a] and Giovanni Gallipoli.[b]

[a]University of Bonn, Department of Economics (michael.j.boehm@uni-bonn.de).
[b]University of British Columbia, Vancouver School of Economics (gallipol@mail.ubc.ca).

3.1 Introduction

The recognition that earnings distributions reflect both worker and firm heterogeneity dates back decades. Robert Willis notably warned about "an imbalance in the human capital literature which has emphasized the supply far more than the demand for human capital" (Willis, 1986). The availability of matched employer–employee records has brought about a renewed interest in firm-level differences (e.g., Card et al., 2013, Lamadon et al., 2022, Song et al., 2018, Sorkin, 2018). A workhorse of this literature is the Abowd et al. (1999, AKM) two-way fixed effect model, which subsumes unobserved heterogeneity of workers and firms into additively separable measures whose contributions to the dispersion of earnings can be transparently quantified. The correlation between firm and worker fixed effects is often interpreted as evidence of nonrandom sorting of workers across employers, or lack thereof. However, several studies (Borovickova and Shimer, 2020, Eeckhout and Kircher, 2011, Hagedorn et al., 2017) caution against drawing inference about match-specific productivity from fixed effect estimates, emphasizing that complementarity is hard to characterize within the boundaries of additively separable models of worker and firm heterogeneity. Such additively separable effects are also unsuitable to examine the endogenous skewness of wages emphasized in matching and assignment models. These considerations inform empirical frameworks that nest flexible matching mechanisms within two-sided unobserved heterogeneity (e.g., Bonhomme et al., 2019, Lentz et al., 2018).

This paper presents novel and direct evidence on worker–firm complementarities, matching, and their effects on earnings. To this end, we link cognitive and noncognitive test scores with population data on Swedish workers and firms, and employ distinct empirical approaches to robustly estimate firm-level returns to skill attributes. Our estimates reveal significant heterogeneity in skill returns, with some firms paying up to 35 log points more than others for similar cognitive and noncognitive attributes.[1] This leads to strong incentives for sorting of workers with different skill endowments across firms. We show that heterogeneous returns, and the induced sorting, materially impact both the level and the distribution of earnings.

The cognitive and noncognitive measures, elicited for almost all Swedish males, have been used in several studies that document their information content[2] and establish their relation to distinct productive attributes (Edin et al., 2022). In our high-dimensional estimation of firm returns, we employ these measures in conjunction with alternative approaches to address limited mobility biases in the estimation of returns and of their quadratic forms. One approach builds on clustering methods (Bonhomme et al., 2019) whereby we group firms into 100 classes based on the earnings and skills of their employees. The other delivers estimates of quadratic forms of the parameters of interest at the individual (non-grouped) firm level after explicitly correcting for biases (Kline et al., 2020). Each approach imposes different sample restrictions and assumptions. Results, however, are remarkably robust in the sense that the relative importance of skill returns, as opposed to conventional measures of firm heterogeneity based on fixed effects, is stable and does not depend on the approach or specific implementation choices. For either approach, estimation of different layers of firm heterogeneity requires significant computational work, which we discuss below.

To motivate our focus on the heterogeneity of skill returns, we begin by estimating standard fixed-effect models separately for high versus low skill workers. The hypothesis that earnings premia at a given firm are the same across skill levels is clearly rejected for both cognitive and noncognitive traits. Informed by this finding, we develop an empirical specification that flexibly allows for granular returns to each skill attribute. To facilitate comparisons to existing work, the specification is derived from a monopsonostic model of labour demand (Card et al., 2018, Lamadon et al., 2022). The model delivers a first-order approximation for a general wage function in which skill $\times$ firm interaction terms reflect heterogeneous returns, while firm intercepts capture skill-independent premia. As we show, standard Mincer returns are a key part of the empirical representation despite being subsumed into worker fixed effects.

Our estimates reveal considerable dispersion in returns across firms in either skill dimension, and relatively more in the cognitive one. The correlation between returns to different skills is

[1] As we show, skill premia can be derived in labor market models with two-sided heterogeneity. Establishing their empirical prevalence and implications is, however, demanding in terms of data requirements and estimation.

[2] For example, Lindqvist and Vestman (2011) show that the military test scores are highly significant in predicting earnings and unemployment conditional on any rich set of control variables. Fredriksson et al. (2018) use them to identify the effects of job–skill mismatch on labor mobility and life-cycle wage growth. See also our Appendix A.1.1.

positive but weak; this suggests that collapsing cognitives and noncognitives into a single composite index might be restrictive when examining complementarities and sorting. Estimates show that returns heterogeneity induces material gains from the assignment of workers to firms, generating earnings gaps of the same order of magnitude as those induced by firm intercepts.

To gauge the intensity of sorting we employ analytical notions developed in multidimensional assignment problems (Lindenlaub, 2017). Several testable restrictions implied by positive assortative matching are supported in the data. We find that the assignment of more able workers to high return employers stochastically dominates (in first-order) the assignment of lower skilled workers (Lindenlaub and Postel-Vinay, 2020). Further corroboration of assortative matching is obtained by projecting firm-level returns onto skill measures (for a discussion of such projection methods, see Kline et al., 2020) as well as by regressing skills onto estimates of returns for different firm clusters. Sorting occurs along both skill dimensions but is stronger in the cognitive one where firm heterogeneity is larger.

The heterogeneity in returns, and the induced sorting, have significant but uneven effects on the moments of the earnings distribution. First, we show that matching increases aggregate efficiency and raises average earnings compared to a counterfactual random allocation of workers to firms. Moreover, earnings differences between different skill levels are strongly convexified by sorting. That is, earnings at the top are magnified by the interaction of skills and returns, while earnings of middle-to-low skill workers suffer a relative decline because they are frequently matched with low-return firms. Compared to random assignment, intermediate skill workers suffer more than the lowest-ability ones since the latter would hardly benefit from higher returns due to their meager skill endowments. Consistent with other studies (e.g., Bonhomme et al., 2019, Borovickova and Shimer, 2020, Hagedorn et al., 2017, Lamadon et al., 2022, Lentz et al., 2018), we find that match effects raise earnings levels and dispersion. A notable impact of worker–firm complementarities is on the skewness of earnings, which become more convex in skill levels. Such effects have long been discussed in the theoretical literature (Becker and Chiswick, 1966, Becker et al., 2018, Lindenlaub, 2017, Sattinger, 1993).

To validate the baseline findings we consider a few extensions and sensitivity checks. Notably, we find that estimates of the relative magnitude of skill returns do not visibly change with the number of firm classes when using the clustering approach. The same is true when we implement alternative sampling restrictions in the firm-level estimation with bias correction of quadratic-forms. Moreover, controlling for industry- or occupation-specific effects illustrates that returns heterogeneity across firms is quantitatively large even within narrow sectors and occupations.

To probe the nature of firm differences, and in keeping with our emphasis on direct measures, we link information from their balance sheets to the main data and show that employers exhibiting high cognitive returns have significantly different capital composition, with more intangible and intellectual assets (as opposed to physical capital) per worker. Moreover, after merging additional firm survey responses, we find that these firms invest more heavily in R&D and introduce product and process innovations more frequently. This lends support to the view that production and organizational arrangements play a key role in shaping the distribution of skill returns.

Our findings support the hypothesis that substantial worker–firm complementarities exist, that they bring about assortative matching, and that they influence earnings. In doing so, we draw attention to a less explored but important dimension of firm heterogeneity. More generally, we find direct evidence of efficient, albeit imperfect, skill assignment across employers. The use of skill measures complements existing studies of worker–firm interactions and presents a transparent counterpart as it does not require tight model restrictions for the identification of unobserved attributes. Resorting to informative skill proxies facilitates the measurement of gains from matching because pecuniary returns are not themselves used to recover skills ranks. This is especially advantageous when establishing which workers benefit or lose from returns' heterogeneity and sorting, as well as to identify the impact of complementarities on skewness.

One aspect that can play an important role in the imperfect assignment of skills to jobs is their multidimensional and bundled nature. A longstanding literature has examined selection and wages in settings where workers are endowed with multiple skills (see early work in Heckman and Scheinkman, 1987, Mandelbrot, 1962, Rosen, 1983). Our estimates suggest that cognitive and noncognitive returns heterogeneity have independent impacts on earnings, thus providing further motivation for research on the implications of workers' inability to separately rent out their skills to the highest bidder (Choné and Kramarz, 2021, Edmond and Mongey, 2021, Lindenlaub, 2017, Skans et al., 2022).

Finally, our findings provide new evidence on the nature and evolution of returns to skills in the labor market. Previous work has shown that both cognitive and noncognitive attributes shape individual outcomes (e.g., Heckman et al., 2006, Lindqvist, 2012). Moreover, the *average* gains from these skills have changed over time (Beaudry et al., 2016, Deming, 2017, Edin et al., 2022). Our analysis implies that each worker's return will depend on both their skill and their match with an employer. For this reason, conventional measures of Mincerian returns are not equivalent to the averages of individual returns across employers. Rather, assortative matching can tangibly change overall skill premia in the economy while inducing uneven and nonmonotonic effects over the skill range. This points to the need to explore the determinants of firm heterogeneity in skill returns as well as its implications for matching over time.[3]

The paper is organized as follows. Section 3.2 describes the data and the methods used to account for incidental parameter biases in the estimation of firm-level variables. Section 3.3 derives the empirical specification within a labour market model with two-sided heterogeneity and presents

[3]In the context of unconditional firm premia, captured in wage intercepts, influential work has related firm heterogeneity to wage growth and inequality (e.g., Card et al., 2018, Lamadon et al., 2022), or linked it back to primitives such as market structure, institutions, and policy (De Loecker and Eeckhout, 2021, Dustmann et al., 2022).

new estimates on firm heterogeneity in skill returns. Section 3.4 illustrates how workers match with firms based on skills and returns, and provides different tests of the assortative matching hypothesis. Section 3.5 examines finer implications of heterogeneity and worker sorting for the distribution of earnings, and of gains and losses, relative to random assignment of skills. Section 3.6 explores the sensitivity of key results to alternative estimation approaches; in the same section we use ancillary information about firm heterogeneity (from surveys and balance sheets) to examine the correlation of firm returns with capital composition, production arrangements, and innovation activities. The last section concludes.

3.2 Data and Preliminary Evidence

3.2.1 Matched Earning Records and Skill Measures

Our data source consists of annual employer–employee matched records for the whole population of Swedish workers and firms during 1990–2017, including earnings, industry, occupation, and worker characteristics such as age, gender, and education. A key strength of these data are cognitive and noncognitive military enlistment tests that can be linked to individual workers. The tests were mandatory before 2007 and are available for almost 90 percent of males, across birth cohorts, in our sample.

The cognitive score is similar to an IQ measure and is assessed through tests covering logic, verbal, spatial, and technical comprehension. The noncognitive score is from a semi-structured interview with a certified psychologist who assesses willingness to assume responsibility, independence, outgoing character, persistence, emotional stability, and initiative.

Prior research shows that these scores are highly significant at predicting workers' earnings and other labor market outcomes (e.g., Edin et al., 2022, Fredriksson et al., 2018, Lindqvist and Vestman, 2011), on their own as well as conditionally on each other and any rich set of control variables. Cognitive and noncognitive measures are recorded on a standard-nine (Stanine) scale, which approximates the Normal distribution and facilitates comparisons across birth cohorts.[4] In online Appendix A.1.1 we discuss these tests in detail and show that, while assessed at age 18–19, their scores are strongly associated with earnings over the entire life-cycle.

Due to the availability of test scores, we restrict the sample to males aged 20–60 with nonmissing scores. We also restrict attention to firms that employ an average of at least ten male workers over five years or more. We focus on estimates from 1999–2008 but results are similar in alternative samples (1990–1999 and 2008–2017). The 1999–2008 sample consists of approximately 26,000 firms and 1,100,000 workers.

[4] Measures are standardized for each birth year. A score of 5 denotes the middle 20 percentiles of the population taking the test. Scores of 6, 7, and 8, are given to the next 17, 12, and 7 percentiles, and the score of 9 to the top 4 percent of individuals. Scoring below 5 is symmetric.

Our dataset reports both organization and workplace identifiers. To identify "firms" we use the workplace with the highest income that year, since workplace is closest to the notion of a production unit and is consistent with existing work (e.g., Card et al., 2013). We also use the annual labor income at the firm, which is available for all workers and includes bonuses and performance pay, as our measure of earnings throughout. Details and descriptive statistics are in Section A.1 of the online Appendix.

In Section 3.6 we link information on firms' financial accounts (from a commercial data provider) and innovation activities (from the Swedish version of the European Community Innovation Survey). These data are reported at the organization level and, in the case of multi-workplace firms, we coarsen estimates to that level of aggregation.

3.2.2 Estimation of High-Dimensional Effects Models

Our analysis requires estimation of models with many fixed effects for firms and workers as well as firm-specific returns to skill measures in matched employer–employee records. We then compute variance components of these parameters or project them onto firm observable characteristics. These steps require restrictions on the data samples and empirical methods we adopt.

Connected sets. To identify model parameters, firms need to be connected to each other through worker mobility in the final sample. This entails working with a connected component of the firm–worker graph (Abowd et al., 2002, Bonhomme et al., 2020). Distinct connected sets may exist within a large sample of employment matches and empirical analyses often focus on the largest set (or 'maximally connected subgraph'). When considering different skill levels (say high and low cognitive skills) the requirement is that we use a set which is connected for each skill level ("dual" or "double" connected in Card et al., 2016, Kline et al., 2020, respectively). As we show below, the connectedness restrictions become less stringent when observations are defined at the level of firm clusters rather than individual firms.

Limited mobility bias. While connectedness leads to unbiased identification of model parameters, researchers are usually interested in variance components. These are in general biased because of sampling error in individual parameter estimates that enter the variance components in a quadratic form. The squared sampling error is thus not mean zero and may not converge to zero as the number of firms increases. Intuitively, the bias arises from an insufficient number of movers into and out of the firm, hence "limited mobility bias", and it tends to overstate variances and understate covariances (Andrews et al., 2008). The magnitude of the bias is inversely related to the degree of connectivity of the firm-worker graph, with the graph being disconnected as limiting case (Jochmans and Weidner, 2019). For further details, see Bonhomme et al. (2020, Section 3).

Since we are interested in the dispersion and correlation of skill return parameters, our analysis is potentially subject to the limited mobility bias. Fortunately, the literature on panel data has made good progress in addressing this problem. One approach, suggested in Bonhomme, Lamadon, and Manresa (2019), defines the relevant level of firm unobserved heterogeneity as the "class" of a firm, corresponding to a cluster of similar employers. While the class can be made arbitrarily close to an individual firm, this may not be desirable because the number of job movers per firm will become smaller and result in an incidental parameters bias (i.e., reinstate the limited mobility problem). Under the assumptions of this approach, unobservable firm heterogeneity operates at the level of firm classes. The latter can be estimated in a first step through k-means clustering based on earnings and skills within each firm. This achieves two objectives: first, it enhances tractability; second, it delivers well-centered and accurate estimates of the contributions of worker and firm heterogeneity to earnings dispersion. Clustering trades off restrictions on the dimensionality of the underlying groups for increased connectedness between firm classes. Notably, this method does not require shedding observations to generate a connected set.

A different approach builds on variance component estimators designed for unrestricted linear models with heteroscedasticity of unknown form. This removes the bias by resorting to leave-out estimators of the variances of errors from the linear model. For each observation, an estimate of the error variance is obtained from a sample where that worker–firm match observation is left out. The leave-out procedure delivers unbiased estimates in finite samples (see Kline, Saggio, and Sølvsten, 2020) and facilitates tests of linear restrictions. It can be implemented as a simple variance component estimator plus a bias correction consisting of observation-specific error variances. The leave-out strategy to estimate the linear model parameters requires that firms remain connected by worker mobility when any single mover is dropped. This involves pruning the original sample to ensure that the connectedness condition is met by all leave-out subsamples.[5] Given the large scale computations involved in the estimation of leave-out quadratic forms, which are executed at the individual firm level, we resort to the random projection method of Achlioptas (2003). This reduces dimensionality and computation time (see Appendix A.2).

Implementation of the clustering and bias correction approaches. Implementation of the grouping approach requires clustering firms into classes. We define 100 classes using a k-means algorithm based on average earnings as well as average cognitive and noncognitive skills of workers. Having a sufficiently large set of classes accommodates rich heterogeneity and ensures stability while still delivering a major dimension reduction. Using information beyond wages has been proposed in the structural literature (Bagger and Lentz, 2019, Bartolucci et al., 2018, Eeckhout and Kircher, 2011, Hagedorn et al., 2017). In our implementation this is further motivated by the theoretical

[5] We use Python NetworkX to identify the articulation points of the worker–firm graph and trim it to construct the double leave-one-out connected set. See Appendix A.1.2 for details on the construction of estimation samples.

restriction that firm-specific production arrangements affect both the skill composition of the workforce and their wages. Alternative clustering criteria (e.g., adding within-firm dispersion of skills and wages, or employment levels) as well as alternative numbers of classes deliver similar results (Section 3.6.3). The availability of skill measures makes it feasible to estimate specifications that feature firm effects in both levels and returns. Previous work has shown that class membership and fixed effects can be accurately estimated with sufficiently many workers. Using skill proxies also avoids incidental parameter biases in estimated returns due to few panel observations per worker.

Implementation of the bias correction approach relies on the leave-one-out double-connected set of firms. We prune the sample to contain firms that remain connected along both skill dimensions (cognitive and noncognitive) for different levels (high and low) when each single observation is dropped. The implementation accounts for correlation of error terms within an individual's spell at a given employer (Kline et al., 2020). This is done by averaging the data to the worker–firm match level. The resulting leave-match-out set is double-connected (in both skill dimensions) and smaller than the original sample but allows for estimation at the individual firm level. The extensive size of the Swedish population data assuages concerns about sample sizes. Appendix A.2 discusses theoretical details of each approach and their implementation. In Table A.1 we report statistics for the underlying samples.

3.2.3 A First Glance at Skill Returns

We begin by examining whether labor market returns entail two firm-specific components: (i) a base wage common to all workers within a firm, irrespective of their attributes; (ii) a skill return. The hypothesis of returns heterogeneity, in its most basic form, can be tested with binary skill levels (high or low test scores). Therefore we consider a well-known specification (Abowd et al., 1999, AKM) where firm fixed effects for high and low skill workers are allowed to differ.

For the purposes of this section we construct subsamples corresponding to the largest connected sets of, respectively, high and low ability workers and we condition on firms that are in both of these sets (double-connectedness in skill levels). Since the analysis is carried out separately for cognitive and noncognitive attributes, this is relatively straightforward and does not require that firms be linked through mobility of both skill dimensions. However, in the following sections we examine set connectedness for the case where multiple skills are considered in the same specification.

We classify workers into high cognitive (Stanine $C = \mathbb{1}[c > 5]$) and high noncognitive ($N = \mathbb{1}[n > 5]$). Then, to exclude potential serial correlation within employment spells from estimated standard errors, we select observations within a two-year set and separately estimate linear binary models of worker and firm effects of the form:

$$\log(w_{ijt}) = \mu_i^S + \theta_j^S + \varepsilon_{ijt}, \tag{3.1}$$

where $S \in \{C=0, C=1\}$ or $S \in \{N=0, N=1\}$ indicates the skill subsample while t takes on the values of the two years selected out of the 1999 to 2008 period. Figure 3.1 plots results for $t \in \{2004, 2007\}$. We use non-adjacent years (in fact, we employ pairs two years apart) to mitigate the impact of partial employment spells during contiguous years when workers switch firms. Various other year pairs are reported in Appendix A.2.3.

As noted before, even in connected samples one should be wary of incidental parameter bias due to limited worker mobility. A simple comparison of firm effects (θ_j^S) for high and low skill workers illustrates this point and shows that the statistics of firm effects (like their variances and correlations) can be biased if identified from few moves of workers into and out of each firm. Panels (a)–(b) in Figure 3.1 plot a scatter of estimated firm fixed effects for high-skill (x-axis) and low-skill (y-axis) workers. The samples consist of firms that are in the leave-one-out connected sets of both high and low ability workers. Each panel refers to a given skill attribute, covering the years 2004 and 2007. Panel (a) shows results for cognitive skills (9,268 firms) while Panel (b) plots those for noncognitives (10,208 firms). The figures show that ignoring estimation biases results in firm effects for high and low skill workers that are positively but weakly correlated within firms. The regression slope from mechanically projecting $\theta_j^{S=0}$ onto $\theta_j^{S=1}$ is 0.31 for cognitive traits and 0.35 for noncognitive. We refer to these slopes as the "plug-in" estimates and employ our two approaches to account for the attenuation biases.

The bias correction raises estimated slopes to 0.63 and 0.81, respectively. Under the null hypothesis of no heterogeneity in skill returns, however, the slopes should be statistically indistinguishable from one and the scatters should align along the dashed 45-degree lines. This is not the case, as the bias-corrected test statistics of equal firm effects for high and low skill workers have z-values above 4 for both cognitive and noncognitive returns. We therefore reject the hypothesis that firm effects are independent of worker skills. In fact, all our estimates indicate slopes that are well below one. Table A.2 in the appendix reports additional tests, which similarly reject the null hypothesis of homogeneous returns in several alternative samples.[6]

Panels (c)–(d) of Figure 3.1 show results when grouping firms into 100 clusters – as explained in Section 3.2.2 (see Bonhomme et al., 2019). Estimates of the slopes (0.66 for cognitive and 0.85 for noncognitive) are remarkably similar to those obtained using the quadratic form correction. Also in this case, the null hypotheses that firm effects are the same for high- and low-skill workers are strongly rejected, further discarding the notion of a homogeneous return to skills across firms. Tests for additional year pairs lead to similar conclusions and are presented in Table A.2.

[6] All tests of parameters of equation (3.1) are based on an upper bound for the estimated error variance var(ε_{ijt}). This leads to conservative test statistics compared to the split-sample estimate in Figure 1 of Kline et al. (2020). Joint tests of the equal effects hypothesis across more than two periods are unfeasible as they introduce issues with clustering of errors at the firm level and no robust procedure is currently available to handle such issues. We thank Raffaele Saggio for feedback and discussions about implementing these tests.

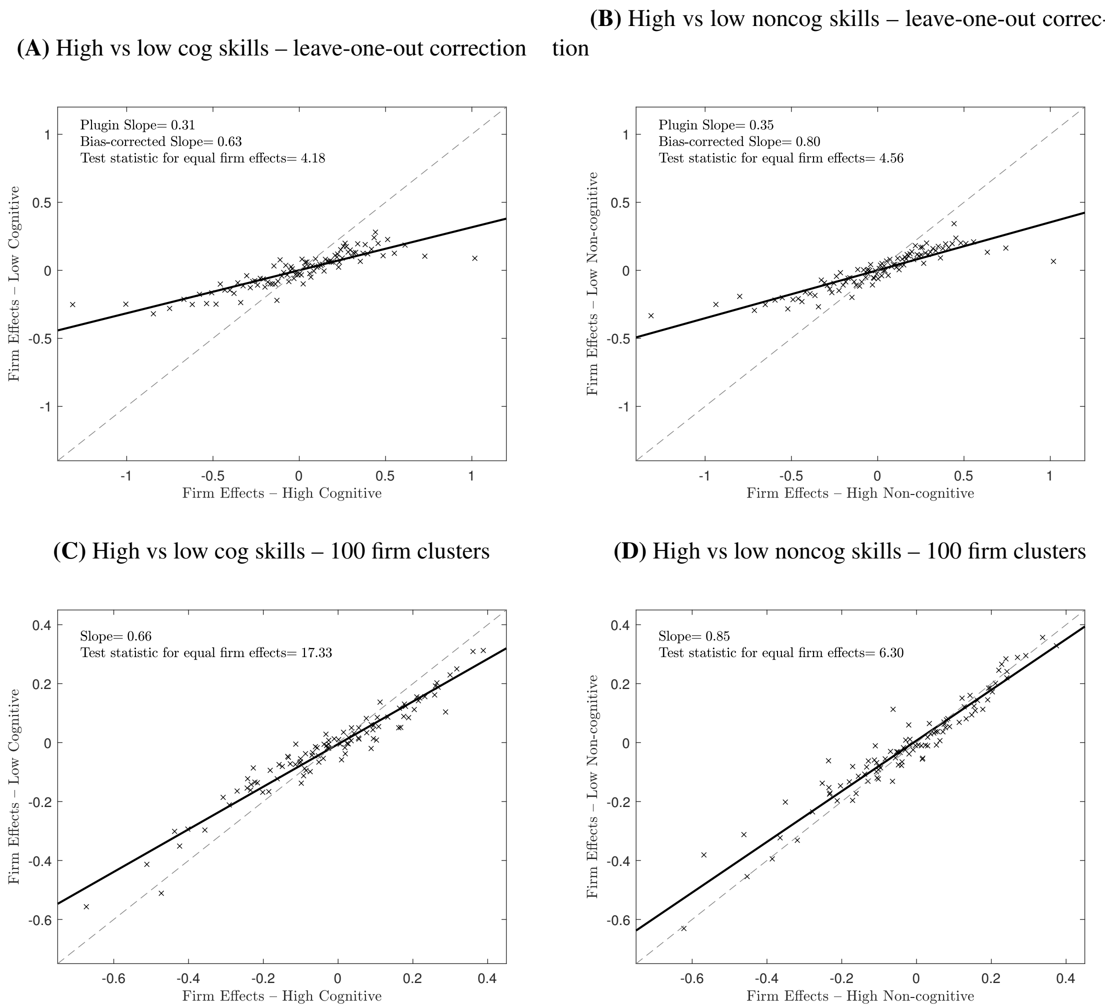


Figure 3.1: Firm effects heterogeneity: cognitive and noncognitive skills.

Top panels: Figures plot the averages of firm effects for low-skill workers ($\theta_j^{S=0}$) against the averages of firm effects for high skill workers ($\theta_j^{S=1}$), where $S \in \{C,N\}$. All sets of firm effects are demeaned. The sample in panel (a) consists of 9,268 firms that are leave-one-out connected in both high and low cognitive skills; in panel (b) we use 10,208 firms connected in both high and low noncognitive skills. The "plugin slope" is the coefficient from a person-year weighted projection of $\theta_j^{S=0}$ onto $\theta_j^{S=1}$. The "bias-corrected slope" adjusts the plug-in slope for attenuation bias by multiplying its value by the ratio of the plug-in estimate of the person-year weighted variance of $\theta_j^{S=1}$ to the bias-adjusted estimate of the same quantity. "Test Statistic" refers to the realization of $\hat{z}_{H_0}/\sqrt{\hat{var}(\hat{z}_{H_0})}$ where $\hat{z}_{H_0}$ is the quadratic form associated with the null hypothesis that the firm effects are equal across skill groups. From Theorem 2 in Kline et al. (2020), $\hat{z}_{H_0}/\sqrt{\hat{var}(\hat{z}_{H_0})}$ converges to a $N \sim (0,1)$ under the null hypothesis that $\theta_j^{S=0} = \theta_j^{S=1}$ for, respectively, all 9,268 and 10,208 firms.
Bottom panels: These figures plot the averages of firm effects for low-skill workers ($\theta_j^{S=0}$) against the averages of firm effects for high skill workers ($\theta_j^{S=1}$), where $S \in \{C,N\}$. Firm effects are estimated for 25,783 firms grouped into 100 clusters based on workers' average cognitive skill, noncognitive skill, and earnings. Firm effects are demeaned. The "Test Statistic" is for the null hypothesis that estimated firm effects are equal across skill groups.
Sample restriction: years 2004 and 2007 only. Tests for other year pairs are in Appendix Table A.2.

3.3 Quantifying Variation in Skill Returns

The previous section emphasizes the significant differences in firm intercepts by skill level. However, to accurately examine the extent of variation in skill returns an empirical framework is needed that allows for granular differences in skill bundles while controlling for other sources of heterogeneity. To this end, we derive a richer empirical baseline from a simple model of demand for productive skills. Our theoretical restrictions aid in the interpretation of estimated parameters and facilitate comparisons with existing work.

3.3.1 Skill Demand by Heterogeneous Firms

We embed return heterogeneity in a model in which firms choose how many workers to hire based on demand for their output. We let firms differ in four dimensions: (i) cognitive returns; (ii) noncognitive returns; (iii) demand in their output market, where they have varying degrees of monopoly power; and (iv) cost of labor in the input market, driven by differences in non-pecuniary firm characteristics valued by employees.

Monopoly power in the output market implies a *skill-independent* firm surplus, which underpins the cross-sectional variation in firm base-wages reflected in fixed effects. On the other hand, firm-specific labor supply curves (input market heterogeneity) imply rents for both workers and firms (Card et al., 2018, Lamadon et al., 2022). These assumptions are sufficient to characterize the components of firm wage premia. To this structure we overlay a production technology with heterogeneous skill returns. Derivations are in Appendix A.3.

Production Complementarities

Consider an environment with two heterogeneous sides (workers, firms), with a measure one of workers who differ in their observable cognitive (c) and noncognitive (n) abilities. Firms are indexed by j and workers by the vector (c,n) of their skills. Firm j matched with a (c,n) worker produces output $y = f_j(c,n)$. Assuming technology is constant returns to scale (CRS) in worker headcounts, a j firm matched with k workers of type (c,n) produces $k \times f_j(c,n)$, while a j firm matched with one (c_1,n_1) and one (c_2,n_2) worker produces $f_j(c_1,n_1) + f_j(c_2,n_2)$. [7] Hence the total output of firm j hiring fraction $q_j(c,n)$ of total (c,n) type workforce is

$$y_j = \int f_j(c,n) q_j(c,n) \, \mathrm{d}G(c,n). \tag{3.2}$$

where G is the population measure of different worker types in the economy.

[7] This is a variation of Eeckhout and Kircher (2018)'s assortative matching production setup for large firms and multiple skill inputs. The production function is defined at the level of the match (see Lise and Robin, 2017).

Labor Supply

A worker's utility from being matched with a specific firm depends on his wage plus a preference shock. For worker i of type (c,n), the utility of working at firm j with wage $w_j(c,n)$ is

$$u_{ij}(c,n) = \beta \log(w_j(c,n)) + \nu_{ij} \tag{3.3}$$

where ν_{ij} is the idiosyncratic preference for working at firm j. Shocks ν_{ij} are independent draws from a Type I Extreme Value distribution. This specification could be easily expanded to add firm-level variation in average amenities (Sorkin, 2018).

Workers choose the firms that give them the highest utility and, using standard arguments (McFadden, 1974), the share $q_j(c,n)$ of type (c,n) workers who choose firm j has logit form

$$\log(q_j(c,n)) = \log(h(c,n)) + \beta \log(w_j(c,n)). \tag{3.4}$$

Equation (A.6) describes the upward sloping labor supply faced by firm j. The intercept $h(c,n)$ is determined in equilibrium and guarantees market clearing so that each worker gets a job,

$$h(c,n) = \left[\int w_j(c,n)^{\beta} \, \mathrm{d}F(j) \right]^{-1} \tag{3.5}$$

where $F(\cdot)$ is the measure describing the distribution of firms in the economy.

Technology and Wages

Given the simple structure outlined above, firm j's output is given by (3.2).

Final good. Each firm's output is an intermediate input for a final good Y produced by a representative firm through a CES technology, $Y = \left[\sum_{j=1}^{J} \varphi_j y_j^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$ where $\sigma > 1$ is the elasticity of substitution of intermediates. Each intermediate's share parameter φ_j is the marginal contribution of y_j to output Y and can be interpreted as the output market power of a firm.

Wages. In Appendix A.3 we provide an analytical characterization of firm-specific wages offered to each skill set and define a stationary equilibrium in the labour market. We also show that a firm's optimal behavior implies:

$$w_j(c,n) = \underbrace{\frac{\beta}{1+\beta}}_{\text{Monops.Markdown}} \times \underbrace{T_j \varphi_j \frac{\sigma-1}{\sigma} \left(\frac{1}{y_j} \right)^{\frac{1}{\sigma}}}_{\text{Marg.Revenue}} \times \underbrace{e^{\Delta_j(c,n)}}_{\text{Skill Productivity}} . \tag{3.6}$$

The wage paid by firm j reflects different aspects of market structure and technology. The monopsonistic firm sets wages at a fraction $\frac{\beta}{1+\beta}$ of the marginal revenue generated by the worker. In turn, the marginal revenue is an increasing function of a firm's output market share φ_j and of its total factor productivity T_j. The latter parameter is normalized to $T_j = f_j(L,l)$, which is the output in firm j of a worker with the lowest cognitive and noncognitive ability. The premium $\Delta_j(c,n) = \log(f_j(c,n)/f_j(L,l))$ is the log output in firm j of a (c,n) type worker relative to a worker with the lowest skill endowment (L,l). The premium associated to the skill vector (c,n) depends on the firm's production technology and on (c,n)'s marginal contribution to output. Equation (A.13) is explicitly derived in Appendix A.3.

In logs, wages are the sum of a common level effect, a firm intercept and a skill return,

$$\log(w_j(c,n)) = \alpha + \Lambda_j + \Delta_j(c,n). \tag{3.7}$$

To obtain an empirical counterpart, we do not restrict the functional form of $f(\cdot)$, and hence of $\Delta_j(c,n)$, but rather use a first-order approximation that delivers a simple bilinear relationship for worker i in firm j.[8] Making the worker index i explicit, the empirical wage representation is:

$$\log(w_{i,j}(c,n)) = \mu_i + \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i, \tag{3.8}$$

where λ_j^0 is the baseline wage that a worker with the lowest endowment in both the cognitive and noncognitive dimension earns in firm j. Gradients λ_j^c and λ_j^n are firm-specific marginal returns, above and beyond the baseline return λ_j^0. Finally, as we show below, the individual intercepts μ_i partly reflect the average (Mincerian) returns to a worker's skill endowments.

Normalizations. It is well known that specifications like (3.8) require linear restrictions on firm effects, since these are only identified relative to a reference firm (or set of firms). It follows that we can identify parameters up to a set of unknown constants $(\kappa_0, \kappa_\text{c}, \kappa_\text{n})$, such that:

$$\begin{aligned} \lambda_j^0 &= \Lambda_j - \kappa_0 \\ \lambda_j^c &= \frac{\partial \Delta_j(c,n)}{\partial c} - \kappa_\text{c} \\ \lambda_j^n &= \frac{\partial \Delta_j(c,n)}{\partial n} - \kappa_\text{n} \\ \mu_i &= \alpha + \kappa_0 + \kappa_\text{c} \cdot c_i + \kappa_\text{n} \cdot n_i \end{aligned} \tag{3.9}$$

[8]For instructive discussions of log-additive firm effects in wage specifications with bundled skills, see Choné and Kramarz, 2021. We explore higher order approximations featuring non-linear returns but the extra flexibility makes little difference. Notably, this type of worker–firm complementarities can be micro-founded by restricting attention to the labor composition alone (e.g., learning and cooperation of workers as in Jarosch et al., 2021).

We set $\kappa_0 = \bar{\Lambda}$, $\kappa_c = \frac{\partial \bar{\Delta}(c,n)}{\partial c}$, and $\kappa_n = \frac{\partial \bar{\Delta}(c,n)}{\partial n}$, where the reference values $(\bar{\Lambda}, \frac{\partial \bar{\Delta}(c,n)}{\partial c}, \frac{\partial \bar{\Delta}(c,n)}{\partial n})$ correspond to the average employment-weighted firm effects. This normalization is quite conservative, since central moments yield the lowest variance of firm heterogeneity (intuitively, they minimize squared deviations).[9]

Unlike models with degenerate skill returns, firm premia are not restricted to be equal across skill groups. Under the model's null hypothesis, within-firm wage variation is a function of worker skill differences as firms with higher λ_j^c or λ_j^n exhibit higher skill premia.

3.3.2 Estimating Skill Returns

The empirical analysis relies on a sample of firms connected through worker mobility along both skill dimensions over the 1999–2008 period. The baseline representation becomes

$$\log(w_{ijt}) = \mu_i + \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i + X_{it} b_t + \varepsilon_{ijt}, \tag{3.10}$$

where λ_j^0 are skill-independent base earnings, λ_j^c and λ_j^n are skill gradients, and μ_i are worker fixed effects. To flexibly account for life-cycle and time variation by skill, we control for interactions of skill type, age, and years, denoted as $X_{it} b_t$ in (3.10).

Identification of firm effects. It is important to emphasize what we can, and cannot, identify from (3.10). While the availability of worker-level skill proxies provides a transparent way to estimate the distribution of firm returns, the distribution of workers' efficiency and firm returns could be jointly identified even in the absence of direct skill measures (Bonhomme et al., 2019, Lamadon et al., 2022). The latter approach would require the assumption that workers moving to a firm are not of similar quality as workers moving out of that firm. While such an assumption is easier to maintain when workers are ranked on a single index, it becomes less suitable in the presence of multiple skill dimensions where no unique ranking of workers is available. The general identification problem in these settings is that workers with different skill mixes may exhibit similar overall productivity. With direct proxies for different skills, returns λ_j^s for $s \in \{c, n\}$ can be identified upon a firm switch by the differential earning changes of workers with different skill levels. Thus the key requirement is that a sufficient number of such switches is observed. Identification of firm intercepts λ_j^0 is also obtained from earnings changes following firm switches.

Interpreting parameters. The level and dispersion of worker fixed effects μ_i partly reflect skill endowments. That is, μ_i includes the average skill return that a worker would get in any firm.

[9]More generally, estimates of central moments tend to be more robust relative to those of the extrema of the firm effects' distribution, which may suffer from non-trivial estimation error.

Moreover, the empirical implementation of μ_i as a fixed effect flexibly accounts for residual dimensions of workers' skills that are priced homogeneously in the market.

We normalize the Stanine scores to take values on the unit interval. Setting a unit upper bound for skills is convenient because each skill return λ_j^s can be interpreted as the earnings gap separating the highest and lowest worker types.[10]

The linear restrictions on firm effects imply that the lowest skill workers gain no employer premium above and beyond firm intercepts. Therefore, for the subset of workers with the lowest skill endowments ($c = 0, n = 0$), equation (3.10) reduces to a standard specification with firm fixed effects λ_j^0, time-varying controls $X_{it}b_t$, and worker fixed effects μ_i. For other skill types, (3.10) augments the double fixed-effect specification by explicitly allowing for heterogeneous returns to skills. By design, if we were to restrict attention to a single skill dummy S over a two year interval, with no other control variables, estimation of (3.10) would collapse back to the binary model in (3.1) where $\lambda_j^0 = \theta_j^{S=0}$ and $\lambda_j^s = \theta_j^{S=1} - \theta_j^{S=0}$.

Interactions of skill, year, and age dummies (in $X_{it}b_t$) flexibly account for potential variation in average skill returns and significantly reduces computation times.[11] Conditional on the latter, worker fixed effects absorb time-invariant residual skill components, as discussed above.

Estimation. As discussed in Section 3.2.2, we report estimates for both the non-clustered leave-out samples and the clustered firm samples. When using the quadratic-form correction, the leave-out samples are defined so that each observation is a unique worker–firm match. That is, we collapse the data to the worker–firm level by averaging variables (earnings, age, time) within each spell a worker has at a given firm. This makes the estimator robust to serial correlation within clusters of observations and yields conservative variance estimates. To employ the group-level estimator, we use the k-means algorithm and partition firms into 100 clusters. The clustering is based on average cognitive and noncognitive scores of employees and on their earnings, consistent with the observation that different production arrangements lead to systematic variation in skill composition within firms. Results are robust to alternative clustering approaches.

[10] The transformation is $(Stanine - 1)/8$ and the distribution of normalized skills is carried over from the Stanine distribution. Normalized scores for c and n are defined on the grid $[0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1]$. Our sampling restrictions have little impact on distribution moments relative to the population of test takers: e.g., $\bar{c} = 0.54$, $\bar{n} = 0.52$, $\text{sd}(c) = 0.24$, $\text{sd}(n) = 0.21$, $\text{corr}(c,n) = 0.36$.

[11] For example, estimation on the leave-out sample takes about 20–30 hours using Python and the JLA approximation. Adding stratified controls raises computation time proportionally to the number of additional parameters. Allowing for time-varying returns to education does not affect results. Life-cycle profiles (by skill and time) are accounted for by the $cognitive \times noncognitive \times age \times year$ group interaction in $X_{it}b_t$. Dummies for $s \leq 0.25$, $0.375 \leq s \leq 0.625$, $0.75 \leq s$ for $s \in \{c,n\}$ are interacted with each other and age groups 20–25, 26–32, 33–42, 42–60 as well as two-year period dummies 1999–2000, 2001–2002, 2003–2004, 2005–2006, 2007–2008.

Table 3.1: Standard deviations of firm parameters, estimates from firm-level sample with quadratic-form correction and from clustering approach.

	Standard deviations		*Standard deviations* $\times$ *(*$90^{th}-10^{th}$ *skill percentile)*	
	firm-level (1)	grouped (2)	firm-level (3)	grouped (4)
sd(λ_j^0)	0.22	0.10		
sd(λ_j^c)	0.15	0.08	*0.11*	*0.06*
sd(λ_j^n)	0.10	0.05	*0.07*	*0.04*
cumulative (cog+noncog score)			*0.19*	*0.10*
# unique firms	19,085	25,783		

Notes: The first two columns show standard deviations of parameters λ_j^0, λ_j^c, and λ_j^n estimated in (3.10). Column (1) quadratic-form corrects variances of the parameters estimated at the individual firm level and takes the square root. Column (2) assigns firms into 100 groups according to their average earnings and average c and n scores using the k-means algorithm. It then estimates (3.10) on this grouped data. Columns (3) and (4) multiply the estimated standard deviations with differences of skills between the 90^{th} (c_i and n_i of 0.875) and 10^{th} (c_i and n_i of 0.125) percentile. Estimation period: 1999–2008.

3.3.3 Estimates of Firm Parameters

Table 3.1 reports estimates of firm returns from specification (3.10) when skills are free to vary over their granular range (i.e., $c_i \in [0, 0.125, .., 1]$). While we initially focus on the dispersion of firm parameters, heterogeneity in skill returns has meaningful implications also for other moments of the earnings distribution through behavioural responses that result in assortative matching patterns. The latter effects are examined in the following sections.

Column (1) shows estimates for the leave-out (non-grouped) sample with bias correction. The first line, $\text{sd}(\lambda_j^0) = 0.22$, highlights that skill-independent premia vary significantly across employers, confirming the well-established relevance of such fixed effects. The estimates in the lines below document a less known layer of firm heterogeneity. In particular, they show that the standard deviations of skill returns are substantial, with $\text{sd}(\lambda_j^n) = 0.10$ for noncognitive skills and $\text{sd}(\lambda_j^c) = 0.15$ for cognitive ones.

Column (2) reports estimates based on the grouped-firms approach. As expected, standard deviations are lower since dispersion within each cluster is restricted to zero by treating all elements within it as a single representative employer. Nonetheless, while delivering a more conservative estimate of the absolute impact of skill returns heterogeneity, variation remains substantial. And, perhaps more interestingly, the relative magnitudes of returns are unchanged as the values of $\text{sd}(\lambda_j^0)$, $\text{sd}(\lambda_j^c)$ and $\text{sd}(\lambda_j^n)$ all approximately halve. The finding of constant *relative* magnitudes is robust throughout the analysis, indicating that estimates of the proportional contribution of each layer of firm heterogeneity do not depend on the estimation method.

A double differencing thought experiment. To convey the magnitude of skill premia, in columns (3) and (4) of Table 3.1 we consider thought experiments whereby workers with different skills are parachuted from their original firm to a different one in which returns are one standard deviation larger. The hypothetical gains of such transitions are reported for high skill workers relative to low skill workers (90^{th} versus 10^{th} percentiles of skills). Based on bias-corrected firm-level sample estimates, moving to a firm that sits just a standard deviation higher in cognitive returns would result in a gain of 11 log points for a worker at the 90th cognitive percentile ($c_i = 0.875$) compared to a worker in the 10th percentile ($c_i = 0.125$). These are considerable differences in the gains from job mobility and, as discussed in Section 3.5, they are elicited through positive assortative matching.

Heterogeneity in noncognitive returns is somewhat lower but still economically significant. Parachuting a worker at the 90th percentile of n_i into a firm that is a standard deviation higher in noncognitive returns raises their earnings gap relative to someone at the 10th percentile of n_i by seven log points. Jointly, a one-standard deviation change in both cognitive and noncognitive returns for workers at the 90th, rather than the 10th, percentile of each skill bring about an impact that is roughly as large as that of firm intercepts (see the cumulative effect in the last line of Table 3.1). The relative magnitude of the joint impact is similar for either of the estimation approaches.[12]

The finding of significant dispersion in skill returns is also robust in several respects. For example, Appendix A.4.1 shows that the bias correction approach in the leave-observation-out sample (rather than the leave-match-out sample) delivers even higher dispersion of skill returns. In sensitivity checks we also show that, when varying the number of firm clusters in the grouping estimator, the relative magnitude of skill returns and firm intercepts is unchanged.

The cross-section of skill returns. To characterize the cross-sectional distribution of firm returns we adopt as a baseline the estimates in column (2) of Table 3.1. Estimates based on the leave-out bias correction indicate even larger returns heterogeneity.

Figure 3.2 shows histograms of cognitive and noncognitive returns in the cross-section of firm clusters. As described earlier, the average λ_j^c and λ_j^n are normalized to zero. Return heterogeneity is significant in both dimensions although larger for cognitive traits, since $\mathrm{sd}(\lambda_j^{\mathrm{c}}) = 0.080$ and $\mathrm{sd}(\lambda_j^{\mathrm{n}}) = 0.048$.

Dispersion is stable across time periods, with $\mathrm{sd}(\lambda_j^{\mathrm{c}}) = 0.095$ and $\mathrm{sd}(\lambda_j^{\mathrm{n}}) = 0.052$ in 1990–1999 and $\mathrm{sd}(\lambda_j^{\mathrm{c}}) = 0.074$ and $\mathrm{sd}(\lambda_j^{\mathrm{n}}) = 0.048$ in 2008–2017 (see Appendix A.4.1). Edin et al. (2022)

[12] The joint estimated gains are 19 and 10 log points, respectively. By comparison, moving to a firm where λ_j^0 is one standard deviation larger raises a worker's earnings by 22 log points in the firm-level and 10 log points in the grouped estimates.

show that the *average* return to noncognitive skills increased while that of cognitive skills declined (see also Beaudry et al., 2016, Deming, 2017, for the U.S.). Our analysis suggests that, at the same time, the *heterogeneity* of skill returns across firms did not change differentially for cognitive and noncognitive skills.

(A) Cognitive, $\mathrm{sd}(\lambda^{c}) = 0.080$ **(B)** Noncognitive, $\mathrm{sd}(\lambda^{n}) = 0.048$

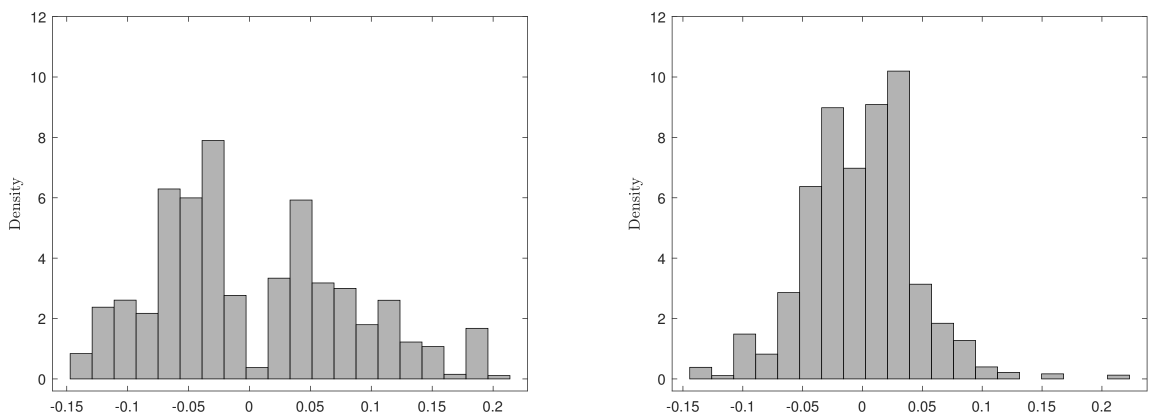


Figure 3.2: Histograms of Firm Returns (20 Bins)

Notes: Estimates of λ^{c} (left panel) and λ^{n} (right panel), based on 100 firm clusters weighted by employment. $\mathrm{corr}(\lambda^{c}, \lambda^{n}) = 0.083$. Grouped estimator for period: 1999–2008.

The employment-weighted correlation of returns among firm clusters, $\mathrm{corr}(\lambda^{c}, \lambda^{n})$, is positive at 0.083.[13] Imperfect correlation lends support to the hypothesis that firm heterogeneity is genuinely multidimensional and that parameters can be independently identified through observable proxies that account for the skill-dependent ranking of workers.

Earnings gaps and skill premia. The plots in Figure 3.2 show that cognitive returns are concentrated between -15 and $+20$ log points. Relative to a worker from the 10th percentile of skills, a worker from the 90th percentile who moves from the bottom to the top of the returns distribution would gain 25 extra log points in earnings. That is, the difference in the cognitive premium between these workers is the skill difference $(0.875 - 0.125 = 0.75)$ multiplied by 35 log points. Complementarity of skills and returns implies that the earning function should be convex over skills because large earning effects accrue from matching high c workers to high λ^{c} firms. Noncognitive returns can also add significantly to these earning differences. It follows that the impact of returns heterogeneity on the distribution of earnings hinges on the intensity of assortative matching and, in Section 3.4, we derive testable restrictions to gauge the prevalence of assortative matching in

[13]Using the firm-level estimates of column (1) in Table 3.1, the bias-corrected correlation is 0.27.

data. Then, in Section 3.5, we examine how firm heterogeneity, and the responses it elicits, shape the earnings distribution, and contrast our estimates to a counterfactual with random assignment of workers.

3.4 Matching

How much do cognitive and noncognitive traits matter for the assignment of workers to employers? And how do they affect the distribution of earnings? To study these questions we analytically characterize worker–firm matching in a setting with multiple skill attributes (Lindenlaub, 2017).

First, we introduce notation. Firms differ in three dimensions: their earnings intercept (λ_j^0) as well as cognitive (λ_j^{c}) and noncognitive (λ_j^{n}) returns. We define a matching function $\mu(\overline{c}_j, \overline{n}_j) = (\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})$, which maps a firm's average worker skills into its returns. One can show that, under the assumption of upward sloping firm-specific labor supplies (equation A.6), the matching function μ should be increasing in $\overline{c}_j$ and $\overline{n}_j$, and multidimensional PAM should hold as defined below. In what follows we examine the empirical content of these restrictions.

3.4.1 Sorting Patterns

Assortative matching, whether positive (PAM) or negative (NAM), is characterized by the properties of the matching function's derivatives. In matching problems with one dimensional heterogeneity this boils down to the sign of a single derivative. With multiple attributes, all elements of the Jacobian play a role.

Definition 6. *The sorting pattern is locally PAM if, for given $(\overline{c}, \overline{n})$, the following holds:*

$$\text{(a) } \frac{\partial \lambda_j^{\mathrm{c}}}{\partial \overline{c}_j} > 0; \qquad \text{(b) } \frac{\partial \lambda_j^{\mathrm{n}}}{\partial \overline{n}_j} > 0; \qquad \text{(c) } \frac{\partial \lambda_j^{\mathrm{c}}}{\partial \overline{c}_j}\frac{\partial \lambda_j^{\mathrm{n}}}{\partial \overline{n}_j} - \frac{\partial \lambda_j^{\mathrm{n}}}{\partial \overline{c}_j}\frac{\partial \lambda_j^{\mathrm{c}}}{\partial \overline{n}_j} > 0.$$

Hence, to examine assortative matching we focus on the Jacobian of the matching function:

$$\frac{\mathrm{d}\mu(\overline{c}_j, \overline{n}_j)}{\mathrm{d}(\overline{c}_j, \overline{n}_j)} = \begin{bmatrix} \frac{\partial \lambda_j^{\mathrm{c}}}{\partial \overline{c}_j} & \frac{\partial \lambda_j^{\mathrm{n}}}{\partial \overline{c}_j} \\ \frac{\partial \lambda_j^{\mathrm{c}}}{\partial \overline{n}_j} & \frac{\partial \lambda_j^{\mathrm{n}}}{\partial \overline{n}_j} \end{bmatrix} \tag{3.11}$$

The Matching Jacobian in data. Intuitively, it does not matter for the empirical test of PAM whether firms choose workers or vice versa. This means that there are different ways to test the sorting hypothesis. We pursue two alternative routes, consistent with the previous analysis. First, we consider sorting regressions based on the Jacobian matrix defined above:

$$\begin{aligned} \lambda_j^{\mathrm{c}} &= d_{1\mathrm{c}} + d_{2\mathrm{c}}\overline{c}_j + d_{3\mathrm{c}}\overline{n}_j + e_j^{\mathrm{c}} \\ \lambda_j^{\mathrm{n}} &= d_{1\mathrm{n}} + d_{2\mathrm{n}}\overline{c}_j + d_{3\mathrm{n}}\overline{n}_j + e_j^{\mathrm{n}}. \end{aligned} \tag{3.12}$$

Table 3.2: Projection of Individual Firms' Returns onto their Average Skills.

	Dependent Variables:					
	(1)		(2)		(3)	
	λ_j^{c}	λ_j^{n}	λ_j^{c}	λ_j^{n}	λ_j^{c}	λ_j^{n}
$\overline{c}_j$	0.29	-0.41	0.29	–0.41	0.16	–0.44
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
$\overline{n}_j$	0.15	0.61	0.15	0.61	0.40	0.56
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
# firms	19,085		19,085		19,085	
Controls	No		# employees		No	
Weights	No		No		# employees	

Notes: The table reports sorting coefficients d_2 and d_3 from estimating (3.12) with individual firm λ_j^{c} and λ_j^{n}. Projections of individual coefficients in estimation period 1999–2008. Standard errors are corrected to account for the first-stage estimates of the outcome variable as in Kline et al. (2020, Section 4).

The linear forms in (3.12) are similar to the projections of fixed effect onto firm characteristics used in the applied literature (Kline et al., 2020). A strength of this specification is that, under general assumptions, the regression parameters can be correctly estimated from a cross-section of individual non-grouped firms. If returns are measured with error, having λ_j^{c} and λ_j^{n} on the left-hand-side avoids biases in the d-parameters of (3.12). One can then use these linear projections to test for PAM in the cross-section of individual firms; this is true even if other statistics, such as the R^2, are potentially biased. One caveat is that, while point estimates from these regressions are generally unbiased, standard errors must be corrected for the correlation across the first-stage estimates of the outcome variable (firm parameters).[14]

Table 3.2 reports estimates from projections in (3.12), obtained from non-grouped firm-level data (employees' cognitive and noncognitive skills are averaged into firm-specific $\overline{c}_j$ and $\overline{n}_j$). It is apparent that PAM cannot be rejected since own-partial derivatives and the determinant of the Jacobian are positive throughout. The coefficients on $\overline{c}_j$ for λ_j^{c} are only about half as large as on $\overline{n}_j$ for λ_j^{n}. Flipping this around, $\overline{c}_j$ responds more to a given difference in returns, which implies stronger sorting on cognitive traits. Below we present additional evidence of uneven sorting patterns.

An alternate test of skill sorting. A different route to test PAM is based on a matching Jacobian where average skills $\overline{c}_j$ and $\overline{n}_j$ are projected onto firm returns. This builds on a definition of the matching function that maps firm gradients into worker characteristics, and provides a way to examine matching patterns where skill sorting in each dimension depends on both of the employer's

[14]We use the correction proposed in equation (7) of Kline et al. (2020) to construct adjusted standard errors.

Table 3.3: Projection of Average Skills onto Grouped Returns.

	Dependent Variables:					
	(1)		(2)		(3)	
	$\overline{c}_j$	$\overline{n}_j$	$\overline{c}_j$	$\overline{n}_j$	$\overline{c}_j$	$\overline{n}_j$
λ_j^{c}	1.21	0.58	1.18	0.55	1.15	0.53
	(0.08)	(0.07)	(0.07)	(0.06)	(0.07)	(0.05)
λ_j^{n}	–0.15	0.61	–0.05	0.71	–0.14	0.61
	(0.11)	(0.08)	(0.10)	(0.07)	(0.11)	(0.07)
R^2	0.676	0.542	0.712	0.612	0.752	0.648
# firms	25,783		25,783		25,783	
Controls	No		λ_j^0, # employees		λ_j^0	
Weights	No		No		# employees	

Notes: Column (1) reports sorting coefficients δ_2 and δ_3 from estimating (3.13). The specification in column (2) additionally controls for intercepts λ^0 and for the firms' total employment headcounts. Column (3) weights the observations by the firm's number of employees. Each firm is one observation. Robust standard errors clustered at the level of the 100 firm groups are in parentheses. Grouped estimator for period 1999–2008.

returns. In practice, we posit $\mu(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}) = (\overline{c}_j, \overline{n}_j)$ and test Jacobian conditions[15] using the projections in (3.13). It is important to recognize that, if the λ parameters are measured with error due to limited mobility, estimation of (3.13) may deliver biased point estimates. To mitigate such concerns, we adopt a grouped-firm approach and project average skills (cognitive or noncognitive) onto the 100 cluster-specific returns. The grouping does not hardwire the relationships in (3.13) since cluster-level returns are free to vary. Table 3.3 reports estimates of the Jacobian for the following specifications:

$$\begin{aligned} \overline{c}_j &= \delta_{1\mathrm{c}} + \delta_{2\mathrm{c}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{c}}\lambda_j^{\mathrm{n}} + \varepsilon_j^{\mathrm{c}} \\ \overline{n}_j &= \delta_{1\mathrm{n}} + \delta_{2\mathrm{n}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{n}}\lambda_j^{\mathrm{n}} + \varepsilon_j^{\mathrm{n}}. \end{aligned} \tag{3.13}$$

The regressions in (3.13) deliver the best linear approximation to the conditional expectations of $\overline{c}_j$ and $\overline{n}_j$. For instance, $\mathrm{E}(\overline{c}_j|\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}) = \delta_{1\mathrm{c}} + \delta_{2\mathrm{c}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{c}}\lambda_j^{\mathrm{n}}$, so that the parameter $\delta_{2\mathrm{c}}$ is the expected value of the top-left element $\left(\frac{\partial \overline{c}_j}{\partial \lambda_j^{\mathrm{c}}}\right)$ of the Jacobian taken over the sample of all firms. Similar arguments hold for $\delta_{3\mathrm{c}}$ and gradients in the second line of (3.13).

The positive and highly significant $\delta_{2\mathrm{c}}$ and $\delta_{3\mathrm{n}}$ in Table 3.3 imply that the own-derivative conditions for PAM are satisfied for both c and n. The Jacobian is also positive definite, with determinant $\delta_{2\mathrm{c}}\delta_{3\mathrm{n}} - \delta_{3\mathrm{c}}\delta_{2\mathrm{n}}$ larger than zero. This lends additional support to the hypothesis that PAM holds over the 1999–2008 period in our large sample of firms.

[15] The Jacobian becomes $\frac{\mathrm{d}\mu(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})}{\mathrm{d}(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})} = \begin{bmatrix} \frac{\partial \overline{c}_j}{\partial \lambda_j^{\mathrm{c}}} & \frac{\partial \overline{c}_j}{\partial \lambda_j^{\mathrm{n}}} \\ \frac{\partial \overline{n}_j}{\partial \lambda_j^{\mathrm{c}}} & \frac{\partial \overline{n}_j}{\partial \lambda_j^{\mathrm{n}}} \end{bmatrix}$.

The positive δ_{2n} in equation (3.13) indicates substantial cross-sorting of high n workers to high λ_j^c firms, which occurs when skill endowments are correlated. High c workers who sort into high cognitive return firms also have a higher endowment of n skills. Consistent with this observation, own-sorting in the c dimension is strong, as shown by the large δ_{2c} estimates and in Figure 3.3 below. Cross-sorting of c to high λ_j^n firms is not present and sorting in the n dimension is substantially weaker (see also Figure 3.3 below).[16]

Results do not change after controlling for firm-specific employment size and firm intercepts λ^0, as shown in column (2). Neither do they change when weighting by employment size as shown in column (3). Between 54 and 68 percent of the skill variation between firm clusters is accounted for by differences in estimated λ^c and λ^n returns alone.[17] When we weigh firms by their employment size and control for λ^0, the explained variation rises to 65–75 percent.

3.4.2 The Distribution of Workers over Returns

If workers with a high endowment of a particular skill match more frequently with firms with high returns to that skill (in the sense of first-order stochastic dominance or FOSD), then sorting is positive along that dimension (Lindenlaub and Postel-Vinay, 2020). A way to visualize such patterns is to compare the cumulative distribution functions (CDF) of returns for separate sets of workers (say, high versus low cognitive skills).

First-order stochastic dominance. Figure 3.3 illustrates sorting patterns along either cognitive or noncognitive attributes, using the grouped-firm estimates. The top panel plots the CDF for workers in three coarse skill-specific ranks (low, medium or high). The CD functions are defined over the ordered set of estimated firm returns.[18]

In Panel (a) we condition on medium noncognitive skills and show that workers with higher cognitive attributes match with higher cognitive returns λ_j^c. The CDF of high cognitive workers dominates all other types, and the CDF of medium cognitive workers dominates the CDF of low cognitive workers. Panel (b) shows FOSD patterns across ranks of noncognitive attributes (n), holding cognitive attributes fixed at the medium rank. Sorting patterns on noncognitive traits are less striking but clearly discernible.

[16] Correlation of λ_j^c and λ_j^n would affect cross-sorting estimates if we had not controlled for each respective other indirect return in (3.13).

[17] A feature of the grouping approach is that the R^2 is essentially between firm classes, since average skills vary little within k-means clusters. Indeed, averaging $\overline{c}_j$ and $\overline{n}_j$ within groups and running the regression for group averages (i.e., 100 observations) gives only a tiny increase in R^2. Nonetheless, it is remarkable that returns can explain so much of the cross-clusters skill variation.

[18] To ease exposition, we coarsen the skill levels to low ($c,n \leq 0.25$), mid ($0.25 < c,n < 0.75$), and high ($c,n \geq 0.75$). Returns are estimated through the clustering approach. This is graphically convenient as it restricts variation on the x-axis. Estimates based on leave-out-samples without clustering deliver similar insights. Additional FOSD plots are in Appendix A.4.3.

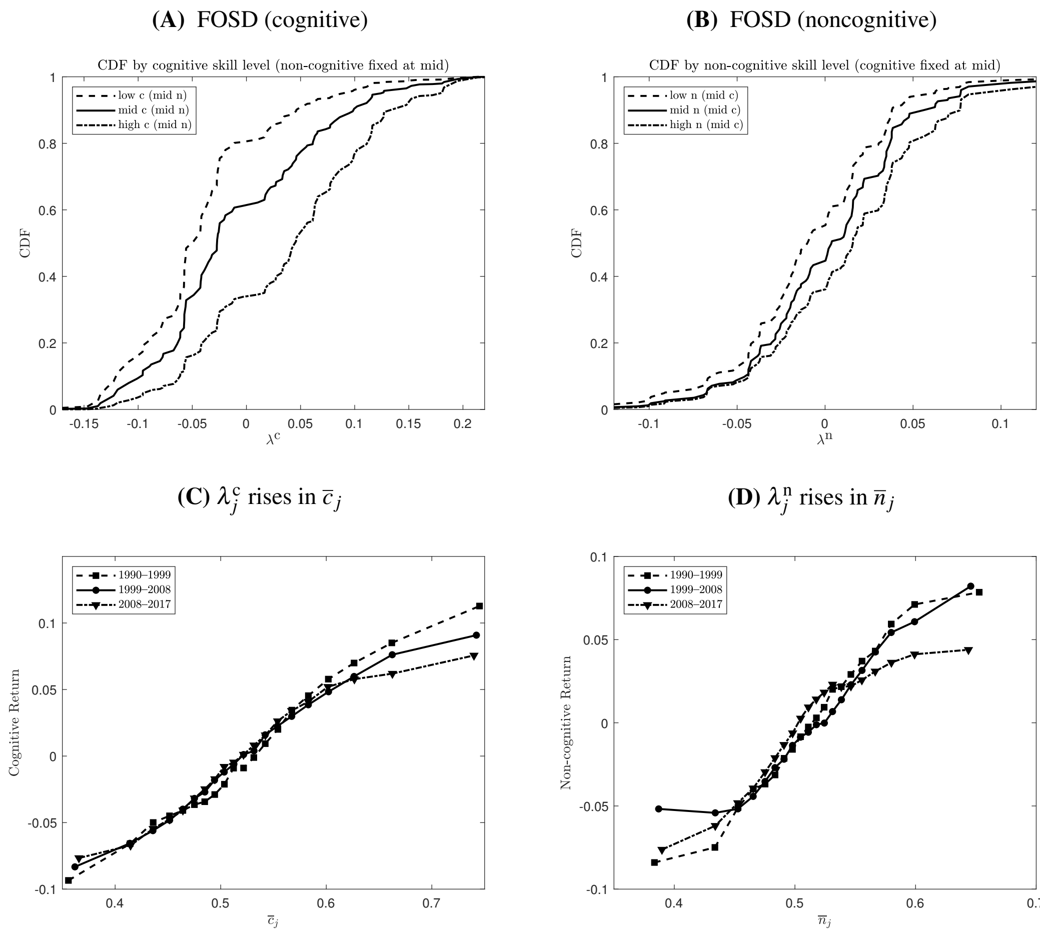


Figure 3.3: Distributions of firm returns for different sets of worker skills.

Notes: Results from the grouped estimator. Panels (a) and (b) show cumulative distribution functions for workers with low ($c,n \leq 0.25$), mid ($0.25 < c,n < 0.75$), or high ($c,n \geq 0.75$) skill ranks over the range of firm returns. Period: 1999–2008. FOSD: first-order stochastic dominance.

Panels (c) and (d) show binned scatterplots of firm-specific skill returns (vertical axis) with average skills (horizontal axis) for three ten-year estimation periods: 1 (1990–1999), 2 (1999–2008), 3 (2008–2017).

In the bottom panels of Figure 3.3 we use an alternative way to visualize the distribution of skills over returns by plotting skill returns over within-firm average skills. These measures of central tendency confirm that returns increase monotonically with skill endowments, consistent with PAM. Between-firm differences in average skills are larger in the cognitive dimension, as expected given the higher dispersion of λ^c relative to λ^n and the stronger sorting incentives. Similar patterns hold for different estimation periods, suggesting that workers consistently sort across firms based on their attributes.

3.5 Complementarities and Earnings

Experiments where workers are parachuted into firms with higher returns, like the ones in Section 3.2.3, are not wholly informative about the actual impact of complementarities on the earnings distribution due to the non-random nature of firm assignment. In what follows, we cast earnings differences due to firm heterogeneity in terms of deviations from cross-sectional means that explicitly account for assortative matching.

3.5.1 Effects on the Distribution of Earnings

Equation (3.10) emphasizes that, after controlling for confounding effects, the return for worker i with skill bundle $s_i = (c_i, n_i)$ working in firm j can be represented as:

$$\log(w_j(s_i)) = \underbrace{\mu_i}_{\substack{\text{(a)}\\ \text{Person effect}\\ \text{(incl. Mincer returns)}}} + \underbrace{\lambda_j^0}_{\substack{\text{(b)}\\ \text{Firm intercept}}} + \underbrace{\lambda_j^{\mathrm{c}}\bar{c} + \lambda_j^{\mathrm{n}}\bar{n}}_{\substack{\text{(c)}\\ \text{Firm returns effect}}} + \underbrace{\lambda_j^{\mathrm{c}}\widetilde{c}_i + \lambda_j^{\mathrm{n}}\widetilde{n}_i}_{\substack{\text{(d)}\\ \text{Match effect}}}, \quad (3.14)$$

where $\widetilde{x}_i$ denotes the deviation of skill x_i from its cross-sectional average $\bar{x}$.

Equation (3.14) has an intuitive interpretation: the term (a) contains the homogeneous Mincerian return $\kappa_{\mathrm{c}} c_i + \kappa_{\mathrm{n}} n_i$, which is often estimated in survey data when skill measures are available (the κ loadings are defined in Section 3.3); component (b) is a firm fixed effect that captures constant differences above and beyond the Mincerian return in (a). The elements (c) and (d) reflect, respectively, the direct impact of firm return hetorogeneity on the earnings of an average-skill person and the more nuanced effect of assortative matching. Terms (c) and (d) add up to the premium $\lambda_j^{\mathrm{c}} c_i + \lambda_j^{\mathrm{n}} n_i$ and jointly subsume firm returns that vary with worker skills. The expected value of the (c) term in (3.14) is nil because $\mathrm{E}(\lambda_j^{\mathrm{c}}) = \mathrm{E}(\lambda_j^{\mathrm{n}}) = 0$. In contrast, the expected value of component (d) can be different from zero as it reflects the per capita wage gains due to assortative matching of workers to firms.

We note that, if skill measures were not available, components (b) and (c) would get conflated into the firm fixed effect, while the skill dependent variation would be absorbed within the person fixed effect μ_i. Separate identification of the impacts of heterogeneous returns and match-quality in summands (c) and (d) can be obtained only when proxies of skill endowments are available.

Lastly, through variance decompositions (Appendix A.4.2) it is possible to show that, if firm heterogeneity is restricted to fixed effects, a share of the earnings variance due to heterogeneous skill returns is improperly attributed to employer intercepts as if they were independent of skills.

Components of permanent heterogeneity. The impact of the components of equation (3.14) on earnings dispersion is summarized in Panel A of Table 3.4 where we present estimates for both the clustering approach and the firm-level estimation with bias correction.

Table 3.4: Contributions of Firm Heterogeneity to Dispersion and Levels of Earnings

Panel (A)	Dispersions: firm-level (1)	grouped (2)	Panel (B)	Levels ($\times$100): firm-level (3)	grouped (4)
sd(μ_i)	0.49	0.43		—	—
sd(λ_j^0)	0.22	0.10		—	—
sd($\lambda_j^c c_i$)	0.09	0.05	E($\lambda_j^c c_i$)	0.66	0.75
sd($\lambda_j^n n_i$)	0.06	0.03	E($\lambda_j^n n_i$)	0.17	0.13
sd($\lambda_j^c c_i + \lambda_j^n n_i$)	0.12	0.06	E($\lambda_j^c c_i + \lambda_j^n n_i$)	0.83	0.88
# unique firms	19,085	25,783		19,085	25,783

Notes: Panel (A) shows the dispersion of each summand in equation (3.14), i.e., the standard deviations of person and firm intercepts, and the standard deviations of the products of returns and skills. Panel (B) shows the averages of the last two summands in equation (3.14), i.e., the contribution of matching to average earnings in the economy (through complementarity gains). Firm-level estimates in column (3) are based on the observation-level, rather than the match-level, leave-out sample to capture the gains from matching in the population of workers. The averages of person and firm intercepts are uninformative due to the normalization of firm parameters and are omitted from Panel (B). Estimation period: 1999–2008.

In line with other studies, unobserved worker heterogeneity has a strong impact on earnings through fixed effects μ_i. The latter include the average returns to skills. The contribution of the heterogeneous components of skill returns to earnings dispersion is between 55% and 60% of that of firm fixed effects.

Returns heterogeneity and sorting lead, on average, to higher earnings. The latter gains can be measured through the covariance of skills and firm returns. For example, if we consider cognitive skills, we have that $\mathrm{E}(\lambda_j^c c_i) = \mathrm{cov}(\lambda_j^c, c_i) = \mathrm{cov}(\lambda_j^c, \bar{c}_j)$. This equivalence confirms that sorting determines the intensity of the average gain accruing from returns' heterogeneity.[19] Panel (B) of Table 3.4 shows estimates of the average gain from match effects, which are between 0.8 and 0.9 log points. The larger gains from heterogeneity in cognitive returns reflect the stronger sorting in that dimension, as documented in Section 3.4.

3.5.2 The Uneven Gains from Sorting

The gains from sorting are unevenly distributed and non-monotonic. They are positive and large for high skill workers, absent for the least skilled workers and negative for a wide range of intermediate

[19] The sorting parameters estimated in equation (3.12) are, in essence, just this gain standardized by the underlying variance of average skills across firms, $\frac{\mathrm{cov}(\lambda_j^c, \bar{c}_j)}{\mathrm{var}(\bar{c}_j)}$. The equality $\mathrm{E}(\lambda_j^c c_i) = \mathrm{cov}(\lambda_j^c, c_i)$ follows from the fact that excess skill returns have zero mean, that is $\mathrm{E}(\lambda_j^c) = 0$.

skills. These patterns can be illustrated by taking expectations of equation (3.14) *after conditioning* on a given skill level.

For brevity, we discuss gains from cognitive skills but similar arguments hold for noncognitives. Given cognitive value c_i, the full earnings gain from sorting is

$$\underbrace{c_i \cdot \mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)}_{\text{Full sorting gain}} = \underbrace{\bar{c} \cdot \mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)}_{\text{Firm returns effect}} + \underbrace{\widetilde{c}_i \cdot \mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)}_{\text{Match effect}}, \tag{3.15}$$

where c_i is split into average $\bar{c}$ and deviation $\widetilde{c}_i$. Since the distribution of returns faced by each individual depends on their skill level, the expected return from firm heterogeneity changes non-linearly with skills. Estimates based on the clustering approach (column 1 of Table 3.5) illustrate that the marginal expected return $\mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)$ is increasing in c_i and thus deviates from the unconditional average, which is normalized to zero. The difference in expected marginal returns between top and bottom cognitive skill workers is almost 12 log points $(6.74-(-5.13)=11.87)$.

Marginal returns conditional on skills. The empirical distribution of gains is summarized in column (2) of Table 3.5. Top cognitive workers vastly benefit from higher conditional returns, which lead to earnings 7 log points higher than if they were matched with the average firm. To illustrate how much this return matters for skill premia, it is useful to consider a simple example where we compare the sorting gains gap between a top worker $(c_i = 1)$ and a low-middle (level 4 in Table 3.5, $c_i = 0.375$), which is 8 log points. The raw earnings difference between these two workers is on average 30 log points in our sample; this gap is reduced to $(30-8) = 22$ log points when sorting effects are taken out. Thus, sorting adds more than 1/3 ($\frac{8}{22}$) to the baseline gap and significantly amplifies between-skill earning differences.

Non-monotonicity of gains. Column (2) of Table 3.5 shows that gains are not monotonic in c_i. In particular, workers with low-to-middle skills lose out compared to a hypothetical situation where everyone is matched with the average return. To understand these losses, and why they wane as c_i approaches zero, equation (3.15) breaks down skill returns into a "return effect" $\bar{c} \cdot \mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)$ and a "match effect" $\widetilde{c}_i \cdot \mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)$.

Estimates of the return effect $\bar{c} \cdot \mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)$ are shown in column (3) of Table 3.5 and reflect the gain that a worker i, whose skill endowment is equal to the cross-sectional average, derives from being assigned to different expected returns. Hence the return effects measure the impact of firm heterogeneity net of any complementarity gains. Since high skill workers sort into high return firms, and low skill workers into low return firms, estimates of the return effects grow monotonically with skills. This raises inequality compared to a random allocation and results in a zero-sum redistribution of returns, as evidenced by the aggregate nil effect reported in the bottom row of

Table 3.5: Gains from sorting across returns λ_j^c for different cognitive skill levels.

	$\mathrm{E}(\lambda_j^c \mid c_i)$ (1)	Full gain (2)	Return effect (3)	Match effect (4)	$\mathrm{E}(\lambda_j^0 \mid c_i)$ (5)
skill level (c_i):					
1 (lowest, $c_i = 0$)	*-5.13*	0.00	-2.75	2.75	-2.00
2	*-4.61*	-0.58	-2.47	1.89	-1.51
3	*-3.75*	-0.94	-2.01	1.07	-1.45
4	*-2.61*	-0.98	-1.40	0.42	-1.28
5 (median, $c_i = 0.5$)	*-0.85*	-0.42	-0.45	0.03	-0.69
6	*1.10*	0.69	0.59	0.10	0.15
7	*2.98*	2.24	1.60	0.64	1.33
8	*4.86*	4.25	2.60	1.65	2.70
9 (highest, $c_i = 1$)	*6.74*	6.74	3.61	3.13	3.83
Aggregate	*0.00*	0.75	0.00	0.75	0.00

Notes: Gains are multiplied by 100 (i.e., in log points) for readability. All returns are differences relative to a scenario with no heterogeneity in firm returns. Estimates are based on the grouping approach. Sample period: 1999–2008. Column (1): expected marginal return conditional on skill. Column (2): total gain from sorting. Column (3): gain from sorting for the average-skill worker. Column (4): gain from sorting in excess of an average-skill worker with the same employer. Column (5): gain from sorting into intercepts.

column (3) in Table 3.5.

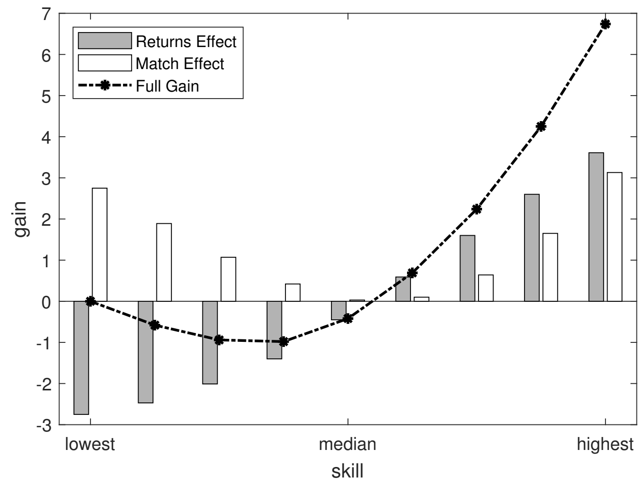
In contrast, the match effects $\widetilde{c}_i \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ in column (4) raise earnings in the aggregate by eliciting incremental gains from worker-firm complementarity.[20] Unsurprisingly, match effects are large at the higher end of the skill distribution, where earnings are magnified compared to a random allocation (3.1 log points match effect for $c_i = 1$; 1.65 for $c_i = 0.875$).

Large match effects are also detected among low skill workers (e.g., 2.7 for $c_i = 0$; 1.9 for $c_i = 0.125$) since, as shown in (3.14), match effects are defined as deviations from the average-worker gains. That is, match quality effects measure returns in excess of those experienced by an average skill worker with the same employer. The result then follows from the observation that average skill workers experience a steeper loss from being matched with a low quality firm due to the higher opportunity cost of the mismatch.

Firm-specific intercepts and gains from sorting. The last column of Table 3.5 shows the wage gains from matching with alternative intercepts λ_j^0, conditional on skill types c_i. These gains have a zero sum due to the lack of complementarity between skills and firm intercepts. Nonetheless, the

[20] Both components are defined as surplus relative to the baseline in which firm heterogeneity is absent and all returns are equal to the population average. Hence, both positive and negative gains must be interpreted relative to a scenario where each worker is given the average return or, equivalently, where workers are randomly allocated across firms.

Figure 3.4: Gains from sorting across returns λ_j^c for different cognitive skill levels.



Notes: Gains are multiplied by 100 (i.e., in log points) for readability. All returns are differences relative to a scenario with no heterogeneity in firm returns. Estimates are based on the grouping approach with detailed numbers in Table 3.5. Sample period: 1999–2008.

differential assignment of workers across firms (and, hence, across λ_j^0) raises earning differences by an extent comparable to that due to sorting on returns in column (2). This reinforces overall inequality between skill levels as more able workers also tend to populate higher intercept firms. The purely redistributive nature of this effect induces, however, little or no additional convexification in the ability-wage space.[21]

A graphical representation. Figure 3.4 offers a concise summary of the distribution of sorting gains and their components. Workers with the lowest skills exhibit positive match quality effects because they do not lose like the average worker from being matched to a low return employer. By the same token, the gains turn negative for low to intermediate skill workers, who would benefit from matching with high return firms but are not assigned to such firms. These individuals would be better off in a world with no firm heterogeneity in skill returns. For workers with above average skills, both the return effects and the match effects are positive, which results in large gains at the top. Complementarities are key to deliver an earnings schedule that is convex in skills.

Estimated gains outweigh losses and matching raises aggregate earnings. A simple way to assess the intensity of matching in the data is to benchmark it against the maximum gain it could

[21] Convexity of earnings is only due to skill complementarities. Further evidence of this point comes from the observation that, after conditioning on noncognitive skills n_i, the sorting across λ_j^0 results in modest effects that offset rather than reinforce the skewness of earnings across noncognitive endowments (see Appendix Figure A.4).

Table 3.6: Moments due to skill returns under random versus actual sorting.

	Mean ×100		Standard deviation		Skewness	
	Random	Actual	Random	Actual	Random	Actual
	(1)	(2)	(3)	(4)	(5)	(6)
$\lambda_j^c c_i$	0.00	0.75	0.05	0.05	0.52	0.90
$\lambda_j^n n_i$	0.00	0.13	0.03	0.03	0.34	0.68
$\lambda_j^c c_i + \lambda_j^n n_i$	0.00	0.88	0.05	0.06	0.28	0.55

Notes: Central moments of distribution of skill returns assuming either the actual allocation or a counterfactual where workers are randomly assigned to firms. Mean earnings $\mu \equiv \mathrm{E}(\lambda_j^c c_i + \lambda_j^n n_i)$ and sub-components rise due to matching in column (2) compared to (1). Dispersion $\sigma \equiv \mathrm{sd}(\lambda_j^c c_i + \lambda_j^n n_i)$ rises modestly in column (4) compared to random assignment (3). Skewness $\widetilde{\mu}_3 \equiv \mathrm{E}[(\lambda_j^c c_i + \lambda_j^n n_i - \mu)/\sigma]^3$ in the actual is almost twice as large relative to random assignment (last two columns). Estimates based on the grouping approach. Sample period: 1999–2008.

generate, given the estimated return and skill dispersion. Adopting this metric, assortative matching in the cognitive dimension generates 0.75 log points (Panel B, Table 3.4) as compared to a hypothetical maximum of 1.9 log points.[22] This simple calculation lends support to the view that the observed allocation of skills across employers, while imperfect, does deliver some of the gains associated to efficient matching. Gains from matching along noncognitive returns are smaller, yet they boost the aggregate match quality gain to 0.88 log points. All the estimates of the gains from sorting are robust to alternative normalizations of skills and returns (see Appendix A.4.4).

Random assignment of returns. It is informative to compare the distribution of estimated skill returns to the one obtained under random assignment of workers to firms. We construct the random assignment counterfactual by sample-weighting all skill types within a firm according to their population share and we are careful to preserve the empirical firm size distribution. Table 3.6 illustrates the findings from this exercise by juxtaposing the first three moments of the empirical distribution to those obtained under random assignment.

Columns (1) and (2) report first moments. This reproduces the aggregate gains reported before, i.e., average log earnings effects are the same when randomly allocating workers or assigning them to the average firm. Columns (3) and (4) show that the standard deviations of skills premia are only marginally different: this is not surprising if one considers that higher between-skill inequality in the non-random allocation, seen in Figure 3.4, is offset by declines in within-skill inequality due to the similarity of worker skills within firms. The muted changes in the second moment of

[22] Match effects are maximized when the correlation $\mathrm{corr}(\lambda_j^c, c_i) = \frac{\mathrm{cov}(\lambda_j^c, c_i)}{\mathrm{sd}(\lambda_j^c)\mathrm{sd}(c_i)} = 1$. Our grouped estimates imply an upper bound for match effects in the cognitive dimension of $\mathrm{sd}(\lambda_j^c) \times \mathrm{sd}(c_i) = 0.08 \times 0.24 = 0.019$.

the distribution point to an important and subtle distinction highlighted in the theoretical literature (Becker and Chiswick, 1966, Lindenlaub, 2017, Sattinger, 1993), which emphasizes how the most conspicuous changes induced by production complementarities may occur in the third moment of the earnings distribution. Columns (5) and (6) in Table 3.6 suggest this is indeed the case in our worker-firm sample, where the skewness of log earnings is twice as large under the non-random assignment of workers to firms.

More generally, high skill workers are not often observed in low return firms while differences in skill returns have little effect on low endowment workers. The latter observation translates into a fairly concentrated left tail of the earnings distribution when compared to random assignment. In contrast, the nonlinear effects from matching high skill individuals to high return firms result in a substantial thickening of the right tail of the earnings distribution, as shown in Figure 3.4.

To sum up, heterogeneity in skill returns provides a natural way to interpret the asymmetries in the distribution of earnings and reconcile models of sorting with the well-established evidence on between-firm variation.[23] Since the distribution of firm sizes is unchanged in our counterfactuals, sorting has no effect on the moments of firm intercepts λ_j^0, which are the same under the actual and random allocations.

3.6 Extensions and Robustness

Firm heterogeneity in skill returns encourages sorting and affects the earnings distribution. One may, however, question to what extent the assignment of workers to jobs occurs along the industry and occupation dimensions. This motivates a robustness exercise where we explicitly test for return heterogeneity within narrowly defined industry and occupation groups.

In addition, and to aid interpretation of our baseline findings, we examine the correlation of skill returns with a subset of firm-level measurements. This is facilitated by external data about firms' balance sheets, capital composition and innovation activities that can be linked to our sample of employers. The latter measures convey information about the nature of production arrangements that may underpin firm differences in skill returns.[24]

Finally, we examine the robustness of estimates under the clustering approach to alternative choices about the number of firm classes and of variables used for grouping firms.

[23] Bonhomme et al. (2019) present an analogous counterfactual where workers are randomly allocated to firms. Our estimated gains from skill complementarities are of similar magnitude when compared to their match effects between unobserved worker and firm types. We find larger effects in the aggregate (almost 1% of earnings vs 0.5%). Part of the difference is accounted for by the more pronounced earnings convexification in our estimates, which disproportionately benefits workers with higher skill endowments.

[24] We focus on clustered firm returns (100 k-means groups) for brevity. Results for the leave-out firm-level samples are consistent with what we emphasize in this section (Appendix A.5).

3.6.1 Industry and Occupation Specific Skill Returns

We begin by assessing whether skill returns simply reflect sector and job characteristics. We do so by adding to the baseline specification (3.10) a full set of industry × occupation interactions with cognitive and noncognitive skills. Detailed estimates are reported in Appendix A.5.1. We find that fine industry and occupation-specific skill returns, or returns that vary by industry × occupation group, account for a minor share of firm-specific heterogeneity. Sorting of both cognitive and noncognitive skills across returns remains strong. Results confirm that significant skill returns heterogeneity occurs at the firm level (as opposed to the more aggregate industry or occupation level). This remains true after conditioning on rather fine occupation measures.

Aggregating to industry or occupation. While most of the heterogeneity occurs at the firm level, some industries or occupations may still exhibit higher skill returns on average. In Appendix A.5.1 we explore this possibility by projecting the baseline λ_j^c and λ_j^n estimates on a broad set of industry-sector indicators and employment shares by occupation group. These projections are similar to those used to test the PAM hypothesis through the equations in (3.12) and one can show that they deliver generally unbiased point estimates. We find that high cognitive returns are frequent in the business services and IT sector as well as in firms with a large share of professional occupations. Noncognitive returns tend to be higher in the personal services sector and in firms that have large shares of managerial, technical and services/sales jobs. These results hold in the firm-level leave-out samples and in the clustered samples.

3.6.2 Capital Composition, Innovation, and Skill Returns

Next, we link balance sheet and innovation data to the sample of employers. This lends some insight into potential sources of skill return heterogeneity.

Balance sheets and capital composition. Since differences in capital composition reflect systematic aspects of productive and organizational structure, we use balance sheet data to measure tangible and intangible capital per employee (as well as finer components). An advantage of the Swedish institutional setting is that a majority of private sector firms are limited liability corporations with publicly available financial statements. We thus aggregate the workplaces at the organization level where this information is reported. We refer to these aggregates as "firms" from now.

Table 3.7 reports estimates for projections of skill returns onto firms' tangible and intangible capital components. We focus on cognitive returns from the group-level estimates. Results for noncognitives are in Appendix A.5.2. To account for zero-value observations for finer capital items in the balance sheets, we use the inverse hyperbolic sine (arcsinh) transform.[25]

[25] The arcsinh approximates $\log(2x_j) = \log(2) + \log(x_j)$. Estimates are interpreted as semi-elasticities (unit changes)

Table 3.7: Projection of Group Returns onto Firm Capital Composition.

	Dependent variable: $\lambda_j^c \times 100$					
	(1)	(2)	(3)	(4)	(5)	(6)
Tangible assets	-0.83	-0.30	-0.53			
	(0.12)	(0.06)	(0.15)			
Buildings, Land, Machinery				-0.92	-0.35	-0.59
				(0.13)	(0.07)	(0.15)
Other tangible assets				0.13	0.05	0.04
				(0.07)	(0.05)	(0.14)
Intangible assets	1.02	0.63	0.85			
	(0.11)	(0.06)	(0.11)			
Patents, licences, capt. R&D				1.17	0.70	0.72
				(0.12)	(0.07)	(0.14)
Goodwill and other intangibles				0.52	0.36	0.52
				(0.09)	(0.06)	(0.09)
R-squared	0.10	0.33	0.08	0.11	0.33	0.09
Number of firms	14,339	14,339	14,339	14,339	14,339	14,339
Sector fixed effects	No	Yes	No	No	Yes	No
Employment weighted	No	No	Yes	No	No	Yes

Notes: Projections of cognitive skill returns onto capital components per employee, using firms' balance sheets. Tangible fixed assets comprise of buildings and land; machinery and equipment; and other. Intangible fixed assets include capitalized expenditure on research and development; patents, licenses, and concessions; goodwill; and other. All variables are transformed using inverse hyperbolic sine, i.e., $\text{arcsinh}(x_j) = \log\left(x_j + \sqrt{x_j^2 + 1}\right)$. The dependent variable λ_j^c is multiplied by 100. Estimates are based on the sample of clustered firms; period 1999–2008. Robust standard errors clustered at the level of each of the 100 firm groups.

Capital composition is strongly associated to cognitive returns. Column (1) of Table 3.7 shows that tangible assets vary negatively with skill returns but intangible assets exhibit a strong positive correlation. Column (4) illustrates that the negative relationship holds strong for physical capital (buildings, land, and machinery) and the positive relationship is especially intense for intellectual capital (patents, licences, and capitalized *R&D* expenses). The notion that intangible capital and intellectual property are complementary to high skilled labor within a firm is consistent with production arrangements that leverage innovation. Relatively high physical assets and machinery, on the other hand, are more frequent in firms that exhibit lower returns to cognitive skills.

These relationships are robust in several respects: they hold within industry sectors of the economy (columns (2) and (4) of Table 3.7) and if we weight with firm employment size (columns (3)

for very small values of the transformed variable x_j, and as elasticities for larger values. See Bellemare and Wichman (2020) and note to Table 3.7. Findings are robust to alternative approaches; Appendix A.5.2 shows that similar results hold at the intensive margin (log transform of capital items) and at the extensive margin (firms with high cognitive returns are more likely to report nonzero intangible assets).

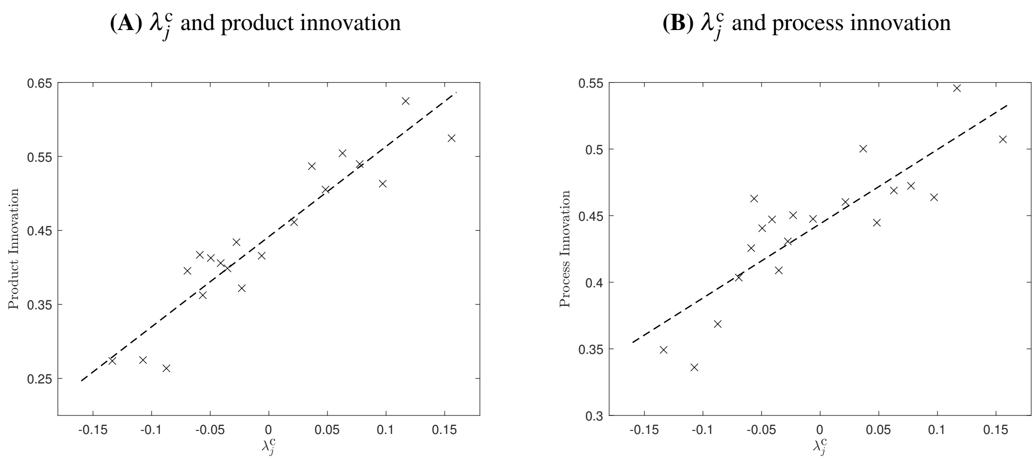


Figure 3.5: Cognitive skill returns and firm innovation.

Notes: The figure plots a binscatter of firms' innovation activities against cognitive skill returns (group-level estimates during 1999–2008). Innovation activities are measured as indicators whether a firm has conducted any product (including service, Panel a) or process (including organizational, Panel b) innovations. This information is from various waves of a representative firm survey (European Community Innovation Survey, CIS). We average the responses (i.e., indicators) for the waves 1998–2000, 2002–2004, 2004–2006, 2006–2008, 2008–2010 relevant to our estimation period. Underlying the plots are 4,138 unique firms. Regression slopes, controlling for a quadratic in firm employment, are $\beta = 1.21$ (clustered S.E. $= 0.13$) and $\beta = 0.55$ (clustered S.E. $= 0.10$) for product and process innovations, respectively.

and (6)). Appendix A.5.2 shows that they hold in the leave-out firm-level samples as well as when using dummy indicators (or logs) instead of the arcsinh transformation. Perhaps unsurprisingly if one considers production arrangements, firms that employ intangible and intellectual assets have substantially higher cognitive skill returns.[26] As we show in Appendix Table A.8, results for noncognitive skills are less pronounced and returns are modestly higher in firms with more physical capital. This lends support to the notion that skills should be modeled separately rather than collapsed into a single index.

Measures of innovation activities. Next, we use responses in the Swedish version of the European Community Innovation Survey (CIS) to study the relationship between skill returns and innovation activities. In each wave of the CIS, a representative sample between 2,000 and 5,000 firms reports whether they conducted any product (including new services) or process (including organizational structure) innovations in the survey year or the preceding two years. Lindner et al. (2021) show that the CIS provides direct, reliable, and broad measures for different types of firm-level technological change.

After linking the CIS survey responses to the administrative sample of employers, in Figure 3.5

[26] Even controlling for capital composition in equation (3.10), or allowing for interactions between capital and skill in parallel to occupation-specific skill returns in Section 3.6.1, has little impact on the heterogeneity of firm-specific skill returns that we uncover.

we plot bin scatters of dummies (taking value one in the presence of product/process innovations in the firm) versus cognitive skill returns.[27] Firm innovation activities are positively, and almost linearly, associated with estimates of cognitive returns. This is especially apparent in the case of product innovations where, moving from the lowest to the highest λ_j^c firms, the share of firms which introduce such innovations rises from about 25 to 65 percent. For process innovations the relationship is fainter and only borderline significant when we also condition on product innovation (Table A.10). However, innovation activities still differ by twenty percentage points between firms with the lowest and the highest skill returns.

In Appendix A.5.2 we show how results are qualitatively robust to alternative firm-level estimation approaches or when controlling for industry fixed effects. Moreover, innovation expenditures in the CIS survey are also larger for higher λ_j^c firms (especially in-house research and development), suggesting that high cognitive return firms differ in their ability to bring forward innovations. This evidence lends support to existing studies linking cognitive skills to worker level innovation activities (Aghion et al., 2017, Bell et al., 2018).

3.6.3 Changing the Cluster Design

When using estimators based on firm clusters, one question is whether results are sensitive to the grouping strategy. Next, we examine differences in the estimated contribution of firm heterogeneity to earnings dispersion under alternative assumptions about the number of firm classes and about the observables used to classify them.

Using only ten firm classes (Bonhomme et al., 2019, Lamadon et al., 2022) marginally lowers the absolute contribution of firm heterogeneity while raising the importance of skill returns relative to the intercepts. However, results remain similar to the baseline. After adding additional observables to the clustering criterion (namely, firm employment and the standard deviations of both earnings and skills within the firm), estimates of firm effects are comparable and in line with the benchmark.[28] If we discard information about worker skills and only use data about within firm earnings to define firm classes (see Bonhomme et al., 2019, Lamadon et al., 2022), we also find that estimates of skill returns' contribution do not change significantly relative to the case where many observables are used. Estimates from these exercises are in Appendix Table A.11.

Changing the number of clusters. The baseline grouping approach, with one hundred clusters, delivers conservative estimates of the contribution of firm heterogeneity to earnings dispersion. To

[27] We plot the raw relationship after controlling for (a quadratic in) employment, since the probability of engaging in any innovation rises with a firm's size. The controls do not substantively affect results. The corresponding relationships for noncognitives are weaker and reported in Appendix A.5.2.

[28] This check is performed using 100 firm classes. We also experiment with including value added, which is however reported at the organization level, and find similar results.

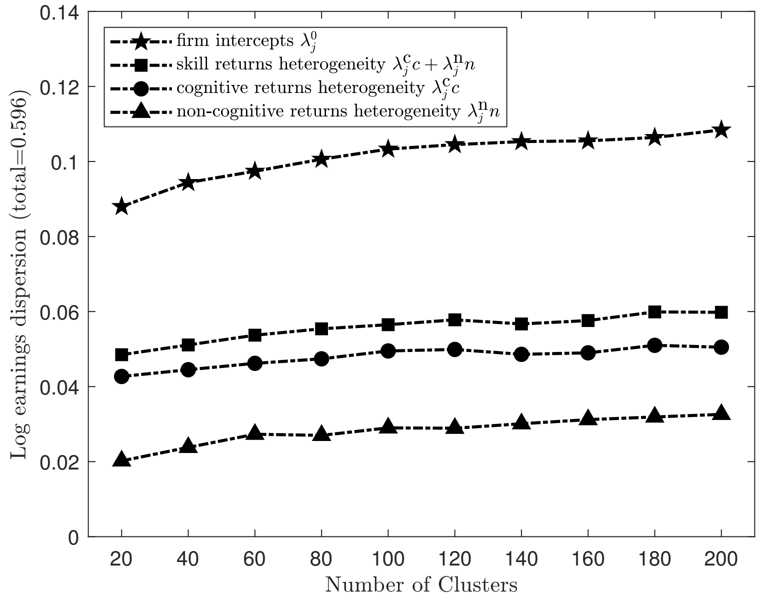


Figure 3.6: Dispersion due to firm heterogeneity (log earnings), by number of k-means groups.

Notes: The figure shows the earnings variation due to firm intercepts sd(λ^0), cognitive skill returns sd($\lambda_j^c c_i$), noncognitive skill returns sd($\lambda_j^n n_i$), and overall skill returns sd($\lambda_j^c c_i + \lambda_j^n n_i$) when we re-estimate the model with different numbers of k-means clusters. Estimation period: 1999–2008.

illustrate how restrictions on the number of clusters affect estimates of firm effects, Figure 3.6 shows the standard deviation of log earnings that is attributed to different layers of firm heterogeneity when we increase the number of clusters from 20 to 200. Under the assumption of only twenty firm clusters, the impact of skill return heterogeneity is substantial, with a contribution of 5 log points to dispersion as opposed to 9 log points due to firm intercepts. Increasing the number of clusters results in higher estimates of the impact of firm heterogeneity on overall inequality. As one might expect, the absolute values of firm effects estimated from finer clusters become similar to those obtained using the quadratic-form correction with no firm grouping. More notable, and perhaps less expected, is that the relative contribution of each layer of heterogeneity is stable throughout.

When no clustering is imposed and estimates are adjusted using the bias correction method, the absolute impact of firm heterogeneity on earnings is larger but the relative impact of different components (intercept vs skill returns) is effectively unchanged (see Table 3.4). This is consistent with findings in Table 3.1 and indicates that the two empirical approaches deliver comparable estimates of the relative contribution of skill return heterogeneity to firm-level variation.

3.7 Conclusion

By examining distinct dimensions of firm heterogeneity, we estimate the extent of skill return variation across employers and present direct evidence of worker–firm complementarities. The analysis relies on alternative empirical approaches in administrative employer–employee population records that we link to high-quality information about the cognitive and noncognitive attributes of workers. Each approach imposes different restrictions to mitigate biases and carry out the large computational exercises. Nonetheless, estimates of the relative impact of skill returns, as opposed to conventional measures of firm heterogeneity based on fixed effects, are stable irrespective of the implementation.

Our key findings can be summarized as follows: (1) Similar skills command substantially different returns across firms. These differences occur along both the cognitive and non-cognitive skill dimension, but firm-level returns to each attribute only weakly correlate with one another. (2) Returns heterogeneity generates incentives for sorting and we document that, indeed, workers with larger endowments of cognitive and noncognitive skills populate firms with higher returns to those attributes. The intensity of sorting in each skill dimension depends on the dispersion of that skill's return across firms; as dispersion grows, so does the incentive for skilled workers to seek a better match. (3) The gains from sorting across employers are unevenly distributed and non-monotonic in worker skills. High skill workers benefit from heterogeneity in returns while the least productive workers experience little loss from low-return employers. Considerable costs are borne by workers with intermediate skills who have a nontrivial opportunity cost of being matched with low quality firms.

More generally, we find evidence of economically meaningful complementarities between workers and firms, and of positive assortative matching in multiple skill dimensions. Sorting has material implications for earnings. In particular, the earnings distribution becomes more skewed due to the matching of high skills to high returns. By the same token, allocative efficiency improves and the average economy-wide skill premium rises.

A central lesson from studies of employer effects over the past two decades is that firm-level productivity and rent-sharing are key determinants of wage growth and inequality (see Card et al., 2018, Lamadon et al., 2022, for a summary). This has spawned interest in the role of market factors, institutions, and policies that may underpin such dependence (e.g., De Loecker and Eeckhout, 2021, Dustmann et al., 2022, Jäger et al., 2021). One implication of our findings is that labor market returns to skill attributes are nuanced and not solely driven by the supply of skills or by aggregate changes in demand (e.g., general skill biases in technical progress). Rather, the composition and evolving demands of firms play an important mediating role for the economy-wide returns to skills.

Using information from firms' balance sheets, we present ancillary evidence that firms with high cognitive returns engage in more innovation and hold more intellectual capital. This points to the presence of underlying complementarities in production and raises the possibility that mismatch between skills and firms may hamper the gains from innovation and productivity advances (Aghion et al., 2017, Bell et al., 2018). Evidence about the covariation of skill returns with intellectual capital and innovation suggests that a fruitful direction for further research may be to explicitly examine the nature and determinants of the extensive heterogeneity in skill returns.

Chapter 4

Universal Basic Income and the City

This chapter is a joint work with Jack Favilukis[a] and Stijn Van Nieuwerburgh.[b]

[a]Sauder School of Business, University of British Columbia (jack.favilukis@sauder.ubc.ca).
[b]Graduate School of Business, Columbia University (svnieuwe@gsb.columbia.edu).

4.1 Introduction

The increase in job automation and the potentially related rise in income inequality has led to calls for Universal Basic Income (UBI), or a fixed payment to every legal resident. Prominent figures, including presidential candidate Andrew Yang and entrepreneur Elon Musk, have endorsed UBI. Since UBI is largely untested, and would drastically change existing U.S. fiscal policy, it would likely lead to many consequences, both intended and not. This paper studies some of these consequences, with a particular focus on real estate and the urban environment.

To study this question, we start with a model identical to Favilukis et al. (2018) and calibrated to New York City. We extend the model by allowing for UBI – a guaranteed, fixed transfer of cash ($5,000 per year in our baseline model) to every household. UBI is financed by income taxes which are, in their nature, distortionary. Distortionary income taxes typically reduce the incentive to work, resulting in households choosing more leisure hours and fewer working hours. This leads to lower city-wide output, income, and consumption. At the same time, UBI is a net transfer from wealthier to poorer households. Since poorer households have higher marginal utility, this is a transfer to those who need money the most.

Under an equally weighted social utility measure, UBI is welfare improving in our baseline model. This is because, despite creating large distortions through the taxes needed to finance UBI, it provides a large benefit to those who need it most. Of course, this result crucially depends on a society's desire for equality and redistribution (proxied by the risk aversion parameter in our model), and on the elasticity of labor supply, which measures how strongly labor supply responds to distortionary income taxes. As we reduce desire for equality (risk aversion), the benefit of UBI is weakened while the costs remain unchanged, leading to smaller welfare gains, and even a welfare loss if risk aversion is sufficiently low. Similarly, as we make labor supply more elastic, the distortionary cost of UBI rises while the benefit is unchanged, also leading to smaller welfare gains.

The most important results for the housing market are that the lower income for those at the top leads to lower rents and prices per square foot, and to a lower home ownership rate. The housing expenditure to income ratio falls as well, suggesting that affordability improves as the real cost of housing falls. However, despite the lower cost of housing, less housing is built and used, since the city is, on average, poorer. This alleviates concerns that UBI would lead to inflation of house prices and rents.

The way UBI is financed is crucial for both the welfare effects and the effects on the city. This is because different financing methods lead to different distortions. We explore two more progressive methods of financing UBI (a more progressive income tax and a wealth tax). Perhaps surprisingly, we find that they lead to less positive welfare effects than the baseline model because too much of the tax burden falls on the high productivity households, distorting their labor and saving decisions. On the other hand, a corporate tax leads to a more positive welfare effect, this depends on the assumptions that corporate owners are outside of the city and do not change their behavior in response to the tax.

Finally, we study a case where UBI is funded exogenously, rather than by local income taxes. While this is not feasible globally, it is analogous to several recent experiments with UBI. For example, private donations funded a $500 per month UBI program for randomly selected lower income households in Stockton, CA, with similar programs starting in New Orleans, Gary, IN, and Durham, NC. We find that this leads to relatively small increases in rents and house prices across the city.

There are also interesting effects by neighborhood, but they depend on the progressivity of the tax system. The central core can become more or less dense, more or less wealthy, or younger or older relative to the rest of the city. The more progressive the tax changes used to pay for UBI, the higher the likelihood that higher income people leave the city center, making the core less wealthy and more dense as lower income people consume less housing per person.

It is important to note our model's limitations. First, our model is analogous to a small open economy where the interest rate is exogenous – this is reasonable when modeling a single city but is unlikely to be true if UBI is implemented nationally. Second, the population in our model is fixed and we do not allow in or out migration, other than to the suburbs. In the real world, UBI implemented in a single city may lead to in migration from other cities. Third, the model's assumptions about preference for equality (risk aversion) and the effect of distortions on labor supply (Frisch elasticity) are crucial. We discuss these in Section 4.3.4.

4.2 Model

Our model is very similar to Favilukis et al. (2018). Here we explain the model's main features. The model is described in full detail in Favilukis et al. (2018).

We model a metropolitan area that consists of two zones, the central business district (zone 1) and the rest of the metropolitan area (zone 2). Working-age households who live in zone 2 commute to zone 1 for work. Commuting entails both an opportunity cost of time and a financial cost. Finally, zones provide different levels of amenities. Zones have different sizes, captured by limits on the maximum amount of housing that can be built. Building becomes especially expensive as the city's housing stock gets close to the maximum limit.

The city is populated by overlapping generations of risk averse households who face idiosyncratic labor productivity risk and mortality risk. They make dynamic decisions on location, non-housing and housing consumption, labor supply, tenure status (own or rent), savings in bonds, primary housing, investment property, and mortgage debt. The rental stock is owned by local households, who rent it to other locals.

Since households cannot perfectly hedge labor income and longevity risk, markets are incomplete. This incompleteness opens up the possibility for redistribution policies to provide insurance. Progressive tax-and-transfer and social security systems capture important existing insurance mechanisms, with UBI going above and beyond these mechanisms. The model generates a rich cross-sectional distribution over age, labor income, tenure status, housing wealth, and financial wealth. This richness is paramount to understanding both the distributional and aggregate implications of housing affordability policies.

On the firm side, the city produces tradable goods and residential housing in each zone, subject to decreasing returns to scale. As a zone approaches its maximum buildable housing limit, construction becomes increasingly expensive, and the housing supply elasticity falls. Wages, house prices, and market rents are determined in the city's equilibrium, to clear the labor market, the housing supply (construction) and demand market, and the rental market, respectively. The interest rate is exogenous and comes from outside of the city. By Walras's law, because the interest is exogenous, the consumption market does not clear. In other words, the city's net bond demand may be positive or negative, and the city may consume more or less than it produces, the difference financed by the interest on the net bond position. In this sense, the city we model is analogous to a small, open economy in the international economics literature.

We extend the Favilukis et al. (2018) model by allowing for UBI. In particular, every household receives a fixed amount D per period, with the total amount of UBI payments funded by an increase in taxes. Following Heathcote et al. (2017), the income tax in our model is captured by two parameters, λ, which is related to the level of taxes, and τ, which is related to tax progressivity. We experiment with different combinations of λ and τ to fund the extra spending on UBI. We also experiment with alternative ways to fund UBI, such as a wealth tax or a corporate tax, and we describe those below.

4.3 Results

4.3.1 Baseline model fit

We calibrate the model to the New York metropolitan area, designating Manhattan as the urban core, or zone 1, and the rest of the metropolitan area (MSA) as zone 2. Our calibration targets key features of the data, including the relative size of Manhattan versus the rest of the MSA, the income distribution in the New York MSA, observed commuting times and costs, the housing supply elasticity, current zoning laws, the current size and scope of the affordable housing system, and the current federal, state, and local tax-and-transfer system. The baseline model generates realistic income, wealth, and home ownership patterns over the life-cycle for various percentiles of the income distribution. It matches both income and wealth inequality, as well as house price and rent levels for the MSA. The model generates a large wedge between the prices and rents in the two zones. The details of the calibration are described in Favilukis et al. (2018).

Figure 4.2 shows household income, wealth accumulation, and home ownership over the life cycle. Households in the model look very much like the data for these quantities. When we break these quantities into low, middle, and high income households, the quantities also look like the data. Thus, the model is able to quantitatively capture a household's behavior throughout its life cycle, as well the high degree of inequality we see in the data. In particular, as in the data, all households begin with low income and little wealth. Households' income rises and peaks around age 50. Households accumulate wealth, with wealth peaking around age 65, just as households retire, at which point they deaccumulate wealth. At the peak, high income (top 25%) households have roughly 2.5 times as much wealth as the average household. Households also start off renting, but shift towards ownership through their 30s and 40s, reaching a peak ownership rate around 80% in their late middle ages. The ownership rate of low income (bottom 25%) of households is significantly below that of the average household.

Figure 4.1 shows the distribution of house sizes. The model (left panel) matches the data (right panel) quite well, even though these moments are not targeted by the calibration. The size distribution of owner-occupied housing is shifted to the right from the size distribution of renter-occupied housing units in both model and data.

Table 4.1 compares various real estate related statistics for the model and for New York City. Here too, the model fits the data well. The model matches the average price-to-rent ratio of 17.8

by construction. Households spend roughly 23% of their income on housing, although this ratio is much higher for renters. A significant fraction is rent burdened, that is, they spend more than 30% of their income on rent. 11% of the population lives in Zone 1 (Manhattan), these households tend to be younger, and have much higher incomes. Zone 1 is significantly denser than Zone 2, with smaller unit sizes, higher rents, higher prices, and a lower ownership rate.

4.3.2 UBI

We focus on a UBI of $5,000 per year given to every household. This quantity is financed by income taxes. We follow Heathcote et al. (2017) and choose an income tax schedule that captures the observed progressivity of the U.S. tax code in a parsimonious way. Net taxes are given by the function $T(\cdot)$: $T(y^{tot}) = y^{tot} - \lambda(y^{tot})^{1-\tau}$. The parameter τ governs the progressivity of the tax and transfer system, in the baseline model, we set $\tau = 0.17$ to match the average income-weighted marginal tax rate of 34% for the U.S. The parameter λ governs the level of the tax and transfer system, in the baseline model, we set $\lambda = 0.75$ to match state and local government spending to aggregate income in the NY metro area, equal to 15-20%.

To finance UBI in our baseline case, while keeping τ fixed, we adjust λ such that total government spending is exactly equal to total government spending in the no-UBI model, plus the cost of UBI. This requires $\lambda = 0.7080$. Figure 4.3 plots the marginal tax rate, as a function of household income, for the no-UBI model, the UBI model where taxes are raised (approximately) evenly across the board to pay for UBI (λ financing), and a model with more progressive taxation, discussed below. The tax increase required to pay for a $5,000 per household UBI is significant, with a family earning $180,000 seeing its tax rate rise from 33% to 43%.

While our focus is on the real estate effects of UBI, it is useful to first consider its effects in general. Almost any welfare analysis of UBI will produce two key forces which will work against each other: a redistributionary force which raises average welfare, and a distortionary incentive force which lowers average welfare.

First, UBI leads to redistribution of after-transfer income towards low income households. While all households receive $5,000, in dollar terms, low income households see only a small increase in taxes, while high income households see large increases. If low income households also have the highest marginal utility, as is the case in many economic models, then they will also see large gains in welfare. High income households, with lower marginal utility, will see smaller losses in welfare. In simpler terms, low income households need an extra $5,000 much more than high income households do. Thus, this first force typically leads to welfare gains if welfare is computed using an equal weighted social welfare function (as it is computed here), and even larger gains under a Rawlsian social welfare function. Of course, UBI is not Pareto optimal because high income households lose.

Second, UBI must be paid for by taxation, and most real world taxes are distortionary. In our model, UBI is paid for by distortionary taxes on total income. This distorts the labor supply decision, causing households to work fewer hours than they otherwise would, leading to lower aggregate income output, income, and consumption in the city. Since total income includes capital income, this also distorts the saving decision. Households save a lower fraction of their income, leading to lower aggregate investment and wealth. Thus, this second force typically leads to welfare losses.

Whether the net effect on welfare is positive or negative depends on the relative strengths of these two forces. For example, if society puts a high value on equality, then the first channel would be quantitatively more important. In our model, preference for equality is controlled by the risk aversion coefficient. In our baseline calibration, the risk aversion is 5, somewhat higher than usually used in economics, but lower than many estimates from financial markets. On the other hand, if labor supply (and more generally effort and output) are very elastic, then the second channel would be quantitatively more important. The Frisch elasticity of labor supply in our model is around 1.1, which is above estimates from micro data, but below estimates from macro data.[1]

Table 4.2 presents the welfare changes caused by UBI. We present the overall welfare change, as well as the change by age, wealth, and income group. In our baseline model, 38% of the population is better off from UBI, while 62% is worse off. Those better off are the poorest and least productive households; they also tend to be either young or old, as opposed to middle-aged. Because the marginal utility of those who are better off is very high, the average welfare effect is positive. In consumption equivalent units, the average household is 1.09% better off in the model where the additional taxes are evenly distributed.

To highlight the importance of preference for equality and labor supply elasticity, Table 4.2 also presents results for models with a lower preference for equality ($\gamma = 4$ and $\gamma = 2$), and a model with a more elastic labor supply (Frisch Elasticity of 2.0). The patterns by age, wealth, and income are qualitatively similar in all models. However, as preference for equality is reduced, the average welfare effect falls and becomes slightly negative when $\gamma = 2$. Note that as γ falls, there is little change in the welfare cost to high productivity households because the distortionary effects are not changing. However, the welfare benefit to low productivity households is reduced, leading to a lower average welfare effect. On the other hand, when labor becomes more elastic, all income groups see a shift down in their welfare change, since distortionary taxes affect everyone.

Although average utility rises in the baseline model, the distortions from higher taxes are quite strong. As a result, households consume too much leisure and too few goods relative to what they

[1] There is disagreement among economists about this number, see Keane and Rogerson (2012). Estimates from microeconomic data suggest these elasticities are small, between 0 and 0.5. On the other hand, estimates from macroeconomic data suggest that they are large, between 1 and 2. These differences may be due to adjustments at the intensive versus the extensive margins, or to human capital accumulation.

would prefer. As shown in Table 4.4, hours fall by 4.1% in the baseline model; effective hours (hours weighted by productivity) falls by far less, 1.7%, suggesting that the fall in hours is greater among the unproductive households, which is, to a degree, efficient. This happens because many low productivity households do not feel the need to work much after receiving the $5,000 UBI. This fall in hours is significantly larger in the model with more elastic labor supply. The fall in labor supply results in higher per-hour wages (though not higher total incomes) as businesses face labor shortages, resulting in a 1.5% fall in average income.

Since the tax is on total income, not just on labor income, higher taxes also disincentivize saving. Wealth falls by 15%, significantly more than income, leading to a 14% fall in the wealth-to-income ratio. As we explain in the discussion, part of this large fall can be attributed to the open economy style model, which assumes that the interest rate is set exogenously to the city. This is a reasonable assumption when focusing on a single city, as we do here, but not when studying the full general equilibrium effects of UBI, as one would want to do if implementing it at a national level. Wealth falls for two reasons: for low income households a safe universal basic income stream makes precautionary saving far less, while for high income households, holding wealth becomes less attractive because of higher taxes on capital income.

Of course, lower income and lower wealth result in lower consumption. Non-durable consumption falls by 5.0%. Housing consumption (dwelling size) falls by 3.3%. Lower demand for housing leads to lower rents, which fall by 3.0%, and lower house prices, which fall by 3.1%. The ratio of housing expenditures to income also falls, implying that housing becomes more affordable. This is because of the non-linearity in the cost of housing construction. Since the aggregate housing demand is lower, the city no longer needs to flatten mountains, fill in lakes, or build extra tall skyscrapers, reducing the average cost of construction. Despite more affordable housing, there is a sharp 16.2% drop in ownership. Thus, within this model, concerns that UBI would lead to soaring house prices are misplaced. Distortionary taxation makes the city poorer overall, leading to lower demand for housing and lower housing costs.

Income and consumption inequality fall. The Gini coefficient for after-transfer income falls from 0.475 in the baseline model to 0.460; the Gini coefficient for consumption falls from 0.387 to 0.369. Wealth inequality actually rises with UBI because low income households expect a safe source of income and engage in less precautionary savings.

Interesting changes also occur in the spatial dimension. In the baseline model the population of the city center (Zone 1) falls while the population of the suburbs rises. This is because the financial costs of commuting are no longer as onerous for lower income households, who prefer to move out to the suburbs and consume larger dwellings where the price per square foot is much smaller. We conjecture that subsidizing commuting or building better public transport would have a similar effect. These households are replaced by relatively more productive and wealthier households. Due to this composition effect, the average dwelling size and ownership rate in the city center actually rise, while they fall elsewhere.

We also solve a model where the UBI is doubled to \$10,000. This model looks qualitatively similar, but with stronger effects than the baseline model. For example, hours, income, and house prices all fall by about twice as much. The welfare increase is more than double that of the baseline model.

4.3.3 Alternative ways of paying for UBI

In the baseline model, UBI is financed by increasing the income tax, with the increase being (approximately) even, that is, there is no change to the progressivity of the tax code. In this section we investigate alternative ways of financing UBI. The welfare results are presented in Table 4.3 and changes to other moments are in the bottom panel of Table 4.4.

In the first experiment, we pay for UBI by making the tax code more progressive by increasing τ to 0.22. We simultaneously adjust $\lambda = 0.6662$ such that, as before, total government spending is exactly equal to total government spending in the baseline model, plus the cost of UBI. In the second experiment, we raise the entirety of the cost of UBI through a wealth tax. In the third experiment, we raise half that of the required revenue through a wealth tax, and the other half through a corporate tax. In the fourth experiment, the UBI is financed by some exogenous organization and requires no increase in local taxes. While this case is clearly not feasible on a large scale (general equilibrium), we include it because it is analogous to several experiments carried out in the real world. In all cases, we keep all other parameters, including risk aversion, labor supply elasticity, and the size of the UBI exactly as in the baseline model. The only difference is the financing method.

The welfare change in the progressive model is positive but about half that in the baseline model. This is somewhat surprising, since UBI itself adds progressivity to government transfers, and UBI is welfare improving. The reason is that with a more progressive tax, the bulk of the extra expense falls on the most productive households, which significantly raises their marginal tax rates. The fall in hours is twice as large as in the baseline model, but the fall in effective (productivity adjusted) hours is more than three times as large because the high productivity households drastically cut hours. As a result, the welfare loss of the high productivity group is 12.46%, compared to 4.62% in the baseline model. House prices and rents fall by almost twice as much as in the baseline model.

The wealth tax experiment also has a positive welfare change, but smaller than the baseline model. There are fewer distortionary effects on hours worked, as a result hours fall by less than the baseline model, and effective hours actually rise. However, wealth accumulation falls by twice as much as in the baseline model, as households would prefer to consume too early over paying additional wealth taxes. Households in the top wealth decile see a welfare loss of 23.9%, compared to 8.2% in the baseline model. House prices and rents fall by about the same amount as in the baseline model.

The welfare increase is significantly larger in a model where a mix of wealth and corporate taxes funds UBI. However, caution must be taken here. First, the businesses in our model are owned by corporations outside of the city, thus, while additional corporate taxes may lead these businesses to hire less locally, leading to lower welfare – indeed, hours fall relative to the wealth only model – the reduction in welfare of the owners of these businesses is not part of our welfare calculation. Second, capital is fixed in our model and it is usually optimal to tax whichever factor cannot get out of the way. In the real world, businesses might invest less or relocate to another city.

In our final experiment, UBI is financed outside of the city, requiring no local tax increase. This is analogous to several UBI experiments implemented in the real world, such as the €560 monthly payment to some unemployed Finns and the $500 monthly payment to low income residents of Stockton, CA. Not surprisingly, since this imposes no costs on the locals, this case has the largest welfare benefits, about 7% in consumption equivalent units. The benefit to the lowest income workers is comparable in magnitude to the other cases because these workers did not see much of an increase in taxes. However, middle and even high income workers now see small benefits as well, since they do not experience any tax increases. In total, only 3% of the population is worse off, because unlike the other cases, UBI leads to a rent and house price increase of approximately 1%, which hurts some households. UBI also leads to a decrease in working hours and wealth accumulation. This increase is smaller than in the baseline case because there are no longer distortionary effects of tax increase. Rather, this happens because households are generally wealthier and choose to consume more leisure.

These different types of taxes also have different effects on the spatial dimension. For example, when the tax is targeted at the most wealthy (the progressive and wealth tax experiments) then high productivity and high wealth households tend to leave the city center for the suburbs. For example, with a progressive tax, the average productivity of those living in the city center falls by 9.4%, and the average income falls by 25.8%; with a wealth tax the average productivity falls by 5% but the average wealth by 15%. Despite this, the population of the city center actually grows because they are replaced by lower income and lower wealth households who consume less real estate per household, leading to more density. This is in contrast to the baseline model, where lower income households move to the suburbs, leading to city center density falling, and wealth and productivity rising. On the other hand, when the tax burden is spread more evenly (the corporate tax and exogenous funding experiments), the spatial effects are more similar to the baseline model.

4.3.4 Discussion

In this section we discuss what may be missing from the model and other loose ends.

In our model, interest rates are fixed and exogenous to the model. This is a natural assumption when considering an individual city. However, if UBI is instituted at a wider scale, then a model where interest rates are endogenous, and where there is a link between aggregate consumption and production is perhaps more appropriate. In such a model, the reduction in saving due to higher taxes would lead to higher interest rates as capital becomes scarce. This, in turn, would attenuate some of the reduction in saving and the effect on wealth.

In our model, the city's population is fixed. If a single U.S. city implements UBI, low income households may migrate in. If the city also finances this UBI through taxes, then whoever bears the burden of the tax – whether that be businesses, high income, or high wealth individuals – may move out. This may lead to an unsustainable spiral where more and more of the tax base moves out, replaced by needy individuals. Any real world UBI implementation at a city level would have to consider this. This may be less of an issue if UBI is implemented in a large country like the U.S. since both in- and out-migration are more difficult.

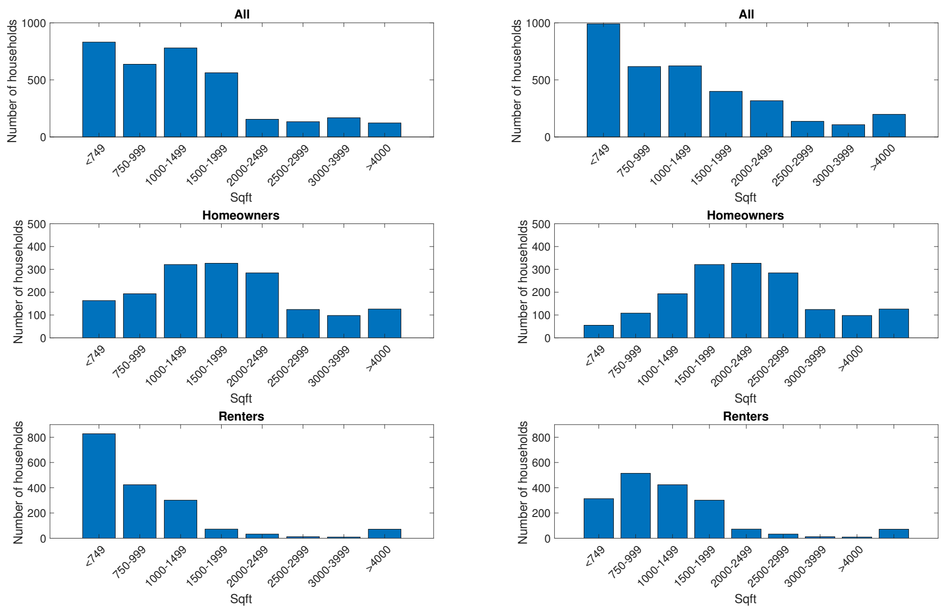
Above, we experimented with labor supply elasticity. Closely related to the labor supply elasticity is the issue of entrepreneurship and creation of new technologies. In our model, workers are endowed with productivity, and they choose hours, which converts productivity into output. While this is an adequate description of the real world production process for many workers, there are other choices that real world workers make. For example, they may choose effort or the risk-return trade-off of the project they work on. They may also choose to work on producing output today, as workers choose in our model, or producing knowledge, which may improve output tomorrow, which is outside of our model. The tax code likely affects these choices in different ways that cannot all be captured even if we match the Frisch labor supply elasticity.

Similarly, since our model has no entrepreneurship, our model also cannot speak to borrowing constraints that entrepreneurs face. These may be important for marginal utility. In our model, the marginal utility is very closely associated with income and wealth. That is, the poorest households are most in need of an extra \$5,000. However, consider for example, an entrepreneur who is relatively wealthy, but who has a great idea that requires significant financing. If this idea has a very high expected return on capital, then the marginal benefit of an extra \$5,000 for this entrepreneur may be higher than that of a low income household.

4.4 Conclusion

We study the effect of universal basic income (UBI), financed by distortionary income taxes, with a particular focus on real estate and the effects on the urban environment. For our chosen parameters, UBI is a net positive for local welfare using the average social welfare measure, although it negatively affects approximately 50% of the population. Because of distortions to the labor and investment decisions, the city is poorer on the net, which leads to lower rents and house prices. Additionally, if the tax used to finance UBI is very progressive, then wealthier households leave the city center for the outer suburbs.

Figure 4.1: House size distribution in Model (L) and Data (R)



Notes: Left panel: model. Right panel: data. Data source: American Housing Survey for the New York MSA, U.S. Census Bureau, 2015.

Table 4.1: New York Metro Data Targets and Model Fit

		Data		Model	
		metro	ratio zone 1/zone 2	metro	ratio zone 1/zone 2
1	Households (thousands)	7124.9	0.12	7124.9	0.12
2	Avg. hh age, cond. age > 20	47.6	0.95	47.4	0.87
3	People over 65 as % over 20	19.1	0.91	21.8	0.98
4	Avg. house size (sqft)	1445	0.59	1448	0.63
5	Avg. pre-tax lab income ($)	124091	1.66	124325	1.69
6	Home ownership rate (%)	51.5	0.42	57.4	0.79
7	Median mkt price per unit ($)	510051	3.11	506592	2.34
8	Median mkt price per sqft ($)	353	5.24	348	3.58
9	Median mkt rent per unit (monthly $)	2390	1.65	2432	1.82
10	Median mkt rent per sqft (monthly $)	1.65	2.78	1.67	2.78
11	Median mkt price/median mkt rent (annual)	17.79	1.89	17.36	1.29
12	Mkt price/avg. income (annual)	3.99	1.71	4.08	1.38
13	Avg. rent/avg. income (%)	23.0	1.00	23.5	1.07
14	Avg. rent/income ratio for renters (%)	42.1	0.81	28.8	0.91
15	Rent burdened (%)	53.9	0.79	45.5	0.49
16	% Rent regulated of all housing units	5.57	2.77	5.25	2.87

Notes: Columns 2-3 report the values for the data of the variables listed in the first column. Column 3 reports the ratio of the zone 1 value to the zone 2 value in the data. Column 5 reports the same ratio in the model.

Figure 4.2: Life-cycle income, wealth, HO: Data (L) vs. Model (R)

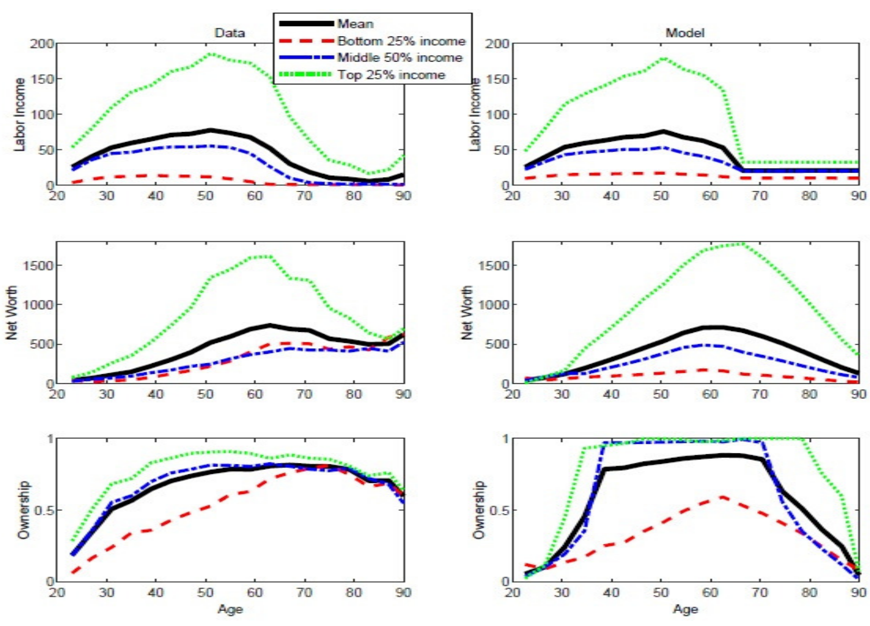


Figure 4.3: Marginal tax rates

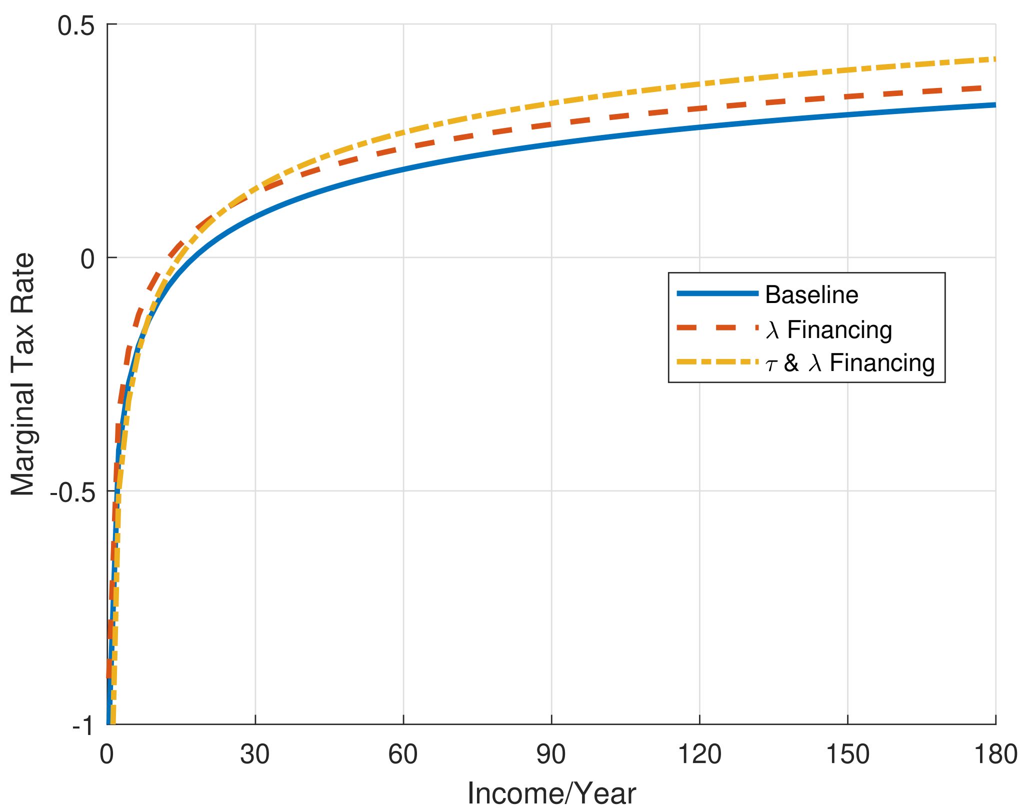


Table 4.2: UBI change in welfare: Alternative utility functions

This table presents the changes in welfare associated with an introduction of UBI across models with different utility functions. For each utility case, the baseline model is calibrated to NYC, and then $5,000 UBI is introduced. The "base" case is the baseline calibration in the text, where risk aversion $\gamma = 5$. We also present two cases with lower risk aversion ($\gamma = 4$ and $\gamma = 2$) and a case with a higher Frisch elasticity of labor supply. We present welfare by age, wealth, and productivity group.

Welfare change by age											
	All	24	32	40	48	56	64	72	80	88	96
base	1.08	4.11	0.86	-0.25	-0.79	-0.96	0.55	1.72	3.80	9.00	11.31
$\gamma = 4$	0.80	3.31	0.65	-0.34	-0.89	-1.07	0.13	1.21	3.74	9.51	11.63
$\gamma = 2$	-0.15	1.66	0.06	-0.78	-1.52	-2.03	-1.77	-0.62	3.86	11.14	13.92
Fr.El.=2	0.13	2.97	0.60	-0.58	-1.16	-1.48	-0.65	-0.15	0.98	3.63	4.78

Welfare change by wealth decile											
	All	1	2	3	4	5	6	7	8	9	10
base	1.08	8.82	9.29	10.52	3.43	0.66	0.15	-3.00	-5.10	-5.74	-8.21
$\gamma = 4$	0.80	7.80	8.39	10.66	3.39	0.41	-0.12	-3.28	-5.22	-5.85	-8.15
$\gamma = 2$	-0.15	5.56	7.20	8.92	3.24	-0.04	-1.31	-4.32	-6.03	-6.37	-8.41
Fr.El.=2	0.13	5.66	5.73	8.84	3.16	1.31	0.03	-2.81	-4.72	-6.34	-9.57

Welfare change by skill group					
	All	1	2	3	4
base	1.08	17.05	2.14	-2.76	-5.62
$\gamma = 4$	0.80	16.45	1.70	-2.91	-5.67
$\gamma = 2$	-0.15	13.96	0.55	-3.35	-6.09
Fr.El.=2	0.13	12.03	2.02	-2.63	-6.53

Table 4.3: UBI change in welfare

This table presents the changes in welfare associated with an introduction of UBI across models with alternative ways of paying for UBI. In each case, the baseline model is calibrated to NYC, and then $5,000 UBI is introduced. In the baseline case, UBI is financed by an across the board increase in tax rates (λ). In the progressive case, UBI is financed by an increase in tax progressivity (τ) with the remainder by across the board increase (λ). In the wealth case, UBI is financed by a wealth tax. In the "wealth+corp" case, UBI is financed by a mix of wealth and corporate taxes. In the "exogenous" case, UBI is financed by a benefactor outside the city, and costs locals nothing. The "UBI×2" is identical to the baseline case, but the size of UBI is doubled. We present welfare by age, wealth, and productivity group.

Welfare change by age											
	All	24	32	40	48	56	64	72	80	88	96
base	1.08	4.11	0.86	-0.25	-0.79	-0.96	0.55	1.72	3.80	9.00	11.31
progr.	0.46	7.12	1.58	-0.46	-1.85	-2.77	-2.14	-1.17	1.16	6.67	9.15
wealth	0.60	7.32	2.34	-0.22	-2.13	-3.29	-1.84	-0.88	1.28	7.14	9.40
wealth+corp	3.91	9.29	4.98	3.17	1.94	1.14	1.93	2.57	4.37	9.34	11.46
exogenous	7.01	10.91	7.31	6.23	5.58	5.18	5.73	6.24	7.84	11.94	13.99
UBI×2	2.94	8.16	2.05	0.30	-0.60	-0.64	2.74	5.05	8.38	15.44	18.38

Welfare change by wealth decile											
	All	1	2	3	4	5	6	7	8	9	10
base	1.08	8.82	9.29	10.52	3.43	0.66	0.15	-3.00	-5.10	-5.74	-8.21
progr.	0.46	13.82	13.31	13.21	4.80	1.11	0.26	-4.08	-8.54	-11.90	-17.36
wealth	0.60	13.67	14.67	14.54	6.83	2.32	1.39	-3.12	-8.27	-12.11	-23.87
wealth+corp	3.91	14.50	14.62	14.34	6.89	4.03	3.11	-0.19	-2.81	-4.58	-10.77
exogenous	7.01	15.16	14.73	14.42	7.51	5.95	4.98	2.81	1.84	1.57	1.17
UBI×2	2.94	16.44	16.56	19.25	8.82	3.25	2.06	-3.24	-8.12	-10.37	-15.22

Welfare change by skill group					
	All	1	2	3	4
base	1.08	17.05	2.14	-2.76	-5.62
progr.	0.46	24.84	3.20	-4.68	-12.46
wealth	0.60	21.63	3.83	-3.52	-12.19
wealth+corp	3.91	22.31	5.29	-0.11	-4.68
exogenous	7.01	23.43	6.94	3.14	1.56
UBI×2	2.94	32.77	5.31	-4.04	-10.44

Table 4.4: UBI change in quantities

This table presents changes in various quantities from an introduction in UBI for each of the models we solve. We separate the changes into Zone 1, Zone 2, and All.

	Base			$\gamma = 4$			$\gamma = 2$			Frisch Elast. = 2		
	Z1	Z2	All	Z1	Z2	All	Z1	Z2	All	Z1	Z2	All
Pop. Frac	-2.75	0.32	0.00	-4.78	0.57	0.00	-1.71	0.20	0.00	3.27	-0.38	0.00
Avg Age	1.63	-0.22	0.00	9.22	-0.91	0.00	5.31	-0.51	0.00	-5.93	0.74	0.00
Prod.	2.25	-0.21	0.00	5.02	-0.57	0.00	2.54	-0.37	0.00	1.66	-0.51	0.00
Hours	-6.43	-4.06	-4.43	-9.36	-3.69	-4.53	-4.91	-4.66	-4.74	1.22	-7.32	-6.21
Eff. Hours	–	–	-2.39	–	–	-2.46	–	–	-2.57	–	–	-4.32
Wage	–	–	0.61	–	–	0.62	–	–	0.64	–	–	1.30
Income	-8.39	-0.03	-1.50	-7.13	-0.19	-1.55	-0.83	-1.61	-1.62	4.80	-4.61	-2.85
Wealth	5.94	-18.03	-15.29	13.51	-19.14	-15.73	-0.40	-18.21	-16.15	-15.93	-17.13	-16.62
Cons.	-5.25	-4.96	-4.95	-5.19	-4.97	-5.01	-5.40	-5.35	-5.39	-7.35	-6.34	-6.43
Unit Size	2.52	-3.82	-3.28	4.80	-4.09	-3.32	1.60	-3.91	-3.48	-3.53	-4.77	-4.75
Ownership	7.94	-17.31	-15.24	11.41	-18.71	-16.21	3.62	-19.42	-16.83	-12.37	-13.05	-12.95
Rent	-2.30	-2.60	-2.95	-2.33	-2.65	-3.31	-2.44	-3.04	-3.15	-2.06	-3.35	-2.53
Price	-2.29	-2.62	-3.10	-2.34	-2.69	-3.59	-2.41	-2.97	-3.16	-2.06	-3.36	-2.30
	Progressive Tax			Wealth Tax			Wealth and Corp. Tax			Exogenous Funding		
	Z1	Z2	All	Z1	Z2	All	Z1	Z2	All	Z1	Z2	All
Pop. Frac	2.84	-0.34	0.00	0.28	-0.03	0.00	-4.27	0.50	0.00	-5.59	0.66	0.00
Avg Age	-8.13	0.85	0.00	1.22	-0.10	0.00	3.76	-0.44	0.00	7.17	-0.76	0.00
Prod.	-9.40	1.78	0.00	-4.99	1.03	0.00	2.75	-0.17	0.00	4.11	-0.30	0.00
Hours	-11.30	-9.95	-10.05	1.28	-3.34	-2.77	-4.66	-3.90	-4.09	-7.74	-4.49	-4.96
Eff. Hours	–	–	-8.12	–	–	0.46	–	–	-1.42	–	–	-2.76
Wage	–	–	2.62	–	–	-0.68	–	–	0.27	–	–	0.99
Income	-25.84	-1.69	-5.51	2.57	0.38	0.73	0.50	-0.50	-0.62	-1.59	-1.06	-1.47
Wealth	-2.90	-29.80	-26.50	-15.33	-33.10	-30.97	2.49	-18.54	-16.19	14.64	-4.80	-2.70
Cons.	-12.49	-11.47	-11.59	-6.74	-5.22	-5.34	-1.75	-1.29	-1.34	2.69	2.54	2.59
Unit Size	-3.51	-8.17	-7.95	-0.57	-3.30	-3.13	4.22	-1.70	-1.14	5.72	0.13	0.72
Ownership	-12.24	-18.46	-18.00	-9.40	-27.76	-26.26	0.24	-19.01	-17.40	4.86	-12.65	-11.16
Rent	-5.29	-5.97	-5.38	-2.61	-3.16	-2.98	-0.38	-0.28	-0.97	2.00	2.74	1.66
Price	-5.30	-6.03	-5.24	-2.58	-3.16	-2.93	-0.37	-0.33	-1.25	1.69	2.34	0.94

Bibliography

J. M. Abowd, F. Kramarz, and D. N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999. → pages 50, 57

J. M. Abowd, R. H. Creecy, F. Kramarz, et al. Computing person and firm effects using linked longitudinal employer-employee data. Technical report, Center for Economic Studies, US Census Bureau, 2002. → page 55

D. Acemoglu and D. Autor. Chapter 12 - skills, tasks and technologies: Implications for employment and earnings. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 4, Part B, pages 1043 – 1171. Elsevier, 2011. → page 133

D. Acemoglu, V. M. Carvalho, A. E. Ozdaglar, and A. Tahbaz-Salehi. The Network Origins of Aggregate Fluctuations. *Econometrica*, 80(5):1977–2016, 2012. ISSN 0012-9682. doi:10.2139/ssrn.1947096. → pages 3, 13

D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608, 2015. ISSN 00028282. doi:10.1257/aer.20130456. → page 3

D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Microeconomic origins of macroeconomic tail risks. *American Economic Review*, 107(1):54–108, 2017. ISSN 00028282. doi:10.1257/aer.20151086. → pages 3, 13

D. Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of computer and System Sciences*, 66(4):671–687, 2003. → pages 56, 119

P. Aghion, U. Akcigit, A. Hyytinen, and O. Toivanen. Living the american dream in finland: The social mobility of inventors. Mimeo., 2017. → pages 83, 86, 114

K. R. Ahern. Network Centrality and the Cross Section of Stock Returns. *SSRN Electronic Journal*, 2013. ISSN 1556-5068. doi:10.2139/ssrn.2197370. → pages 3, 28

K. R. Ahern and J. Harford. The importance of industry links in merger waves. *Journal of Finance*, 69(2):527–576, 2014. ISSN 00221082. doi:10.1111/jofi.12122. → page 3

F. Allen and A. Babus. Networks in finance. *The network challenge: strategy, profit, and risk in an interlinked world*, 367, 2009. → page 39

M. J. Andrews, L. Gill, T. Schank, and R. Upward. High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3):673–697, 2008. → pages 55, 117

E. Atalay. How important are sectoral shocks? *American Economic Journal: Macroeconomics*, 9 (4):254–280, 2017. ISSN 19457715. doi:10.1257/mac.20160353. → page 3

J. Bagger and R. Lentz. An empirical model of wage dispersion with sorting. *The Review of Economic Studies*, 86(1):153–190, 2019. → pages 56, 120

A. V. Banerjee. A simple model of herd behavior. *The quarterly journal of economics*, 107(3): 797–817, 1992. → page 49

S. Banerjee, S. Dasgupta, and Y. Kim. Buyer-supplier relationships and the stakeholder theory of capital structure. *Journal of Finance*, 63(5):2507–2552, 2008. ISSN 00221082. doi:10.1111/j.1540-6261.2008.01403.x. → page 4

D. R. Baqaee. Cascading Failures in Production Networks. *Econometrica*, 86(5):1819–1838, 2018. ISSN 0012-9682. doi:10.3982/ecta15280. → pages 3, 13

D. R. Baqaee and E. Farhi. The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica*, 87(4):1155–1203, 2019. ISSN 0012-9682. doi:10.3982/ecta15202. → pages 3, 13

G. Barlevy and P. Veronesi. Information acquisition in financial markets. *The Review of Economic Studies*, 67(1):79–90, 2000. → page 38

J. N. Barrot and J. Sauvagnat. Input specificity and the propagation of idiosyncratic shocks in production networks. *Quarterly Journal of Economics*, 131(3):1543–1592, 2016. ISSN 15314650. doi:10.1093/qje/qjw018. URL http://qje.oxfordjournals.org/content/early/2016/05/03/qje.qjw018{%}5Cnhttp://qje.oxfordjournals.org/content/early/2016/05/03/qje.qjw018.full.pdf. → pages 2, 3

C. Bartolucci, F. Devicienti, and I. Monzón. Identifying sorting in practice. *American Economic Journal: Applied Economics*, 10(4):408–38, 2018. → pages 56, 120

P. Beaudry, D. A. Green, and B. M. Sand. The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, 34(S1):199–247, 2016. → pages 53, 67

G. S. Becker and B. R. Chiswick. Education and the distribution of earnings. *The American Economic Review*, 56(1/2):358–369, 1966. → pages 52, 79

G. S. Becker, S. D. Kominers, K. M. Murphy, and J. L. Spenkuch. A theory of intergenerational mobility. *Journal of Political Economy*, 126(S1):S7–S25, 2018. → page 52

A. Bell, R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen. Who Becomes an Inventor in America? The Importance of Exposure to Innovation*. *The Quarterly Journal of Economics*, 134(2):647–713, 11 2018. ISSN 0033-5533. doi:10.1093/qje/qjy028. URL https://doi.org/10.1093/qje/qjy028. → pages 83, 86

M. F. Bellemare and C. J. Wichman. Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61, 2020. → page 81

B. Bernard, A. Capponi, and J. E. Stiglitz. Bail-ins and bailouts: Incentives, connectivity, and systemic stability. *Journal of Political Economy*, 130(7):000–000, 2022. → page 3

J. Bianchi. Overborrowing and systemic externalities in the business cycle. *American Economic Review*, 101(7):3400–3426, 2011. → page 4

S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026, 1992. → page 49

M. Böhm, D. Metzger, and P. Strömberg. Since you're so rich, you must be really smart: Talent, rent sharing, and the finance wage premium. Technical report, ECONtribute Discussion Paper, 2022. → page 112

S. Bonhomme and E. Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184, 2015. → page 120

S. Bonhomme, T. Lamadon, and E. Manresa. Discretizing unobserved heterogeneity. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2019-16), 2017. → page 117

S. Bonhomme, T. Lamadon, and E. Manresa. A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739, 2019. → pages 50, 51, 52, 56, 58, 63, 79, 83, 115, 117, 120, 127, 143, 144

S. Bonhomme, K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler. How much should we trust estimates of firm effects and worker sorting? Technical report, National Bureau of Economic Research, 2020. → pages 55, 117, 127

K. Borovickova and R. Shimer. High wage workers work for high wage firms. Working Paper, 2020. → pages 50, 52

D. Card, J. Heining, and P. Kline. Workplace heterogeneity and the rise of west german wage inequality*. *The Quarterly Journal of Economics*, 128(3):967–1015, 2013. → pages 50, 55, 111

D. Card, A. R. Cardoso, and P. Kline. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, 131(2): 633–686, 2016. → page 55

D. Card, A. R. Cardoso, J. Heining, and P. Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70, 2018. → pages 51, 53, 60, 85

V. M. Carvalho and A. Tahbaz-Salehi. Production networks: A primer. *Annual Review of Economics*, 11:635–663, 2019. → page 39

P. Choné and F. Kramarz. Matching workers' skills and firms' technologies: From bundling to unbundling. Technical report, Center for Research in Economics and Statistics, 2021. → pages 53, 62

P. Colla and A. Mele. Information linkages and correlated trading. *The Review of Financial Studies*, 23(1):203–246, 2010. → page 38

D. Cossin and H. Schellhorn. Credit risk in a network economy. *Management Science*, 53(10): 1604–1617, 2007. ISSN 00251909. doi:10.1287/mnsc.1070.0715. → page 3

J. De Loecker and J. Eeckhout. Global market power. 2021. Working Paper. → pages 53, 85

D. J. Deming. The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640, 2017. → pages 53, 67

J. Dow, I. Goldstein, and A. Guembel. Incentives for information production in markets where prices affect real investment. *Journal of the European Economic Association*, 15(4):877–909, 2017. → page 38

C. Dustmann, A. Lindner, U. Schönberg, M. Umkehrer, and P. Vom Berge. Reallocation effects of the minimum wage. *The Quarterly Journal of Economics*, 137(1):267–328, 2022. → pages 53, 85

D. Easley, J. Kleinberg, et al. *Networks, crowds, and markets*, volume 8. Cambridge university press Cambridge, 2010. → page 39

P.-A. Edin and P. Fredriksson. Linda-longitudinal individual data for sweden. Technical report, Working Paper, Department of Economics, Uppsala University, 2000. → page 111

P.-A. Edin, P. Fredriksson, M. Nybom, and B. Öckert. The rising return to non-cognitive skill. *American Economic Journal: Applied Economics*, 2022. → pages 51, 53, 54, 66

C. Edmond and S. Mongey. Unbundling labor. Technical report, Working Paper, 2021. → page 53

J. Eeckhout and P. Kircher. Identifying sorting in theory. *The Review of Economic Studies*, 78(3): 872–906, 2011. → pages 50, 56, 120

J. Eeckhout and P. Kircher. Assortative matching with large firms. *Econometrica*, 86(1):85–132, 2018. → page 60

E. Eyster and M. Rabin. Naive herding in rich-information settings. *American economic journal: microeconomics*, 2(4):221–43, 2010. → page 49

J. Favilukis, P. Mabille, and S. Van Nieuwerburgh. Out-of-town investors and city welfare. Working Paper Columbia GSB and UBC Sauder, October 2018. → pages 87, 89, 90

L. Feng and M. S. Seasholes. Correlated trading and location. *The Journal of finance*, 59(5): 2117–2144, 2004. → page 37

P. Fredriksson, L. Hensvik, and O. N. Skans. Mismatch of talent: Evidence on match quality, entry wages, and job mobility. *American Economic Review*, 108(11):3303–38, 2018. → pages 51, 54

X. Gabaix. The Granular Origins of Aggregate Fluctuations. *Econometrica*, 79(3):733–772, 2011. ISSN 0012-9682. doi:10.2139/ssrn.1111765. → page 3

D. Garcia and G. Strobl. Relative wealth concerns and complementarities in information acquisition. *The Review of Financial Studies*, 24(1):169–207, 2011. → page 38

D. Garcia and J. M. Vanden. Information acquisition and mutual funds. *Journal of Economic Theory*, 144(5):1965–1995, 2009. → page 38

M. Grinblatt and M. Keloharju. How distance, language, and culture influence stockholdings and trades. *The Journal of Finance*, 56(3):1053–1073, 2001. → page 37

S. J. Grossman and J. E. Stiglitz. On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408, 1980. → page 38

M. Hagedorn, T. H. Law, and I. Manovskii. Identifying equilibrium models of labor market sorting. *Econometrica*, 85(1):29–65, 2017. → pages 50, 52, 56, 120

B. Han and L. Yang. Social Networks, Information Acquisition, and Asset Prices. *Management Science*, 59(6):1444–1457, 2013. ISSN 0025-1909. doi:10.1287/mnsc.1120.1678. → pages 38, 39, 49

B. Han, D. Hirshleifer, and J. Walden. Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, pages 1–42, 2018. → page 49

K. Y. Hansen, J. J. Heckman, and K. J. Mullen. The effects of schooling and ability on achievement test scores. *Journal of Econometrics*, 121:39–98, 2004. → page 114

J. Heathcote, K. Storesletten, and G. L. Violante. Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics*, 132(4):1693–1754, 2017. → pages 89, 91

J. Heckman and J. Scheinkman. The importance of bundling in a gorman-lancaster model of earnings. *The Review of Economic Studies*, 54(2):pp. 243–255, 1987. → page 53

J. J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–481, 2006. → pages 53, 114

C. A. Hennessy and D. Livdan. Debt, bargaining, and credibility in firm–supplier relationships. *Journal of Financial Economics*, 93(3):382–399, 2009. → page 4

B. Herskovic. Networks in Production: Asset Pricing Implications. *Journal of Finance*, 73(4): 1785–1818, 2018. ISSN 15406261. doi:10.1111/jofi.12684. → pages 3, 13

B. Herskovic, B. Kelly, H. Lustig, and S. V. Nieuwerburgh. Firm volatility in granular networks. *Journal of Political Economy*, 128(11):4097–4162, 2020. ISSN 1537534X. doi:10.1086/710345. → page 3

H. Hong, J. D. KUBIK, and J. C. STEIN. Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers. *The Journal of Finance*, 60(6):2801–2824, 2005. ISSN 1540-6261. doi:10.1111/j.1540-6261.2005.00817.x. → page 37

M. Horvath. Sectoral shocks and aggregate fluctuations. *Journal of Monetary Economics*, 45(1): 69–106, 2000. ISSN 03043932. doi:10.1016/S0304-3932(99)00044-6. → page 3

Z. Ivković and S. Weisbenner. Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices. *The Review of Financial Studies*, 20(4): 1327–1357, 2007. → page 37

M. O. Jackson. *Social and economic networks*. Princeton university press, 2010. → page 39

M. O. Jackson. An overview of social networks and economic applications. *Handbook of social economics*, 1:511–585, 2011. → page 39

G. Jarosch, E. Oberfield, and E. Rossi-Hansberg. Learning from coworkers. *Econometrica*, 89(2): 647–676, 2021. → page 62

K. Jochmans and M. Weidner. Fixed-effect regressions on network data. *Econometrica*, 87(5): 1543–1560, 2019. → page 55

S. Jäger, B. Schoefer, and J. Heining. Labor in the boardroom. *Quarterly Journal of Economics*, 136(2):669–725, 2021. → page 85

M. Kacperczyk, S. Van Nieuwerburgh, and L. Veldkamp. Rational attention allocation over the business cycle. Technical report, National Bureau of Economic Research, 2009. → page 38

M. Keane and R. Rogerson. Micro and macro labor supply elasticities: A reassessment of conventional wisdom. *Journal of Economic Literature*, 50(2), 2012. → page 92

P. Kline, R. Saggio, and M. Sølvsten. Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898, 2020. → pages 51, 52, 55, 56, 57, 58, 59, 69, 116, 117, 119, 120

A. Kraus and R. H. Litzenberger. a State-Preference Model of Optimal Financial Leverage. *The Journal of Finance*, 28(4):911–922, 1973. ISSN 15406261. doi:10.1111/j.1540-6261.1973.tb01415.x. → page 4

T. Lamadon, M. Mogstad, and B. Setzler. Imperfect competition, compensating differentials, and rent sharing in the us labor market. *American Economic Review*, 112(1):169–212, 2022. → pages 50, 51, 52, 53, 60, 63, 83, 85, 127, 143, 144

M. T. Leary and M. R. Roberts. Do Peer Firms Affect Corporate Financial Policy? *Journal of Finance*, 69(1):139–178, 2014. ISSN 00221082. doi:10.1111/jofi.12094. → page 28

R. Lentz, S. Piyapromdee, and J.-M. Robin. On worker and firm heterogeneity in wages and employment mobility: Evidence from danish register data, 2018. Working Paper. → pages 50, 52

I. Lindenlaub. Sorting multidimensional types: Theory and application. *The Review of Economic Studies*, 84(2):718–789, 2017. → pages 52, 53, 68, 79

I. Lindenlaub and F. Postel-Vinay. Multidimensional sorting under random search. Working Paper, 2020. → pages 52, 71

A. Lindner, B. Muraközy, B. Reizer, and R. Schreiner. Firm-level technological change and skill demand. *Working Paper*, 2021. → page 82

E. Lindqvist. Height and leadership. *Review of Economics and Statistics*, 94(4):1191–1196, 2012. → page 53

E. Lindqvist and R. Vestman. The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, 3(1): 101–128, 2011. → pages 51, 54, 112, 114

J. Lise and J.-M. Robin. The macrodynamics of sorting between workers and firms. *American Economic Review*, 107(4):1104–35, 2017. → pages 60, 122

J. Long and C. Plosser. Real business cycles. *Journal of Political Economy*, 91(1):39–69, 1983. doi:10.4324/9780203443965.ch18. → page 13

G. Lorenzoni. Inefficient credit booms. *The Review of Economic Studies*, 75(3):809–833, 2008. → page 3

M. D. Maggio, F. Franzoni, A. Kermani, and C. Sommavilla. The relevance of broker networks for information diffusion in the stock market. *Journal of Financial Economics*, 134(2): 419–446, 2019. ISSN 0304-405X. doi:10.1016/j.jfineco.2019.04.002. → pages 37, 38

B. Mandelbrot. Paretian distributions and income maximization. *The Quarterly Journal of Economics*, 76(1):57–85, 1962. → page 53

D. McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142, 1974. → pages 61, 122

A. Mele and F. Sangiorgi. Uncertainty, information acquisition, and price swings in asset markets. *The Review of Economic Studies*, 82(4):1533–1567, 2015. → page 38

S. V. Nieuwerburgh and L. Veldkamp. Information acquisition and under-diversification. *The Review of Economic Studies*, 77(2):779–805, 2010. → page 38

H. N. Ozsoylev and J. Walden. Asset pricing in large information networks. *Journal of Economic Theory*, 146(6):2252–2280, 2011. → page 38

H. N. Ozsoylev, J. Walden, M. D. Yavuz, and R. Bildik. Investor networks in the stock market. *The Review of Financial Studies*, 27(5):1323–1366, 2014. → page 37

J. Peress. Wealth, information acquisition, and portfolio choice. *The Review of Financial Studies*, 17(3):879–914, 2004. → page 38

J. Peress. The tradeoff between risk sharing and information production in financial markets. *Journal of Economic Theory*, 145(1):124–155, 2010. → page 38

I. Prostakova. Capital structure decisions in the supplier-customer network. In *31st Australasian Finance and Banking Conference*, 2018. → page 4

A. A. Robichek and S. C. Myers. Problems in the Theory of Optimal Capital Structure. *The Journal of Financial and Quantitative Analysis*, 1(2):1, 1966. ISSN 00221090. doi:10.2307/2329989. → page 4

S. Rosen. A note on aggregation of skills and labor quality. *The Journal of Human Resources*, 18 (3):425–431, 1983. → page 53

M. Sattinger. Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2):pp. 831–880, 1993. ISSN 00220515. URL http://www.jstor.org/stable/2728516. → pages 52, 79

J. H. Scott. A Theory of Optimal Capital Structure. *The Bell Journal of Economics*, 7(1):33, 1976. ISSN 0361915X. doi:10.2307/3003189. → page 4

R. J. Shiller and J. Pound. Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior & Organization*, 12(1):47–66, 1989. → page 37

O. N. Skans, P. Choné, and F. Kramarz. When workers? skills become unbundled: Some empirical consequences for sorting and wages. *Working Paper*, 2022. → page 53

J. Song, D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter. Firming up inequality. *The Quarterly Journal of Economics*, 134(1):1–50, 2018. → page 50

I. Sorkin. Ranking firms using revealed preference. *Quarterly Journal of Economics*, 133(3): 1331–1393, 2018. → pages 50, 61, 122

J. E. Stiglitz. Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Take-Overs. *Bell J Econ Manage Sci*, 3(2):458–482, 1972. ISSN 00058556. doi:10.2307/3003033. → page 4

S. Titman. The effect of capital structure on a firm's liquidation decision. *Journal of Financial Economics*, 13(1):137–151, 1984. ISSN 0304405X. doi:10.1016/0304-405X(84)90035-7. → page 4

S. Titman and R. Wessels. The Determinants of Capital Structure Choice. *The Journal of Finance*, 43(1):1–19, 1988. ISSN 15406261. doi:10.1111/j.1540-6261.1988.tb02585.x. → page 4

S. Van Nieuwerburgh and L. Veldkamp. Information immobility and the home bias puzzle. *The Journal of Finance*, 64(3):1187–1215, 2009. → page 38

R. E. Verrecchia. Information acquisition in a noisy rational expectations economy. *Econometrica: Journal of the Econometric Society*, pages 1415–1430, 1982. → page 38

R. J. Willis. Wage determinants: A survey and reinterpretation of human capital earnings functions. In *Handbook of labor economics*, volume 1, pages 525–602. Elsevier, 1986. → page 50

Appendix A

Firm Heterogeneity in Skill Returns

A.1 Data and Samples Construction

A.1.1 Data

Base sample. The main data source for our analysis is the *Longitudinal Integrated Database for Health Insurance and Labor Market Studies* (LISA) by Statistics Sweden (SCB). LISA contains employment information (such as employment status, organization and workplace identifiers, industry and, from 2001, occupation), tax records (including labor and capital income) and demographic information (such as age, education) for all individuals 16 years of age and older domiciled in Sweden. LISA starts in 1990, with the most recent data including 2017.

Our measure of earning returns is annual labor income from the employer with highest recorded earnings. This is available for all workers, not top-coded, and includes end-of-year bonuses and performance pay. LISA reports a unique identifier for each individual's "company of employment", a so-called organization number, as well as a workplace identifier, which is the combination of organization number, employment location, and industry. To be consistent with the earning measure, and with the firm literature (see, among others, Card et al., 2013), we use the workplace with the highest earnings in a given year as the worker's "firm".

We keep workers dependently employed in the private nonprimary sector who earn above the *Prisbasbelopp* (the minimum amount of earnings that qualifies for the earnings-related part of the public pension system; see also Edin and Fredriksson, 2000). In 2008, the *Prisbasbelopp* was 41,000 kr or approximately 6,200 USD. We drop all observations with incomplete data (missing test scores, age, or workplace) and restrict the sample to 20–60 year old males. This process results in a sample of approximately 1 million unique workers, 26 thousand firms, and 6.6 million worker $\times$ year observations for the main sample period of 1999–2008.[1] Colum (1) of Table A.1 reports summary statistics for the resulting sample.

Measures of cognitive and noncognitive traits. A strength of our data source is that we have access to extensive and consistent measures of workers' cognitive and noncognitive attributes. This information comes from military enlistment tests, which were mandatory for Swedish males before 2007 and typically taken between age 18 and 19. In the early 2000s, Sweden started requiring progressively fewer males to do military service. The service was abolished in 2010. Before 2007, however, all males were required to take the military enlistment tests and test scores are available for almost 90 percent of males born up to the 1980s (e.g., see Figure A.1 in Böhm et al., 2022).

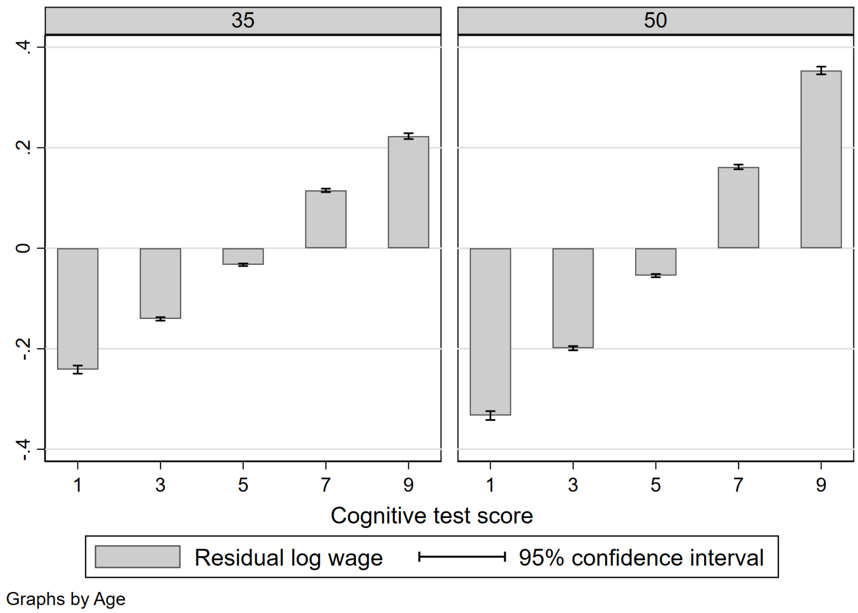
One might worry that certain individuals could deliberately perform badly on these tests to avoid military service. There are, however, several pieces of evidence suggesting this was not a major problem. In particular, we emphasize that employers routinely put considerable weight on military service performance and anecdotal evidence suggests that some positions – like being an officer in the navy – were important for the networks individuals would obtain; a substantial fraction of individuals working in influential positions within Swedish society went through these military service assignments. Consistent with these observations, and perhaps more importantly, military test scores have been shown to significantly predict future earnings at long intervals after the tests, as well as other labor market outcomes such as managerial positions and incidence of unemployment (see, e.g., Lindqvist and Vestman, 2011).

The enlistment process for military service spans two days and evaluates a person's medical and physical status as well as cognitive and mental abilities. We use the tests of cognitive and noncognitive ability, which are well established in the labor economics literature, for our analysis. The test of cognitive ability consists of four different parts (logic, verbal, spatial, and technical comprehension), each of which is constructed from 40 questions. These are aggregated into an overall score. The test is a rich measure of general competence and intelligence and it has a stronger fluid IQ component than the American AFQT, which focuses more on crystallized IQ. The aggregate cognitive score ranges from the integer value 1 (lowest) to 9 (highest), according to a STANINE (standard nine) scale that approximates a Normal distribution with a mean of 5 and standard deviation of 2 (meaning that a gap of two scores covers a standard deviation).

Noncognitive ability is assessed through a 25-minute semi-structured interview by a certified psychologist. Individuals are graded on, among others, their willingness to assume responsibility, independence, outgoing character, persistence, emotional stability, and power of initiative (Swedish National Service Administration, referenced by, among others, Lindqvist and Vestman, 2011). The psychologist weighs these components together and assigns an overall noncognitive score on a

[1] We also document results for two alternative periods, 1990–1999 and 2008–2017.

(A) Cognitives



(B) Noncognitives

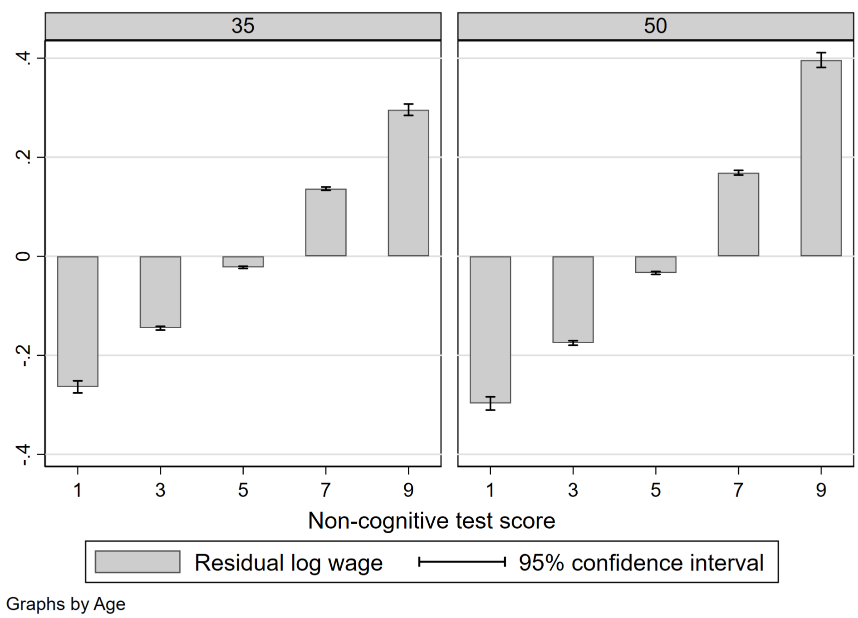


Figure A.1: Average Earnings of Males at Age 35 and 50, by Test Score Group.

Notes: Earnings for different test score ranks {1,3,5,7,9}; values are residualized using full age × year dummy interactions. Sample period: 1990–2017. 95% confidence intervals indicated by brackets.

STANINE scale. Lindqvist and Vestman (2011), on p. 108f, discuss in detail how the noncognitive score is related to, and different from, other measures often used in the literature on personality and labor market outcomes. Rather than assessing a unique trait, the noncognitive score assesses the ability to function in a demanding environment (military combat). Previous work provides robust evidence that these traits are also rewarded in the labor market.[2]

Test scores and later life outcomes. An important advantage of the military test scores is that they allow for a professional standardized measurement of different ability dimensions over a large population. Military enlistment scores are by design exogenous and predetermined with respect to individuals' career choices. Although cognitive and noncognitive ability are not fixed, they are hard for individuals to manipulate after late childhood or early adulthood (Hansen et al., 2004, Heckman et al., 2006). Crucially, as we show in Figure A.1, the tests are strongly associated to labor market outcomes and accurately predict earnings several decades later. Figure A.1 compares the earnings of workers with different STANINE scores in our sample (residualized using full age $\times$ year dummy interactions) and documents highly significant differences at ages 35 and 50, across both cognitive and noncognitive competencies. These plots emphasize the lasting informational content of the tests and their relevance for long term labor market outcomes. Strong significance at long lags is not always the case with ability tests in survey data and is partly due to the fine-grained and homogeneous nature of the procedures used to elicit different attributes, resulting in measures that can be mapped into earnings for the whole population of interest over its working life cycle.[3]

A.1.2 Estimation Samples

Clustered estimation: sample and firm grouping. We concentrate on the largest set of firms that are connected via worker mobility. This corresponds to moving from column (1) to (2) in Table A.1, and is in fact not strictly necessary: for estimating clusters only mobility between firm classes is required, a condition almost trivially satisfied here. Nonetheless, we keep with existing literature and require connected firms; this is not a consequential sample restriction, as shown in Table A.1. The latter finding indicates that even our initial restrictions are enough to lead to a sample of relatively large and well-connected firms. Overall, there are 25,783 unique firms and 510,077 workers who move between firms at least once during 1999–2008 in column (2) of the table.

[2] Individuals who score sufficiently high on the cognitive test are also evaluated for leadership ability, again on a STANINE scale. The leadership score is meant to capture the suitability to become an officer. Since leadership is only assessed for a subset of individuals, we focus on cognitive and noncognitive ability in our analysis. Noncognitive ability and leadership ability are also highly correlated; in our data the correlation is above 0.8, while the correlation of cognitive and noncognitive is 0.3.

[3] Aghion et al. (2017) further show that cognitive military test scores similar to ours strongly predict whether an individual becomes an inventor in Finland, another important later in life outcome and closely related to our analyses in Section 3.6.2.

Next, we employ the k-means algorithm (see also Bonhomme et al., 2019, Section 4) to group firms into 100 clusters. We do this by using variation in mean earnings, mean cognitive, and mean noncognitive skills, which reflect the dimensions of firm heterogeneity that we are interested in. In particular, differing technologies should lead to variation in both firms' skill composition and earnings. We estimate model (3.10) using this sample and the definition of firm clusters (i.e., the j subscripts refer to the 100 clusters). Results are reported in Column (1) of Table 3.1. Section 3.6.3 in the paper and associated Appendix A.5.3 examine robustness with respect to alternative clustering criteria as well as to the number of firm classes.

Table A.1: Summary statistics for the estimation samples

	Full sample (1)	Largest connected (2)	Leave-one-out (3)	Match-level (4)
Number of observations	6,610,567	6,609,865	3,267,381	1,188,618
Number of stayers	578,146	578,146	-	-
Number of movers	510,077	510,077	477,424	477,424
Number of firms	25,839	25,783	19,085	19,085
Average log annual earnings	7.84	7.84	7.83	7.83
StDev log annual earnings	0.60	0.60	0.64	0.71
Average cognitive skill	5.28	5.28	5.44	5.44
Average noncognitive skill	5.13	5.13	5.23	5.23
Average age	37.32	37.32	36.35	36.35

Notes: Summary statistics for successively more restricted samples. Column (1) are all males aged 20–60 with nonmissing employer, earnings, and test scores 1999–2008 at firms that exist at least five years with at least ten sample workers on average. Column (2) extracts the largest connected set of firms and their employees. Column (3) extracts the leave-one-out connected set of firms and removes workers who stay in the same firm in all years they are observed. Column (4) collapses the column (3) sample to worker–firm matches (summary statistics weighted by underlying frequencies). Earnings are real annual labor income in 2008 Swedish kronor. Cognitive and noncognitive scores are in Stanine scale. Our estimation samples are in bold font, (2) for clustering and (4) for the bias-correction approach.

Bias-correction estimation: leave-one-out match-level samples. The estimation of variance components with the bias correction requires a set of firms that are leave-one-out connected by mobility of high and low skill workers in both the cognitive and noncognitive dimension. We meet this condition by only sampling firms that are leave-one-out connected through: (i) low skill workers ($c \leq 5, n \leq 5$), (ii) low in one and high in the other dimension workers ($c \geq 6, n \leq 5$ or $c \leq 5, n \geq 6$), and (iii) generally high-skill workers ($c \geq 6, n \geq 6$). A leave-one-out connected set of firms remains connected when any one worker is removed. This requires finding the workers that constitute cut vertices or articulation points in the corresponding bipartite network (Kline et al., 2020, Computational Appendix 2.1).

The algorithm to construct our estimation sample works as follows:

Step 1: We use Python's NetworkX package to identify the articulation points of the worker–firm graph, remove them and find the largest connected set that remains, then add back those articulation points that are connected to this largest leave-one-out connected set.

Step 2: We identify the largest leave-one-out connected set separately for the three skill groups (i)–(iii) and only keep those firms that are in the intersection of these sets.

We repeat Steps 1 and 2 until there is no reduction in the size of the graph (i.e., the three largest leave-one-out connected sets coincide). This final set is leave-one-out connected for the three skill groups.

We estimate the model at the worker–firm match level to account for potential serial correlation within worker–firm employment spells. That is, we collapse the data to means and drop workers who stay in the same firm throughout the period, since in the match-level estimation these do not contribute to identifying the firm effects. We thereby follow exactly Kline et al. (2020, Appendix A)'s recommendations for estimating variance components in panels of $T > 2$.

The final firm-level sample to estimate (3.10) is summarized in Column (4) of Table A.1. This consists of 19,084 unique firms and 477,423 mover workers within the firm-level sample. The leave-one-out connectedness requirement increases employer size as it reduces the number of firms (26%) relatively more than the number of workers (7%). However, these reductions seem to have moderate effects. The average and dispersion of earnings do not change much but workers in the firm-level sample with larger firms are slightly younger and more skilled. The smaller number of observations in the match-level sample, without stayers and collapsed to the worker–firm match level, also reduces the computational burden (see footnote 6 below). For comparison, we also show results for the leave-one-observation-out sample in Table A.3 and, as expected, estimated dispersions of firm returns are substantially larger. In that sense, the match-level results in the main text are conservative.

A.2 Overview of Econometric Methods

Throughout the paper we use high-dimensional firm effects specifications featuring firm-specific returns to cognitive and noncognitive skills. Estimates from these models are employed to study quadratic forms of model parameters. The baseline linear model is[4]

$$\log(w_{ijt}) = \mu_i + \lambda_j^0 + c_i \cdot \lambda_j^c + n_i \cdot \lambda_j^n + \varepsilon_{ijt}.$$

Of particular economic interest is the set of second moments of firm and worker specific parameters. For instance, in the standard double fixed effect model, one might interpret $\text{cov}(\mu, \lambda^0)$ as a measure of sorting of high-type workers into high-type firms. However, the naive plug-in estimates of these moments are prone to biases. In fact, developing unbiased estimators of such quadratic forms is the object of several papers in the firm heterogeneity literature (Andrews et al., 2008, Bonhomme et al., 2019, 2020, Kline et al., 2020). Since our interest i in studying similar second moments, in what follows we briefly overview some details about the methods we employ to estimate firm effects.[5]

A.2.1 Estimating Bias-Corrected Quadratic Forms

We begin by rewriting our baseline specification as:

$$\begin{aligned} \log(w_{ijt}) &= \mu_i + \lambda_j^0 + c_i \cdot \lambda_j^c + n_i \cdot \lambda_j^n + \varepsilon_{ijt}, \\ &\equiv \mathbb{X}_{ij}\beta + \varepsilon_{ijt} \end{aligned} \tag{A.1}$$

where $\beta = [\mu; \lambda^0; \lambda^c; \lambda^n]' \equiv [\mu_1, ..., \mu_I;\ \lambda_1^0, ..., \lambda_J^0;\ \lambda_1^c, ..., \lambda_J^c;\ \lambda_1^n, ..., \lambda_J^n]'$ is the parameter vector and $\mathbb{X}_{ij} = [\mathbf{1}_i,\ \mathbf{1}_j,\ c_i\mathbf{1}_j,\ n_i\mathbf{1}_j]$ is the data matrix.

The symbol $\mathbf{1}_i$ denotes a $I \times 1$ indicator vector whose elements are all zero except the i^{th} coordinate (corresponding to worker i) which is set to 1. Similarly $\mathbf{1}_j$ is a $J \times 1$ indicator vector for firm j.

Kline et al. (2020) suggest an unbiased estimator for arbitrary quadratic forms involving the coefficients of (A.1) in the form of $\beta' A \beta$, for given matrix A. By appropriately choosing the A matrix, one can recast all the second moments of firm parameters λ_j^0, λ_j^c, and λ_j^n into quadratic expressions of the form $\beta' A \beta$.

Constructing quadratic forms. We begin by definining three row vectors associated to different firm parameters: $\mathbb{X}_{ij}^0 = [0_{1\times I},\ \mathbf{1}_j,\ 0_{1\times J},\ 0_{1\times J}]$, $\mathbb{X}_{ij}^c = [0_{1\times I},\ 0_{1\times J},\ \mathbf{1}_j,\ 0_{1\times J}]$, and $\mathbb{X}_{ij}^n = [0_{1\times I},\ 0_{1\times J},\ 0_{1\times J},\ \mathbf{1}_j]$,

[4] In the specifications studied in the main body we also include a broad set of control variables which are ignored here for notational simplicity.

[5] For in-depth discussions of these estimators see Kline et al. (2020) and Bonhomme et al. (2017, 2019).

where i identifies worker and j is firm. Also, we let $\mathbb{X}$ denote the matrix that results from vertically stacking all the observations in row vector $\mathbb{X}_{ij}$. Then, $\mathbb{X}^0$, $\mathbb{X}^c$, and $\mathbb{X}^n$ denote the matrices that result from vertically stacking $\mathbb{X}^0_{ij}$, $\mathbb{X}^c_{ij}$ and $\mathbb{X}^n_{ij}$. Finally, we define

$$A^0 = \frac{1}{\sqrt{N}} \left(\mathbb{X}^0 - \overline{\mathbb{X}}^0 \right)$$
$$A^c = \frac{1}{\sqrt{N}} \left(\mathbb{X}^c - \overline{\mathbb{X}}^c \right)$$
$$A^n = \frac{1}{\sqrt{N}} \left(\mathbb{X}^n - \overline{\mathbb{X}}^n \right)$$

where $\overline{\mathbb{X}}^0 = \frac{1}{N}[0_{N\times I},\ 1_{N\times J},\ 0_{N\times J},\ 0_{N\times J}]$, $\overline{\mathbb{X}}^c = \frac{1}{N}[0_{N\times I},\ 0_{N\times J},\ 1_{N\times J},\ 0_{N\times J}]$, and $\overline{\mathbb{X}}^n = \frac{1}{N}[0_{N\times I},\ 0_{N\times J},\ 0_{N\times J},\ 1_{N\times J}]$. One can use the A matrices above to estimate second moments of interest, e.g. $\text{VAR}(\lambda^0) = \beta'(A^{0\prime}A^0)\beta$ or $\text{COV}(\lambda^c, \lambda^n) = \beta'(A^{c\prime}A^n)\beta$. In what follows we set $A = A_1'A_2$ to estimate $\theta = \beta'A\beta$, where A_1 and A_2 could be any of A^0, A^c, and A^n (depending on which moments we are interested in).

Plug-in estimator. The plug-in estimator $\hat{\theta}_{\text{PI}} = \hat{\beta}'A\hat{\beta}$ can be obtained by simply using the OLS estimates of $\hat{\beta}$ in the quadratic form defining θ. However, the plug-in estimator is biased and its expected value is

$$\mathbb{E}[\hat{\theta}_{\text{PI}}] = \theta + \text{trace}(A \times \text{VAR}[\hat{\beta}]) = \theta + \sum_{k=1}^{N} B_{kk}\sigma_k^2 \tag{A.2}$$

where $S = \mathbb{X}'\mathbb{X}$, B_{kk} is the k-th diagonal element of $B = \mathbb{X}S^{-1}AS^{-1}\mathbb{X}'$ corresponding to observation k, and σ_k^2 is the variance of error term of observation k. Therefore, the bias in the plug-in estimator can be corrected by using unbiased estimates of σ_k^2, which is the route we take when estimating the model at the level of individual firms.

Bias-corrected quadratic forms. We use leave-k-out OLS estimators of β, denoted by $\hat{\beta}_{-k}$, that are obtained from a sample where the observation k is excluded. This delivers an unbiased estimator of σ_k^2 such that

$$\hat{\sigma}_k^2 = y_k(y_k - x_k\hat{\beta}_{-k}), \tag{A.3}$$

where y_k is the dependent variable (i.e. log earnings) of observation k and x_k is the corresponding independent variables vector (i.e. row k of $\mathbb{X}$). Using the $\hat{\sigma}_k^2$ above, we compute the bias corrected estimator of θ as

$$\hat{\theta}_{KSS} = \hat{\beta}'A\hat{\beta} - \sum_{k=1}^{N} B_{kk}\hat{\sigma}_k^2. \tag{A.4}$$

Large Scale Computations. Estimating $\hat{\theta}_{KSS}$ is computationally expensive for large data-sets with many estimated parameters such as ours. Like Kline et al. (2020), we use a variant of the random projection method of Achlioptas (2003, known as Johnson-Lindenstrauss Approximation, or JLA) to estimate the $\hat{\sigma}_k^2$ and B_{kk} required in the estimation of $\hat{\theta}_{KSS}$. JLA suggests the following approximation:

$$\hat{P}_{kk} = \frac{1}{p}||R_P \mathbb{X} S^{-1} x_k||^2$$
$$\hat{B}_{kk} = \frac{1}{p}(R_B A_1 S^{-1} x_k)'(R_B A_2 S^{-1} x_k)$$
$$\hat{\sigma}^2_{k,JLA} = \frac{y_k(y_k - x_k\hat{\beta})}{1-\hat{P}_{kk}}(1 - \frac{1}{p}\frac{3\hat{P}_{kk}^3 + \hat{P}_{kk}^2}{1-\hat{P}_{kk}}),$$

where $p \in \mathbb{N}$ is a number much smaller than the total number of estimated parameters. That is, we can achieve a material reduction in the dimensionality of the prolem. The $R_P, R_B \in \{-1,1\}^{p \times N}$ are random matrices of order $p \times N$ featuring elements equal to +1 and -1 with equal probabilities. This makes computations significantly faster when parameters are estimated at the level of individual firms.[6]

A.2.2 Cluster-Based Estimation

Models with two sided heterogeneity rely on job movers to identify the unobserved firm and worker parameters. In typical employer–employee linked data sets the number of job movers per firm tends to be small, which leads to the well known limited mobility bias in quadratic forms of these estimates. To alleviate this problem, group based estimates have been suggested in the literature. In this approach, firm parameters are assumed to only vary across groups or clusters of firms, rather than individual firms. Under this assumption about the underlying data generating process and further assuming that the number of groups is limited, the number of job moves per group of firms is sufficiently large, which alleviates the small sample bias concern.

Partitioning returns across clusters. To adapt this framework to our setting, we begin by rewriting the baseline specification as

$$\log(w_{ijt}) = \mu_i + \lambda^0_{g(j)} + c_i \cdot \lambda^c_{g(j)} + n_i \cdot \lambda^n_{g(j)} + \varepsilon_{ijt},$$

[6]Estimating the bias-corrected second moments of parameters in model (3.10) on the data in Column (4) of Table A.1 takes about 20–30 hours using Python and the JLA approximation with $p = 50$ depending on the Swedish server's workload. Setting $p = 50$ is in line with Kline et al. (2020) and we have tested that further increasing p does not change our results.

where $g : \{1,...,J\} \rightarrow \{1,...,K\}$ is a partitioning function that maps firm j into cluster $g(j)$ that the firm j belongs to, and K is the total number of groups. These groups could in principle be the individual firms, i.e., $g(j) = j$, but only in models with a reduced number of groups is the limited mobility bias less of a concern.

Two-step estimation. We estimate the model in two steps i.e. (i) partition firms into K disjoint groups (ii) estimate the model using the firm groups. Bonhomme and Manresa (2015) show that a k-means estimator can consistently identify the firm classes up to a relabeling of groups. In the first step, as discussed in Section 3.2.2, we use the average earnings as well as average cognitive and noncognitive traits of their workers to group firms. Intuitively, the earnings and average skills in firms with identical intercepts and returns should be the same and one could then use these observed firm variables to define separate firm classes. The structural literature advocates going beyond earnings when clustering firms (Bagger and Lentz, 2019, Bartolucci et al., 2018, Eeckhout and Kircher, 2011, Hagedorn et al., 2017), since a classification may fail to be identified when two firm classes have identical earnings distributions in the cross section.[7]

Once firm groups are defined, firm and worker parameters are identified (up to the normalization discussed in Section 3.3.1 of the main text) under the assumptions of serial conditional independence of earnings and random job mobility (Bonhomme et al., 2019), and estimated using panel regressions in conjunction with skill proxies.

A.2.3 Testing Equality of Firm Effects across Worker Skills

In Section 3.2.3, and Figure 3.1, we choose the years 2004 and 2007 to test the equality of firm effects for high versus low skilled workers. Two years are selected to exclude potential serial correlation within employment spells due to estimated standard errors (see Kline et al., 2020, Computational Appendix 2.5).

Years are non-adjacent, in order to remove partial employment years when workers switch firms, while not too far apart to minimize any potential changes in firm effects over time. The sample is selected on firms that are leave-one-out connected in both high and low levels of the respective skill dimension ("double-connected").

To gauge robustness, we also replicate the analysis for alternative duplets of years. Like in Section 3.2.3, we focus on testing the null hypothesis that firm effects are equal for high and low skills groups; the hypotheses are separately tested for cognitive and noncognitive skills. Table A.2 shows the resulting test statistics and sample sizes of the respective double-connected individual firms for several year pairs using the bias-correction approach. The last column reports the corresponding

[7] For example, a firm class may have higher intercepts and the other higher returns but worker sorting is such that observed earnings are the same. See also discussion in Bonhomme et al. (2019, page 14).

test statistics among 100 firm classes using clustering as in Figure 3.1(c)–(d).

Table A.2: Tests for equality of firm effects by high- versus low-skill workers (by year combination and cognitive / noncognitive)

Year origin (1)	Year destination (2)	Skill (3)	Test Statistic Firm-level (4)	# Firms (5)	Test Statistic Grouped (6)
1999	2002	C	3.66	8,757	9.22
1999	2002	N	2.49	9,766	12.45
2000	2003	C	2.60	8,653	8.66
2000	2003	N	0.39	9,648	8.14
2001	2004	C	2.76	7,922	9.93
2001	2004	N	1.78	8,941	7.65
2002	2005	C	0.60	7,904	10.83
2002	2005	N	3.50	8,772	6.96
2003	2006	C	4.04	8,335	13.88
2003	2006	N	0.85	9,258	7.00
2004	2007	C	4.18	9,269	17.33
2004	2007	N	4.56	10,209	6.30
2005	2008	C	3.26	9,846	10.74
2005	2008	N	2.54	10,825	5.38

Notes: Table A.2 expands on Figure 3.1 to show test statistics associated with the null hypothesis that firm effects ($\theta_j^{S=0}$) and ($\theta_j^{S=1}$) are equal across skill level, where skill $S \in \{C,N\}$. Test statistic for firm-level bias-adjusted estimates as in Figure 3.1(a)–(b) are shown in column (4). The associated number of double-connected firms in each of the skill types and year combinations are reported in column (5). The last column reports the corresponding test statistic among 100 firm classes using the clustering approach as in Figure 3.1(c)–(d).

A.3 A Labor Market with Two-Sided Heterogeneity and Heterogeneous Skills

To examine the interaction of employer and employee heterogeneity we develop an empirically tractable model featuring workers with different cognitive and noncognitive abilities. We consider a static setting with a continuum of firms, each producing its own distinct product using labor. All firms benefit from more able workers, but each firm exhibits an idiosyncratic return to skills. Firm-specific skill returns act as a force for sorting of high-skill workers into high-return firms, something that the matching literature has long emphasized. These layers of heterogeneity are embedded in a labor market where employers choose how many workers to hire based on the demand for their output. Equilibrium implies that the labor market clears.

A.3.1 Production and Market Structure

There is a measure one of workers who differ in their observable cognitive (c) and noncognitive (n) abilities and we let $G(c,n)$ denote the measure describing the distribution of worker types in the economy. A worker's utility from being matched with a specific firm depends on the wage they receive from that firm plus an idiosyncratic preference shock. For worker i of type (c,n), the utility of working at firm j with wage $w_j(c,n)$ is

$$u_{ij}(c,n) = \beta \log(w_j(c,n)) + \nu_{ij} \tag{A.5}$$

where ν_{ij} captures an idiosyncratic preference for working at firm j. We assume that shocks ν_{ij} are independent draws from a Type I Extreme Value distribution. This specification could be expanded adding firm-level variation of average amenities as in Sorkin (2018).

Given wages, workers choose the firms that give them the highest utility. Using standard arguments (McFadden, 1974), the share $q_j(c,n)$ of type (c,n) workers who choose firm j has a logit form

$$\log(q_j(c,n)) = \log(h(c,n)) + \beta \log(w_j(c,n)). \tag{A.6}$$

Equation (A.6) delivers the upward sloping labor supply equation faced by firm j, with elasticity of supply β. The intercept $h(c,n)$ is determined in equilibrium and guarantees market clearing (every worker gets a job), that is

$$h(c,n) = \left[\int w_k(c,n)^{\beta}\ \mathrm{d}F(k)\right]^{-1} \tag{A.7}$$

where $F(\cdot)$ is the probability measure describing the distribution of firms in the economy.

As in Lise and Robin (2017), the production function is defined at the level of the match and we do not model complementarity between workers within a firm. A worker of type (c,n) employed at firm j produces according to $f_j(c,n)$, where the function f_j describes the output from the firm-worker match. Technology is CRS and a firm's output is the sum of all employees' products.[8] Firm j's total output is

$$y_j = \int f_j(c,n) q_j(c,n) \, \mathrm{d}G(c,n). \tag{A.8}$$

In the output market, firms face a downward sloping demand curve for their products. Firm j's inverse demand is

$$\log(p_j) = \log(\varphi_j) - \frac{1}{\sigma}\log(y_j) \tag{A.9}$$

where p_j is product price, y_j is output, φ_j is a firm-specific (inverse) demand intercept, and σ is the output demand elasticity with respect to price.

The firm's problem. Given output demand and labor supply curves, a firm decides how many workers to hire for each skill type. Firm j's profit maximization problem is:

$$\begin{aligned} \max_{q_j(c,n)} \quad & p_j y_j - \int w_j(c,n) q_j(c,n) \, \mathrm{d}G(c,n) \\ s.t. \quad & y_j = \int f_j(c,n) q_j(c,n) \, \mathrm{d}G(c,n) \\ & \log(p_j) = \log(\varphi_j) - \tfrac{1}{\sigma}\log(y_j) \\ & \log(q_j(c,n)) = \log(h(c,n)) + \beta \log(w_j(c,n)) \end{aligned} \tag{A.10}$$

This problem has a closed form solution, with equilibrium wages in firm j

$$w_j(c,n) = \frac{\left(\frac{\beta}{1+\beta}\right)^{\frac{\sigma}{\sigma+\beta}} f_j(c,n) \left(\frac{\sigma-1}{\sigma}\varphi_j\right)^{\frac{\sigma}{\sigma+\beta}}}{\left[\int f_j(c,n)^{1+\beta} h(c,n) \, \mathrm{d}G(c,n)\right]^{\frac{1}{\sigma+\beta}}} \tag{A.11}$$

A.3.2 Base Pay and Skill Premia: Mapping Model to Firm Wages

Firms' production choices can be characterized along the two input dimensions (cognitive and noncognitive). Every worker has a type within the set (c,n), with the first letter denoting cognitive level and the second noncognitive level. The wage premium associated to skill bundle (c,n) in firm j is

$$e^{\Delta_j(c,n)} = \frac{f_j(c,n)}{f_j(L,l)} \tag{A.12}$$

for all (c,n). This corresponds to the wage relative to the low-type worker (L,l), $\left(\frac{w_j(c,n)}{w_j(L,l)}\right)$, since everything else in the wage equation (A.11) cancels. The premium $e^{\Delta_j(c,n)}$ is proportional to the

[8] Additive separability is often assumed in matching models with one-to-many sorting. In the empirical section we show how this technology specification delivers an accurate approximation of returns to different skill types. While convenient, the separability assumption is not crucial for our findings about sorting and returns heterogeneity.

(measurable) productivity of a (c,n) worker in firm j relative to a baseline worker of type (L,l). The parameter $\Delta_j(c,n)$ subsumes two sources of variation: (1) the skill endowment bundle (c,n), and (2) the return to that bundle in firm j. By definition, $\Delta_j(L,l) = 0$ and one can redefine baseline match productivity in firm j as $T_j = f_j(L,l)$, which is the output of workers of type (L,l). Using T_j and $\Delta_j(c,n)$, we write the output of firm j as $y_j = T_j \sum_{(c,n)} e^{\Delta_j(c,n)} q_j(c,n) \mathrm{d}G(c,n)$, where $\mathrm{d}G(c,n)$ with some abuse of notation denotes the total number of (c,n) type workers, and recast the profit maximization as a choice over a discrete set of skill bundles (c,n).

Optimal hiring behavior in the discrete maximization problem implies:

$$w_j(c,n) = \underbrace{\frac{\beta}{1+\beta}}_{\text{Monops.Markdown}} \times \underbrace{\frac{\sigma-1}{\sigma} \varphi_j T_j \left(\frac{1}{y_j}\right)^{\frac{1}{\sigma}}}_{\text{Marg.Revenue}} \times \underbrace{e^{\Delta_j(c,n)}}_{\text{Skill Productivity}} \tag{A.13}$$

This expression captures different aspects of market structure. The marginal revenue is an increasing function of the firm's output demand φ_j. However, the monopsonistic firm sets wages at a fraction $\frac{\beta}{1+\beta}$ of the marginal revenue generated by the worker, with the fraction approaching one in more competitive markets where the labor supply elasticity β is larger. An extra unit of skill rescales marginal revenues proportionally to the firm's skill return $\Delta_j(c,n)$.

In log form, the equilibrium wage lends theoretical underpinning the empirical specifications in the paper. That is:

$$\log(w_j(c,n)) = \alpha + \Lambda_j + \Delta_j(c,n). \tag{A.14}$$

The intercept $\alpha \equiv \log\left(\frac{\beta}{1+\beta}\frac{\sigma-1}{\sigma}\right)$ is common across firms and skills, while $\Lambda_j \equiv \log\left(\varphi_j T_j y_j^{-\frac{1}{\sigma}}\right)$ is the firm-specific baseline wage, which does not vary with worker skills; $\Delta_j(c,n)$ is a *firm-specific return to skill bundle* (c,n). Under the model's null hypothesis, the firm's demand intercept φ_j is subsumed in the fixed effect component Λ_j.

Optimal behavior implies that firms with higher returns to (c,n)-type skills tend to hire a larger share of (c,n)-type workers. This observation suggests that firms with similar returns to a skill type can be grouped together based on their share of workers with that particular type.

A.4 Additional Estimation Results

A.4.1 Results from Alternative Samples and Estimation Approaches

Columns (1) and (2) in Table A.3 show the standard deviations of individual firm effects when we do not apply the quadratic-form correction (plug-in values) or when the sampling entails leaving single observations (worker–year) out, rather than the whole worker–firm spell as we do in Table 3.1 of the paper. As expected, when comparing to the baseline results, both these alternative specifications result in more pronounced firm heterogeneity. In this sense, our baseline estimates provide a conservative view of firm return variation. Details about the different sampling approaches (e.g. leaving out one worker-firm observation rather than the whole match) are discussed in the Appendix Section A.1.2. Colums (3) and (4) in Table A.3 show results when we re-estimate the model using

Table A.3: Standard deviations of firm parameters in alternative estimations.

	Firm-level (1999–2008):		Grouped (alt. periods):	
	Plug-in	Leave-obs-out	1990–1999	2008–2017
	(1)	(2)	(3)	(4)
sd(λ_j^0)	0.32	0.22	0.10	0.09
sd(λ_j^c)	0.40	0.21	0.10	0.07
$\times$ $90^{th} - 10^{th}$ *pct, cog score (c)*	*0.30*	*0.15*	*0.07*	*0.06*
sd(λ_j^n)	0.39	0.17	0.05	0.05
$\times$ $90^{th} - 10^{th}$ *pct, noncog score (n)*	*0.30*	*0.13*	*0.04*	*0.04*
$\times$ $90^{th} - 10^{th}$ *pct, cumulative (c+n)*	*0.59*	*0.28*	*0.11*	*0.09*
# unique firms	19,085	19,085	20,484	22,079

Notes: The table shows standard deviations of parameters λ_j^0, λ_j^c, and λ_j^n estimating (3.10) in alternative specifications and periods. Column (1) are plug-in estimates at the firm-level without quadratic-form correction. Column (2) quadratic-form corrects the firm-level variances leaving one observation (i.e., worker in a given year) rather than match (i.e., worker–firm spell) out at a time. Estimation period: 1999–2008. Columns (3) and (4) show the firm-clustered estimates in alternative periods 1990–1999 and 2008–2017. Otherwise notes to Table 3.1 apply.

the clustering approach for alternative sample periods. Dispersion of firm returns is slightly higher in 1990–1999 than in the baseline estimation period in Table 3.1. It is slightly lower in 2008–2017, alongside a lower standard deviation of firm intercepts. Overall, the dispersion of firm parameters appears remarkably stable over time.

A.4.2 Variance Accounting

To facilitate comparisons of our findings to existing work, it is useful to characterize the contribution of different layers of firm heterogeneity to the overall variance of earnings. We perform this exercise for either of the two estimation approaches (bias-correction and clustering); we also report similar decompositions for standard AKM estimators that do not explicitly account for firm-level heterogeneity in skill returns. Finally, to illustrate robustness of results under each approach, we carry out the analysis for the full sample (where each worker–firm match is observed for possibly multiple periods) and for the collapsed match-level samples (where each worker–firm match represents the average value over possibly several periods in which worker and employer are jointly observed).

Table A.4: Variance decomposition of log earnings (shares $\times$ 100). Clustered firms approach with one hundred classes.

	$\frac{\text{var}(\alpha_{it})}{\text{var}(log(w_{ij}))}$ (1)	$\frac{\text{var}(\psi_{ij})}{\text{var}(log(w_{ij}))}$ (2)	$\frac{2\text{cov}(\alpha_{it},\psi_{ij})}{\text{var}(log(w_{ij}))}$ (3)		
Full sample				Obs. (million)	6.48
Full model	60.8	3.8	11.2	total	75.8
AKM	61.0	3.7	11.0	total	75.7
Match-level collapsed sample				Obs. (million)	1.19
Full model	62.0	4.5	14.0	total	80.5
AKM	62.1	4.4	13.9	total	80.4

Notes: Decomposition of the percentage in log earnings variance explained based on estimates from specification (3.10). We subsume worker-only contributions in $\alpha_{it} \equiv \mu_i + X_{it}b_t$ and firm/worker contributions in $\psi_{ij} \equiv \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i$. We group firms into 100 clusters following the clustering approach as described in the text. AKM is alternative without heterogeneous firm returns; here it means fixed effects for the firm groups, not the individual firms. Estimation period: 1999–2008.

To control for variation due to worker-only components, we define $\alpha_{it} \equiv \mu_i + X_{it}b_t$. This is consistent with the normalization adopted in Section 3.3.1 of the paper, where μ_i contains average returns $\kappa_c \cdot c_i + \kappa_n \cdot n_i$ across firms, and α_{it} accounts for both observed and unobserved worker-level variation. The firm-related component (possibly capturing interactions with worker skills) is defined as $\psi_{ij} \equiv \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i$. In a standard AKM specification this latter component reduces to firm fixed effects.

Table A.4 shows the variance accounting exercise when we estimate the baseline equation

Table A.5: Variance decomposition of log earnings (shares × 100). Variance correction approach individual firm estimates (bias-corrected).

	$\frac{\text{var}(\alpha_i)}{\text{var}(log(w_{ij}))}$ (1)	$\frac{\text{var}(\psi_{ij})}{\text{var}(log(w_{ij}))}$ (2)	$\frac{2\text{cov}(\alpha_i,\psi_{ij})}{\text{var}(log(w_{ij}))}$ (3)		
Leave-one-out sample				Obs. (million)	3.27
Full model	49.4	8.4	5.8	total	63.6
AKM	42.7	8.0	5.3	total	55.9
Match-level collapsed sample				Obs. (million)	1.19
Full model	47.8	7.1	11.0	total	65.9
AKM	42.9	7.6	8.2	total	58.8

Notes: Decomposition of the percentage in log earnings variance explained based on estimates from specification (3.10). We capture worker-only contributions in $\alpha_i \equiv \mu_i$ and firm/worker contributions in $\psi_{ij} \equiv \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i$. Estimation period: 1999–2008.

(3.10) using 100 firm clusters for the 1999–2008 period. Results are similar for the 1990–1999 and 2008–2017 estimation periods, and comparable to grouping-based implementations for Sweden (Bonhomme et al., 2019) and the U.S. (Lamadon et al., 2022). As often found, worker-level heterogeneity accounts for much of the total earnings variation while the covariance between α and ψ is the second largest contributor to total variation (consistent also with Bonhomme et al., 2020, who study several countries). Perhaps most interesting is the observation that one would obtain similar results when restricting the specification to a standard AKM with no skill interactions. This suggests that the economically significant heterogeneity in returns would be mistakenly attributed to employer and worker fixed effects. This is especially concerning when interpreting employer fixed effects as skill independent earnings shifts.

Table A.5 shows the variance accounting exercise when the coefficients in (3.10) are estimated using the bias-correction approach. Since this approach adjusts the quadratic forms for worker-only variation downward, the contribution from worker fixed effects is somewhat lower (relative to the clustering approach) although it remains the largest by far. Consistent with the estimates reported in Table 3.1 of the paper, the direct impact of firm heterogeneity on total variation is also larger but remains a smaller share of the total. Due to the downward rescaling of the quadratic forms, total explained variation is lower than for the clustered estimation. The comparison to the restricted AKM specification confirms that return heterogeneity is mistakenly conflated into separate employer and worker fixed effects. This concern becomes even more relevant when we observe that, under the full unrestricted model, the covariation of worker and firm effects also become larger.

A.4.3 Implications for Matching of Workers with Firms

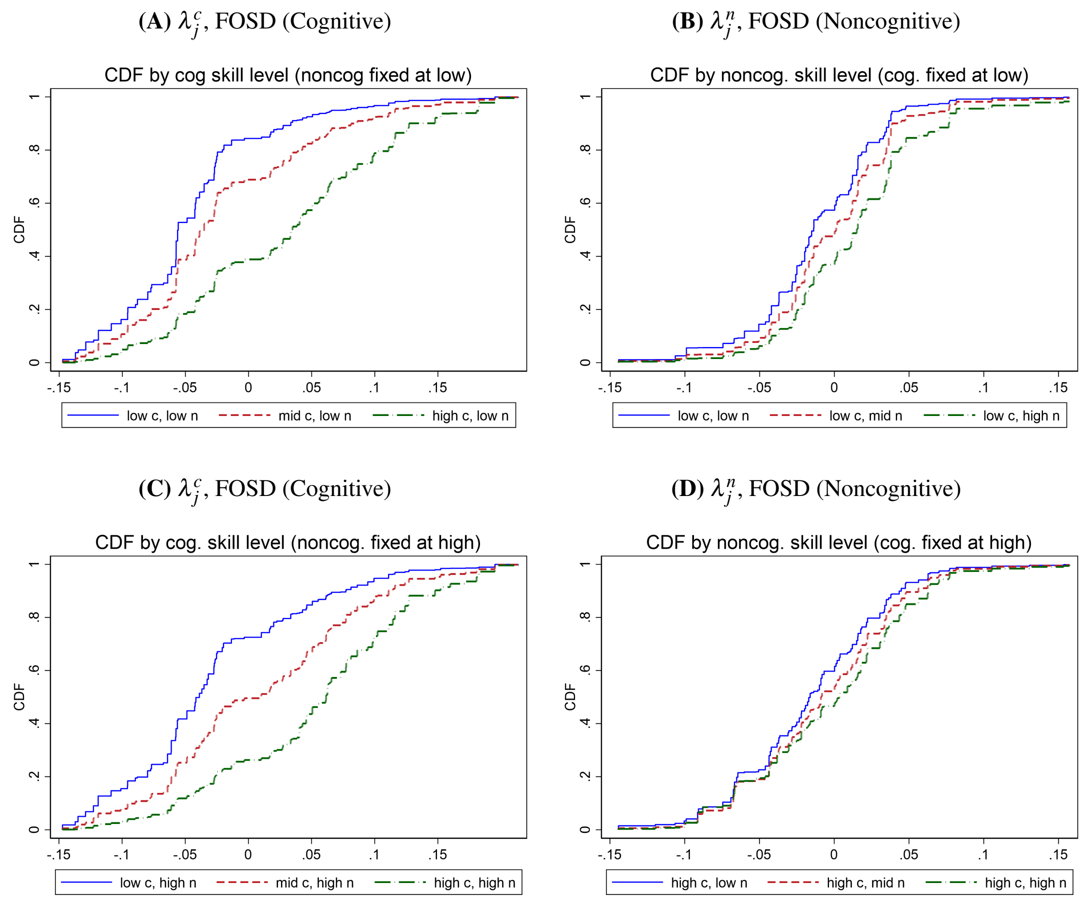


Figure A.2: Distribution of firm returns for different sets of worker skills.

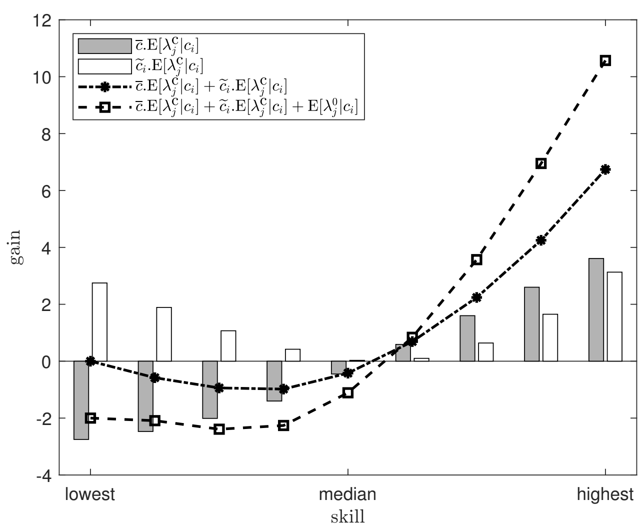
Notes: The figure shows cumulative distribution functions for workers with low ($c, n \leq 0.25$), mid ($0.25 < c, n < 0.75$), or high ($c, n \geq 0.75$) skill ranks over the range of firm returns. Period: 1999–2008. Results from the grouped estimator. FOSD: first-order stochastic dominance.

A.4.4 The Uneven Gains from Sorting

Robustness to rescaling of skills and returns. The sorting gains discussed in the main text are robust to rescaling of skills and returns since (i) multiplication of c_i by a non-zero factor would lead to a proportional change in the λ_j^c estimates as these would be scaled down by the same factor, leaving the product $\lambda_j^c c_i$ unchanged. (ii) Shifting the level of skills, by adding a constant x to c_i, leaves λ_j^c unchanged and shifts firm intercepts to $\lambda_j^0 - \lambda_j^c x$. Returns from working in firm j become $\lambda_j^c(c_i + x)$ but this is offset by $\lambda_j^0 - \lambda_j^c x$.

The total sorting gain, corresponding to the sum of both intercepts and returns ($\lambda_j^0 + \lambda_j^c c_i$), is hence fully invariant. This cumulative effect, calculated as the sum of columns (2) and (6) in Table 3.5, induces even larger inequality and skewness across the range of skill levels. Match effects are completely unaffected by rescaling, since they are defined relative to the demeaned $\widetilde{c}_i$.

Figure A.3: Gains from sorting across returns λ_j^c for different cognitive skill levels.



Notes: Gains are multiplied by 100 (i.e., in log points) for readability. All returns are differences relative to a scenario with no heterogeneity in firm returns. Estimates are based on the grouping approach with detailed numbers in Table 3.5. Sample period: 1999–2008.

The dashed line in Figure A.3 shows the total sorting gain in the cognitive dimension, that is $\mathrm{E}(\lambda_j^0 \mid c_i) + c_i \cdot \mathrm{E}(\lambda_j^c \mid c_i)$. This induces even wider earning differences between skill levels and retains the strong convexity. The average effect, i.e., the aggregate gain from matching, is exactly the same as for the thick dotted line $c_i \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ already seen in the main text.

Gains from sorting on noncognitive returns. Table A.6 reports the effects from the sorting of noncognitive attributes n_i across noncognitive returns λ_j^{n}. These effects are comparatively smaller than in the cognitive dimension, which reflects the lower dispersion of noncognitive returns across firms (see Section 3.3) and the weaker sorting in that dimension (see Section 3.4). Nonetheless, there is clear evidence of sorting also in the noncognitive dimension.

Table A.6: Gains from sorting across returns λ_j^{n} for different noncognitive skill levels.

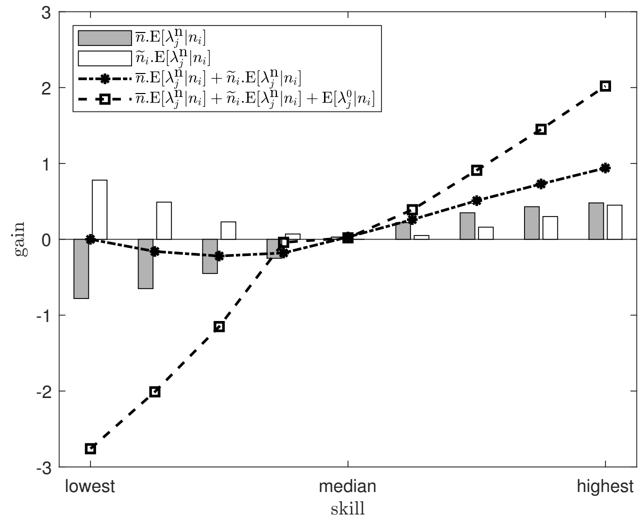
	$\mathrm{E}(\lambda_j^{\mathrm{n}} \mid n_i)$ (1)	Full gain (2)	Returns effect (3)	Match effect (4)	$\mathrm{E}(\lambda_j^{0} \mid n_i)$ (5)
skill level (n_i):					
1 (lowest, $n_i = 0$)	*-1.52*	0.00	-0.78	0.78	-2.76
2	*-1.26*	-0.16	-0.65	0.49	-2.01
3	*-0.87*	-0.22	-0.45	0.23	-1.15
4	*-0.48*	-0.18	-0.25	0.07	-0.04
5 (median, $n_i = 0.5$)	*0.06*	0.03	0.03	0.00	0.02
6	*0.42*	0.26	0.22	0.05	0.39
7	*0.68*	0.51	0.35	0.16	0.91
8	*0.84*	0.73	0.43	0.30	1.45
9 (highest, $n_i = 1$)	*0.94*	0.94	0.48	0.45	2.02
Aggregate	*0.00*	0.13	0.00	0.13	0.00

Notes: Gains are multiplied by 100 (i.e., in log points) for readability. All returns are differences relative to a scenario with no heterogeneity in firm returns. Estimates are based on the grouping approach. Sample period: 1999–2008. Column (1): expected marginal return conditional on skill. Column (2): total gain from sorting. Column (3): gain from sorting for the average-skill worker. Column (4): gain from sorting in excess of an average-skill worker with the same employer. Column (5): gain from sorting into intercepts.

Column (1) in Table A.6 shows that workers with higher noncognitive endowments sample from a distribution of employers with higher returns. Moving from $n_i = 0$ to $n_i = 1$ there is a 2.5 log points difference in $\mathrm{E}(\lambda_j^{\mathrm{n}} \mid n_i)$. This again leads to non-monotonic gains, since high-skill workers benefit the most from the sorting whereas the lowest-skill workers would benefit (or lose) little from any skill returns. The workers who experience steep losses are those with intermediate skills since they would gain from matching with high return firms but are not assigned to such firms. Match effects in column (4) reflect the complementarity of high-skill workers with high-return firms, and of low-skill workers with low-return firms, as well as the induced sorting. These are again positive and raise aggregate earnings by 0.13 log points. In the last column of Table A.6, inequality is further increased by the sorting of noncognitive attributes n_i over λ_j^{0} intercepts.

Figure A.4 represents these effects visually. The earnings differences between skill levels are clearly convexified by the sorting (thick dotted line), albeit the convexification is not as pronounced as for cognitive traits. Interestingly, sorting over intercepts reverses this convexification (dashed

Figure A.4: Gains from sorting across returns λ_j^{n} for different noncognitive skill levels.



Notes: Gains are multiplied by 100 (i.e., in log points) for readability. All returns are differences relative to a scenario with no heterogeneity in firm returns. Estimates are based on the grouping approach with detailed numbers in Table A.6. Sample period: 1999–2008.

line), since the least skilled workers face particularly low λ_j^0 (see column (5) of Table A.6). As we emphasized in the main body, and as we see here, the purely redistributive fixed effects (due to sorting into firm intercepts with no complementarity) do not in general induce skewness of the earnings distribution.

A.5 Extensions and Robustness

A.5.1 Industries and Occupations

Using occupation and industry identifiers we can assess whether return heterogeneity is genuinely firm-specific. To this purpose we add industry and occupation interactions with cognitive and noncognitive skills to the specification (3.10). That is, $X_{it}b_t$ now contains $\lambda_o^c \cdot c + \lambda_o^n \cdot n$ as additional controls where each o indexes one industry or occupation cell.

Table A.7 reports the results, with the first column referring to the baseline specification from the main text for comparison. In column (2) we add industry-specific cognitive and noncognitive skill returns (for 19 different sectors). The contributions of firm intercepts and of returns heterogeneity to earnings dispersion decline very slightly – from 0.10 to 0.09 for sd(λ_j^0) and from 0.06 to 0.05 for sd($\lambda_j^c c_i + \lambda_j^n n_i$). The overall effects remain similar. Column (3) adds detailed five-digit industries, with up to 586 separate returns for each skill dimension; also in this case, the contributions of firm-level parameters to overall dispersion remain stable.

Occupation information is only available in the LISA data from 2001 onward (and only partially before then) so that the estimation sample shrinks. This can be seen, e.g., in the lower number of unique firms in the bottom row of Table A.7.

Introducing occupation-specific returns has more influence on the firm-level parameters. In column (4) of Table A.7 we allow for heterogeneous returns for eight major occupation groups (similar to those used in Acemoglu and Autor, 2011). In this specification the standard deviations of baseline cognitive and noncognitive returns, as well as their contributions to earnings dispersion, decline by about one third compared to the benchmark in column (1). This partly reflects variation in production arrangements within firms; to the extent this variation underpins firm-specific skill returns, it is natural to expect it to be captured by occupation-specific returns. Put differently, the firm-level occupation make-up is one of the primitives accounting for firm heterogeneity in skill returns and, therefore, is a legitimate component of the total firm return. Finer occupations in column (5) and even industry-sector × occupation-group interactions in column (6), which proxy for specific jobs in a firm, have little additional effect on the contribution of firm heterogeneity to earnings dispersion.[9]

Sorting patterns. Figures A.5 and A.6 show that the patterns of skill sorting across returns are effectively unchanged when we control for industry and occupation-specific interactions. The range

[9]While results would not be much different than the industry-sector × occupation-group specification, we refrain from explicitly reporting estimates of detailed industry × detailed occupation-specific returns estimates. The reason is that this has additionally more than 21 thousand nonmissing cell-specific returns (almost as many as there are firms) for each skill dimension and thus reinstates an incidental parameter bias problem that the group-level estimation shown here circumvents.

Table A.7: Dispersion of estimated effects under industry / occupation controls.

	Main (1)	Sector (2)	Industry (3)	Occup-Group (4)	Occupation (5)	Sec×OccGr (6)
sd(μ_i)	0.43	0.43	0.43	0.41	0.40	0.40
sd(λ_j^0)	0.10	0.09	0.09	0.10	0.09	0.09
sd(λ_j^c)	0.08	0.08	0.07	0.05	0.05	0.05
sd(λ_j^n)	0.05	0.05	0.04	0.04	0.04	0.04
sd($\lambda_j^c c_i$)	0.05	0.05	0.04	0.03	0.03	0.03
sd($\lambda_j^n n_i$)	0.03	0.03	0.03	0.03	0.02	0.02
sd($\lambda_j^c c_i + \lambda_j^n n_i$)	0.06	0.05	0.05	0.04	0.04	0.04
# unique firms	25,783	25,783	25,783	23,999	24,168	23,973

Notes: Parallel to Tables 3.1 and 3.4, this table shows standard deviations of worker and firm effects but controlling for industry- or occupation-specific skill returns in equation (3.10). Column (1) repeats our specification from the main text without such controls. Column (2) adds broad industry sector specific skill returns (19 unique values per skill dimension). Column (3) adds detailed industry specific skill returns (up to 586 unique values per skill dimension). Column (4) adds broad occupation group specific skill returns (8 values, these groups can be seen in Figure A.7). Column (5) adds detailed occupation specific skill returns (113 values). Column (6) adds industry-sector × occupation-group specific skill returns (152 values). Group-level estimates in period: 1999–2008.

of variation of firm-level returns is only slightly smaller, in line with the reduction of dispersion in Table A.7. Sorting across firms remains strong and remarkably robust over the skill range.

We conclude that firm-level differences are an important source of skill return heterogeneity. Accounting for industry and occupation heterogeneity provides further evidence of the large differences that persist at the firm level; these differences do not reflect purely sectoral or occupational variation. Rather, we find that even within the same narrow industries and occupations, skills command significantly different returns across employers.

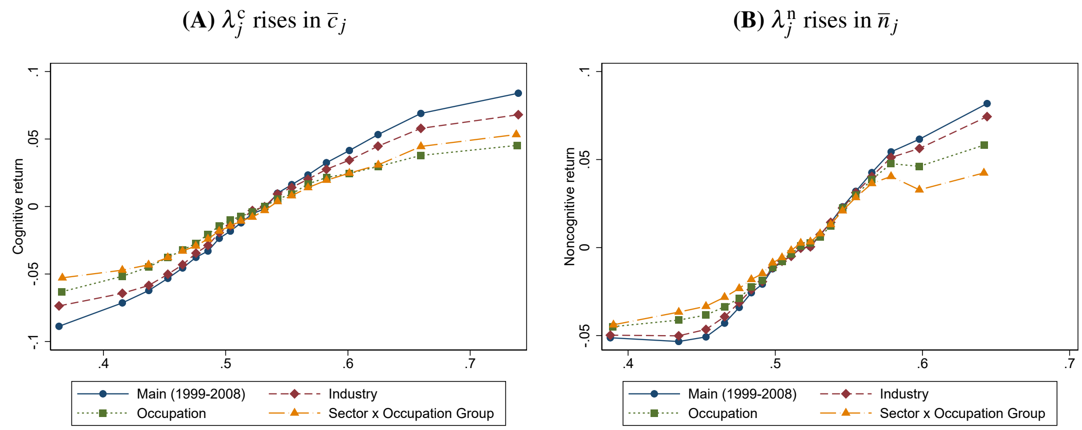
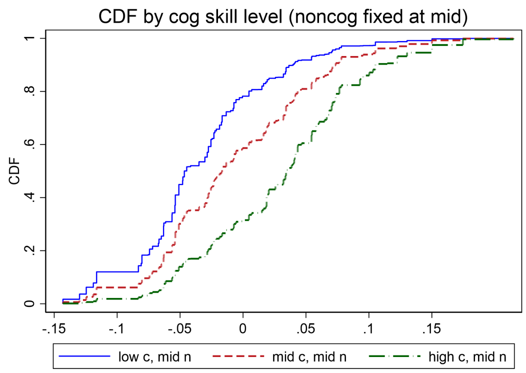


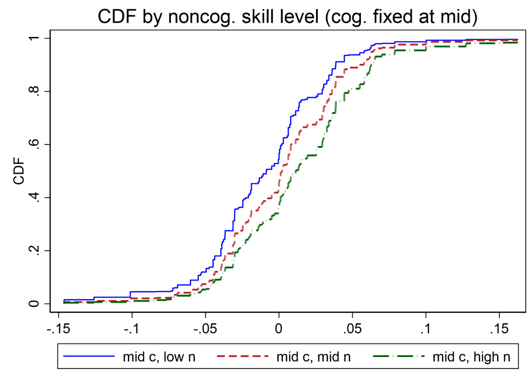
Figure A.5: Average skill by estimated return under different industry / occupation controls.

Notes: Parallel to Figure 3.3, the figure plots binned scatterplots of firm-specific skill returns (vertical axis) with average skills (horizontal axis) for the baseline specification shown in the main text; additionally controlling for detailed industry specific skill returns (up to 586 unique values per skill dimension) in equation (3.10); controlling for detailed occupation specific skill returns (113 values); and for industry-sector × occupation-group specific skill returns (152 values). Group-level estimates in period: 1999–2008.

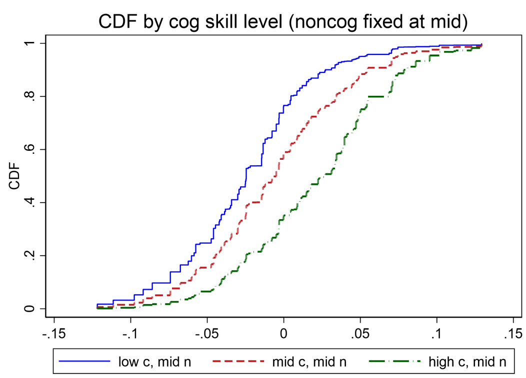
(A) λ_j^c, detailed industry



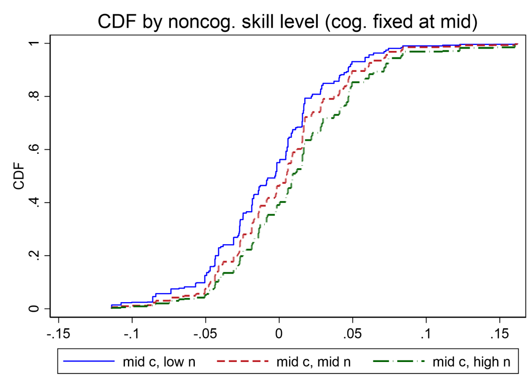
(B) λ_j^n, detailed industry



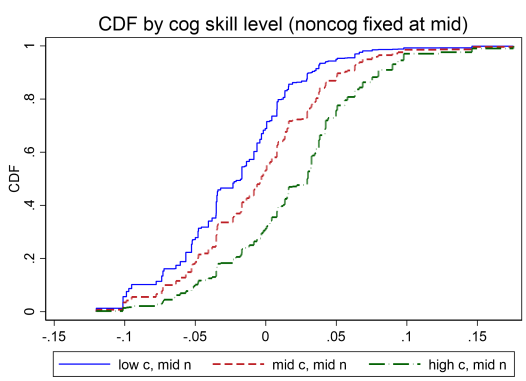
(C) λ_j^c, detailed occupation



(D) λ_j^n, detailed occupation



(E) λ_j^c, industry-sector $\times$ occupation-group



(F) λ_j^n, industry-sector $\times$ occupation-group

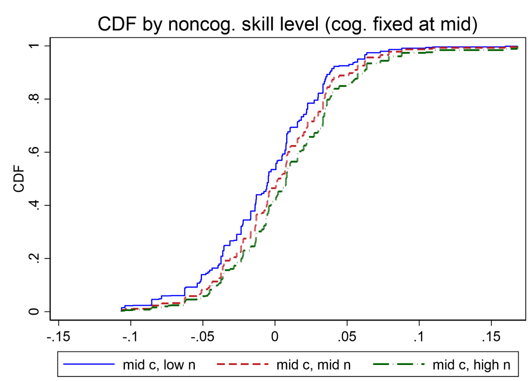


Figure A.6: FOSD sorting under different industry / occupation controls.

Notes: Parallel to main text Figure 3.3, this figure shows the cumulative distribution functions for workers with different skill ranks over the range of firm returns. The returns are now estimated controlling for industry- or occupation-specific skill returns in equation (3.10). The top row controls for detailed industry specific skill returns (up to 586 unique values per skill dimension); the middle row for detailed occupation specific skill returns (113 values); and the bottom row for industry-sector $\times$ occupation-group specific skill returns (152 values). FOSD: first-order stochastic dominance. Group-level estimates in period: 1999–2008.

Aggregating returns to the industry and occupation level. Whereas most of the heterogeneity occurs at the firm level, one may ask which industries or occupations exhibit higher skill returns on average. To answer this question, we first consider linear projections of baseline estimates of λ_j^c and λ_j^n on a full set of seven industry sector dummies. The projections are similar to those described in (3.12), where $\bar{c}_j$ and $\bar{n}_j$ are replaced by sector dummies, and yield the average cognitive and noncognitive return in the respective industry compared to the omitted "Manufacturing" sector.

Figure A.7A summarizes the results for the group-level estimates in the form of a coefficients plot. Cognitive returns are especially high in the business services and IT sector, noncognitive returns tend to be higher in wholesale and personal service related activities. By contrast, cognitive returns are rather low in the omitted manufacturing sector itself (represented by the zero line) and in utilities, transport, and services. Noncognitive returns in addition are remarkably low in business services and IT.

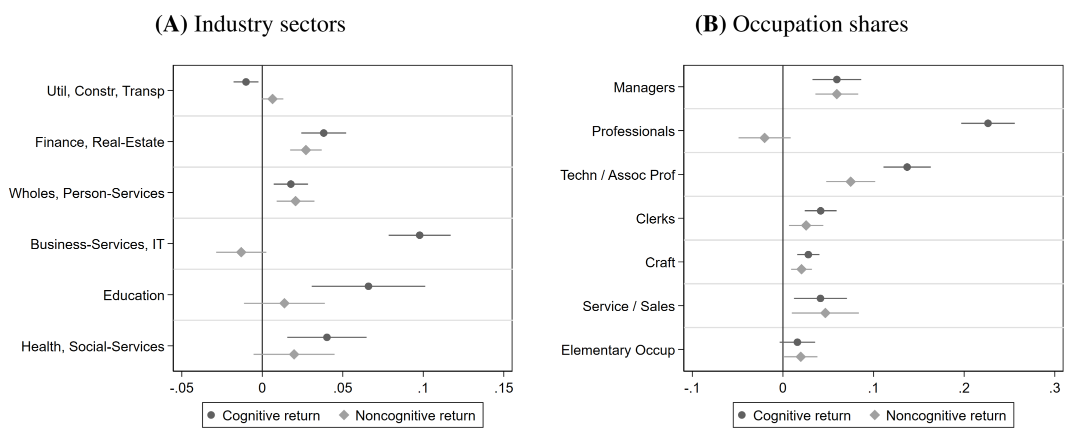


Figure A.7: Skill returns by industry and occupation composition.

Notes. Panel (a): coefficients from the projection of skill returns λ_j^c and λ_j^n onto seven broad industry-sector dummies. Sector dummies add up to one and the omitted sector is "Manufacturing", i.e., coefficients indicate difference in average skill return compared to average in manufacturing ($\lambda_j^c = -0.029$ and $\lambda_j^n = -0.004$ in that sector). Panel (b): coefficients from the projection of λ_j^c and λ_j^n onto a full set of eight broad occupation employment shares in each firm. Occupation group shares sum to one and the omitted group is "Operators / Assemblers". Returns are estimated for 100 firm classes. 95% confidence intervals based on robust standard errors clustered at the level of firm classes.

Figure A.7B shows corresponding results from a linear projection of estimates of λ_j^c and λ_j^n onto an exhaustive set of employment shares for eight broad occupation groups. The baseline omitted occupation are "Operators / Assemblers", a large manufacturing-type occupation group. As in Table A.7, and likely because they can vary within firms, occupations are somewhat more related to cognitive and noncognitive returns. That is, firms with large shares of professional, technical, and clerical workers have significantly higher cognitive returns compared to operator/assembler workers. Firms with larger shares of managers, technical workers, and services/sales workers have both high cognitive and noncognitive returns. As for business services and the IT sector, noncognitive skill returns are low among firms with a high share of professional workers.[10]

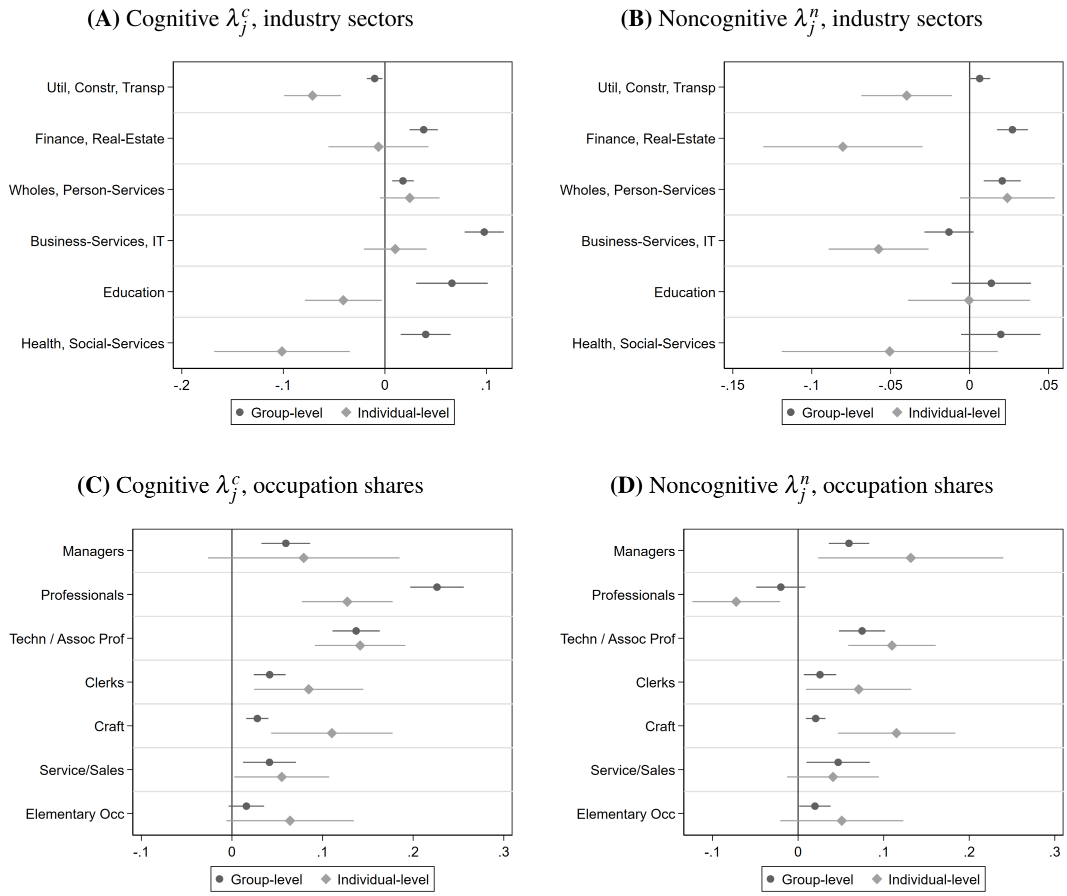


Figure A.8: Skill returns by industry and occupation, firm- versus group-level estimates.

Notes. See note to Figure A.7. Here we additionally plot the projections of firm-level λ_j^c and λ_j^n estimates onto broad industry sector dummies and occupational employment shares, and then compare them to the projections of group-level estimates from Figure A.7 separately by skill dimension.

Finally, results are robust to alternatively considering the firm-level estimates of λ_j^c and λ_j^n from the (smaller) leave-match-out sample. This is shown in Figure A.8, next to the group-level estimates. This approach is less precise and has wider confidence intervals but remains broadly consistent with the group level projections. These exercises suggest that firms in certain industries, and with certain occupations, differentially reward particular skills. Yet, while such variation exists,

[10]These low noncognitive returns are consistent with the cross-sorting we found in Section 3.4 if the very high cognitive returns attract very cognitively able professionals to those firms. The professionals also have high noncognitive skills but they accept the low noncognitive returns in the ”business / professional services” firms in exchange for the exceptional returns on their cognitives.

skill returns (even conditional on, say, a given occupation) vary substantially across firms. In fact, occupation composition can itself be an outcome partly driven by return differences across firms.

A.5.2 Capital Composition, Innovation, and Skill Returns

Balance sheets and capital components. We use a commercial data product, the "Serrano" database provided by Bisnode AB, that collects and cleans information about each firm's financials. Up to now, and consistent with prior work, we have referred to workplaces as "firms". However, for the balance sheet analysis we aggregate workplaces up to the organization level (a broader notion of "firms") for which both financial accounts and innovation activity are reported. Since there are multi-workplace corporations, this reduces the number of observations by about one third (see, e.g., Table 3.7).

Table A.8: Projection of Group Returns onto Firm Capital Composition.

	Dependent variable: λ_j^c or λ_j^n $\times 100$					
	(1)	(2)	(3)	(4)	(5)	(6)
Tangible assets	-3.69		-1.06		0.14	
	(0.87)		(0.15)		(0.09)	
Buildings, Land, Machinery		-2.86		-0.80		0.07
		(0.81)		(0.21)		(0.10)
Other tangible assets		-2.40		0.25		0.12
		(0.35)		(0.10)		(0.05)
Intangible assets	2.95		0.97		0.03	
	(0.36)		(0.10)		(0.09)	
Patents, licences, capt. R&D		3.49		0.64		-0.24
		(0.44)		(0.12)		(0.10)
Goodwill and other intangibles		1.57		0.53		0.19
		(0.27)		(0.13)		(0.06)
Number of firms	14,339	14,339	5,496	862	14,339	14,339
Dependent variable	λ_j^c	λ_j^c	λ_j^c	λ_j^c	$\boldsymbol{\lambda_j^n}$	$\boldsymbol{\lambda_j^n}$
Independent variables as	**dummy**	**dummy**	**logs**	**logs**	arcsinh	arcsinh

Notes: Results from regressions of skill returns onto capital components per employee from firms' balance sheets. Observations (firms) are unweighted with no further control variables. Columns (1) and (2) use dummies for whether the firm reports a positive value of the respective capital item as opposed to zero. Columns (3) and (4) take logs of the items' values. Columns (5) and (6) use noncognitive instead of the cognitive return as dependent variable with independent variables in arcsinh as in the main text Table 3.7. Dependent variables λ_j^c, λ_j^n multiplied by 100. Grouped estimates in period: 1999–2008. Robust standard errors clustered at the level of the 100 firm groups.

Table A.8 shows the projections of skill returns estimated at the group level onto firm capital components per employee in various robustness specifications. First we employ alternatives to the arcsinh transformation of balance sheet items in the main text. Columns (1) and (2) use dummies, which take the value of one when a firm reports a positive value of the respective capital item as opposed to zero (missing values are still removed). We observe that tangible assets, and in particular physical capital, is significantly negatively associated with cognitive returns whereas intangible assets, and especially intellectual capital, is significantly positively related. As a flip-side of this "extensive margin", we also study the "intensive margin" where we use log transformations of the balance sheet items. As discussed, the number of non-missing observations now drops and especially so for the detailed distinctions within tangible and intangible assets in column (4). Nonetheless, qualitatively and statistically (as well as in terms of coefficient sizes) the results are comparable to those based on the arcsinh transformation in the main text.

Table A.9: Projection of Firm-Level Returns onto Firm Capital Composition.

	Dependent variable: $\lambda_j^c \times 100$ from firm-level estimates					
	(1)	(2)	(3)	(4)	(5)	(6)
Tangible assets	-0.45	-0.56	-0.88			
	(0.36)	(0.41)	(0.19)			
Buildings, Land, Machinery				-0.32	-0.45	-0.59
				(0.44)	(0.46)	(0.25)
Other tangible assets				-0.34	-0.38	-0.55
				(0.40)	(0.41)	(0.22)
Intangible assets	1.03	0.76	0.57			
	(0.32)	(0.33)	(0.16)			
Patents, licences, capt. R&D				1.33	1.03	0.75
				(0.46)	(0.47)	(0.21)
Goodwill and other intangibles				0.46	0.33	0.37
				(0.38)	(0.38)	(0.19)
Number of firms	10,258	10,258	10,258	10,258	10,258	10,258
Sector fixed effects	No	Yes	No	No	Yes	No
Employment weighted	No	No	Yes	No	No	Yes

Notes: Firm-level estimates of λ_j^c in period 1999–2008. Robust standard errors in parentheses. Other than that, see note to Table 3.7.

Table A.9 shows that the projection results onto capital components are remarkably robust even if we instead use the firm-level estimates of cognitive returns. Finally, we note that the relationships with noncognitive returns are weaker, as shown in columns (5) and (6) of Table A.8. If anything, patents, licences, and capitalized R&D appear slightly negatively related to noncognitive returns (goodwill and other intangibles positively). Overall, these results are consistent with firms exhibiting heterogeneous production arrangements, whereby capital and employment structure vary substantially and lead to different returns to skill attributes, with the stronger impacts holding in the cognitive skill dimension.

Innovation output. In what follows we conduct further tests on the data from the Community Innovation Survey (CIS). Similar to the preceding analyses, the first part of Table A.10 shows estimates from projecting cognitive skill returns onto both product and process innovations. As in the main text, we control for a quadratic in employment, since the probability of engaging in innovation rises with the firm's size. The results on product innovations remain strong, whether or not we use group-level (column 1) or firm-level (column 3) return estimates or we control for industry sector fixed effects (i.e., the 19 unique ones from Section A.5.1).

Table A.10: Projection of Skill Returns onto Firm Innovation Activities.

	Dependent variable: $\lambda_j^c \times 100$					
	(1)	(2)	(3)	(4)	(5)	(6)
Innovation output:						
Product innovation	3.73	2.89	7.77			
	(0.53)	(0.37)	(2.61)			
Process innovation	0.29	0.48	4.09			
	(0.30)	(0.25)	(2.73)			
Innovation spending:						
Intramural R&D				0.39	0.30	0.46
				(0.06)	(0.04)	(0.27)
Extramural R&D				-0.01	0.06	0.03
				(0.04)	(0.03)	(0.32)
Acquisition of machinery				-0.14	-0.02	0.05
				(0.04)	(0.03)	(0.27)
Other external knowledge				0.14	0.06	0.37
				(0.04)	(0.03)	(0.31)
Number of firms	4,138	4,138	3,344	3,857	3,857	3,123
Sector fixed effects	No	Yes	No	No	Yes	No
Estimates (level)	Group	Group	Firm	Group	Group	Firm

Notes: The first three columns report estimates from regressions of cognitive skill returns onto indicators for product and process innovations (as defined in the text and note to Figure 3.5), controlling for a quadratic in firm employment size. Column (1) uses group-level returns estimates, column (2) adds industry sector fixed effects, and column (3) uses firm-level returns estimates. The last three columns regress returns onto firms' innovation expenditure items, which are $\text{arcsinh}(x_j) = \log\left(x_j + \sqrt{x_j^2 + 1}\right)$ transformed. Otherwise specifications (4)–(6) are parallel to (1)–(3). Returns estimated in period 1999–2008. Robust standard errors in parentheses and clustered at the level of the 100 firm classes for the grouped estimates.

The relationship between skill returns and process innovations gets weaker when we condition on product innovations, and it is only borderline significant.

Specific innovation activities. Next, we examine firms' CIS-reported expenditures on specific types of innovation activities. This is, again, done by using the arcsinh transformation. Column (4) of Table A.10 shows that, consistent with the preceding findings, high cognitive returns firms spend significantly more on intramural (or in-house) research and development. They also spend somewhat more on purchasing external knowledge, and somewhat less on specific machinery. These findings are robust to adding industry sector fixed effects or using estimates of firm-level returns in columns (5) and (6).

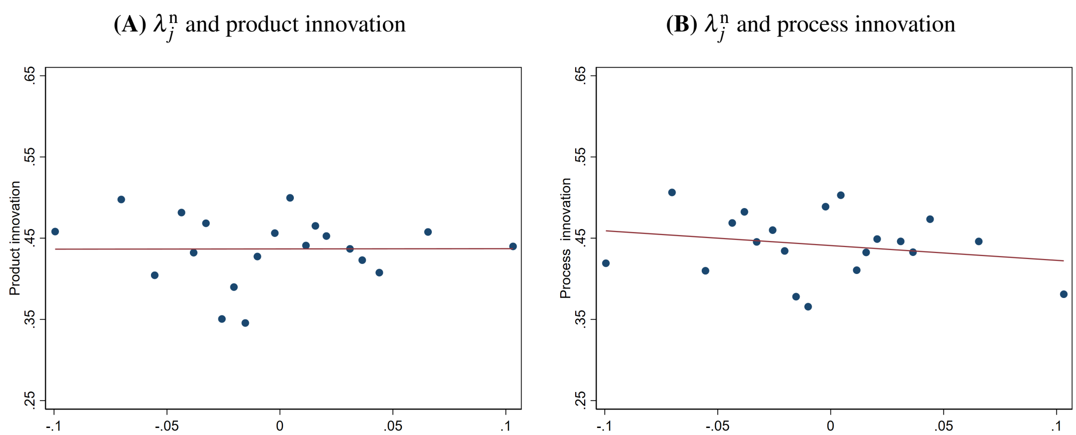


Figure A.9: Noncognitive skill returns and firm innovation.

Notes: The figure plots a binscatter of firms' innovation activities against noncognitive skill returns (group-level estimates during 1999–2008). Innovation activities are measured as indicators whether a firm has conducted any product (including service, Panel a) or process (including organizational, Panel b) innovations. This information is from various waves of a representative firm survey (European Community Innovation Survey, CIS). We average the responses (i.e., indicators) for the waves 1998–2000, 2002–2004, 2004–2006, 2006–2008, 2008–2010 relevant to our sample period. Underlying the plots are 4,138 unique firms. Regression slopes, controlling for a quadratic in firm employment, are $\beta = 0.00$ (clustered S.E. $= 0.32$) and $\beta = -0.18$ (clustered S.E. $= 0.19$) for product and process innovation, respectively.

Lastly, Figure A.9 shows the baseline binned-scatter plot of skill returns vis-a-vis product and process innovations for the noncognitives. Broadly in line with our prior findings, there is no detectable relationship and, in contrast to λ_j^{c}s, the λ_j^{n}s do not actually predict higher innovation activity.

A.5.3 Clustering Strategies and Number of Firm Clusters

In what follows, we document how the dispersion (standard deviations) of firm-level parameters, and their contributions to earnings dispersion, vary under alternative restrictions on the number of clusters as well as on the observables used for the clustering.

Table A.11 illustrates the key results. For comparison column (1) replicates the baseline specification in the main text. In column (2) we use only 10 (rather than the baseline 100) firm clusters; this number is the same as in the main analyses of Bonhomme et al. (2019), Lamadon et al. (2022). The contributions of firm intercepts and skill return heterogeneity to earnings dispersion marginally

Table A.11: Alternative clustering specifications.

	Main (1)	10 clusters only (2)	Adding variables (3)	Earning dist. only (4)
sd(μ_i)	0.43	0.44	0.43	0.43
sd(λ_j^0)	0.10	0.07	0.10	0.11
sd(λ_j^c)	0.08	0.06	0.07	0.06
sd(λ_j^n)	0.05	0.03	0.04	0.03
sd($\lambda_j^c c_i$)	0.05	0.04	0.04	0.04
sd($\lambda_j^n n_i$)	0.03	0.02	0.02	0.02
sd($\lambda_j^c c_i + \lambda_j^n n_i$)	0.06	0.04	0.05	0.05
# unique firms	25,783	25,783	25,711	25,783

Notes: Adding to the evidence in Tables 3.1 and 3.4, this table shows standard deviations of worker and firm effects under alternative clustering specifications. Column (1) repeats the baseline specification from the main text, for comparison. Column (2) shows estimates when using the means of earnings, cognitive and noncognitive skills within each firm but just ten clusters. Column (3) shows results for 100 clusters after adding standard deviations of earnings, cognitive and noncognitive skills and firm employment size as additional clustering variables. Column (4) shows results when we only use quantiles of the earnings distribution (10th, 30th, 50th, 70th, and 90th) within each firm for clustering and we impose 100 groups. All group-level estimates are based on the sample period: 1999–2008.

declines, while the *relative* contribution of returns rises. When we use a richer set of clustering variables – including firm employment size as well as the standard deviations of earnings and cognitive and noncognitive skills, in addition to the means of these variables – returns heterogeneity does not change significantly either in relative or absolute terms, as shown in column (3). Finally, column (4) shows that restricting the clustering strategy to earnings alone, through the use of quantiles of the earnings distribution within each firm (see Bonhomme et al., 2019, Lamadon et al., 2022), does not materially change the estimated impact of returns heterogeneity when compared to the other robustness checks. In all alternative specifications we also confirm the presence of positive assortative matching patterns.

In the main text, Figure 3.6 shows estimates of the impact of firm heterogeneity for a range alternative restrictions on the number of clusters (which are let to vary between 20 and 200). The relative contributions of intercepts and skill returns change only marginally, lending further support to the results obtained under the baseline cluster design.