Discussion Paper Series – CRC TR 224

Discussion Paper No. 555
Project B 03

Big Data and Inequality

Carl-Christian Groh[1]

June 2024

[1]University of Bonn, Email: carlchristian.groh@gmail.com

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Big Data and Inequality*

Carl-Christian Groh†

April 8, 2024

Abstract

This paper studies the distributional consequences of the increasing importance of (big) data in modern economies. I consider a simple theoretical model in which firms produce output using capital and labor. Firms can hire labor on the spot market, but must choose their capital stock for a given period in advance and under uncertainty regarding their future profitability. Access to data resolves this uncertainty, thereby primarily increasing the aggregate demand for and the remuneration of capital. Furthermore, the increased demand for capital crowds out labor demand by reducing the price of the output goods, which reduces aggregate labor income. By an analogous logic, the rising availability of data can also increase the skill premium, given that firms can adjust their unskilled labor input more easily than their skilled labor input.

Keywords: inequality, big data, uncertainty, wages
JEL Classification: D21, D81, E24, J24

*I would like to thank Benny Moldovanu, Oliver Pfäuti, and Laura Veldkamp for insightful comments. Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project B03) is gratefully acknowledged.

†Groh: Institute for Microeconomics, University of Bonn. Lennéstrasse 37, 53113 Bonn, Germany (email: carlchristian.groh@gmail.com)

1 Introduction

Modern economies increasingly revolve around (big) data, i.e., digitized information about consumers and market conditions. The share of firms that apply data-driven decision making is increasing rapidly (Galdon-Sanchez et al., 2022; Brynjolfsson et al., 2023) and the economic value of data is enormous (Abis & Veldkamp, 2023; Statista, 2024). While access to data serves many purposes, a central feature of data that distinguishes it from existing technologies is that it enables firms to predict future outcomes. Utilized as an input of artificial intelligence (AI) processes, data allows firms to accurately predict key components of their future profitability such as firm-level demand (Chong et al., 2016; Bajari et al., 2019; Fildes et al., 2022), their operating costs (Ajit, 2016; Choudhury et al., 2021), and their rivals' strategic decisions (Rani et al., 2023). In fact, around 75% of US manufacturing firms are utilizing data to implement such forms of predictive analytics (Brynjolfsson & McElheran, 2019). This distinguishing feature of data, together with its central role in modern economies, necessitates a tailored approach to understand the societal impacts of data.

In this paper, I study the distributional consequences of these developments: How is the economic value generated by data distributed among workers, capital owners, and entrepreneurs? The increasing importance of (big) data is a key technological trend in modern economies, the relevance of which will only increase further over time as advancements in AI unlock the full predictive power of data. Understanding how these developments impact the well-being of all members of society is hence of first-order importance. The relevance of this endeavour is reinforced by the strong upward trends in income inequality (OECD, 2011; UNDP, 2019; World Bank, 2021) and the capital share (Rognlie, 2016; Burdín et al., 2022) in OECD countries, which constitute important policy concerns (European Union, 2018).

I demonstrate that the increasing prevalence of data can be a powerful magnifier of inequality. I consider a simple theoretical model with the key feature that it takes firms longer to adjust their capital stock than their labor input, which is empirically well-established (Nalewaik & Pinto, 2015; Meier, 2020; Oh & Yoon, 2020). In the absence of data, any firm thus chooses its capital input for a given period under uncertainty regarding its profitability in that period, while this uncertainty is resolved when the firm decides how many workers to hire. By endowing firms with a forecast of their future profitability, access to data thus primarily raises the aggregate capital demand, which translates into a higher capital share and lower labor share. Moreover, the higher capital demand crowds out labor demand by reducing the price of output goods, so total labor income may be reduced by the availability of data. By similar arguments, the availability of data raises the skill premium because skilled labor is less adjustable than unskilled labor (Autor et al., 2006; Ghaly et al., 2017).

Formally, I consider a stylized two-period model in which a unit mass of firms produce output using capital and labor, which are inelastically supplied. Production only happens in the second period and the firms are heterogeneous in their second-period profitability. Crucially, capital accumulation is subject to a time-to-build friction, i.e. firms must choose the amount of capital they utilize in the second period one period in advance. By contrast, labor is hired on the spot market in the second period.

Access to data endows firms with a forecast of their future profitability when they choose their capital inputs. I formalize this by considering two different economies: An *economy without data* and an *economy with data*. In the economy without data, any firm is made aware of its second-period profitability only at the beginning of that period, i.e. before choosing its labor input, but after setting its capital demand. In the economy with data, any firm knows the realization of its idiosyncratic profitability ex ante, i.e. before choosing its capital input. Thus, firms in the economy with data condition their capital input choices on their future profitability, which firms in the economy without data cannot do. I abstract from competition between firms and assume that the underlying distribution of profitability is the same in both economies.

I study how the aggregate labor and capital demands differ across the two economies, which establishes how the prevalence of data affects aggregate labor (capital) income as well as the labor (capital) share.[1] The impact of data is governed by two key features of the economy, namely (1) the degree of substitutability between capital and labor in production and (2) the price elasticity of the demand for the output goods produced by firms.

Consider a benchmark in which capital and labor are perfect substitutes and the demand for the output goods is completely price inelastic. Then, the aggregate labor income is the same in both economies, i.e. workers receive none of the economic value generated by data. This holds by the following logic: If the input factors are perfectly substitutable and there are no price effects, the labor demand of any firm only depends on its profitability. Because the distribution of profitability is the same in the economy with and without data, the aggregate labor demand (and hence, the wage) is unaffected by the availability of data.

By contrast, the aggregate demand for capital is higher in the economy with data. This holds by the following logic: Firstly, any firm in the economy without data will demand the same amount of capital as a firm in the economy with data whose next-period profitability is equal to its expected value. Intuitively, this is because a firm without access to data can do no better than to set its capital demand to match its expected profitability. Secondly, the expected capital demand of a firm with data is higher than its capital demand at the

[1] Aggregate labor (capital) income is the product of the equilibrium wage rate (interest rate) and the supply of labor (capital). The labor (capital) share is the ratio of labor (capital) income to total income.

expected profitability, since a firm's optimal capital choice is convex in its profitability (given that capital has diminishing returns in production). Thus, the aggregate demand for capital is higher in the economy with data, which means that the labor (capital) share will be lower (higher) in the economy with data.

The difference in the labor shares across the two economies becomes smaller, the more complementary the two input factors are.[2] This is because superior access to data affects a firm's labor demand through its positive effect on the expected capital input. As the input factors become more complementary, an increase in a firm's capital demand will trigger a larger increase in the firm's labor demand. If capital and labor are strongly complementary, the increase in the aggregate capital demand triggered by the availability of data is thus accompanied by a substantial increase in the aggregate demand for labor (and thus, the wage). Hence, the impact of data availability on the labor share is mitigated.

When the demand for output goods is price elastic, the availability of data reduces total labor income through two further channels. First, any increase in the firms' expected capital demand triggered by access to data crowds out labor demand. This is because a firm that utilizes more capital will produce more, which reduces the price of the firm's output good and, by implication, its incentives to hire workers. Secondly, the availability of data enables firms to produce more efficiently, i.e. to require lower input amounts per unit of output. Holding the level of aggregate output fixed, the availability of data thus reduces the aggregate demand for both input factors. As the demand for output goods becomes more price elastic, the increase in aggregate output triggered by the availability of data is reduced. This implies that both the aggregate capital and labor demand may be lower in the economy with data. If this is the case, both the labor share and the capital share are lower in the economy with data and the value of data is captured by entrepreneurs.

Thereafter, I move beyond the analysis of capital/labor inequality and study how the availability of data affects the skill premium, which is defined as the ratio of the wages of skilled and unskilled workers. I set up a model in which a unit mass of firms with heterogeneous profitabilities produce output in two periods using unskilled and skilled labor. While unskilled labor can be freely chosen in both periods, a firm faces adjustment costs for skilled labor in the second period — this is in line with the empirical evidence which suggests that skilled labor is more difficult to adjust because of significant hiring and firing costs (Autor et al., 2006; Ghaly et al., 2017). Thus, a firm's second-period profitability affects its optimal choice of skilled labor in the first period. As before, I consider an *economy without data* and an *economy with data*.

[2]Hubmer (2023) documents that capital and labor are gross substitutes, which confirms the empirical relevance of the channels I highlight.

The skill premium in the economy with data is higher if skilled and unskilled labor are sufficiently substitutable. This is because adjustment costs only pertain to skilled labor, which means that access to superior data primarily raises the demand for skilled labor. If a firm's demand for skilled labor is independent of its demand for unskilled labor (which is the case when these inputs are perfectly substitutable), the relative wage of skilled workers is thus higher if data is available. As unskilled and skilled labor become more complementary, the strength of this effect becomes weaker, since any increase in the aggregate demand for skilled labor is accompanied by an increase in the aggregate demand for unskilled labor.

Related Literature: To the best of my knowledge, this is the first paper which studies how the availability of data (modelled as signals about firms' future profitabilities) affects the prevailing level of economic inequality.

My work is most closely related to the rapidly growing literature that studies how digitization shapes macroeconomic outcomes. Veldkamp & Chung (2019) provide an overview of the role of data in the economy. Eeckhout & Veldkamp (2022) show that data can be a source of market power because access to superior data reduces a firm's risk, thereby incentivizing it to reduce its marginal costs and attain scale. Farboodi & Veldkamp (2022) study growth in the data economy, devoting explicit attention to the data accumulation process. Acemoglu et al. (2022) show that data markets are not efficient in the presence of data externalities, i.e. when an individual's data reveals information about others. Bergemann & Bonatti (2022) study how access to data grants platforms market power. Groh & Pfäuti (2023) integrate data into a simple macroeconomic model of business cycles and monetary policy.[3] None of these papers study the effect of data on inequality as measured by the skill premium or the capital/labor shares.

Within this literature, the papers that are most closely related to mine are Arvai & Mann (2022), Bughin (2023), Babina et al. (2023), and Abis & Veldkamp (2023), given that they study topics related to inequality. Arvai & Mann (2022) show that households with greater income consume more digitally manufactured goods and that such goods feature lower price inflation. Bughin (2023) studies how AI affects employment through automation and by boosting innovation. Babina et al. (2023) empirically analyse the relationship between AI adoption and the within-firm workforce composition.[4] The authors document that firms which invest more into AI development also feature an increase in the share of skilled workers they employ. Abis & Veldkamp (2023) study how data and labor combine to generate

[3] Glocker & Piribauer (2021) empirically document that increases in the amount of sales that are conducted through digital retail reduce the real effects of monetary policy.

[4] Hubmer (2023) shows how technological progress and the evolution of preferences shape the labor share.

knowledge and establish that new developments in AI and big data reduce the labor share of income in the production of knowledge.[5] In contrast to my work, these papers do not consider the role of data as I define it, namely as signals that firms receive about their future profitabilities.

Moreover, my paper relates to the research on the role of uncertainty for firm behaviour. The seminal contribution of Bloom (2009) shows that increases of uncertainty reduce hiring and investment. Kumar et al. (2023) provide causal evidence that increases in perceived uncertainty lead firms to reduce their employment, capital stock, and sales. Recent contributions (Fischer et al., 2021; Theophilopoulou, 2022; Belianska, 2023) empirically document that increases in macroeconomic uncertainty go along with reductions in the skill premium.

The work on uncertainty differs from my paper because previous work in this field does not consider data as I define it, namely as signals about future profitabilities. Increases in macroeconomic uncertainty are always modelled as increases in the second moment of the aggregate profitability (or productivity) distribution. Crucially, endowing firms with access to data does not have the same effects as reducing macroeconomic uncertainty. For example, endowing firms with access to data will reduce the aggregate demand for capital and labor if the demand for the output goods is sufficiently price elastic, while the opposite holds true for reductions of macroeconomic uncertainty. Moreover, recent evidence suggests that reductions of uncertainty increase firm employment to a greater extent than the capital stock (Kumar et al., 2023). By contrast, the availability of data induces an increase of the aggregate capital demand, relative to the aggregate labor demand.

Finally, my work is related to the literature on rational inattention, which was pioneered by Sims (2003). This is because my framework can be viewed as a model of rational inattention, in which access to data changes a firm's cost of acquiring information about its future profitability. However, there are substantial differences in focus and setup: Generally speaking, papers in this literature establish how rational inattention can account for inertia in macroeconomic outcomes, which is not the focus of my work. In terms of setup, my work is most closely related to Charoenwong et al. (2022) and Gondhi (2023), who consider models in which firms receive signals about their idiosyncratic profitability draws. I am not aware of any paper which studies the relationship of rational inattention and inequality.

Outline: The rest of the paper proceeds as follows: In section 2, I present my theoretical model, which is analysed in section 3. In section 4, I present and solve the alternative framework concerning the skill premium. I conclude in section 5.

[5]Eden & Gaggl (2018) and Jaimovich et al. (2020) study how automation shapes macroeconomic outcomes. Bessen et al. (2022) argue that automatisation will increase the quality of labor demanded by firms.

2 Framework

I consider a two-period model with time periods indexed by $t \in \{1, 2\}$. A unit mass of heterogeneous firms indexed i produce output ($Y_{2,i}$) in period 2 using capital ($K_{2,i}$) and labor ($N_{2,i}$) according to the following technology:

$$Y_{2,i} = y(K_{2,i}, N_{2,i}), \tag{1}$$

where $y(K_{2,i}, N_{2,i})$ is some twice continuously differentiable function that is increasing in both arguments. The demand for firm i's good is expressed by the function $D(Q_i|A_{2,i})$, which gives the highest possible price at which the firm can sell Q_i units of its good:

$$D(Q_i|A_{2,i}) = A_{2,i} - \eta Q_i \tag{2}$$

The random variable $A_{2,i}$ reflects the idiosyncratic profitability of firm i. The parameter η captures the prevailing level of demand elasticity. When $\eta = 0$, demand is fully inelastic and the price any firm sets will just be equal to its profitability $A_{2,i}$. I impose that $A_{2,i} \sim F$, where F is a continuous distribution with support $[\underline{A}, \bar{A}]$.

There is no production in period 1 and any firm pays an interest rate r per unit of capital it uses and a wage w per unit of labor it employs. Thus, the total profits of any firm take the following form:

$$A_{2,i}y(K_{2,i}, N_{2,i}) - \eta\big(y(K_{2,i}, N_{2,i})\big)^2 - wN_{2,i} - rK_{2,i} \tag{3}$$

Every firm chooses its capital stock $K_{2,i}$ in period 1. By contrast, labor is hired on the spot market, i.e. any firm chooses its labor input $N_{2,i}$ in period 2. Capital and labor are inelastically supplied, where $\bar{K}$ is the exogenous supply of capital and $\bar{N}$ the supply of labor.

Data enables firms to predict their future profitabilities. I formalize this by considering two different economies, namely *an economy without data* and an *economy with data*. In the economy with data, every firm knows its realization of $A_{2,i}$ at the beginning of period 1. In the economy without data, any firm is just made aware of its profitability realization at the beginning of period 2. Thus, firms in the economy with data condition their capital input choices on their future profitability, which firms in the economy without data cannot do.

The capital demand of firms in the economy with data is thus expressed by a function $K^d(A_{2,i}, p)$, where I define $p = (w, r)$ as the vector of input prices. By contrast, the capital demand of any firm in the economy without data is the same and expressed by $K^{nd}(p)$. In both economies, firms choose their labor inputs after observing their profitabilities. Thus, the optimal labor input of firms in either economy takes the same form and is given by a function $N_2^*(A_{2,i}, K_{2,i}, p)$, where $K_{2,i}$ is the amount of capital used by the firm.

Key objects of analysis will be the aggregate demands for capital and labor in the two economies. I define the aggregate demand for capital and the aggregate demand for labor in the economy with data as $\bar{K}^d(p)$ and $\bar{N}^d(p)$, respectively. These are given by:

$$\bar{K}^d(p) = \int_{\underline{A}}^{\bar{A}} K^d(A_{2,i}, p) dF(A_{2,i}) \tag{4}$$

$$\bar{N}^d(p) = \int_{\underline{A}}^{\bar{A}} N_2^*(A_{2,i}, K^d(A_{2,i}, p), p) dF(A_{2,i}) \tag{5}$$

The aggregate demand for capital in the economy without data, which I refer to as $\bar{K}^{nd}(p)$, is equal to the function $K^{nd}(p)$ as defined previously. The aggregate demand for labor in the economy without data is expressed by the function $\bar{N}^{nd}(p)$, where:

$$\bar{N}^{nd}(p) = \int_{\underline{A}}^{\bar{A}} N_2^*(A_{2,i}, K^{nd}(p), p) dF(A_{2,i}) \tag{6}$$

The economy reaches an equilibrium if two conditions are satisfied: (1) Every firm i choses its capital and labor inputs to maximize its expected profits and (2) the interest rate r and the wage rate w clear the capital and labor markets, respectively.

As is standard, I define the aggregate labor income in an economy as the product of the equilibrium wage and the labor supply. The aggregate capital income is defined as the product of the equilibrium interest rate and the capital supply. The labor (capital) share is defined as the ratio of labor (capital) income and total income, i.e. the expected revenue of firms.

I refer to the model I have just outlined as the *capital/labor framework*. Throughout the following analysis, I omit the firm-level index i for expositional convenience.

3 Analysis

3.1 Price inelastic demand

In this section, I abstract from any price effects by assuming that the demand for the firm's output goods is perfectly price inelastic, i.e. that $\eta = 0$. This allows me to establish how the effect of data on the welfare of different groups in the economy is shaped by the degree of substitutability between capital and labor in production. The main message of this section is the following: The labor share will be lower in the economy with data (and hence, data exacerbates capital-labor inequality) if labor and capital are sufficiently substitutable.

I begin by providing analytical results for two benchmark production functions, namely (1) a perfect substitutes production function $y(K_2, N_2) = (K_2)^{\alpha_K} + (N_2)^{\alpha_N}$ with $\alpha_K \in (0, 1)$ and $\alpha_N \in (0, 1)$ as well as (2) a Cobb-Douglas production function $y(K_2, N_2) = (K_2)^{\alpha_K}(N_2)^{\alpha_N}$ with $\alpha_K + \alpha_N < 1$. The two production functions differ in the degree to which the input factors are complementary: In the first production function, the two inputs are perfect substitutes. By contrast, the inputs are complementary under the Cobb-Douglas technology. One can establish the following results:

Proposition 1 (Data & labor/capital shares: Input complementarity)
Consider the capital/labor framework and suppose that the demand for any firm's good is price inelastic (i.e. $\eta = 0$):

- *If $y(K_2, N_2) = (K_2)^{\alpha_K} + (N_2)^{\alpha_N}$, aggregate labor income is identical in the economies with and without data. Aggregate capital income and the capital share are higher in the economy with data. The profit share is higher in the economy with data if $\alpha_N > \alpha_K$.*
- *If $y(K_2, N_2) = (K_2)^{\alpha_K}(N_2)^{\alpha_N}$, the labor, capital, and profit shares are identical in the two different economies.*

To understand these results, suppose firstly that $y(K_2, N_2) = (K_2)^{\alpha_K} + (N_2)^{\alpha_N}$. Because the demand for the output good is price inelastic and the inputs are perfect substitutes, the optimal labor demand of any firm is independent of the firm's capital choice and pinned down by the following first-order condition:

$$A_2 \alpha_N \big(N_2^*(A_2, w)\big)^{\alpha_N - 1} - w = 0 \tag{7}$$

Both in the economy with and without data, any firm's labor demand thus only depends on its profitability A_2 and the wage rate. Since the underlying distribution of profitability is the same in both economies, the aggregate labor demand function is unaffected by the availability of data to firms. By implication, the equilibrium wage rate (and aggregate labor income) is the same in the two economies.

By contrast, the aggregate capital demand will be strictly larger in the economy with data (for any given interest rate). To see this, note that any firm in the economy without data sets its capital stock to solve the following maximization problem:

$$\max_{K_2} \Big\{ \mathbb{E}\big[A_2 y(K_2, N_2^*(A_2, w)) - w N_2^*(A_2, w) - r K_2\big] \Big\} \tag{8}$$

Any firm in the economy with data knows its idiosyncratic profitability realization when deciding its capital input. As a result, the firm chooses its capital stock to solve the following maximization problem:

$$\max_{K_2} \left\{ A_2 y(K_2, N_2^*(A_2, w)) - w N_2^*(A_2, w) - r K_2 \right\} \tag{9}$$

The result that the aggregate capital demand is strictly higher in the economy with data (for any given factor prices) then follows from two observations: First, the maximization problem of any firm in the economy without data is identical to the maximization problem of a firm in the economy with data whose second-period profitability is equal to its expected value. Thus, the aggregate capital stock in the economy without data is equal to the optimal capital stock of said firm in the economy with data, i.e. $\bar{K}^{nd}(p) = K^d(\mathbb{E}[A_2], p)$. Secondly, the optimal capital choice of firms in the economy with data is a convex function in the firm's profitability, because capital has diminishing returns in production. This means that the aggregate capital demand in the economy with data, namely $\mathbb{E}[K^d(A_2, p)]$, is strictly higher than $K^d(\mathbb{E}[A_2], p)$, i.e. the aggregate capital demand in the economy without data.

Thus, the aggregate demand for capital is higher in the economy with data if capital and labor are perfect substitutes. This implies that the equilibrium interest rate (and by implication, aggregate capital income) will be higher in the economy with data. Because labor income is identical in both economies and capital income increases disproportionately when data is available to firms, the labor share will be lower in the economy with data and the capital share will be higher in the economy with data. Finally, note that the profit share is higher in the economy with data if and only if $\alpha_N > \alpha_K$. Intuitively, this is because the ordering of these parameters governs how much of total revenue entrepreneurs have to pay to capital owners. When α_K is relatively small, entrepreneurs can appropriate a large share of the revenue generated by the higher capital usage in the economy with data. Given that estimates of α_K are in the range $[0.3, 0.4]$ while estimates of α_N are around 0.5, this suggests that entrepreneurs also disproportionately benefit from data.

When the production technology takes the Cobb-Douglas form, the labor share is equal to α_N and the capital share is equal to α_K in both economies. Because total output is higher in the economy with data, aggregate labor income and capital income are thus raised by the availability of data to firms.

Taken together, these results establish that the availability of data will exacerbate capital-labor inequality (which can be proxied by the capital share) if capital and labor are sufficiently substitutable. This is because the availability of data to firms primarily increases the demand for capital by resolving the uncertainty firms face regarding their future profitability, which only plays a role when firms set their capital inputs. Endowing firms with data will only increase their labor demand if labor and capital are complements in production. The associated increase of the aggregate labor demand implied by this second-order effect is larger, the more complementary the two inputs are.

In the following, I provide some numerical results which underline this result further: As during the previous analysis, I abstract from price effects (i.e. η is set to 0). By contrast, I now consider generalized CES-type production functions that take the following form:

$$y(K_2, N_2) = \left[\alpha_K (K_2)^\gamma + \alpha_N (N_2)^\gamma\right]^{\nu/\gamma} \tag{10}$$

In the following, I plot the equilibrium capital and labor shares in the two different economies for different combinations of α_N, α_K, and γ. I consider an example in which $A_2 \sim U[0, 2]$ and $v = 0.5$. A given graph corresponds to a fixed combination of $\alpha_K \in \{0.2, 0.4\}$ and $\alpha_N = 0.45$. The focus of this analysis is the degree of complementarity between inputs, which is governed by γ. I consider $\gamma \in (0.2, 0.5)$, which I plot on the x-axis of each graph. Higher values of γ represent higher substitutability of the input factors. The equilibrium wage rate in the economy with data (without data) is plotted in dark (light) blue. The interest rate in the economy with data (without data) is plotted in dark (light) red.

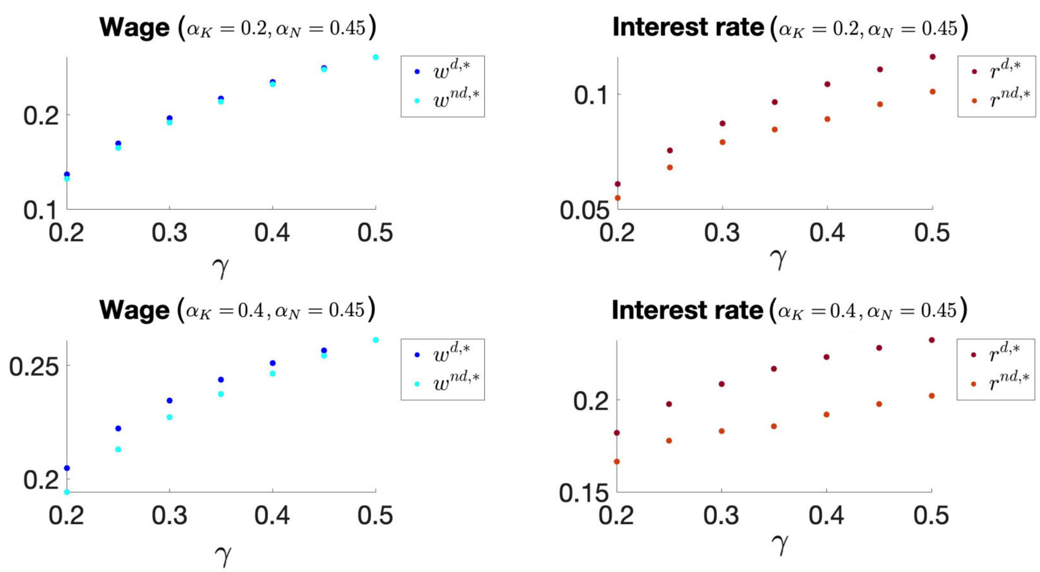

Figure 1: Equilibrium factor prices & input substitutability

The previous insights are reinforced: As the substitutability between the input factors rises, the difference in the ratio of interest rate and wage between the two economies increases.

3.2 Price elastic demand

In this section, I consider how the availability of data to firms affects capital/labor inequality by changing the prices of the output goods produced by firms. Formally, I now consider general values of $\eta > 0$, while I had restricted attention to price-inelastic demand ($\eta = 0$) in the previous section.

There are two main take-aways from this section: First, any increase in the aggregate capital demand induced by the availability of data crowds out labor demand. This is because a firm which employs more capital produces more, which leads to a reduced output price, thereby reducing this firm's incentives to hire labor. Secondly, the availability of data may lead to a reduction in both the aggregate capital and labor demand. This is because the availability of data allows firms to produce more efficiently, i.e. use less inputs to produce a given amount of output. While access to data also entails a positive effect on the expected output of firms (as outlined in the previous subsection), the magnitude of this effect is small when demand is very price elastic. Then, the ability to produce more efficiently using data will mean that both aggregate factor demands are reduced by the availability of data.

I begin by formalizing the first point for a simple example in which capital and labor are perfect substitutes:

Proposition 2 (Crowding out of labor demand)
Suppose $\eta > 0$ and $y(K_2, N_2) = (K_2)^{0.5} + (N_2)^{0.5}$. For any wage and interest rate, aggregate labor demand is smaller in the economy with data and aggregate capital demand is larger in the economy with data.

This implies that the equilibrium wage rate (and thus, total labor income) will be lower in the economy with data, while the equilibrium interest rate (and total capital income) is larger in the economy with data. This is visualized in the following example, in which I consider a production technology $y(K_2, N_2) = (K_2)^{0.5} + (N_2)^{0.5}$ and impose that $A_2 \sim U[0, 2]$. Different values of η are plotted on the x-axis of the graph. The equilibrium wage and interest rate in the economy with data (economy without data) for a given η are plotted in blue (red):

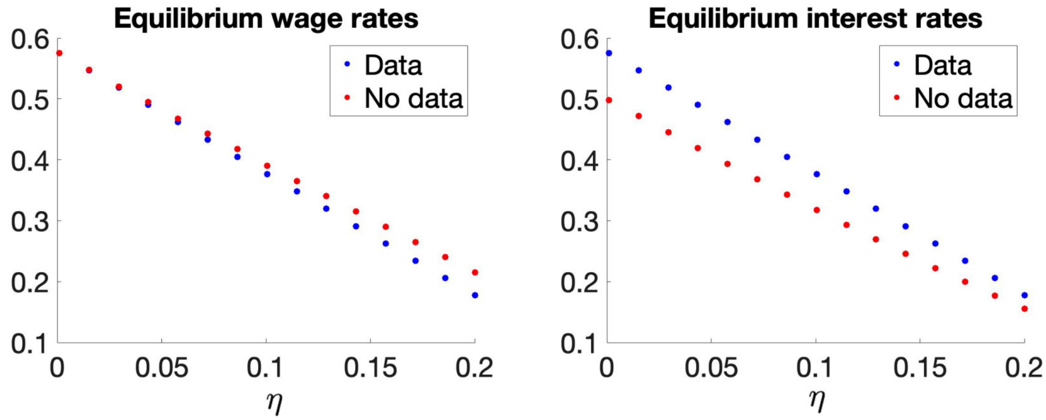

Figure 2: Equilibrium factor prices & price elasticity

In the following, I now formalize the second aforementioned channel, namely that both factor demands can be reduced by the availability of data because this enables firms to use inputs more efficiently. In other words, total labor demand can be lower in the economy with data even when the crowding-out effect is absent:

Proposition 3 (Production efficiency & data)
Suppose $\eta > 0$ and $y(K_2, N_2) = (K_2)^{0.5}(N_2)^{0.5}$. For any wage and interest rate, aggregate labor and capital demand are smaller in the economy with data. Thus, the equilibrium wage and interest rate are smaller in the economy with data, while total output is larger.

In a nutshell, the availability of data entails two opposing effects on the aggregate demand for capital. First, the fact that firms' capital demand in the economy with data is convex in profitability (which holds true if capital has a diminishing marginal product) raises the aggregate capital demand in the economy with data, relative to the economy without data. This was the key channel discussed in the previous subsection and is the only relevant channel when there are no price effects. Second, the availability of data enables firms to produce more efficiently, i.e. to use less inputs to produce a given amount of output. This effect reduces the aggregate demand for both inputs in the economy with data, relative to the economy without data. If the demand for the output good is very price elastic, the first channel is relatively weak, because the price reduction induced by an increase of output makes it less attractive for firms to increase their capital and labor inputs. Then, the second channel will dominate, i.e. the aggregate demand for both factors will be lower in the economy with data.

The relative strength of these two effects is not only determined by the price elasticity of demand for the output goods (η), but also by the returns to scale in production. This is formalized in proposition 3, where I have considered a production function with constant returns to scale. Under this specification, capital no longer has diminishing returns in production (absent any price effects), which renders the first channel mute. In other words, the strength of the first effect (which raises the aggregate demand for capital in the economy with data, relative to the aggregate capital demand in the economy without data) is greater when the production function exhibits smaller returns to scale, because this makes the firms' capital demand in the economy with data more convex in profitability.

This intuition is underlined in the following numerical exercise, where I consider Cobb-Douglas production functions $y(K_1, N_1) = (K_1)^{\alpha_K}(N_1)^{\alpha_N}$. I fix $\alpha_N = 0.45$ and consider two different $\alpha_K \in \{0.2, 0.4\}$. Every graph corresponds to a given combination of α_K and α_N and different levels of η are plotted on the x-axis of the respective graphs. The equilibrium wage rate in the economy with data (without data) is plotted in dark (light) blue. The interest rate in the economy with data (without data) is plotted in dark (light) red.

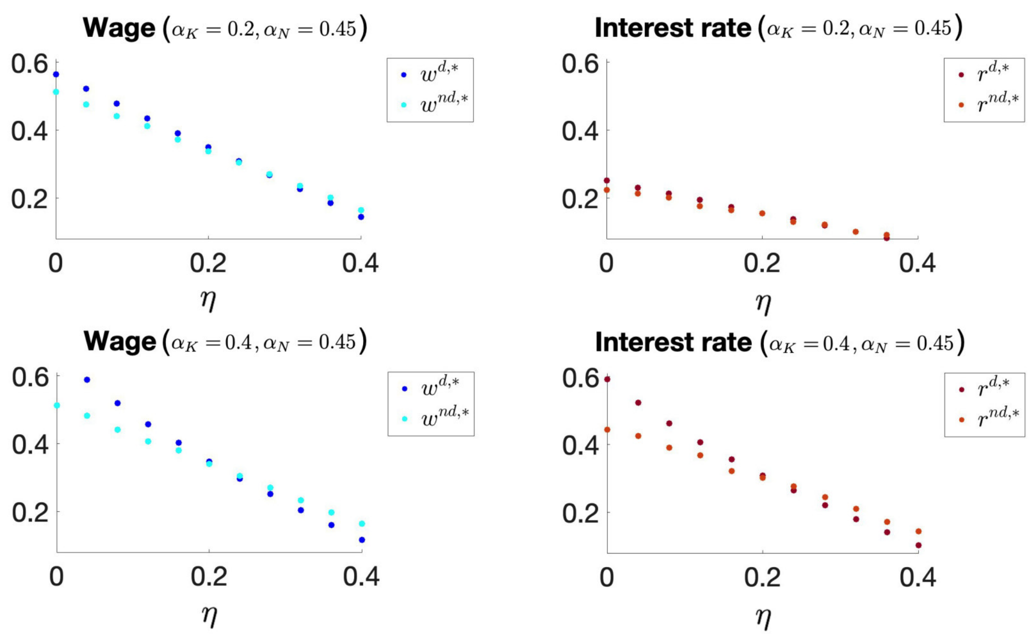

Figure 3: Equilibrium factor prices — the role of α and η

These graphs reinforce two ideas: Firstly, increases in the price elasticity of demand reduce the wage and the interest rate in the economy with data, relative to the economy with data. This is because the positive effect of data on the factor prices through the expansion of output is mitigated as demand gets more price elastic. Secondly, as the production technology moves closer towards constant returns to scale, the factor prices in the economy with data become more likely to be lower than the factor prices in the economy without data (at a given price elasticity of demand). This is because the expansion of output induced by the availability of data becomes weaker as the production technology moves closer to constant returns to scale, given that capital becomes less convex in profitability.

4 Data and the skill premium

4.1 Model

I consider a unit mass of heterogeneous firms (a firm-level index is omitted for ease of exposition) that operate in two time periods $t = \{1, 2\}$. In each period, any firm produces output using skilled (H_t) and unskilled labor (L_t). A firm's revenue takes the following form:

$$Y_t = y(L_t, H_t), \tag{11}$$

where $y(L_t, H_t)$ is some twice continuously differentiable function that is increasing in both arguments and A_t denotes the profitability of the firm in period t. In each period $t \in \{1, 2\}$, any firm chooses how much unskilled labor (L_t) and skilled labor (H_t) it wishes to hire in this period. Firms do not discount future profits.

The initial profitability of firms (A_1) follows some arbitrary continuous distribution F with support $[\underline{A}_1, \bar{A}_1]$. The second-period profitability of any firm (A_2) satisfies the following:

$$A_2 = A_1 + \epsilon_2, \tag{12}$$

where $\epsilon_2 \sim G$ and G is some arbitrary continuous distribution with $\mathbb{E}[\epsilon_2] = 0$.

The wage received by skilled workers is denoted by w_H and the wage received by unskilled workers is denoted by w_L. I assume, for simplicity, that both wages are constant over time. This can be microfounded by assuming that workers can work in period 1 and 2 and that their utility from leisure is linear, which implies that the wage of a given worker type will be the same in both periods (in equilibrium). I assume that both types of labor are inelastically supplied, where $\bar{L}$ denotes the exogenously given supply of unskilled labor and $\bar{H}$ denotes the exogenously given supply of skilled labor.

The firm chooses its inputs while facing adjustment costs. Based on the empirical evidence, I impose that the costs of adjusting skilled labor are higher, which I formalize as follows: Unskilled labor can be freely adjusted in both periods, whereas skilled labor can only be freely adjusted in period 1. A firm which employed H_1 units of skilled labor in period 1 will face the following total adjustment costs in period 2:

$$C(H_2, H_1) = \kappa\varphi(|H_2 - H_1|), \tag{13}$$

where H_2 is the amount of skilled labor used by this firm in period 2 and $\varphi(.)$ is an arbitrary strictly increasing function. The parameter κ, which I assume to be weakly positive, governs the degree of adjustment costs faced by any firm.

Given this setup, the second-period profits of any firm are given by:

$$\Pi_2 = A_2 y(L_2, H_2) - w_L L_2 - w_H H_2 - \kappa\varphi(|H_2 - H_1|) \tag{14}$$

The first-period profits of any firm are:

$$\Pi_1 = A_1 y(L_1, H_1) - w_L L_1 - w_H H_1 \tag{15}$$

Data enables firms to predict their future productivities. As before, I formalize this by considering two different economies, namely *an economy without data* and an *economy with data*. In either economy, every firm learns its realization of A_1 at the beginning of period 1, i.e. before choosing L_1 and H_1. In the economy with data, every firm further knows its realization of A_2 at the beginning of period 1, i.e. it knows its profitability in both periods when making its labor input choices in the first period. In the economy without data, no firm has access to this information and must (at the beginning of period 1) form expectations about its future profitability based on its knowledge of A_1 and the distribution of ϵ_2.

Thus, the first-period skilled labor demand of any firm in the economy without data is a function $H_1^{nd}(A_1; w)$ and depends on the wage levels $w = (w_L, w_H)$ and the firm's initial profitability (A_1). The first-period skilled labor demand of any firm in the economy with data also conditions on A_2 and is given by a function $H_1^d(A_1, A_2; w)$. The function describing the optimal second-period skilled labor demand is the same in either economy and expressed by the function $H_2^*(H_1, A_2, w)$. In either period, firms choose their unskilled labor input to maximize the following:

$$A_t y(H_t, L_t) - w_L L_t, \tag{16}$$

the solution to which is described by a function $L^*(H_t, A_t; w)$.

Key objects of analysis will be the aggregate labor demands, which determine the skill premium, namely the ratio w_H/w_L. I define the aggregate skilled labor demands in the *economy without data* in periods 1 and 2 as $\bar{H}_1^{nd}(w)$ and $\bar{H}_2^{nd}(w)$, respectively. These are given by:

$$\bar{H}_1^{nd}(w) = \int_{\underline{A}_1}^{\bar{A}_1} H_1^{nd}(A_1; w) dF(A_1) \tag{17}$$

$$\bar{H}_2^{nd}(w) = \int_{\underline{A}_1}^{\bar{A}_1} \int_{\underline{\epsilon}}^{\bar{\epsilon}} H_2^*(H_1^{nd}(A_1; w), A_1 + \epsilon; w) dG(\epsilon) dF(A_1) \tag{18}$$

Analogously, the first-period and second-period aggregate skilled labor demands in the *economy with data*, which I define as $\bar{H}_1^d(w)$ and $\bar{H}_2^d(w)$, are given by

$$\bar{H}_1^d(w) = \int_{\underline{A}_1}^{\bar{A}_1} \int_{\underline{\epsilon}}^{\bar{\epsilon}} H_1^d(A_1, A_1 + \epsilon; w) dG(\epsilon) dF(A_1) \tag{19}$$

$$\bar{H}_2^d(w) = \int_{\underline{A}_1}^{\bar{A}_1} \int_{\underline{\epsilon}}^{\bar{\epsilon}} H_2^*(H_1^d(A_1, A_1 + \epsilon; w), A_1 + \epsilon; w) dG(\epsilon) dF(A_1) \tag{20}$$

The aggregate unskilled labor demands in the economy without data can be defined analogously — they are the expected value of a firm's unskilled labor demand, which is given by the function $L^*(H_t, A_t; w)$ in either economy. The distinction between the economies is that $H_t = H_t^d(\cdot)$ is inserted when studying the economy with data and $H_t = H_t^{nd}(\cdot)$ is inserted when considering the economy without data.

The total aggregate labor demand (summed over both periods) for unskilled and skilled labor is denoted by $\bar{L}^j = \bar{L}_1^j + \bar{L}_2^j$ and $\bar{H}^j = \bar{H}_1^j + \bar{H}_2^j$, respectively, where $j \in \{d, nd\}$ captures the economy under consideration. I refer to the model I have just outlined as the *skilled/unskilled labor framework*.

4.2 Analytical results

To fix ideas, suppose that the costs of adjusting the skilled labor supply in period 2 are prohibitively high, i.e. $\kappa \to \infty$. This implies that any firm will always optimally choose its second-period skilled labor input to be equal to the input level it chose in the first period, i.e. that any firm only makes a meaningful choice regarding its skilled labor input in the first period.

Once again, I consider two different simple production functions $y(L_t, H_t)$, namely (i) the perfect substitutes production function $y(L_t, H_t) = (L_t)^{\alpha_L} + (H_t)^{\alpha_H}$ with $\alpha_L \in (0,1)$ and $\alpha_H \in (0,1)$ and (ii) the Cobb-Douglas production function $y(L_t, H_t) = (L_t)^{\alpha_L}(H_t)^{\alpha_H}$ with $\alpha_L + \alpha_H < 1$.

Recall that firms face no adjustment costs pertaining to unskilled labor. Thus, firms always choose the statically optimal value of unskilled labor, independent of whether they are in the economy with or without data. This statically optimal input choice $L^*(H_t, A_t; w)$

satisfies the following first-order condition:

$$A_t \frac{\partial y(L^*(\cdot), H_t)}{\partial L_t} - w_L = 0 \tag{21}$$

I now characterize the optimal first-period labor input choices of firms in the economy without data. Conditional on a fixed level of A_1, any firm in the economy without data faces the same optimization problem when choosing its skilled labor input in the first period, namely:

$$\max_{H_1} \left\{ A_1 H_1^{\alpha_H} - w_H H_1 - w_L L^*(H_1, A_1; w) + \mathbb{E}_\epsilon \big[A_2 H_1^{\alpha_H} - w_H H_1 - w_L L^*(H_1, A_1; w) \big] \right\} \tag{22}$$

Now consider the first-period decision problem of a firm with access to data, which knows both its first-period profitability and second-period profitability. Conditional on its productivities A_1 and A_2, this firm's optimization problem is given by:

$$\max_{H_1} \left\{ A_1 H_1^{\alpha_H} - w_H H_1 - w_L L^*(H_1, A_1; w) + \big[A_2 H_1^{\alpha_H} - w_H H_1 - w_L L^*(H_1, A_2; w) \big] \right\} \tag{23}$$

In the following, I establish how access to data shapes the demands for skilled and unskilled labor, and how this depends on the degree of complementarity between the two input factors:

Proposition 4 (Data & the skill premium: Input complementarity)
Consider the skilled/unskilled labor framework. Suppose that $\kappa \to \infty$.

- *If $y(L_t, H_t) = (H_t)^{\alpha_H} + (L_t)^{\alpha_L}$, then $\bar{H}^d(w)/\bar{L}^d(w) > \bar{H}^{nd}(w)/\bar{L}^{nd}(w)$ holds for any w. Thus, the equilibrium skill premium is strictly higher in the economy with data.*
- *If $y(L_t, H_t) = (H_t)^{\alpha_H}(L_t)^{\alpha_L}$, then $\bar{H}^d(w)/\bar{L}^d(w) = \bar{H}^{nd}(w)/\bar{L}^{nd}(w)$ holds for any w. Thus, the equilibrium skill premium is the same in the two economies.*

The intuition behind this result is the following: Recall that a firm faces adjustment costs pertaining to the amount of skilled labor it hires in period 2, while unskilled labor can be freely adjusted. Because of this, increases in the level of uncertainty a firm faces will disproportionately reduce its demand for skilled labor. Given that access to data resolves the uncertainty a firm faces, the aggregate demand for skilled workers will be higher in the economy with data than in the economy without data.

The extent to which this affects the ratio of labor demands depends on the degree of complementarity/substitutability between the two input factors. If $y(L_t, H_t) = H_t^{\alpha_H} + L_t^{\alpha_L}$, then unskilled and skilled labor are perfectly substitutable. This means that increases in a firm's demand for skilled labor will not affect its demand for unskilled labor. Under this production function, the demand for unskilled labor will thus be the same in the economies with and without data. Since the aggregate demand for skilled labor is higher in the economy with data, the availability of data increases the skill premium.

When the production technology is Cobb-Douglas, data no longer impacts the skill premium. In a nutshell, this is because of the strong complementarity between inputs that exists for this production technology. In the economy with data, aggregate demand for skilled labor is still higher than in the economy without data. If skilled and unskilled labor demand are complementary, however, the aggregate demand for unskilled labor is also higher in the economy with data. For the given production technology, the increase in both aggregate factor demands is proportional, so the skill premium in both economies will be the same.

4.3 Numerical analysis

In the following, I provide some numerical analysis which establishes that the previous insights also hold true when considering non-prohibitive adjustment costs and general CES-type production functions given by:

$$y(L_t, H_t) = \left[0.5 L_t^{\alpha_L \sigma} + 0.5 H_t^{\alpha_H \sigma}\right]^{1/\sigma}, \tag{24}$$

where the parameter σ governs the degree of complementarity between the inputs. Throughout the numerical analysis, I consider linear adjustment costs, i.e. the function $\varphi(\cdot)$ is:

$$\varphi(x) = x \tag{25}$$

As before, the optimal level of unskilled labor in any period $t \in \{1, 2\}$, which is given by the function $L^*(H_t, A_t; w)$, must solve the following optimization problem:

$$\max_{L_t} \left[A_t y(L_t, H_t) - w_L L_t\right] \tag{26}$$

The second-period skilled labor input of any firm maximizes the following objective function:

$$\Pi^2(H_2, H_1; A_2, w) = A_2 y(L_2^*(H_2, A_2; w), H_2) - w_L L_2^*(H_2, A_2; w) - w_H H_2 - \kappa |H_2 - H_1| \tag{27}$$

In the economy without data, a firm with initial profitability A_1 maximizes the following objective function through choice H_1:

$$\Pi^{1,nd}(H_1; A_1, w) = A_1 y(L_1^*(H_1, A_1; w), H_1) - w_L L_1^*(H_1, A_1; w) - w_H H_1 + \int_{\underline{\epsilon}}^{\bar{\epsilon}} \Pi^2(H_2, H_1; A_2, w) dG(\epsilon) \tag{28}$$

In the economy with data, a firm with productivities A_1 and A_2 maximizes the following objective function through choice of H_1:

$$\Pi^{1,d}(H_1; A_1, A_2, w) = A_1 y(L_1^*(\cdot; w), H_1) - w_L L_1^*(\cdot 1; w) - w_H H_1 + \Pi^2(H_2; H_1, A_2) \tag{29}$$

In the following, I consider the following particular example: Any firm's initial profitability is non-stochastic and given by $A_1 = 0.5$. The second-period profitability shock follows $\epsilon_2 \sim U[-0.1, 0.1]$. Moreover, I set $\alpha_L = 0.3$, $\alpha_H = 0.4$, and fix the wage levels at $w_L = 0.02$ and $w_H = 0.04$. I plot the ratios $H^{nd}(w)/L^{nd}(w)$ and $H^d(w)/L^d(w)$ for different values of $\sigma \in (0.4, 1)$ (on the x-axis of each graph) and two different $\kappa \in \{1, 100\}$, which are visualized in separate graphs.

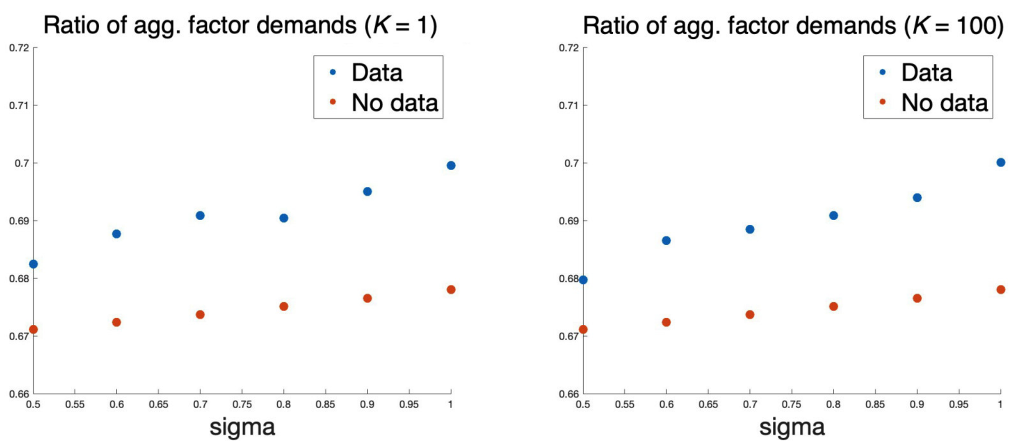

Figure 4: Factor demands with and without data

As σ increases, skilled and unskilled labor become more substitutable. The graph underlines the previously documented intuition by demonstrating the following: When the inputs are more substitutable, granting firms access to data leads to a greater increase in the relative factor demand for skilled labor and by extension, to a larger rise in the skill premium.

5 Conclusion

Modern economies increasingly revolve around (big) data, which fundamentally differs from existing technologies because it enables firms to predict their future profitabilities. In this paper, I have studied the distributional consequences of these technological developments: How is the enormous economic value generated by data distributed amongst capital owners, workers, and entrepreneurs? I consider a simple theoretical model in which firms produce output using capital and labor. The key feature of the model is the empirically well-established fact that it takes longer for firms to adjust their capital stock than their labor inputs. This means that firms set their capital inputs under uncertainty regarding their future profitability, which is resolved when firms decide how many workers to hire in a given period.

I have demonstrated that the increasing availability of data can be a powerful magnifier of inequality. This is because data endows firms with a forecast of their future profitability when they choose their capital inputs. Thus, the availability of data primarily raises the aggregate demand for and the remuneration of capital. Aggregate labor income only increases when labor and capital are complements in production. When the demand for output goods is price elastic, any increase in the aggregate demand for capital crowds out labor demand by reducing the price of the output goods. Furthermore, the availability of data enables firms to produce more efficiently (i.e. they require less inputs to produce a given amount of output). Thus, the availability of data may reduce the aggregate demand for both factors when consumer demand is sufficiently price elastic, in which case the economic value generated by data is appropriated by entrepreneurs.

A Proofs

Proof of Proposition 1:

Part 1: Perfect substitutes production function.

Part 1a: Calculating aggregate labor demand and total labor income.

Given that the optimal labor choices are independent of the capital choices, any firm will always set is labor demand equal to the statically optimal labor input, namely:

$$N^*(A_2) = \left(\frac{\alpha_N A_2}{w}\right)^{\frac{1}{1-\alpha_N}} \tag{30}$$

Thus, aggregate labor demand is the same in both economies. This must be equal to $\bar{N}$, so the equilibrium wage w^* must be the same in both economies. To calculate this, note that the aggregate labor demand in either economy is given by:

$$N^d(w) = N^{nd}(w) = \big(\alpha_N/w\big)^{\frac{1}{1-\alpha_N}}\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_N}}\big] \tag{31}$$

Defining $\bar{N}$ as the aggregate labor supply, the wage rate (in either economy) will solve:

$$w^* = \alpha_N\left(\frac{\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_N}}\big]}{\bar{N}}\right)^{1-\alpha_N} \tag{32}$$

Total labor income is given by $w^*\bar{N}$, i.e. is given by the following in both economies:

$$I^{N,d} = I^{N,nd} = \bar{N}\alpha_N\left(\frac{\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_N}}\big]}{\bar{N}}\right)^{1-\alpha_N} = \alpha_N(\bar{N})^{\alpha_N}\Big(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_N}}\big]\Big)^{1-\alpha_N} \tag{33}$$

Part 1b: Equilibrium interest rates and capital income shares.

Consider a firm in the economy without data. This firm maximizes the following profit function through choice of its capital:

$$\Pi^{nd}(K_2) = \mathbb{E}\big[A_2K_2^{\alpha_K} - rK_2\big] \tag{34}$$

The first-order condition reads:

$$\mathbb{E}[A_2]\alpha_K(K_2)^{\alpha_K-1} - r = 0 \tag{35}$$

Thus, the capital stock in the economy without data is:

$$\bar{K}^{nd} = \big(\alpha_K/r\big)^{\frac{1}{1-\alpha_K}}\big(\mathbb{E}[A_2]\big)^{\frac{1}{1-\alpha_K}} \tag{36}$$

By analogous arguments, the individual firm capital stock in the economy with data is:

$$K^d(A_2) = \big(\alpha_K A_2/r\big)^{\frac{1}{1-\alpha_K}} \tag{37}$$

This means that the aggregate capital stock in the economy with data is given by:

$$\bar{K}^d = \big(\alpha_K/r\big)^{\frac{1}{1-\alpha_K}}\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big] \tag{38}$$

Note that the aggregate capital supply is given by $\bar{K}$, which implies that the equilibrium interest rate in the economy with data (which I denote by $r^{d,*}$) is given by:

$$\big(\alpha_K/r^{d,*}\big)^{\frac{1}{1-\alpha_K}}\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big] = \bar{K} \iff r^{d,*} = \alpha_K\left(\frac{\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big]}{\bar{K}}\right)^{1-\alpha_K} \tag{39}$$

By analogous arguments, the interest rate in the economy without data is given by:

$$r^{nd,*} = \alpha_K\left(\frac{\big[\mathbb{E}(A_2)\big]^{\frac{1}{1-\alpha_K}}}{\bar{K}}\right)^{1-\alpha_K} \tag{40}$$

Thus, capital income in the economy with and without data is given by:

$$I^{K,d} = \alpha_K(\bar{K})^{\alpha_K}\left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big]\right)^{1-\alpha_K} \quad ; \quad I^{K,nd} = \alpha_K(\bar{K})^{\alpha_K}\big[\mathbb{E}(A_2)\big] \tag{41}$$

<u>Part 1c:</u> Equilibrium profits and profit shares

We begin by considering the economy with data. In this economy, the profits any firm makes in the production period are given by:

$$\Pi^d(A_2) = A_2\big(K^d(A_2)\big)^{\alpha_K} + A_2\big(N^*(A_2)\big)^{\alpha_K} - w^*N_2^*(A_2) - r^{d,*}K^d(A_2)$$

The aggregate profits (from capital) in the economy with data are given by:

$$\bar{\Pi}^{K,d} = \mathbb{E}\left[A_2\left(\frac{\alpha_K A_2}{r^{d,*}}\right)^{\frac{\alpha_K}{1-\alpha_K}}\right] - \alpha_K\left(\frac{\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K}}\right]}{\bar{K}}\right)^{1-\alpha_K}\bar{K} =$$

$$\big(1-\alpha_K\big)(\bar{K})^{\alpha_K}\left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big]\right)^{1-\alpha_K} \tag{42}$$

The aggregate profits from capital in the economy without data are given by:

$$\bar{\Pi}^{K,nd} = \mathbb{E}\left[A_2\big(\alpha_K/r^{nd,*}\big)^{\frac{\alpha_K}{1-\alpha_K}}\big(\mathbb{E}[A_2]\big)^{\frac{\alpha_K}{1-\alpha_K}}\right] - \alpha_K\left(\frac{\big[\mathbb{E}(A_2)\big]^{\frac{1}{1-\alpha_K}}}{\bar{K}}\right)^{1-\alpha_K}\bar{K} =$$

$$\big(1-\alpha_K\big)(\bar{K})^{\alpha_K}\big[\mathbb{E}(A_2)\big] \tag{43}$$

Because $\frac{1}{1-\alpha_K} > 1$, we have:

$$\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big] > \big[\mathbb{E}\big(A_2\big)\big]^{\frac{1}{1-\alpha_K}} \implies \left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big]\right)^{1-\alpha_K} > \mathbb{E}\big[A_2\big] \tag{44}$$

Aggregate profits from labor are the same in both economies and given by:

$$\Pi^{L,d} = \Pi^{L,nd} = \mathbb{E}\left[A_2\left(\frac{\alpha_N A_2}{w}\right)^{\frac{\alpha_N}{1-\alpha_N}}\right] - w^*\bar{N} = (1-\alpha_N)\big(\bar{N}\big)^{\alpha_N}\left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_N}}\big]\right)^{1-\alpha_N}$$

<u>Part 1d:</u> Capital, labor, and profit shares

The capital share in the economy with data is given by:

$$S^{K,d} = \frac{\alpha_K(\bar{K})^{\alpha_K}\left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big]\right)^{1-\alpha_K}}{(\bar{K})^{\alpha_K}\left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big]\right)^{1-\alpha_K} + \big(\bar{N}\big)^{\alpha_N}\left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_N}}\big]\right)^{1-\alpha_N}} = \frac{\alpha_K}{1+Z^{K,d}}, \tag{45}$$

where I have defined

$$Z^{K,d} := \frac{\big(\bar{N}\big)^{\alpha_N}\left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_N}}\big]\right)^{1-\alpha_N}}{(\bar{K})^{\alpha_K}\left(\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K}}\big]\right)^{1-\alpha_K}} \tag{46}$$

By similar arguments, the capital share in the economy without data is given by:

$$S^{K,nd} = \frac{\alpha_K(\bar{K})^{\alpha_K}\mathbb{E}[A_2]}{(\bar{K})^{\alpha_K}\mathbb{E}[A_2] + \left(\bar{N}\right)^{\alpha_N}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N}} = \frac{\alpha_K}{1+Z^{K,nd}}, \tag{47}$$

where I have defined

$$Z^{K,nd} := \frac{\left(\bar{N}\right)^{\alpha_N}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N}}{(\bar{K})^{\alpha_K}\mathbb{E}[A_2]} \tag{48}$$

Note that:

$$\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K}}\right]\right)^{1-\alpha_K} > \mathbb{E}[A_2] \iff Z^{K,d} < Z^{K,nd} \tag{49}$$

From this, it follows that the capital share in the economy with data is strictly larger. The labor share in the economy with data is given by:

$$S^{N,d} = \frac{\alpha_N\left(\bar{N}\right)^{\alpha_N}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N}}{(\bar{K})^{\alpha_K}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K}}\right]\right)^{1-\alpha_K} + \left(\bar{N}\right)^{\alpha_N}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N}} = \frac{\alpha_N}{1+(1/Z^{K,d})} \tag{50}$$

Moreover, the labor share in the economy without data is given by:

$$S^{N,nd} = \frac{\alpha_N\left(\bar{N}\right)^{\alpha_N}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N}}{(\bar{K})^{\alpha_K}\mathbb{E}[A_2] + \left(\bar{N}\right)^{\alpha_N}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N}} = \frac{\alpha_N}{1+(1/Z^{K,nd})} \tag{51}$$

It follows that the labor share in the economy with data is smaller.

Now, we consider the profit share in total GDP. In the economy with data, this is:

$$S^{\pi,d} = \frac{(1-\alpha_K)(\bar{K})^{\alpha_K}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K}}\right]\right)^{1-\alpha_K} + (1-\alpha_N)\left(\bar{N}\right)^{\alpha_N}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N}}{(\bar{K})^{\alpha_K}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K}}\right]\right)^{1-\alpha_K} + \left(\bar{N}\right)^{\alpha_N}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N}} \tag{52}$$

In general, the profit share in the economy with data can be written as:

$$S^{\pi,d} = (1-\alpha_K)\frac{1}{1+Z^{K,d}} + (1-\alpha_N)\frac{1}{1+(1/Z^{K,d})} = \frac{(1-\alpha_K)+(1-\alpha_N)Z^{K,d}}{1+Z^{K,d}} \tag{53}$$

By similar arguments, the profit share in the economy without data can be written as:

$$S^{\pi,nd} = (1-\alpha_K)\frac{1}{1+Z^{K,nd}} + (1-\alpha_N)\frac{1}{1+(1/Z^{K,nd})} = \frac{(1-\alpha_K)+(1-\alpha_N)Z^{K,nd}}{1+Z^{K,nd}} \tag{54}$$

The profit share in the economy with data is higher if and only if $\alpha_N > \alpha_K$. This holds by the following logic:

$$S^{\pi,d} > S^{\pi,nd} \iff \frac{(1-\alpha_K)+(1-\alpha_N)Z^{K,d}}{1+Z^{K,d}} > \frac{(1-\alpha_K)+(1-\alpha_N)Z^{K,nd}}{1+Z^{K,nd}}$$

$$\iff$$

$$(1-\alpha_K)\big(Z^{K,nd}-Z^{K,d}\big) > (1-\alpha_N)\big(\underbrace{Z^{K,nd}-Z^{K,d}}_{>0}\big) \iff \alpha_N > \alpha_K \tag{55}$$

Part 2: Cobb-Douglas production function

Part 2a: Preliminaries

Regardless of whether a firm had access to data in period 1 or not, it will maximize the following objective function through choice of N_2 in period 1:

$$\Pi^1(N_2; K_2, A_2) = A_2(K_2)^{\alpha_K}(N_2)^{\alpha_N} - wN_2 \tag{56}$$

Thus, the optimal labor input in period 2 must satisfy the following:

$$A_2(K_2)^{\alpha_K}(\alpha_N)N_2^{\alpha_N-1} - w = 0 \iff N_2^*(A_2, K_2) = \left(\frac{\alpha_N A_2 K_2^{\alpha_K}}{w}\right)^{\frac{1}{1-\alpha_N}} \tag{57}$$

Part 2b: Aggregate factor inputs (and factor prices) in the economy with data

Now consider the initial period optimization problem of a firm in the economy with data. Given its knowledge of A_2, this firm chooses its capital stock K_2 to maximize the following:

$$\Pi^d(K_2) = A_2 K_2^{\alpha_K} N_2^{\alpha_N} - rK_2 - wN_2 \tag{58}$$

Taking the first-order condition and plugging in the optimal labor input yields:

$$A_2 \alpha_K (K_2)^{\alpha_K - 1} \left(\frac{\alpha_N A_2 K_2^{\alpha_K}}{w} \right)^{\frac{\alpha_N}{1-\alpha_N}} - r = 0 \tag{59}$$

The firm-level capital stock in the economy with data is hence given by:

$$K^d(A_2) = \left((\alpha_K / r)^{1-\alpha_N} (\alpha_N / w)^{\alpha_N} A_2 \right)^{\frac{1}{1-\alpha_K-\alpha_N}} \tag{60}$$

Thus, the aggregate capital stock in the economy with data is given by:

$$\bar{K}^d = \left((\alpha_K / r)^{1-\alpha_N} (\alpha_N / w)^{\alpha_N} \right)^{\frac{1}{1-\alpha_K-\alpha_N}} \mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}} \right] \tag{61}$$

I now compute the individual and aggregate labor inputs. For a given firm with fixed A_2, the firm-level labor input (in the economy with data) is:

$$N^d(A_2) = \left(\frac{\alpha_N A_2 (K^d(A_2))^{\alpha_K}}{w} \right)^{\frac{1}{1-\alpha_N}} = \left[\frac{\alpha_N A_2}{w} \left((\alpha_K / r)^{1-\alpha_N} (\alpha_N / w)^{\alpha_N} A_2 \right)^{\frac{\alpha_K}{1-\alpha_K-\alpha_N}} \right]^{\frac{1}{1-\alpha_N}} =$$

$$\left[(\alpha_K / r)^{\alpha_K} (\alpha_N / w)^{1-\alpha_K} \right]^{\frac{1}{1-\alpha_K-\alpha_N}} \left(A_2 \right)^{\frac{1}{1-\alpha_K-\alpha_N}} \tag{62}$$

Thus, the aggregate labor demand in the economy with data is:

$$\bar{N}^d = \left[\left(\alpha_K / r \right)^{\alpha_K} \left(\alpha_N / w \right)^{1-\alpha_K} \right]^{\frac{1}{1-\alpha_K-\alpha_N}} \mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}} \right] \tag{63}$$

Now, we plug these results into the market clearing conditions, which jointly pin down the equilibrium wage and interest rate. The labor market clearing condition reads:

$$\bar{N} = \bar{N}^d \iff \bar{N} = \left[\left(\alpha_K / r \right)^{\alpha_K} \left(\alpha_N / w \right)^{1-\alpha_K} \right]^{\frac{1}{1-\alpha_K-\alpha_N}} \mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}} \right] \iff$$

$$w = \left[\left(\alpha_K \right)^{\alpha_K} \left(\alpha_N \right)^{1-\alpha_K} \right]^{\frac{1}{1-\alpha_K}} (r)^{\frac{-\alpha_K}{1-\alpha_K}} \left[\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}} \right] \right]^{\frac{1-\alpha_K-\alpha_N}{(1-\alpha_K)}} (\bar{N})^{-\frac{1-\alpha_K-\alpha_N}{(1-\alpha_K)}} \tag{64}$$

From the capital market clearing condition, we have:

$$\bar{K} = \bar{K}^d \iff \bar{K} = \left((\alpha_K / r)^{1-\alpha_N} (\alpha_N / w)^{\alpha_N} \right)^{\frac{1}{1-\alpha_K-\alpha_N}} \mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}} \right] \tag{65}$$

Taking the ratio of the two market clearing conditions yields:

$$\frac{\bar{K}}{\bar{N}} = \frac{\Big[(\alpha_K/r)^{1-\alpha_N}(\alpha_N/w)^{\alpha_N}\Big]^{\frac{1}{1-\alpha_K-\alpha_N}} \mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\big]}{\Big[\big(\alpha_K/r\big)^{\alpha_K}\big(\alpha_N/w\big)^{1-\alpha_K}\Big]^{\frac{1}{1-\alpha_K-\alpha_N}} \mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\big]} \iff w = \big(\bar{K}/\bar{N}\big)\big(\alpha_N/\alpha_K\big)r \quad (66)$$

Plugging this into the market clearing condition for labor yields:

$$w = \Big[\big(\alpha_K\big)^{\alpha_K}\big(\alpha_N\big)^{1-\alpha_K}\Big]^{\frac{1}{1-\alpha_K}} (r)^{\frac{-\alpha_K}{1-\alpha_K}} \Big[\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\big]\Big]^{\frac{1-\alpha_K-\alpha_N}{(1-\alpha_K)}} (\bar{N})^{-\frac{1-\alpha_K-\alpha_N}{(1-\alpha_K)}} \iff$$

$$r^{*,d} = \alpha_K\left(\frac{(\bar{N})^{\alpha_N}}{(\bar{K})^{1-\alpha_K}}\right)\Big[\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\big]\Big]^{1-\alpha_K-\alpha_N} \quad (67)$$

Plugging this into the expression for the wage-interest rate ratio yields the following:

$$w = \big(\bar{K}/\bar{N}\big)\big(\alpha_N/\alpha_K\big)r^{d,*} = \big(\bar{K}/\bar{N}\big)\big(\alpha_N/\alpha_K\big)\alpha_K\left(\frac{(\bar{N})^{\alpha_N}}{(\bar{K})^{1-\alpha_K}}\right)\Big[\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\big]\Big]^{1-\alpha_K-\alpha_N}$$

$$\iff$$

$$w^{d,*} = \alpha_N\left(\frac{(\bar{K})^{\alpha_K}}{(\bar{N})^{1-\alpha_N}}\right)\Big[\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\big]\Big]^{1-\alpha_K-\alpha_N} \quad (68)$$

Aggregate capital income in the economy with data is given by:

$$I^{K,d} = \alpha_K(\bar{N})^{\alpha_N}(\bar{K})^{\alpha_K}\Big[\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\big]\Big]^{1-\alpha_K-\alpha_N} \quad (69)$$

Aggregate labor income in the economy with data is given by:

$$I^{N,d} = \alpha_N(\bar{K})^{\alpha_K}(\bar{N})^{\alpha_N}\Big[\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\big]\Big]^{1-\alpha_K-\alpha_N} \quad (70)$$

Part 2c: Aggregate profits in the economy with data.

Total production in the economy with data is given by:

$$Y^d = \mathbb{E}\Big[A_2\big(K^d(A_2)\big)^{\alpha_K}\big(N^d(A_2)\big)^{\alpha_N}\Big] =$$

$$\left[\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\right]\right]\left(\alpha_K\right)^{\frac{\alpha_K}{1-\alpha_K-\alpha_N}}\left(\alpha_N\right)^{\frac{\alpha_N}{1-\alpha_K-\alpha_N}}\left(r\right)^{\frac{-\alpha_K}{1-\alpha_K-\alpha_N}}\left(w\right)^{\frac{-\alpha_N}{1-\alpha_K-\alpha_N}} \tag{71}$$

Plugging in the factor prices yields:

$$Y^d = \left[\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}}\right]\right]^{1-\alpha_K-\alpha_N}\left(\bar{N}\right)^{\alpha_N}\left(\bar{K}\right)^{\alpha_K} \tag{72}$$

Part 2d: Aggregate factor inputs (and factor prices) in the economy without data

Now consider the initial period optimization problem of a firm in the economy without data. This firm chooses its capital stock K_2 to maximize the following objective function:

$$\Pi^{nd}(K_2) = \mathbb{E}_{A_2}\left[A_2 K_2^{\alpha_K} N_2^{\alpha_N} - rK_2 - wN_2\right] \tag{73}$$

The envelope theorem still applies. The derivative of the optimal labor input with respect to the capital stock must be zero. Thus, taking the relevant first-order condition yields:

$$\alpha_K (K_2)^{\alpha_K - 1}\int_{\underline{A}}^{\bar{A}} A_2\left(\frac{\alpha_N A_2 K_2^{\alpha_K}}{w}\right)^{\frac{\alpha_N}{1-\alpha_N}} dF(A_2) - r = 0 \iff$$

$$K^{nd} = \left((\alpha_K/r)^{1-\alpha_N}(\alpha_N/w)^{\alpha_N}\right)^{\frac{1}{1-\alpha_K-\alpha_N}}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{\frac{1-\alpha_N}{1-\alpha_K-\alpha_N}} \tag{74}$$

This is also equal to the aggregate capital stock in the economy without data, namely $\bar{K}^{nd}$.

Using the previous results, we can calculate the individual firm labor demand in the economy without data. For a given firm with fixed A_2, this is:

$$N^{nd}(A_2) = \left(\frac{\alpha_N A_2 (\bar{K}^{nd})^{\alpha_K}}{w}\right)^{\frac{1}{1-\alpha_N}} =$$

$$(\alpha_K/r)^{\frac{\alpha_K}{1-\alpha_K-\alpha_N}}(\alpha_N/w)^{\frac{1-\alpha_K}{1-\alpha_N-\alpha_K}}\left(A_2\right)^{\frac{1}{1-\alpha_N}}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{\frac{\alpha_K}{1-\alpha_K-\alpha_N}} \tag{75}$$

Using this, one can calculate the aggregate labor demand in the economy without data.

$$\bar{N}^{nd} = \left[(\alpha_K/r)^{\alpha_K}(\alpha_N/w)^{1-\alpha_K}\right]^{\frac{1}{1-\alpha_N-\alpha_K}}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{\frac{1-\alpha_N}{1-\alpha_K-\alpha_N}} \tag{76}$$

Using these expressions, we can pin down solutions for the equilibrium wage and interest rate in the economy without data. Taking the ratio of the market-clearing conditions yields:

$$\frac{\bar{K}}{\bar{N}} = \frac{\left[(\alpha_K/r)^{1-\alpha_N}(\alpha_N/w)^{\alpha_N}\right]^{\frac{1}{1-\alpha_K-\alpha_N}}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{\frac{1-\alpha_N}{1-\alpha_K-\alpha_N}}}{\left[(\alpha_K/r)^{\alpha_K}(\alpha_N/w)^{1-\alpha_K}\right]^{\frac{1}{1-\alpha_N-\alpha_K}}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{\frac{1-\alpha_N}{1-\alpha_K-\alpha_N}}} \iff w = \frac{\bar{K}}{\bar{N}}\frac{\alpha_N}{\alpha_K}r \quad (77)$$

Now examine the labor market clearing condition, which reads:

$$\bar{N} = \left[(\alpha_K/r)^{\alpha_K}(\alpha_N/w)^{1-\alpha_K}\right]^{\frac{1}{1-\alpha_N-\alpha_K}}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{\frac{1-\alpha_N}{1-\alpha_K-\alpha_N}} \iff$$

$$r^{*,nd} = \alpha_K\left(\frac{(\bar{N})^{\alpha_N}}{(\bar{K})^{1-\alpha_K}}\right)\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N} \quad (78)$$

Plugging this back in to the expression generated above yields the following solution for the equilibrium wage in the economy without data:

$$w^{nd,*} = \alpha_N\left(\frac{(\bar{K})^{\alpha_K}}{(\bar{N})^{1-\alpha_N}}\right)\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N} \quad (79)$$

<u>Part 2d:</u> Total output in the economy without data

The output of any firm is given by:

$$Y^{nd}(A_2) = A_2\left(\bar{K}^{nd}\right)^{\alpha_K}\left(N^{nd}(A_2)\right)^{\alpha_N} =$$

$$(A_2)^{\frac{1}{1-\alpha_N}}(\alpha_K/r)^{\frac{\alpha_K}{1-\alpha_K-\alpha_N}}(\alpha_N/w)^{\frac{\alpha_N}{1-\alpha_K-\alpha_N}}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{\frac{\alpha_K}{1-\alpha_K-\alpha_N}} \quad (80)$$

Thus, aggregate output in the economy without data is given by:

$$\bar{Y}^{nd} = (\alpha_K/r)^{\frac{\alpha_K}{1-\alpha_K-\alpha_N}}(\alpha_N/w)^{\frac{\alpha_N}{1-\alpha_K-\alpha_N}}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{\frac{1-\alpha_N}{1-\alpha_K-\alpha_N}} \quad (81)$$

Plugging in the factor prices in the economy without data yields:

$$\bar{Y}^{nd} = (\bar{N})^{\alpha_N}(\bar{K})^{\alpha_K}\left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}}\right]\right)^{1-\alpha_N} \quad (82)$$

<u>Part 2e:</u> Capital and labor shares in the two economies.

Aggregate capital income in the economy with data is given by:

$$I^{K,d} = \alpha_K (\bar{N})^{\alpha_N} (\bar{K})^{\alpha_K} \left[\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}} \right] \right]^{1-\alpha_K-\alpha_N} \tag{83}$$

Aggregate labor income in the economy with data is given by:

$$I^{N,d} = \alpha_N (\bar{K})^{\alpha_K} (\bar{N})^{\alpha_N} \left[\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}} \right] \right]^{1-\alpha_K-\alpha_N} \tag{84}$$

Total production in the economy with data is:

$$Y^d = \left(\bar{N}\right)^{\alpha_N} \left(\bar{K}\right)^{\alpha_K} \left[\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_K-\alpha_N}} \right] \right]^{1-\alpha_K-\alpha_N} \tag{85}$$

Thus, the capital share ($S^{K,d}$) and the labor share ($S^{N,d}$) satisfy $S^{K,d} = \alpha_K$ and $S^{N,d} = \alpha_N$.

Now consider the economy without data. Aggregate capital income in this economy is:

$$I^{K,nd} = \alpha_K (\bar{N})^{\alpha_N} (\bar{K})^{\alpha_K} \left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}} \right] \right)^{1-\alpha_N} \tag{86}$$

Agregate labor income in the economy without data is:

$$I^{N,nd} = \alpha_N \left(\frac{(\bar{K})^{\alpha_K}}{(\bar{N})^{1-\alpha_N}} \right) \left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}} \right] \right)^{1-\alpha_N} \bar{N} = \alpha_N (\bar{N})^{\alpha_N} (\bar{K})^{\alpha_K} \left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}} \right] \right)^{1-\alpha_N} \tag{87}$$

Total output is given by:

$$\bar{Y}^{nd} = (\bar{N})^{\alpha_N} (\bar{K})^{\alpha_K} \left(\mathbb{E}\left[(A_2)^{\frac{1}{1-\alpha_N}} \right] \right)^{1-\alpha_N} \tag{88}$$

Thus, the capital share ($S^{K,nd}$) and the labor share ($S^{N,nd}$) in the economy without data also satisfy $S^{K,d} = \alpha_K$ and $S^{N,d} = \alpha_N$. ■

Proof of Proposition 2:

Part 1: Calculation of aggregate factor demands.

Part 1a: The economy with data.

Suppose that the demand (expressed as a price function) is given by $A_2 - \eta Q_i$, where the production function is given by $y(K_2, N_2) = (K_2)^{0.5} + (N_2)^{0.5}$.

In period 2, the firm thus chooses its labor demand to maximize the following:

$$A_2\big((K_2)^{0.5} + (N_2)^{0.5}\big) - \eta\big((K_2)^{0.5} + (N_2)^{0.5}\big)^2 - wN_2 \tag{89}$$

The first-order condition pinning down the optimal labor choice in period 2 reads:

$$0.5A_2(N_2)^{-0.5} - \eta(N_2)^{-0.5}\big((K_2)^{0.5} + (N_2)^{0.5}\big) - w = 0 \iff$$

$$N^*(A_2, K_2) = \left(\frac{0.5A_2 - \eta\sqrt{K_2}}{\eta + w}\right)^2 \tag{90}$$

Now consider the optimal capital choice of a firm in the economy with data. This firm maximizes the following through choice of K_2:

$$A_2\big((K_2)^{0.5} + (N^*(A_2, K_2))^{0.5}\big) - \eta\big((K_2)^{0.5} + (N^*(A_2, K_2))^{0.5}\big)^2 - rK_2 \tag{91}$$

Taking the first-order condition and plugging in the optimal labor choice yields:

$$0.5A_2(K_2)^{-0.5} - \eta(K_2)^{-0.5}\left((K_2)^{0.5} + \frac{0.5A_2 - \eta\sqrt{K_2}}{\eta + w}\right) - r = 0 \iff$$

$$K^d(A_2) = \left(\frac{0.5A_2 w}{\eta(r + w) + rw}\right)^2 \tag{92}$$

As a result, the aggregate capital stock in the economy with data is given by:

$$\bar{K}^d = \left(\frac{0.5w}{\eta(r + w) + rw}\right)^2 \mathbb{E}[(A_2)^2] \tag{93}$$

The individual-firm labor demand in the economy with data is given by:

$$N^d(A_2) = N^*(A_2, K^d(A_2)) = \left(\frac{0.5A_2 - \eta(K^d(A_2))^{0.5}}{\eta + w}\right)^2 =$$

$$\left(\frac{1}{\eta + w}\right)^2 \left[0.5A_2 - \eta\left(\frac{0.5A_2 w}{\eta(r + w) + rw}\right)\right]^2 = \left(\frac{0.5}{\eta + w}\right)^2 \left[A_2 - \eta\left(\frac{A_2 w}{\eta(r + w) + rw}\right)\right]^2 \tag{94}$$

The aggregate labor demand in the economy with data is given by:

$$\bar{N}^d = \mathbb{E}[N^d(A_2)] = \left(\frac{0.5}{\eta + w}\right)^2 \mathbb{E}\left[\left(A_2 - \left(\frac{\eta w}{\eta(r+w)+rw}\right)A_2\right)^2\right] =$$

$$\left(\frac{0.5}{\eta + w}\right)^2 \left[\mathbb{E}\big[(A_2)^2\big] + \left(-2 + \frac{\eta w}{\eta(r+w)+rw}\right)\left(\frac{\eta w}{\eta(r+w)+rw}\right)\mathbb{E}\big[(A_2)^2\big]\right] \tag{95}$$

<u>Part 1b:</u> The economy without data.

The optimal labor demand in the economy without data takes the same form as above. Thus, any firm maximizes the following through its choice of capital K^{nd}:

$$\Pi^{nd} = \int \Bigg[\underbrace{A_2\big((K^{nd})^{0.5} + (N^*(A_2, K^{nd}))^{0.5}\big) - \eta\big((K^{nd})^{0.5} + (N^*(A_2, K^{nd}))^{0.5}\big)^2 - rK^{nd}}_{\Pi(A_2, K^{nd})} \Bigg] dA_2 \tag{96}$$

Thus, the first-order condition reads:

$$\frac{\partial \Pi^{nd}}{\partial K^{nd}} = \int \frac{\partial \Pi(A_2, K^{nd}, N_2)}{\partial K^{nd}} dA_2 = 0 \tag{97}$$

This holds by the envelope theorem. Note that:

$$\frac{\partial \Pi(A_2, K^{nd}, N_2)}{\partial K^{nd}} = 0.5A_2(K^{nd})^{-0.5} - \eta(K^{nd})^{-0.5}\big((K^{nd})^{0.5} + (N^*(A_2, K^{nd}))^{0.5}\big) - r =$$

$$0.5A_2\left(\frac{w}{w+\eta}\right)(K^{nd})^{-0.5} - \frac{r(\eta + w) + \eta w}{\eta + w} \tag{98}$$

Plugging this back into the first-order condition allows one to pin down the optimal capital stock in the economy without data:

$$\frac{\partial \Pi^{nd}}{\partial K^{nd}} = \mathbb{E}\left[0.5A_2\left(\frac{w}{w+\eta}\right)(K^{nd})^{-0.5} - \frac{r(\eta + w) + \eta w}{\eta + w}\right] = 0 \iff$$

$$\bar{K}^{nd} = \left(\frac{0.5w}{rw + \eta(r+w)}\right)^2 \big(\mathbb{E}[A_2]\big)^2 \tag{99}$$

Using this, we can calculate the individual-firm labor supply in the economy without data:

$$N^{nd}(A_2) = N^*(A_2, \bar{K}^{nd}) = \left(\frac{0.5A_2 - \eta(\bar{K}^{nd})^{0.5}}{\eta + w}\right)^2 =$$

$$\left(\frac{0.5}{\eta+w}\right)^2\left((A_2)^2 - 2\left(\frac{\eta w}{rw+\eta(r+w)}\right)A_2\mathbb{E}[A_2] + \left(\frac{\eta w}{rw+\eta(r+w)}\right)^2\big(\mathbb{E}[A_2]\big)^2\right) \quad (100)$$

The aggregate labor demand in the economy without data is given by:

$$\bar{N}^{nd} = \left(\frac{0.5}{\eta+w}\right)^2\left[\mathbb{E}\big[(A_2)^2\big] + \underbrace{\left(-2 + \frac{\eta w}{rw+\eta(r+w)}\right)}_{<0}\left(\frac{\eta w}{rw+\eta(r+w)}\right)\big(\mathbb{E}[A_2]\big)^2\right] \quad (101)$$

Part 2: Comparison of aggregate factor demands.

We begin by comparing the aggregate labor demands. Aggregate labor demand in the economy with data is smaller if and only if

$$\bar{N}^d(w,r) < \bar{N}^{nd}(w,r) \iff$$

$$\left(-2+\frac{\eta w}{\eta(r+w)+rw}\right)\mathbb{E}\big[(A_2)^2\big] < \left(-2+\frac{\eta w}{rw+\eta(r+w)}\right)\big(\mathbb{E}[A_2]\big)^2 \iff \mathbb{E}\big[(A_2)^2\big] > \big(\mathbb{E}[A_2]\big)^2$$

The last inequality holds by construction. Thus, aggregate labor demand will be (for a given combination of wage and interest rate) lower in the economy with data.

Now consider the aggregate demand for capital. For any given wage and interest rate, the aggregate capital demand in the economy with data will be larger. This holds because:

$$\bar{K}^d(w,r) > \bar{K}^{nd}(w,r) \iff \mathbb{E}[(A_2)^2] > \big(\mathbb{E}[A_2]\big)^2$$

■

Proof of Proposition 3:

Suppose that the demand (expressed as a price function) is given by $p(Q_2) = A_2 - \eta Q_2$, where A_2 is the firm-level profitability. Suppose that the production function is $y(K_2, N_2) = (K_2)^{0.5}(N_2)^{0.5}$. Revenue is thus $p(Q_2)Q_2 = A_2Q_2 - \eta(Q_2)^2$. Total profits are given by:

$$A_2\big((K_2)^{0.5}(N_2)^{0.5}\big) - \eta\big((K_2)^{0.5}(N_2)^{0.5}\big)^2 - wN_2 - rK_2 \quad (102)$$

Part 1: The economy with data

Part 1a: Calculating the second-period labor demand.

In period 2, the firm chooses its labor demand to optimize the following function:

$$\Pi^2(N_2|K_2, A_2) = A_2\big((K_2)^{0.5}(N_2)^{0.5}\big) - \eta K_2 N_2 - wN_2 - rK_2 \tag{103}$$

The first-order condition that pins down the optimal labor choice reads:

$$0.5A_2(K_2)^{0.5}(N_2)^{-0.5} - \eta K_2 - w = 0 \iff N^*(A_2, K_2) = \left(\frac{0.5A_2(K_2)^{0.5}}{\eta K_2 + w}\right)^2 \tag{104}$$

Part 1b: Aggregate factor demands in the economy with data.

Now consider the optimization problem of a firm in the economy with data. Through its choice of capital, this firm maximizes the following profit function:

$$\Pi^d(K_2|A_2) = A_2\big((K_2)^{0.5}(N^*(\cdot))^{0.5}\big) - \eta\big((K_2)^{0.5}(N^*(\cdot))^{0.5}\big)^2 - wN^*(A_2, K_2) - rK_2 \tag{105}$$

The first-order condition on the optimal capital choice yields:

$$K^d(A_2) = \frac{0.5A_2\sqrt{w/r} - w}{\eta} \tag{106}$$

Plugging this into the expression for the optimal labor input yields an expression for the firm-level labor demand in the economy with data:

$$N^d(A_2) = N^*(A_2, K^d(A_2)) = \left(\frac{0.5A_2}{\eta K^d(A_2) + w}\right)^2 K^d(A_2) =$$

$$\left(\frac{0.5A_2}{0.5A_2\sqrt{w/r}}\right)^2\left(\frac{0.5A_2\sqrt{w/r} - w}{\eta}\right) = \frac{0.5A_2\sqrt{w/r} - w}{\eta(w/r)} \tag{107}$$

The aggregate capital demand in the economy with data is given by:

$$\bar{K}^d = \mathbb{E}\left[\frac{0.5A_2\sqrt{w/r} - w}{\eta}\right] = \frac{(0.5\sqrt{w/r})\mathbb{E}[A_2] - w}{\eta} \tag{108}$$

The aggregate labor demand in the economy with data is given by:

$$\bar{N}^d = \mathbb{E}\left[\frac{0.5A_2\sqrt{w/r} - w}{\eta(w/r)}\right] \iff \bar{N}^d = \frac{(0.5\sqrt{w/r})\mathbb{E}[A_2] - w}{\eta(w/r)} \tag{109}$$

Part 1c: Total production in the economy with data.

Now I calculate total production in the economy with data. To do so, I begin by calculating output (revenue) of any individual firm, which is given by:

$$Y^d(A_2) = A_2\big((K^d(A_2))^{0.5}(N^d(A_2))^{0.5}\big) - \eta\big((K^d(A_2))^{0.5}(N^d(A_2))^{0.5}\big)^2 =$$

$$\frac{1}{\eta}\left[0.25(A_2)^2 + 0.5A_2\big(\sqrt{w/r}\big)r - 0.5A_2\big(\sqrt{r/w}\big)w - rw\right] = (1/\eta)\big[0.25(A_2)^2 - rw\big] \tag{110}$$

Thus, aggregate output in the economy with data is given by:

$$\bar{Y}^d = (1/\eta)\big[0.25\mathbb{E}\big[(A_2)^2\big] - rw\big] \tag{111}$$

Part 1d: Factor prices in the economy with data.

Now let's calculate the factor prices in the economy with data. These are pinned down by the following two market clearing conditions:

$$\bar{K} = \frac{(0.5\sqrt{w/r})\mathbb{E}[A_2] - w}{\eta} \quad ; \quad \bar{N} = \frac{(0.5\sqrt{w/r})\mathbb{E}[A_2] - w}{\eta(w/r)} \tag{112}$$

Taking the ratio of the market clearing conditions yields:

$$\frac{\bar{K}}{\bar{N}} = \frac{1}{\frac{1}{\frac{w}{r}}} \iff \frac{\bar{K}}{\bar{N}} = \frac{w}{r} \iff w = \frac{\bar{K}}{\bar{N}}r \tag{113}$$

Plugging this into the market clearing condition for capital yields that:

$$\bar{K} = \frac{(0.5\sqrt{w/r})\mathbb{E}[A_2] - w}{\eta} \iff \eta\bar{K} = 0.5\left(\frac{\bar{K}}{\bar{N}}\right)^{0.5}\mathbb{E}[A_2] - \frac{\bar{K}}{\bar{N}}r \iff$$

$$r^{d,*} = 0.5\left(\frac{\bar{N}}{\bar{K}}\right)^{0.5}\mathbb{E}[A_2] - \eta\bar{N} \tag{114}$$

Analogously, we obtain:

$$w^{d,*} = 0.5\left(\frac{\bar{K}}{\bar{N}}\right)^{0.5} \mathbb{E}[A_2] - \eta\bar{K} \tag{115}$$

Total labor income in the economy with data is given by:

$$I^{d,N} = 0.5\mathbb{E}[A_2]\big(\bar{K}\big)^{0.5}\big(\bar{N}\big)^{0.5} - \eta\bar{K}\bar{N} \tag{116}$$

Part 2: The economy without data.

<u>Part 2a:</u> Aggregate factor demands in the economy without data.

As in the economy without data, the statically optimal labor choice solves:

$$N_2^*(A_2, K_2) = \left(\frac{0.5A_2(K_2)^{0.5}}{\eta K_2 + w}\right)^2 \tag{117}$$

Every firm in the economy without data maximizes the following through choice of $\bar{K}^{nd}$.

$$\Pi^{nd}(K^{nd}) = \mathbb{E}\left[A_2\big((K^{nd})^{0.5}(N^*(\cdot))^{0.5}\big) - \eta\big((K^{nd})^{0.5}(N^*(\cdot))^{0.5}\big)^2 - wN^*(A_2, K_2) - rK^{nd}\right] \tag{118}$$

The associated first-order condition reads:

$$\frac{\partial\Pi^{nd}}{\partial K^{nd}} = \mathbb{E}\left[0.5A_2(K^{nd})^{-0.5}(N^*(A_2, K^{nd}))^{0.5} - \eta(N^*(A_2, K^{nd})) - r\right] = 0 \iff$$

$$\bar{K}^{nd} = \frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta} \tag{119}$$

The firm-level labor demand in the economy without data is:

$$N^{nd}(A_2) = \left(\frac{0.5A_2(K^{nd})^{0.5}}{\eta K^{nd} + w}\right)^2 =$$

$$\left[\frac{0.5}{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r})}\right]^2\left(\frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta}\right)\big(A_2\big)^2 \tag{120}$$

Thus, the aggregate labor demand in the economy without data is given by:

$$\bar{N}^{nd} = \frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta(w/r)} \tag{121}$$

<u>Part 2b</u>: Aggregate output in the economy without data.

The firm level output in the economy without data is given by:

$$Y^{nd}(A_2) = A_2\big((\bar{K}^{nd})^{0.5}(N^{nd}(A_2))^{0.5}\big) - \eta\big((\bar{K}^{nd})^{0.5}(N^{nd}(A_2))^{0.5}\big)^2 \tag{122}$$

Thus, we have:

$$Y^{nd}(A_2) = A_2\Bigg[\Bigg(\frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta}\Bigg)\Bigg(\frac{0.5}{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r})}\Bigg)(A_2)\Bigg] -$$
$$\eta\Bigg[\Bigg(\frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta}\Bigg)\Bigg(\frac{0.5}{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r})}\Bigg)(A_2)\Bigg]^2 \tag{123}$$

Thus, the aggregate output in the economy without data is given by:

$$\bar{Y}^{nd} = \mathbb{E}[Y^{nd}(A_2)] =$$
$$\Bigg(\frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta}\Bigg)\Bigg(\frac{0.5}{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r})}\Bigg)\mathbb{E}[(A_2)^2] -$$
$$\eta\Bigg(\frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta}\Bigg)^2\Bigg(\frac{0.5}{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r})}\Bigg)^2\mathbb{E}[(A_2)^2]$$
$$=$$
$$(1/\eta)\big[0.25\big(\mathbb{E}[(A_2)^2]\big) - rw\big] \tag{124}$$

<u>Part 2c:</u> Equilibrium in the economy without data

Now, we calculate the market clearing factor prices in the economy with data. These are pinned down by the following two market clearing conditions:

$$\bar{N} = \frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta(w/r)} \quad ; \quad \bar{K} = \frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta} \tag{125}$$

Taking the ratio of the two first-order conditions yields:

$$\frac{\bar{N}}{\bar{K}} = \frac{\frac{0.5\left(\mathbb{E}[(A_2)^2]\right)^{0.5}(\sqrt{w/r})-w}{\eta(w/r)}}{\frac{0.5\left(\mathbb{E}[(A_2)^2]\right)^{0.5}(\sqrt{w/r})-w}{\eta}} \iff w = \frac{\bar{K}}{\bar{N}}r \tag{126}$$

Plugging this into the market clearing condition for capital yields the following:

$$\bar{K} = \frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}(\sqrt{w/r}) - w}{\eta} = \frac{0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\big(\bar{K}/\bar{N}\big)^{0.5} - \frac{\bar{K}}{\bar{N}}r}{\eta} \iff$$

$$r^{nd,*} = 0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\left(\frac{\bar{N}}{\bar{K}}\right)^{0.5} - \eta\bar{N} \tag{127}$$

Similarly, one can calculate the equilibrium wage rate in the economy without data. This is given by:

$$w^{nd,*} = \left(\frac{\bar{K}}{\bar{N}}\right)\left(0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\left(\frac{\bar{N}}{\bar{K}}\right)^{0.5} - \eta\bar{N}\right) \iff$$

$$w^{nd,*} = 0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\left(\frac{\bar{K}}{\bar{N}}\right)^{0.5} - \eta\bar{K} \tag{128}$$

This means that the aggregate labor income in the economy without data is given by:

$$I^{nd,N} = 0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\big(\bar{K}\big)^{0.5}\big(\bar{N}\big)^{0.5} - \eta\bar{K}\bar{N} \tag{129}$$

Part 3: Comparison of aggregate labor income and aggregate capital income.

The wage and labor income in the economy with data are given by:

$$w^{d,*} = 0.5\left(\frac{\bar{K}}{\bar{N}}\right)^{0.5}\mathbb{E}[A_2] - \eta\bar{K} \quad ; \quad I^{d,N} = 0.5\mathbb{E}[A_2]\big(\bar{K}\big)^{0.5}\big(\bar{N}\big)^{0.5} - \eta\bar{K}\bar{N} \tag{130}$$

The wage and labor income in the economy without data is given by:

$$w^{nd,*} = 0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\left(\frac{\bar{K}}{\bar{N}}\right)^{0.5} - \eta\bar{K} \tag{131}$$

$$I^{nd,N} = 0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\big(\bar{K}\big)^{0.5}\big(\bar{N}\big)^{0.5} - \eta\bar{K}\bar{N} \tag{132}$$

Since $\mathbb{E}[(A_2)^2] > \big(\mathbb{E}[A_2]\big)^2$, aggregate labor income is higher in the economy without data.

Now I consider the aggregate capital income. Recall that the interest rate and capital income in the economy with data are given by:

$$r^{d,*} = 0.5\left(\frac{\bar{N}}{\bar{K}}\right)^{0.5}\mathbb{E}[A_2] - \eta\bar{N} \quad ; \quad I^{d,K} = 0.5\big(\bar{K}\big)^{0.5}\big(\bar{N}\big)^{0.5}\mathbb{E}[A_2] - \eta\bar{N}\bar{K} \tag{133}$$

The interest rate and the capital income in the economy without data were given by:

$$r^{nd,*} = 0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\left(\frac{\bar{N}}{\bar{K}}\right)^{0.5} - \eta\bar{N} \quad ; \quad I^{nd,K} = 0.5\big(\mathbb{E}[(A_2)^2]\big)^{0.5}\big(\bar{K}\big)^{0.5}\big(\bar{N}\big)^{0.5} - \eta\bar{N}\bar{K} \tag{134}$$

This means that capital income in the economy with data is also smaller.

For a given wage and interest rate, output in the two economies are the same. Because both factor prices in the economy with data are smaller, total output is larger.

■

Proof of Proposition 4 :

Part 1: Perfect substitutes production function.

Part 1a: For any combination of factor prices, skilled labor demand is higher in the economy with data.

Consider any (w, r). First, I show that $H_1^d(A_1, \mathbb{E}[A_2|A_1]; w) = H_1^{nd}(A_1, w)$ holds true for any A_1. Thereafter, I show that $H_1^d(A_1, A_2; w)$ is a convex function in A_2. Together, these arguments imply that the expected skilled labor demand of any firm with a fixed A_1, which I call $\bar{H}_1^d(A_1, w)$, is higher in the economy with data because:

$$\bar{H}_1^d(A_1, w) = \mathbb{E}_\epsilon\big[H_1^d(A_1, A_1 + \epsilon; w)\big] \underbrace{>}_{\text{Point (2)}} H_1^d(A_1, A_1 + \mathbb{E}[\epsilon]; w) \underbrace{=}_{\text{Point (1)}} H_1^{nd}(A_1; w) \tag{135}$$

To see why point (1) holds true, note that the optimal unskilled labor input choice is independent of the skilled labor input choice. Thus, $H_1^d(A_1, A_1 + \mathbb{E}[\epsilon]; w_H)$ is a maximizer of the following objective function:

$$A_1(H_1)^{\alpha_H} - w_H H_1 + \big(\mathbb{E}[A_2|A_1](H_1)^{\alpha_H} - w_H H_1\big) \tag{136}$$

Moreover, $H_1^{nd}(A_1, w_H)$ maximizes the following objective function:

$$A_1(H_1)^{\alpha_H} - w_H H_1 + \Bigg(\underbrace{\int_{\underline{\epsilon}}^{\bar{\epsilon}} (A_1 + \epsilon)(H_1)^{\alpha_H} dG(\epsilon)}_{=\mathbb{E}[A_2|A_1](H_1)^{\alpha_H}} - w_H H_1 \Bigg) \tag{137}$$

Thus, the two objective functions take the same form and the first result follows.

To see why point (2) holds true, note that $H_1^d(A_1, A_2; w_H)$ is a convex function in A_2, because it can be computed as follows:

$$H_1^d(A_2; w) = \left(\frac{\alpha_H(A_1 + A_2)}{2w_H} \right)^{1/(1-\alpha_H)} \tag{138}$$

This is a convex function in A_2, since it is increasing in A_2 and $\frac{1}{1-\alpha_H} > 1 \iff 1 > 1 - \alpha_H$. Aggregate first period high skilled labor demand in the economy with data is:

$$\bar{H}_1^d(w) = \int_{\underline{A}_1}^{\bar{A}_1} \bar{H}_1^d(A_1; w) dF(A_1) \tag{139}$$

Aggregate first-period high skilled labor demand in the economy without data is:

$$\bar{H}_1^{nd}(w) = \int_{\underline{A}_1}^{\bar{A}_1} H_1^{nd}(A_1; w) dF(A_1) \tag{140}$$

Because $\bar{H}_1^d(A_1; w) > \bar{H}_1^{nd}(A_1; w)$ holds for any A_1, we thus have $\bar{H}_1^d(w) > \bar{H}_1^{nd}(w)$.

Because adjustment costs are prohibitively high, the skilled labor demand of any firm must be identical in both periods. Thus, the aggregate first period high skilled labor demand in the economy with data will be given by:

$$\bar{H}^d(w) = 2 \int_{\underline{A}_1}^{\bar{A}_1} \bar{H}_1^d(A_1; w) dF(A_1) \tag{141}$$

Total aggregate high skilled labor demand in the economy without data is:

$$\bar{H}^{nd}(w) = 2 \int_{\underline{A}_1}^{\bar{A}_1} H_1^{nd}(A_1; w) dF(A_1) \tag{142}$$

Because $\bar{H}_1^d(A_1; w) > \bar{H}_1^{nd}(A_1; w)$ holds for any A_1, we thus have $\bar{H}^d(w) > \bar{H}^{nd}(w)$.

<u>Part 1b:</u> For every combination of factor prices, unskilled labor demand is the same in both

economies.

The unskilled labor input choices are independent of the skilled labor input choices. Given that there are no adjustment costs for unskilled labor, these will be set to be the statically optimal unskilled labor input choices in either economy, which are given by:

$$L^*(A_t, w) = \left(\frac{\alpha_L A_t}{w_L}\right)^{1/(1-\alpha_L)} \tag{143}$$

Thus, aggregate unskilled labor demand is the same in both economies.

<u>Part 1c:</u> Equilibrium wages

Given that unskilled and skilled labor demands are independent, the equilibrium w_L^* only depends on the aggregate unskilled labor demand. This is the same in both economies, so this wage w_L^* is the same. In the economy with data, high skilled labor demand is higher, so w_H^* must be higher and the result follows.

Part 2: Cobb-Douglas production function

<u>Part 2a:</u> Aggregate labor demands in the economy with data.

Conditional on H_t, the low skilled labor choice (which is not subject to adjustment costs) must maximize the following objective in either economy:

$$A_t(H_t)^{\alpha_H}(L_t)^{\alpha_L} - w_L L_t \tag{144}$$

Thus, the optimal low skilled labor choice in periods $t \in \{1, 2\}$ satisfies:

$$\alpha_L A_t(H_t)^{\alpha_H}(L_t)^{\alpha_L - 1} - w_L = 0 \iff L_t^*(H_t) = \left(\frac{\alpha_L A_t(H_t)^{\alpha_H}}{w_L}\right)^{\frac{1}{1-\alpha_L}} \tag{145}$$

We have supposed that $\kappa \to \infty$, so any firm will always set its second-period skilled labor input equal to H_1. In the economy with data, the total profits of any firm are thus given by:

$$A_1(H_1)^{\alpha_H}\left(L_1^*(H_1)\right)^{\alpha_L} - w_H H_1 - w_L L_1^*(H_1) + \left[A_2(H_1)^{\alpha_H}\left(L_2^*(H_1)\right)^{\alpha_L} - w_H H_1 - w_L L_2^*(H_1)\right] \tag{146}$$

The first-order condition on H_1 reads:

$$\alpha_H A_1(H_1)^{\alpha_H-1}\big(L_2^*(H_1)\big)^{\alpha_L} - w_H + \Big[\alpha_H A_2(H_1)^{\alpha_H-1}\big(L_2^*(H_1)\big)^{\alpha_L} - w_H\Big] = 0 \tag{147}$$

Plugging in the optimal unskilled labor supply yields:

$$\alpha_H A_1(H_1)^{\alpha_H-1}\bigg(\frac{\alpha_L A_1(H_1)^{\alpha_H}}{w_L}\bigg)^{\frac{\alpha_L}{1-\alpha_L}} - w_H + \bigg[\alpha_H A_2(H_1)^{\alpha_H-1}\bigg(\frac{\alpha_L A_2(H_1)^{\alpha_H}}{w_L}\bigg)^{\frac{\alpha_L}{1-\alpha_L}} - w_H\bigg] = 0$$

$$\Longleftrightarrow$$

$$H_1^d(A_1,A_2) = \Big((\alpha_H/2w_H)^{1-\alpha_L}(\alpha_L/w_L)^{\alpha_L}\Big)^{\frac{1}{1-\alpha_L-\alpha_H}}\Big((A_1)^{\frac{1}{1-\alpha_L}} + (A_2)^{\frac{1}{1-\alpha_L}}\Big)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}} \tag{148}$$

In both periods, any firm will set the same skilled labor demand. Thus, total aggregate skilled labor demand in the economy with data is given by the following:

$$\bar{H}^d = 2\Big((\alpha_H/2w_H)^{1-\alpha_L}(\alpha_L/w_L)^{\alpha_L}\Big)^{\frac{1}{1-\alpha_L-\alpha_H}}\mathbb{E}_{A_1,\epsilon}\bigg[\Big((A_1)^{\frac{1}{1-\alpha_L}} + (A_2)^{\frac{1}{1-\alpha_L}}\Big)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}}\bigg] \tag{149}$$

Now consider aggregate unskilled labor demand in the economy with data. To calculate this, we calculate the labor demand of any individual firm:

$$L^d(A_1,A_2) = \bigg(\frac{\alpha_L A_1(H_1)^{\alpha_H}}{w_L}\bigg)^{\frac{1}{1-\alpha_L}} + \bigg(\frac{\alpha_L A_2(H_1)^{\alpha_H}}{w_L}\bigg)^{\frac{1}{1-\alpha_L}} =$$

$$\big(\alpha_L/w_L\big)^{\frac{1}{1-\alpha_L}}\bigg[\Big((\alpha_H/2w_H)^{1-\alpha_L}(\alpha_L/w_L)^{\alpha_L}\Big)^{\frac{1}{1-\alpha_L-\alpha_H}}\bigg]^{\frac{\alpha_H}{1-\alpha_L}}\Big((A_1)^{\frac{1}{1-\alpha_L}} + (A_2)^{\frac{1}{1-\alpha_L}}\Big)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}} \tag{150}$$

Thus, aggregate unskilled labor demand is given by the following:

$$\bar{L}^d = \bigg(\frac{\alpha_L}{w_L}\bigg)^{\frac{1}{1-\alpha_L}}\Bigg[\Bigg(\bigg(\frac{\alpha_H}{2w_H}\bigg)^{1-\alpha_L}\bigg(\frac{\alpha_L}{w_L}\bigg)^{\alpha_L}\Bigg)^{\frac{1}{1-\alpha_L-\alpha_H}}\Bigg]^{\frac{\alpha_H}{1-\alpha_L}}\mathbb{E}_{A_1,\epsilon}\bigg[\Big((A_1)^{\frac{1}{1-\alpha_L}} + (A_2)^{\frac{1}{1-\alpha_L}}\Big)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}}\bigg] \tag{151}$$

Part 2b: Aggregate labor demands in the economy without data.

By similar calculations, one can obtain the skilled labor demand in the economy without data. For a given A_1, a firm in the economy with data maximizes the following objective function through choice of H_1:

$$A_1(H_1)^{\alpha_H} L_1^{\alpha_L} - w_H H_1 - w_L L_2^*(H_1) + \mathbb{E}_\epsilon \Big[A_2(H_1)^{\alpha_H} \big(L_2^*(H_1)\big)^{\alpha_L} - w_H H_1 - w_L L_2^*(H_1) \Big] \tag{152}$$

The first-order condition reads:

$$\alpha_H A_1(H_1)^{\alpha_H - 1} \left(\frac{\alpha_L A_1 (H_1)^{\alpha_H}}{w_L} \right)^{\frac{\alpha_L}{1-\alpha_L}} - w_H + \mathbb{E}_\epsilon \left[\alpha_H A_2 (H_1)^{\alpha_H - 1} \left(\frac{\alpha_L A_2 (H_1)^{\alpha_H}}{w_L} \right)^{\frac{\alpha_L}{1-\alpha_L}} - w_H \right] = 0$$

$$\Longleftrightarrow$$

h re

$$\bar{H}_1^{nd}(A_1) = \Big((\alpha_H/2w_H)^{1-\alpha_L} (\alpha_L/w_L)^{\alpha_L} \Big)^{\frac{1}{1-\alpha_L-\alpha_H}} \Big((A_1)^{\frac{1}{1-\alpha_L}} + \mathbb{E}_\epsilon \big[(A_2)^{\frac{1}{1-\alpha_L}} \big] \Big)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}} \tag{153}$$

The skilled labor input of any firm will be the same in both periods. Thus, aggregate skilled labor input is given by:

$$\bar{H}^{nd} = 2\mathbb{E}_{A_1,\epsilon} \Bigg[\Big((\alpha_H/2w_H)^{1-\alpha_L} (\alpha_L/w_L)^{\alpha_L} \Big)^{\frac{1}{1-\alpha_L-\alpha_H}} \Big((A_1)^{\frac{1}{1-\alpha_L}} + \mathbb{E}_\epsilon \big[(A_2)^{\frac{1}{1-\alpha_L}} \big] \Big)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}} \Bigg]$$

$$\Longleftrightarrow$$

$$\bar{H}^{nd} = 2\Big((\alpha_H/2w_H)^{1-\alpha_L} (\alpha_L/w_L)^{\alpha_L} \Big)^{\frac{1}{1-\alpha_L-\alpha_H}} \mathbb{E}_{A_1} \Bigg[\Big((A_1)^{\frac{1}{1-\alpha_L}} + \mathbb{E}_\epsilon \big[(A_2)^{\frac{1}{1-\alpha_L}} \big] \Big)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}} \Bigg] \tag{154}$$

The unskilled labor demand of any firm in the economy without data is:

$$\left(\frac{\alpha_L A_1 (\bar{H}_1^{nd}(A_1))^{\alpha_H}}{w_L} \right)^{\frac{1}{1-\alpha_L}} + \left(\frac{\alpha_L A_2 (\bar{H}_1^{nd}(A_1))^{\alpha_H}}{w_L} \right)^{\frac{1}{1-\alpha_L}} \tag{155}$$

In the economy without data, aggregate unskilled labor demand is thus:

$$\bar{L}^{nd} = \mathbb{E}_{A_1,\epsilon} \left[\left(\frac{\alpha_L A_1 (\bar{H}_1^{nd}(A_1))^{\alpha_H}}{w_L} \right)^{\frac{1}{1-\alpha_L}} + \left(\frac{\alpha_L A_2 (\bar{H}_1^{nd}(A_1))^{\alpha_H}}{w_L} \right)^{\frac{1}{1-\alpha_L}} \right] =$$

$$\left(\frac{\alpha_L}{w_L} \right)^{\frac{1}{1-\alpha_L}} \left[\left(\left(\frac{\alpha_H}{2w_H} \right)^{1-\alpha_L} \left(\frac{\alpha_L}{w_L} \right)^{\alpha_L} \right)^{\frac{1}{1-\alpha_L-\alpha_H}} \right]^{\frac{\alpha_H}{1-\alpha_L}} \mathbb{E}_{A_1} \Bigg[\Big((A_1)^{\frac{1}{1-\alpha_L}} + \mathbb{E} \big[(A_2)^{\frac{1}{1-\alpha_L}} \big] \Big)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}} \Bigg] \tag{156}$$

Part 2c: The ratio of factor demands (and the skill premium) is the same in both economies.

The ratio of the factor demands in the economy without data is:

$$\frac{\bar{H}^{nd}(w)}{\bar{L}^{nd}(w)} =$$

$$\frac{2\Big((\alpha_H/2w_H)^{1-\alpha_L}(\alpha_L/w_L)^{\alpha_L}\Big)^{\frac{1}{1-\alpha_L-\alpha_H}}\mathbb{E}_{A_1}\left[\left((A_1)^{\frac{1}{1-\alpha_L}}+\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_L}}\big]\right)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}}\right]}{(\alpha_L/w_L)^{\frac{1}{1-\alpha_L}}\left[\Big((\alpha_H/w_H)^{1-\alpha_L}(\alpha_L/w_L)^{\alpha_L}\Big)^{\frac{1}{1-\alpha_L-\alpha_H}}\right]^{\frac{\alpha_H}{1-\alpha_L}}\mathbb{E}_{A_1}\left[\left((A_1)^{\frac{1}{1-\alpha_L}}+\mathbb{E}\big[(A_2)^{\frac{1}{1-\alpha_L}}\big]\right)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}}\right]}$$

$$\iff$$

$$\frac{\bar{H}^{nd}(w)}{\bar{L}^{nd}(w)} = \frac{\alpha_H}{\alpha_L}\frac{w_L}{w_H} \tag{157}$$

The ratio of factor demands in the economy with data is given by:

$$\frac{\bar{H}^{d}(w)}{\bar{L}^{d}(w)} =$$

$$\frac{2\Big((\alpha_H/2w_H)^{1-\alpha_L}(\alpha_L/w_L)^{\alpha_L}\Big)^{\frac{1}{1-\alpha_L-\alpha_H}}\mathbb{E}_{A_1,\epsilon}\left[\left((A_1)^{\frac{1}{1-\alpha_L}}+(A_2)^{\frac{1}{1-\alpha_L}}\right)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}}\right]}{(\alpha_L/w_L)^{\frac{1}{1-\alpha_L}}\left[\Big((\alpha_H/2w_H)^{1-\alpha_L}(\alpha_L/w_L)^{\alpha_L}\Big)^{\frac{1}{1-\alpha_L-\alpha_H}}\right]^{\frac{\alpha_H}{1-\alpha_L}}\mathbb{E}_{A_1,\epsilon}\left[\left((A_1)^{\frac{1}{1-\alpha_L}}+(A_2)^{\frac{1}{1-\alpha_L}}\right)^{\frac{1-\alpha_L}{1-\alpha_L-\alpha_H}}\right]}$$

$$\iff$$

$$\frac{\bar{H}^{d}(w)}{\bar{L}^{d}(w)} = \frac{\alpha_H}{\alpha_L}\frac{w_L}{w_H} \tag{158}$$

Because the ratio of factor demands must equal the ratio of labor supplies, the skill premium is uniquely pinned down in both economies and must be equal.

■

References

Abis, S., & Veldkamp, L. (2023). The changing economics of knowledge production.

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, *14*(4), 218–256.

Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms. *Algorithms*, *4*(5).

Arvai, K., & Mann, K. (2022). Consumption inequality in the digital age.

Autor, D. H., Donohue III, J. J., & Schwab, S. J. (2006). The costs of wrongful-discharge laws. *The Review of Economics and Statistics*, *88*(2), 211–231.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2023). Firm investments in artificial intelligence technologies and changes in workforce composition.

Bajari, P., Chernozhukov, V., Hortaçsu, A., & Suzuki, J. (2019). The impact of big data on firm performance: An empirical investigation. *AEA Papers and Proceedings*, *109*, 33–37.

Belianska, A. (2023). Macroeconomic uncertainty and capital-skill complementarity.

Bergemann, D., & Bonatti, A. (2022). Data, competition, and digital platforms.

Bessen, J., Denk, E., & Meng, C. (2022). The remainder effect: How automation complements labor quality.

Bloom, N. (2009). The impact of uncertainty shocks. *econometrica*, *77*(3), 623–685.

Brynjolfsson, E., Jin, W., & Wang, X. (2023). *Information technology, firm size, and industrial concentration* (Tech. Rep.). National Bureau of Economic Research.

Brynjolfsson, E., & McElheran, K. (2019). Data in action: data-driven decision making and predictive analytics in us manufacturing.

Bughin, J. (2023). Does artificial intelligence kill employment growth: the missing link of corporate ai posture. *Frontiers in Artificial Intelligence*, *6*.

Burdín, G., De Rosa, M., Vigorito, A., & Vilá, J. (2022). Falling inequality and the growing capital income share: Reconciling divergent trends in survey and tax data. *World Development*, *152*, 105783.

Charoenwong, B., Kimura, Y., Kwan, A., & Tan, E. (2022). Capital budgeting, uncertainty, and misallocation. *Uncertainty, and Misallocation*.

Chong, A. Y. L., Li, B., Ngai, E. W., Ch'Ng, E., & Lee, F. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations & Production Management*, *36*(4), 358–383.

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal*, *42*(1), 30–57.

Eden, M., & Gaggl, P. (2018). On the welfare implications of automation. *Review of Economic Dynamics*, *29*, 15–43.

Eeckhout, J., & Veldkamp, L. (2022). *Data and market power* (Tech. Rep.). National Bureau of Economic Research.

European Union. (2018). Addressing income inequality through development cooperation :a quick guide.

Farboodi, M., & Veldkamp, L. (2022). A model of the data economy.

Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, *38*(4), 1283–1318.

Fischer, M. M., Huber, F., & Pfarrhofer, M. (2021). The regional transmission of uncertainty shocks on income inequality in the united states. *Journal of Economic Behavior & Organization*, *183*, 887–900.

Galdon-Sanchez, J. E., Gil, R., & Uriz-Uharte, G. (2022). The value of information in competitive markets: The impact of big data on small and medium enterprises.

Ghaly, M., Anh Dang, V., & Stathopoulos, K. (2017). Cash holdings and labor heterogeneity: The role of skilled labor. *The Review of Financial Studies*, *30*(10), 3636–3668.

Glocker, C., & Piribauer, P. (2021). Digitalization, retail trade and monetary policy. *Journal of International Money and Finance*, *112*, 102340.

Gondhi, N. (2023). Rational inattention, misallocation, and the aggregate economy. *Journal of Monetary Economics*, *136*, 50–75.

Groh, C.-C., & Pfäuti, O. (2023). Data, business cycles, and the cyclical effectiveness of monetary policy.

Hubmer, J. (2023). The race between preferences and technoloy. *Econometrica*, *91*(1), 227–61.

Jaimovich, N., Saporta-Eksten, I., Siu, H. E., & Yedid-Levi, Y. (2020). The macroeconomics of automation: Data, theory, and policy analysis. *NBER working paper series*.

Kumar, S., Gorodnichenko, Y., & Coibion, O. (2023). The effect of macroeconomic uncertainty on firm decisions. *Econometrica*, *91*(4), 1297–1332.

Meier, M. (2020). Supply chain disruptions, time to build, and the business cycle.

Nalewaik, J., & Pinto, E. (2015). The response of capital goods shipments to demand over the business cycle. *Journal of Macroeconomics*, *43*, 62–80.

OECD. (2011). Society at a glance 2011: Oecd social indicators.

Oh, H., & Yoon, C. (2020). Time to build and the real-options channel of residential investment. *Journal of Financial Economics*, *135*(1), 255–269.

Rani, R., Kumar, S., Patil, R., & Pippal, S. (2023). A machine learning model for predicting innovation effort of firms. *Int. J. Electr. Comput. Eng*, *13*, 4633–4639.

Rognlie, M. (2016). Deciphering the fall and rise in the net capital share: accumulation or scarcity? *Brookings Papers on Economic Activity*, *2015*(1), 1–69.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690.

Statista. (2024). *Big data market size revenue forecast worldwide from 2011 to 2027.* Retrieved 2024-02-26, from `https://www.statista.com/statistics/254266/global-big-data-market-forecast/`

Theophilopoulou, A. (2022). The impact of macroeconomic uncertainty on inequality: An empirical study for the united kingdom. *Journal of Money, Credit and Banking*, *54*(4), 859–884.

UNDP. (2019). Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century. *Human Development Report 2019*.

Veldkamp, L., & Chung, C. (2019). Data and the aggregate economy. *Journal of Economic Literature*.

World Bank. (2021). *Gini index - by country.*