

THE UNIVERSITY OF CHICAGO

Social Science Division

**Analysis of Innovation and Inequality
with Microdata**

by

Koichi Onogi:

A paper submitted in partial fulfillment of the
requirements for the Master of Arts degree in
the Master of Arts in Computational Social Science

Faculty Advisor: Dr. Kanit (Ken) Kuevibulvanich

Preceptor: Dr. Sanja Miklin

June, 2023

Analysis of Innovation and Inequality with Microdata

Koichi Onogi

May 1, 2023

Abstract

This paper investigates the effects and relationship of innovation and industry concentration on top-income inequality in various sectors. It contends that innovation plays a critical role in the rise of income inequality and that comprehending the relationship between market mechanisms and inequality is necessary for elucidating the structures that underlie these economic phenomena. The study aims to establish links between economic factors related to innovation and different types of inequality while determining possible negative impacts of innovation. The study utilizes microdata from CPS IPUMS, patents, and an existing industry concentration dataset and shows that innovation might increase top income inequality. Additionally, the paper offers a theoretical explanation of how innovation could cause top-income inequality but not in general wage inequality

1 Introduction

It is widely recognized that serious inequality is a major problem facing the economy, but the role of innovation in rising inequality has generated little attention. As a basic principle of the market, firms and enterprises have always faced challenges to survive. They are often confronted with complex and high-risk choices and continuously experiment to ensure their sales in the future. This competition to survive often leads to innovation, which frequently enriches people's lives. Thus, innovation is considered a driver of economic activity, and for this reason, market competition plays a crucial role in our lives. However, given market pressures, there might be a negative side to innovation as well. When highly-productive companies gain their competitive advantages by providing products with lower prices and/or of higher quality than their competitors, they receive greater revenues/profits compared to their competitors. As a result, they seek to keep expanding or sustain their market positions by investing further towards increasing their productivity and realizing innovation (Porter, 1985)[15]. Gradually, this cycle of investment and innovation would lead top-income workers, most of whom work in innovation-related positions, to stand out and receive even higher wages than other workers. Moreover, this degree of inequality would augment as the industry market has a higher expected return by innovation because the cycle of innovation and inequality would be faster. Additionally, this tendency should correlate with revenue concentration by a few firms that can invest further under the winner-take-all mechanism. However, The exact mechanisms by which innovation-driven inequality grows have received surprisingly little attention in economics research.

Another gap in existing work on inequality has to do with the lack of attention to the many measures of inequality. Previous research has attempted to examine the factors of inequality by quantifying its components (Song et al. 2019[17], Hubmer and Restrepo, 2021[10]), or has worked to elucidate the potential mechanism through which current inequality has emerged, such as the connection with the decreasing labor share (Autor et al, 2020)[3].

However, it's important to note that inequality is measured by multiple methods, including Gini coefficients, the variance of natural logs, the top-median ratio, and the 90:10 ratio. Each measurement

grasps different aspects of dispersion in the distribution, thus providing different insights. Yet, most researchers do not fully focus on what kinds of wage inequality measurements are appropriate to use; they are mainly using those measurements merely to check the robustness of their models, which could be an issue in accurately grasping the results and connections of economic factors. Although the existing work has found some components to have more effects than others, there is not much research on how those components affect each inequality measurement separately.

This paper aims to clarify connections between economic factors related to innovation and different types of inequality while showcasing an approach to quantifying the negative impact of innovation. To do so, this study examines the relationship between innovation, investment, and industry concentration on top-income inequality across different industry sectors, by utilizing microdata from CPS IPUMS and other industry data, such as patents, investments in IT and R&D, and industry concentration. Although a tremendous amount of research has been conducted to investigate the causes and factors of these trends, as well as trends in economic inequality, not much research has successfully quantified the negative impact of those investments, innovation, and industry concentration. This paper also provides a theoretical explanation of how innovation could cause top-income inequality and not general inequality and contribute to determining the overall quantitative relevance of innovation in top-income inequality. In section 2, I introduce background literature. In section 3, I develop the theoretical background of this research. Section 4 describes the datasets this paper uses, and section 5 conducts the first empirical analyses using FE to examine the effects of the variables. Lastly, I conduct another supplemental analysis with covariates.

2 Related Literature

A few existing research already challenge to reveal that innovation has a correlation and partial causation on the top income inequality, utilizing aggregate datasets, which is different from this research. Aghion et al. (2018)[2], *Innovation and Top Income Inequality show positive correlations between measures of innovation and top income inequality*, while the correlations between innovation and broad inequality measures are insignificant. They also used IV 2SLS analysis to find partial causation of innovation on the top income inequality. I would like to utilize their Schumpeterian growth model. Also, Kwon et al. (2023)[13] *100 Years of Rising Corporate Concentration*, with historical data on market concentration for a century, provide evidence that market concentration is closely correlated with investment in research and development (R&D) and information technology (IT). I want to investigate the relationship between patent creation and top-income inequality. Some research (Aghion et al. (2018)[2] and Gompers et al. (2023)[8]) utilized micro-level data to investigate the mechanism and impacts of innovation. Still, they did not focus on the specific relationship with wage/income inequality.

Some research investigated to elucidate the wage/income disparity mechanism with the relationship of innovation. However, it does not aim to elucidate the impact of innovation on income/wage disparity. Autor et al. (2020)[3] claim that the emergence of superstar firms, highly productive platform companies, and the decline in the labor share can be linked to changes in the behavior of large dominant employers, who are now relying more on domestic outsourcing to contract a wider range of tasks previously performed in-house to third-party firms and independent contractors. These tasks may include clerical work, logistics, food services, and janitorial work, among others, and this phenomenon, also known as the "fissuring" of the workplace, has a direct impact on the labor share by excluding a large number of workers from the wage premia paid by high-wage employers to rank-and-file workers. Consequently, it can reduce the bargaining power of both in-house and outsourced workers in occupations threatened by outsourcing and increase labor market competition, potentially growing wage inequality.

Although Autor et al. (2020)[3]'s story well explains the link between innovation and a decrease

in the labor force, they might have been able to describe the connection to wage inequality. Those trends in outsourcing and job democratic change might remove some workers from the labor force and potentially decrease wages for those workers by an iota. Still, it does not necessarily mean that higher competition in the labor force increases the overall wage disparity. Song et al. (2018)[17] indirectly provide empirical support for this mechanism by showing that wage inequality among firms (wage premium) contributes more to entire wage inequality than within-firm inequality. Song et al. (2018)[17] examined two components of wage inequality, the wage gap by occupations and the average wage gap by a firm, utilizing a database of the U.S. Social Security Administration (SSA), which has a matched employer-employee database. The paper concludes 2/3 of the entire wage inequality in the market comes from inequality by firms. Cortes and Tschopp (2020) expand this theoretical idea to labor markets and assert that current increasing price sensitivities advance both market concentration and wage inequality, at least for among-firm wage inequality. This effect is theoretically well established, but empirical research would be limited due to the unavailability of price-sensitivity measurement.

Although Autor et al. (2020)[3] investigated negative impact of innovation on income inequality, they used aggregate state level data and did not use micro level data. On the other hand, though some research utilized micro data, the research focus and region are not aligning with this paper's interest in the negative impact of innovation on wage/income. Bilal and Lhuillier (2021) conducted empirical research with French administrative data and found that "more productive firms are more likely to outsource to save on higher wage premia" and asserted that those outsourced workers receive lower wages than other workers. Goldschmidt and Schmieder (2017)[7] also conducted similar empirical research with German Social Security Records. They find that, given a rapid increase in outsourcing in Germany since the 1990s, "wages in outsourced jobs fall by approximately 10–15 % relative to similar jobs that are not outsourced." Although some factors associated with nations might affect the external validity in the US, both research claims two common results: the tendency of highly productive firms to implement outsourcing and the decrease in wages for those outsourced workers.

Some research examined that the job allocation and skill dynamics could impact the wage and income distribution. Acemoglu and Autor (2011)[1] claimed that one of the potential causes of increasing wage inequality could be "broad-based increases in employment in high-skill and low-skill occupations relative to middle-skilled occupations (i.e., job "polarization") Hémous and Olsen (2022) created an exogenous growth model with automation and horizontal innovation, and they show how those factors affect unequal income distribution. Babina et al. (2021)[5] found that companies with high growth in sales tend to invest in Artificial Intelligence (AI)-related areas. The authors further claim that those investments in advanced technologies and innovation like AI can contribute to their sales expansion and create this continuous cycle of the largest firms investing the most in sale-expanding factors to maintain or improve their positions. This cycle might attract more and more investment from firms and increase wages for those who are involved in the cycle. Autor and Dorn (2013)[4] examined the US's overall job and wage dynamics. They find that labor specialized in routine tasks has been allocated to information technology and service occupations, and they "experienced earnings growth at the tails of the distribution." This implies that some portions of workers are concentrated into a specific section of IT service occupation (employment polarization) and receive relatively low wages (wage polarization). Some of these effects will be discussed later in this paper.

Some empirical focused on the decrease in labor share and job allocation, which effects income disparity. Hubmer and Restrepo 2021[10]) discuss the decline of the US labor share, particularly in the manufacturing and retail sectors, while the labor share of firms in these sectors has increased. The authors propose a model where firms incur fixed costs to automate tasks, and in response to lower capital prices, larger firms automate more tasks, reducing the overall labor share. In contrast, smaller firms continue to use labor-intensive technology, leading to a rising labor share. Using their model, the authors decompose the labor share decline, and sales concentration rise into parts driven by lower

capital prices and reallocation to higher-markup firms. They find that reallocation played a minor role in explaining the labor share decline in manufacturing and some role in retail and other sectors between 1982-2012.

De Loecker et al. (2019) analyze the trend of increasing market power in the United States economy, which has been accompanied by a rise in profits and a decline in labor's share of income. Market power refers to the ability of a company to charge higher prices or reduce output without facing significant competition. The authors find that average markups (the ratio of price to marginal cost) have risen significantly across all industries in the US since the 1980s. They argue that this trend is largely driven by an increase in firms' ability to differentiate their products and by a decrease in the number of firms operating in each industry due to factors such as technological change and increased regulatory barriers to entry. The paper explores the macroeconomic implications of rising market power, finding that it can lead to higher profits, lower investment and innovation, and reduced economic growth. The authors suggest that policies aimed at promoting competition and reducing market power, such as antitrust enforcement and regulatory reform, could have positive effects on economic outcomes.

3 Models

This theoretical model tries to explain how innovation, specifically industrial sectors with high innovation rates, could cause top income inequality, utilizing some structural modeling in economic studies. To build a general Schumpeterian growth model of the relationship between innovation and top wage inequality, this paper follows mainly Aghion et al. (2018)[2], and also this paper refers partially Jones and Kim (2018)[11] *A Schumpeterian Model of Top Income Inequality*. These theoretical narratives for the connection between innovation and top-income inequality try to explain that the industrial sectors with high innovation rates, meaning that they experience innovation more frequently than other sectors, would increase the share of top income earners, namely top income inequality. The model can be divided into three parts. Labor, innovation, and how the two influence each other.

First, the model provides the fundamental rules of labor. The model assumes two types of labor: entrepreneurs and workers, and top entrepreneurs set product prices in each market and markups as long as they successfully maintain their market shares or gain new market shares by their innovation. Entrepreneurs here are described in broader concepts: any labor whose wages are decided based on how well they do in their markets. For example, the ideas of entrepreneurs include not only Silicon Valley startup founders but also sports players, actors, and even product managers whose bonus heavily depends on their products. Therefore, if they produce innovation, which creates a vast productivity or sales increase, they receive their benefits based on the markups they develop through the innovation. All industrial sectors here have monopolistic competition, meaning the price of products is greatly influenced by the top firms or entrepreneurs with innovation.

Second, there are two types of innovation: productive and defensive innovation. In both cases, entrepreneurs would experience high markups and high wages, and thus, importantly, entrepreneurs enjoy the benefits by both types of innovation. Productive innovation creates a productivity increase in each sector, and defensive innovation is one that the current top players make to block entrants from creating innovation. This defensive innovation explains current trends that an increasing number of patents does not necessarily align with the increase in productivity in each industry.

Finally, the combination of these two settings explains the rationale behind our main point. Through innovation, entrepreneurs can influence the price of products and increase their benefits, which increases the relative share of income compared to the workers whose wages are not affected by those creative activities. When certain industries experience a high innovation rate, more frequency

in the innovation cycle, or higher return on R&D investment, they should experience a higher portion of top-income earners' share, which is higher top-income inequality.

3.1 Setup and Production(Cobb–Douglas technology)

First, we set population, type of labor, and Cobb–Douglas production function. Population - M

Two types of labor - entrepreneur and worker (ratio 1 to L)

- entrepreneur population - M/(1+L)

- worker population - ML/(1+L)

$$\ln Y_t = \int_0^{M/(1+L)} \frac{1+L}{M} \ln y_{it} di \quad (1)$$

where y_{it} is the amount of intermediate input i used for final production at date t . The number of product lines $M/(1+L)$ scales up with population size. Therefore, the final good sector spends the same amount, \tilde{Y}_t on all intermediates

$$p_{i,t} y_{it} = \tilde{Y}_t = \frac{1+L}{M} Y_t \text{ for all } i \quad (2)$$

where p is price of the intermediate

and each intermediate i is produced by a monopolist who faces a competitive fringe, using a linear production function

$$y_{it} = q_{it} l_{it} \quad (3)$$

where l is the amount of labor hired and q is labor productivity

3.2 Innovation

Second, we set two types of innovation parameters.

Productive innovation - $q_{i,t} = \eta_H q_{i,t-1}$ where $\eta_H > 1$ is productivity growth by the innovation. At the end of period t , other firms can partly imitate the (now incumbent) innovator's technology so that, in the absence of a new innovation in period $t+1$, the technological lead enjoyed by the incumbent firm in sector i shrinks from η_H to η_L with $1 < \eta_L < \eta_H$.

Defensive Innovation - The incumbent may instead undertake a “defensive innovation” which does not increase productivity but ensures maintaining a technological lead of η_H . That is, a defensive innovation prevents potential competitors from using a technology which is too close to the incumbent's

3.3 Income shares of entrepreneur and worker

Finally, we set income share of each labor type. Profit and labor share are calculated as following.

Marginal Cost = $MC_{it} = \frac{w_t}{q_{i,t}}$

The price charged by the incumbent producer - $p_{i,t} = \frac{w_t \eta_{it}}{q_{i,t}}$ where $\eta \in \{\eta_H, \eta_L\}$

Profits - $\Pi_{it} = (p_{it} - MC_{it}) y_{it} = \frac{\eta_{it}-1}{\eta_{it}} \tilde{Y}_t$ Profits are higher whenever the technological leader has recently innovated (no matter the type of innovation, productive or defensive)

$$\Pi_{H,t} = \pi_H \tilde{Y}_t > \Pi_{L,t} = \pi_L \tilde{Y}_t \quad (4)$$

with

$$\pi_H \equiv \frac{\eta_H - 1}{\eta_H}, \pi_L \equiv \frac{\eta_L - 1}{\eta_L}$$

The fraction of high-mark-up sector - μ_t
Then, Entrepreneur Share

$$\frac{\mu_t \Pi_{H,t} + (1 - \mu_t) \Pi_{L,t}}{\tilde{Y}_t} = 1 - \frac{\mu_t}{\eta_H} - \frac{1 - \mu_t}{\eta_L} \quad (5)$$

Worker share

$$\frac{w_t L}{\tilde{Y}_t} = \frac{\mu_t}{\eta_H} + \frac{1 - \mu_t}{\eta_L} \quad (6)$$

We restrict our attention to the case where $\eta_L - 1 > 1/L$, which ensures that w is less than Π_L for any value of μ_t , so that top incomes are earned by entrepreneurs. As a result, the entrepreneur share of income is a proxy for top income inequality

Since mark-ups are larger in sectors with new technologies, aggregate income shifts from workers to entrepreneurs in relative terms whenever the share of product lines with new technologies μ_t increases. By the law of large numbers this share is equal to the probability of an (unblocked) innovation in any intermediate sector.

Since mark-ups are larger in sectors with new technologies, aggregate income shifts from workers to entrepreneurs in relative terms whenever the share of product lines with new technologies μ_t increases. By the law of large numbers this share is equal to the probability of an (unblocked) innovation in any intermediate sector. Formally, we have:

productive innovation rate by Incumbent - X_{It}

productive innovation rate by Entrants - X_{Et} entry barrier - z

$$\mu_t = X_{It} + (1 - z)X_{Et} \quad (7)$$

Higher innovation rate (on either incumbent or entrants) would increase μ , which increase the entrepreneur share

3.4 innovation to industry concentration

Additionally, this section tries to explain how industry sales concentration and income inequality should be correlated by referring the model built at Kwon et al. (2023)[13].

Based on the standard nested CES demand structure, where there is a continuum of firms in industry k indexed by $i \in [0, N_k]$. In other words, a firm i in industry k faces demand:

$$y_{i,k} = Y_k \cdot \left(\frac{p_{i,k}}{P_k} \right)^{-\sigma}, \quad (8)$$

where $p_{i,k}$ is the price, and $P_k^{1-\sigma} = \int_0^{N_k} p_{i,k}^{1-\sigma} di$ is the aggregate price index for industry k , with N_k being the mass of firms in industry k . The aggregate demand for industry k is given by:

$$Y_k = \bar{Y} \left(\frac{P_k}{\bar{P}} \right)^{-\epsilon}, \quad (9)$$

with the aggregate price index $\bar{P}^{1-\epsilon} = \int_0^1 P_{k,t}^{1-\epsilon} dk$.

Firms pay an entry cost κ to enter the market. After entry, each firm i observes its idiosyncratic productivity, q . Depending on the realization of its idiosyncratic productivity, q firm has three options:

1. Exit immediately.
2. Operate with old technology:

3. Operate with new technology:

Net Present Value of Investment

$$\begin{aligned}\Pi(q_i) &= \underbrace{\sum_{t=1}^{\infty} \frac{1}{R^t} \pi_{Lt}(a_i)}_{\text{Discounted Profit}} - \underbrace{\phi}_{\text{Investment}}, \\ \Pi'(q_i) &= \underbrace{\sum_{t=1}^{\infty} \frac{1}{R^t} \pi'_{Ht}(a_i)}_{\text{Discounted Profit}} - \underbrace{\Phi(\eta_H)}_{\text{Investment}}.\end{aligned}\tag{10}$$

the firm i in industry k with unit cost $\frac{1}{q_i}$ charges a price of $\frac{1+u}{q_i}$ faces demand: Markup: $u = \frac{P-MC}{MC} = \eta_H - 1$

$$y_{i,k} = Y_k \cdot \left(\frac{1+u}{\eta_{Ht}} \frac{1}{P_k} \right)^{-\sigma}.\tag{11}$$

This pins down the input choice $L_{i,k}^*$ at:

$$L_{i,k}^* = \frac{1}{q_i} y_{i,k} = \frac{1}{q_i} Y_k \cdot \left(\frac{1+u}{\eta_{Ht}} \frac{1}{P_k} \right)^{-\sigma}.\tag{12}$$

The expressions for the firm that adopts the new technology follows similarly. Then, one can derive the following expression for Π and Π' :

$$\begin{aligned}\Pi(q_i) &= \frac{R}{R-1} \cdot \frac{u}{(1+u)^\sigma} Y_k \cdot P_k^\sigma a_i^{\sigma-1} - \phi, \\ \Pi'(q_i) &= \frac{R}{R-1} \cdot \frac{u}{(1+u)^\sigma} Y_k \cdot P_k^\sigma (h \cdot a_i)^{\sigma-1} - \phi \cdot \eta_H^h.\end{aligned}\tag{13}$$

There will be three groups of firms in equilibrium: 1) the most productive firms adopt the new technology, 2) the next productive firms operate with the old technology, and 3) the least productive firms exit immediately. In equilibrium, there exists two thresholds q^* and q^{**} , defined by:

$$\begin{aligned}\Pi(q^*) &= 0 \iff \phi = \frac{R}{R-1} \frac{u}{(1+u)^\sigma} Y_k \cdot P_k^\sigma (q^*)^{\sigma-1}, \\ \Pi(q^{**}) &= \Pi'(q^{**}) \iff q^{**} = \left(\frac{\eta_H^h - 1}{\eta_H^{\sigma-1} - 1} \right)^{1/(\sigma-1)} q^*.\end{aligned}\tag{14}$$

In equilibrium, firms with $q_i < q^*$ exit, firms with $q^* \leq q_i \leq q^{**}$ use the old technology, and firms with $q_i \geq q^{**}$ use the new technology

Second, let $S_t(q_i)$ be the per-period revenue, and $\tilde{\pi}_t(q_i) = \frac{\max\{\pi_t(q_i), \pi'_t(q_i)\}}{S_t(q_i)}$ be the profitability of the firm with idiosyncratic productivity q_i in equilibrium.

Let dF^* be the (normalized) distribution of a_i conditional on $q_i \geq q^*$, and let Q^* be given by

$$Q^* = \left(\int_{q^*}^{q^{**}} a^{\sigma-1} dF^*(q) + \eta_H^{\sigma-1} \int_{q^{**}}^{\infty} q^{\sigma-1} dF^*(q) \right)^{\frac{1}{\sigma-1}}.\tag{15}$$

Then,

$$S_t(q_i) = \begin{cases} \left(\frac{q_i}{Q^*} \right)^{\sigma-1} \frac{P_k \cdot Y_k}{N_k} & q^* \leq q_i \leq q^{**}, \\ \left(\frac{\eta_H \cdot q_i}{Q^*} \right)^{\sigma-1} \frac{(P_k \cdot Y_k)}{N_k} & q^{**} \leq q_i. \end{cases}\tag{16}$$

Furthermore, $\tilde{\pi}_t(q_i) = \frac{u}{1+u}$: the profitability of a firm corresponds one-to-one with the markup u . In

particular, it does not depend on the technology index η .

Ultimately, assuming the Pareto distribution, we can obtain the formula for measuring industry concentration based on sales. To match our findings from the real world, we determine the percentage of total sales contributed by the top 1% of companies. To make things easier, we suppose that the parameters are such that all of the top 1% of firms use new technology.

Then, the concentration measure is given by:

$$\zeta_{1\%} = \frac{\eta_H^{\sigma-1} \int_{\alpha q^*}^{\infty} q_i^{\sigma-1} dF^*(q)}{\int_{q^*}^{q^{**}} q_i^{\sigma-1} dF(q) + \eta_H^{\sigma-1} \int_{q^{**}}^{\infty} q_i^{\sigma-1} dF^*(q)}, \quad (17)$$

where α is a constant that only depends on the Pareto parameter κ . Note that for aggregate sales to be finite, we need $\kappa > \sigma - 1$. We can then obtain the concentration ratio.

The concentration ratio (top 1% sales share) is given by:

$$\zeta_{1\%} = C \cdot \frac{\eta_H^{\sigma-1}}{1 + (\eta_H^{\sigma-1} - 1)^{\frac{\kappa}{\sigma-1}} (\eta_H^h - 1)^{1 - \frac{\kappa}{\sigma-1}}}, \quad (18)$$

where C is a constant independent of η_H . Therefore, high productivity growth, which creates high innovation rate, should increase the industry concentration.

$$\zeta_{1\%} = C \cdot \frac{\eta_H^{\sigma-1}}{1 + (\eta_H^{\sigma-1} - 1)^{\frac{\kappa}{\sigma-1}} (\eta_H^h - 1)^{1 - \frac{\kappa}{\sigma-1}}} \quad (19)$$

We can log-linearize the formula

Taking the natural logarithm of both sides:

$$\ln(\zeta_{1\%}) = \ln(C) + \ln\left(\frac{\eta_H^{\sigma-1}}{1 + (\eta_H^{\sigma-1} - 1)^{\frac{\kappa}{\sigma-1}} (\eta_H^h - 1)^{1 - \frac{\kappa}{\sigma-1}}}\right)$$

Simplifying the right-hand side:

$$\ln(\zeta_{1\%}) = \ln(C) + \ln(\eta_H^{\sigma-1}) - \ln\left(1 + (\eta_H^{\sigma-1} - 1)^{\frac{\kappa}{\sigma-1}} (\eta_H^h - 1)^{1 - \frac{\kappa}{\sigma-1}}\right)$$

Using the approximation $\ln(1+x) \approx x$ for small x :

$$\ln(\zeta_{1\%}) \approx \ln(C) + (\sigma - 1) \ln(\eta_H) - [(\eta_H^{\sigma-1} - 1)^{\frac{\kappa}{\sigma-1}} (\eta_H^h - 1)^{1 - \frac{\kappa}{\sigma-1}}] \quad (20)$$

The resulting equation shows high productivity growth, which creates high innovation rate, should increase the industry concentration.

4 Data

For Wage data in industry and occupational level, I use IPUMS CPS data, which has micro-data from the monthly U.S. labor force survey, the Current Population Survey (CPS). The data set provides integrated data from different census and surveys across time. Each record is done for a person with numerical categories. The large scale samples are weighted to certain categories, which allowed researchers to conduct large data research over time. The integration of the data set would enable us to manage variables relatively freely, study historical changes, and analyze individuals data.

With the data from 1962 to 2021, this paper used incwage, which shows annual earning by only firm salaries. For this data and impute incwagelong and oincwage for missing values. The incwagelong is variable for annual salary from longest job an individual works in the year, oincwage is all other

earning besides the longest job. The data has 7.4 billion samples with 28,299 dollars for average earning by an individual between 1968 and 2021 with standard deviation of 44308.5. The minimum and maximum wage earnings among samples is 1 dollar and 3.2 million dollars.

There are another variables INCBUS, which is non-farm business income. The measurement seems to align with a part of our modeling. However, the data does not specify the type of business income and does not include dividends or capital interest. Another variable *inctot*, which is the total income, in addition to the wage earnings and business income, includes other sources of income, such as social security, pension, dividends, interests, and other governmental assistance. This contains unnecessary information of governmental assistance. Due to these reasons, this paper uses wage earnings.

For data in industry concentration, R&D, patent data by industry, I use a data set created by Kwon et al (2023)[13]. They extract data from "the Statistics of Income (SOI) and the associated Corporation Source Book published annually by the IRS" (Kwon et al (2023)[13]). This data set contains panel data of sales, assets, and net income of top 0.1, 1, 10, 50 % firms in market and industry. I use data for the entire market from 1962 - 2018 and data for sub-sector, which has concentration data by industry, for 1968-2013. The shorter periods for sub-sector data is due to "an update in its privacy guidelines (IRS Publication 1075)" (Kwon et al (2023)[13]). Patent data has two types. One is total number of patent and the other is the breakthrough patents, which are measured as the number of patents that had significant impacts to other patents innovation. The measurement is compiled through previous text analyses on patent datasets used in Kwon et al.(2023)[13].

Most data sets have the category of CPS Major Industry Records, and I categorize eight major industry groups based of the category referring Kwon et al (2023)[13], and the categories would be industry keys to combine the datasets.

5 Analysis

5.1 Equation

$$Y_{it} = X_{it}\beta + \eta_i + \epsilon_{it} \dots \text{Industry FE}$$

$$Y_{it} = X_{it}\beta + \eta_i + \theta_t + \epsilon_{it} \dots \text{Industry and Year FE}$$

where Y is a top 1% wage share and X is number of patents (innovation rate by industry X in section 3.3) or revenue concentration (ζ in section 3.4) in industry i at year t . η is the time-invariant component of the error term (industry-level fixed effect), and θ is the common unobserved time component of the error term (year-level fixed), and ϵ is error term that captures the remaining unobserved characteristics. Standard errors are Driscoll and Kraay (1998).

5.2 Measurement

Top income inequality is measured by a share of top 1 percentile income earner among entire wage earnings by sector. The Gini coefficient is a measurement method to check the statistical dispersion of distribution, specifically used to measure inequality. Zero in Gini coefficient means perfect equality where everyone receives the same amount, while one in Gini coefficient means perfect inequality one person takes everything. The industry concentration is measured by the revenue share of the top 0.1, 1, 10, and 50 percentile firms in the industry among the entire industry revenue share. Patent data is number of patents and breakthrough patents. This paper divided the market into eight different sectors: Agriculture, Mining, Manufacturing, Finance, Utility, Trade, Service, and Construction. The measurement is different from labor concentration, which measures the ratio of labor the top firms

hire among all labor force (Manning 2020)[14]

5.3 Regression Results

Table 1 shows the result of regression analysis(Ordinary Least Square (OLS)) with the top income inequality and innovation measurement, which is number of patents and breakthrough patents. The first four results on the left side only include industry fixed effects, and the other four results on the right side include industry and year fixed effects. Adding fixed effects refers to separating out the effects that are specific to a particular time or industry. This approach allows us to isolate the impact of a particular factor without being influenced by other time or industry-specific trends. For instance, in economics, one might want to study the relationship between a country's GDP and its level of government spending. However, without accounting for the specific year or industry, it would be difficult to isolate the true relationship between the two variables. By adding fixed effects, such as controlling for the specific year, we can remove the effects of yearly trends and focus on the underlying relationship between GDP and government spending. Additionally, adding industry and year fixed effect separately is one way to check the robustness of the correlation. Table 1 shows that consistent positive correlation between top 1 percent income and number of patents with industry and year FE. Having stars shows that the correlation is statistically significant.

Similar to Table 1, Table 2 also shows consistent positive correlation between top 1 percent wage earnings and industry concentration with industry and year FE

Table 3 shows the correlation between the share of managers and professional occupations, instead of top 1%, with numbers of patents to see the case when the assumption in section 3.3 does not hold. It shows somewhat consistent positive correlations with numbers of patents. Table 1-3 matches our model settings.

Unlike table 1, 2, and 3, Table 3, the result of the OLS regression analysis with the general inequality measurement (Gini coefficients) and innovation measurement with number of patents, show statistically insignificant results.

Interestingly, table 4, the result of the OLS regression analysis with Gini coefficients and industry concentration, shows that, with only industry FE, sales concentration of top firms by main industry sector have statistically significant positive coefficients with Gini coefficients of wage inequality. However, the sign becomes negative in most cases by adding the year FE. This implies that the effects of year dummy variables absorb the impacts of industry concentration to wage inequalities. The correlation between the two variables is not robust and consistent, and the two variables are highly likely to have just two upward trends over time. Several regression analyses are conducted, including replacing industry revenue concentration with industry assets concentration of top firms or changing the number of sectors. However, the results have a similar tendency with not much consistency with different types of fixed effect settings.

The result of table 1 and 4, in terms of having consistent result with top income inequality but not other general inequality measurement, align with the results of Aghion et al. (2018)[2]

6 Discussion and Conclusion

This paper provides a theoretical background with empirical support for how innovation causes top-income inequality. The initial analysis, aligned with the theoretical model, shows that an increase

Table 1: Regression results with Top Wage Earning Share

	(1)	(2)	(3)	(4)	(5)	(6)
	Top 1% Wage Earning Share(Inequality)					
Log Breakthrough Patent	0.0145*		0.00770	0.00448**		0.00318**
	(0.006)		(0.005)	(0.001)		(0.001)
Log Patent		0.0526***	0.0333**		0.0169***	0.0110**
		(0.007)	(0.011)		(0.003)	(0.003)
Observations	175	210	175	175	210	175
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Regression results with Top Wage Earning Share

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Top1%Wage Earning Share							
Top 0.1% receipt share	0.103**				0.00332			
	(0.038)				(0.011)			
Top 1% receipt share		0.186***				0.186***		
		(0.051)				(0.051)		
Top 10% receipt share			0.296***				0.296***	
			(0.077)				(0.077)	
Top 50% receipt share				1.272***				1.272***
				(0.297)				(0.297)
Observations	336	336	336	336	336	336	336	336
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	No	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Wage Share of Managers and Professional Occupations

	Wage Income Share of Managers/Professionals			
Log Breakthrough Patent	0.0758***		0.00749	
	(0.009)		(0.005)	
Log Patent		0.145***		0.0855**
		(0.027)		(0.027)
Observations	175	210	175	210
Industry FE	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Gini Coefficient and Patent Data

	Gini Coefficient of Wage Income			
Log Breakthrough Patents	0.00530		-0.0145*	
	(0.004)		(0.005)	
Log Patents		0.0252		0.0320*
		(0.013)		(0.013)
Observations	210	210	210	210
Industry FE	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Regression results with Gini

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Gini Coefficient Wage Earning							
Top 0.1% receipt share	0.143***				0.00528			
	(0.019)				(0.030)			
Top 1% receipt share		0.147***				-0.207***		
		(0.019)				(0.037)		
Top 10% receipt share			0.183***				-0.396***	
			(0.031)				(0.045)	
Top 50% receipt share				0.331*				-2.414***
				(0.148)				(0.207)
Observations	368	368	368	368	368	368	368	368
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	No	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in innovation and productivity would increase the top income inequality, which is an income share of entrepreneurs, but not necessarily increase general income inequality measurement like Gini coefficient, which aligns with previous research

A disadvantage of the Gini coefficient is that it fails to account for any changes in the structure of a population. In other words, the measurement could show the same degree of inequality for a few of the population receiving extremely high income, and a large portion receiving relatively high income, given less inequality among the lower income population for both cases. In the theoretical section, we divided the labor into two sections. The ratio of those two types of labor influences a lot with degrees in other general inequality measurements like Gini. If the ratio of income earners who experience sharp increases in wages compared with others is high, Gini would show a higher increase in inequality. However, the measurement would not really respond to distributional changes in a few top workers' earnings. Therefore, the measurement could fail to grasp the structural change in income distribution given the innovation. However, the top income inequality shows a positive correlation, which aligns with our theoretical prediction.

The main limitation of this paper's analysis is of its data. Income inequality is operationalized through the wage variable, but this might not be the best variable to measure the top income inequality. Further research should also run an analysis of other data sources, such as the Survey of Consumer Finance(SCF). The dataset collects data of households' income through wage and capital, which is closer to our theoretical settings though the dataset also faces some limitations, such as a limited number of available years of data and data level of households but not of individuals. In terms of operationalizing innovation, further research should add various different measurements, such as R&D investment and IT investment. For the general inequality measurement as well, this paper used Gini coefficient, but we could also use other inequality measurements, such as coefficient of variance, variance of log, and 90-50 ratio. Especially, a variance of log could be used to decompose the inequality by categories like occupation, firm, and industry. Further research should also provide calibration, utilizing the equation 1, 7, and 20, to see if those theoretical models and empirical data match. Another limitation is necessity of theoretical implication of those effects as Syverson[18] mentions those concentration measurements do not necessarily measure market power and they tend to be outcomes of other economic trends and factors. Though this research only focused on correlation with innovation, further research might need to clarify the connection between those concentration measurements and other economic factors and identify the implications of those connections.

For the next step, researchers can quantify the impact of innovation growth on wages by controlling other covariates in order to determine the overall quantitative relevance of innovation in top income inequality. Ideally, further research should link the above dataset with Lightcast (previously called BurningGlass) job postings data, which has near universe data of job postings. The dataset would allow us to extract job postings information, such as title, salary, posted time, and job description. Combining the job postings data with the industrial investment and patents data sets would create a novel dataset that can be utilized to further scrutinize the impact of innovation on income inequality. Some other job posting platforms, such as glassdoor, can be also utilized to see the dynamic change in firms' job demands (Karabarbounis and Pinto 2019)[12]. To conduct those empirical research, Haltiwanger et al (2022)[9] provides some methodologies for decomposing components of inequalities in industries and occupation. Wage dispersion could be divided into dispersion in mean wages of each industry and dispersion in wages within each industry. Further research also should decompose the trends and effects of innovation on wages in different dimensions, such as skills and geography (Tambe and Cappelli 2020[19], Bilal and Lhuillier 2021[6]). Rossi-Hansberg et al (2018)[16] finds the distinctive features in local and national revenue concentrations.

Combined with complementary research on the negative impacts of top income inequality on eco-

conomic stability and wellness, this research would support scrutinizing the negative impact of innovation or other types of productivity increase and provide statistical evidence for policy insights, such as a new structure frame for anti-trust legislation. If innovation, contrary to popular belief, has negative impacts by increasing top income inequality, policymakers and economist must consider how they can mitigate those negative impacts while encouraging the positive aspects of economic growth.

References

- [1] D. Acemoglu and D. Autor. Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 1043-1171, 2011.
- [2] Akcigit U. Bergeaud A. Blundell R. Aghion, P. and D. Hemous. Innovation and top income inequality. *The Review of Economic Studies*, 86(1):1-45, 2018.
- [3] D. Autor, D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen. The fall of the labor share and the rise of superstar firms*. *The Quarterly Journal of Economics*, 135:2, 2020.
- [4] D. H. Autor and D Dorn. The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review*, 103(5):1533-1597, 2013.
- [5] T. Babina, A. Fedyk, A. X. He, and J. Hodson. *Artificial Intelligence, firm growth, and industry concentration*. SSRN Electronic Journal, 2020.
- [6] A. Bilal and H. Lhuillier. Outsourcing, inequality and aggregate output. *SSRN Electronic Journal*, 2021.
- [7] D. Goldschmidt and J. F. Schmieder. The rise of domestic outsourcing and the evolution of the german wage structure*. *The Quarterly Journal of Economics*, 132(2):1165-1217, 2017.
- [8] Kaplan S. and; Mukharlyamov V. Gompers, P. The market for ceos: evidence from private equity. 2023.
- [9] C. Haltiwanger, R. Hyat, and J. Spletzer. Industries, mega firms, and increasing inequality. *NBER working paper*, 104, 2022.
- [10] J. Hubmer and P. Restrepo. Not a typical firm: The joint dynamics of firms, labor shares, and capital-labor substitution. *SSRN Electronic Journal*, 2021.
- [11] C. I. Jones and J. Kim. A schumpeterian model of top income inequality. *Journal of Political Economy*, 126(5):1785-1826, 2018.
- [12] M. Karabarbounis and S. Pinto. What can we learn from online wage postings? evidence from glassdoor. *Economic Quarterly*, 104, 2019.
- [13] S. Y. Kwon, Y. Ma, and K. Zimmermann. *100 years of rising corporate concentration*. SSRN Electronic Journal, 2021.
- [14] A. Manning. Monopsony in labor markets: A review. *ILR Review*, 74:1, 2020.
- [15] M. E. Porter. Competitive strategy: Core concepts. *In Competitive advantage: Creating and sustaining superior performance*, 1, 1985.
- [16] E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter. Diverging trends in national and local concentration, 2018.
- [17] J. Song, D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter. Firming up inequality*. *The Quarterly Journal of Economics*, 134:1, 2018.

- [18] C. Syverson. Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, 33:3, 2019.
- [19] P. Tambe, X. Ye, and P. Cappelli. Paying to program? engineering brand and high-tech wages. *Management Science*, 66(7):3010–3028, 2020.