IARIW 2024

IARIW 2024

Thursday, August 22 – Friday, August 30

Measuring Hourly Wage of Self-employed Workers

David Jang

(Bureau of Economic Analysis)

Paper prepared for the 38th IARIW General Conference
August 26-30, 2024

THEME 6: Output and Productivity Measures in Dynamic Industries

Time: Wednesday, August 28, 2024 [17:30-18:30 GMT]

Measuring Wages of Self-Employed Workers*

David W. Jang†

July 31, 2024

Abstract

The income of self-employed workers is a combined income from their labor and capital input. Self-employed workers' hourly wages are often assumed to be comparable to those of salaried workers because it is hard to pick labor compensation from the mixed income of the self-employed. To overcome this challenge, I examine the hourly income of owner and non-owner self-employed workers and measure the hourly labor wage of self-employed workers. With the newly measured self-employed workers' wages, I augment the US labor share data and BEA/BLS integrated industry-level production accounts.

JEL Codes: E01, E24

*This version is prepared for the 38th IARIW General Conference

†Bureau of Economc Analysis. The views expressed in this paper are solely those of the author and not necessarily those of the U.S. Bureau of Economic Analysis or the U.S. Department of Commerce. Email: david.jang@bea.gov.

1 Introduction

In this study, I propose a novel approach for measuring the hourly wage of self-employed (SE) workers by leveraging the US survey datasets that provide information on self-employed individuals' income. Then, I incorporate this data to improve the aggregate data series produced by the U.S. statistical agencies. In contrast to the hourly wage of salaried workers, there is a major challenge in measuring the hourly wage of SE workers. The earnings of self-employed individuals encompass both the rewards for their work efforts and the returns from the business assets they have invested in (Krueger (1999)). Since survey responses regarding the income of self-employed workers are mixed incomes of their labor and capital, it has been challenging to measure the hourly labor compensation for self-employed workers. To overcome this issue, many approaches have been utilized to complete the official statistics such as the U.S. Labor Share data by the Bureau of Labor Statistics and BEA-BLS Integrated Industry-level Production Accounts (KLEMS). In this paper, I present a new measure of the hourly wage of SE workers and show how this new measure compares to the traditional methods.

Currently, the hourly wages of SE workers are often inferred from the hourly wages of salaried workers. In the case of the U.S Labor Share data (U.S. Bureau of Labor Statistics (2024)), the hourly wage of self-employed workers is often assumed to be the same as that of salaried workers as explained in Giandrea and Sprague (2017). Also, the Integrated Industry-level Production Accounts (KLEMS) assume that self-employed workers and salaried workers earn the same hourly wages once cross-classified by the demographics of the workers as mentioned in Samuels and Varghese (2022). In both cases, this assumption is made because of the lack of information on self-employed workers' hourly wages in the source data they use.

The Census Bureau Survey of Income and Program Participation (SIPP) and the Federal Reserve Board Survey of Consumer Finances (SCF) are two surveys that provide information on the hourly wages of self-employed workers. SIPP shows job-level income and hours worked data for self-employed workers[1]. The advantage of SIPP data over SCF is that SIPP has a larger sample size with a higher frequency. The self-employment wages are recorded on a monthly frequency whereas the SCF is conducted in a triennial frequency.

Even with these surveys and their attempt to record the income and work hours of SE workers, it is not straightforward to measure the hourly wages of SE workers because their answer is the combined amount of income for their labor and capital input. In this paper, I propose a two-step process to determine the labor part of the hourly wages of self-employed workers. Firstly, I estimate the mixed income of self-employed workers using the SIPP data at annual frequency. Secondly, I discount the mixed income of self-employed workers to remove the capital compensation leaving only the labor compensation.

The second step is available thanks to the unique feature of SCF data in terms of self-employed workers which is that the pool of self-employed workers could be divided into two distinct groups depending on the business ownership status. Using the ownership indicator variable, I estimate the hourly wage premium of business owners. Then, I use this estimated wedge to measure the hourly labor wage of self-employed workers in SIPP data. In other words, the way I decompose the hourly income of self-employed workers into hourly wage and hourly capital income is to compare self-employed workers who own their businesses to those who do not.

In practice, the SCF asks their survey respondents two separate questions regarding their employment and business ownership status. The first question asks whether they are self-employed workers or not. The second question asks whether their household owns at least one business or not. Among self-employed workers, about 17% of them record that their household does not own any business. The self-employed workers who respond that they don't own a business are people who work independently for somebody else. Certain forms

[1]In contrast, the Current Population Survey (CPS) and American Community Survey (ACS) both do not ask about the income of self-employed workers. CPS Annual Social and Economic Supplements (ASEC) does include annual business income but asks for the total hours worked for all jobs instead of the hours dedicated to the business, making it impossible to calculate the hourly wages put into the business. SIPP and SCF data are free from this issue.

of self-employment such as contractors might not involve assets or obligations that survey respondents would typically classify as constituting a business. Since self-employed workers who do not own businesses would not utilize a significant amount of capital for their work, they provide a counterfactual labor income for those who own businesses if they only use their labor for their business.

2 Hourly Wages of Self-Employed Workers

2.1 Picking the Labor Wage from the Mixed Income

In this part of the paper, I pick the labor hourly wages of self-employed workers from the information on their mixed income of labor and capital. The first part of the solution is to estimate the business ownership premium, which will then be used inversely as a discount factor to convert mixed income into labor income. This conversion is crucial for accurately calculating the labor hourly wages of self-employed workers. The discount factor will be the inverse of the ownership premium if we assume that the ownership premium is coming from business owners' capital input. In the appendix, I show different ownership premiums by industry and their net worth to support this claim in table 3 and table 4.

For estimating the ownership premium of self-employed workers, I focus on non-farm business sectors excluding workers in farm and public sectors. In the publicly available version of the SCF data, specific four-digit industry codes are not provided. Instead, industries are categorized into seven sectors, denoted by the variable showing the broad sector. Among these seven sectors, I exclude workers in farm and public sectors using the industry information. In addition, I include data for both the head of the household and his spouse, if applicable[2]. Since the income reported in the survey varies by pay frequency, the data were cleaned in terms of hourly wages using the variables hours worked and weeks worked per

[2]The SCF survey is conducted at the household level, so the information of the head and spouse are included in the same observation with different variables. I separated them into different observations for my analysis

year variables[3]. I also remove workers who are not self-employed for the main job.

Table 1 shows a brief table of summary statistics of two different worker groups in self-employment shown in the SCF data. The first column of table 1 is for non-owner self-employment workers and the second column is for owner self-employed workers. I use main job wages. The hourly wage includes regular pay and bonus pay. Wages are winsorized at the top and bottom one percent. Overall, owner-self-employed workers earn much higher hourly wages, are older, consist more of male workers, are more likely to be married, and show longer job tenure.

Table 1: Summary Statistics by Group

	Non-owners	Owners
Hourly wages	68.14 (197.93)	232.16 (409.34)
Age	49.93 (14.63)	53.03 (11.85)
Female ratio	0.41 (0.49)	0.29 (0.45)
Married ratio	0.66 (0.47)	0.85 (0.36)
Tenure years	13.18 (12.52)	18.13 (12.82)
n	10626	53597

The data is pooled from SCF 1989 to SCF 2016. All wages are in 2012 dollars

I compare these two groups of self-employed workers to estimate the ownership premium and subsequently determine the labor share of self-employment wages. The non-owner self-employment workers do not possess a significant amount of capital to categorize their work arrangements as owning a business. In contrast, owners own and borrow capital to operate their businesses. By controlling for observable characteristics, the wage difference between these two groups comes from the capital input. The following equation outlines the regression for comparing owner and non-owner self-employed workers and measures the capital share of self-employment wages:

[3]For example, I use variable X4112 for the income from the main job of household heads. However, this variable represents income for varying frequencies which is reported in variable X4113. Divide the income variable by the appropriate unit of hours for each income frequency to calculate the hourly wages

$$log(Y_{it}) = \beta_0 + \beta_1 \cdot \text{Ownership}_{it} + \beta_2 \cdot X_{it} + \beta_3 \cdot D_{it} + \varepsilon_{it} \tag{1}$$

For this pooled OLS regression, Y_{it} represents an hourly wage for individual i in year t, and β_1 is the estimated ownership premium using the variable of the business ownership indicator. X_{it} denotes a set of control variables including age, age-squared, sex, education, tenure, tenure-squared, and industry. I also include year and firm size controls denoted as $_{it}$ whose inclusion varies by specifications. [4]. The result is reported in table 2.

For the pooled sample of self-employed workers in SCF, the business ownership premium is shown in the first row of table 2 and ranges between $e^{0.388} - 1 \approx 47.4\%$ and $e^{0.404} - 1 \approx 49.8\%$. Therefore, under the assumption that the business ownership premium shows the capital share of self-employed wages, the labor share of self-employed workers is between $\frac{1}{e^{0.404}} \approx 66.8\%$ and $\frac{1}{e^{0.388}} \approx 67.8\%$. Surprisingly, this number aligns very well with the traditional literature examining the labor share of the aggregate economy.

After estimating the labor share of self-employed workers' wages, I compare the wages of self-employed workers to salaried workers using the SIPP data. Because there is a discount rate ready to apply to mixed income to transform it into labor income, SIPP with its rich number of observations comes in useful. With SCF, the wages of self-employed workers are only observable every three years. The purpose of this procedure is to gather the best information about the mixed-income wage of self-employed workers and apply our result in table 2. Here, I chose 67% for the labor share of the self-employed wages, which is a middle value among the estimated coefficients in table 2 and discounted the observed hourly wages by multiplying 0.67 by them. To be clear, I use the business income from the SIPP survey as the income of self-employed workers.

For sample selection, I only include workers in non-farm business sectors as I did in SCF

[4]Race is not included because the spouse's race information is included starting from the 1998 SCF survey

Table 2: Estimated Ownership Premium

	(1)	(2)	(3)	(4)
Ownership	0.388***	0.402***	0.390***	0.404***
	(0.0176)	(0.0162)	(0.0175)	(0.0161)
Age	0.00652	0.00234	0.00400	-0.000340
	(0.00352)	(0.00324)	(0.00351)	(0.00322)
Age^2	0.0000101	0.0000534	0.0000293	0.0000746*
	(0.0000330)	(0.0000303)	(0.0000329)	(0.0000301)
Female	-0.665***	-0.548***	-0.669***	-0.551***
	(0.0141)	(0.0130)	(0.0140)	(0.0129)
Married	0.178***	0.0933***	0.184***	0.0989***
	(0.0156)	(0.0144)	(0.0155)	(0.0143)
Tenure	0.0501***	0.0351***	0.0490***	0.0337***
	(0.00165)	(0.00153)	(0.00165)	(0.00152)
Tenure^2	-0.000583***	-0.000443***	-0.000569***	-0.000421***
	(0.0000362)	(0.0000333)	(0.0000360)	(0.0000331)
Firm size	No	Yes	No	Yes
Year	No	No	Yes	Yes
N	55259	55259	55259	55259
adj. R^2	0.346	0.447	0.353	0.455

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

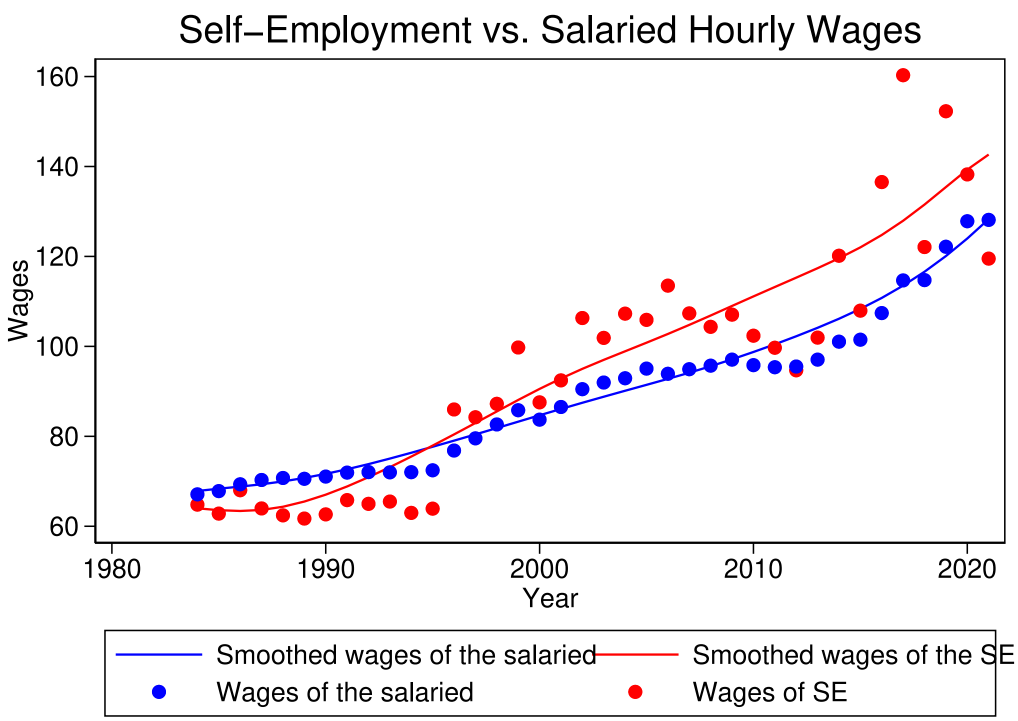
All wages are in 2012 dollars. Samples are from SCF 1983 to 2016 pooled together

data[5]. One thing to note is that the 2004 SIPP panel started regarding contingent workers as

[5]In terms of SIPP industry codes of salaried workers, I excluded 0170 (Crop production), 0180 (Animal production), 0190 (Forestry except logging), 0270 (Logging), 0280 (Fishing, hunting, and trapping), and 0290 (Support activities for agriculture and forestry). For businesses, I exclude a category of agriculture, forestry, fishing, and hunting (observations with a value of 1 in the TBSIND1 code for the corresponding years. For other years, I use the equivalent variable instead because the variable name changes over time). Only primary job information is used in the analysis. The wage is winsorized at the bottom and top one percent

workers running businesses. Therefore, I removed the contingent workers from the sample. If I did not, the hourly wage of self-employed workers shows a sudden drop in the year 2003. Before the 2004 SIPP panel, it was hard to find information on the wages of contingent workers. The wages are in real terms in 2012 dollars where the PCE price index was used to calculate them.

Figure 1: Real Wages over Time



The data is from SIPP data of non-farm business workers from the year 1984 to 2021. The hourly wages are deflated in 2012 dollars using the PCE price index. Smoothed values are calculated using the Lowess smoothing method

Figure 1 shows the wage change over time for salaried workers and self-employed workers. The blue and red dots each show the hourly wages of the salaried and self-employed wages. Since the wages seem to be very volatile over different survey panel years of SIPP, I report the smoothed wages in solid lines. Overall, before the mid-1990s, salaried workers used to show higher hourly wages but self-employed workers have shown a steeper incline in their wages making the self-employed workers earn more than salaried workers starting in 1995.

Even though the time series only goes back to 1984, the conjecture in the Elsby et al. (2013) that assuming self-employed wages as the same as salaried workers would have exaggerated the actual wages in the 1980s.

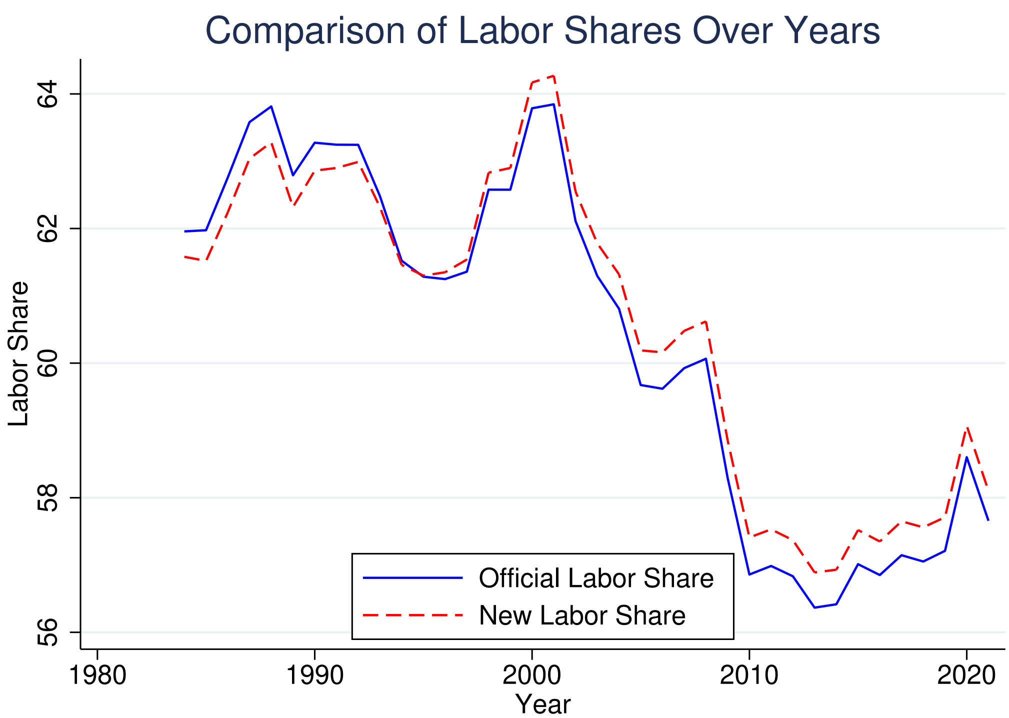
2.2 Implications on the US Labor Share

In this section, I suggest a method of building an alternative labor share data series that utilizes the self-employed wages data that I built in the previous section without an imputation process that equates self-employment and salary work wages. The headline labor share published by BLS is a data series that assumes that self-employed workers earn the same wage as salaried workers. Giandrea and Sprague (2017) mention that an attempt has been made to refine the imputation of self-employment wages by dividing workers into occupation groups. Instead, I build an alternative labor share data series without imputing self-employed wages and utilize the data points in fig. 1. The official and alternative labor share series are shown in fig. 2

The solid line in fig. 2 is the official BLS headline labor share and the dashed line is the labor share data without imputing self-employed workers. Instead of imputing it, a two-step procedure is adopted to incorporate the result in fig. 1. Firstly, I calculate the ratio of self-employed wages to salaried wages using data from fig. 1. Secondly, I multiply the ratio by the non-farm business sector wages and again multiply it by self-employed workers' total hours to calculate the aggregate self-employed income. The difference between the two lines in fig. 2 is that the solid line uses self-employment income imputed from the assumption that self-employed workers earn the same hourly wage as salaried workers, while the dashed line does not rely on this assumption. Instead, use the ratio between self-employed and salaried wages that I obtained from fig. 1 and multiply it by the wages of salaried workers used in the computation of the official labor share data[6]. The benefit of this method is that I utilize the wage data of the salaried workers used in the official publication which is more reliable

[6]Since the raw SIPP wages are very volatile over panel years, I calculate the ratio with smoothed values

Figure 2: Labor Share Measures Using the New SE Wages data



The labor share is shown from the year 1984 to 2021 because 1984 is the furthest year that the SIPP data goes back to.

than the wage levels I calculated directly from the SIPP data.

There are two major points to note from fig. 2. The first finding is that the decline in the labor share over the period is much less pronounced for the new labor share measure. This is because the hourly wage of the self-employed workers has increased more steeply than that of the salaried workers. This implies that the assumption about wages of self-employed workers taken to calculate the labor share may have made the labor share appear to be declining faster than it is currently. The second finding is that the alternative labor share is lower than the official numbers in the periods before 1995. The difference between these two series in 1984 is 0.37%p and goes up to 0.54%p in 1987. Then, the new labor share measure exceeds the official numbers in 1995 and is 0.44%p higher in 2021.

2.3 Implications on the US KLEMS

Another data product that assumes that self-employed workers earn the same hourly wage as salaried workers is the BEA-BLS Integrated Industry-level Production Accounts (KLEMS). One of the main building blocks needed to build the TFP growth rate in the US KLEMS is the labor input. In calculating the labor input, the hours of workers are weighted by their labor compensation, because the same hour of labor input should be considered different between high-productivity and low-productivity workers. The labor compensation is calculated for each demographic and industry group. However, the survey responses of self-employed workers' labor compensation is mixed income of capital and labor inputs, which makes it inappropriate weight to be used. Therefore, the official US KLEMS cross-classifies workers by sex, age, education, and industry to overcome this issue (Samuels and Varghese (2022)). Then, it is assumed that self-employed workers earn the same hourly wages as salaried workers in the same group.

Alternatively, for each cross-classified group, I utilize the ratio of self-employed wages to salaried wages using the data from fig. 1. Instead of using the salaried wage for self-employed workers, I multiply the salaried wage by the ratio. This changes the labor compensation of the worker groups, which eventually leads to TFP growth rates that are different from the official US KLEMS data.

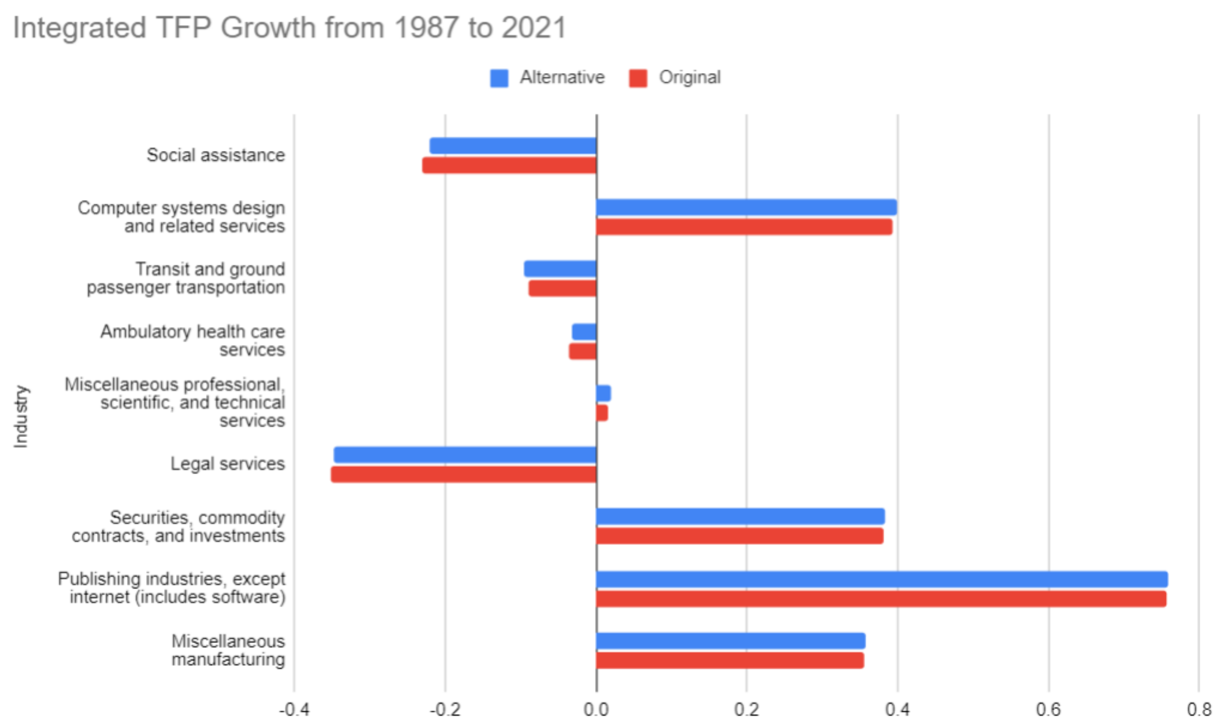


Figure 3: Integrated TFP growth rates from 1987 to 2021 by industries

The labor share is shown from the year 1984 to 2021 because 1984 is the furthest year that the SIPP data goes back to.

Figure 3 shows the TFP growth rates by industry over the period between 1987 and 2021 for the top 10 industries that show the biggest discrepancies between official TFP growth rates and the alternative TFP growth rates in terms of %p difference. For example, the social assistance industry showed a negative 23.1% productivity growth in the official data over three decades. However, it only showed a negative 22.0% productivity growth in the alternative data showing about a 1.1%p discrepancy. For the insurance carriers and related activities industry, it was a 0.3%p discrepancy implying that the other 53 industries show a very small gap between the alternative and official TFP growth rates.

An interesting finding is that for 9 out of the 10 industries shown in the fig. 3, the TFP growth rates were higher once measured without the equal wage assumption between salaried and self-employed workers. Among 59 industries, 32 industries showed higher growth rates, 17 industries showed the same growth rates, and 10 industries showed lower growth rates with the alternative measure than the original measure[7]. It would be interesting to investigate further how integrating the higher wage growth rates of self-employed workers than those of salaried workers resulted in an improvement in the TFP growth rates across industries. My current conjecture is that this is caused by increased hourly wages of self-employed workers combined with lowered total hours worked from them[8]. The high wage growth of self-employed workers put increased weight on their hours in calculating aggregate labor input growth but their total hours input did not show as much increase[9]. Therefore, the industry-level aggregate labor input growth is lower in the alternative measure than the original one, resulting in a higher TFP growth rate in some industries.

3 Conclusion

In conclusion, this study presents a practical method to understand the income composition of self-employed workers, addressing a persistent challenge in official statistics. By analyzing data from surveys such as the SCF and the SIPP, this research sheds light on the mixed nature of self-employment earnings, suggesting a way to disentangle labor income and returns from business capital. Through a comparative analysis of self-employed individuals who own businesses and those who do not, this study provides insights into the hourly wage premium associated with business ownership, offering a clearer picture of self-employment income dynamics. Using the larger sample size and higher frequency data from SIPP complements the findings of SCF, improving our understanding of hourly wage dynamics among self-employed workers.

[7] I excluded farm, forestry, state, and federal government industries from the original US KLEMS show 63 industries focusing on 59 non-farm business industries

[8] From 1987 to 2021, total hours input from salaried workers increased 39 percent but those from self-employed workers decreased 12 percent according to the data used to compute the US KLEMS data

[9] Russell et al. (2021) show how to calculate the labor input measure used in the calculation in the US KLEMS in detail

These findings have practical implications for those seeking to refine economic indicators and inform labor market analysis. By disentangling wage and capital components of self-employment income, this study contributes to more accurate economic measurements, which can inform policy decisions and theoretical frameworks. This research lays the groundwork for further investigation into the complicated nature of self-employment compensation,

This research is not without its shortcomings. First, it is assumed that the labor share of self-employed workers does not change over time. Also, it is required to investigate the accuracy of the labor share estimate with different estimation methods such as a matching estimator. Another venue to examine is to check how well the measure of self-employed labor income from this research aligns with the asset approach to calculate proprietors' labor compensation equal to the difference between the proprietor' income and the non-corporate capital income.

Even with its caveats, this research aims to improve official statistics by constructing direct estimates of self-employed wages. I plan to further enhance the measurement of self-employed wages by dividing SIPP workers into different groups and applying separate self-employment-to-salaried wage ratios to each group.

References

Elsby, Michael WL, Bart Hobijn, and Ayşegül Şahin (2013) "The decline of the US labor share," *Brookings Papers on Economic Activity*, 2013 (2), 1–63.

Giandrea, Michael D and Shawn A Sprague (2017) "Estimating the US labor share," *Monthly Labor Review*.

Krueger, Alan B (1999) "Measuring labor's share," *American Economic Review*, 89 (2), 45–51.

Russell, Matt, Jon Samuels, Michael Sverchkov, Jerin Varghese, Corby Garner, and Mike Jadoo (2021) "A BEA-BLS Integrated Approach for Adjusting Industry Labor Input Measures for Labor Composition,"Technical report, Bureau of Labor Statistics and Bureau of Economic Analysis, Joint work by BLS and BEA economists.

Samuels, Jon and Jerin Varghese (2022) "Introducing Demographic Labor Market Data into the US National Accounts0F1."

U.S. Bureau of Labor Statistics (2024) "Nonfarm Business Sector: Labor Share for All Workers [PRS85006173]," https://fred.stlouisfed.org/series/PRS85006173, retrieved from FRED, Federal Reserve Bank of St. Louis; July 19, 2024.

Appendix A.

In table 3, the ownership premium is estimated separately by industry. The regression specification is the same as table 2. Then, table 4 shows the average net worth among businesses that report positive net worth. It should be noted that the manufacturing and professional service sectors show a high business worth and an ownership premium at the same time. It sheds light on how the ownership premium is correlated with the net worth of the business. Although the net worth of the business does not show how much capital is present in the business, it is a close variable provided in the SCF survey.

Table 3: Labor share estimation by industry

	(1) Mining and Cons. $ln(wage)$	(2) Manu. $ln(wage)$	(3) Trade $ln(wage)$	(4) Prof. service $ln(wage)$	(5) Gen. service $ln(wage)$
Ownership	0.187***	0.643***	0.342***	0.449***	0.338***
	(0.0477)	(0.0690)	(0.0587)	(0.0379)	(0.0239)
Age	-0.0155	0.0279*	-0.0129	0.00388	0.00880
	(0.00919)	(0.0126)	(0.00979)	(0.00772)	(0.00538)
Age^2	0.000273**	-0.000210	0.000167	0.0000286	-0.0000243
	(0.0000906)	(0.000115)	(0.0000918)	(0.0000694)	(0.0000514)
Female	-0.541***	-0.770***	-0.693***	-0.622***	-0.652***
	(0.0573)	(0.0494)	(0.0358)	(0.0310)	(0.0198)
Married	0.249***	-0.138*	0.126**	0.266***	0.200***
	(0.0410)	(0.0586)	(0.0441)	(0.0344)	(0.0220)
Tenure	0.0485***	0.0529***	0.0768***	0.0500***	0.0433***
	(0.00483)	(0.00572)	(0.00445)	(0.00341)	(0.00249)
$Tenure^2$	-0.000613***	-0.000409***	-0.000894***	-0.000572***	-0.000606***
	(0.000107)	(0.000117)	(0.0000942)	(0.0000724)	(0.0000572)
N	6176	5248	7779	13386	22670
adj. R^2	0.386	0.298	0.337	0.330	0.366

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From left to right, mining and construction, manufacturing, trade, professional service, and general service. All wages are in 2012 dollars

Table 4: Summary Statistics for Business Worth by Industry (in Millions)

	Mining and Cons.	Manu.	Trade	Prof. Service	Gen. Service
Mean	7.36	10.60	6.67	11.10	2.91
Standard deviation	43.10	54.30	39.00	47.00	23.20
N	8,089	9,765	12,238	20,004	37,647

Note: Means and standard deviations are in millions of dollars. All dollars are in 2012 dollars.