The Impact of Self-employed on Productivity Measurement in the Netherlands

Hugo de Bondt
(Statistics Netherlands / CBS)
h.debondt@cbs.nl

Paper prepared for the 37th IARIW General Conference

August 22-26, 2022

Session 4B-1, Measuring the Self-Employed, Contract Work, and Gig Workers I

Time: Wednesday, August 24, 2022 [14:00-15:30 CEST]

The impact of self-employed on productivity measurement in the Netherlands

Hugo de Bondt[1]

Abstract:
This paper discusses several different estimates on labor income of self-employed used in productivity measurement, and present these estimates using data on the Netherlands. Labor inputs of self-employed are an often overlooked part of productivity measurement. The main reason is that explicit labor income for this group of workers is not measured in national accounts. What's more, their share of the economy was largely stable over time in developed countries, but in countries such as the Netherlands a rise in their numbers and in share of the work force had occurred. To shed more light on this issue this paper presents an overview of possible estimates of self-employed labor income into three "families" and presents five of those possible estimates. It is found that imputing based on average employee compensation gives the highest estimates and a direct measure gives the lowest estimate. Results based on mixed income fall right in between. Adding new volume estimates based on the background characteristics and combining with the labor income estimates show a strong impact of the new volume on the contribution of self-employed labor in the years before the financial crisis. After the financial crisis the weight impact was more pronounced. A final result show the changing endogenous rates of return to capital, for each of the five different results on self-employed labor income. Imputing on average employee compensation lowers the rate of return, and direct measures give the highest rates, with mixed income based measures offering a middle ground.

Part of this work is kindly financed by Eurostat under the ESTAT 2019 PA4-C-GPA grant. The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

[1] Statistics Netherlands, Henri Faasdreef 312, 2492 JP The Hague, The Netherlands, e-mail: h.debondt@cbs.nl

1. Introduction

Self-employed and their labor input form a relatively sparsely discussed part of productivity measurement and productivity statistics. Although Diewert called for a better measurement of self-employed in this context, it has not resulted in a large body of research into this topic (Diewert 2008). The main difficulty lies in the estimation of their labor input, both in terms of income as well as labor volume. The system of national accounts recognizes that the income of self-employed constitutes several parts, i.e. a remuneration for capital, for labor and for entrepreneurship, but gives no practical solution for separating each item. It notes (SNA 2008, UN 2009, 7.9): "[…] the balancing item is described as mixed income because it implicitly contains an element of remuneration for work done by the owner, or other members of the household, that cannot be separately identified from the return to the owner as entrepreneur. In many cases, though, the element of remuneration may dominate the value of mixed income."

Primarily, statistical agencies, research institutes and universities rely on the concepts put forward in the OECD manual in their studies of productivity (OECD 2001). The manual, like the SNA, notes the difficulty of separating the labor part from the capital part of mixed income. Most likely, one of the main reasons for not studying the variable more closely is due to the unknown nature of the labor income of self-employed and overall greater importance of volume changes. To shed more light on this issue, and using data on the Netherlands, this paper provides estimates of self-employed labor income for 5 different methods and the contribution to value added growth in a growth accounting framework. The volume growth of labor inputs and the impact on productivity measurement is shown, by moving from a fairly simple set-up for the contribution of self-employed labor inputs to a more sophisticated one. This more sophisticated measure in this case is an application of the labor services concept which not only reflects the change in hours worked, but also reflects the shift in background characteristics of self-employed such as sex, age and level of education. Final results show the different endogenous rates of return to capital for each of the different methods of estimating labor income.

This paper is set-up as follows: part one discusses the related field of study of labor income and the growing importance of self-employed labor income for the Netherlands. Part two and three explain the concept of labor services and the growth accounting framework used in this paper. Part four shows a conceptual breakdown of different methods for estimating the labor income into three distinct "families". Part five, six and seven show the empirical set-up, data and results. Part eight shows the impact on the contribution of self-employed labor by moving from a relatively simple set of weights and volumes to a more sophisticated one. Part nine finally presents the endogenous rates of return to capital associated with each of the 5 estimates of self-employed labor income.

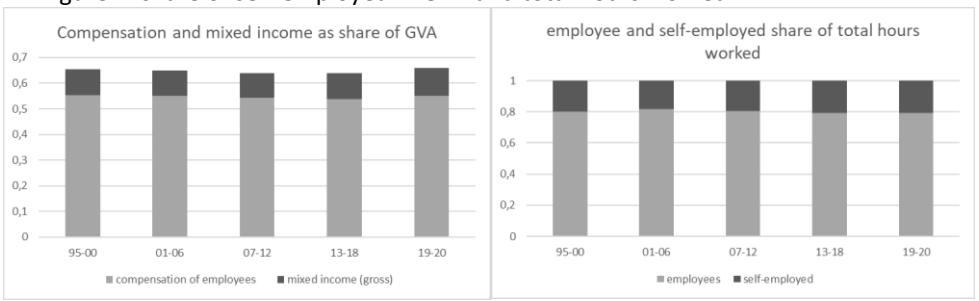
1.1 Self-employed and the study of the labor share of income

The study of the labor share, and especially it's supposed decline compared to the mid and late 20th century have dealt with estimates of labor income of self-employed extensively. Several authors have studied the topic, and used different methods. Authors making cross-country comparisons are much more limited in their data and consequently also in the methods that can be applied. Cho et al. use average wages of employees to impute the labor income, but also

adjust by 50% for the fact that self-employed income is usually lower per unit of labor input (Cho et al. 2017). Cette et al. use average hourly wages to impute, without any further adjustment (Cette et al. 2020). Pionnier and Guidetti also impute but assume the labor income per head are the same for employees and self-employed (Pionnier and Guidetti 2015). Finally Izyumov and Vahaly list several possibilities for estimating the labor income of self-employed based on a literature study. They find either a variant of imputed wages, or a measure of adjusted mixed income. Mixed income can be adjusted by a fixed share, e.g. 2/3, ½, or no adjustment thus assigning all mixed income to labor, or doing the contrary and assigning all of mixed income to capital (Izyumov and Vahaly 2015).
Studies focusing on a single country usually have more detail at hand and can therefore apply more sophisticated methods, e.g. Fukao and Perugini use both mixed income and compensation of employees and further adjust to give an estimate for Japan (Fukao and Perugini 2021). Van den Bergen et al. study several options for the Netherlands, but eventually find net mixed income the most suitable for studying the development of the labor share (Van den Bergen et al. 2017). Freeman uses imputation based on employee average wages to study the development of the US labor share, but adjusts for background characteristics (Freeman 2011). Mendieta-Muñoz et al. study the US labor share by assuming that the labor and capital shares for self-employed are the same as in the corporate sector (Mendieta-Munoz et al. 2021).

The choice of a suitable method depends on the data availability, but also on economic circumstances. The Netherlands, which serves as the further focus of this paper, is an example. It recorded a strong growth in the number of self-employed and their hours worked, however the growth of mixed income did not grow at the same pace. The difference between the two growth rates made a method of imputation with average employee compensation less suitable. Average employee compensation did still show growth, even after the global financial crisis whereas total mixed income (of self-employed) stagnated. The reasons behind this trend are the increasing flexibilisation of the labor market, and the lower rates of income taxes and social security compared to employees (OECD 2021a). These circumstances have spurred the growth in total hours worked, but not in the labor share of income. Measured as a percentage of hours worked, the self-employed increased from 18 % in the period 01-06 to 21 % in the periods 13-18 and 19-20. Covid has had a strong and negative influence on self-employed, due to their large numbers in affected industries such as hotels and restaurants and services. Their share in total hours dropped, but the mixed income share in gross value added remained stable, helped by government subsidies.

Figure 1. share of self-employed in GVA and total hours worked



1.2 Self-employed and labor services

The concept of labor services (cf. Timmer et al. 2007) or QALI[2] is widely used in productivity measurement. Most applications however focus on splitting the hours worked and compensation of employees, without explicitly taking into account the labor services of self-employed. Labor services are generally regarded as a combination of an hours worked component and a compositional change component. The compositional change quantifies the quality of the hours worked, i.e. not all hours worked are treated equally[3]. It is derived as the ratio of a weighted labor volume and the change of total hours worked (Timmer et al. (2007), Statistics Netherlands (2015)). The weights needed to calculate the labor services are usually based on level of education and age, as proxies for skill and experience. On other occasions sex is also taken into account, as well as the economic activity. Data sources for these breakdowns are available for employees, but usually not for self-employed. The main reason for this lack of data relates to the concept of labor income of self-employed. Since that income is unknown, or not measured, the weights are also unknown. Several ways of getting around this omission have been put in practice. Statistics Netherlands for instance have assumed the same compositional change as employees in the growth accounts tables (Statistics Netherlands 2015). Eurostat estimates the labor services using, among other sources, the labor force survey (LFS), which contains information on self-employed. In a final step they gross up the weights to total compensation of employees and hours worked of both employees and self-employed, conforming to official NSI-publications (Eurostat, no date). This can however be biased due to the concentration of self-employed in specific industries.

1.3 Growth accounting and the contribution of self-employed

A general presentation of the value added growth accounting formula, adapted from Balk (2018), reads:

$$Q_{VA}(1,0) = ITFPROD_{VA}(1,0) \cdot Q_{KL}(1,0) \qquad (1)$$

Where $Q_{VA}(1,0)$ denotes the quantity index of value added between periods 1 and 0, $ITFPROD_{VA}(1,0)$ is the index of multi-factor productivity in a value added growth accounting framework and $Q_{KL}(1,0)$ denotes the quantity index of capital and labor. Industry subscripts have been left out for matters of clarity and convenience, but the equations apply to any level of economic activity. Because data in this paper uses Dutch growth accounts tables, this paper uses a Laspeyres-type quantity change, which conforms to method used for the Dutch growth accounts. This quantity change of inputs is then:

$$Q_{KL}(1,0) = \alpha \cdot Q_L(1,0) + (1-\alpha) \cdot Q_K(1,0) \qquad (2)$$

Where α denotes the share of labor in total capital and labor inputs, $Q_L(1,0)$ denotes the quantity index of labor and $Q_K(1,0)$ denotes the quantity index of capital. Because of the Laspeyres-type indices the share of labor is calculated as:

$$\alpha = \frac{L^0}{L^0 + K^0} \qquad (3)$$

[2] Quality adjusted labor input or quality adjusted labor index.

[3] The term quality can have a normative meaning and is usually best avoided. See also Timmer et al. (2007). The OECD voiced similar concerns, which led them to advise against the use of sex as a dimension of labor inputs in productivity studies (OECD 2021c).

Where L^0 and K^0 denote the nominal inputs of labor and capital in the base period (0) respectively. The quantity of labor consists of the contribution of the quantity change of employees and the contribution of self-employed:

$$Q_L(1,0) = \beta \cdot Q_{SE}(1,0) + (1-\beta) \cdot Q_{EMP}(1,0) \quad (4)$$

Where $Q_{SE}(1,0)$ and $Q_{EMP}(1,0)$ denote the volume index of self-employed and employees respectively. The β denotes the share of self-employed in total labor inputs and is defined as:

$$\beta = \frac{L^0_{SE}}{L^0_{SE}+L^0_{EMP}} \quad (5)$$

Where L^0_{SE} and L^0_{EMP} denote the labor income of self-employed and compensation of employees in the base period respectively.
In this paper, the contributions to value added are shown, which results in[4]:

$$\frac{\Delta VA}{VA^0} = \gamma \cdot \frac{\Delta L_{SE}}{L^0_{SE}} + \delta \cdot \frac{\Delta L_{EMP}}{L^0_{EMP}} + \theta \cdot \frac{\Delta K}{K^0} + CTFPROD_{VA} \quad (6)$$

Where the weights sum up to 1:

$$\gamma + \delta + \theta = \frac{L^0_{SE}}{L^0+K^0} + \frac{L^0_{EMP}}{L^0+K^0} + \frac{K^0}{L^0+K^0} = 1 \quad (7)$$

Labor volumes can be calculated as weighted index of hours worked, using a more complete set of background characteristics (QALI):

$$Q_{SE}(1,0) = \sum_j s_{SE,j,t} \cdot Q_{H,SE,j}\,(1,0) \quad (8)$$

Where the subscripts j denote the QALI-characteristics such as industry, sex, age and level of education, H indicates hours worked and t denotes the period, in this case the base period. The share of each background characteristic in the total is denoted by s.

As can be derived from equations 1-6, the assumptions and method for calculating the labor income of self-employed impact the contribution of labor to value added growth and multi-factor productivity growth.

1.4 Families of self-employed labor income

The little amount of attention to the labor income of self-employed in the context of productivity measurement has resulted in a less well understood impact of this group of workers on productivity. As opposed to the studies on the labor share, in growth accounting often more simple assumptions are used to calculate the labor income of self-employed. Compilers of these statistics generally do not provide much data on self-employed and the impacts of different methods or assumptions are generally not available. A small survey, shown in table 1, on the labor income in different growth accounts systems across Europe and in international datasets show that most countries and institutes impute the labor income of self-employed based on hourly compensation or (annual) compensation per employee. Some

[4] Due to the use of the Laspeyres index, the contribution of total factor productivity (CTFPROD) is close but unequal to the index of total factor productivity, i.e.: $CTFPROD_{VA} \approx ITFPROD_{VA} - 1$

countries, the Conference Board and the Penn World tables use mixed income to derive the labor income.

Table 1

Method for estimating labor income of self-employed	Country or institute
imputed, hours worked	Eurostat, EUKLEMS, FI, IT, PL, DK, NO, SE
imputed, total number	OECD, DE, IE, CH
mixed income	Conference Board, Penn World Tables (#), NL, GB, BE

mixed income is the preferred method; Sources: Eurostat - Eurostat (2021); EUKLEMS - Jäger (2017); Conference Board - De Vries and Erumban (2017); Penn World Tables - Inklaar and Timmer (2013); OECD - OECD (2021b); for countries the websites of NSI's have been consulted; Netherlands method presented in this paper.

Grouping the methods in this manner does not give a complete fine-grained picture; within each group different methods are used. Italy for instance excludes assisting family members, the UK and Penn World Tables impute by applying the ratio of compensation of employees and gross operating surplus to split gross mixed income. The Netherlands and Belgium both use a scaling factor or pro rata adjustment for labor and capital inputs.
A less researched option is to directly calculate labor income of self-employed from microdata sources. This option is much less researched given the conceptual limitations, i.e. is it possible to directly extract labor income, and more practical issues such as coverage, exhaustiveness, non-response etc.? However, the OECD note that “improved estimates can only be obtained from surveys of the self-employed or from census questions that specifically target the self-employed” (OECD 2001). This yields a total of three ‘families’ for calculating labor compensation. They are coined families, because within each family different variants can be calculated. It is probably impossible to list all different possibilities, but generally each compiler comes across a range of similar choices when a certain family of estimation is chosen.

For the family of imputed labor income most importantly the average of employee compensation has to be chosen. Hourly, annual (fte), number of jobs or number of employees are each viable options, if the same variable is available for self-employed. Industry breakdowns provide more detail and are often used, but even more detail can be used such as age, sex and education (Freeman 2011). Compilers often choose average compensation, but national accounts and labor accounts sometimes provide wages and salaries, and wage costs.

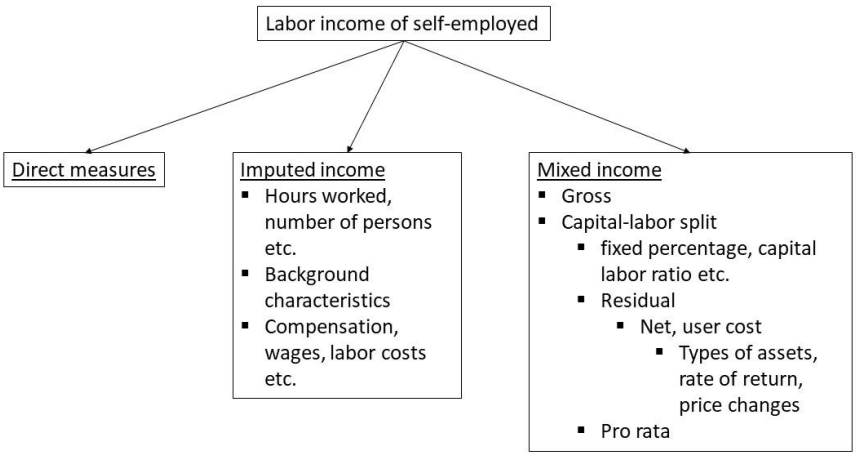
Using mixed income first forces the compiler to choose whether to split the income into labor and capital or not. Not splitting yields gross mixed income, assigning all self-employed income, to labor. Splitting has relatively simple applications, e.g. applying the ratio of compensation of employees to operating surplus or assuming a fixed share. Net mixed income, effectively assigning depreciation of fixed assets to capital and the residual to labor, is still relatively simple because most national accounts from advanced economies provide this data. It is however more difficult than the previous two options, because it usually requires industry estimates of fixed asset stocks for the self-employed. Another residual option is to calculate the user cost of assets and subtract these from gross mixed income. When taking this approach the compiler is faced with estimating user cost of self-employed, entailing choices on which asset types to distinguish, the rate of return and possibly also the price changes of these assets.
Often user cost calculations rely on the stock of fixed assets, conveniently linking to the PIM (perpetual inventory method) employed by statistical offices and other institutions. However, user cost calculations can also be extended to other, non-PIM type, assets such as land and

inventories.[5] The rate of return is most often calculated so that it completely exhausts value added. This means the sum of labor and capital is equal to value added, all expressed in current prices. The rate of return is backed out so that the user cost formula can be applied, i.e. the rental price multiplied with the stock of assets. Statistics Netherlands (Van den Bergen et al. 2008), Switzerland (BFS 2013) and the OECD (OECD 2021b) for example all calculate the user cost by applying exogenous rates of return. These are based on different sources, and do not necessarily result in exhausting value added.
The prices used in the user cost formula express the holding gains or losses associated with owning an asset for a fixed period. These prices can either be sourced from the official statistics, such as the price changes found in the balance sheet of non-financial assets, or a true ex-ante or forward-looking price can be used, where the expected holding gains or losses are calculated instead of realized gains or losses. Switzerland and the OECD for example both calculate these prices based on moving averages of historical prices (BFS 2013 and OECD 2021b).

Even more difficult is the pro-rata adjustment of mixed income (OECD 2001, ABS 2015). This method relies on separate estimates of imputed labor income and user costs, which are then adjusted so that both components sum up to gross mixed income. The compiler that chooses this method has to concern him or herself with the issue of choosing the imputation method, in the same way a compiler that chooses the family of imputed income has to. Also, a choice on the user cost calculation has to be made, as described in the residual case of mixed income.

Schematically, an overview of the possibilities is given in figure 2.

Figure 2. Families of self-employed labor income



1.5 Empirical set-up

Results are calculated and shown for 5 different measures of labor income, based on data available at Statistics Netherlands. They are 1.) direct measures; 2.) imputed with hourly compensation by industry and 3.) imputed with hourly compensation using all background characteristics ("bc"); 4.) net mixed income and 5.) the pro rata adjustment using hours worked

[5] An example of user cost calculations of land can be found in Cho et al. (2017).

and user costs of fixed assets, land and inventories. This distributes the estimates per family of labor income more or less evenly, with one estimate for the direct family, and two estimates each for the imputed and mixed income family.

Direct measures

A direct measure is difficult to find, and would have to come from directly surveying self-employed and asking about their labor and capital income. Still, this split might be subjective or lead to non-response. An possibly interesting source are the mixed income statistics. These statistics are used in the national accounts to calculate (gross) mixed income of households. A suitable variable is the fiscal profits, since all costs, depreciation (fiscal depreciation) and interest received and payed is subtracted from the income. This equals the fiscal profit reported by the self-employed to the tax authorities. It is important to note that those figures on depreciation and interest differ from the results of national accounts figures on depreciation (consumption of fixed capital), which is calculated using a PIM. Moreover, some subgroups of the self-employed are not included in the tax records. Labor accounts adjust for groups not covered, i.e. the illegal and grey economy, family members active within the family enterprise, auxiliary jobs of self-employed, freelancers and domestic help, but does so at the macro level. The adjustments are therefore not available at the micro level, leaving out some groups.

Imputed income

The average compensation of employees is used to estimate the income of self-employed. For both groups, i.e. employees and self-employed, a volume measure is needed to calculate the average. These can be hours worked, full time equivalents, number of jobs or number of persons.

$$L_{i,t,SE} = \frac{L_{t,i,EMP}}{V_{t,i,EMP}} V_{t,i,SE} \quad (9)$$

Subscript i can be either only the industry code or a combination of background characteristics such as age, sex and or level of education. Subscript EMP denotes employees. Summing over all background characteristics (industry or all) gives labor income:

$$L_{t,SE} = \sum_i L_{i,t,SE} \quad (10)$$

Mixed income

Net mixed income is calculated by subtracting the consumption of fixed capital from gross mixed income.

$$L_{i,t,SE} = NMI_{i,t} = GMI_{i,t} - CFC_{i,t,SE} \quad (11)$$

Where L denotes the labor income of self-employed, NMI, GMI and CFC denote net mixed income, gross mixed income and consumption of fixed capital respectively. Subscripts i and t denote the industry and time.

The pro rata estimates presented here include all the hourly imputed results presented in the previous paragraph using only the industry breakdown, and user costs for fixed assets, inventories and land. The user cost of fixed assets are given by the “standard” user cost formula (OECD 2009), and more precisely explained in Van den Bergen et al. (2008):

$$U_{j,t,SE} = NCS_{j,t-,SE} \cdot (r_{t+t-} - i_{j,t+t-} + \delta_j) \quad (12)$$

Where U_j denotes the user cost of fixed asset type j, NCS_{t-} the net capital stock at the start of reporting year t, r_{t+t-} the rate of return applicable to the entire reporting year, $i_{j,t+t-}$ the price change of asset j during the entire reporting year and δ_j the depreciation rate of asset j.

On average, and by national accounts convention, the investments[6] are done half-way through the year, and also generate user costs:

$$V_{j,t,SE} = I_{j,t,SE} \cdot (r_{t+t} - i_{j,t+t} + \left(1 - \sqrt{1 - \delta_j}\right)) \quad (13)$$

The user costs of investments are labelled "V" and the investments are labelled "I". For inventories and land[7] the user costs are given by:

$$U_{k,t,SE} = NCS_{k,t+t-,SE} \cdot r_{t+t-} \cdot \left(1 + i_{k,t+t-}\right) + NCS_{k,t+t,SE} \cdot r_{t+t} \cdot \left(1 + i_{k,t+t}\right) + NCS_{k,tt-,SE} \cdot r_{tt-} \cdot (1 + i_{k.tt-}) \quad (14)$$

Here the subscript k denotes the asset types inventories or land and the net stock is split into three parts: the remaining stock for the entire reporting year (NCS_{t+t-}), the (net) additions or purchases (NCS_{t+t}) and the (net) withdrawals/sales (NCS_{tt-}).
Total user costs of self-employed are then the sum of the three parts:

$$U_{t,SE} = \sum_j U_{j,t,SE} + \sum_j V_{j,t,SE} + \sum_k U_{k,t,SE} \quad (15)$$

Adding up the results from equations 10 and 15 gives the estimate GMI-hat (ABS 2015). The time subscripts are dropped for matters of convenience:

$$\widehat{GMI} = \widehat{L_{SE}} + \widehat{U_{SE}} \quad (16)$$

The pro rata factor is then calculated as:

$$s = \frac{GMI}{\widehat{GMI}} \quad (17)$$

So that the labor income of self-employed is calculated as:

$$L_{SE} = s \cdot \widehat{L_{SE}} \quad (18)$$

1.6 Data

Direct measures

This data is based on the databases that are also used in compiling mixed income in national accounts. The statistical source is the satellite account on self-employed (Satelliet zelfstandige ondernemers, SZO). Its main source is tax record data, which can be linked to the business

[6] In national accounts investments are usually called gross fixed capital formation (GFCF). Given the net definition, i.e. sales and purchases of second-hand assets are taken into account, it is not appropriate to label investments here as GFCF.

[7] The Dutch growth accounts include user costs of mineral and energy reserves, but is only used for the mining and quarrying industry. This industry does not have any self-employed, so the inclusion of these user costs is not relevant for this paper.

register so that it also contains industry codes. Total years available range from 2007 up until 2018. The source data only contains mostly fiscal information on self-employed and is subject to several updates resulting in full coverage of self-employed as considered by the tax authorities. Definitive results are only available at t+3 years, because self-employed can get an extension for their tax declaration. For the groups not covered no separate estimates have been made. This results in an undercoverage, which has to be kept in mind when judging the results.

Imputed income

The data for the imputed compensation are derived from labor accounts data. Here the hours worked, compensation of employees, as well as hours worked of self-employed are available for the years 1995-2018 at several industry classification levels, including industry level breakdown. The hours worked for the self-employed including background characteristics, i.e. sex, industry, education level and age are derived from the tax records on self-employed, the income register[8], income statements and the labor force survey (LFS). This combination of sources is only available for the years 2010 up to and including 2017.

As a final adjustment the hours worked are matched at the officially published data on hours worked by industry. It was however noticed that the data contained a relatively large group of self-employed for which the level of education was unknown. Generally the group for which education was known tended to be highest educated, therefore making proportional adjustment based on the recorded levels of education not feasible. Published results of the LFS by educational attainment were set as a benchmark, which revealed older age groups needed more adjustment. It was assumed that for these groups the coverage of the highly educated is better than for lower levels of education, thereby disproportionate share of low and medium educated are in the unknown category. To compensate for this bias, the number of low and medium educated were raised by 50%. The differences in target population between LFS and national accounts were assumed not to influence the imputation (national vs. concept), i.e. the initial shares from LFS are assumed to be representative for the groups based on national accounts definitions.

For the years 2001 up until 2006, a separate dataset was compiled on hours worked of self-employed along the age-group, education level and sex dimensions, based on LFS. Originally, these data were used in an older vintage of the Dutch growth accounts publication and based on the SIC'93 industry classification. For the purpose of this research, the industry classification was converted from SIC'93 to SIC 2008 and subsequently hours worked were adjusted to comply with published totals on hours worked by labor accounts.

For the remaining years 2007-2009 the integral income survey was used to interpolate. For these years, data on number of self-employed, excluding owner-managers who are on the pay-roll, are used at an aggregate level. These results are then converted to hours worked by using the ratio of persons employed and hours worked from labor accounts. As a final step the results are adjusted to totals of hours worked by industry and sex as published by labor accounts.

The time series (2001-2017) for the full set of background characteristics provides hours worked for self-employed according to the A21-industry classification, sex (male – female), 6 different age groups (<25, 25-<35, 35-<45, 45-<55, 55-<65, >65) and 3 levels of education (low, medium and high).

[8] The income register is itself partially based on tax data but also relies on other sources such as information on subsidies.

Hours worked and compensation of employees according to the age, sex and education breakdown is derived from the price index of labor. Published by labor accounts as a separate statistics, this level of detail is then aggregated to the level of detail of self-employed. These results are also used for the full QALI estimates of self-employed.

Mixed income

This group relies on extracting data from different sources. Net and gross mixed income are available at the national accounts databases at Statistics Netherlands. The pro rata measures use the results from imputed income based on hours worked by industry. The user cost of fixed assets use the data from the Dutch PIM system, and user cost of land and inventories reserves are compiled using data from the balance sheet for non-financial assets. The price changes used for calculating holding gains and losses are directly available in the Dutch PIM and the balance sheet for non-financial assets. The real rate of return used in the user costs calculation for self-employed is the same as the general user cost used in the Dutch growth accounts tables, and vary per year (Van den Bergen et al. 2008).

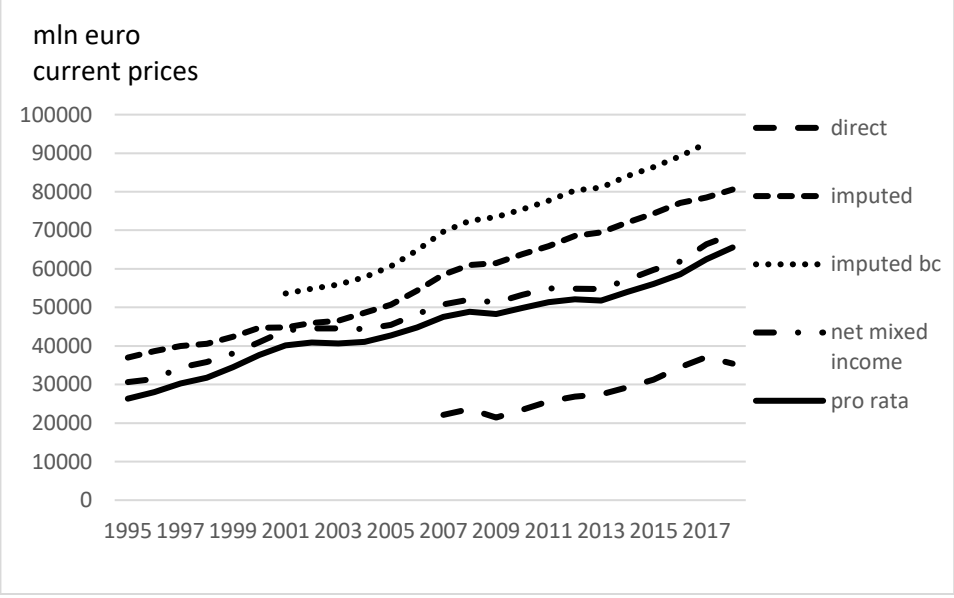
Table 2. real rates of return for user cost calculations

95-00	01-06	07-12	13-18
0.05	0.04	0.04	0.03

1.7 Results

The years covered vary from 2007 – 2018 for the direct estimates, 1995-2018 for the imputed hourly compensation by industry, net mixed income and the pro rata estimates, and 2001-2017 for the imputed measures using all background characteristics. The results are presented at the market sector level, which includes all industries except the real estate sector, public administration, education and households as employer. This aggregate is also used in the growth accounts tables of Statistics Netherlands.

Figure 3. Labor income of self-employed, market sector 1995-2018



The estimates vary significantly, with the imputed income including background characteristics as the highest estimate and the direct estimate as the lowest. The high estimate when including background characteristics is due to the fact that self-employed are much more common in the higher skilled categories. The high amount of hours worked compared to other categories gives them a comparatively high imputed income and push the total of imputed income up. This fact is also reflected in the other measure of imputed income, which is clearly above both mixed income family estimates. The direct estimates are unsurprisingly the lowest, because it only contains data from tax authorities and does not cover groups such as the grey and illegal economy and auxiliary family members.

Table 3. Average yearly contribution to value added and share of total inputs

		95-00	01-06	07-12	13-18
direct	contribution to value added %points	-	-	0.02	0.15
	share of total inputs	-	-	0.06	0.07
imputed	contribution to value added %points	0.05	0.06	0.13	0.32
	share of total inputs	0.15	0.14	0.14	0.16
imputed bc	contribution to value added %points	-	0.16	0.15	0.36
	share of total inputs	-	0.16	0.17	0.18
net mixed	contribution to value added %points	0.04	0.05	0.11	0.27
	share of total inputs	0.13	0.13	0.13	0.13
pro rata	contribution to value added %points	0.04	0.05	0.11	0.25
	share of total inputs	0.12	0.12	0.12	0.13

* 2001 and 2018 unavailable

The contributions are calculated using the same volume measure, the change of total hours worked for the market sector. The contributions from the imputed measures are highest, especially when including background characteristics for the years 2001-2006. This is in part also due to results lacking for 2001, when hours worked declined, which pushed up the average. As expected the direct contributions and shares were lowest, and mixed income and pro rata falling in between. The difference between imputed income and mixed income estimates are attributable to the strong (fiscal) incentives for self-employed. Sectors such as construction and business services have seen a strong increase in their number. Their increase in numbers and hours worked has not been matched by a growth in income similar to the income growth of employees. A preliminary conclusion is therefore that imputing labor income is not a suitable measure for the self-employed. The estimates from the mixed income family do a better job, and their results vary only slightly. Opting for the pro rata estimates has the advantage that it employs the user cost concept used widely in growth accounting, moreover does a better job separating out labor and capital in industries where land and inventories are more important, such as agriculture and trade.

1.8 The impact of self-employed on productivity measurement

Moving from a more simple approach of calculating the contribution of labor with hours worked (by industry) and imputed labor (also by industry) to a more sophisticated measure using the QALI and pro rata results impacts the contribution of labor. Decomposing the shift gives a quantification of each factor, clarifying whether the weights changes or volume changes are more important. Although each of five estimates of labor income can be used, the imputed measure and pro rata measures are selected, because these are put in to practice in growth accounts tables. The target variable is the contribution of self-employed labor to value added growth, so starting points are equations 6 and 8. First to simplify notation, the volume change of labor for self-employed is denoted $\acute{L}$ ($\acute{L} = \frac{\Delta L_{SE}}{L^0_{SE}}$). The decomposition taken from De Haan et

al. (2014) is used, where the first part of the decomposition shows the changes in shares of self-employed weights, i.e. from imputed weights (γ_1) to pro rata weights (γ_2), and the latter part the changes in the volume index used, by moving from a volume index determined by industry weights (L_1) to a full QALI for self-employed (L_2).

$$\gamma_2 \cdot \acute{L_2} - \gamma_1 \cdot \acute{L_1} = 0.5\Delta\gamma \cdot \left(\acute{L_1} + \acute{L_2}\right) + 0.5 \cdot (\gamma_1 + \gamma_2) \cdot \Delta\acute{L} \qquad (19)$$

Where $\Delta\gamma = \gamma_2 - \gamma_1$ and $\Delta\acute{L} = \acute{L_2} - \acute{L_1}$.

Figure 4. Decomposition of the contribution of self-employed labor to value added growth, 2001= 1

Two interesting observations arise from figure 4. First, at the start of the series the cost shift had very little impact on the more sophisticated measure, and the growth shift explained most if not all of the shift. From 2005 on, the cost shift became more important and pushed the contribution of the sophisticated measure downwards. Second, after the financial crisis the importance of the growth shift faded, and the weight shift became much more important. The combination of factors resulted in a contribution of the more sophisticated measure that was greater than the simple measure right up until the financial crisis, but right afterwards it dropped below the more simple measure. The considerable impact of the cost shift was also one of the reasons for changing the official calculations of the labor share in the Netherlands, where imputing was abandoned in favor of adopting net mixed income (Van den Bergen et al. 2017). It clearly shows that the share of self-employed in income in these later years did not rise as fast as their hours worked would have suggested. The small contribution of the growth shift in later years is a reflection of the growth of younger self-employed, whose hours have a relatively smaller weight.

Lastly it should be noted that the L_2 measure only uses the relative weights to sum up a volume index. The share in total inputs (K+L) is determined by the pro rata weights, instead taking the relative volume weights as contribution weights would give a larger share of total inputs and a cost shift that wouldn't decrease the contribution, as can be seen from table 3.

1.9 Self-employed and the rate of return

User cost can either be determined with and exogenous rate of return, or an endogenous rate. The exogenous rate uses for instance government bond rates or, as used in the pro rata calculations in this paper, the internal reference rate of banks. The endogenous calculations rely on an accounting identity where capital income is determined as a residual. Commonly, value added is taken as the total of labor and capital income, so that capital income is the difference between value added and labor income. Another approach is to take the sum of operating surplus and mixed income and subtract the labor income of self-employed. The difference between the two approaches is whether to assign taxes and subsidies directly to one of each production factor, to leave them out, or to assume all taxes and subsidies are part of capital income. For this analysis taxes and subsidies are left out of the discussion, resulting in an endogenous capital income:

$$K_{end} = GOS + GMI - L_{SE} \quad (20)$$

Here GOS denotes the gross operating surplus. Equation 20 shows the impact of the labor income of self-employed on the endogenous capital income and the importance for the rate of return. Usually, the rate of return in the endogenous approach is backed out using information on the capital stock and asset price changes (e.g. Eurostat 2021). Because other types of capital besides fixed assets feature in this paper, a slightly different method is used to back out the rate of return. First the user cost with a zero rate of return is calculated and summed after which the rate can be backed out using equation 20. It should be noted that calculating the rates net of taxes normally yields lower rates than endogenous capital income including taxes.

Two caveats relate to the mining and quarrying industry and the pro rata estimates. For the mining and quarrying industry Statistics Netherlands imputes all residual capital income to the subsoil assets oil and gas (De Haan et al. 2014). This is essentially a combination of exogenous and endogenous capital income, because the fixed assets-user cost are calculated with the rates of return from table 2. The residual is calculated as the difference between gross operating surplus and the exogenous user cost of fixed assets of the mining and quarrying industry. Because no self-employed work in the mining and quarrying industry in the Netherlands, this industry is not relevant for the analysis here and excluded from the market sector. A similar caveat can be made for the pro rata estimates. The method needs user cost for self-employed to be calculated beforehand so that the pro rata factor can be calculated. As a result, an implicit endogenous user cost for self-employed is already available ($K_{SE} = GMI - s \cdot \widehat{L_{SE}}$). The rate of return implied by this method likely differs from the “true” endogenous rate. Adjusting for this complication is not possible, since it would entail calculating two residuals at the same time. For practical matters the results of the pro rata labor income are used without any further adjustments in equation 20.

Table 4. Average yearly endogenous real rates of return, market sector (excl. Mining and Quarrying)

		95-00	01-06	07-12	13-18
all assets	direct	-	-	0.11	0.11
	imputed	0.05	0.07	0.07	0.07
	imputed bc *	-	0.06	0.06	0.06
	net mixed income	0.06	0.07	0.08	0.08
	pro rata	0.07	0.08	0.09	0.08

* 2018 unavailable

The results show the lowest estimate for the rate of return for imputing with background characteristics, reflecting the highest labor income for self-employed. The highest rate of return, calculated with direct estimates, reflects the small labor income for this group. Differences between high and low amount to a maximum of 8 percentage points. The table also show that differences between more common approaches, e.g. imputed and pro rata still differ between 2 and 3 percentage points over the period studied. It is clear from the table that the assumptions for the labor income of self-employed matter. Comparing the results from table 4 to table 2 show much higher rates for the endogenous rates. The results from table 4 should however not be used to call the exogenous rates into question. As Fatica noted, "serious statistical and economic concerns arise for the calculations of the endogenous rates" (Fatica 2017). Balk raises similar concerns over the use of restrictive models which are empirically refuted, such as the restriction of summing all inputs to value added and compiling growth accounts with endogenous rates (Balk 2010). Finally Inklaar argues that endogenous rates of return are only useful when capital income is measured correctly and all capital assets are accounted for, which he finds doubtful in practice (Inklaar 2010). The dependency of endogenous capital income on the self-employed labor income estimates confirm this position and the results on the rates shown in this paragraph provide further support. Put more strongly, it should be questioned whether the estimates of labor income of self-employed should have any influence on the rate of return to capital. On a more practical level applying exogenous rates make a pro rata labor income estimate easier to implement, and prevents the need for estimating two residuals at the same time.

2. Conclusion

This paper addresses the issue of labor income of self-employed for productivity measurement. In contrast to studies of the labor share, self-employed have not received much attention for this field of study. The importance of correct measurement has increased due to their growing importance not only in the Netherlands, but also in other advanced countries. Compilers of productivity statistics should look at the available data and choose a method, usually belonging to a family as presented in this paper, that suits them most. The method chosen impacts productivity measurement through the contribution of labor, which is even more strongly impacted with a QALI type labor volume. What is even more striking is the impact on the endogenous rate of return. One of the goals of this paper is to highlight that link and cast doubt on the usefulness of endogenous rates of return. Next to the objections made by other authors, the dependency of these rates on the labor income are an example of the disadvantages of these rates. Self-employed labor income matters for productivity measurement and therefore more international consensus from compilers of these statistics on the most preferred methods is in order. This can help create more comparable and more accurate productivity statistics.

References

Australian Bureau of Statistics (2015), 'Chapter 19 Productivity measures', Australian system of national accounts. Concepts, sources and methods (https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/australian-system-national-accounts-concepts-sources-and-methods/2020-21/chapter-19-productivity-measures).

Balk, B. (2010), 'An assumption-free framework for measuring productivity change', Review of Income and Wealth 56 Special Issue 1/2010 (https://doi.org/10.1111/j.1475-4991.2010.00388.x)

Balk, B. (2018), 'Empirical productivity indices and indicators', in: The Oxford Handbook of Productivity Analysis (2018, Grifell-Tatjé, Knox-Lovell and Sickles eds.), (https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780190226718.001.0001/oxfordhb-9780190226718-e-2).

Bergen, D. van den, M. van Rooijen-Horsten, M. de Haan and B. Balk (2008), 'Productivity measurement at Statistics Netherlands', CBS working paper The Hague Heerlen 2008 (https://www.cbs.nl/-/media/imported/onze%20diensten/methoden/dataverzameling/aanvullende%20onderzoeksbeschrijvingen/documents/2008/44/2008-03-x41-pub.pdf).

Bergen, van den D., A. Bruil, B. Butler, A. van der Horst, H. Kranendonk, M. Mellens, M. Tanriseven, J. Verbruggen (2017), 'Herziening methode arbeidsinkomensquote', CBS-CPB-DNB paper (https://www.cpb.nl/sites/default/files/omnidownload/CBS-CPB-DNB-paper-Herziening-methode-arbeidsinkomensquote.pdf).

BFS (Bundesamt für Statistik, 2013), 'Multifaktorproduktivität Methodenbericht', BFS Aktuell (https://www.bfs.admin.ch/asset/de/349131).

Cette, G., L. Koehl and T. Philippon, 'Labor share', Economics Letters 188/2020 (https://doi.org/10.1016/j.econlet.2020.108979).

Cho, T., S. Hwang and P. Schreyer (2017), 'Has the labor share declined? It depends.', OECD Statistics Working Papers 2017/01 (https://doi.org/10.1787/2dcfc715-en).

Diewert, E. (2008), 'What is to be done for better productivity measurement', International Productivity Monitor 16/2008 (http://csls.ca/ipm/16/IPM-16-diewert-e.pdf).

Eurostat (no date), 'Methodological document on labour productivity indicators for the EU-28: quality adjusted labour input',
(https://ec.europa.eu/eurostat/documents/7894008/8915486/Methodology_QALI.pdf).

Eurostat (2021), ''Crude' multi-factor productivity (MFP). Methodological note',
(https://ec.europa.eu/eurostat/web/experimental-statistics/multifactor-productivity).

Fatica, S. (2017), 'Measurement and allocation of capital inputs with taxes: a sensitivity analysis for OECD countries', Review of Income and Wealth 63 1/2017
(https://doi.org/10.1111/roiw.12199).

Freeman, R.A. (2011), 'Accounting for the self-employed in labour share estimates: The case of the United States', OECD Science, Technology and Industry Working Papers 2011/04 (https://dx.doi.org/10.1787/5kg0w877vlwg-en).

Fukao, K. and C. Perugini (2021), 'The long-run dynamics of the labor share in Japan', Review of Income and Wealth 67 2/2021 (https://doi.org/10.1111/roiw.12465).

Haan, M. de, E. Veldhuizen, M. Tanriseven and M. Van Rooijen-Horsten (2014), 'The Dutch Growth accounts: measuring productivity with non-zero profits', Review of Income and Wealth 60 Supplement Issue (https://doi.org/10.1111/roiw.12140).

Inklaar, R. (2010), 'The sensitivity of capital services measurement: Measure all assets and the cost of capital', Review of Income and Wealth 56 2/2010 (https://doi.org/10.1111/j.1475-4991.2010.00383.x).

Inklaar, R. and M. Timmer (2013), 'Capital, labor and TFP in PWT8.0', Groningen Growth and Development Centre (https://www.rug.nl/ggdc/productivity/pwt/related-research-papers/capital_labor_and_tfp_in_pwt80.pdf).

Izyumov, A. and J. Vahaly (2015), 'Income shares revisited', Review of Income and Wealth 61 1/2015 (https://doi.org/10.1111/roiw.12072).

Jäger, K. (2017), 'EU KLEMS Growth and Productivity Accounts 2017 Release, Statistical Module', (http://euklems.net/TCB/2018/Metholology_EUKLEMS_2017_revised.pdf).

Mendieta-Muñoz, I., C. Rada and R. von Arnim (2021), 'The decline of the US labor share across sectors', Review of Income and Wealth 67 3/2021 (https://doi.org/10.1111/roiw.12487).

OECD (2001), 'Measuring productivity. Measurement of aggregate and industry-level productivity growth', OECD manual (https://www.oecd.org/sdd/productivity-stats/2352458.pdf).

OECD (2009), 'Measuring capital - OECD manual 2009: second edition', OECD Publishing (https://doi.org/10.1787/9789264068476-en)

OECD (2021a), 'Economic survey of the Netherlands', (https://www.oecd.org/economy/netherlands-economic-snapshot/).

OECD (2021b), 'OECD Productivity Statistics Database Methodological Notes', (https://www.oecd.org/sdd/productivity-stats/OECD-Productivity-Statistics-Methodological-note.pdf?_ga=2.120793098.364331113.1655834469-366768726.1645697810).

OECD (2021c), 'Towards Improved and Comparable Productivity Statistics. A Set of Recommendations for Statistical Policy', (https://doi.org/10.1787/1ae0ec74-en).

Pionnier, P. and E. Guidetti (2015), 'Comparing profit shares in value-added in four OECD countries: Towards more harmonised national accounts', OECD Statistics Working Papers 2015/03 (https://dx.doi.org/10.1787/5js0bsm2g0lt-en).

Statistics Netherlands (2015), 'Productivity and labour composition', in: *ICT and Economic Growth*, (https://www.cbs.nl/en-gb/publication/2015/24/ict-and-economic-growth).

Timmer, M., T. van Moergastel, E. Stuivenwold, G. Ypma, M. O'Mahony and M. Kangasniemi (2007), 'EUKLEMS growth and productivity accounts version 1.0. Part I Methodology', (http://euklems.net/data/EUKLEMS_Growth_and_Productivity_Accounts_Part_I_Methodology.pdf).

United Nations (2009), SNA 2008, (https://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf).

Vries, K. de, and A. Erumban (2017), 'Total economy database. A detailed guide to its sources and methods', Conference Board (https://www.conference-board.org/data/economydatabase/total-economy-database-methodology).