

Hybrids and their status of employment

- how do we handle tax evasion in the name of statistics¹

Fredrik W. Andersson, Statistics Sweden, fredrik.andersson@scb.se

Abstract

To obtain a comprehensive picture of the population's status of employment, it is necessary to develop methods which identify individuals' who are self-employed and classify their status of employment. A problematic group are the hybrids, i.e., those individuals who are at the same time both employees and self-employed. It is neither new nor controversial that self-employed under-report their business income. However, to produce statistics we can allow them to have only one status of employment in each month: i.e., employed or self-employed. But how do we evaluate business income from self-employed individuals in contrast to their income from an ordinary employment position?

The previous studies are based on the fact that Engel curves, i.e., individuals with the same disposable income demand the same amount of a commodity, are offset in parallel between employees and self-employed. This study uses the same theoretical approach, but instead of studying the survey data on food consumption, register data on interest expenditures is used. All measures have their problems and additional study in this area is beneficial to conclude the level of tax evasion. The new method provides a level of under-reporting that is similar to the levels found in previous Swedish studies of the phenomenon of under-reporting.

The empirical result indicates that self-employed people in unincorporated business under-report about 50 percent of their business income. Hence, under-reporting is an aspect which is needed to be considered for an unbiased estimation of status of employment of the hybrids. Out of the total 110 thousand hybrids in 2021 around 15 percent are given self-employed as status of employment.

Keywords: Self-employed; hybrids; status of employment; administrative data; novel approach; under-reporting.

¹ I'm very greatfull to productive comments from Per Engström, participants of a brown-bag seminar at Örebro university, participants of an informal seminar at Swedish Tax Agency and Lucas Gameraov at Statistics Sweden.

1. Introduction

In an era when sample surveys are expensive and respondents are hard to find and extract information from, administrative data sources become more attractive. In addition, due to the process of digitalization, individuals leave more and more frequent administrative footprints after their activities. Hence, monthly statistics on population's statuses becomes possible for Sweden to produce. This unfolds opportunities to follow, e.g., cyclical labor market phenomena on a granular level.

To obtain a comprehensive picture of the population's status of employment, it is necessary to develop methods which identify individuals' who are self-employed and classify their status of employment. A problematic group are hybrids, i.e., those individuals who are at the same time both employees and self-employed. To be able to produce monthly statistics on, e.g., industry statistics, one of these statuses of employment must be chosen. And when we cannot measure the distribution of working time between these two occupations, we are compelled to look at which of these occupations generate most utility, in the form of bring in the highest amount of income. But how do you evaluate self-employed individuals and their business income in contrast to their income from an ordinary employment position? It is neither new nor controversial that self-employed under-report their business income. One SEK in income from a business entity is most likely worth more than one SEK in income from an employment.

In the academic literature, there are several studies that show that entrepreneurs underreport their business income. Studies on Swedish data occur. Internationally, sample surveys are used to study the underreporting phenomenon. They usually use data on food consumption. However, data of food consumption is from an intermittent household survey and samples are small in addition self-employed is a small group of people. Researchers need often to pool household surveys to have sufficient samples sizes.

It is also neither new nor controversial. But individuals who run business activities have more opportunities to under-report their income to the Swedish Tax Agency

than ordinary wage earner has.² Under-reporting can consist of entrepreneurs buying goods and services through his companies but using them privately, partly invoicing black or barter services/goods with other entrepreneurs in their network.

International studies have for some time tried to shed light on and estimate the size of the entrepreneurs' phenomena of under-reporting. Most of these studies are based on Pissarides and Weber's classic work from 1989 (Pissarides and Weber, 1989).

There are also studies that use Swedish Household budget survey data; e.g. Engström and Holmlunds (2009). Their study shows that entrepreneurs generally have a higher food consumption than employees (wage earners) given the same level of disposable income. Under-reporting is greater among entrepreneurs who run sole proprietorships than among entrepreneurs who run limited liability companies. The result is by the author interpreted as entrepreneurs are under-reporting income.

The purpose of this study is to present an alternative approach to studying entrepreneurs' level of under-reporting their entrepreneurial income by using administrative data; and in addition, how we given this information can hopefully more accurately choose a status of employment for the hybrids.

2. Method

2.1. Hybrids - two sides on the same coin

One a problematic group are thus the hybrids, i.e., those individuals who are at the same time both employees and self-employed. It is rather small phenomena on the labor market, around 110 thousand individuals are hybrids.

In the population by labor market status³ two different strings of methods determine the status of employee and self-employed, respectively. In a very simplified explanation. Those persons who during a month receive a wage income from an employer are employees. And a novel algorithm is constructed upon administrative data to identify active self-employed. The algorithm is beyond the purpose of this

² Even employees, probably with artisan professions, have great opportunities to exchange services with each other.

³ In Swedish: Befolkningens arbetsmarknadsstatus (BAS).

paper. Hence, for a small sub-group of self-employed we have some hybrids. Which often is a springboard to the become full time entrepreneurs, especially for women.

2.2. To choose status of employment

An individual will have the status of employment that gives highest income from. Hybrids are thus having income both from an employment and the self-employment. But how do we compare these two incomes? To complicate it even more, income from employment is pre-tax and income from self-employment is post-tax.

Hence, in the process of evaluate status of employment for the hybrids we require numerical values for two ratio: *i*) the F-tax ratio⁴ and *ii*) the level of under-reporting. The ambition is to calculate these ratios on similar approaches and at the same periodicity each year. Therefore, we argue to use a value from a moving average calculated on a lagged three-year period. It is based primarily on creating a less volatile model-based ratio since it becomes less sensitive to changes in a specific year.

2.3. How to evaluate under-reporting

Pissaride's and Weber's method is based on the assumptions that the preferences of the groups are the same and that individuals with the same disposable income demand the same amount of a commodity. In many cases, household spending is used for food. Engström and Holmlund (2009) use data from Statistics Sweden's survey on Household budget survey. To my knowledge sample surveys is used as a rule to study the phenomenon.

The approach of this study is based on the same approach as Pissarides and Weber (1989) and Engström and Holmlund, but instead individuals' interest expenses⁵ are now used to estimate entrepreneurs' propensity to under-report their business income.

Pissarides and Weber's approach is illustrated by the Engel curves in Figure 1. Let r denote interest expenses ($r=\ln(R)$) and y the disposable income ($y=\ln(Y)$) for all

⁴ The F-tax ratio is beyond the scope of this paper. We need this ratio since the income is post tax.

⁵ The interest expenses that are deductible and are reported on each individual's tax return data.

individuals. The two Engel curves shown in the figure below correspond to two log-linear interest expenses - disposable income profiles. A curve for entrepreneurs and a parallel-shifted curve for wage earners. What distinguishes the profiles is the intercept on the y-axis. This difference allows entrepreneurs to theoretically underreport their disposable income compared to wage earners' disposable income and thus end up at the same level of interest expenses, even though we assume their "true" levels of disposable income differ. When preferences of these groups of individuals are the same, the underreporting is equal to $(y^{wage\ earner} - y^{entrepreneur})$.

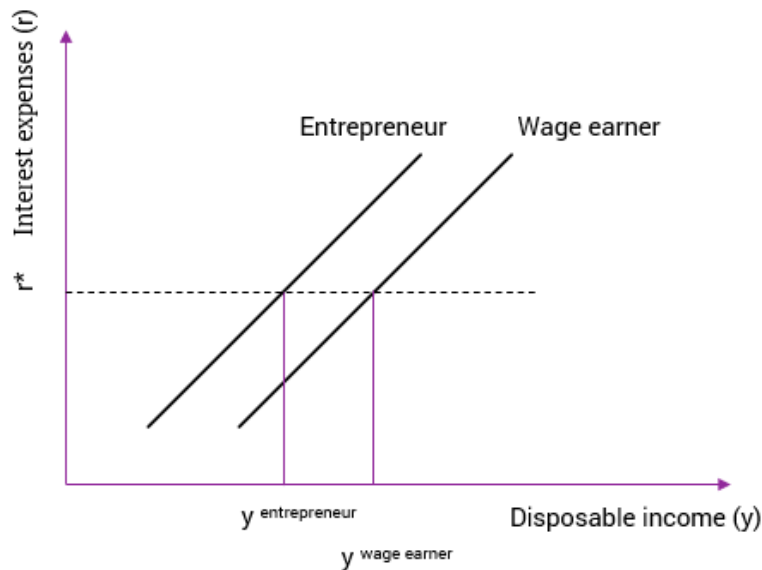
Thus, when data on entrepreneurs and wage earners are pooled into one dataset, the following equation can be estimated:

$$r_i = \mathbf{X}_i\alpha + \beta y_i + \gamma ENT_i + \varepsilon_i, \quad (1)$$

where i denotes individual i and X a vector of characteristics that affect individuals' interest expenses. ENT is a dummy variable (1,0) for whether individual i is entrepreneur (we will later distinguish between self-employed or owner of closely limited liability company) or not, and ε being a random error term. Parameter γ in the model captures the vertical distance between the Engel curves of the two different groups of individuals.

The degree of underreporting (in logs) is the same as $y^{wage\ earner} - y^{entrepreneur} = \gamma/\beta$. This underreporting can also be converted into a percentage $D = (\exp(\gamma/\beta) - 1) * 100$. Alternatively, an under-reporting ratio can be expressed as follows: $k = \exp(\gamma/\beta)$, which is the ratio by which the disposable income of entrepreneurs needs to be multiplied with in order to be equated with the "true" disposable income of entrepreneurs.

Figure 1: Engel curves for wage earners and entrepreneurs



3. Data

The data used in this illustrative example are for 2018 and comes partly from the Register-based labour market statistics (RAMS), partly from the Income and Taxation Register (IoT), and as well as the Apartment and Land Register. All registers are so called total register.

From RAMS a population of entrepreneurs and wage earners are retrieved and in addition their characteristics such as age, gender and education and so on, with the requirement that all individuals must be considered gainfully employed in RAMS. The population consists of employed⁶ individuals who are at one point in time only wage earners (employees) or only registered as entrepreneurs⁷. Individuals with any form of hybrid engagement either as business or as employees are not allowed. From the Taxation Register it is obtained information of individuals' levels of disposable income and interest expenses in addition information about ownership of a holiday residences from the Apartment and Land Register.

⁶ Employed in RAMS.

⁷ Information on the status of occupation is utilized here: self-employed, partners in close companies.

One advantage of this approach is that register data is utilized, which means that the study does not suffer from any non-response problems. At the same time it is less time-consuming and hence cheaper to utilize.

The main analysis's estimation procedure excludes observations that can be characterized as outliers. This means that 10 percent of the observations at both ends of the distribution of disposable income are removed, i.e., observations above the P90 value as well as observations below the P10 value. The main reason is that there exist individuals that have very low level of disposable income but rather high level of interest expenditure. By including these individuals, the model estimates a significant much larger degree of under-reporting. Thus, in the data sample we have 161,288 self-employed and 95,809 owner of closely limited liability companies and almost four million employees.

4. Result

4.1. Descriptive result

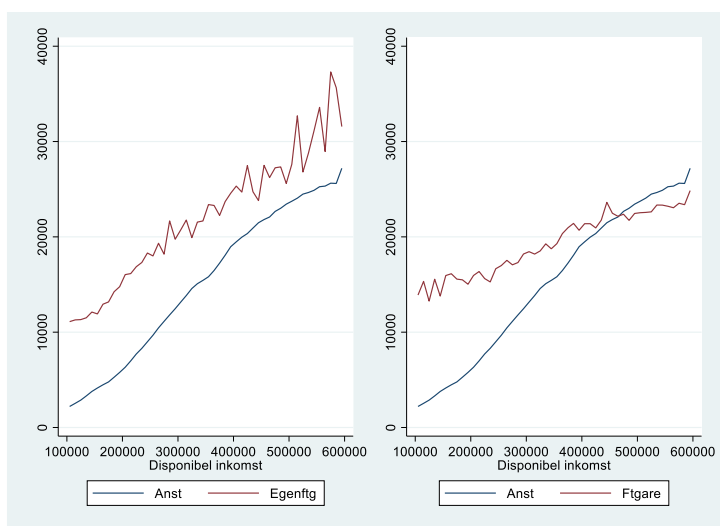
Engström and Holmström (2009) show according to their model that entrepreneurs under-report income but estimating the degree of under-reporting is a non-trivial pathway to undertake. They estimate the under-reporting of self-employed entrepreneurs' income to be around 50 percent and for the shareholders of the limited liability company to be 15-20 percent.⁸ In a later study by Engström and Hagen (2017), they find an under-reporting of 20-25 percent for the self-employed. The Swedish Tax Agency (2020) has also conducted a survey of self-employed people and their under-reporting. By auditing companies, they estimate under-reporting around 25 percent. The Swedish Tax Agency also refers to its own previous study (2006) that showed that self-employed people's under-reporting was 52 percent.

Figure 2 illustrates the relationship between individuals' interest expenses and their disposable income in the eve of their employment classification. More specifically, the average interest expenses for different SEK 10,000 are stated intervals per category

⁸ They estimate the coefficient of underreporting at the household level; for households with at least one contractor's.

(employee (wage earners), self-employed (self-employed) and close limited liability company owners (limited company)). Note that micro data on individual level is used.⁹ As can be shown in the figure 2, similarities can be noted with the theoretical parallel-shifted Engel curves in Figure 1, especially for self-employed people. Regardless of the level of disposable income, the interest expenses of the self-employed are constantly higher than the interest expenses of employees at the same level of disposable income. This obvious pattern indicates that self-employed people do underreport their business income. The higher variance depends on fewer observations.

Figure 2: Interest expenses and disposable income for employees, self-employed (left) and closely limited liability company owners (right)," in SEK.



Source: Statistics Sweden, own calculations.

The tendency is also the same for the owners of closely limited companies, but only within a range of disposable income. However, why do owners of closely limited company have another borrowing behavior compared to the earlier depicted borrowing behavior of self-employed individuals. From Figure 2 it is possible to see that above a certain threshold level of disposable income the Engel curves is no longer parallel-shifted. Does disposable income above that level reflect a group of more successful entrepreneurs who have had time to repay the loan stock to a greater extent? Or do more successful entrepreneurs stop under-reporting their

⁹ If the study were to be done at the household level, disposable income would increase

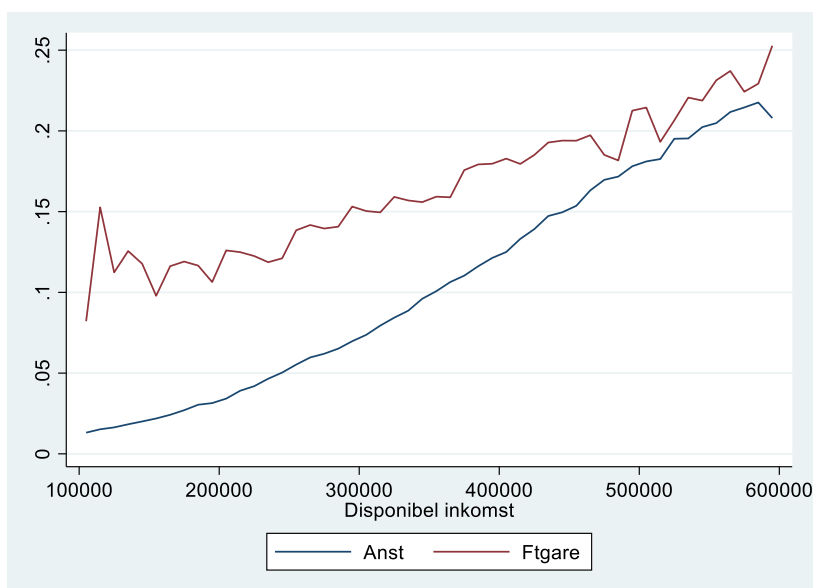
entrepreneurial income? The Swedish Tax Agency (2020) believes that closely limited company partners in companies with less turnover under-report because they cannot find evidence of the same behavior in the companies with higher turnover.

The notion now is to identify if there exists information of an individual consumption goods, from another total register, to measure the phenomenon of under-reporting.

4.1.1 The share who owns holiday residences

The idea is to study if the share of holiday residences differs like the assumed parallel-shifted Engel curves between wage earners and owner of closely limited companies. Owning a holiday residence cannot be seen as one of the basic needs of mankind and is definitely more of the nature of luxury consumption. And a weakness of current use of data is that we do not know if the holiday home is bought or passed down from previous generations and if it differs between the two groups of individuals. But from the depicted results shown in Figure 3, there undoubtedly is an indication that at the same level of disposable income a higher proportion of closely limited company owners own holiday residences compared to wage earners. The result in Figure 3 is more in line with the assumption of owner of closely limited companies under-reporting compared to the results from interest expenditures in Figure 2.

Figure 3: Share of holidays' residence owners and disposable income (SEK) for employees and owners of closely limited liability companies (owners).



Source: Statistics Sweden and own calculations.

4.2. Econometric results

The next step is to use an econometric approach to modelling interest expenditure on disposable income and control for difference in individuals' characteristics such as: age, sex, educational level, living status and industry belonging. An ordinary least square approach is used to estimate a model equal to equation 1. The econometric results strengthen the earlier signs that self-employed people under-report their incomes.¹⁰ The results from the regressions are presented in the appendix (A1 and A2).

From the coefficients in Tables A1 and A2 the level of under-reporting (k) is calculated according to: $k = \exp(\gamma/\beta)$. Given the econometric results we cannot reject that owner of legal entities do in *facto* under-report their business-related income – a kind of tax evasion. As shown in Table 1, the degree of under-reporting varies with the legal form of the business as well as the level of human capital (formal level of education) the entrepreneur has. There is an indication that entrepreneurs with higher degree of human capital tend to under-report a higher degree.

The under-reporting of owners of closely limited liability companies is more complicated. One reason might be that companies in this category are more heterogeneous than unincorporated business in terms of number of employees and turnover. Larger closely limited liability companies are more open to the public, owners of these companies may have more to lose if they are caught of under-reporting. When the under-reporting is estimated on the whole sample owners of closely limited liability companies (between P10 and P90), owners of closely limited liabilities companies tend to have a level of under-report at 1.6 percent. As we remember from Figure 2 the non parallel-shifted line of interest expenditure we re-run the estimation process with an impose a disposable income in the range of 100 and 370 thousand SEK. Now under-reporting of corporate income of 16.0 percent is obtained instead.

¹⁰ The degree of underreporting is looking to be stable. The same estimate was made for 2015 on data on all self-employed persons and joint-owners of limited liability companies. The rate was then estimated at 50.6 and 3.2 percent, respectively.

Table 1: The degree of under-reporting in 2018, as a percentage.

Group	All	Level of education		
		Pre-upper secondary	Upper secondary	Post-secondary
Self employed	52.9	50.3	52.9	56.4
Closely limited liability company	2.5	5.4	1.2	12.0

Source: Statistics Sweden and own calculations.

If we now instead of the individuals' interest expenses as dependent variable use the proposed alternative measure, i.e., the ownership of holiday residence the estimation of the econometric model generates a 48.8 percent level of under-reporting for self-employed. Thus, this level is in the vicinity of earlier estimated level (52.9). However, using ownership of holiday residence as a dependent variable for owners of closely limited liability companies gives a level of under-reporting of 18.7 percent, compared to 2.5. This indicates a much higher level of under-reporting.

4.3. Sensitivity analyzes

4.3.1. Entrepreneurs and their borrowing of capital

One weakness of the micro data of individuals interest payment approach is that the data in no way indicates whether and if so, to what extent interest expenses are linked to individuals' entrepreneurial activities. The notion is that ordinary wage earners borrow capital only for their private consumption. Do wage earners change their borrowing behavior after they become self-employed? To follow up this we identified in total 1 092 individuals who were wage earners in 2017 and became self-employed in 2018. The interest expenditure for this group increased with a median value of 2 SEK.¹¹

Non the less, one of the few things that business owners can really mortgage is their private real estate. Regardless of this weakness, the results implies that when Statistics Sweden's level of disposable income is the same for self-employed and

¹¹ With a mean value of 368 SEK between observations P10 and P90. Slightly lower numbers were obtained from a similar approach but for the year 2015 to 2016.

wage earners, self-employed have higher interest expenses. In purely private economic terms, such a situation does not seem favorable for the self-employed, since in that case the wage earner has more liquid funds in hand than the self-employed, *ceteris paribus*. The only way to personally compensate for higher interest expenses is for self-employed to have income that is not accounted for. Hence, there is a plausible argument that self-employed under-report part of their business income. Another side of the story would be that self-employed are "forced" into their entrepreneurship because as wage earners they should have had lower interest expenses show the same level of disposable income and hence be able to raise their standard of living, *ceteris paribus*.

4.3.2 Incentives for higher or lower level of under-reporting

Given rational behavior of self-employed, do under-reporting follow economic incentives? In the case of a period of sick leave, Swedish social insurance office, Försäkringskassan, compensates individuals for the incapacity for work that arises. Individuals' level of compensation is roughly 80 percent of current income, up to a certain income ceiling. For the self-employed, the assessment is made on the basis of a forecasted future income. Andersson and Vejsiu (2019) study whether individuals' likelihood of being on sick leave affects their choice to become entrepreneurs. Now a similar line of argument is being used, but with the hypothesis that self-employed people who have a high probability of being on sick leave have incentive to under-report at a less degree. The opposite is true for self-employed people with a low probability of being on sick leave. They have incentives to under-report their business income to a higher degree.

To apply this strategy, we need in a first step to estimate the probability of being on sick leave at the individual level. In a second step the self-employed are split into two sub-groups: one group with low level of probability and a second group with a high probability of being on sick leave. Self-employed with a low probability of being on sick leave are those individuals in the lowest quartile of the probability distribution to be on sick leave while those with a high probability are those in the highest quartile.

The calculated level of under-reporting in Table 2 show that under-reporting is in the line with the notion of rational economic behavior of self-employed. Hence,

individuals' who are self-employed people with a low probability of being on sick leave have a 75.1% under-reporting while the corresponding number is 49.6% for individuals with a high probability of being on sick leave.

Table 2: The degree of under-reporting by high and low probability of being on sick leave, percent.

Probability of being on sick leave	Self employed
Low	75.1
High	49.6

Source: Statistics Sweden and own calculations.

5. Discussion of under-reporting and its consequences in statistical accounting

In this study, it's shown that under-reporting is complex and depends on many aspects. But for the sake of simplicity, we say that under-reporting for 2018 is equal to 1.53 for self-employed. This implies that income from a self-employed of 1 SEK is equal to 1.53 SEK income for a wage earner.

In the next section we give a brief context of self-employed hybrids and their status of employment and how we handle tax evasion in the name of statistics.

5.1. Adjusting for under-reporting and post-tax income

If one only use the degree of under-reporting for a single year, there is a likelihood to obtain a rather volatile numerical value of under-reporting that may create a rather unstable decision process for status of employment of the hybrids. Therefore, when we use the algorithm for identifying self-employed individuals', it is better to feed the algorithm with a coefficient that is more of sluggish character. Therefore, we use a moving average method to calculate the coefficient.

The applicable coefficient for, e.g., 2019 will therefore use values for the years of 2015 to 2017. The lag structure depends on the use of administrative data of disposable income and interest expenditure. In the first step we calculate a product from of the ratio of under-reporting and F-tax ratio. Each years' products is calculated in the same procedure and these products constitute the input values for the moving

average.¹² The moving average for 2019 is hence the average of the values: 3.18; 3.15 and 3.08. This implies that the input coefficient of the algorithm is 3.14 for the year 2019. Hence, the status of employment for the hybrids is then decided by the following decision rule:

$$I_{self-employed} * \Omega \geq I_{employment} \rightarrow \text{status of employment: } self-employed$$

$$I_{self-employed} * \Omega < I_{employment} \rightarrow \text{status of employment: } employees$$

where I is the income either from the self-employment or employees and Ω (Ω is equal to 3.14 for 2019) is the calculated input coefficient reported in Table 4.

Table 4: Calculating the input coefficient of the algorithm

Year	Ratio		Product of (2) & (3)	Input coefficient algorithm (Ω)	mars year t
	Under-reporting	F-tax			
2015	1.53	2.08	3.18		2015
2016	1.56	2.02	3.15		2016
2017	1.54	2.00	3.08		2017
2018	1.53	1.96	3.00		2018
2019	1.51	1.89	2.85	3.14	2019
2020	1.45	2.01	2.91	3.08	2020
2021				2.98	2021
2022				2.92	2022

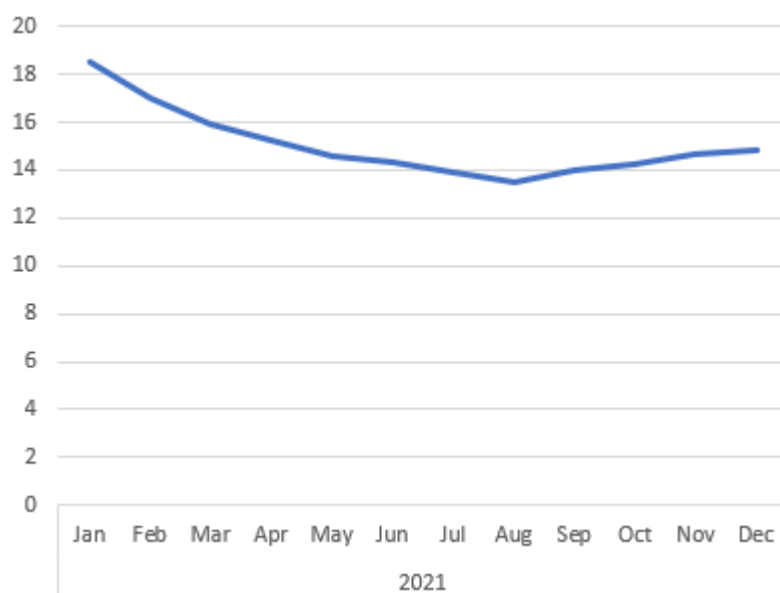
Source: Statistics Sweden and own calculations.

¹² The calculations are made for the respective annual totals of F-tax notifications for F-taxes that are between the 10th and 90th percentile of the F-tax distribution.

5.2. Around 15 percent of the hybrids will be classified as self-employed

At any given month under 2021 there were around 110 thousands hybrids on the Swedish labor market, with a slightly decreasing number over the year. Figure 4 depicts the share of hybrids that become classified as self-employed and around 15 percent is thus thought to be self-employed in the new statistical source Population by Labour market status. As also shown in Figure 4, there seems to be some kind of calendar or seasonal effect that may affect the share of hybrids that is classified as self-employed.

Figure 4: The share of hybrids that is classified as self-employed under 2021, percent.



Source: Statistics Sweden and own calculations.

6. References

Andersson F W and A Vejsiu (2019), "Vem lämnar en (o)trygg anställning för att bli egenföretagare eller aktiebolagsdelägare", *Ekonomisk debatt*, vol 47, pp 18–30.

Engström, P and J Hagen (2017), "Income underreporting among the self-employed: a permanent income approach", *European Economic Review*, vol 92, pp 92–109.

Engström, P and B Holmlund (2009), "Tax evasion and self-employment in a high-tax country: evidence from Sweden", *Applied Economics*, vol 41, pp 2419–2430.

Pissarides, C and G Weber (1989), “An expenditure-based estimate of Britain's black economy”, *Journal of Public Economics*, vol 39, pp 17–32.

Swedish Tax Agency (2020), “Black labor income: calculations based on the Tax Agency's audits”.

7. Appendix

Table A1: Regression results (OLS), self-employed 2018

	All	Level of education		
		Pre-upper secondary	Upper secondary	Post- secondary
Ln Disp. Income	3.273*** (480.426)	3.732*** (199.098)	3.348*** (387.526)	2.607*** (203.092)
Self-employed	1.390*** (121.272)	1.521*** (55.346)	1.422*** (100.547)	1.165*** (41.286)
Woman	0.233*** (56.322)	0.141*** (10.452)	0.220*** (42.697)	0.275*** (37.319)
Age	0.450*** (442.256)	0.258*** (93.761)	0.487*** (392.901)	0.430*** (175.917)
Age^2	-0.00492*** (-424.919)	-0.00276*** (-88.924)	-0.00535*** (-374.920)	-0.00470*** (-175.182)
Post- secondary	-0.539*** (-71.624)			
Upper secondary	0.0191*** (3.006)			
Own house	2.203*** (491.654)	1.614*** (122.568)	1.858*** (339.675)	3.531*** (367.500)
Condominium	2.029*** (354.216)	1.020*** (54.525)	1.757*** (244.219)	3.252*** (304.101)
Industries	yes	yes	yes	yes
Constant	-44.75*** (-575.072)	-46.65*** (-225.055)	-46.05*** (-465.390)	-37.47*** (-237.151)
Number of obs.	3553693	389184	2250454	1005332
R ²	0.275	0.243	0.282	0.293

t statistics in parentheses

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

*Table A2: Regression results (OLS), owners of closely limited liability companies
2018*

	All	Level of education		
		Pre-upper secondary	Upper secondary	Post- secondary
Ln Disp. Income	2.151*** (286.551)	2.856*** (96.481)	2.319*** (240.050)	1.596*** (118.706)
Owner	0.0522*** (5.433)	0.154*** (4.813)	0.0273** (2.334)	0.180*** (9.254)
Woman	0.0812*** (19.627)	0.0402** (2.484)	0.0392*** (7.379)	0.135*** (18.848)
Age	0.390*** (311.518)	0.285*** (74.951)	0.411*** (273.399)	0.371*** (133.903)
Age^2	-0.00435*** (-316.608)	-0.00310*** (-75.685)	-0.00460*** (-275.434)	-0.00413*** (-138.930)
Post- secondary	-0.556*** (-67.524)			
Upper secondary	-0.00341 (-0.463)			
Own house	2.470*** (479.746)	1.903*** (122.805)	2.073*** (331.377)	3.513*** (309.972)
Condominium	2.293*** (373.891)	1.308*** (61.838)	1.942*** (252.756)	3.324*** (269.522)
Industries	yes	yes	yes	yes
Constant	-29.19*** (-310.944)	-36.08*** (-97.525)	-31.39*** (-259.926)	-22.99*** (-131.565)
Number of obs.	3107457	295581	1914247	867993
R ²	0.190	0.136	0.182	0.228

t statistics in parentheses

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