

IZA DP No. 1971

Employees Who Become Self-Employed: Do Labour Income and Wages Have an Impact?

Pernilla Andersson
Eskil Wadensjö

February 2006

Employees Who Become Self-Employed: Do Labour Income and Wages Have an Impact?

Pernilla Andersson

SOFI, Stockholm University

Eskil Wadensjö

*SOFI, Stockholm University
and IZA Bonn*

Discussion Paper No. 1971
February 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Employees Who Become Self-Employed: Do Labour Income and Wages Have an Impact?*

This paper analyzes the self-employment decision among Swedish-born male employees. The main objective of the paper is to investigate the impact of the relation between the actual and the predicted income on the probability to become self-employed. The predicted income is calculated from a standard income regression with controls for age, education, family status, family background and place of residence. By construction of a ratio between the actual and the predicted income we identify three groups of employees: (1) employees who have an actual income lower than the predicted income (underpaid), (2) employees with an actual income close to the predicted one, and (3) employees with an actual income higher than the predicted one (overpaid). The first question is if individuals who are “overpaid” or “underpaid” are more likely to become self-employed than those who are paid as we can expect. Our main finding is that employees who receive an income that differs from the one predicted by the income regression are more likely to become self-employed. We also analyse the effect of the ratio on four different measures of success as self-employed: income from self-employment, number of employees, turnover of the firm, and the probability to have a firm registered as a limited liability company. The general conclusion is that those who performed well as employees are also more successful as self-employed.

JEL Classification: J23, J24, J30, J62

Keywords: self-employment, occupational choice, occupational mobility, labour income, wages

Corresponding author:

Eskil Wadensjö
Swedish Institute for Social Research
Stockholm University
SE 106 91 Stockholm
Sweden
Email: Eskil.Wadensjo@sofi.su.se

* We would like to thank seminar participants at the Swedish Institute for Social Research for helpful comments on an earlier version. Financial support from IFAU is gratefully acknowledged.

1. Introduction

The literature on self-employment and entrepreneurship has grown substantially during the last decade. One of the most studied issues is why some individuals and not others become self-employed. Individual characteristics such as age, education, marital status, gender, place of residence, ethnical background, and unemployment experience have been put forward as explanatory variables by economists.¹ In studies by non-economists, psychological factors such as need for achievement, internal locus of control, above average risk-taking propensity and a tolerance of ambiguity have been analysed.²

Several studies have analysed the impact of individual unemployment experience on the probability of becoming self-employed. The hypothesis behind these studies is that self-employment can be a way of escaping unemployment. In a way, individuals are pushed into self-employment. Martinez-Granado (2002), for example, estimates labour market transitions between employment and unemployment to self-employment using the British Household Panel Survey, and find that the unemployed are more likely than the employed to become self-employed but that the probability declines with the duration of unemployment.

In this paper we do not analyse the decision of the unemployed to become self-employed, but focus on the factors that affect the willingness of wage-earners, i.e. those who already have a job, to become self-employed. We restrict the analysis in this paper to only cover employed males aged between 20 and 60 years who were born in Sweden. We are particularly interested in the impact of the relative economic reward on the probability to become self-employed, i.e. the impact of the income or the wage an employee receives compared to the income he could be expected to receive as a wage-earner given his characteristics. Expected income is defined as the predicted income calculated from a standard Mincer equation. We calculate the ratio

¹ See Blanchflower (2004) and Parker (2004) for surveys of earlier research.

² Amit, Glosten and Muller (1993) identify these four factors as interesting and relevant regarding the decision to become self-employed. See also Parker (2004) for a summary of the research on psychological factors.

between the actual income and the predicted income as a measure of success as a wage-earner given characteristics. Some employees have an income that is lower than what is predicted by their observed characteristics (age, education, marital status, place of residence), others receive an income that is higher than the income predicted by the Mincer equation. We call the first group “underpaid” and the second group “overpaid”. The third group comprises those who receive an income which is approximately the same as the expected income.

The questions we ask in this paper are: (1) Which of these three groups are more likely to become self-employed, those who are paid less than expected, those who are paid more than expected, or those who are paid as we can expect them to be? (2) Which of these groups, once they have become self-employed, will perform best in terms of (i) income from self-employment, (ii) number of employees, (iii) turnover in the firm, and (iv) probability to have the firm registered as a limited liability company (LLC)?

For the first question we pose two hypotheses for what results to expect. The first is that individuals with a low quotient (low actual income compared to the predicted one) may turn to self-employment as a reaction to the disappointment of the low income as wage earner. In this way they may be pushed into self-employment. The second is that a high value of the quotient may be a result of high capability on the part of the individual. If the employee is not fully rewarded due to redistributive elements in the wage setting (a solidaristic wage policy or an employer policy of smaller wage differences than productivity differences to avoid conflicts) people may leave in an attempt to get a higher reward for their skills. They may in this way be pulled into self-employment.

Regarding the second question, a lower income than what is predicted by the Mincer equation can be a result of that the individual is not “suited” for being an employee. He might be longing for high levels of self-determination and might not feel very comfortable with having a boss. These individuals might be good entrepreneurs and it is possible that they perform

better than other groups when having a firm of their own. On the other hand, a low relative income may be the result of a low capability compared to his characteristics both as wage-earner and as self-employed leading to a poor result as self-employed.

Individuals who are paid much better than we can expect from their observed characteristics may have very high competency and it is likely that they are high performers not only as wage earners but also as self-employed.

Our study contributes to the literature on self-employment in several ways. The data used and introduced below, is unique in that it contains information on all employed individuals who lived in Sweden in 1998 and on all individuals who left their employment as a wage-earner to become self-employed in the period 1999-2002. This data enables us to use two measures of earnings. For the whole population we have information on the annual income from labour and for a part of the population we also have information on the hourly wage. Thus far, nobody has been interested in the impact of the economic reward an employee receives put in relation to what can be expected as a wage-earner for the decision to become self-employed. This paper asks a new question.

What implications can the results of this study have outside academia? The most interesting part for non-academics is probably the result regarding the performance of the different groups, or “types”, of wage-earners that become self-employed. Policy makers and representatives from industry argue that new small firms and their entrepreneurs will be the engine behind economic growth in the future. If the self-employed who were very successful as wage-earners out perform the other two groups, this would mean that many “high quality” entrepreneurs are to be found inside firms. An implication of this is that skilled employees should be encouraged to start firms.

In this paper we use unique register data for Sweden covering the period 1998 to 2002. For 1998 we have information on all individuals in Sweden between 20 and 60 years of age who were employed and estimate labour income and wage regressions based on those observations. Some of these individuals became self-employed in the following years and for them we have information for all years between 1998 and 2002. This design of the data enables us to proceed with the analysis in three steps: (1) by using the data for 1998 we estimate a Mincer type income equation, calculate the predicted income and construct the ratio between the actual and the predicted income, (2) we estimate a probit regression for the probability to be self-employed in 2002 and include controls for the ratio, and (3) we investigate the effects of the ratio on the individual's performance as self-employed among those who were self-employed in 2002 and were employees in 1998.

We have information for the whole population on annual labour incomes in 1998 and for a part of the population we also have information on hourly wages. Having access to both these measures we can consider both the impact of receiving an annual income that is higher (lower) than what is predicted by the income equation and the impact of receiving an hourly wage that is higher (lower) than what is predicted by the wage equation. A low annual income is probably to a large extent explained by that the individual has not worked full-time or full-year. Some of them may involuntarily have worked part-time or part-year. It is reasonable to believe that some of these individuals might turn to self-employment to be able to work more. A higher (lower) hourly wage than what is predicted by the Mincer equation can only be interpreted in terms of a diverging reward per unit of time.

The rest of this paper is organized as follows; in section two we present earlier research on the topic of why people become self-employed and discuss and motivate our key variable in more detail. The data and estimation strategy are described in section three. Section four reports on the results of the first step of the analysis and presents a description of the individuals in the

different ratio-groups. In section five the second and third steps of the analysis are presented, and finally in section six our findings are summarized and we draw some conclusions.

2. Why do people become self-employed?

2.1 Factors behind the decision to become self-employed³

The starting point of the analysis is the question of why wage earners choose to become self-employed. When an individual is comparing the alternative of becoming self-employed with other alternatives, the expected economic outcomes of the alternatives are important. Not only the economic outcome, but also other factors might influence the choice, for example working conditions (working hours, flexibility) and the degree of independence.

It is generally not possible to get information about the alternatives people consider when they are deciding between being self-employed and being a wage-earner. Empirical research on the selection into self-employment relies instead mainly on variables which in different ways may be related to the prospects of self-employment in comparison with the prospects of other alternatives. We will present some variables below that are often included in empirical studies and the expected effects of them on the probability of becoming self-employed.

Age is often included among the variables when attempting to explain why some people become self-employed and others not. Higher age in most cases means more experience and more knowledge of how to handle one's own business. This may explain a tendency towards an increasing propensity for self-employment with age. On the other hand, older people on average may be more risk averse and to become self-employed is often a high-risk activity.

³ For recent surveys, see Blanchflower (2004) and Parker (2004).

This may explain a declining propensity to become self-employed with age. This means that the expected sign of age (or labour market experience) is undetermined.⁴

Gender is also regularly included among the explanatory variables. Also here the expected sign is undetermined. Given human capital characteristics, women have lower wages than men which may make self-employment more attractive. On the other hand, the fact that women traditionally do more of the work in the household may make it more difficult to work as self-employed since self-employment often means long working-hours.⁵ The effect of gender on the probability to become self-employed is not considered in this paper.

Variables indicating family status are often included. People who are married have a second bread-winner in the household to a higher extent than those who are not married. It makes it easier to handle a period of low income at the start of self-employment – an argument for that individuals who are married more often become self-employed. On the other hand, to be married may also imply a less risky behaviour. More people are dependent on the individual's income which may be an argument against going from being a wage-earner to self-employment. It means that also here the expected sign of the coefficient of the variable is undetermined.

Education may influence the propensity to become self-employed. However, it is not necessarily so that higher (or lower) education leads to a greater likelihood of becoming self-employed. Some forms of education prepare for self-employment to a higher extent than other forms. Educations that prepare for self-employment can be found on different levels, for example both on the secondary school level (vocational study lines) and on the university level.

⁴ Lin, Picot and Yates (1999), for example, get as a result a declining propensity to enter self-employment with age in a study of Canada.

⁵ Georgellis and Wall (1998) study the propensity to enter self-employment for men and women in Germany and find that some factors have different effects for men and women.

The propensity to establish self-employment may be relatively high among some (but not all) groups of immigrants and groups of other ethnicity than the majority of the population. Self-employment may be a way of escaping employer discrimination, and some (but not all) groups of immigrants may have had a stronger tradition of self-employment in their home country than is prevalent in the host country. On the other hand, due to discrimination, immigrants and ethnic minorities may have difficulties in financing self-employment and have less support of a network than those born in the country. Also here the expected sign of the coefficient is indeterminate.⁶

Independent of ethnicity, some families have a tradition of self-employment. Children receive knowledge about the starting and running of a business from self-employed parents and in some cases inherit businesses from their parents.⁷ Variables representing self-employment of the parents are often included with expected positive signs.

Variables representing wealth are often included in studies of recruitment to self-employment. Personal wealth may facilitate the starting of self-employment and also lead to that it becomes easier to handle variations in activity.⁸

Another variable often included is individual unemployment. Unemployment may make the starting of self-employment a more attractive alternative. It does not mean that a higher unemployment rate in the economy leads to that more people turn to self-employment.⁹ The prospects for self-employment may decline in a down-turn of the economy so that fewer wage-earners turn to self-employment which may outweigh that more unemployed turn to self-employment. In this paper we are studying those who are employed.

⁶ See for example Andersson and Wadensjö (2004) for a comparison of the propensity to enter self-employment among natives and immigrants in Denmark and Sweden.

⁷ See for example Hout and Rosen (2000).

⁸ See for example Taylor (2004).

⁹ Audrechts et al. (2005) call a positive effect of unemployment on the propensity to start self-employment “the refugee effect”; the self-employed are fleeing from unemployment. On the other hand, the establishment of self-employment may lead to a decline of unemployment. They call it the “entrepreneurial effect”.

Another factor that is assumed to affect the decision to become self-employed is the predicted income differential between self-employment and wage-employment.¹⁰ The idea behind this variable is that an individual chooses to become self-employed if the expected utility is higher than the utility of being a wage-earner. From theory we expect that the higher the predicted income from self-employment is in relation to that from being a wage-earner, the higher is the probability to leave employment with a firm and become self-employed. Parker (2004) has summarised the findings of these studies and concludes that the empirical results are mixed regarding the importance of relative earnings. In a study of immigrants in Sweden, Hammarstedt (2005) finds that an increase in the differential between self-employment and wage-employment earnings has a positive impact on the self-employment rate among immigrants.

2.1 The relative wage level and the propensity for becoming self-employed

In our study we concentrate on the factors influencing wage-earners to become self-employed. Why do people leave a job where they receive a wage to become self-employed? One factor may be the compensation they receive as a wage-earner (the actual wage) relative to the wage income to be expected given the human capital characteristics of the individual (the predicted wage). Which factors can then cause an individual's income to diverge, both up and down, from the predicted income given a set of observed characteristics?

It is possible that there are some unobserved characteristics which are correlated with productivity and the wage rate and annual labour income. This may lead to a difference between the predicted and the actual income. A lower or higher wage than the predicted one may be fair given the characteristics, observed and unobserved. But such a difference may

¹⁰ Studies in which this has been analysed are for example Hammarstedt, 2005, Johansson, 2002, Constant and Zimmerman, 2005, Rees and Shah, 1986, Taylor, 1996, Dolton and Makepeace, 1990, and Clark and Drinkwater, 2000.

also be explained by other factors. One factor may be that the wage setting is influenced both by the market and by institutions such as collective bargaining and the employer's wage setting policy.

Another factor could be a mismatch regarding education. During the last ten years unemployment among university educated has increased in Sweden and many highly educated individuals, mostly young, are forced to take an employment for which they are overqualified. These individuals are likely to be those who have an actual income that is lower than the predicted one. This can be labelled as overeducation. Böhlmark (2003) and Tåhlin and Korpi (2005) have studied over and undereducation using the Swedish Level of Living Survey. In both papers it is shown that the overeducated are less satisfied with their job than those who have an occupation corresponding to their education. This might be one reason for why people in this group are likely to become self-employed. On the other hand, during the IT-boom in the late 1990s some very young people with no formal education were employed in this sector alongside highly educated engineers. These individuals are in a formal sense under-qualified for the jobs they performed and are likely to be those who have an actual income that is higher than what is predicted by the income equation.

A low relative reward may, irrespective of the reason, but especially if the difference is considered unfair, make people dissatisfied with their job. People compare themselves to others and a low reward may induce people to look for other opportunities, of which one could be self-employment. So what we expect to find is that people with low compensation compared to what could be expected in light of their characteristics are more likely to become self-employed than the average wage earner is. This could be seen as that it is the "most disappointed" who leave for self-employment.

What about people with a relatively high compensation, compensation higher than could be expected from their characteristics? We may expect that they should be satisfied with their job

– they are paid more for the characteristics we observe than what people are paid in general – and therefore leave their employment for becoming self-employed to a lower extent than those with an average compensation in relation to their characteristics. On the other hand, the high compensation may also be seen as that they have a higher capacity than the average. And high capacity may not be fully rewarded in the labour market. A large wage spread, even if motivated by differences in capability and productivity, may be seen as unfair by the employees and lead to a lower output by making teamwork less efficient. Employers may therefore avoid rewarding differences in capability and productivity to full extent. Wage agreements reached by collective agreements are another factor at least in some parts of the labour market. Wage agreements are in many cases formulated in ways that counteract large wage differentials. That high productivity is not fully compensated may lead to that people try to find other ways of getting compensated for their resources. To become self-employed may be one way of getting such compensation. One hypothesis is that those with a high actual wage compared to the predicted wage, “the best and the brightest”, leave employment to a higher extent than those with average compensation.

Our starting point for the empirical analysis is that the actual relative to the predicted wage of wage earners may be an important factor in the individual’s decision to become self-employed. Our hypothesis is that both those with a low reward – “the most disappointed” – and those with a high reward – “the best and the brightest” – are overrepresented among those who turn to self-employment.

An implication of this reasoning is that the ratio will not have a linear effect on the probability to become self-employed. If we only assumed a linear relationship this would only enable us to test whether it is the most disappointed *or* the best and the brightest that become self-employed. Therefore we specify the probability models so that we can capture the possible non-linear effect of the ratio on the probability to become self-employed.

3. Data and Method

In this study we use register data for Sweden for the years 1998 and 2002. For 1998 we have information on all individuals who had an income from employment. In this paper we restrict the analysis to men between 20 and 60 years of age who were born in Sweden and who were employed in 1998.

Some individuals have done several things during a single year and there is no obvious way to define those who have been employees during the whole year due to lack of information on hours worked. In the database individuals are defined as employed if they performed paid work corresponding to at least one hour during one week in November. Those defined as employed can, for example, have been unemployed, students or on parental leave during some part of the year. To eliminate the main part of those working only part of the year we have put several restrictions on the data for 1998. Individuals who are included in the sample fulfil the following restrictions;

- they were employed in November 1998
- they were not registered as students during the year
- they were not registered as unemployed and they had no days with unemployment benefits
- they had not been in military service
- they had not received a disability pension
- they had not received social security payments equal to or higher than their labour income
- they were not self-employed in 1998 and they had not received any income from self-employment¹¹
- they had an annual labour income in 1998 of at least 20 000 SEK

¹¹ Some individuals who are not coded as self-employed have still received some income from self-employment. This can happen for two reasons: if an individual is employed in November and receives his largest income in this month from an employment with a firm but has a smaller income from self-employment, or if an individual has a job as wage-earner in November but has received an income from self-employment earlier during the year.

- they were not employed in agriculture or fishing
- they were not younger than 20 and not older than 60 years of age in 1998
- they were born in Sweden

The total population consists of about 1.2 million men who were born in Sweden and who were wage earners in 1998. It is for this group that we estimate the income regression in 1998. Of them, 2.8 percent or 34 410 individuals were self-employed in 2002. To be defined as self-employed it is not necessary that an individual has completely left employment with a firm. Becoming self-employed does not have to be a discrete jump from only being an employee to only being self-employed. In our data, an individual is defined as self-employed if the income received from self-employment in November is higher than the income received as a wage earner in November. It can be a smooth transition out of employment with an employer to self-employment. We estimate the probability to become self-employed, not the probability to cease being a wage earner and become self-employed.

We perform the first step of the analysis using two different dependent variables; annual income and hourly wage. Information on annual income exists for the whole population while we only have information on hourly wage for a part of the population. There are about 635 000 individuals in this group. One problem with using the wage information is that it is not random for which individuals this information exists. The wage data is available for all individuals who were employed in the public sector by the 1st of September in 1998. In the private sector this information is only gathered by Statistics Sweden for a sample of firms and organisations. It is estimated that about 50 per cent of the employees in the private sector are included in this register.¹²

In the income and the wage regressions we control for age, marital status, if one has children younger than six years, family background (second generation immigrants), education and

¹² For more detailed information of this statistic see “Strukturlönestatistiken”, Statistics Sweden.

place of residence. The full results of the income and wage regressions that are estimated using OLS, are presented in *Table 2* and discussed in section 4. It is from these two models that we calculate the predicted incomes and wages and construct two different ratios. If the actual annual income (hourly wage) is greater than the predicted income the ratio is higher than one, and if the actual annual income (hourly wage) is lower than the predicted income, the ratio is lower than one.

In the second step, we estimate a probit model for the probability to be self-employed in 2002 and in this model we include the ratio among the explanatory variables. We estimate four different specifications of the model. The first only includes controls for the ratio which is included in two alternative ways. First we divide the population in three groups; (i) individuals with a ratio lower than one, (ii) those with a ratio close to one, and (iii) those with a ratio higher than one.¹³ Using this categorisation we include two dummy variables and set the middle group as reference category. Second, we divide the population according to in which decile of the ratio distribution they are in, hence we have ten different intervals that we control for by including nine dummy variables. In the second specification we control for the actual income (wage) that an individual received in 1998. The reason is that we expect a positive correlation between actual income and the value of the ratio, i.e. that individuals with a high actual income are more likely to be overpaid while individuals with a low actual income are more likely to be underpaid. The third and fourth model corresponds to the first and second but now we add controls for age, marital status, if one has children younger than 6 years, family background (second generation immigrants), education and place of residence.

¹³ To divide the population into three groups we have identified the ratio at the 33rd and the 66th percentile of the ratio distribution. Individuals with a ratio equal to or lower than the ratio at the 33rd percentile are classified as those with a ratio lower than one, individuals with a ratio between the 33rd and the 66th percentile are classified as those with a ratio approximately equal to one and those who have a value of the ratio higher than or equal to the ratio at the 66th percentile are classified as those with a ratio higher than one. Using this method to divide the population into ratio groups gives us three approximately equally large groups.

In the third step of the analysis we look at the effect of the ratio on the outcome for self-employed in 2002. We have four different outcome measures: income from self-employment, number of employees, the turnover in the firm, and whether the firm is a limited liability company or not. Here we control for the same characteristics as before, now measured in 2002, but also for industry. In these models we also control for different “types” of self-employed. One group of self-employed are those who do not start a firm by themselves but take over an existing firm. We consider a self-employed to be of this type if the firm he has in 2002 was started in 1998 or earlier. We also control for whether the firm started in 1999, 2000, 2001 or in 2002. Other groups of self-employed are those who are self-employed at the same firm as where they were employees in 1998.

It is interesting to see if individuals with a low income or wage as a wage earner compared to what they are expected to receive given their characteristics, are also less successful as self-employed. This can be interpreted as that those who are “good” employees in the sense that they receive a higher income than is expected from their characteristics will also be “good” entrepreneurs. On the other hand it is also possible that those with low incomes in relation to what is expected are more likely to be “good” entrepreneurs. In these models we also control for the actual income (wage) that an individual received in 1998.

4. Income Regression in 1998 and the Predicted Income

In this section we present the results of the first step of the analysis, i.e. the Mincer equations for employees in 1998. In *Table 1* we show the mean values of the included covariates and the annual income in 1998 for all employees, for those who are not self-employed in 2002, and for those who are self-employed in 2002. All covariates are measured in 1998. The covariates are age, education (seven categories), marital status, and place of residence. We present the corresponding mean values for the sample of the population for which we have wage information.

There are some noteworthy differences between those who became self-employed and those who did not. Among those who became self-employed, a somewhat lower share had the lowest educational level (primary school less than nine years) and a higher share had higher education. Almost 50 per cent of those who became self-employed lived in one of the three big city areas in 1998, while only 33 per cent of the others lived in these areas.

It is interesting to see if there is a difference in average income in 1998 between the two groups. We find that the average annual income is similar for the two groups; 270 858 SEK¹⁴ and 279 090 SEK. However, there is a large difference in the spread of the incomes. For those who did not become self-employed the standard deviation was 156 210 SEK and for those who became self-employed the standard deviation is much higher, 193 321 SEK. This indicates that those who became self-employed are more heterogeneous than the employees who did not become self-employed.

The corresponding analysis for individuals for whom information on the hourly wage rate exists shows that those who become self-employed on average have higher hourly wages than those who do not. In this population those who become self-employed have higher education than those who did not. This is most likely the result of that employees in the public sector are overrepresented in the wage data and individuals in this group who become self-employed are not entirely representative for all those becoming self-employed. It is clear that the analysis with both dependent variables has its problems and for this reason it is good to keep these difficulties in mind and compare the results throughout the paper.

From the estimated Mincer equations presented in *Table 2* we calculate the predicted annual income/hourly wage. We have not included controls for industry in the regression. If an individual has a lower income than we expect him to have because he is in the “wrong” industry, he is perhaps more likely to become self-employed. We do not want to eliminate this

¹⁴ January 31, 2006, 1 euro = 9.1 SEK

effect by controlling for industry. Instead we expect this effect to appear in the relation between the actual and the predicted incomes.

In *Table 3* we describe the population according to ratio group where the ratio is based on the Mincer equation using annual income.¹⁵ There are four main things to notice. First, the share that becomes self-employed differs between the groups. Slightly more than 4 per cent of those who are “underpaid” become self-employed compared to 2.2 per cent of those who are paid as expected and 2.9 per cent among those who are “overpaid”. Second, in the underpaid group almost 19 percent have three years or more of higher education compared to 16 per cent in the overpaid group. In the middle group, only 10 per cent have this education. This can be interpreted as that individuals with higher education tend to be either under or overpaid to a higher extent than those who are rewarded as expected. Third, there is a small difference in the age distribution between the three groups. A higher share of those who are overpaid are between 20 and 25 years of age while a higher share in the group of underpaid are older than 46 years. Fourth, there is a difference in average annual income between the groups. Among those who are underpaid we find individuals with fairly high incomes even if still lower than what we expect. We find large differences between the groups. The average incomes of the underpaid group is 192 277 SEK compared to 372 442 SEK for the overpaid group. For a more detailed description see *Table 4*.

The second way to control for the effect of the ratio between the actual and the predicted income is by dividing the population into ten groups according to which decile of the ratio distribution they belong. D1, in *Table 3*, consists of the 10 per cent of the population that has the lowest value of the ratio and D10 consists of the 10 per cent with the highest values of the ratio. Individuals in the same decile of the ratio distribution, the 5th and the 7th deciles, are

¹⁵ A corresponding table can be constructed using the results from the Mincer equation where hourly wage is used as the dependent variable. Here we only present the mean values for different groups based on the results from the annual income regression.

split up and belong to both the under and the middle group, and the middle and the over group, respectively.

The ratio is constructed by dividing the log of the actual income with the predicted income from the income regression with log income as the dependent variable. However, to give a detailed description of the relation between the actual and the predicted income we also predict the incomes using the annual income in absolute values.¹⁶ In *Table 4* we make this description for the three ratio categories – under, middle and over – and for the whole population and for those who become self-employed and those who do not, separately. If we start by looking at the average difference between the actual income and the predicted income for the three first groups, one gets an idea of how the table should be read. For the first group which is labelled “under”, the mean difference is -69 399 SEK. This means that on average individuals in this group have income that is about 69 000 SEK lower than expected. In the “middle” group this difference is smaller, 771 SEK.¹⁷ In the group that is labelled “over” the individuals earn on average 118 252 SEK more than what is predicted by the Mincer equation.

When we divide the population into individuals who become self-employed and those who do not, we see that in the group that is underpaid those who became self-employed were more underpaid on average than those who did not. They earned 87 820 SEK less on average than predicted while those who do not become self-employed earn 68 794 SEK less on average than predicted. For individuals who are overpaid we find that those who become self-employed are more overpaid on average than those who do not. The first group had an excess income of about 151 262 SEK and the second had an excess income of 117 263 SEK.

¹⁶ As in *Table 3*, we only show the results for annual income for simplicity’s sake.

¹⁷ In the earlier discussion and motivation we argued that this group consists of individuals who are paid as expected, hence we should expect not to find any difference between the actual and the predicted income. This is not the case but the difference has to be considered to be very small considering that we only look at employees with an annual income higher than 20 000 SEK.

Having this description in mind we now continue the analysis with the second step; modelling the effect of the relation between the actual and the predicted income/wage on the probability to become self-employed.

5. The Probability of Becoming Self-Employed

In *Table 5a* and *5b* we present the results for the probit regressions where we estimate the marginal effects on the probability to be self-employed in year 2002 for belonging to different ratio groups. In *Table 5a* we first have the results when the ratio is based on annual income. In the comparison of the models we focus on what happens when we include controls for actual income. We argued earlier that there is a fairly high correlation between the ratio and the actual income but we are especially interested in the effect of the ratio. When not controlling for actual income, part of the effect of the ratio is due to the effect of the actual income. In the simplest model (*Model I*) we see that individuals who are underpaid have a 1 percentage unit higher probability to become self-employed compared to those who are paid as expected. The probability to become self-employed is 0.7 percentage units higher for individuals who are overpaid. When including the actual income the effect for those who are underpaid decreases and the effect for overpaid increases. The effects are significantly different from each other as is indicated by that the figures are shown in **bold**. The effect of annual income is negative and significantly different from zero meaning that the higher the annual income the less likely an individual is to leave employment with a firm and become self-employed. In the full models we see the same pattern. A likelihood ratio test tells us that the *Model II* is significantly different from *Model I*, and *Model IV* is significantly different from *Model III*. So the conclusion based on the estimates in *Model IV* is that the probability to be self-employed in 2002 is 0.25 percentage units higher for employees with a lower annual income than expected and the probability is 1.2 percentage units higher for those who received a higher income than expected.

In the four corresponding models where we include dummy variables for decile of the ratio distribution we find the same pattern. When controlling for actual income we also here see that the effect at the lower part of the ratio distribution decreases while the effect increases at the upper part of the distribution.

In *Table 5b* the result when the ratio is based on hourly wage is presented. Here the results are slightly different. Comparing *Model III* and *Model IV* we find that the inclusion of actual hourly wage significantly changes the estimated effects but here the change goes in the opposite direction compared to what we found for the inclusion of actual annual income; the effect for individuals who are underpaid increases and goes from negative to positive while the effect for individuals who are overpaid decreases. The correlation between hourly wage and the probability to become self-employed is positive which is also contrary to what we found using annual income. One explanation for this result is that the group for which we have wage information is not representative for the whole population. The wage information exists for all employees in the public sector but only for a sample of the employees in the private sector.¹⁸

From *Model IV* where we control for decile group we can derive the predicted probabilities to become self-employed and for each decile calculate the average predicted probability. These probabilities are plotted in *Figure 1*. Here we see a difference in the relative magnitude of the effects between the top and the bottom. Based on the ratio on the annual income, individuals in the bottom have a higher probability to become self-employed than individuals at the top, while the pattern is the opposite when the ratio is based on the hourly wage. This pattern is also investigated in the regressions by comparing the estimates for those under and overpaid and at the bottom and at the top of the ratio distribution. A t-test shows that the coefficient for the 1st decile is significantly different from the estimate at the 9th decile.

¹⁸ When the regression with annual income is estimated for the population for which information on wages exist, we find that annual income has a significant positive impact on the probability to become self-employed.

The conclusion of this second step of the analysis is that both those who are paid less than expected and those who are paid more than expected have a higher probability to leave employment and become self-employed. Controlling for actual annual income does not change this result but the estimated effects decrease for the underpaid and increase for the overpaid.

6. The Performance of Self-Employed in 2002

The main conclusion from the results presented in section five is that those who are underpaid as well as those who are overpaid compared to their characteristics are more likely to become self-employed compared to those who are paid as expected. In this section we continue the analysis by looking at the impact of the ratio on the performance of the self-employed. We measure their achievement, or success, as self-employed using four outcomes: incomes from self-employment, number of employees, turnover of the firm, and the probability that the firm is registered as a limited liability company. These variables are described in *Table 6* for the whole population and for the three different groups separately. For each outcome, the same types of models as before are estimated. The covariates are now measured in 2002. See *Tables 7-10*.

The main result of the analysis is that those who are overpaid as employees perform better compared to those who are paid as expected. This is the case both when the ratio is based on the annual income and on the hourly wage rate. One explanation for this is that these individuals have a high capacity and perform well not only as wage earners but also as self-employed. We also find that those who are underpaid as employees perform worse in relation to those who are paid as expected also when they are self-employed. Hence, the basic results are that we get a negative effect of being underpaid and a positive effect of being overpaid on the different outcomes.

In the outcome regressions we also estimate the models with and without controlling for the actual income/hourly wage in 1998. We basically see the same pattern in all the models; the negative impact of being underpaid as an employee decreases when controlling for actual income/wage and the positive effect of being overpaid decreases when controlling for actual income/wage. In other words, this means that the difference in performance between the groups decreases when considering the correlation between the outcome and the actual income/wage.

In the following we look at some of the results in more detail. All references are made to the results in *Model IV* which is the model where we control for the full set of covariates including industry, starting year of the firm, and actual income/wage.¹⁹

The first outcome or measure of success we look at is the income from self-employment. These results are presented in *Table 7*. Individuals who are underpaid as employees have on average around 6 per cent lower annual incomes from self-employment than those who are paid as expected. Individuals who belong to the group who have been overpaid as employees in 1998 earn on average 5 per cent higher incomes. We find a highly significant positive impact on the incomes from self-employment. In the estimations where the ratio is based on hourly wage we only find a significant negative effect of being underpaid compared to being paid as expected. We still find a large positive effect of the actual wage the self-employed received as wage-earner in 1998.

The second outcome is number of employees, a variable that has been divided into three categories since the distribution is skewed. Most self-employed do not have any employees. We estimate a multinomial logit model to analyse the effect of being under/over paid as an employee on the number of employees those who become self-employed have in their firms.

¹⁹ In *Table 7-10* we only present the result for the regressions when controlling for the three ratio groups. The results when controlling for decile of the ratio distribution is available from the authors upon request.

In *Table 8a* and *8b* relative risk ratios are presented. A value larger than one means that the risk is higher for individuals in this group to belong to the category in question compared to belonging in the reference category than for individuals who are not included in the group. A relative risk ratio lower than one means that the risk is lower for individuals in this group. The reference category is self-employed with no employees. Here we get the same pattern both when basing the ratio on annual income and basing it on hourly wage. Individuals who were underpaid as employees are less likely to have between 1 and 10 employees, and also less likely to have more than 10 employees. The relative risk for individuals who are overpaid as employees to have more than 10 employees in their firms is 1.4 times higher than the risk for those who are paid as expected.

In the analysis of the turnover in the firm we use the log of this variable.²⁰ See *Table 9*. Those with negative or zero turnover in a year are excluded so this estimation is based on fewer observations than the other regressions. Individuals with a lower annual income than the predicted one have on average about 15 per cent lower turnover in 2002 compared to self-employed who as employees were paid as we can expect. The turnover in the firms run by those who were overpaid as employees is on average 18 per cent higher than the reference group.

When we use the probability to have the firm registered as a limited liability company we find a different pattern in *Model I* and *Model II* than what we found for the other groups. Individuals who were underpaid are actually more likely to have their firm registered as a limited liability company compared to the reference category. But when controlling for the full set of covariates we find a negative effect. See *Table 10*. In *Model IV* we see that the probability to have a LLC is approximately 3.5 percentage units lower for those who were underpaid and 5 percentage units higher for overpaid, compared to those who were paid as

²⁰ Of the self-employed, 770 (2.4 per cent) have a negative or zero turnover during 2002. These individuals are excluded from the regressions.

expected. When the analysis is based on wages we find the same differences but the precision of the estimates have decreased. Both annual income and hourly wage have a positive impact on the probability to start an LLC.

When analysing the impact of the relation between the actual and the predicted income on the four measures of success we find basically the same pattern for all outcomes. Those who are successful as employees in the sense that they are rewarded more than we expect them to be as wage earners, are also more successful as self-employed. If this finding is interpreted in terms of that these individuals are high achievers, then it is this group that policy makers would want to encourage to become self-employed.

7. Summary and Conclusion

Taken together, the results presented in this paper can be interpreted in the following way. Those who are underpaid perform worse than what we can expect from their characteristics. They may be expected to leave to a higher extent than those paid according to their characteristics due to a disappointment with what they get in reward and in that they be pushed into self-employment. We get as a result that this group is overrepresented among those becoming self-employed. Some may be pushed into self-employment even if they are not very skilled entrepreneurs. This may explain the result that they as self-employed have lower incomes, fewer employees, and lower turnover than those who become self-employed who are not underpaid as wage-earners.

Those who are overpaid perform very well as wage earners, better than what we can expect from their characteristics. This group is also overrepresented among those becoming self-employed. This group may be pulled into self-employment. The wages are rigid due to collective bargaining and internal wage policy leading to that some of those who are overpaid are underpaid compared to their productivity. For high achieving individuals, self-employment can be an alternative to perform even better economically. It also seems

reasonable that this group of individuals will also perform well as self-employed. The estimations also indicate that they have higher incomes, more employees, and high turnover as self-employed than the others who become self-employed.

So what can we learn from this analysis that can be useful in the discussion of self-employment and economic growth? One policy implication is that self-employment should be encouraged among high performing individuals on the labour market. These individuals also perform better as self-employed if performance is measured in terms of income, number of employees and turnover. Different ways to encourage this group to become self-employed are discussed by Delmar, Wiklund and Sjöberg (2003).

We conclude this paper by arguing that we have shown that an individual's reward as an employee contributes to explain the probability to become self-employed as well as the performance as self-employed. We have interpreted the relation between the predicted and the actual income as a proxy for the quality of the worker and also as proxy of the quality of the match of the worker and the job and by seeing it this way we have seen that there seems to be both a push and a pull effect into self-employment where those who are pulled in perform better than those who are pushed into this form of employment.

References

- Amit, R, Glosten, L, Muller, E (1993), "Challenges of theory development in entrepreneurship research", *Journal of Mathematical Sociology*, Vol. 30, 815-834.
- Andersson, Pernilla and Wadensjö, Eskil (2004), "Self-employment for immigrants in Denmark and Sweden: A way to economic self-reliance?", IZA discussion paper no. 1130.
- Audrecht, David B, Carree, Martin A, Steel, A J and Turik, Roy (2005), "Does Self-employment Reduce Unemployment?", CEPR Discussion Paper Series no. 5057.
- Blanchflower, David G. (2004), "Self-employment: More may not be better", *Swedish Economic Policy Review*, Vol. 11, no. 2, 15-73.
- Böhlmark, Anders (2003), "Over- and Undereducation in the Swedish Labour Market. Incidence, Wage-effects and Characteristics 1968-2000, work in progress, Swedish Institute for Social Research.
- Clark, Kenneth and Drinkwater, Stephen (2000), "Pushed out or pulled in? Self-employment among ethnic minorities in England and Wales", *Labour Economics*, Vol. 7, 603-628.

- Constant, Amelie and Zimmermann, Klaus F. (2004), "The making of entrepreneurs in Germany: Are native men and immigrants alike?", IZA DP No. 1440.
- Delmar, Frédéric, Wiklund, Johan and Sjöberg, Karin (2003), "The involvement in self-employment among the Swedish science and technology labor force between 1990 and 2000", Working Paper, A2003:017, Swedish Institute for Growth Policy Studies.
- Dolton, Peter H. and Makepeace, Gerald H. (1990), "Self-employment among graduates", *Bulletin of Economic Research*, Vol. 42, 35-53.
- Georgellis, Yannis and Wall, Howard J (1998), "Gender differences in self-employment: panel evidence from the former West Germany", working paper, Department of Economics and Finance, Brunel University.
- Hammarstedt, Mats (2005), "The predicted earnings differential and immigrant self-employment in Sweden", forthcoming in *Applied Economics*.
- Hout, Michael and Rosen, Harvey S. (2000), "Self-employment, family background, and race", *Journal of Human Resources*, Vol. 35, 670-692.
- Johansson, Edvard (2000), "Self-employment and the predicted earnings differential – Evidence from Finland", *Finnish Economic Papers*, Vol. 13, 45-55.
- Korpi, Tomas and Tåhlin, Michael (2005), "Skill Mismatch and Wage Growth", work in progress, Swedish Institute for Social Research.
- Lin, Zhengxi, Picot, Garnett and Yates, Janice (1999), "The Entry and Exit Dynamics of Self-Employment in Canada", *Business and Labor Market Analysis*, no. 134, Statistics Canada.
- Martinez-Granado, M (2002), "Self-employment and labour market transitions: a multiple state model", CEPR Discussion Paper Series no. 3661.
- Parker, Simon C. (2004), *The Economics of Self-Employment and Entrepreneurship*, Cambridge, Cambridge University Press.
- Rees, H and Shah, A (1986), "An empirical analysis of self-employment in the UK", *Journal of Applied Econometrics*, Vol. 1, 95-108.
- Taylor, Mark (1996), "Earnings, independence or unemployment: why become self-employed?", *Oxford Bulletin of Economics and Statistics*, Vol. 58, 253-265.
- Taylor, Mark (2004), "Self-employment in Britain: When, Who and Why?", *Swedish Economic Policy Review*, Vol. 11, no. 2, 139-173.

TABLES

Table 1 Mean values of covariates in 1998 for the whole population, for those who were self-employed in 2002 and for those who were not self-employed in 2002, per cent

Covariates measured in 1998	Wage earners in 1998	Not self-employed in 2002	Self-employed in 2002	Wage earners in 1998	Not self-employed in 2002	Self-employed in 2002
	ANNUAL LABOUR INCOME			HOURLY WAGE		
Share (%)	100.0	97.2	2.8	100.0	98.3	1.07
Age (years)	41.1	41.1	39.9	42.5	42.8	41.2
Age cohorts						
20-25	7.7	7.7	7.4	5.3	5.3	6.2
26-30	12.5	12.4	14.1	10.5	10.5	13.3
31-35	14.9	14.8	17.8	13.8	13.7	16.3
36-40	13.0	12.9	14.7	13.0	13.0	14.2
41-45	13.3	13.3	13.3	14.1	14.1	13.4
46-50	14.1	14.1	12.4	15.6	15.6	11.6
51-55	15.1	15.1	12.6	17.0	17.1	13.0
56-60	9.6	9.6	7.8	10.8	10.8	12.1
Education						
Primary school less than 9 years	7.9	8.0	4.2	7.2	7.2	3.3
Primary school 9(10) years	11.9	11.9	13.2	9.6	9.6	8.6
Upper secondary 2 years or less	32.7	32.7	32.4	30.1	30.1	28.1
Upper secondary more than 2 years	17.7	17.6	19.5	16.6	16.6	17.6
Higher education less than 3 years	14.8	14.8	15.2	16.8	16.7	18.5
Higher education 3 years or more	13.9	13.9	14.9	18.0	17.9	22.2
Post graduate education	1.2	1.2	0.7	1.9	1.9	1.8
Marital status						
Married	47.6	47.6	48.6	50.7	50.7	51.1
Cohabiting	11.5	11.4	13.2	10.7	10.7	12.0
Other	40.9	41.0	38.2	38.6	38.6	36.9
Small children	20.9	20.7	25.7	19.9	19.8	24.2
Family background						
Parents born abroad	2.9	2.9	3.5	2.9	2.9	3.5
One parent born abroad	10.7	10.7	11.0	10.9	10.9	11.1
Place of residence						
Big city areas (Stockholm. Gothenburg. Malmoe)	32.9	32.5	50.0	31.3	31.1	44.8
Sector						
Public	17.5	82.1	94.6	34.2	65.7	72.4
Private	82.5	17.9	5.4	65.8	34.3	27.6
Income in 1998						
Annual income/Hourly wage in 1998 (SEK) (standard deviation)	271 087 (157 363)	270 858 (156 210)	279 090 (193 321)	123.2 (49.6)	123.0 (49.4)	141.0 (69.1)
Number of observations	1 240 544	1 206 134	34 410	635 107	628 329	6 778

Table 2 OLS Income regression. Dependent variable is log of annual income from being a wage-earner in 1998 and hourly wage in 1998. Robust standard errors in parentheses

	Log(income) in 1998	Log(wage) in 1998
Age cohorts		
20-25	Ref.	Ref.
26-30	0.187 (0.001)**	0.093 (0.001)**
31-35	0.259 (0.002)**	0.153 (0.001)**
36-40	0.296 (0.002)**	0.186 (0.001)**
41-45	0.314 (0.002)**	0.204 (0.001)**
46-50	0.327 (0.002)**	0.226 (0.001)**
51-55	0.344 (0.002)**	0.257 (0.001)**
56-60	0.323 (0.002)**	0.260 (0.002)**
Marital status		
Other	Ref.	Ref.
Married	0.107 (0.001)**	0.083 (0.001)**
Cohabiting	0.056 (0.001)**	0.040 (0.001)**
Small children		
	-0.030 (0.001)**	-0.001 (0.001)
Country of origin		
Born in Sweden. parents born in Sweden	Ref.	Ref.
Parents immigrants	-0.014 (0.002)**	-0.003 (0.002)
One parent immigrant	-0.013 (0.001)**	-0.009 (0.001)**
Education		
Primary school less than 9 years	-0.209 (0.001)**	-0.200 (0.001)**
Primary school 9(10) years	-0.150 (0.001)**	-0.126 (0.001)**
Upper secondary 2 years or less	-0.102 (0.001)**	-0.100 (0.001)**
Upper secondary more than 2 years	Ref.	Ref.
Higher education less than 3 years	0.058 (0.001)**	0.062 (0.001)**
Higher education 3 years or more	0.240 (0.001)**	0.241 (0.001)**
Post graduate education	0.407 (0.004)**	0.430 (0.003)**
Place of residence		
Living in one of the three big city areas	0.082 (0.001)**	0.099 (0.001)**
Constant	12.099 (0.001)**	4.488 (0.001)**
Number of observations	1 240 544	635 107
R ² (adjusted)	0.25	0.39

Notes: Standard errors in parentheses; * significant at 5%; ** significant at 1%.

Table 3 Population divided by if they had an income lower than the predicted one (under), an income close to the predicted one (middle), or an income higher than the predicted one (over). Mean values, per cent.

Covariates measured in 1998	Under	Middle	Over
Share (%)	33.0	33.0	34.0
Self-employed in 2002	3.2	2.2	2.9
Age (years)	42.0	40.8	40.6
Age cohorts			
20-25	6.2	6.6	10.1
26-30	10.4	14.0	12.9
31-35	13.9	16.2	14.5
36-40	13.3	13.2	12.5
41-45	14.1	13.0	12.7
46-50	15.5	13.3	13.6
51-55	16.5	14.3	14.5
56-60	10.1	9.4	9.3
Education			
Primary school less than 9 years	6.7	10.0	7.0
Primary school 9(10) years	10.7	13.1	11.9
Upper secondary 2 years or less	30.3	36.4	31.3
Upper secondary more than 2 years	18.3	16.3	18.4
Higher education less than 3 years	15.4	13.8	15.1
Higher education 3 years or more	17.3	9.4	15.0
Post graduate education	1.4	0.8	1.3
Marital status			
Married	52.3	45.5	45.1
Cohabiting with children	10.7	12.7	11.0
Other	37.0	41.7	43.9
Small children	21.1	21.5	20.1
Place of residence			
Big city areas (Stockholm, Gothenburg, Malmoe)	35.8	28.9	34.2
Family background			
Parents born abroad	2.8	2.8	3.1
One parent born abroad	10.8	10.5	10.9
Income			
Income in 1998	192 277	245 471	372 442
(Standard deviation)	(50 557)	(48 014)	(226 223)
Decile in distribution of the ratio			
D1	30.3	-	-
D2	30.3	-	-
D3	30.3	-	-
D4	9.1	21.2	-
D5	-	30.3	-
D6	-	30.3	-
D7	-	18.2	11.8
D8	-	-	29.4
D9	-	-	49.4
D10	-	-	29.4
Number of observations	409 381	409 383	421 780

Table 4 Annual income, predicted annual income, and the difference between them for the three groups.

ALL				
Under (N=409 380)	Mean	Std.Dev.	Min	Max
Income	192 537	50 818	20 016	410 526
Predicted income	261 936	51 197	145 738	460 050
Difference	-69 399	41 453	-402 653	-16 595
Middle (N=409 379)	Mean	Std.Dev.	Min	Max
Income	244 854	47 643	134 075	511 011
Predicted income	245 625	45 685	145 738	460 050
Difference	-771	15 390	-48 210	50 961
Over (N=421 785)	Mean	Std.Dev.	Min	Max
Income	372 788	226 105	164 489	2.30e+07
Predicted income	254 536	56 984	143 686	460 050
Difference	118 252	202 068	17 180	2.26e+07
SELF-EMPLOYED				
Under (N=13 004)	Mean	Std.Dev.	Min	Max
Income	172 864	58 225	23 914	404 274
Predicted income	260 684	50 568	152 554	460 050
Difference	-87 820	54 023	-363 563	-16 869
Middle (N=9 143)	Mean	Std.Dev.	Min	Max
Income	247 838	50 017	140 743	490 248
Predicted income	248 477	47 830	152 554	460 050
Difference	-639	15 730	-46 026	45 265
Over (N=12 263)	Mean	Std.Dev.	Min	Max
Income	415 036	261 222	173 264	1.12e+07
Predicted income	263 773	58 339	152 554	460 050
Difference	151 262	236 911	18 186	1.09e+07
NOT SELF-EMPLOYED				
Under (N=396 376)	Mean	Std.Dev.	Min	Max
Income	193 183	50 427	20 016	410 526
Predicted income	261 977	51 217	145 738	460 050
Difference	-68 794	40 835	-402 653	-16 595
Middle (N=400 236)	Mean	Std.Dev.	Min	Max
Income	244 785	47 586	134 075	511 011
Predicted income	245 560	45 633	145 738	460 050
Difference	-774	15 383	-48 210	50 961
Over (N=409 522)	Mean	Std.Dev.	Min	Max
Income	371 523	224 847	164 489	2.30e+07
Predicted income	254 259	56 921	143 686	460 050
Difference	117 263	200 848	17 180	2.26e+07

Note: Predicted income is calculated as $\exp(\log(\text{predicted income}))$ from the Mincer equation in Table 2. Now the three groups are constructed on the basis of the value of the ratio defined as “annual income/ $\exp(\log(\text{predicted annual income}))$ ”. In all other tables and calculations the ratio is defined as “ $\log(\text{annual income})/\log(\text{predicted annual income})$ ”. For this reason the size of the groups is a bit different from what we otherwise use.

Table 5a Probability to be self-employed in 2002 estimated with probit regressions. Marginal effects are shown. The ratio is based on annual income

Independent variables	MODEL I	MODEL II	MODEL III	MODEL IV	MODEL I	MODEL II	MODEL III	MODEL IV
	No covariates	No covariates	Full set of covariates	Full set of covariates	No covariates	No covariates	Full set of covariates	Full set of covariates
	RATIO BASED ON ANNUAL INCOME				RATIO BASED ON ANNUAL INCOME			
Under	0.0101 (0.0004)**	0.0088 (0.0004)**	0.0085 (0.0004)**	0.0025 (0.0004)**				
Middle	Ref.	Ref.	Ref.	Ref.				
Over	0.0074 (0.0004)**	0.0090 (0.0004)**	0.0062 (0.0004)**	0.0119 (0.0005)**				
Log income in 1998	-	-0.0041 (0.0006)**	-	-0.0153 (0.0007)**	-	-0.0023 (0.0006)**	-	-0.0163 (0.0010)**
DECILE								
1					0.0309 (0.0011)**	0.0289 (0.0011)**	0.0272 (0.0010)**	0.0121 (0.0011)**
2					0.0043 (0.0008)**	0.0039 (0.0008)**	0.0034 (0.0007)**	-0.0007 (0.0007)
3					0.0011 (0.0007)	0.0008 (0.0007)	0.0005 (0.0007)	-0.0017 (0.0006)**
4					0.0011 (0.0007)	0.0010 (0.0007)	0.0008 (0.0007)	-0.0002 (0.0007)
5					Ref.	Ref.	Ref.	Ref.
6					0.0012 (0.0007)	0.0014 (0.0007)	0.0011 (0.0007)	0.0022 (0.0007)**
7					0.0011 (0.0007)	0.0014 (0.0007)*	0.0012 (0.0007)	0.0035 (0.0007)**
8					0.0028 (0.0007)**	0.0032 (0.0008)**	0.0027 (0.0007)**	0.0066 (0.0008)**
9					0.0062 (0.0008)**	0.0070 (0.0008)**	0.0057 (0.0007)**	0.0122 (0.0009)**
10					0.0190 (0.0009)**	0.0214 (0.0011)**	0.0160 (0.0009)**	0.0324 (0.0015)**
Observations	1 240 544	1 240 544	1 240 544	1 240 544	1 240 544	1 240 544	1 240 544	1 240 544
Number of parameters	2	3	21	22	9	10	28	29
*Wald chi ²	720.19	751.97	6342.98	6829.10	3248.27	3253.18	8573.44	8874.11

Notes: ** and * indicate that the coefficients are significantly different from zero at the 1 and 5 percent level of significance, respectively; figures in **bold** indicate that the estimate is significantly different from the corresponding estimate in the previous model. * The model is estimated with robust standard errors and STATA reports the Wald statistics instead of the LR-statistics that is reported when the model is estimated without robust standard errors.

Table 5b Probability to be self-employed in 2002 estimated with probit regressions. Marginal effects are shown. The ratio is based on hourly wage

Independent variables	MODEL I	MODEL II	MODEL III	MODEL IV	MODEL I	MODEL II	MODEL III	MODEL IV
	No covariates	No covariates	Full set of covariates	Full set of covariates	No covariates	No covariates	Full set of covariates	Full set of covariates
	RATIO BASED ON HOURLY WAGE				RATIO BASED ON HOURLY WAGE			
Under	-0.0000 (0.0003)	0.0009 (0.0003)**	-0.0008 (0.0003)**	0.0015 (0.0004)**				
Middle	Ref.	Ref.	Ref.	Ref.				
Over	0.0060 (0.0004)**	0.0024 (0.0004)**	0.0048 (0.0003)**	0.0010 (0.0004)**				
Log wage in 1998		0.0085 (0.0005)**		0.0114 (0.0007)**		0.0055 (0.0006)**		0.0071 (0.0011)**
DECILE								
1					0.0020 (0.0007)**	0.0028 (0.0007)**	0.0001 (0.0006)	0.0024 (0.0008)**
2					-0.0003 (0.0006)	0.0001 (0.0006)	-0.0008 (0.0006)	0.0003 (0.0006)
3					0.0003 (0.0006)	0.0007 (0.0006)	0.0001 (0.0006)	0.0008 (0.0006)
4					-0.0003 (0.0006)	-0.0001 (0.0006)	-0.0003 (0.0006)	0.0001 (0.0006)
5					Ref.	Ref.	Ref.	Ref.
6					0.0010 (0.0006)	0.0008 (0.0006)	0.0009 (0.0006)	0.0005 (0.0006)
7					0.0017 (0.0007)*	0.0011 (0.0006)	0.0015 (0.0006)*	0.0008 (0.0006)
8					0.0034 (0.0007)**	0.0023 (0.0007)**	0.0028 (0.0007)**	0.0015 (0.0007)*
9					0.0062 (0.0008)**	0.0038 (0.0007)**	0.0050 (0.0007)**	0.0025 (0.0007)**
10					0.0148 (0.0010)**	0.0080 (0.0010)**	0.0121 (0.0009)**	0.0056 (0.0012)**
Observations	635 107	635 107	635 107	635 107	635 107	635 107	635 107	635 107
Number of parameters	2	3	21	22	9	10	28	29
Wald chi ²	466.63	782.23	1288.46	1584.50	854.93	940.37	1582.07	1631.19

Notes: ** and * indicate that the coefficients are significantly different from zero at the 1 and 5 percent level of significance, respectively; figures in **bold** indicate that the estimate is significantly different from the corresponding estimate in the previous model.

Table 6 Descriptive statistics for the outcomes of the self-employed.

	All	Under	Middle	Over
Income from self-employment (SEK) ¹	246 599	188 611	227 393	322 470
Number of employees				
No employees	53.7 %	56.7 %	53.5 %	50.7 %
Between 1 and 10 employees	35.4 %	37.2 %	35.6 %	33.3 %
More than 10 employees	10.9 %	6.1 %	10.9 %	15.9 %
Turnover (SEK) ²	6 775 043	4 047 951	5 892 208	10 290 113
Turnover, given positive turnover (SEK) ³	6 943 925	4 155 119	6 017 288	10 297 360
Limited liability company	54.9 %	53.1 %	51.5 %	59.3 %
Number of observations	34 410	13 006	9 148	12 256

¹In total, 32 self-employed have reported zero income from self-employment in 2002.

²Information is missing for a total of 2 661 firms.

³In total, 772 firms reported negative or zero turnover in 2002.

Table 7 Income from self-employment. OLS regressions with log(income) from self-employment as the dependent variable. Robust standard errors in parentheses.

Independent variables	MODEL I	MODEL II	MODEL III	MODEL IV	MODEL I	MODEL II	MODEL III	MODEL IV
	No covariates	No covariates	Full set of covariates, including industry	Full set of covariates, including industry	No covariates	No covariates	Full set of covariates, including industry	Full set of covariates, including industry
	RATIO BASED ON ANNUAL INCOME				RATIO BASED ON HOURLY WAGE			
Under	-0.210 (0.014)***	0.028 (0.016)*	-0.262 (0.013)***	-0.055 (0.015)***	-0.192 (0.045)***	-0.137 (0.045)***	-0.236 (0.040)***	-0.087 (0.044)**
Middle	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Over	0.246 (0.015)***	-0.003 (0.016)	0.221 (0.013)***	0.049 (0.015)***	0.315 (0.040)***	0.134 (0.046)***	0.204 (0.036)***	-0.037 (0.045)
Log income in 1998/log wage rate in 1998		0.572 (0.017)***		0.459 (0.020)***		0.425 (0.057)***		0.725 (0.083)***
Observations	34 378	34 378	34 360	34 360	6 776	6 776	6 774	6 774
R ²	0.03	0.06	0.28	0.30	0.02	0.03	0.27	0.28

Notes: ***, ** and * indicate that the coefficients are significantly different from zero at the 1, 5 and 10 percent level of significance respectively.

Table 8a Number of employees. Multinomial logit. Relative risk ratios are presented. Standard errors in parentheses. Reference category is “No employees”.

Independent variables	MODEL I No covariates		MODEL II No covariates		MODEL III Full set of covariates, including industry		MODEL IV Full set of covariates, including industry	
	1-10 employees	More than 10 employees	1-10 employees	More than 10 employees	1-10 employees	More than 10 employees	1-10 employees	More than 10 employees
RATIO BASED ON ANNUAL INCOME								
Under	0.984 (0.029)	0.527 (0.027)***	1.053 (0.034)	0.660 (0.036)***	0.726 (0.026)***	0.329 (0.019)***	0.920 (0.038)**	0.538 (0.036)***
Middle	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Over	0.987 (0.030)	1.536 (0.067)***	0.921 (0.031)**	1.183 (0.059)***	1.273 (0.046)***	2.149 (0.112)***	1.046 (0.043)	1.399 (0.085)***
Log income in 1998			1.177 (0.039)***	1.782 (0.090)***			1.704 (0.086)***	3.165 (0.255)***
Observations	34 410		34 410		34 392		34 392	
# parameters	4		6		156		158	
LR chi2(#)	644.40		777.97		14993.81		15233.81	

Table 8b The same as above but with hourly wage

Independent variables	MODEL I No covariates		MODEL II No covariates		MODEL III Full set of covariates, including industry		MODEL IV Full set of covariates, including industry	
	1-10 employees	More than 10 employees	1-10 employees	More than 10 employees	1-10 employees	More than 10 employees	1-10 employees	More than 10 employees
RATIO BASED ON HOURLY WAGE								
Under	0.887 (0.072)	0.583 (0.095)***	0.888 (0.073)	0.607 (0.099)***	0.748 (0.070)***	0.430 (0.080)***	0.901 (0.091)	0.559 (0.110)***
Middle	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Over	1.086 (0.077)	1.778 (0.210)***	1.081 (0.090)	1.557 (0.212)***	1.168 (0.095)*	1.846 (0.259)***	0.866 (0.089)	1.243 (0.210)
Log wage in 1998			1.011 (0.105)	1.355 (0.204)**			2.471 (0.450)***	3.379 (0.954)***
Observations	6 778		6 778		6 776		6 776	
# parameters	4		6		148		150	
LR chi2(#)	83.95		88.00		2317.99		2351.10	

Table 9 Turnover in 2002. OLS regression with log(turnover) as the dependent variable. Robust standard errors in parentheses.

Independent variables	MODEL I	MODEL II	MODEL III	MODEL IV	MODEL I	MODEL II	MODEL III	MODEL IV
	No covariates	No covariates	Full set of covariates, including industry	Full set of covariates, including industry	No covariates	No covariates	Full set of covariates, including industry	Full set of covariates, including industry
	RATIO BASED ON ANNUAL INCOME				RATIO BASED ON HOURLY WAGE			
Under	-0.240 (0.024)***	-0.092 (0.026)***	-0.418 (0.019)***	-0.152 (0.022)***	-0.263 (0.057)***	-0.264 (0.058)***	-0.365 (0.049)***	-0.178 (0.054)***
Middle	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Over	0.266 (0.026)***	0.108 (0.028)***	0.406 (0.021)***	0.179 (0.023)***	0.339 (0.054)***	0.344 (0.063)***	0.322 (0.046)***	0.019 (0.057)
Log income 1998, log hourly wage 1998		0.364 (0.028)***		0.601 (0.029)***		-0.013 (0.078)		0.908 (0.109)***
Observations	30 977	30 977	30 977	30 977	5 877	5 877	5 877	5 877
R ²	0.02	0.02	0.39	0.40	0.02	0.02	0.36	0.37

Table 10 Probability to have a LLC. Probit regression where the dependent variable equals one if the firm is registered as an LLC. Marginal effects are presented and robust standard errors are in parentheses.

Independent variables	MODEL I	MODEL II	MODEL III	MODEL IV	MODEL I	MODEL II	MODEL III	MODEL IV
	No covariates	No covariates	Full set of covariates, including industry	Full set of covariates, including industry	No covariates	No covariates	Full set of covariates, including industry	Full set of covariates, including industry
	RATIO BASED ON ANNUAL INCOME				RATIO BASED ON HOURLY WAGE			
Under	0.016 (0.007)**	0.097 (0.007)***	-0.094 (0.009)***	-0.035 (0.010)***	-0.038 (0.016)**	-0.009 (0.016)	-0.077 (0.015)***	-0.049 (0.016)***
Middle	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Over	0.077 (0.007)***	-0.006 (0.008)	0.099 (0.008)***	0.052 (0.009)***	0.129 (0.014)***	0.034 (0.016)**	0.087 (0.014)***	0.033 (0.017)*
Log income, log hourly wage in 1998		0.197 (0.008)***		0.132 (0.012)***		0.223 (0.020)***		0.153 (0.028)***
Observations	34 410	34 410	34 372	34 372	6 778	6 778	6 559	6 559
# parameters	2	3	68	69	2	3	58	59
Wald chi ²	154.08	750.03	5679.29	5760.72	171.52	288.05	1275.98	1287.09

Figure 1 Average predicted probability to be self-employed in 2002 plotted against the value of the ratio.

