RATIO

Ratio Working Paper No. 281

Entrepreneurship and Income Inequality

Daniel Halvarsson*

Martin Korpi**

Karl Wennberg***

*daniel.halvarsson@ratio.se, The Ratio Institute, P.O. Box 3203, SE-103 64 Stockholm, Sweden
**martin.korpi@ratio.se, The Ratio Institute, P.O. Box 3203, SE-103 64 Stockholm, Sweden
***karl.wennberg@liu.se, The Ratio Institute, P.O. Box 3203, SE-103 64 Stockholm and Department of Management and Engineering (IEI) and Institute of Analytical Sociology (IAS) Norrköping, Linköping University, Sweden

Entrepreneurship and Income Inequality

Daniel Halvarsson
Ratio Institute
Stockholm, Sweden
E-mail: daniel.halvarsson@ratio.se

Martin Korpi
Ratio Institute
Stockholm, Sweden
E-mail: martin.korpi@ratio.se

Karl Wennberg **
Institute for Analytical Sociology (IAS)
Linköping University
Norrköping, Sweden
E-mail: karl.wennberg@liu.se
&
Ratio Institute

* We are grateful for helpful comments by Chanchal Balachandran; Frédéric Delmar; Vivek Sundriyal; Jesper Roine; and seminar participants at the Institute of Analytical Sociology (IAS) and the department of Economics at Linköping University, the Ratio Institute, Copenhagen Business School, Instituto Superior Técnico, University of Lisbon, and the DynIPer conference, Marseille, May 26–27, 2016 (Dynamics of Inequalities and their Perception). This work was generously funded by the Swedish Research Council (DNR 340-2013-5460) and Riksbankens Jubileumsfond (DNR M12-0301:1). Any errors remain ours alone.

** Corresponding author.

Keywords: Inequality, entrepreneurship, income decomposition

JEL-Codes: J31, J24, L26, 015

Entrepreneurship and Income Inequality

Abstract: Entrepreneurship research highlights individual entrepreneurship as a simultaneous source of enhanced income mobility for some but a potential source of poverty for others. Research on inequality has furthered new types of models to decompose and problematize various sources of income inequality in modern economies, but attention to entrepreneurship as an increasingly prevalent occupational choice in these models remains scant. This paper seeks to bridge these two literatures by applying regression-based income decomposition among entrepreneurs and paid workers, distinguishing between self-employed (SE) and incorporated self-employed (ISE) individuals in Sweden. We find that the proportion of self-employed in the workforce significantly increases income dispersion by way of widening the bottom end of the distribution, whereas the proportion of incorporated self-employed contributes only marginally to income dispersion at the top end of the distribution.

Keywords: Inequality, entrepreneurship, income decomposition

JEL-Codes: J31, J24, L26, 015

1. Introduction

The last few decades have witnessed a notable rise in both self-employment and small business ownership in most developed economies (Blanchflower, 2000; Steinmetz and Wright, 1989). Much of public policy in this context has been directed toward increasing the supply of entrepreneurs, most often measured as the number of self-employed or the number of businesses in an economy (Shane, 2009). Parallel to this "rise of the entrepreneurial economy" (Audretsch, 2009), scholars have noted a trend of accelerating income inequality and increasingly stratified unemployment (Atkinson, 2003; Autor, 2014; Goldthorpe, 2010). This type of stratification is also apparent in research on entrepreneurship which highlights self-employment as a simultaneous source of enhanced income mobility for some but a potential source of poverty for a large fraction of the self-employed workforce (Åstebro et al., 2011).

This paper is motivated by these trends and addresses two orthogonal theoretical concerns in the literatures on entrepreneurship and economic inequality. First, while entrepreneurship constitutes a source of enhanced income mobility for some, the majority of entrepreneurs earn incomes lower than population average (e.g. Hamilton, 2000; Åstebro et al., 2011). Still, work that probes the relationship between entrepreneurship and overall workforce income inequality remains scant (Van Praag and Versloot, 2007; Wright and Zahra, 2011). Second, recent inequality research has furthered new types of models to decompose and problematize various sources of income inequality in modern economies (e.g. Cowell and Fiorio, 2011; Creedy and Hérault, 2011; Thewissen et al., 2013), but these models have not incorporated entrepreneurship as an increasingly common occupation.

Our paper seeks to bridge these literatures by developing and empirically testing a model of the relationship between entrepreneurship and income inequality. We start by discussing the different ways in which entrepreneurship can contribute to overall income inequality. We then develop an econometric model that lets us decompose aggregate measures of income dispersion into its individual micro-level sources using the generalized entropy (GE) index and distinguishing between the following groups, which are likely to contribute to income inequality in differing ways: workers (W), self-employed entrepreneurs (SE), and incorporated self-employed entrepreneurs (ISE) (as in e.g. Levine and Rubinstein, 2016; Åstebro and Tåg, 2015; Özcan, 2011). In tuning the GE index to different segments of the income distribution, we are able to study change in inequality at different parts of the distribution and the way these changes relate to SE and ISE. The analysis also includes an integrated factor-source decomposition whereby we consider a number of explanatory variables to account for inequality within each of the groups. This approach allows us to test how different explanatory variables relate to within-group inequality as well as the extent to which the same group-specific explanatory variables individually relate to total aggregate-level inequality (Cowell and Fiorio, 2011; Fields, 2003).

Based on analyses of Swedish labor force data, our empirical results reveal that on average, SE entrepreneurs exhibit lower incomes than workers and that this type of entrepreneurship consequently increases inequality by widening the bottom end of the distribution. ISE entrepreneurs, on the other hand, have higher incomes than workers, augmenting inequality from the top end of the income distribution by enhancing the total number of high-income earners in society. The combined contribution of entrepreneurship on aggregate income inequality accounts for a sizable share of inequality depending on the inequality measure. When emphasizing the lower end of the income distribution, SE accounts for about 30% of total inequality and ISE accounts for only a fraction, but the opposite is true when emphasizing upper income levels where ISE accounts for about 10% but the number of SE are trivial. This finding highlights the importance of also considering entrepreneurship—an increasingly prevalent occupation—in studies of income inequality. We extend our sub-group decomposition analysis with a regression-based factor source decomposition analysis. Our key finding here is that years of education and gender differences each account for about 4% of total income inequality. While years of education is prominent in accounting for the inequality of workers, it is also one of the key factors behind inequality among ISE entrepreneurs (but not among SE entrepreneurs), suggesting low returns to education among SE entrepreneurs (Van der Sluis et al., 2005). Gender differences' contributions to inequality are more prominent among workers than among any of the two entrepreneurial groups.

To the best of our knowledge, this paper is the first to assess the effects of entrepreneurship for income inequality using state-of-the art decomposition techniques. It also represents a first attempt at addressing the joint theoretical concern in the literatures on entrepreneurship and economic inequality regarding how changes in the relative number of entrepreneurs and their within-group income dispersion affect aggregate income dynamics in modern economies. Our methodology leans heavily on Cowell and Fiorio (2011) work, which shows how to decompose inequality indices using several sub-groups in a regression framework. This helps explain each group's contribution to overall inequality, providing a clearer picture of the group dynamics that drive inequality at the population level. Further, we are able to pinpoint the significance of a number of explanatory variables commonly used in income regressions, and we assess their individual contribution to sub-group and aggregate inequality.

Our paper is structured as follows: In the next section, we outline the theoretical background and previous literature. Section 3 outlines our empirical strategy and the regression-based decomposition method used to analyze the link between entrepreneurship and income inequality. Section 4 details the data used in the analysis, which is followed by the results in Section 5. The paper concludes with a discussion about the implications for public policy.

2. Entrepreneurship, income dynamics, and inequality

It has been noted in the empirical literature that much of the entrepreneurship in modern economies does not take the form of growing productive firms, but rather of increasing rates of self-employment (Sanandaji and Leeson, 2013; Stam, 2013; Steinmetz and Wright, 1989). However, research has been scant on the possible consequences of this development in terms of income inequality.

A substantial number of studies have investigated income differentials between employees and the self-employed. By and large, this literature has concluded that entrepreneurship generally results in earnings lower than comparable salaried work. For example, the well-cited papers by Blau (1987), Borjas and Bronars (1989), and Evans and Leighton (1989) estimate that the earnings of self-employed individuals are below those of workers and—for the latter two studies—that the distribution of these earnings is considerably skewed downwards (i.e., skewed toward low-income earners). Similarly, in a paper comparing the return to investment for US non-publicly traded equity with that of public equity, Moskowitz and Vissing-Jorgensen (2002) identify a large public equity investment premium and similarly argue that entrepreneurship has poor returns overall.

More recent studies have furthered these findings, confirming that entrepreneurial earnings are below those of comparable salaried workers on average, adding that the overall distribution of entrepreneurial earnings also comes with a substantially “fat” upward tail. Using data from the Canadian Survey of Labor and Income Dynamics (SLID) 1993–1994, Lin et al. (2000) address earnings differences at different quantiles of the earnings distribution and the extent and cyclicality of entry and exit into and out of entrepreneurship. Their study shows that the mean income of self-employed is about 20% lower than among comparable workers across the first four quantiles of the earnings distribution but more than double that of employed workers in the top fifth quantile.

Similarly, using monthly panel data on US male non-farm workers from 1983 to 1986, a well-cited study by Hamilton (2000) finds that most entrepreneurs persist in small businesses despite the fact that they have both lower initial earnings and lower earnings growth relative to employees. Estimating a median worker-entrepreneur earnings differential of around 35% across industries and ruling out the possibility that entrepreneurs have lower abilities on average, the study suggests that non-pecuniary benefits likely explain both entrepreneurial entry and persistence. However, Hamilton also finds support for the “superstar hypothesis”—namely, that entrepreneurial income at the very top of the earnings distribution is highly skewed upward because of a relatively small number very successful high-productivity individuals—and cautions that this pattern is not captured in his estimates of median income differentials.

This outcome is also in line with results from Åstebro et al. (2011) and Levine and Rubinstein (2016). In the first study, utilizing 1998–2004 panel data from the Korean Labor and Income Panel Study to test their model of occupational choice, Åstebro et al. find that entrants into self-employment are drawn disproportionately from both tails of the earnings distribution. However, in contrast to the Hamilton (2000) study, Åstebro et al. also find that the skewed income distribution also reflects the distribution of ability in the workforce, with workers with both above-average and below-average unobserved ability being more likely to engage in entrepreneurship. A recent study by Levine and Rubinstein (2016) also provides evidence in line with this outcome. Using US data from the Current Population Survey, 1994–2010, which separates salaried workers and entrepreneurs in the form of self-employed (SE) and incorporated self-employed (ISE), they pinpoint both the determinants behind the sorting of people into these different types of employment as well as their subsequent earnings. The authors argue that the earlier findings suggesting that entrepreneurs earn below the median income of salaried workers, essentially reflect the fact that there are significantly fewer incorporated entrepreneurs than self-employed, and that this latter group—which indeed does earn lower incomes—tends to dominate estimates of "average outcomes" in earning equations comparing entrepreneurs to workers. In contrast, their results suggest that incorporated entrepreneurs have earnings of around 30% above salaried workers with comparable traits and skills, who in turn earn more than their self-employed counterparts.

Similar to Åstebro et al. (2011), Levine and Rubinstein (2016) find that these outcomes reflect innate ability (i.e., non-cognitive traits and cognitive skills) to a significant degree. Whereas some traits were common to both types of entrepreneurs (e.g., having been engaged in illicit activities as teenagers and risky behavior), incorporated entrepreneurs generally scored higher in learning aptitude tests than self-employed individuals.[1]

Turning to studies that focus specifically of entrepreneurship and income inequality, we note a thinner empirical literature that emphasizes the association between inequality within

[1] These results are also in line with work specifically focusing on necessity entrepreneurship. Necessity entrepreneurs are defined as those who have started their own firms "because they cannot find a suitable role in the world of work, [so] creating a new business is their best available option" (Reynolds et al., 2005). Recent data from the Global Entrepreneurship Monitor suggest that between 3% and 30% of all entrepreneurs in OECD countries fit this category, with high fluctuations over time (Singer et al., 2014). These necessity entrepreneurs have been shown to exhibit limited income mobility as well as low gains in individual productivity (Block and Wagner, 2010; Poschke, 2013).

organizations and employees' transition to entrepreneurship. Rather than exploring possible links to workforce income dispersion per se, these studies have mostly focused on the conditions in which income dispersion within firms may represent a source of upward earnings mobility among individuals choosing to leave these firms for entrepreneurship (Carnahan et al., 2012; Kacperczyk and Balachandran, 2016; Sørensen and Sharkey, 2014). However, a few macro-oriented studies indicate that the greater the number of small firms in an economy, the more unequal is the earnings distribution in that economy (Davis, 2013; Fields and Yoo, 2000). In terms of explanations for this pattern, the literature has mostly provided very broad structural patterns. For example, (Lippmann et al., 2005) provide cross-country evidence on the relationship between workforce income inequality and the rate of entrepreneurship using Global Entrepreneurship Monitor (GEM) data. They find that entrepreneurship rates are higher in countries with significant income inequality and discuss seven structural factors broadly associated with this pattern: level of economic development, government policies, foreign direct investment, service sector growth, increasing labor market flexibility, wealth transfer programs, and variation in worker unionization.

To summarize the literature on earnings differentials between salaried workers and entrepreneurs, rising levels of entrepreneurship has been suggested to potentially increase inequality by expanding the share of either top or bottom income earners in the workforce or both. However, this possibility has been indirectly inferred from equations of earnings differentials between entrepreneurs and workers, and we do not know how entrepreneurship affects total inequality within the workforce as a whole. As for the literature using country comparisons to address the issue, this literature also suggests that entrepreneurship is positively related to inequality but is lacking in three regards: there is a lack of understanding about (1) what dynamics are involved, (2) which parts of the income distribution might be affected, and (3) how different types of entrepreneurship might affect overall inequality.

Resolving this puzzle and filling these gaps in the literature is vital if we are to provide any clear-cut answer on the relationship between entrepreneurship and income inequality. This also forms the basis for our overall research questions: *(1) How does entrepreneurship affect overall workforce income inequality? (2) Which parts of the income distribution are primarily affected and to what extent? (3) Do different types of entrepreneurship affect inequality in differing ways?*

3. Empirical strategy and method

Firstly, since there are numerous ways to conceptualize and measure inequality, each with a unique set of properties, we need to choose a measure appropriate to the research question. The most common measure is certainly the Gini coefficient (Akita et al., 1999). In spite of its prevalence and simplicity, the Gini coefficient does not fit our purposes for two reasons. First, it does not easily decompose into sub-groups and factor sources (Cowell, 2011), and since we want to both gauge the differential contribution of different types of entrepreneurs and salaried workers to aggregate inequality and understand the different factors behind group-level inequality, decomposability is crucial. Second, because we want to gauge these respective groups' contribution to inequality at various ranges of the income distribution, the Gini coefficient is also inappropriate since it mostly reflects incomes in the middle ranges of the distribution.

This requires a more flexible measure. Our choice is the generalized entropy index (GE index) which fulfils both of the requirements mentioned above. It allows us to study inequality within each group separately and assess how much the group as a whole contributes to aggregate inequality. Further, the GE index is defined as a function of a sensitivity parameter $\alpha \in (-\infty, \infty)$, which allows us to adjust the sensitivity of the measure to specific parts of the income distribution. By choosing different values of α when decomposing inequality into sub-groups, we can discern which part of the income distribution is most affected by each separate occupational group. Specifically, the lower the value of α, the more sensitive the GE-index is to dispersion in the lower parts of the income distribution. Conversely, the higher the level of α, the more sensitive the GE index is to dispersion in the upper parts of the income distribution (Cowell, 2000). Although the GE index is defined on a continuous axis, we limit our study to the following subset of values: $\{-1, 0, 1,2\}$. Of particular interest is the relative contribution of SE and ISE to overall inequality when α is tuned to either the bottom or top parts of the income distribution (i.e., $\alpha \in \{-1, 2\}$) as compared to the middle parts of the income distribution (i.e., $\alpha \in \{0,1\}$).[2]

[2] With the possible exception of $\alpha = -1$, $\{0,1,2\}$ constitute the most common values in similar studies using the GE index. Since $\alpha = -1$ turns the focus to the bottom of the income distribution, even more so than $\alpha = 0$, we include it in our analysis.

A formal definition of the GE index allows us to compare its properties to comparable indices. If we let $y = [y_1 \dots y_N]$ represent a vector of incomes for a total of N individuals in the population, the sample analogue of the GE index can then be defined as a function of α by the expression

$$GE(y;\alpha) = \frac{1}{\alpha(\alpha-1)N}\sum_{i=1}^{N}\left[\left(\frac{y_i}{\mu(y)}\right)^{\alpha} - 1\right], \qquad \alpha \in (-\infty,\infty) \cap \{0,1\}. \tag{1}$$

The sum is taken over the α-exponent of individual incomes y_i $(i = 1, \dots, N)$ divided by the mean income of the population $\mu(y)$ equivalent to the mean weighted α-moment of income y. We can interpret the different values of α as follows. Starting with $\alpha = 2$ and $GE(y;2)$, the expression in (1) corresponds to *half the squared coefficient of variation (CV)*, which is given by $\mathrm{var}(y)/2\mu(y)$ and also synonymous with the Hirschman-Herfindahl concentration index (Quintano et al., 2005). When α takes the values 1 or 0, in turn, $GE(y;1)$ and $GE(y;0)$, we need to evaluate the expression at the limit, $\alpha \to 1$ or $\alpha \to 0$. If $\alpha = 1$, the GE index is identical to the so-called Theil index, and when $\alpha = 0$, it corresponds to the mean logarithmic deviation (MLD). In this listed order, these entropy measures are sensitive to changes in top, upper, and middle income ranges, respectively. Lastly, when $\alpha = -1$, the emphasis of the index $GE(y;-1)$ turns to the bottom ranges of the distribution, as it contains the expected value of reciprocal income $(1/y)$.

3.1. Sub-group decomposition of the generalized entropy (GE) index

Having chosen an inequality measure, this section discusses the sub-group decomposability of the GE- index, an additive property that is crucial for evaluating the contribution from each different sub-group to overall inequality.[3]

Let us consider a total of J subgroups ($J = 3$ throughout this study). The GE index can then be divided into two aggregate parts,

$$GE(y;\alpha) = GE_b(y;\alpha) + GE_w(y;\alpha), \tag{2}$$

[3] $GE(y;\alpha)$ are strictly decomposable indices because their between-group components measure the change in overall inequality when group means are equalized while keeping the within-group component constant.

one between-group part $GE_b(y;\alpha)$, which reflects inequality as measured by differences in mean income *between* groups, and one within-group part $GE_w(y;\alpha)$, which reflects the dispersion of individual income *within* the groups. Here, the latter expression, $GE_w(y;\alpha)$, is a weighted sum of the GE index computed for each of the sub-groups ($j = 1, \dots, J$) using

$$GE_w(y;\alpha) = \sum_{j=1}^{J} w_j GE(y_j;\alpha). \tag{3}$$

The weight is in turn defined by $w_j = p_j r_j^{\alpha}$, where $p_j = N_j/N$ corresponds to a population weight, and $r_j = \mu(Y_j)/\mu(Y)$ corresponds to the mean income ratio between sub-group j and the mean income of the population (Fiorio and Cowell, 2011). Substituting this expression for the weight into expression (3) and using (1) yields the following formula for total within-group inequality:

$$GE_w(y;\alpha) = \frac{1}{\alpha^2-\alpha}\sum_{j=1}^{J}\frac{N_j}{N}\left(\frac{\mu(Y_j)}{\mu(Y)}\right)^{\alpha}\frac{1}{N_j}\sum_{i=1}^{N_j}\left[\left(\frac{y_{ji}}{\mu\left(y_j\right)}\right)^{\alpha}-1\right]. \tag{4}$$

This expression thus reflects the share of total inequality $GE(y;\alpha)$ that results from income dispersion within each of the separate sub-groups *combined*. Using the identity in eq. (2) and the expression for $GE_w(y;\alpha)$ in eq. (4), the 'between part' $GE_b(y;\alpha)$ can easily be backed out to

$$GE_b(y;\alpha) = \frac{1}{\alpha^2-\alpha}\sum_{j=1}^{J}\frac{N_j}{N}\left[\left(\frac{\mu(Y_j)}{\mu(Y)}\right)^{\alpha}-1\right], \tag{5}$$

which in turn captures "residual inequality" once within-group inequality is accounted for.[4] The GE index thereby accounts for the differences in mean incomes across the sub-groups—namely, between-group inequality.[5]

---

[4] The 'between part' in eq. (5) is written slightly different here compared to, for example, Fiorio and Cowell (2011), where the population weight is included within the square brackets as follows: $\frac{1}{\alpha^2-\alpha}\left(\sum_{j=1}^{J}\left[\frac{N_j}{N}\left(\frac{\mu(Y_j)}{\mu(Y)}\right)^{\alpha}\right]-1\right)$. Although the two expressions are equivalent, when using this expression, it is not clear how to calculate the contribution to aggregate inequality from one particular group's 'between part' since it leaves out a "residual term" of $\frac{-1}{\alpha^2-\alpha}$ that also needs to be distributed among the groups. Using the expression in (5) neatly takes care of this inconvenience by partitioning the residual between the groups by the amount of $\frac{-N_j}{\alpha^2-\alpha}$.

[5] Using the expression in Footnote 4, similar expressions can be reached for the MLD and Theil index. We do not show separate derivations for these measures here. As long as $\alpha > 1$, it is apparent that for groups with a lower mean income than the population average, the 'between part' contributes negatively to overall inequality.

Note also that the sub-group decomposition in (4) and (5) allows for an alternative decomposition by defining the contribution to total inequality $GE(y;\alpha)$ from a given group j,

$$\widetilde{GE}(y_j;\alpha) = \frac{p_j\left(r_j{}^{\alpha} - 1\right)}{\alpha^2 - \alpha} + w_j GE(y_j;\alpha). \tag{6}$$

The first term in this expression reflects group j's contribution to the between component of inequality, and the second term reflects the group's contribution from its within component. This expression is useful whenever we are interested in the total contribution from one particular sub-group rather than the combined within part or between part of all sub-groups taken together. To get back to aggregate inequality $GE(y;\alpha)$ from this expression, we simply take the sum over the J groups. If not mentioned otherwise, it is this expression we refer to when discussing a sub-group's total contribution to aggregate inequality.

3.2. Factor source decomposition of subgroup within inequality

In this section, we turn to inequality decomposition by factor sources, something that often accompanies sub-group inequality decomposition analyses. For each of the sub-groups, this approach allows to assess how different income determinants contribute to inequality. Traditionally, a factor source analysis takes into account the incomes from different sources, for instance salaried income, capital income, or transfer payments. In our analysis, we rely extensively on Fields (2003) and Fiorio and Cowell (2011), who develop a regression-based framework that allows for a much wider set of possible sources on which to base the analysis. Furthermore, using regression analysis to decompose income inequality, we can lend extensively from the literature that investigates the determinants of income using Mincer-type wage regressions.

The basic approach for our factor source decomposition is as follows. Let the income for individual i $(i = 1, \ldots, N)$ in subgroup j $(j = 1, \ldots, J)$ be split into a sum of K different factor sources (i.e., components):

The converse is true for groups with a mean income larger than the population average. For $\alpha < 1$, the relationship becomes the opposite.

$$y_{ji} = y_{ij1} + y_{ij2} + \cdots + y_{ijK}. \tag{7}$$

Provided that an inequality index denoted by $I(y_j)$ satisfies the six basic assumptions laid out by Shorrocks (1982, see appendix 1), the index can be decomposed into a sum of K inequality components, here denoted by $S_{jk}(y_{jk}, y_j)$ for $k = 1, \dots, K$ as

$$I(y_j) = S_{j1}\left(y_{j1}, y_j\right) + S_{j2}\left(y_{j2}, y_j\right) + \cdots + S_{jK}\left(y_{jK}, y_j\right), \tag{8}$$

where y_{jk} refers to the kth income source for the jth group. In fact, provided the assumptions are fulfilled, this type of decomposition is invariant to the choice of inequality measure $I(y)$ (Shorrocks, 1982). To see this, define the share—that is, the proportional contribution of S_{jk} to $I(y_j)$—by

$$s_{jk} \equiv \frac{S_{jk}(y_{jk}, y_j)}{I\left(y_j\right)}. \tag{9}$$

Since $s_{j1} + s_{j2} + \cdots + s_{jK} = 1$ by construction, multiplying through with $I(y_j)$ shows that s_{jk} becomes the "loading" of factor k to the inequality $I(y_j)$, which is given by,

$$I(y_j) = s_{j1}I(y_j) + s_{j2}I(y_j) + \cdots + s_{jK}I(y_j). \tag{10}$$

For this class of inequality measures (the GE index among them), the function s_{jk} can thus be written in terms of the covariance between the income component y_{jk} and total income y_j divided by the variance of y_j. Thus,

$$s_{jk} = \frac{\sigma(y_{jk}, y_j)}{\sigma^2(y_j)}. \tag{11}$$

This result comes from Shorrocks (1982) and was first used by Fields (2003) to connect income regression analysis with the a priori inequality decomposition methods.

3.3. Combining factor source decomposition with Mincer-type regression

Following Fields (2003), we consider a linear model of income for individual i in group j as

$$y_{ij} = b_{j0} + \sum_{k=1}^{K-1} b_{jk}x_{ijk} + u_{ij}, \tag{12}$$

which includes $K-1$ number of (potentially endogenous) explanatory variables x_{kji}, and an i.i.d. error term u_{ij}. Because regression (7) has the same linear form as the expression in equation (7), this in turn means that a factor source decomposition of the inequality of y_j, $I\left(y_j\right)$ can be accomplished by mapping $b_{jk}x_{ijk}$ to y_{jk} in eq. (7). While the result in Fields (2003) are valid with respect to the full population, here, we keep with the sub-group indexation j as in Fiorio and Cowell (2011). By expanding the covariance, we are better able to see how different parts of the regression equation contributes to (y_j):

$$s_{jk} = b_{jk}^2 \frac{\sigma^2(x_{jk})}{\sigma^2\left(y_j\right)} + b_{jk} \sum_{r \neq k}^{K-1} b_{jr}\, \rho(x_{jr}, x_{jk}) \frac{\sigma(x_{jr})\sigma(x_{jk})}{\sigma^2\left(y_j\right)} + b_{jk}\rho(u_j, x_{jk}) \frac{\sigma(u_j)\sigma(x_{jk})}{\sigma^2\left(y_j\right)}, \tag{13}$$

$$s_{jK} = \frac{\sigma^2(u_j)}{\sigma^2(y_j)} + \sum_{k=1}^{K-1} b_{jk}\, \rho(u_j, x_{jk}) \frac{\sigma(u_j)\sigma(x_{jk})}{\sigma^2(y_j)}. \tag{14}$$

Starting with s_{jk}, the first term gives the direct contribution of $b_{jk}x_{jk}$, the second represents the contribution if x_{jk} is correlated with x_{jr} for $r \neq k$ other explanatory variables (multicollinearity), and the third term represents variation due to endogeneity (i.e., when the x_{jk} term is correlated with the residual term u_j). As for the s_{jK} share, it represents the contribution to inequality from the unobserved part of equation (12) expressed in terms of the residual's direct contribution as well as the variation resulting from any endogenous explanatory variables.

Since all direct contributions are squared, a necessary condition for s_k to be negative is that either x_{jk} is correlated with at least one x_r ($r \neq k$ and is hence multicolinear) or x_{jk} is correlated with the error term and is hence endogenous (Fiorio and Cowell, 2011). On the other hand, if all assumptions in OLS are satisfied, s_{kj} and s_{Kj} are reduced to their respective first terms, $s_{jk} = b_{jk}^2\sigma^2(x_{jk})/\sigma^2\left(y_j\right)$ and $s_{jK} = \sigma^2(u_j)/\sigma^2\left(y_j\right)$.

To find the point estimates of s_{jk} and s_{jK}, we simply run the appropriate regression on (no-log) income and collect the estimates. These are then combined with information about covariance and the correlation matrix to form the sample versions $\hat{s}_k$ and $\hat{s}_K$ of s_k and s_K.

When it comes to the standard deviations of the estimates $\hat{s}_k$ and $\hat{s}_K$, Fiorio and Cowell (2011) suggest a bootstrapping procedure considering the difficulty of computing analytical standard errors. Since the computation outlined in this section involves numerous separate computations and, in our case, almost 3.5 million observations, bootstrapping becomes very time consuming. Fortunately, using a recent result in Bigotta et al. (2015, p.5, theorem 2), we can compute asymptotic standard errors for $\hat{s}_k$ based on the square root of the k^th diagonal elements in the following covariance matrix $\hat{\Sigma}(s_{jk})$:

$$\hat{\Sigma} = \frac{1}{N}\hat{\sigma}^2(u)\frac{\left(I_K \otimes \hat{\beta}^T X^T X\right) L(X^T X)^{-1} L^T \left(I_K \otimes \hat{\beta}^T X^T X\right)}{\left(\hat{\beta}^T X^T X\hat{\beta} + \hat{\sigma}^2(u)\right)^2}. \tag{15}$$

Here, we dispense with the j sub-group indexation, but $\hat{\Sigma}$ should be understood to correspond to the regression of a particular sub-group. I_K is a $K \times K$ identity matrix; $\otimes$ is the Kronecker product; and L is a $K^2 \times K$ selection matrix given by $L = [l_1 \quad \cdots \quad l_K]^T$, where l_k is a $K \times K$ matrix of zeros with 1 on the k^th diagonal entry. To calculate the 95% confidence interval for the k^th factor, we simply use the k^th diagonal element of $\hat{\Sigma}$ to form asymptotical standard errors with the formula

$$s_{jk} = \left[\hat{s}_{jk} \pm 1.96\sqrt{\hat{\Sigma}_{jkk}}\right].$$

3.4. Combining sub-group decomposition with regression-based factor source decomposition

This section is based on Fiorio and Cowell (2011), who present a unifying framework including both sub-group decomposition (as outlined in Section 3.1) and regression-based factor source decomposition (Section 3.3).

Because mean income in group j can be written in terms of a sum of regression outputs in the form $\mu(y_j) = \sum_{k=1}^{K-1} b_{jk}\mu(x_{jk})$, the inequality contribution from group j given by equation (6) can be restated as follows:

$$\widetilde{GE}(y_j;\alpha) = \frac{p_j}{\alpha^2-\alpha}\left(\left[\frac{\sum_{k=1}^{K-1} b_{jk}\mu(x_{jk})}{\sum_{k=1}^{K-1} b_{jk}\mu(x_k)}\right]^{\alpha} - 1\right) + w_j \sum_{k=1}^{K} GE(y_j;\alpha)s_{jk}. \qquad (16)$$

In the first expression gauging between-group inequality, the term in the nominator refers to the parameters from a regression restricted to the individuals in sub-group j, whereas the denominator gives the parameters from a regression on the full population. The second term which captures within-group inequality, is now expressed as a weighted sum by the $K-1$ factor loadings corresponding to the explanatory variables and by loading s_{jK} from the term u_{ijt}.

Except for the population shares calculated directly from the data, all information regarding the sub-group factor source decomposition are provided by regressing income on the set of explanatory variables in equation (12).

4. Data and descriptive statistics

4.1. Data

The empirical test for our model is based on microdata from Sweden for the years 2005 and 2013. The Swedish economy has one of the world's lowest rates of income inequality. However, inequality in Sweden has increased between 1985 and the early 2010s, similar to most OECD countries (OECD, 2015).[6] During the same period, the country has seen increasing rates of entrepreneurship in the form of self-employment and newly registered

[6] Our inequality estimates are somewhat different from the OECD's country analyses which report a very slight increase in Gini for market income from 2004 to 2011 (0.369 to 0.371). OECD subsequently change their definition of market income in 2012 to include a more detailed breakdown of household income transfers, as well as a revised definition of household income. Specifically, a main source of divergence between the OECD's estimates of income inequality and ours rely on their usage of equivalized income data (by the square root of household size), and constant prices, whereas we use individual nominal income data. In our analysis we also focus on workforce inequality whereas OECD analyses include those unemployment and outside the workforce. Taken together, the aggregate OECD statistics are fairly consistent with our slightly declining Gini for market income between 2005 and 2013 (0.318 to 0.309, see Table 2, below on page 23).

firms, making Sweden an interesting case to probe the role of entrepreneurship in income inequality.

Our paper relies on data from the LISA database, which includes all individuals residing in Sweden aged 16 and older. The LISA data comes from governmental registers and is maintained for research purposes by Statistics Sweden. The data contains a wealth of demographic and income-related information and is generated from a number of sources, including individual tax statements, birthplace registries, and school records. The database offers information on employment as well as industrial and occupational structures, and it tracks flows in the labor market. While income-related information dates back to 1990, the database did not include necessary entrepreneurial income and occupational data until more recently. In our analysis we therefore focus on two cross-sections, 2005 and 2013, the first and last years the LISA database included comprehensive data on the income and occupation variables used in our analysis. These two data points also represent a full economic cycle. Using the methodology described in the previous section, we can account for both the *level* of inequality for each of these years as well as any increase or decrease that occurred over the nine-year period.

We sample all individuals in the workforce between 25 and 65 years of age in the respective years for which labor market data is available. From this data, we exclude individuals that are not associated with an employing organization, such as sailors and seasonal workers. The sample used for analysis comprises 3,659,414 individuals in 2005 and 3,772,742 individuals in 2013. Except for age, the only criteria we use to systematically exclude individuals from the sample if they reported zero income on their income statement. We also exclude a handful of individuals at the very top of the income distribution (see discussion below).

This rich data further enables us to distinguish between two types of entrepreneurs; those individuals who are self-employed (SE) in a private business and those who are self-employed in an incorporated firm (ISE) (Blanchflower, 2000). Since we include all individuals with income statements above zero, our category for salaried workers needs to include individuals who were not only employed at the time of measurement (which is the month of November for each separate calendar year) but who also worked sometime during that same year.

Consistent with government classifications, we define an entrepreneur as an individual whose main source of income comes from a company in which he or she has a majority ownership stake and works full time (Folta et al., 2010).[7] With this classification, we use information from two different sources. First, we use information from government register data (LISA) regarding the sources from which an individual derives the largest share of his or her income.[8] Based on these data, we only include entrepreneurs reporting their own business as the source of their majority income. Part-time entrepreneurs whose income stems primarily from paid employment are coded as workers. Second, we use information from RAMS (Swedish labor force register data) on whether the entrepreneur considers him- or herself "active" in the sense that he or she works at least 600 hours a year in his or her own business. Unless entrepreneurs reports actively doing business in this manner, he or she is put into the worker category.

This rather strict definition of what constitutes entrepreneurship may reduce the count of people in both the SE and ISE categories, but it comes with the benefit that we can be fairly certain that running their respective business is these entrepreneurs' main occupation.

The primary income variable in our models is *market income*, defined as the sum of gross wage income + net income from an active business + capital income. All three variables are included in market income since while SE entrepreneurs receive 100% of their earnings in the form of net income from an active business, ISE entrepreneurs receive their earnings both as gross wage income (from their business) and capital income (Alstadsæter and Jacob, 2015; Edmark and Gordon, 2013).

Although our main focus is on market income, in a supplementary analysis, we also utilize detailed data on *disposable income* to account for entrepreneurships' effects on inequality, post taxes, and government transfers and as a robustness test to ensure that our results are not tainted by the potential problem of tax evasion among entrepreneurs (Engström and Holmlund, 2009). Disposable income is measured by Statistics Sweden by equalized disposable household income, potentially giving different members of the household different consumption weights. Each family member's personal disposable income is multiplied with an individual consumption weight (as calculated by Statistics Sweden) and then divided by the family's total consumption weight. Disposable income includes factor incomes, such as gross wages, business-related income (net deficit), and net capital profits, as well as taxable

[7] As of October 1, 2010, the minimum equity required when forming an incorporated business was lowered from 100,000 SEK to 50,000 SEK.
[8] In this classification, reported business income is weighted by a factor of 1.6 to compensate for the fact that business income compared to workers is lower in terms of the hours spent working.

(and non-taxable) transfers, such as rehabilitation compensation, pensions, and child allowances (e.g., housing benefits, social security, and study allowances).

In addition to the two entrepreneurship variables discussed above, other explanatory variables included in the regression are *age* and *age squared, job tenure* and *job tenure squared, job changes,* and *years of education* (e.g. Folta et al., 2010; Yamauchi, 2001; Åstebro et al., 2011). All individuals living in Sweden receive a personal identification number based on their date of birth. We use this information to calculate an individual's age (in years) as well as the squared term. Job tenure and job changes are computed from LISA and are defined as the number of years of experience at an individual's workplace and the number of workplace switches since 1990.[9] Years of education is the most common operationalization of general human capital in the entrepreneurship and inequality literatures (Arum and Müller, 2004; Cowell and Fiorio, 2011; Van Praag et al., 2013). Our variable is created from educational codes (available for all individuals in the LISA register) that provide information on the length of an individual's highest-attained education (commensurate with the International Standard Classification of Education [ISCED] 97). Further, we control for the number of *children living at home, marital status,* and *gender* (1 = male; 0 = female). In unreported models (available upon request), we consider additional controls, such as industry and regional dummies. Although adding controls for industry and regional differences help us account for the variation in wage, we do not include these dummies in the main results of our model since an important point of getting an education is to move to better paying industries and/or regions, not only to advance within the same industry or region.

Finally, we note that our analysis of income inequality is sensitive from the very top incomes in each year. We took this approach because these individuals severely skew the income distribution, which is in violation of OLS assumptions. Still, we do not want to exclude too many individuals with top incomes because they may reflect important differences between W, SE, and ISE. This approach is also motivated by the recent literature's interest in top income earners (see e.g. Atkinson et al., 2011; Quadrini, 1999; Roine and Waldenström, 2008). Thus, there is a tradeoff between dropping more high-income earners, which increases

[9] To account for potential bias arising from left censoring the *job tenure* variable, in unreported robustness tests, we replicate the results from the factor source regression analyses in Tables 5 and 6 with an additional dummy variable taking the value 1 for those individuals with the maximum years of job tenure (Wennberg et al., 2010). These results—available upon request—are consistent with the results reported here.

R^2, and limiting the sensitivity of the inequality measures to top incomes. We settle on a fairly generous restriction by keeping 99.9999% of the full sample, excluding 366 and 377 individuals with the highest market income in the years 2005 and 2013, respectively.

In the standard Mincer-type regression, the dependent variable is usually expressed in logarithmic form. However, in our context of regression decomposition, using log-income to generate proportional weights means that the decomposition of inequality refers to the inequality of *log-income* which is much less informative and not standard in the decomposition literature. Therefore, we opt to estimate our Mincer regressions with income expressed in absolute rather than logarithmic form.

4.2. Descriptive statistics

Table 1 shows the descriptive statistics for three sub-groups of the population—salaried workers (W), self-employed (SE), and incorporated self-employed (ISE)—for all variables in the years 2005 and 2013. Over the period, we see that the number of SE, as a share of the total workforce, decreases from 3.86% to 3.32%, whereas the share of ISE increases from 2.22% to 2.47%. Most variables display fairly moderate changes across the two time periods, which is to be expected when working with large sample sizes. However, there are a few noteworthy differences: the average share of married/cohabiting individuals decreases over the time period for all groups, and even though SE entrepreneurs exhibit somewhat lower levels of education compared to W and ISE entrepreneurs in both periods (a result often found in other studies, see e.g. Robinson and Sexton, 1994), average years of education for all groups increases.[10] Finally, with little change over time, both SE and ISE entrepreneurs are predominately men with a higher average age of around 48 and 47, respectively, compared to W who are considerably younger (44) on average.

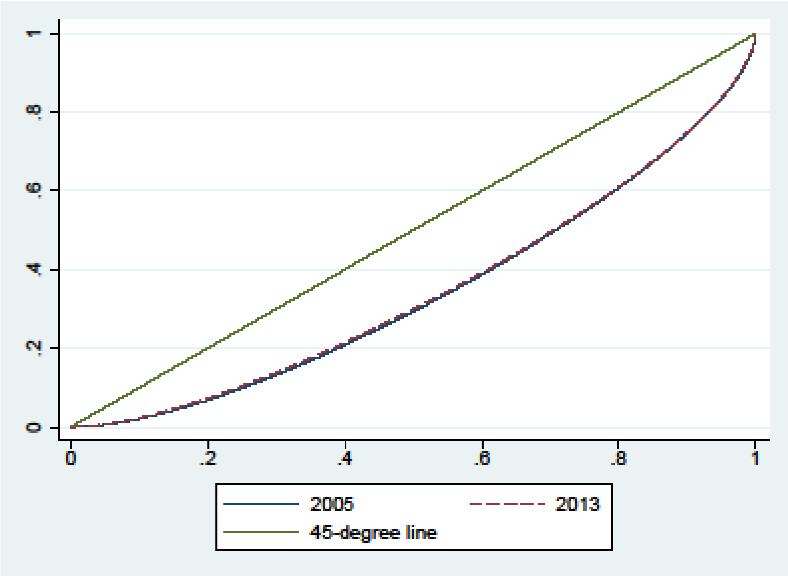
[10] The fact that average job tenure increases from 2005 to 2013 is mainly due to the fact that the tenure variable is left-censored in 1990, which means that the maximum tenure allowed in 2005 is 15 years compared to 23 years in 2013. The same holds for the variable for job changes, which was also calculated from 1990 onward.

Table 1. Descriptive statistics: Workers (W), Self-employed entrepreneurs (SE) and Incorporated self-employed entrepreneurs (ISE)

	2005				*2013*			
Variables by each occupational group:	*Mean*	*Sd.*	*Min*	*Max*	*Mean*	*Sd.*	*Min*	*Max*
Workers (W)								
Market income (in 100s Swedish krona)	2.453.165	2.113.554	0.909	1.14150.68	2.827.492	2.252.876	0.812	108.732.242
Age	43.996	10.875	25	64	44.08	10.944	25	64
Age squared (demeaned)	132.879	134.89	0.033	520.699	135.425	142.842	0.001	530.632
Job tenure	6.241	5.294	0	15	7.535	7.092	0	23
Job changes	1.938	1.693	0	14	2.735	2.176	0	20
Children living at home	0.976	1.095	0	12	1.008	1.085	0	14
Gender (1=Men)	0.507	0.5	0	1	0.503	0.5	0	1
Marital status (1=Married/cohabitant)	0.488	0.5	0	1	0.468	0.499	0	1
Years of education	12.332	2.312	9	20	12.728	2.312	9	20
Obs.	3.437.021 (93.92% of the workforce)				3.554.138 (94.21% of the workforce)			
Self-employed (SE)								
Market income (in 100s Swedish krona)	1.728.648	2.470.781	0.909	111.326.797	1.861.262	2.555.928	0.812	106.108.648
Age	47.918	10.42	25	64	48.155	10.531	25	64
Age squared (demeaned)	108.589	108.348	0.033	520.699	110.911	115.275	0.001	530.632
Job tenure	6.465	5.291	0	15	8.839	6.877	0	23
Job changes	1.75	1.6	0	12	2.489	2.076	0	18
Children living at home	0.987	1.141	0	12	0.986	1.114	0	11
Gender (1=Men)	0.686	0.464	0	1	0.66	0.474	0	1
Marital status (1=Married/cohabitant)	0.559	0.497	0	1	0.52	0.5	0	1
Years of education	11.393	2.06	9	20	11.755	2.107	9	20
Obs.	141.261 (3.86% of the workforce)				125.293 (3.32% of the workforce)			
Incorporated self-employed (ISE)								
Market income (in 100s Swedish krona)	3.480.648	4.159.77	0.909	113.410.617	3.909.648	3.860.34	0.812	107.112.789
Age	47.251	9.705	25	64	47.262	9.384	25	64
Age squared (demeaned)	94.515	97.33	0.033	520.699	88.656	100.552	0.001	530.632
Job tenure	8.261	5.068	0	15	8.927	7.059	0	23
Job changes	1.741	1.588	0	13	3.043	2.189	0	16
Children living at home	1.06	1.104	0	10	1.146	1.091	0	10
Gender (1=Men)	0.795	0.404	0	1	0.788	0.409	0	1
Marital status (1=Married/cohabitant)	0.618	0.486	0	1	0.582	0.493	0	1
Years of education	11.951	2.245	9	20	12.376	2.25	9	20
Obs.	81.132 (2.22% of the workforce)				93.311 (2.47% of the workforce)			

Turning our attention to the descriptive statistics of aggregate inequality, Figure 1 shows Lorentz curves for market income inequality in 2005 and 2013, and Table 2 displays a set of common inequality measures. The Lorentz curves show a slight inward shift between the two years, indicating that the overall dispersion of market income actually becomes less unequal over the period. The Gini coefficient decreased from 31.8 to 30.9, which corresponds to a 2.8% decrease (Table 2). In our supplementary analysis of disposable income, the Lorentz curve display a slight outward shift (see Figure A1 in Appendix 2).

Figure 1. Lorentz curve for workforce market income, 2005 and 2013



The slight inward shift and decrease in overall income inequality in Figure 1 represent the aggregate changes within occupational groups and at different parts of the income distribution. We attend to these changes in Table 2, which shows different income percentile ratios as well as the GE index and Gini index for the separate sub-groups of workers (W), self-employed (SE), and incorporated entrepreneurs (ISE) in the years 2005 and 2013.

Table 2 contains the equivalent measures of inequality in market income for workers (W) in Columns 1 and 4, for SE entrepreneurs in Columns 2 and 5, and for ISE entrepreneurs in Columns 3 and 6. Beginning with workers, we see a slight decrease in inequality over the period across almost all inequality measures. The decrease may be slight but is nevertheless present. The exceptions are for inequality as measured by percentile ratios p90/p50, which display a slight increase of 1%.

Table 2. Income inequality statistics of market income

	2005				*2013*			
	W	*SE*	*ISE*	*Total*	*W*	*SE*	*ISE*	*Total*
Inequality measure:	*(1)*	*(2)*	*(3)*	*(3.5)*	*(4)*	*(5)*	*(6)*	*(7)*
p90/p10	4.073	15.321	4.483	4.331	3.805	14.948	4.483	4.047
p90/p50	1.748	2.494	2.148	1.765	1.765	2.348	2.058	1.783
p50/ p10	2.33	6.143	2.086	2.451	2.156	6.367	2.178	2.268
p75/p25	1.771	3.716	1.891	1.813	1.719	3.583	1.976	1.754
GE(-1)	0.73	3.858	0.433	0.906	0.59	5.203	0.407	0.829
GE(0)	0.202	0.527	0.26	0.219	0.185	0.526	0.233	0.201
GE(1)	0.198	0.446	0.322	0.212	0.185	0.429	0.264	0.197
GE(2)	0.371	1.021	0.714	0.405	0.317	0.943	0.487	0.341
Gini				0.318				0.309

Note: All inequality measures are computed as raw figures for each of the subgroups, without weights that accounts for their contribution to the aggregate income inequality. The subgroups are created such that *W + SE + ISE = total workforce* (population), where the SE and ISE groups correspond to the number of individuals presented in Table 1, and W = total workforce – (SE + ISE).

Table 2 reveals similar developments for SE entrepreneurs as for W, with decreasing inequality registered for all measures except p50/p10 and GE(-1). This would suggest some form of contraction of top incomes, whereas bottom incomes become more dispersed. Capturing the low end of the income distribution, GE(-1) increased by a total of 34.9%.

As for SE, inequality among ISE entrepreneurs increase for p50/p10 as well as for p75/p25 by 4.5%. However, none of the GE indices reflect this increase.

5. RESULTS

5.1. Sub-group decomposition of the GE index

To further probe what lies behind the changes in levels of inequality presented in Table 2, we model outcomes in two stages. First, for our two cross-sections 2005 and 2013, we use the decomposability of the GE index to disaggregate inequality $GE(y;\alpha)$ into between-group and within-group parts for the various sub-groups (W, SE, and ISE), probing how these groups contribute to inequality at different parts of the income distribution at different points in time. Second, using the Mincer regression framework outlined above, we estimate the contribution of our explanatory variables to within and overall inequality for each separate year. Finally, we discuss how the impact of these explanatory variables changes over time.

As for the first step in this analysis, instead of reporting the contributions in terms of inequality points, we calculate the percentage contributions for each of the terms in $\widetilde{GE}(y_j;\alpha)$

(eq. 16), which facilitates interpretation. Dividing both sides of $\widetilde{GE}(y_j;\alpha)$ with aggregate inequality $GE(y;\alpha)$, we define $\Lambda\left(y_j;\alpha\right)$ as

$$\Lambda\left(y_j;\alpha\right) \equiv \frac{\widetilde{GE}\left(y_j;\alpha\right)}{GE(y;\alpha)} = \underbrace{\frac{p_j\left(r_j{}^{\alpha}-1\right)}{GE(y;\alpha)(\alpha^2-\alpha)}}_{\Lambda_b\left(y_j;\alpha\right)} + \underbrace{\frac{w_j GE\left(y_j;\alpha\right)}{GE(y;\alpha)}}_{\Lambda_w\left(y_j;\alpha\right)}. \qquad (17)$$

The term $\Lambda_b(y_j;\alpha)$ represents the percentage contribution to aggregate inequality from group j's between part, and $\Lambda_w(y_j;\alpha)$ represents the contribution from its within part. The total contribution from a specific group is given by $\Lambda(y_j;\alpha)$. These calculations are presented for market income in Table 3, corresponding to the rows labeled *Between*, *Within*, and *Total*. The table shows separate calculations for each of the inequality indices—*GE(-1), GE(0), GE(1),* and *GE(2)*—with index values for the different sub-groups presented in the first row. The results in Columns 1–4 are based on income data for the year 2005, and the results in Columns 5–8 are based on income data for the year 2013.

We also calculate the contributions to overall inequality when the between- and within-group parts are summed across groups. By taking the sum over all J groups, the equation takes the following identity:

$$\sum_{j=1}^{J} \Lambda\left(y_j;\alpha\right) = \sum_{j=1}^{J} \left[\Lambda_b\left(y_j;\alpha\right) + \Lambda_w\left(y_j;\alpha\right)\right] = 1. \qquad (18)$$

Based on the rules of decomposition, summing all contributions amounts to 1 (100%). In Table 3, this represents the sum of the entries in the row for *Total,* which is presented in the column for *Total*. The sums across the groups' between and within contributions are calculated using the terms $\sum_{j=1}^{J}\Lambda_b\left(y_j;\alpha\right)$ and $\sum_{j=1}^{J}\Lambda_w\left(y_j;\alpha\right)$, which are presented in the column *Total* as they sum the corresponding rows for *Between* and *Within.*

As a start, looking at GE(-1), the total contribution to inequality from W, SE, and ISE is 75.44%, 24.18%, and 0.38%, respectively. Hence, when emphasizing the bottom parts of the income distribution, the largest share of inequality comes from salaried workers (W). However, considering their relatively small size, a mere 3.86% in 2005 (Table 1), the contribution from SE entrepreneurs to total inequality is sizable. Further, using GE(-1), it is the within-inequality component of each group that accounts for almost all of the total inequality.

Table 3: Sub-group decomposition in percentage points of $GE(y;\alpha)$ for market income, 2005 and 2013

	2005				*2013*			
	W (1)	*SE* (2)	*ISE* (3)	*Total* (4)	*W* (5)	*SE* (6)	*ISE* (7)	*Total* (8)
GE(-1)	0.73	3.858	0.433	0.906	0.59	5.203	0.407	0.829
Between: $\Lambda_b(y_j;\alpha)$	-0.11	0.887	-0.363	0.414	-0.107	1.035	-0.415	0.512
Within: $\Lambda_w(y_j;\alpha)$	75.548	23.292	0.746	99.586	66.993	31.616	0.878	99.488
Total: $\Lambda(y_j;\alpha)$	75.438	24.179	0.383	100	66.886	32.651	0.463	100
GE(0)	0.202	0.527	0.26	0.219	0.185	0.526	0.233	0.201
Between: $\Lambda_b(y_j;\alpha)$	-0.908	6.131	-3.562	1.661	-0.883	6.877	-4.01	1.983
Within: $\Lambda_w(y_j;\alpha)$	86.419	9.284	2.636	98.339	86.472	8.682	2.862	98.017
Total: $\Lambda(y_j;\alpha)$	85.511	15.416	-0.927	100	85.589	15.559	-1.148	100
GE(1)	0.198	0.446	0.322	0.212	0.185	0.429	0.264	0.197
Between: $\Lambda_b(y_j;\alpha)$	0.939	-4.468	5.227	1.698	0.903	-4.63	5.672	1.945
Within: $\Lambda_w(y_j;\alpha)$	87.791	5.726	4.786	98.302	88.7	4.768	4.586	98.055
Total: $\Lambda(y_j;\alpha)$	88.729	1.258	10.012	100	89.604	0.138	10.258	100
GE(2)	0.371	1.021	0.714	0.405	0.317	0.943	0.487	0.341
Between: $\Lambda_b(y_j;\alpha)$	0.492	-2.387	2.794	0.898	0.522	-2.754	3.336	1.105
Within: $\Lambda_w(y_j;\alpha)$	86.355	4.85	7.896	99.102	88.106	3.998	6.792	98.895
Total: $\Lambda(y_j;\alpha)$	86.847	2.463	10.69	100	88.628	1.244	10.128	100

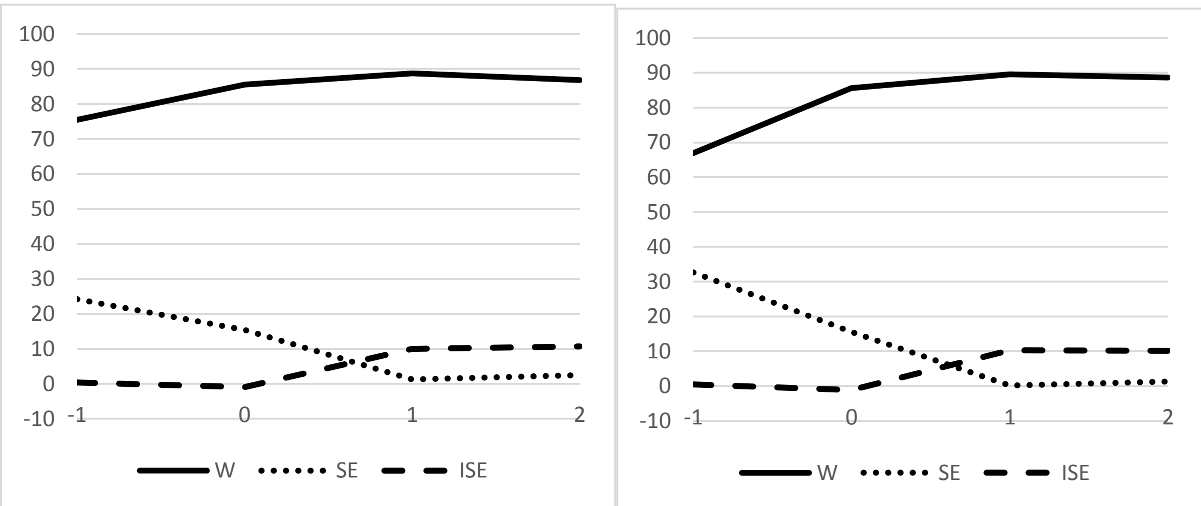
Note: The table shows the percentage contribution (1=100%) of the between- and within-inequality component from the sub-groups workers (W), self-employed (SE), and incorporated entrepreneurs (ISE). Separate contributions are calculated for the various GE-indices with $\alpha = \{-1,0,1,2\}$. The inequality levels for each of the subgroups are calculated using eq. (16) with the appropriate weights. These inequality levels differs from those presented in Table 2 that comprise raw calculations based on equation (1) applied to the restricted sample.

Turning the sensitivity parameter gradually to higher levels of the income distribution (from an α of -1 to 0, and 1 to 2), we see that the total contributions from entrepreneurs as a group (i.e., SE and ISE) creates a U-shaped relationship. Beginning at 24.562 for GE(-1), entrepreneurs' contribution to inequality shrinks to 14.849% for GE(0) and 11.27% for GE(1) but increases again for GE(2) to 13.153%. As we increase the value of α, there are two other interesting tendencies that stand out. The first is that the between-inequality component plays a more prominent role for GE(0) and GE(1) in the middle ranges of the distribution. The second notable tendency is a reversal between SE and ISE in terms of their respective total contribution to inequality. ISE accounts for an increasingly larger share of inequality, whereas SE accounts for a decreasing share, except for in the GE(2) measure. For the GE(2) measure, the total percent contribution to inequality from W, SE, and ISE is to 86.85%, 2.46% and 10.69%, respectively.

The same patterns are even more pronounced when we perform an equivalent decomposition analysis for market income in 2013. That is, the combined effect of entrepreneurship is U-shaped, going from small to larger values of α. A significant part of the contribution stems from SE when we focus on the bottom of the income distribution. Comparing the two time periods, the relative contribution of entrepreneurship (SE + ISE) to total inequality, as measured at the bottom of the income distribution $(\Lambda(y; -1)$, increases substantially by 8.55%. Further, this change is almost completely driven by changes in the within-inequality component for SE entrepreneurs.

We can illustrate the above findings, including the income dynamics for each of the W, SE, and ISE groups, by plotting the total percentage contribution to aggregate income inequality for increasing values of α. This plot is shown in Figure 2 (showing market income inequality, 2005 and 2013). In the figure, the contribution from W to overall income inequality is represented by the solid line, SE is represented by the dotted, and ISE is represented by the dashed line. The figure, for both years, illustrates how a downward shift in the inequality contribution for W as measured by GE(-1) corresponds to an upward shift for the SE group while the contribution from ISE to inequality at low income levels remains largely unchanged.

Figure 2. Proportional contribution to overall inequality in market income from workers, self-employed entrepreneurs, and incorporated entrepreneurs



(i) Market income 2005 (ii) Market income 2013

5.2. Regression estimates from factor sources of income inequality, 2005 and 2013

The final step in our analysis concerns estimating an income equation based on the same sample and sub-groups as in our decomposition analysis above. Our regression model is based on equation (12), where y_{it} = *market income*. We run separate regressions for each of the three sub-groups ($j = 1, \dots, 3$) of W ($j = 1$), SE ($j = 2$), and IES ($j = 3$). To shed light on recent trends, the model is estimated for the two cross-sections presented in Tables 6 and 7 for 2005 and 2013, respectively. We use the type of OLS model specified by Fiorio and Cowell (2011), and both tables show individual-level coefficients (in Columns 1–3) as well as each coefficient's proportional contribution to income inequality s_{jk} (Columns 4–6).

We should note that the proportional contributions to inequality for our explanatory variables refer to the combined "price" and "quantity" effect of each variable. Using years of education as an example, this means that we do not separate between the price effect corresponding to the estimated coefficient, which captures the effect of an additional year of education in each separate group, and the quantity effect—namely, the effect of the average level of education among individuals in each respective group.[11]

First, in decomposing inequality to its various individual-level components $\hat{b}_{kj}\mu(x_{kj})$, it is important to remember that the regression approach is limited by the amount of variance in income explained by the regression run for each sub-group. Thus, for a given sub-group j, adding up the contributions of all explanatory variables (s_{jk}) amounts to the R^2 of each respective regression. The unexplained part of the model $(1 - R^2)$ equals the proportional contribution of the residual (s_{jK}) computed for the residual in equation (14). These estimates are provided in the bottom row of each table and show how much of the variance in income inequality that cannot be attributed to the factor sources that we include in our analysis.

Also, for further clarification, s_{jk} represents the raw contribution to inequality computed for a particular sub-group. Since each group varies in size, this means that s_{jk} needs to be reweighted to assess the contribution to aggregate inequality $\Lambda(y_j; \alpha)$.[12] Focusing on the

[11] Since we cannot ensure that our explanatory variables are strictly exogenous, interpretation of outcomes should also be done with some measure of caution.

[12] In terms of percentage points, we showed in Table 3 that within-group inequality contributed 98.3% and 98.1% of aggregate inequality for $\alpha = 1$ in 2005 and 2013, respectively. For other values of α, the within-group share is even higher.

within-group inequality component shown above to account for most of the aggregate inequality, s_{kj} is then scaled by

$$w_j \frac{GE(y_j;0)}{GE(y;0)}, \tag{19}$$

where w_j is the weight function defined in connection to expression (3). The different weight schemes computed for $GE(y_j;\alpha)$ with sensitivity parameters $\alpha = \{-1,0,1,2\}$ are presented in Table 4 below.

Table 4: Inequality weights (w_j) for the GE index

	2005			*2013*		
GE-index	*W*	*SE*	*ISE*	*W*	*SE*	*ISE*
GE(-1)	0.937	0.055	0.016	0.94	0.05	0.018
GE(0)	0.939	0.039	0.022	0.942	0.033	0.025
GE(1)	0.941	0.027	0.032	0.944	0.022	0.034
GE(2)	0.943	0.019	0.045	0.946	0.014	0.047

The aggregate inequality variance s_{kj} applies to all inequality indices, not solely to the various GE indices, provided the basic assumptions in Shorrocks (1982) are satisfied (see Appendix 1). In order to investigate the contribution to aggregate inequality of a particular explanatory variable in a specific sub-group, one needs to specify a given index in order to define the scaling function.

Beginning with inequality in the year 2005, Table 5 shows separate regression models of the main contributors to within-group inequality for the three groups: W, SE, and ISE. Columns 1–3 display Mincer-type estimates (OLS), regressing our explanatory variables on individuals' market income, whereas Columns 4–6 show the effects of the same explanatory variables for overall income inequality in each specific group. Almost all variables are statistically significant in the regressions, which is not surprising given the large number of observations.

For most sub-groups, three explanatory variables stand out as the major contributors to inequality. These are years of education, gender (1 = Male), and age. The variable years of education shows a marginal effect in the individual-level estimates of 22,533 SEK higher yearly income for each additional year of education.[13] Among SE and ISE entrepreneurs, each

[13] 1 Swedish krona= roughly 0.11US$ and 0.1€, August 2016.

additional year translates into a higher income of 10,986 and 29,972 SEK, respectively. Although ISE showed the largest marginal effect of education, inspecting the proportional contributions of education for overall income inequality in each of the groups (Columns 4–6), we observe another pattern: education contributes the most to inequality among W with 5.07%, whereas it contributes only 0.59% for SE entrepreneurs and 2.42% for ISE entrepreneurs. As much as these results inform us on the contribution to the inequality of $GE_w(y_j, \alpha)$—that is, to the within-group inequality of each sub-group—their estimated contribution to aggregate inequality—that is, to $GE(y, \alpha)$—still need to be computed.

Table 5: Regression results for market income in 2005 and percentage contribution to $GE(y_j; \alpha)$

	Regression estimates			$100 \times s_{kj}$		
	W	*SE*	*ISE*	*W*	*SE*	*ISE*
Variables	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Age	15.951***	5.151***	20.624***	0.953	0.092	0.283
	(0.163)	(0.917)	(2.064)	(0.000)	(0.003)	(0.004)
Age square	-1.277***	-0.812***	-1.501***	0.994	0.18	0.189
	(0.010)	(0.063)	(0.157)	(0.000)	(0.001)	(0.001)
Job tenure	24.227***	27.570***	18.197***	0.312	0.332	0.048
	(0.297)	(1.866)	(4.471)	(0.000)	(0.001)	(0.002)
Job changes	70.268***	46.425***	7.118	0.169	0.021	-0.002
	(0.886)	(5.143)	(12.530)	(0.000)	(0.001)	(0.001)
No. children	-36.747***	15.911***	23.893	0.016	0.001	0.007
	(1.172)	(5.757)	(16.823)	(0.000)	(0.001)	(0.001)
Gender	972.655***	555.585***	889.601***	4.809	1.01	0.61
	(2.245)	(14.071)	(31.585)	(0.000)	(0.001)	(0.002)
Marital Status	149.939***	76.653***	119.245***	0.277	0.019	0.053
	(2.252)	(14.823)	(33.620)	(0.000)	(0.001)	(0.001)
Years of education	225.328***	109.860***	299.721***	5.074	0.593	2.417
	(0.606)	(3.983)	(7.722)	(0.000)	(0.004)	(0.004)
Immigrant	-393.522***	-606.254***	-900.441***	0.178	0.409	0.083
	(3.851)	(16.601)	(65.419)	(0.000)	(0.000)	(0.000)
Constant	-1660.521***	-338.908***	-1888.220***			
	(13.269)	(71.781)	(148.455)			
Obs.	3437021	141261	81132			
R-sq.	0.128	0.027	0.037			
Res. (=1-R-sq)				87.218	97.344	96.311

NOTE: Standard errors in parenthesis with significance levels *** p<0.01, ** p<0.05, * p<0.1. Columns (1) to (3) gives the robust OLS results from each of the groups W, SE and ISE, run separately. Columns (4) to (6) show the proportional contributions of the corresponding explanatory variable to the total within-group inequality for W, SE and ISE respectively. Because of the large sample sizes, almost all percentage contributions are strongly significant. For this reason we display their standard errors without asterisks denoting level of significance.

To gauge the contribution of each explanatory variable for overall income inequality, we compute their estimated contribution based on the Theil index of $G(y, 1)$ using sub-group and year-specific Theil index values and weights from Tables 3 and 4, respectively. In 2005, the scaling factor amounts to 0.878 for W ($0.941 \times 0.198/0.212$), 0.056 for SE ($0.027 \times 0.443/0.212$), and 0.049 for ISE ($0.032 \times 0.322/0.212$). Once multiplied with the weight, education among W accounts for 4.46% of overall within inequality ($\Lambda_w(y;1)$), 0.03% for SE, and 0.12% for ISE.[14] Although, education among SE and ISE accounts for a sizable portion of these groups' within inequality, the contribution to aggregate inequality stemming from differences in education level naturally drops because the two groups are small in size compared to W. What is more, to know the total contribution of an explanatory variable to aggregate inequality, we can simply add the weighted contributions of that variable from all groups. For years of education, this amounts to $4.61\%(= 4.459\% + 0.034\% + 0.117\%)$ (out of 100%).[15]

Turning to the gender variable, which takes the value 1 if the individual is male and 0 otherwise, we find the largest marginal effect for W. Among W, it appears that men earn 97,266 SEK more on average than women.[16] The gender income difference is the smallest for SE of 55,585 SEK, and it falls in between for ISE with a difference of 88,960 SEK. The proportional contribution of gender to within-group inequality is even more pronounced for W, with 4.81% for W compared to 1.01% for SE and 0.61% for ISE. Rescaled as contributions to aggregate within inequality $\Lambda_w(y;1)$, the gender composition of each group contributes 4.23% from W, 0.06% from SE, and 0.03% from ISE. We can also add these last percentage-point contributions to get the total workforce inequality contribution attributable to gender differences in income. Doing this, we find that gender accounts for 4.31% of total aggregate (within) income inequality.

Table 5 also shows that age is positively associated with income for all groups, which, according to the age squared estimate, is decreasing for W and ISE. For SE, on the other hand, the effect from age is increasing as the entrepreneurs grow older. Using the scaling factor, we

---

[14] We use $\Lambda_w(y;1)$ instead of $\Lambda(y;1)$ because we focus on the decomposition of within-group inequality. Strictly speaking, since $b_{kj}\mu(x_{kj})$ is also contained in $\Lambda_b(y;1)$, the decomposition results in the regression analysis only account for the contribution of $b_{kj}\mu(x_{kj})$ to $\Lambda(y;1)$ via $\Lambda_w(y;1)$. Given the comparatively small contribution of $\Lambda_b(y;1)$ to $\Lambda(y;1)$, however, the impact from any particular $b_{kj}\mu(x_{kj})$ on $\Lambda_b(y;1)$ is almost negligible and therefore left out when presenting the results.

[15] The contribution from education may seem small in magnitude and is about half the size as the result in Fiorio and Cowell (2011), who estimate the contribution from college education among Finnish males (females) to account for 3.1% (3.8%) of within gender-group inequality in 2004.

[16] 1 SEK = roughly 0.11 USD and 0.1 euro as of August 2016. The large gender income gap may be partly attributed to women more often working part time (Angelov et al., 2013).

see that the total contribution of age to income inequality in each group amounts to 0.84% for W, 0.01% for SE, and 0.01% for ISE. By adding the scaled contributions, we get a total contribution from age to $\Lambda_w(y;1)$ of 0.86%.

An important reason for why we find larger contributions to overall inequality from W can be seen from the R^2 value, which shows that the empirical model explains 12.8% of the variance in income for W. For SE and ISE, however, R^2 amounts to 2.7% and 3.7%, respectively. This means that the residual contribution to within group inequality, as captured by the error term in the model, contains most of the variation that accounts for inequality and more so for SE and ISE than for W. Earnings regressions commonly produce lower R^2 for entrepreneurs than for wage workers (Åstebro and Chen, 2014). For example, Hamilton (2000) finds that the highest R^2 for entrepreneurs is 0.07, which is much higher than what we report in Table 5. Income among entrepreneurs is known to be affected more by unobserved ability than income among workers (Åstebro et al., 2011)

The contribution of the error term for overall income inequality within groups is presented in the bottom row in Columns 4–6 and accounts for 87.22%, 97.34%, and 96.31% for inequality for W, SE, and ISE, respectively. Since the residual enters the model just as any other explanatory variable, we can also compute its total weighted contribution to aggregate within inequality, which amounts to 86.863 (i.e., the contribution to $\Lambda_w(y;1)$ emanating from $u_{Wi} + u_{SEi} + u_{ISEi}$ in expression (12)). Thus, adding up all weighted contributions, including the residual, amounts to the aggregate within inequality term $\Lambda_w(y;1) = 98.302$.

By repeating the empirical analysis for a later period, we are able to investigate changes in the determinants of inequality between the two time periods. Table 6 presents identical models to those of Table 5 but for the 2013 sample. To begin with, we observe that years of education, gender, and age all still play a significant role in accounting for inequality in 2013. Continuing with the Theil index, the scaling factors for 2013 amount to 0.886 for W ($0.944 \times 0.185/0.197$), 0.051 for SE ($0.022 \times 0.429/0.185$), and 0.049 for ISE ($0.034 \times 0.264/0.185$). Our subsequent analysis and discussion focus on the scaled contribution to aggregate income inequality when reporting the differences for 2013 compared to 2005.

Table 6. Regression results for market income 2013 and contribution to $GE(\mathrm{y_j};\alpha)$

	Regression estimates			$100 \times s_{kj}$		
	W	*SE*	*ISE*	*W*	*SE*	*ISE*

Variables	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Age	21.309***	1.904*	12.559***	1.544	0.035	0.208
	(0.174)	(1.012)	(1.686)	(0.00)	(0.006)	(0.004)
Age square	-1.244***	-0.757***	-2.184***	1.258	0.186	0.433
	(0.011)	(0.068)	(0.129)	(0.00)	(0.001)	(0.001)
Job tenure	21.283***	25.491***	44.617***	0.51	0.442	0.566
	(0.232)	(1.616)	(2.577)	(0.00)	(0.002)	(0.001)
Job changes	63.239***	44.848***	13.706*	0.403	0.091	-0.009
	(0.699)	(4.451)	(7.188)	(0.00)	(0.002)	(0.001)
No. children	-12.198***	33.609***	62.088***	-0.009	0.014	0.056
	(1.241)	(6.985)	(13.552)	(0.00)	(0.001)	(0.001)
Gender	960.778***	462.213***	835.323***	3.809	0.707	0.591
	(2.396)	(15.190)	(28.301)	(0.00)	(0.002)	(0.001)
Marital Status	168.084***	78.253***	201.475***	0.349	0.022	0.146
	(2.400)	(16.044)	(26.231)	(0.00)	(0.001)	(0.001)
Years of education	240.263***	69.692***	307.739***	4.886	0.178	2.693
	(0.603)	(4.041)	(6.298)	(0.00)	(0.005)	(0.003)
Immigrant	-396.395***	-463.896***	-592.389***	0.298	0.354	0.095
	(3.380)	(16.911)	(59.261)	(0.00)	(0.00)	(0.00)
Constant	-1858.759***	365.506***	-1566.253***			
	(13.961)	(72.788)	(126.334)			
Obs.	3,554,138	125,293	93,311			
R-sq.	0.13	0.02	0.048			
Res. (=100-R^2*100)				86.953	97.971	95.222

NOTE: Standard errors in parenthesis with significance levels *** p<0.01, ** p<0.05, * p<0.1. Columns (1) to (3) gives the robust OLS results from each of the groups W, SE and ISE, run separately. Columns (4) to (6) show the proportional contributions of the corresponding explanatory variable to the total within-group inequality for W, SE and ISE respectively. Because of the large sample sizes, almost all percentage contributions are strongly significant. For this reason we display their standard errors without asterisks denoting level of significance.

Looking first at the role of education for aggregate inequality in Table 6, we see that for W, it accounts for 4.33% of aggregate inequality compared to SE and ISE, for which education contributes with very little to aggregate inequality: 0.01% and 0.13%, respectively. These results show that returns to formal education among entrepreneurs in Sweden are comparatively low (Van Praag et al., 2013). Considering the effect from education across all groups in 2013, education is found to account for 4.47% of aggregate workforce inequality. Since education accounts for 4.61% in 2005, the results for 2013 suggest that education as a factor explaining the development of income inequality decreases slightly. This declining trend can be found among W and SE. For ISE entrepreneurs, however, the importance of education for income inequality is lower in 2013 than 2005.

A similar pattern is observed with regard to the contribution of age and gender to income inequality. The weighted percentage contribution from gender in 2013 is 3.38% for W, 0.04% for SE, and 0.03% for ISE, which adds up to 3.44% in total. For age, we have 1.37% for W, 0.002% for SE, and 0.01% for ISE, which together amounts to 1.38%. Whereas the total contribution to workforce income inequality from gender differences in wages decreases from 4.41% in 2005 to 3.44% in 2013, the contribution from the workforce age distribution increases from 0.86% in 2005 to 1.38% in 2013.

One explanation for this particular development is hinted at in the bottom row of Table 6. For each of the three occupational groups, we see that the part of the variance in income that is not explained by the model decreases for SE and increases for W and ISE, which directly impacts the size of the estimates for the partial contributions s_{kj}.

6. Summary and Discussion.

This paper outlines an approach that seeks to problematize and probe the ways in which entrepreneurship may contribute to income inequality. Using recently developed regression-based decomposition models and microdata for the total workforce in Sweden 2005 and 2013, we gauge inequality in three workforce groups: workers (W), self-employed (SE) and incorporated self-employed (ISE). By estimating inequality both within and between each of these sub-groups, our model provides a clear picture of the group dynamics that drive inequality at the workforce level. By tuning the entropy-based inequality indices to different segments of the income distribution we are able to assess at which income level our two categories of entrepreneurs have the most impact. At a second stage of the analysis, our regression based decomposition enables us to pinpoint the significance of each individual level explanatory factor and their contribution to both subgroup- and overall inequality.

Starting with overall inequality, our data show that for the 2005–2013 period inequality development has been quite stable, and that its direction of change differs to some extent depending on which income measure we use. On the one hand, using market income (income measured before taxes and transfers) we see a very slight decrease in inequality. On the other hand, using disposable income (i.e., income after taxes and transfers) we instead see a very moderate increase. These developments are almost exclusively related to changes in the middle or the bottom end of the income distribution (i.e., using measures like the Gini coefficient and GE(-1,0)), which suggests that the income distribution is augmented by more people with lower levels of disposable incomes.

As to the specific roles of entrepreneurship for workforce inequality, our main conclusions are:

- The two types of entrepreneurship, self-employed (SE) and incorporated self-employed (ISE), together have a distinct effect on total inequality in both 2005 and 2013. By using an inequality measure that emphasizes the bottom-half of the income distribution, we find that entrepreneurs account for around 30 % of inequality, and when using inequality measures that emphasize the top-half of the income distribution their combined contribution is around 10 %. These findings suggest that entrepreneurs (SE and ISE) do in fact increase income inequality, by disproportionately affecting income at the bottom and the top end of the income distribution.
- In the occupational sub-group decomposition analysis, where we assess the contribution of SE and ISE separately, we discover a distinct pattern suggesting that SE accounts for most of the 30 % observed for entrepreneurs at inequality for bottom-level incomes, and that ISE accounts for most of the 10 % observed for entrepreneurs at inequality for top-level incomes. Hence, our results point towards a polarizing effect. Although entrepreneurs as a group thus seem to play a decisive role, the mechanism by which entrepreneurship affects workforce inequality differs depending on whether we look at SE or ISE.
- While total change in income inequality between 2005 and 2013 is moderate, our data show that an increase in the numbers of self-employed (SE) over the period contributes positively to the increase in explained bottom-end inequality in 2013. This type of entrepreneurship is thus a distinctive factor in explaining the spread and variance in bottom-end income over the period. However, none of these changes translate into changes in total workforce inequality.
- The aggregate effects of both types of entrepreneurship for overall workforce income inequality are similar in magnitude to more conventional factors, such as relative educational group size, suggesting that although entrepreneurship may not represent an exclusive explanation for changing inequality in contemporary economies, it is a

factor that should be explored along with others more common explanatory variables in the inequality literature.

- This conclusion is further strengthened by the fact that, in our mincer-type regression estimates, education explains very little of within-group variation for both our groups of entrepreneurs, much less than for employed workers. The share of entrepreneurs within the workforce thus captures variation in inequality that only to a limited extent can be substituted for an ordinary explanatory variable such as level of education or share of higher educated. This finding also corroborates and extends results in studies from other countries that returns to formal education among entrepreneurs are low. Success in economic terms, or for that matter failure, seems to be determined by other factors.
- A limitation of these results are that they rely on non-logarithmic Mincer income regression to link determinants of income at the individual level to overall inequality, and are thus sensitive to incomes at the absolute top end of the income distribution (p. 19). Very high incomes – and the resulting inequality – are thus not explained by the same factors used in this model (Bihagen et al., 2013).

Our model and results come with some limitations that of course can and should be probed in future research efforts. Specifically, after the abolition of wealth taxation in Sweden in 2005, the microdata do not include measures of wealth. As prior studies suggest that entrepreneurship is a key element in understanding wealth concentration (Buera, 2009; Quadrini, 1999), future research could for example expand on our model by examining wealth inequality instead of market- and disposable income inequality. Finally, since the decomposition technique that we utilize does not easily lend itself to direct causal interpretation, and entrepreneurship is possibly endogenous to some extent (inequality can also further entrepreneurship), future research could benefit from using regulatory changes or other quasi-experimental settings to gauge the causal relationship between rates of entrepreneurship and levels of income inequality (Kerr and Nanda, 2009).

References

Angelov, N., P. Johansson, E. Lindahl. 2013. Is the persistent gender gap in income and wages due to unequal family responsibilities? IFAU Working paper 2013:3. Institute for Evaluation of Labour Market and Education Policy (IFAU).

Akita, T., Lukman, R.A., Yamada, Y., 1999. Inequality in the distribution of household expenditures in Indonesia: A Theil decomposition analysis. The Developing Economies, 37(2), 197-221.

Alstadsæter, A., Jacob, M., 2015. Dividend taxes and income shifting. Scandinavian Journal of Economics.

Arum, R., Müller, W., 2004, The reemergence of self-employment: a comparative study of self-employment dynamics and social inequality. Princeton Univ Pr.

Atkinson, A.B., 2003. Income inequality in OECD countries: Data and explanations. CESifo Economic Studies, 49(4), 479-513.

Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. Journal of economic literature, 49(1), 3-71.

Audretsch, D.B., 2009. The entrepreneurial society. The Journal of Technology Transfer, 34(3), 245-254.

Autor, D.H., 2014. Skills, education, and the rise of earnings inequality among the" other 99 percent". Science, 344(6186), 843-851.

Bigotta, M., Krishnakumar, J., Rani, U., 2015. Further results on the regression-based approach to inequality decomposition with evidence from India. Empirical Economics, 48(3), 1233-1266.

Bihagen, E., Nermo, M., Stern, C., 2013. Class origin and elite position of men in business firms in Sweden, 1993–2007: The importance of education, cognitive ability, and personality. European Sociological Review, 29(5), 939-954.

Blanchflower, D.G., 2000. Self-employment in OECD countries. Labour Economics, 7(5), 471-505.

Blau, D.M., 1987. A time-series analysis of self-employment in the United States. Journal of Political Economy, 95(3), 445-467.

Borjas, G.J., Bronars, S.G., 1989. Consumer Discrimination and Self-Employment. Journal of Political Economy, 97(3), 581-605.

Buera, F.J., 2009. A dynamic model of entrepreneurship with borrowing constraints: theory and evidence. Annals of finance, 5(3-4), 443-464.

Carnahan, S., Agarwal, R., Campbell, B.A., 2012. Heterogeneity in turnover: The effect of relative compensation dispersion of firms on the mobility and entrepreneurship of extreme performers. Strategic Management Journal, 33(12), 1411-1430.

Cowell, F., 2011, Measuring inequality. Oxford University Press.

Cowell, F., Fiorio, C., 2011. Inequality decompositions—a reconciliation. The Journal of Economic Inequality, 9(4), 509-528.

Cowell, F.A., 2000, Measurement of inequality. In: A.B. Atkinson, F. Bourguignon (Eds.). Handbook of Income Distribution Elsevier, North-Holland, pp. 87-166.

Creedy, J., Hérault, N., 2011. Decomposing inequality and social welfare changes: the use of alternative welfare metrics.

Davis, G.F., 2013. After the Corporation. Politics & Society, 41(2), 283-308.

Edmark, K., Gordon, R.H., 2013. The choice of organizational form by closely-held firms in Sweden: tax versus non-tax determinants. Industrial and Corporate Change, 22(1), 219-243.

Engström, P., Holmlund, B., 2009. Tax evasion and self-employment in a high-tax country: evidence from Sweden. Applied Economics, 41(19), 2419-2430.

Evans, D.S., Leighton, L.S., 1989. Some Empirical Aspects Of Entrepreneurship. The American economic review, 79(3), 519-535.

Fields, G.S., 2003, Accounting for income inequality and its change: A new method, with application to the distribution of earnings in the United States. Solomon W. Polachek (ed.) Worker Well-Being and Public Policy (Research in Labor Economics) Emerald Group, pp. 1-38.

Fields, G.S., Yoo, G., 2000. Falling labor income inequality in Korea's economic growth: Patterns and underlying causes. Review of Income and Wealth, 46(2), 139-159.

Folta, T.B., Delmar, F., Wennberg, K., 2010. Hybrid Entrepreneurship. Management Science, 56(2), 253-269.

Goldthorpe, J.H., 2010. Analysing social inequality: a critique of two recent contributions from economics and epidemiology. European Sociological Review, 26(6), 731-744.

Hamilton, B.H., 2000. Does entrepreneurship pay? An empirical analysis of the returns of self-employment. Journal of Political Economy, 108(3), 604-631.

Kacperczyk, A.J., Balachandran, C., 2016. Vertical and horizontal comparisons and mobility outcomes: Evidence from the Swedish micro-data MIT Sloan School of Management.

Kerr, W., Nanda, R., 2009. Financing constraints and entrepreneurship, National Bureau of Economic Research.

Levine, R., Rubinstein, Y., 2016. Smart and illicit: Who becomes an entrepreneur and do they earn more? Quarterly journal of economics.

Lin, Z., Picot, G., Yates, J., 2000. The entry and exit dynamics of self-employment in Canada. Small Business Economics, 15(2), 105-125.

Lippmann, S., Davis, A., Aldrich, H.E., 2005. Entrepreneurship and inequality. Research in the Sociology of Work, 15, 3-31.

Moskowitz, T.J., Vissing-Jorgensen, A., 2002. The returns to entrepreneurial investment: A private equity premium puzzle?, National Bureau of Economic Research.

OECD, 2015. OECD Income inequality data update: Sweden. http://www.oecd.org/sweden/OECD-Income-Inequality-Sweden.pdf.

Quadrini, V., 1999. The importance of entrepreneurship for wealth concentration and mobility. Review of Income and Wealth, 45(1), 1-19.

Quintano, C., Castellano, R., Regoli, A., 2005, The Contribution of Self-Employment to Income Inequality. A Decomposition Analysis of Inequality Measures by Sources and Subgroups for Italy, 1998–2002. International Conference in Memory of Two Eminent Social Scientists: C. Gini and M. O. Lorenz., Siena, Italy.

Reynolds, P., Bosma, N., Autio, E., Hunt, S., De Bono, N., Servais, I., . . . Chin, N. 2005. Global entrepreneurship monitor: Data collection design and implementation 1998–2003. Small Business Economics, 24(3), 205-231.

Robinson, P.B., Sexton, E.A., 1994. The effect of education and experience on self-employment success. Journal of Business Venturing, 9(2), 141-156.

Roine, J., Waldenström, D., 2008. The evolution of top incomes in an egalitarian society: Sweden, 1903–2004. Journal of Public Economics, 92(1), 366-387.

Sanandaji, T., Leeson, P.T., 2013. Billionaires. Industrial and Corporate Change, 22(1), 313-337.

Shane, S., 2009. Why encouraging more people to become entrepreneurs is bad public policy. Small Business Economics, 33(2), 141-149.

Shorrocks, A.F., 1982. Inequality decomposition by factor components. Econometrica, 50(1), 193-211.

Stam, E., 2013. Knowledge and entrepreneurial employees: a country-level analysis. Small Business Economics, 41(4), 887-898.

Steinmetz, G., Wright, E.O., 1989. The fall and rise of the petty bourgeoisie: changing patterns of self-employment in the postwar United States. American journal of sociology, 973-1018.

Sørensen, J.B., Sharkey, A.J., 2014. Entrepreneurship as a Mobility Process. American Sociological Review, 79(2), 328-349.

Thewissen, S., Wang, C., Van Vliet, O., 2013. Sectoral Trends in Earnings Inequality and Employment: International Trade, Skill-Biased Technological Change, or Labour Market Institutions?

Van der Sluis, J., Van Praag, M., Vijverberg, W., 2005. Entrepreneurship selection and performance: A meta-analysis of the impact of education in developing economies. The World Bank Economic Review, 19(2), 225-261.

Van Praag, C.M., Versloot, P.H., 2007. What is the value of entrepreneurship? A review of recent research. Small Business Economics, 29(4), 351-382.

Van Praag, M., van Witteloostuijn, A., van der Sluis, J., 2013. The higher returns to formal education for entrepreneurs versus employees. Small Business Economics, 40(2), 375-396.

Wennberg, K., Wiklund, J., DeTienne, D., Cardon, M. 2010. Reconceptualizing entrepreneurial exit: Divergent exit routes and their drivers. Journal of Business Venturing, 25(4), 361-375.

Wright, M., Zahra, S., 2011. The other side of paradise: Examining the dark side of entrepreneurship. Entrepreneurship Research Journal, 1(3).

Yamauchi, F., 2001. Does inequality of labor earnings emerge in young days or later?: Labor earnings dynamics and learning about individual ability in heterogeneous society. Journal of Economic Behavior & Organization, 44(4), 413-434.

Åstebro, T., Chen, J., 2014. The entrepreneurial earnings puzzle: Mismeasurement or real? Journal of Business Venturing, 29(1), 88-105.

Åstebro, T., Chen, J., Thompson, P., 2011. Stars and misfits: Self-employment and labor market frictions. Management Science, 57(11), 1999-2017.

Åstebro, T., Tåg, J., 2015. Jobs Incorporated: Incorporation Status and Job Creation. IFN Working paper 1059, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2576044.

Özcan, B., 2011. Only the lonely? The influence of the spouse on the transition to self-employment. Small Business Economics, 37(4), 465-492.

Appendix 1: Shorrocks' assumptions for income decomposition

Shorrocks' (1982) six assumptions for computing inequality measures have long been central in the literature (Bigotta et al., 2015). The six assumptions posit inequality as a function $I(\boldsymbol{y})$ and are listed below:

ASSUMPTION 1: (i) $I(\boldsymbol{y})$ is continuous and symmetric. (ii) $I(\boldsymbol{y}) = 0$ iff $\boldsymbol{y} = \mu\boldsymbol{e}$, for $\boldsymbol{e} = (\mathbf{1}, \dots, \mathbf{1})$**,** where μ is the mean income.

ASSUMPTION 2: (i) $S_k(\boldsymbol{y^1}, \dots, \boldsymbol{y^K}; K)$ is continuous in $\boldsymbol{y^k}$**.** (ii) That $S_{\boldsymbol{\pi_k}}(\boldsymbol{y^1}, \dots, \boldsymbol{y^K}; K) = S_k(\boldsymbol{Y^{\pi_1}}, \dots, \boldsymbol{Y^{\pi_K}}; K)$, where $\boldsymbol{\pi_1}, \dots, \boldsymbol{\pi_K}$ is some permutation of $1, \dots, K$.

ASSUMPTION 3: That $S_1(\boldsymbol{y^1}, \dots, \boldsymbol{y^K}; K) = S_1(\boldsymbol{y^1}, \boldsymbol{y} - \boldsymbol{y^1}; 2) = S(\boldsymbol{y^1}, \boldsymbol{y})$, which means independence of the level of disaggregation.

ASSUMPTION 4: That decomposition is consistent, hence $\sum_k S_k(\boldsymbol{y^1}, \dots, \boldsymbol{y^K}; K) = \sum_k S(\boldsymbol{y^k}, \boldsymbol{y}) = I(\boldsymbol{y})$.

ASSUMPTION 5: (i) Of population symmetry, that $\boldsymbol{P}$ is some n by n permutation matrix. $S(\boldsymbol{y^k P}, \boldsymbol{yP}) = S(\boldsymbol{y^k}, \boldsymbol{y})$. (ii) And normalization for equal factor distribution, which means that $S(\mu^{\boldsymbol{k}}\boldsymbol{e}, \boldsymbol{y}) = 0$, $\forall\, \mu_k$.

ASSUMPTION 6: That for all permutation matrices, $S(\boldsymbol{y^1}, \boldsymbol{y^1} + \boldsymbol{y^1 P}) = S(\boldsymbol{y^1 P}, \boldsymbol{y^1} + \boldsymbol{y^1 P})$.

Appendix 2: Robustness tests for disposable income (akin to Tables 2–6 and Figure 2)

Figure A1 plots and Table A1 presents the descriptive statistics for disposable income, which are akin to the plot and descriptive statistics provided for market income in Figure 1 and Table 1 in Section 4. We see that the relationship between mean income and the standard deviation of disposable income remains intact when compared to market income. Mean disposable income is the highest for ISE followed by W and lastly SE. The standard deviation is highest for ISE followed by SE and lastly W. These relationships are observed for both 2005 and 2013.

Figure A1. Lorentz curve for workforce market income, 2005 and 2013

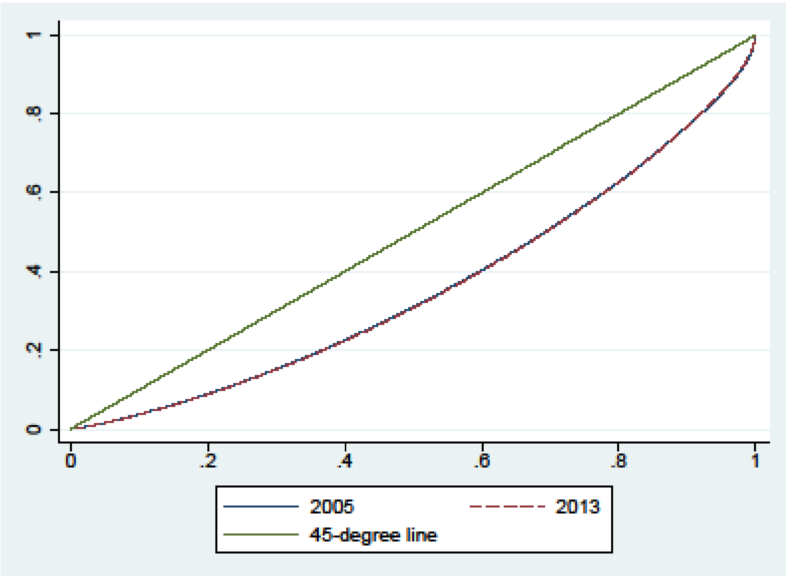


Table A1. (akin to Table 1). Descriptive statistics: Disposable income among entrepreneurs

	2005				*2013*			
	Mean	*Sd.*	*Min*	*Max*	*Mean*	*Sd.*	*Min*	*Max*
Variables:	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Number of Workers (W)								
Disposable income	1565.963	1341.788	0.909	516446.031	1925.266	1595.889	0.812	1154628.375
Self-employed (SE)								
Disposable income	1358.659	1941.492	0.909	237640.344	1671.789	1900.219	0.812	249922.531
Incorporated entrepreneurs (ISE)								
Disposable income	2203.164	3013.582	5.455	358868.906	2712.824	2655.294	22.729	263005.594

In Table 2, we see that inequality for market income generally declines over the 2005–2013 period. For disposable income, total inequality increases according to most measures shown in Table A2. The exception are GE(1) and GE(2), which instead show that inequality decreases. For entrepreneurs, the picture is less uniform. Whereas W largely reflects changes in total inequality, increases in inequality for SE and ISE seem to be most pronounced in the lower income ranges, although the total distribution seems to be more elongated as suggested by an increase in p90/p10. The Gini coefficient also increases slightly by less than 1%.[17]

Table A2 (akin to Table 2). Income inequality statistics of disposable income

	2005				*2013*			
	W	*SE*	*ISE*	*Total*	*W*	*SE*	*ISE*	*Total*
Inequality measure:	*(1)*	*(2)*	*(3)*	*(3.5)*	*(4)*	*(5)*	*(6)*	*(7)*
Disposable Income								
p90/p10	3.263	4.816	4.055	3.333	3.334	4.868	4.176	3.403
p90/p50	1.714	2.073	2.01	1.730	1.742	2.049	2.033	1.759
p50/ p10	1.904	2.324	2.018	1.923	1.914	2.376	2.054	1.934
p75/p25	1.872	2.27	1.98	1.892	1.893	2.284	2.057	1.912
GE(-1)	0.151	0.508	0.256	0.171	0.157	0.699	0.225	0.181
GE(0)	0.13	0.257	0.234	0.139	0.133	0.252	0.201	0.14
GE(1)	0.155	0.311	0.317	0.167	0.152	0.276	0.24	0.161
GE(2)	0.367	1.021	0.935	0.414	0.344	0.646	0.479	0.362
Gini				0.283				0.285

In Tables A3 and Figure A1, we present the same decomposition as in Table 3 and Figure 2 in the main paper but for disposable income instead of market income. The results are qualitatively similar to those for market income. W accounts for roughly 80%–90% of aggregate inequality. Looking at Figure A1 (i) and (ii), we see that the inequality group dynamics between SE and ISE largely remain for disposable income over the period. SE (ISE) stands for most of workers' residual contribution when inequality is computed with emphasis on the lower (upper) tail of the disposable income distribution.

Table A3 shows that when using disposable income, ISE appears to increase aggregate inequality by amending incomes to the upper end of the distribution for both disposable and

[17] As with measures based on market income, our inequality estimates for disposable income differ from those of the OECD (see footnote no. 6, page 15). Compared to the OECD's country analyses which report a marked increase in Gini for disposable income from 0.236 to 0.268 for the years 2004 and 2011, our data shows a much more marginal increases in Gini for disposable income between 2005 and 2013 (0.283 to 0.285). The differences are due to us using non-weighted, individual data while the OECD using constant prices and normalize disposable income by the square root of household size, including also those unemployment and outside the workforce.

market income. The contribution from ISE in 2005 and 2013 to market income inequality for GE(2) of 10.69% and 10.128% is close to the contribution to disposable income inequality of 12.41% and 9.73%.

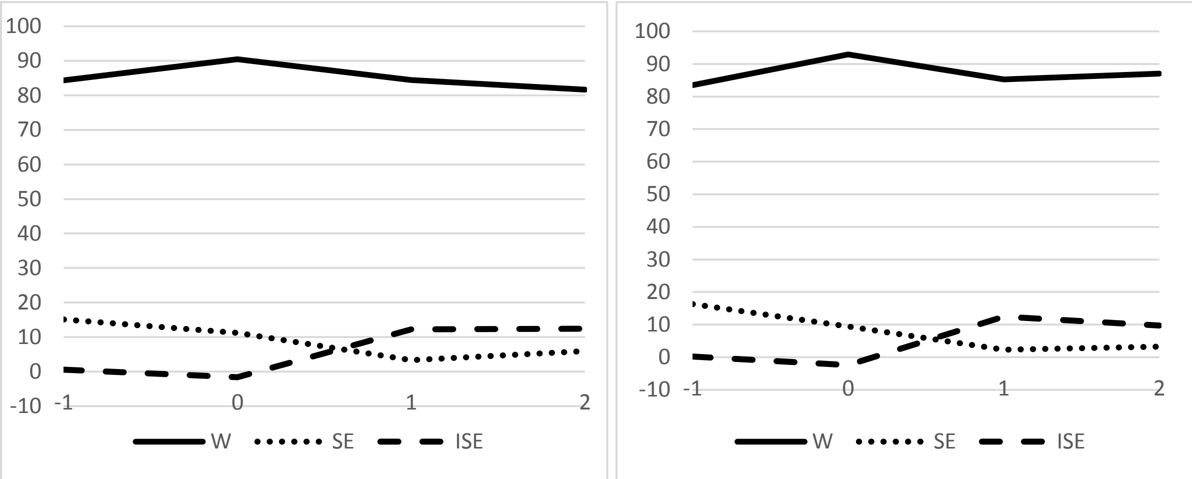
Table A3 (akin to Table 3): Sub-group decomposition in percentage points of $GE(y;\alpha)$ for disposable income, 2005 and 2013

	2005				*2013*			
	W (1)	*SE* (2)	*ISE* (3)	*Total* (4)	*W* (5)	*SE* (6)	*ISE* (7)	*Total* (8)
GE(-1)	0.151	0.508	0.256	0.171	0.157	0.699	0.225	0.181
Between: $\Lambda_b(y_j;\alpha)$	1.077	1.778	-1.862	0.993	1.496	1.452	-1.956	0.991
Within: $\Lambda_w(y_j;\alpha)$	83.331	13.302	2.374	99.007	81.96	14.851	2.197	99.009
Total: $\Lambda(y_j;\alpha)$	84.408	15.079	0.513	100	83.456	16.303	0.241	100
GE(0)	0.13	0.257	0.234	0.139	0.133	0.252	0.201	0.14
Between: $\Lambda_b(y_j;\alpha)$	2.632	4.043	-5.371	1.304	3.844	3.475	-5.94	1.378
Within: $\Lambda_w(y_j;\alpha)$	87.853	7.115	3.728	98.696	89.129	5.953	3.539	98.622
Total: $\Lambda(y_j;\alpha)$	90.485	11.158	-1.643	100	92.973	9.428	-2.401	100
GE(1)	0.155	0.311	0.317	0.167	0.152	0.276	0.24	0.161
Between: $\Lambda_b(y_j;\alpha)$	-2.189	-2.918	6.285	1.179	-3.338	-2.62	7.268	1.31
Within: $\Lambda_w(y_j;\alpha)$	86.691	6.219	5.911	98.821	88.584	4.923	5.183	98.69
Total: $\Lambda(y_j;\alpha)$	84.502	3.302	12.197	100	85.247	2.303	12.45	100
GE(2)	0.367	1.021	0.935	0.414	0.344	0.646	0.479	0.362
Between: $\Lambda_b(y_j;\alpha)$	-0.881	-1.179	2.579	0.519	-1.484	-1.169	3.294	0.64
Within: $\Lambda_w(y_j;\alpha)$	82.549	7.103	9.829	99.481	88.504	4.423	6.432	99.36
Total: $\Lambda(y_j;\alpha)$	81.668	5.924	12.408	100	87.02	3.254	9.726	100

Note: The table shows the percentage contribution (1=100%) of the between and within-inequality component from the sub-groups workers, SE, and ISE. Separate contributions are calculated for the various GE-indices with $\alpha = \{-1,0,1,2\}$. The inequality levels for each of the subgroups are calculated using eq. (16) with the appropriate weights. These inequality levels differs from those presented in Table 2 that comprise raw calculations based on equation (1) applied to the restricted sample.

Table A3 also depicts a corresponding shift for SE compared to ISE but by increasing inequality in the lower-end distribution of disposable income. Our main results in Table 3 in Section 5 show that SE contributes to market income inequality for GE(-1) in 2005 and 2013 by 24.18% and 32.65%. This result is substantially higher than the contribution of SE to disposable income inequality of 15.08% and 16.30%, as shown in Table A3 above. The difference, we suspect, can be attributed to transfer payments, which should counteract inequality, especially at the lower end of the disposable income distribution. In Figure A1, we plot the total percentage contribution to income inequality for each of the three groups at different levels of the sensitivity parameter α. Overall, the figure confirms the findings in Table A3 and the interchangeable roles played by SE and ISE at the bottom and top of the disposable income distribution.

Figure A2 (akin to Figure 2). Proportional contribution to overall inequality in disposable income from workers (W), self-employed (SE), and incorporated entrepreneurs (ISE).



(i) Disposable income 2005 (ii) Disposable income 2013

The next three tables replicate Tables 4–6 in the main paper using disposable income instead of market income. Again, we use Fiorio and Cowell's (2011) method to decompose within-group inequality $\Lambda_w(y_j; \alpha)$ to a number of explanatory variables (i.e., factor sources).

Table A4: Inequality weights (w_j) for the GE index

	Workers	SE	ISE	Workers	SE	ISE
	Disposable income, 2005			Disposable income, 2013		
GE(-1)	0.943	0.045	0.016	0.947	0.038	0.018
GE(0)	0.939	0.039	0.022	0.942	0.033	0.025
GE(1)	0.936	0.033	0.031	0.937	0.029	0.035
GE(2)	0.932	0.029	0.044	0.931	0.025	0.049

We also estimate the same regressions for disposable income (akin to Tables 5 and 6). Beginning with disposable income in 2005, we observe some key differences to market income. First, there are lower contributions from years of education by 1.511% for W, 0.40% for SE, and with 0.90% for ISE. For W and ISE, this is less than half the size of the percentage contribution for market income. For SE, the difference in the contribution from education appears to be small. Although age and gender are also important contributors to disposable income inequality, age appears to become more prominent for disposable income and gender less so. The largest difference observed between the two income measures is number of children. This result likely stems from individuals with children, especially women, restricting their labor supply, and/or the effects of governmental transfers to parents

(Angelov et al. 2013). While number of children hardly contributes anything to market income inequality, it accounts for a total of 3.75%, 0.82%, and 1.89% of the within-group inequality of disposable income among W, SE, and ISE, respectively.

Table A5: Regression results for disposable income in 2005 and contribution to $GE(\mathrm{y_j};\alpha)$

	Regression estimates disp inc.			% contrib. to $GE(y_j;\alpha)$, ($s_k \times 100$)		
Variable	Workers	Self-emp.	Inc. emp.	Workers	Self-emp.	Inc. emp.
Age	17.234***	24.407***	30.333***	2.16	1.838	1.297
	(0.125)	(0.730)	(1.520)	(0.000)	(0.003)	(0.006)
Age square	-0.127***	0.385***	-0.02	0.054	0.020	0.000
	(0.008)	(0.058)	(0.117)	(0.000)	(0.001)	(0.001)
Job tenure	7.140***	-6.899***	8.369***	0.104	-0.027	0.063
	(0.200)	(1.497)	(2.829)	(0.000)	(0.001)	(0.003)
Job Changes	38.566***	20.043***	7.694	0.077	-0.003	-0.014
	(0.761)	(4.055)	(7.976)	(0.000)	(0.001)	(0.002)
Nr. children	-221.486***	-119.543***	-338.088***	3.754	0.822	1.893
	(0.730)	(4.337)	(11.608)	(0.000)	(0.001)	(0.001)
Gender	411.694***	284.386***	390.040***	2.315	0.37	0.197
	(1.365)	(11.572)	(32.871)	(0.000)	(0.001)	(0.002)
Marital Status	-35.128***	54.164***	-109.299***	0.021	0.023	0.021
	(1.492)	(11.446)	(25.393)	(0.000)	(0.001)	(0.002)
Years of education	84.789***	68.917***	141.049***	1.511	0.403	0.9
	(0.392)	(3.482)	(5.898)	(0.000)	(0.003)	(0.005)
Immigrant	-175.163***	-352.961***	-373.377***	0.122	0.314	0.041
	(5.029)	(10.792)	(41.752)	(0.000)	(0.000)	(0.000)
Constant	-309.122***	-711.336***	-874.420***			
	(10.185)	(65.211)	(112.025)			
Obs.	3437021	141261	81132			
R-sq.	0.101	0.038	0.044			
Res. (=1-R-sq)				89.882	96.239	95.602

NOTE: Standard errors in parenthesis with significance levels *** p<0.01, ** p<0.05, * p<0.1. Columns (1) to (3) gives the robust OLS results from each of the groups W, SE and ISE, run separately. Columns (4) to (6) show the proportional contributions of the corresponding explanatory variable to the total within-group inequality for W, SE and ISE respectively. Because of the large sample sizes, almost all percentage contributions are strongly significant. For this reason we display their standard errors without asterisks denoting level of significance.

As for the changes in contributions, going from 2005 to 2013 for disposable income, the trends for education (decreasing except for ISE), gender (decreasing), and age (increasing) all mirror the findings for market income found in Tables 5–6 in the main results. As for number of children, the contribution to both W and SE remains roughly the same, but looking at the ISE group, number of children contributes 1.89% in 2005 and 3.03% in 2013 to the inequality within ISE.

Table A6: Regression results for disposable income in 2013 and contribution to $GE(\mathrm{y_j};\alpha)$

	Regression estimates disp inc.			% contrib. to $GE(y_j;\alpha)$, $(s_k \times 100)$		
Variable	Workers	Self-emp.	Inc. emp.	Workers	Self-emp.	Inc. emp.
Age	20.724***	24.011***	30.515***	2.348	2.024	1.803
	(0.175)	(0.717)	(1.167)	(0.000)	0.003	0.004
Age square	-0.126***	0.332***	-0.473***	0.073	-0.015	0.02
	(0.014)	(0.050)	(0.094)	(0.000)	0.001	0.001
Job tenure	8.360***	1.575	21.001***	0.225	0.029	0.485
	(0.192)	(1.239)	(1.833)	(0.000)	0.001	0.001
Job Changes	43.323***	39.194***	27.173***	0.256	0.132	-0.072
	(1.045)	(3.747)	(4.904)	(0.000)	0.001	0.001
Nr. children	-260.481***	-127.916***	-400.454***	3.413	0.881	3.053
	(0.702)	(5.001)	(9.106)	(0.000)	0.001	0.001
Gender	477.461***	264.432***	526.854***	2.079	0.393	0.485
	(1.602)	(11.674)	(23.224)	(0.000)	0.001	0.001
Marital Status	-25.315***	98.628***	-36.805**	0.011	0.045	0.008
	(1.729)	(10.967)	(17.580)	(0.000)	0.001	0.001
Years of education	101.963***	59.813***	172.994***	1.354	0.298	1.584
	(0.405)	(2.863)	(4.420)	(0.000)	0.003	0.003
Immigrant	-214.028***	-374.660***	-253.985***	0.219	0.541	0.062
	(6.798)	(13.888)	(36.675)	(0.000)	(0.000)	(0.000)
Constant	-402.542***	-397.503***	-1025.189***			
	(15.987)	(50.366)	(89.404)			
Obs.	3554138	125293	93311			
R-sq.	0.1	0.043	0.074			
Res. (=1-R-sq)				90.022	95.671	92.573

NOTE: Standard errors in parenthesis with significance levels *** p<0.01, ** p<0.05, * p<0.1. Columns (1) to (3) gives the robust OLS results from each of the groups W, SE and ISE, run separately. Columns (4) to (6) show the proportional contributions of the corresponding explanatory variable to the total within-group inequality for W, SE and ISE respectively. Because of the large sample sizes, almost all percentage contributions are strongly significant. For this reason we display their standard errors without asterisks denoting level of significance.