# Entrepreneurship and income inequality

Daniel Halvarsson [a], Martin Korpi [b,*], Karl Wennberg [c]

[a] Ratio Institute, P.O. Box 3203, 103 64, Stockholm, Sweden
[b] Ratio Institute, Stockholm, Sweden
[c] Institute for Analytical Sociology (IAS), Linköping University, 60174 Norrköping, Sweden

## ARTICLE INFO

## ABSTRACT

Entrepreneurship research highlights 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, but attention to entrepreneurship as an increasingly prevalent occupational choice in these models remains scant. This paper seeks to bridge these two literatures using 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 increases income dispersion by way of widening the bottom end of the distribution, whereas the proportion of incorporated self-employed contributes to income dispersion at the top end of the distribution. Implications for research are discussed.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

The last few decades have witnessed a notable rise in both self-employment and small business ownership in most developed economies, and much public policy in this context has been directed toward increasing the supply of entrepreneurs (Blanchflower, 2000; Steinmetz and Wright, 1989). At the same time, many countries have experienced increasing income inequality and stratified unemployment (Atkinson, 2003; Autor, 2014; Goldthorpe, 2010). Despite these two parallel economic trends, studies that probe the direct relationship between entrepreneurship and overall workforce inequality remain scant in the still largely separate literatures on entrepreneurship and inequality (Van and Versloot, 2007; Wright and Zahra, 2011). In the entrepreneurship literature, many studies have documented a two-pronged effect of entrepreneurship on individual income, arguing that entrepreneurship is a source of enhanced income mobility for some but results in lower-than-average incomes for the large fraction of the self-employed workforce (e.g., Hamilton, 2000; Åstebro et al., 2011). However, attention to the potential implications of these patterns for macro-level income distribution has been scant in the entrepreneurship literature. In the inequality literature, recent studies have developed new types of models to decompose and problematize different sources of income inequality (e.g., Cowell and Fiorio, 2011; Creedy and Hérault, 2011; Thewissen et al., 2013), but these studies have tended to overlook entrepreneurship as one of these potential sources. This oversight is problematic since entrepreneurship is an increasingly prevalent occupational choice in many economies (Audretsch, 2009).

In this paper, we seek to address this gap in the literature, which also forms the basis of our overall research questions: (1) Does entrepreneurship affect overall workforce income inequality? (2) Which aspects of the income distribution affect overall inequality and to what extent? (3) What individual characteristics among entrepreneurs (compared to employed workers) may account for the effects observed in Questions 1 and 2?

To account for the fact that self-employed entrepreneurs are over-represented in both tails of the earnings distribution (Åstebro et al., 2011), we allow for two types of entrepreneurs—those who are self-employed in sole proprietorships (SE) and those who are self-employed in incorporated businesses (ISE)—and we compare these two groups with salaried workers (W). While entrepreneurs are found to earn less than salaried workers on average, recent studies, such as Levine and Rubinstein (2017), Åstebro and Tåg (2017), and Özcan (2011), have suggested that even though this is true for SE-entrepreneurs, ISE-entrepreneurs tend to earn relatively more even when compared to employees with comparable traits and skills. Thus, in distinguishing between SE and ISE, we allow for the possibility that different types of entrepreneurship, or different types of entrepreneurs, may differ in their effects on total income distribution.[1]

Using highly detailed data on the full Swedish workforce and capitalizing on recent inequality research (e.g., Cowell and Cowell and Fiorio, 2011; Creedy and Hérault, 2011; Thewissen et al., 2013), we develop a novel empirical approach to answer these research questions. Central to our analysis is our choice to use a flexible inequality measure—the generalized entropy (GE) index—to fine tune our decomposition analysis to different segments of the income distribution. Using this measure, we begin the analysis by studying whether changes in aggregate inequality are systematically related to SE, ISE, and W and which parts of the distribution are affected. Next, we present an integrated factor-source decomposition analysis in which we consider a number of micro-level sources to account for the inequality among SE, ISE, and W. Specifically, 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 relate to aggregate inequality in the total workforce (Cowell and Fiorio, 2011; Fields, 2003).

Based on this analysis, we find that on average, SE-entrepreneurs have lower incomes than workers (W), whereas ISE-entrepreneurs tend to earn relatively more than both these two groups. When considering the total income distribution in the workforce, we find that SE- and ISE-entrepreneurs contribute substantially to income inequality, but using different GE indices, we find that this relationship is U-shaped. While the prevalence of SE-entrepreneurs increases inequality by widening the bottom end of the distribution, ISE-entrepreneurs augment inequality from the top of the distribution by enhancing the total number of high-income earners in the workforce. In particular, when tuning the GE index to lower parts of the income distribution, SE accounts for about 30% of total inequality and ISE contributes only a fraction, but the opposite is true for the upper end of the income distribution. When tuning the GE index to the top of the distribution, we instead find that ISE accounts for about 10% of total inequality and SE accounts for only a fraction. Both types of entrepreneurship, however, capture very little variation in the middle of the income distribution.

These results highlight the importance of considering entrepreneurship in studies of income inequality and demonstrate that inequality measures emphasizing middle-range income, such as the commonly used Gini coefficient, may not capture these effects well. In regard to our factor-source decomposition analysis, a key finding is that individuals' education and gender differences each account for about 4% of total income inequality. While the variable years of education explains inequality among workers and among ISE-entrepreneurs, this is not so among SE-entrepreneurs, suggesting low returns to education among SE-entrepreneurs (Van der Sluis et al., 2005).[2] Further, gender differences in earnings contribute more to inequality among workers than among either SE or ISE.

To the best of our knowledge, this paper is the first to assess the effects of entrepreneurship on income inequality using state-of-the-art decomposition techniques. It also represents a first attempt to address 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. While a host of studies have shown that entrepreneurs generally have a much wider distribution of earnings compared to workers, as far as we know, no one has actually examined the implications of this for the wider distribution of income in the economy as a whole. Our results suggest that entrepreneurship, at least at the tails of the distribution, affects the economy-wide distribution of earnings in a similar magnitude as more conventional factors included in most models and studies gauging inequality.

Our methodology leans heavily on the work of Cowell and Fiorio (2011), which decomposes inequality indices using several sub-groups embedded in a regression framework. This approach 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 while also assessing their individual contribution to sub-group and aggregate inequality.

---

[1] We also chose this division because in most countries, including Sweden, these two forms of entrepreneurship are subject to different tax laws (Edmark and Gordon, 2013). As shown in Tables A7 and A8 in the appendix, the type of firms started by SE- and ISE-entrepreneurs also differ systematically in terms of entrepreneurial team size (Table A7), firm size (Table A7), and industry classification (Table A8). Recent studies have also suggested differences in terms of innate ability among SE- and ISE-entrepreneurs (Levine and Rubinstein, 2017, Åstebro et al., 2011), a point we return to later. For a lengthy discussion and an outline of differences along similar lines using US data, see Herranz et al. (2009) and De Nardi et al. (2007).

[2] Robustness tests for a sub-group of the 2005 cohort presented in Footnote 17 suggest that income inequality among SE-entrepreneurs may be explained to a large extent by innate ability as proxied by high school grades and aptitude tests.

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

The empirical literature has noted that 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 downward (i.e., skewed toward low-income earners). Similarly, in a paper comparing the return on 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 but 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 individuals 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 even though 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. 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.

Other recent studies have vindicated and extended the stylized pattern provided by Hamilton (2000). Åstebro et al. (2011) use data from the Korean Labor and Income Panel Study 1998–2004 to test a model of occupational choice, finding that entrants into self-employment are drawn disproportionately from both tails of the earnings distribution. However, in contrast to the Hamilton study, Åstebro et al. also find that the skewed income distribution 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. Another recent study by Levine and Rubinstein (2016) uses US data from the Current Population Survey 1994–2010 to separate salaried workers and entrepreneurs in the form of SE and ISE, pinpointing both the determinants behind the sorting of people into these different types of employment and their subsequent earnings. Levine and Rubinstein argue that earlier findings suggesting that entrepreneurs earn below the median income of salaried workers essentially reflect the facts that there are significantly fewer incorporated entrepreneurs than self-employed entrepreneurs and that this latter group—which indeed tends to exhibit lower incomes—tends to dominate estimates of "average outcomes" in earning equations comparing entrepreneurs to workers. In contrast, their results suggest that incorporated entrepreneurs' earnings are in the magnitude of 30% above salaried workers with comparable traits and skills, who in turn earn more than their self-employed counterparts. As suggested by Åstebro et al. (2011), Levine and Rubinstein (2017) find that these outcomes reflect innate ability (i.e., non-cognitive traits and cognitive skills) to a significant degree. Whereas some traits are common of both types of entrepreneurs (e.g., engaging in illicit activities as teenagers and risky behavior), incorporated entrepreneurs score higher in learning aptitude tests than self-employed individuals.[3]

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

---

[3] These results are in line with work focusing on "necessity entrepreneurship." Necessity entrepreneurs are defined as those who have started their own firm "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). The Global Entrepreneurship Monitor suggests 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).

among individuals choosing to leave these firms for entrepreneurship (Carnahan et al., 2012; Kacperczyk and Balachandran, 2017; Sørensen and Sharkey, 2014). However, a few macro-oriented studies have indicated that the greater the number of small firms in an economy, the more unequal the earnings distribution in that economy is (Davis, 2013; Cobb and Lin, 2017; Fields and Yoo, 2000). In terms of explaining this pattern, the literature has mostly provided very broad structural explanations. 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 has argued that rising levels of entrepreneurship may increase inequality by expanding the share of either top or bottom income earners—or both—in the workforce. However, this possibility is 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. Studies explicating a potential relationship between entrepreneurship and income inequality have tended to focus on income inequality within firms—most often measured by the traditional Gini coefficient—and its effects on workers' mobility to entrepreneurship. Studies using country comparisons have suggested that entrepreneurship is positively related to aggregate income inequality, but they have been silent in regard to which parts of the income distribution might be affected and how different types of entrepreneurship may affect overall inequality. In this literature, there has also been no attempt of any kind to link individual-level developments to either within-group inequality or changes in aggregate outcomes. These empirical puzzles form the basis of our research questions, to which we now turn.

## 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 begin by choosing an inequality 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 gauge both the different contributions of entrepreneurs and salaried workers to aggregate inequality as well as how different factors contribute to inequality, decomposability is crucial. Second, since we want to estimate of our sub-groups' respective contribution to inequality at various ranges of the income distribution, the Gini coefficient is also inappropriate since it mostly reflects variation in middle range income.

We thus need a more flexible measure and the generalized entropy index (GE-index) fulfils both requirements, allowing for the study of inequality within each group separately and also for an assessment of how much each group contributes to the aggregate. Further, the GE-index is defined as a function of a sensitivity parameter $\alpha \in (-\infty, \infty)$, which allows us to adjust its sensitivity to specific parts of the distribution. By choosing different values of $\alpha$ 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 $\alpha$, the more sensitive the GE-index is to dispersion in the lower parts of the income distribution. Conversely, the higher the level of $\alpha$, 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 the real line, 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 $\alpha$ is tuned to either the bottom or top parts of the distribution (i.e., $\alpha \in \{-1, 2\}$) as compared to the middle (i.e., $\alpha \in \{0, 1\}$).[4]

To better describe the properties of the GE-index we include a formal definition. Let $y = [y_1 \ldots y_N]$ be a vector of incomes for a total of $N$ individuals in the population. A sample analogue of the GE-index can then be defined as a function of $\alpha$ by the expression

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

The sum is taken over individual incomes $y_i$ ($i = 1, \ldots, N$) with exponent $\alpha$ scaled by the mean income of the population $\mu(y)$. We can interpret the different values of $\alpha$ 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 $\text{var}(y)/2\mu(y)$ and is synonymous with the Hirschman-Herfindahl concentration index (Quintano et al., 2005). When $\alpha$ takes the values 1 or 0, in turn, $GE(y; 1)$ and $GE(y; 0)$ is evaluated in the limit as $\alpha \to 1$ and $\alpha \to 0$, respectively. At $\alpha = 1$, the GE-index is identical to the so-called Theil index, and at $\alpha = 0$, it corresponds to the mean logarithmic deviation (MLD). In this listed order, these entropy measures are sensitive to changes in the top, upper-middle, and lower-middle income ranges, respectively. Lastly, at $\alpha = -1$, the emphasis of the index $GE(y; -1)$ is on the bottom ranges of the distribution, as it contains the expected value of reciprocal income ($1/y$).

---

[4] 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.

### 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 sub-group to overall inequality.[5]

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

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

one between-group part $GE_b(y; \alpha)$, which reflects inequality as measured by variance in mean income *between* different groups, and one within-group part $GE_w(y; \alpha)$, which reflects the dispersion of individual income within each separate group. Here, the latter expression, $GE_w(y; \alpha)$, is a weighted sum of the GE-index computed for each of the sub-groups ($j = 1, \ldots, J$) 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 ratio between mean income in sub-group $j$ and the mean income of the population (Cowell and Fiorio, 2011). Substituting this expression for the population weight into expression (3) and using (1) yields the following formula for total within-group inequality:

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

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 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.[6] The GE-index thereby accounts for the differences in mean incomes across the sub-groups—namely, between-group inequality.[7]

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 $j$, as

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

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 to the within component. This expression is useful whenever we are interested in the total contribution from one particular sub-group rather than the combined within- 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 particular sub-group's total contribution to aggregate inequality.

---

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

[6] The "between part" in Eq. (5) is written slightly different here compared to Cowell and Fiorio (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, it is less clear, using their expression, how to calculate the contribution to aggregate inequality from one particular group's "between part" since their equation leaves out the "residual term" of $\frac{-1}{\alpha^2 - \alpha}$ that also needs to be distributed among the groups. Using the expression in (5) instead, it neatly takes care of this inconvenience by portioning the residual term between the groups by the amount of $-N_j/N(\alpha^2 - \alpha)$.

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

### 3.2. Factor source decomposition of a sub-group's within inequality

In this section, we turn to inequality decomposition by factor sources, something that often accompanies a sub-group decomposition analysis. Traditionally, a factor-source analysis is used to decompose inequality into income sources such as salaried income, capital income, or transfer payments. To allow for a much wider set of possible factors we rely on Fields (2003) and Cowell and Fiorio (2011) who develop a regression-based approach to inequality decomposition. When specifying the inequality model, this approach allows us to include determinants of income typically used in Mincer-type wage regressions but are rarely incorporated in studies on inequality.

The regression-based approach for factor-source decomposition can be described as follows. Let the income for individual $i$ $(i = 1, \ldots, N)$ in sub-group $j$ $(j = 1, \ldots, J)$ be split into a sum of $K$ different factor sources (i.e., components):

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

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

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

where $y_{jk}$ refers to the $k^{\text{th}}$ income source for the $j^{\text{th}}$ 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 the proportional contribution of $S_{jk}$ to $I\left(y_j\right)$ by,

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

Since $s_{j1} + s_{j2} + \ldots + s_{jK} = 1$ by construction, multiplying through with $I\left(y_j\right)$ shows that $s_{jk}$ becomes the loading of factor $k$ to the inequality $I\left(y_j\right)$ decomposed into

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

Furthermore, for this class of inequality measures (the GE-index among them), the term $s_{jk}$ can be expressed in terms of the covariance between the income component $y_{jk}$ and total income $y_j$ divided by the variance of $y_j$, as follows

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

This result comes from theorem 3 of Shorrocks (1982) and was first used by Fields (2003) to connect income regression analysis with traditional factor-source decomposition of income inequality. In the next section we present a basic income model and show how the result in Eq. (11) can be used to decompose the inequality of income into $K - 1$ explanatory sources using linear regression analysis.

### 3.3. Factor source decomposition using regression analysis

We consider a linear model of income for individual $i$ in group $j$ given by

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

which includes $K - 1$ number of (potentially endogenous) explanatory variables $x_{kji}$, and an i.i.d. error term, $u_{ij}$. Because the model has the same linear form as the expression in Eq. (7) we can decompose the inequality of $y_j$, i.e. $I\left(y_j\right)$, into factor sources by mapping $b_{jk}x_{ijk}$ in Eq. (12) to $y_{jk}$ in Eq. (7). By expanding the covariance in Eq. (11), Fields (2003) shows that different parts of the linear income model are related to $s_{jk}$ as follows (Cowell and Fiorio, 2011)

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

where the first term gives the direct contribution of $b_{jk}x_{jk}$ to $I\left(y_j\right)$; the second term is a sum and represents the contribution to $I\left(y_j\right)$ from multicoliniarity, i.e. if $x_{jk}$ is correlated with $x_{jr}$ for the $r \neq k$ other explanatory variables; and the third term represents the contribution to $I\left(y_j\right)$ due to endogeneity, i.e., if the $x_{jk}$ term is correlated with the residual term $u_j$. As for

the last term, $s_{jK}$, it represents the contribution to inequality from the unobserved part of Eq. (12) expressed in terms of the residual's direct contribution as well as the variation resulting from any endogenous explanatory variables, given by[8]

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

To find the sample version of $s_{jk}$ and $s_{jK}$, we estimate the (no-log) income model with OLS and collect the point estimates. These are combined with information from the covariance matrix of all variables included in the empirical model to form estimates $\hat{s}_k$ and $\hat{s}_K$.

When it comes to the standard deviations of the estimates $\hat{s}_k$ and $\hat{s}_K$, Cowell and Fiorio (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 less attractive. In Bigotta et al. (2015) the authors provide analytical standard errors for $\hat{s}_k$. With our inclusion of 3-digit industry codes along with dummies for local-labor markets, even the calculations of these become too cumbersome with ordinary statistical software and computational power.[9] We therefore show decomposition results without the attached standard errors. This is in line with much of the previous literature using regression decomposition techniques in which the reporting of significance levels has yet to become standard.

### 3.4. Decomposition of within-group inequality in several sub-groups

In the last section we kept the sub-group indexation $j$ in the term $s_{jk}$ and $s_{jK}$ in Eqs. (13) and (14) in line with Cowell and Fiorio (2011) who show that it is possible to reconcile within-group decomposition of inequality into factors sources with the sub-group decomposition presented in Section 3.1, using regression analysis and OLS. In particular, using regression coefficients expressing $\mu (y_j) = \sum_{k=1}^{K-1} b_{jk} \mu (x_{jk})$, Eq. (6) they show that the inequality contribution from group $j$ to total inequality $I(y)$ can be restated as follows,

$$\tilde{GE} (y_j; \alpha) = \frac{p_j}{\alpha^2 - \alpha} \left( \left[ \frac{\sum_{k=1}^{K-1} b_{jk} \mu (x_{jk})}{\sum_{k=1}^{K-1} b_k \mu (x_k)} \right]^\alpha - 1 \right) + w_j \sum_{k=1}^{K} GE (y_j; \alpha) s_{jk}. \tag{15}$$

The first expression still captures the group-component to between-group inequality, but here this is expressed in terms of parameters from two separate regressions, one restricted to individuals in sub-group $j$ and one unrestricted on the full population. As before, the second term captures within-group inequality, but here expressed as the population weighted ($w_j$) sum of $K$ number of inequality components each scaled with the loading factor $s_{jk}$. Except for the population shares that can be 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 Eq. (12) for each of the $j$ subgroups.

## 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 increased markedly between 1985 and the early 2010s (OECD, 2015).[10] During the same period, the country saw increasing rates of entrepreneurship in the form of ISE, making Sweden an interesting case to probe the role of entrepreneurship in income inequality.

---

[8] Since all direct contributions are squared, a necessary condition for $s_k$ to be negative is that $x_{jk}$ is either correlated with at least one of the $x_r$ ($r \neq k$) resulting in multicollinearity, or it is correlated with the error term resulting in endogeneity (Cowell and Fiorio, 2011). On the other hand, if all standard assumptions in OLS are satisfied (i.e., no multicollinearity and no endogeneity) $s_{kj}$ and $s_{Kj}$ are reduced to their respective first terms, $s_{jk} = b_{jk}^2 \sigma^2 (x_{jk}) / \sigma^2 (y_j)$ and $s_{jk} = \sigma^2 (u_j) / \sigma^2 (y_j)$.

[9] When computing Bigotta et al. (2015) standard errors for our smaller samples, we find them to be very small in magnitude. Except for a few cases when the OLS estimate is insignificant, and the percentage contribution is miniscule, $s_{jk}$ tend to be highly significant.

[10] Our inequality estimates are somewhat different from the OECD's country analyses, which report a slight increase in Gini for market income from 2004 to 2011 (0.369–0.371). OECD changed their definition of market income in 2012 to a more detailed breakdown of household income transfers and a revised definition of household income. A main source of divergence between the OECD's estimates of income inequality and ours is due to 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 of non-zero market income, whereas OECD analyses also include those unemployed and outside the workforce. Taken together, the aggregate OECD statistics are fairly consistent with ours that show a slightly declining Gini coefficient for market income between 2005 and 2013 (0.318–0.309, see Table 2 below).

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 does not include 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 including 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 the *level* of inequality for each of these years as well as probe any change that occurred over the nine-year period.

We use all individuals in the workforce between 25 and 64 years of age in the respective years for which labor market data is available. From this data, we then exclude a number of individuals who are not associated with an employing organization, such as sailors and seasonal workers. The sample used for analysis comprises 3,619,132 individuals in 2005 and 3,746,272 individuals in 2013. Except for age, the only criteria we use to exclude individuals from the sample is when an individual 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 enables us to distinguish between two types of entrepreneurs: those individuals who are self-employed (SE) in a private business (sole proprietorships) and those who are self-employed in an incorporated business (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 corresponds to 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).[11] 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.[12] Based on these data, we only include entrepreneurs reporting their own business as the source of the majority of their 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 h a year in his or her own business. Unless entrepreneurs actively report doing business in this manner, they are put into the worker category.

This rather strict definition of what constitutes entrepreneurship may reduce the number of individuals in both our SE and ISE categories, but it ensures that we can be certain that running their respective business is the main occupation of these entrepreneurs.

Similar to studies on inequality by, for example, the OECD (2015), the primary income variable in our models is *market income*, defined as the sum of gross wage income plus net income from an active business plus capital income. All three variables are included in market income to be able to better compare income from both types of entrepreneurship: while SE-entrepreneurs receive 100% of their earnings in the form of net income from an active business, ISE-entrepreneurs receive their earnings as both gross wage income (from their business) and capital income (Alstadsæter and Jacob, 2016; Edmark and Gordon, 2013).

Although our main focus in the paper is on market income, in a supplementary analysis, we also estimate our models using *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 overly affected 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 both factor incomes, such as net wages, business-related income (net deficit), and net capital profits, and taxable and non-taxable transfers, such as rehabilitation compensation, pensions, and child allowances (e.g., housing benefits, social security, and study allowances).

In the empirical model, we consider the following explanatory variables: *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

---

[11] Incorporation processes and tax rates differ across countries and have implications for entrepreneurs' choice of legal form (see e.g. De Nardi et al., 2007; Edmark and Gordon, 2013; Herranz et al., 2009). These processes has been compared by e.g. World Bank scholars in terms of differences in are significant differences in the incorporation process for Sweden and other nations in terms of *cost, time* and *coverage* of incorporation, suggesting that (i) *costs* for incorporating a firm in Sweden was significantly higher than comparable countries in 2005, but not in 2013 after the minimum equity required when forming an incorporated business was lowered from 100,000 to 50,000 SEK as of October 1, 2010, (ii) *time* required to incorporate a firm in Sweden is shorter than most countries (Djankov et al., 2002), and (iii) *coverage* of incorporation is comparable to other nations, however Edmark and Gordon (2013) note that although lower-income individuals face relatively neutral incentives, higher-income households face tax incentives to incorporate.

[12] 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.

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 gained at a workplace and the number of workplace switches since 1990.[13] Years of education is the most common operationalization of general human capital both 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* (1 = married/cohabitant) and *gender* (1 = male; 0 = female) as well as industry (using three-digit NACE codes) and regional differences. Our region dummy variable is based on Statistics Sweden's 2006 definition, which asserts that Sweden can be divided into 79 separate local labor markets.[14]

Finally, we note that our analysis of income inequality is sensitive to the very top incomes in each year. We therefore exclude a number of top income earners who severely skew the income distribution, the inclusion of which also violates the ordinary least squares (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, and including most individuals within this category is also motivated by the literature's recent interest in top income earners and the debate regarding the top one percent (see e.g. Atkinson et al., 2011; Quadrini, 1999; Roine and Waldenström, 2008). Thus, there is a tradeoff between dropping top income earners, which increases $R^2$ and the relevance of our study in terms of encompassing the entire distribution of wage income earners (including all top earners' results in zero $R^2$ for our Mincer regressions for non-logarithmic market income). We settle on a fairly generous restriction by keeping 99.9999% of the full sample, merely excluding 366 and 377 individuals with the highest market income in 2005 and 2013, respectively.[15]

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 decomposition weights has the consequence that the decomposition then corresponds to 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%. As noted in Footnote 11, this partly reflects the changes in minimum equity required for forming an incorporated business in 2010.

Most variables display fairly moderate changes across the two time periods, which is to be expected when working with population-sized datasets. 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 [e.g., see Robinson and Sexton, 1994]), average years of education for all groups increases.[16] 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. When comparing SE and ISE entrepreneurs, we note that their demographics are quite similar, albeit with a higher average job tenure for ISE entrepreneurs. Furthermore, the proportion of immigrants among SE entrepreneurs is 7.0% (10.1%) in 2005 (2013) but among ISE entrepreneurs is only 1.6% (1.8%) in 2005 (2013). Average education is only slightly higher among ISE entrepreneurs.

Turning our attention to the descriptive statistics of aggregate inequality, Fig.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 decreases 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 displays a slight outward shift (see Fig. A1 in Appendix B).

Since the two Lorentz curves are close together, they are difficult to differentiate visually. In the right panel of Fig. 1(b), we therefore plot the vertical distance between the two curves, computed as the difference between the curves for 2013 and

---

[13] 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.

[14] For detailed information on industry and local labor market classification criteria, see Eurostat (2008) and SCB (2010), respectively. These controls marginally decreased the sample size by 40,282 (1%) of all individuals in 2005 and 26,470 (0.7%) of all individuals in 2013 since some individuals' lack data on industry affiliation.

[15] Coefficient estimates of our models are not overly sensitive to this cut-off but the more restrictive our sample definition, the larger the increase in $R^2$. For example, dropping the top 0.5 % in terms of market income (20, 000 individuals) nearly doubles $R^2$ for all three categories of individuals examined (Tables available upon request).

[16] 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.

**Table 1**
Descriptive statistics: workers (W), self-employed entrepreneurs (SE) and incorporated self-employed entrepreneurs (ISE).
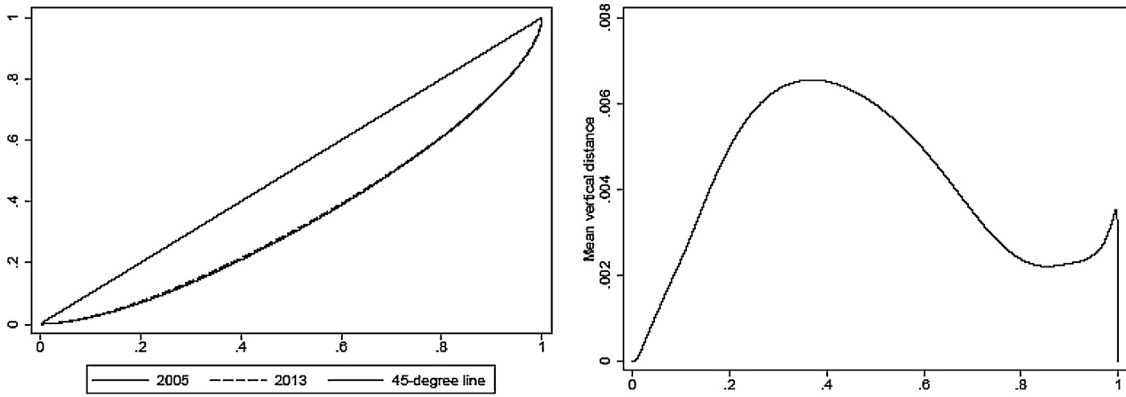
| Occupational groups | 2005 | | | | 2013 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | St.dev | Min | Max | Mean | St.dev | Min | Max |
| **Workers (W)** | | | | | | | | |
| Market income (in 100 s Swedish krona) | 2,453.165 | 2,13.554 | 0.909 | 1,14150.68 | 2,827.492 | 2,252.876 | 0.812 | 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 |
| Immigrants | 0.037 | 0.188 | 0 | 1 | 0.064 | 0.245 | 0 | 1 |
| 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 |
| Immigrants | 0.069 | 0.253 | 0 | 1 | 0.102 | 0.302 | 0 | 1 |
| Obs. | 141,261 *(3.86% of the workforce)* | | | | 125,293 *(3.32% of the workforce)* | | | |
| **Incorporated self-employed (ISE)** | | | | | | | | |
| Market income (in 100 s 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 |
| Immigrants | 0.016 | 0.126 | 0 | 1 | 0.33 | 0.178 | 0 | 1 |
| Obs. | 81,132 *(2.22% of the workforce)* | | | | 93,311 *(2.47% of the workforce)* | | | |

2005, respectively. The increasing vertical distance reflects a smaller area between the 45° line and the Lorentz curve, i.e. a decrease in inequality. From the figure, we also see that this decrease mainly involves the bottom 40 percent of individuals in the income distribution (for which the vertical distance between the two curves is the largest). For income levels above the 40th percentile the distance is still positive however smaller in magnitude.

The slight inward shift and decrease in overall income inequality in Fig. 1 represents the aggregate changes across all occupational groups. We attend to these changes in Table 2, which shows the Gini coefficient, percentile ratios, and GE indices for our separate sub-groups of W, SE and ISE in the years 2005 and 2013.

Beginning with workers (W), we see a slight decrease in inequality over the period across almost all inequality measures. The decrease may be small but is nevertheless present. The exceptions are for inequality as measured by the percentile ratio p90/p50, which displays a slight increase of 1%.

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 suggests some form of contraction of top incomes, whereas bottom incomes become more dispersed among SE. The latter appears significant as the GE(-1)-index increases by a total of 34.9%. As for ISE, inequality among this group of entrepreneurs increases for p50/p10 as well as for p75/p25 by 4.5%. None of the GE indices, however,

a. Lorentz curves

b. Vertical distance between Lorentz curves

**Fig. 1.** Lorentz curve for workforce market income in 2005 and 2013.

**Table 2**
Income inequality statistics of market income for W, SE and ISE in 2005 and 2013.

| | 2005 | | | Total | 2013 | | | Total |
|---|---|---|---|---|---|---|---|---|
| | W | SE | ISE | | W | SE | ISE | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 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 | 0.260 | 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 |

*Notes*: 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).

reflect this increase for ISE, suggesting that the Gini is an imperfect measures when it comes to distributions marked with 'fat tails' such as is prevalent in data sets on entrepreneurs.

## 5. Results

### 5.1. Sub-group decomposition of the GE index

To further examine 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 the inequality of $GE(y; \alpha)$ into between-group and within-group parts for our sub-groups W, SE, and ISE, probing their contribution to inequality at different parts of the income distribution at different points in time. Second, using the 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.

In the first step in this analysis, instead of reporting the contributions in terms of inequality points, we calculate the percentage contributions for each term of $\tilde{GE}(yj, \alpha)$ in Eq. (15), which greatly facilitates interpretation. Dividing both sides of $\tilde{GE}\left(y_j; \alpha\right)$ with aggregate inequality, $GE(y; \alpha)$, we define $\Lambda\left(y_j; \alpha\right)$ as

$$\Lambda(y_j; \alpha) \equiv \frac{\tilde{GE}(y_j; \alpha)}{GE(y; \alpha)} = \underbrace{\frac{p_j([\frac{\sum_{k=1}^{K-1} b_{jk}\mu(x_{jk})}{\sum_{k=1}^{K-1} b_k\mu(x_k)}]^{\alpha} - 1)}{(\alpha^2 - \alpha)GE(y; \alpha)}}_{\Lambda_b(y_j; \alpha)} + w_j \underbrace{\frac{GE(y_j; \alpha)}{GE(y; \alpha)}}_{\Lambda_w(y_j; \alpha)} \sum_{k=1}^{K} s_{jk}. \tag{16}$$

**Table 3**

Percentage decomposition of $GE(y; \alpha)$ into the sub-groups W, SE and ISE for market income, 2005 and 2013.

| | 2005 | | | Total | 2013 | | | Total |
|---|---|---|---|---|---|---|---|---|
| | W | SE | ISE | | W | SE | ISE | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| GE(-1) | 0.73 | 3.858 | 0.433 | 0.906 | 0.59 | 5.203 | 0.407 | 0.829 |
| Between: $\Lambda_b\left(y_j; \alpha\right)$ | −0.11 | 0.887 | −0.363 | 0.414 | −0.107 | 1.035 | −0.415 | 0.512 |
| Within: $\Lambda_w\left(y_j; \alpha\right)$ | 75.548 | 23.292 | 0.746 | 99.586 | 66.993 | 31.616 | 0.878 | 99.488 |
| Total: $\Lambda\left(y_j; \alpha\right)$ | 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: $\Lambda_b\left(y_j; \alpha\right)$ | −0.908 | 6.131 | −3.562 | 1.661 | −0.883 | 6.877 | −4.01 | 1.983 |
| Within: $\Lambda_w\left(y_j; \alpha\right)$ | 86.419 | 9.284 | 2.636 | 98.339 | 86.472 | 8.682 | 2.862 | 98.017 |
| Total: $\Lambda\left(y_j; \alpha\right)$ | 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: $\Lambda_b\left(y_j; \alpha\right)$ | 0.939 | −4.468 | 5.227 | 1.698 | 0.903 | −4.63 | 5.672 | 1.945 |
| Within: $\Lambda_w\left(y_j; \alpha\right)$ | 87.791 | 5.726 | 4.786 | 98.302 | 88.7 | 4.768 | 4.586 | 98.055 |
| Total: $\Lambda\left(y_j; \alpha\right)$ | 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: $\Lambda_b\left(y_j; \alpha\right)$ | 0.492 | −2.387 | 2.794 | 0.898 | 0.522 | −2.754 | 3.336 | 1.105 |
| Within: $\Lambda_w\left(y_j; \alpha\right)$ | 86.355 | 4.85 | 7.896 | 99.102 | 88.106 | 3.998 | 6.792 | 98.895 |
| Total: $\Lambda\left(y_j; \alpha\right)$ | 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 $\alpha = \left\{-1, 0, 1, 2\right\}$. The inequality levels for each of the subgroups are calculated using Eq. (16) with the appropriate weights. These inequality levels differ from those presented in Table 2 that comprise raw calculations based on Eq. (1) applied to a restricted sample.

The term $\Lambda_b\left(y_j; \alpha\right)$ represents the percentage contribution to aggregate inequality from group $j$'s between part and $\Lambda_w\left(y_j; \alpha\right)$, the contribution from its within part. The total contribution from a specific group is given by $\Lambda\left(y_j; \alpha\right)$. 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 equivalent data for 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:

$$\Lambda(y; \alpha) = \sum_{j=1}^{J} \Lambda\left(y_j; \alpha\right) = \sum_{j=1}^{J} \left[\Lambda_b\left(y_j; \alpha\right) + \Lambda_w\left(y_j; \alpha\right)\right] = 1. \tag{17}$$

Based on the rules of decomposition, summing all contributions amounts to 1 (i.e., 100%). In Table 3 above, this represents the sum of the entries in the row for Total, 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 entries.

As a start, looking at $GE(-1)$, the total contributions to inequality from W, SE, and ISE are 79.06%, 20.53%, and 0.42%, 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 share of the total workforce—a mere 3.64% 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.

Turning the sensitivity parameter to higher levels of the income distribution (from an $\alpha$ of −1 to 0, and 1–2), we see that the total contributions to inequality from entrepreneurs as a group (i.e., SE and ISE together) form a U-shaped relationship. Beginning at 20.94% for $GE(-1)$, entrepreneurs' contribution to inequality shrinks to 13.24% for $GE(0)$ and 11.23% for $GE(1)$ but increases again for $GE(2)$ to 13.0%. As we increase the value of $\alpha$, there are two other patterns 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 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 the $GE(2)$ measure. For this measure, the total percentage contributions to inequality from W, SE, and ISE are 86.98%, 2.13%, and 10.87%, respectively.
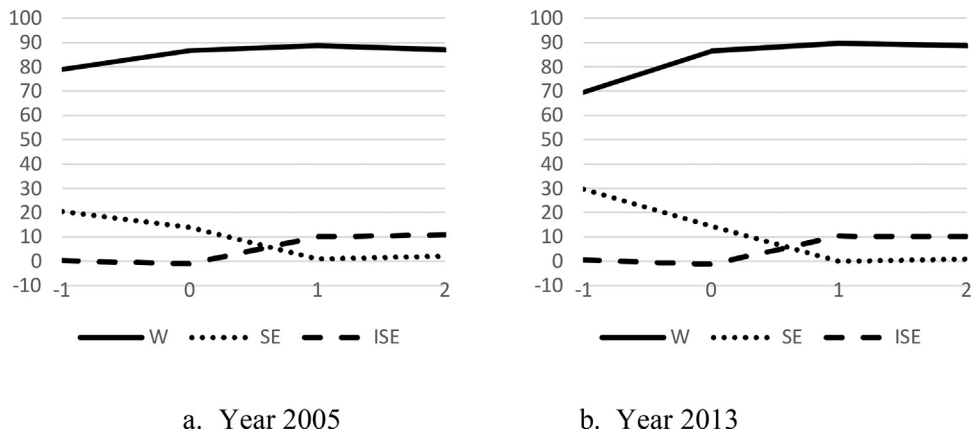
a.  Year 2005　　　　　　　b.  Year 2013

**Fig. 2.** Proportional contribution from workers (W), SE- and ISE entrepreneurs to overall inequality (100%) of market income.

This same pattern is even more pronounced in the equivalent decomposition analysis for 2013. That is, the combined effect of entrepreneurship is U-shaped, going from small to larger values of $\alpha$. A significant part of the contribution stems from SE when we focus on the bottom end 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 percentage points from 2005 to 2013. Further, this increase is almost completely due to changes in the within-inequality component for SE-entrepreneurs.

We can illustrate the above findings regarding income dynamics for W, SE, and ISE by plotting the total percentage contribution to aggregate income inequality for different values of $\alpha$. The plots are shown in Fig. 2, with the result for 2005 in (i) and for 2013 in (ii). The contribution from W to overall income inequality is here represented by a solid line, and the contributions of SE and ISE are represented by a dotted line and a dashed line, respectively. For both years, the figure illustrates how a downward shift in the inequality contribution from W, as measured by $GE(-1)$, corresponds to an upward shift for SE, whereas the ISE contribution at low income levels remains largely unchanged.

### 5.2. Estimates from regression-based decomposition of income inequality for sub-groups W, SE, and ISE

The final step in our analysis is to estimate an income model for the same sample of sub-groups as in our decomposition analysis above. Here, our empirical model is based on Eq. (12), with the dependent variable $y_{it}$ given by (non-logarithmic) market income in levels. We run separate regressions for each of the three sub-groups of workers (W), self-employed entrepreneurs (SE), and incorporated self-employed entrepreneurs (ISE). As above, the model is estimated for two cross sections: 2005 and 2013. The results are presented in Table 4 for 2005 and Table 5 for 2013, with robust OLS estimates shown in Columns 1–3 and the corresponding proportional contribution to income inequality, given by $s_{jk}$ from Eq. (13), shown in 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 attempt to separate the price effect corresponding to the estimated coefficient, which captures the effect of one additional year of education in each group, from the quantity effect, which results from the average level of education among individuals in each respective group.[17]

First, in decomposing inequality to income factors $\hat{b}_{kj}\mu\left(x_{kj}\right)$, it is important to remember that the regression approach is limited by the explanatory power of the model, given by the $R^2$. Thus, for a given sub-group, adding up the contributions ($s_{jk}$) amounts to the $R^2$. The unexplained part of the model ($1 - R^2$) equals the proportional contribution of the residual,

---

[17] Since we cannot ensure that our explanatory variables are strictly exogenous, interpretation of our outcomes should generally be done with some measure of caution. A problem with our data is of course that we do not have information on underlying ability. In other words, when estimating our models using our current data, we cannot control for self-selection and do not know to which extent the outcomes of our two entrepreneurial categories reflect organizational form (SE or ISE) or variation in the capabilities of the entrepreneurs themselves other than to the extent that our observables, primarily education levels, indeed do reflect underlying ability. To better gauge this issue, as a robustness test, we have rerun our model on a different 2005 sample of individuals for which we have additional data on results from military service conscription tests regarding intelligence and psychologic ability (i.e., very good proxies for underlying ability) as well as high school grades. The results using this much smaller dataset indicate that our study does suffer from endogeneity problems in regard to the effect of education for SE-entrepreneurs (years of education is not statistically significant for SE-entrepreneurs when controlling for innate ability in terms of our two aptitude tests and GPA), this is less of a concern for ISE-entrepreneurs or workers. In other words, using our model, we are likely much safes to infer causality in regard to ISE- as compared to SE-entrepreneurs. These robustness tests are available from the authors upon request.

**Table 4**
Regression results for market income in 2005 and percentage contribution to within-group inequality.

| | Regression estimates | | | $100 \times s_{kj}$ | | |
|---|---|---|---|---|---|---|
| | W (1) | SE (2) | ISE (3) | W (4) | SE (5) | ISE (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, and where + is significant according to a 95% confidence interval. Columns (1) to (3) gives the robust OLS results from each of the groups W, SE and ISE, estimated 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.

**Table 5**
Regression results for market income in 2013 and percentage contribution to within-group inequality.

| | Regression estimates | | | $100 \times s_{kj}$ | | |
|---|---|---|---|---|---|---|
| | W (1) | SE (2) | ISE (3) | W (4) | SE (5) | ISE (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-R² * 100) | | | | 86.953 | 97.971 | 95.222 |

given by $s_{jK}$ in Eq. (14). These estimates are provided in the bottom row of each table and show how much of the variance in income inequality cannot be attributed to the factors (explanatory variables) included in our analysis.

Also, for further clarification, $s_{jk}$ represents the raw contribution to inequality computed for a particular sub-group. Since the groups W, SE, and ISE differ in size, the proportional contribution attributed to a single factor, $s_{jk}$, for a particular group

**Table 6**
Inequality weights ($w_j$) for GE-indexindices.

| | 2005 | | | 2013 | | |
|---|---|---|---|---|---|---|
| | W | SE | ISE | | W | SE |
| | (1) | (2) | (3) | | (4) | (5) | (6) |
| $GE(-1)$ | 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 |

*Note*: Standard errors in parenthesis with significance levels *** p < 0.01, ** p < 0.05, * p < 0.1, and where + is significant according to a 95% confidence interval. 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.

must be scaled to assess its contribution to aggregate inequality, here captured by $\Lambda(y;\alpha)$ in Eq. (17).[18] The scale factor is given by

$$w_j \frac{GE(y_j;\alpha)}{GE(y;\alpha)}, \tag{18}$$

where $w_j$ is the weight defined in connection to the expression in Eq. (3) and presented in Table 6. As for the terms $GE(y_j;\alpha)$ and $GE(y;\alpha)$, they are given in Table 3 above for different values of $\alpha$.

Although the proportional contributions to income inequality for a given income factor ($s_{kj}$) is invariant to the inequality measure used, provided the basic assumptions formulated by Shorrocks (1982) are satisfied (see Appendix A), we see from the expression of the scaling factor that it is necessary to specify a given index to determine how much a factor source in one sub-group contributes to the aggregate inequality of $\Lambda(y;\alpha)$. Thus, when calculating the contribution to aggregate inequality for W + SE + ISE from some group-specific $\hat{b}_{kj}\mu(x_{kj})$, we chose the relatively common $G(y,1)$ index (Theil index) using sub-group- and year-specific values from Table 3 and inequality weights from Table 6.

Beginning with the results for inequality in 2005, Table 4 shows that almost all variables are statistically significant in the regressions, which is not surprising given the large number of observations. Common for all sub-groups, there are three explanatory variables that stand out as major contributors to inequality. These are years of education, gender (1 = Male), and age. The variable for years of education shows a marginal effect in the individual-level estimates of 24,571 SEK higher yearly income for each additional year of education.[19] Among SE- and ISE-entrepreneurs, each additional year of education translates into a higher income of 7616 SEK and 22,790 SEK, respectively.[20] Although we find almost as large a marginal effect of education on earnings among ISE as among W, inspecting the proportional contributions of education to income inequality within each of the sub-groups (Columns 4–6), we observe how education in percentage terms contributes substantially more to inequality among W (with 5.06%), whereas the corresponding number for SE- and ISE-entrepreneurs is only 0.38% and 1.67%, respectively. As much as these results inform us about the contribution to inequality of $GE(y_j,\alpha)$—that is, to the within-group inequality of each sub-group—their estimated contribution to aggregate inequality in terms of percentage—that is, out of $\Lambda(y;\alpha) = 100\%$—still needs to be computed.

For 2005, the scaling factor given in Eq. (18) for the Theil index amounts to 0.884 for W ($0.942 \times 0.197/0.210$, 0.053 for SE ($0.026 \times 0.426/0.210$), and 0.049 for ISE ($0.032 \times 0.322/0.210$) for $GE(y;1)$. Once multiplied with the weight, education among W accounts for 4.47% ($0.884 \times 5.056$) of overall within inequality given by $\Lambda(y;1)$, 0.02% ($0.053 \times 0.38$) for SE, and 0.08% ($0.049 \times 1.671$) for ISE. 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 entrepreneurial groups are small compared to W. To gauge the total contribution from one explanatory variable to aggregate inequality, we can add the weighted contributions from all groups. For years of education, this amounts to 4.57% ($4.47 + 0.02 + 0.08$) out of 100%.[21]

---

[18] The contribution to aggregate inequality only refers to the contribution to $\Lambda_w(y;\alpha)$. Hence, if an income factor should be important for the determination of the between-inequality part $\Lambda_b(y;\alpha)$, we do not account for it here. However, in Table 3, we show in terms of percentage points that $\Lambda_w(y;\alpha)$ accounts for 98.3% and 98.1% of aggregate inequality for $\alpha = 1$ in 2005 and 2013, respectively. Any contribution of a single factor in a given group to aggregate inequality $\Lambda(y;\alpha)$ that goes via $\Lambda_b(y;\alpha)$ is therefore negligible.

[19] 1 Swedish krona (SEK) was $0.11 and 0.11€ in November 2005 and $0.15 and 0.11€ in November 2013.

[20] In our robustness tests using 462,242 individuals in 2005 for which we have data on high school GPA and innate ability as measured by conscription records, we find returns to education which is 40% lower for W and ISE and almost 80% lower for SE. Comparatively, returns to one point higher GPA (on a 1–100 scale) yields a return in earning of about 1005 SEK for W, 748 SEK for SE, and 1784 SEK for ISE while returns to one point higher 'psychological ability' (on a 1–9 scale) is 8346 for W, 4135 for SE, and 8162 for ISE. It thus appears that a significant proportion of returns to education – almost half for W and ISE and more than three-quarters for SE, comes from high-ability individuals having longer education.

[21] As a comparison, the contribution from education is roughly in line with the results in Cowell and Fiorio (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.

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 85,064 SEK more on average than women. The gender income difference is the least for SE, with men earning 45,075 SEK more on average. For ISE, the gender difference is the largest and amounts to 97,081 SEK. The proportional contribution of gender to within-group inequality is, however, most pronounced for W with 3.89% compared to 0.72% for SE and 0.61% for ISE. Rescaled as contributions to aggregate within inequality $\Lambda(y; 1)$, the gender composition of each group accounts for 3.41% $(0.884 \times 3.857)$ from W, 0.04% $(0.053 \times 0.724)$ from SE, and 0.03% $(0.049 \times 0.605)$ from ISE. We can also add these percentage-point contributions to get the total contribution from gender differences to workforce inequality. Doing this, we find that gender accounts for 3.48% $(3.41 + 0.04 + 0.03)$ of total aggregate (within) income inequality.[22]

The OLS estimates in Table 4 also show that age is positively associated with income for all groups but declines for individuals with higher age from the negative age-squared effect. Age also contributes positively to inequality for both W and ISE, whereas for SE the contribution to inequality is negative and therefore likely merely a consequence of multicollinearity. Using the scaling factor, we can calculate the contribution of age for W and ISE to aggregate income inequality, which amounts to 1.05% $(0.884 \times 1.184)$ for W and 0.01% $(0.048 \times 0.204)$ for ISE, giving a total contribution from age to $\Lambda(y; 1)$ of 1.06%.

With some exceptions, we find generally larger contributions to inequality from W compared to SE and ISE. One explanation is that the $R^2$ value is higher for W, where the empirical model explains 17.5% of the variance in market incomes. In comparison, for SE and ISE, $R^2$ only amounts to 6.2% and 6.3%, 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 the inequality for both SE and ISE. This finding is in line with Åstebro and Chen (2014), who note that earnings regressions commonly produce lower $R^2$ for entrepreneurs than for wage workers. Also Hamilton (2000) notes that the highest $R^2$ in a series of earnings equations among a representative sample of US entrepreneurs was 0.07. One likely interpretation of these low $R^2$ values is that compared to workers, entrepreneurial earnings are more strongly affected by unobserved abilities (Åstebro et al., 2011). This conclusion is vindicated by the robustness test in footnote 13, which shows that the effect of education for SE entrepreneurs is not statistically significant when including proxies for innate ability.

For the groups W, SE, and ISE, the error term contributes 82.519%, 93.848%, and 93.672%, respectively, to each group's inequality. Since the residual enters the model just as any other explanatory variable, we can also compute its total weighted contribution to aggregate within-inequality (i.e., the contribution to $\Lambda(y; 1)$), which amounts to 82.51% $(0.884 \times 82.519 + 0.053 \times 93.848 + 0.049 \times 93.672)$. Thus, adding up all weighted contributions, including the residual, amounts to the aggregate within-inequality term of $\Lambda_w(y; 1) = 98.6 (100 \times (0.884 + 0.053 + 0.049))$, which leaves 1.4% for the between component $\Lambda_b(y; 1)$ to satisfy the identity in Eq. (17).

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 5 presents identical models to those of Table 4 but for the sample in 2013. 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, we compute the scaling factors of 0.889 for W $(0.942 \times 0.184/0.195)$, 0.045 for SE $(0.021 \times 0.414/0.195)$, and 0.046 for ISE $(0.034 \times 0.264/0.195)$. Our subsequent analysis and discussion focus on the scaled contribution to aggregate income inequality when reporting the differences for 2013 as compared to 2005.

Looking first at the role of education for aggregate inequality in Table 5, we find that for W, it accounts for 4.10% $(0.889 \times 4.612)$ of aggregate inequality compared to SE and ISE, for which education still contributes little to aggregate inequality, 0.004% $(0.045 \times 0.095)$ and 0.06% $(0.046 \times 1.395)$, respectively. Considering the combined effect from education across all groups in 2013, education is found to account for 4.16% $(4.10 + 0.004 + 0.06)$ of aggregate workforce inequality. Since education accounts for 4.57% 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 has increased since 2005, potentially reflecting a positive selection out of SE-entrepreneurship and into ISE-entrepreneurship for high-ability individuals following the changes in minimum equity required for forming an incorporated business in 2010.

A similar pattern is observed with regard to the contribution of gender and age to income inequality. The weighted percentage contribution from gender in 2013 is 2.50% $(0.889 \times 2.813)$ for W, 0.02% $(0.045 \times 0.42)$ for SE, 0.03% $(0.046 \times 0.601)$ for ISE, which adds up to 2.55% in total. For age, we have 1.70% $(0.889 \times 1.91)$ for W and 0.003% $(0.046 \times 0.059)$ for ISE, which together amounts to 1.703%. The contribution of age for SE remains negative.

Whereas the total contribution to workforce income inequality from gender differences in wages decreases from 3.48% in 2005–2.55% in 2013, the contribution from the workforce age distribution increases from 1.06% in 2005–1.703% in 2013.

One statistical explanation for this particular development is hinted at in the bottom row of Table 5. 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 $s_{kj}$.

---

[22] Our estimated gender income gap may at first glance seem large compared to conventional estimates. However, the main source of our larger estimate is related to gender differences in full-time employment (we do not exclude part-time workers), and we employ a generous restriction in terms of including those within the absolute top of the income distribution where men are over-represented. For gender income differentials using full-time equivalent income, see SCB (2014).

## 6. Summary and conclusions

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 in 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 an entropy-based inequality index 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 sub-group and overall inequality.

Starting with overall inequality, our data show that for the 2005–2013 period, inequality development is 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 (i.e., 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 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 income.

Regarding the specific roles of entrepreneurship for workforce inequality, the following are our main conclusions:

- 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. 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, we find that 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, forming a U-type relationship.
- Saying this, it is important to note that inequality measures emphasizing the middle of the income distribution (e.g., the Gini coefficient), by construction, do not capture this relationship well. Methodologically, researchers interested in further probing this connection are well advised to use measures that to a larger extent capture changes taking place in the tails of the distribution. Otherwise, these effects are most likely to go undetected or under-reported.
- In the occupational sub-group decomposition analysis, in which we assess the contribution of SE and ISE separately, we discover a distinct pattern suggesting that SE accounts for almost all of the bottom third in the income distribution among entrepreneurs and that ISE accounts for almost all of the top 10% in the income distribution among entrepreneurs. Hence, our results point toward a polarizing effect. Thus, although entrepreneurs as a group 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 in our data, our data show that a decrease in the numbers of self-employed (SE) over the period coincides with an increase in explained bottom-end inequality in 2013, possibly reflecting a selection out of SE for high-ability entrepreneurs when the cost of incorporating a firm was halved in 2010. As argued by Edmark and Gordon (2013), while lower-income individuals face relatively neutral incentives, higher-income entrepreneurial households may face tax incentives to incorporate their firms. Our analysis, however, suggests that none of these changes translate into any notable impact on total workforce inequality.
- Of theoretical consequence, our study demonstrates that 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, included as controls in most models of changing income inequality. This finding suggests that although entrepreneurship does not, of course, exclusively explain changing inequality in contemporary economies, it is a factor that should be explored along with other more common explanatory variables in the inequality literature.
- Noteworthy, this conclusion is further strengthened by the fact that education explains limited within-group variation in earnings among both our groups of entrepreneurs. This indicates that the share of entrepreneurs within the workforce captures variation in inequality that can be substituted only to a limited extent for standard variables in income inequality studies, such as level of education or share of more highly education individuals in the workforce. This finding also corroborates and extends results in studies from other countries that the returns to formal education among entrepreneurs are low. Economic success for entrepreneurs seems to be determined by factors not typically accounted for in traditional inequality studies.

Our new model and results come with some limitations that offer opportunities for future research. One limitation is that the results rely on non-logarithmic Mincer income regressions 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 (see p. 17 for a discussion). Very high incomes—and the resulting inequality—are thus not explained by the standard Mincer factors used in our model (cf. Bihagen et al., 2013). Future research may seek to study how entrepreneurship affects wealth levels in the economy. As Swedish wealth taxation was abolished in 2005, our microdata do not include measures of wealth. Since prior studies have suggested that entrepreneurship is a key element in understanding wealth concentration (Buera, 2009; Frid

et al., 2016; Quadrini, 1999), future research could, for example, expand on our model by examining wealth inequality instead of income inequality. Future studies would also benefit from cross-country comparisons using microdata on income and occupations. Finally, the decomposition technique that we utilize does not easily lend itself to direct causal interpretation. From a macro perspective, entrepreneurship is potentially endogenous to the extent that inequality may also affect the level of 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, 2011). Finally, we note in Table 7 in the appendix that ISE-entrepreneurs tend to run significantly larger firms than SE-entrepreneurs. Future research may seek to bridge the role of entrepreneurship in income inequality to the role of workplace reorganization and the shrinking firm-size distribution in modern economies (Bresnahan et al., 2002; Cobb and Lin, 2017). To the extent that the rise in entrepreneurship among SE- and ISE-entrepreneurship is driven by incumbent firms' outsourcing of work to subcontractors, income inequality among entrepreneurs represents a topic of relevance for several strands of economic research.

## Acknowledgements

## Appendix A. Supplementary tables and figures

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jebo.2017.11.003.

## References

Åstebro, T., Chen, J., 2014. The entrepreneurial earnings puzzle: mismeasurement or real? J. Bus. Ventur. 29 (1), 88–105.
Åstebro, T., Tåg, J., 2017. Gross, net, and new job creation by entrepreneurs. J. Bus. Ventur. Insights 8, 64–70.
Åstebro, T., Chen, J., Thompson, P., 2011. Stars and misfits: self-employment and labor market frictions. Manage. Sci. 57 (11), 1999–2017.
Akita, T., Lukman, R.A., Yamada, Y., 1999. Inequality in the distribution of household expenditures in Indonesia: a Theil decomposition analysis. Dev. Econ. 37 (2), 197–221.
Alstadsæter, A., Jacob, M., 2016. Dividend taxes and income shifting. Scand. J. Econ. 118 (4), 693–717.
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., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.
Atkinson, A.B., 2003. Income inequality in OECD countries: data and explanations. CESifo Econ. Stud. 49 (4), 479–513.
Audretsch, D.B., 2009. The entrepreneurial society. J. Technol. Transfer 34 (3), 245–254.
Autor, D.H., 2014. Skills, education, and the rise of earnings inequality among the" other 99%". 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 Econ. 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. Eur. Sociol. Rev. 29 (5), 939–954.
Blanchflower, D.G., 2000. Self-employment in OECD countries. Labour Econ. 7 (5), 471–505.
Blau, D.M., 1987. A time-series analysis of self-employment in the United States. J. Polit. Econ. 95 (3), 445–467.
Block, J.H., Wagner, M., 2010. Necessity and opportunity entrepreneurs in Germany: characteristics and earnings differentials. Schmalenbach Bus. Rev. 62, 154–174.
Borjas, G.J., Bronars, S.G., 1989. Consumer discrimination and self-employment. J. Polit. Econ. 97 (3), 581–605.
Bresnahan, T.F., Brynjolfsson, E., Hitt, L.M., 2002. Information technology, workplace organization, and the demand for skilled labor: firm-level evidence. Quart. J. Econ. 117 (1), 339–376.
Buera, F.J., 2009. A dynamic model of entrepreneurship with borrowing constraints: theory and evidence. Ann. 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. Strateg. Manage. J. 33 (12), 1411–1430.
Cobb, J.A., Lin, K.-H., 2017. Growing apart: the changing firm-size wage premium and its inequality consequences. Organ. Sci. 28 (3), 429–446.
Cowell, F., Fiorio, C., 2011. Inequality decompositions—a reconciliation. J. Econ. Inequality 9 (4), 509–528.
Cowell, F.A., 2000. Measurement of inequality. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. Elsevier, North-Holland, pp. 87–166.
Cowell, F., 2011. Measuring Inequality. Oxford University Press.
Creedy, J., Hérault, N., 2011. Decomposing Inequality and Social Welfare Changes: the Use of Alternative Welfare Metrics (Melbourne Institute Working Paper Series No. 8/11).
Davis, G.F., 2013. After the corporation. Polit. Soc. 41 (2), 283–308.
De Nardi, M., Doctor, P., Krane, S., 2007. Evidence on entrepreneurs in the United States: data from the 1989–2004 survey of consumer finances. Econ. Perspect. Q IV, 18–36.
Djankov, S., LaPorta, R., Lopez-de-Silanes, F., Shleifer, A., 2002. The regulation of entry. Quart. J. Econ. 117 (1), 1–37.
Edmark, K., Gordon, R.H., 2013. The choice of organizational form by closely-held firms in Sweden: tax versus non-tax determinants. Ind. Corp. Change 22 (1), 219–243.
Engström, P., Holmlund, B., 2009. Tax evasion and self-employment in a high-tax country: evidence from Sweden. Appl. Econ. 41 (19), 2419–2430.

Eurostat, 2008. Rev. 2—Statistical Classification of Economic Activities in the European Community. Office for Official Publications of the European Communities, Luxemburg.

Evans, D.S., Leighton, L.S., 1989. Some empirical aspects of entrepreneurship. Am. Econ. Rev. 79 (3), 519–535.

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

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. In: Solomon Polachek, W. (Ed.), Worker Well-Being and Public Policy (Research in Labor Economics) Emerald Group. , pp. 1–38.

Folta, T.B., Delmar, F., Wennberg, K., 2010. Hybrid entrepreneurship. Manage. Sci. 56 (2), 253–269.

Frid, C.J., Wyman, D.M., Coffey, B., 2016. Effects of wealth inequality on entrepreneurship. Small Bus. Econ. 47 (4), 895–920.

Goldthorpe, J.H., 2010. Analysing social inequality: a critique of two recent contributions from economics and epidemiology. Eur. Sociol. Rev. 26 (6), 731–744.

Hamilton, B.H., 2000. Does entrepreneurship pay? An empirical analysis of the returns of self-employment. J. Polit. Econ. 108 (3), 604–631.

Herranz, N., Krasa, S., Villamil, A.P., 2009. Small firms in the SSBF. Ann. Finance 5, 341–359.

Kacperczyk, A.J., Balachandran, C., 2017. Vertical and horizontal comparisons and mobility outcomes: evidence from the Swedish micro-data. MIT Sloan School Manage., Forthcoming.

Kerr, W.R., Nanda, R., 2011. Financing constraints and entrepreneurship. In: Audretsch, D., Heblich, S. (Eds.), Handbook of Research on Innovation and Entrepreneurship. , pp. 88–103.

Levine, R., Rubinstein, Y., 2017. Smart and illicit: who becomes an entrepreneur and do they earn more? Quart. J. Econ. 132 (2), 963–1018.

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

Lippmann, S., Davis, A., Aldrich, H.E., 2005. Entrepreneurship and inequality. Res. Sociol. Work 15, 3–31.

Moskowitz, T.J., Vissing-Jorgensen, A., 2002. The returns to entrepreneurial investment: a private equity premium puzzle? Nat. Bur. Econ. Res. 92 (4), 745–778.

OECD, 2015. OECD Income Inequality Data Update: Sweden. http://www.oecd.org/sweden/OECD-Income-Inequality-Sweden.pdf.

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

Poschke, M., 2013. 'Entrepreneurs out of necessity': a snapshot. Appl. Econ. Lett. 20 (7), 658–663.

Quadrini, V., 1999. The importance of entrepreneurship for wealth concentration and mobility. Rev. Income 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. In: 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., et al., 2005. Global entrepreneurship monitor: data collection design and implementation 1998–2003. Small Bus. Econ. 24 (3), 205–231.

Robinson, P.B., Sexton, E.A., 1994. The effect of education and experience on self-employment success. J. Bus. Ventur. 9 (2), 141–156.

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

Sørensen, J.B., Sharkey, A.J., 2014. Entrepreneurship as a mobility process. Am. Sociol. Rev. 79 (2), 328–349.

SCB, 2010. Lokala arbetsmarknader–egenskaper, utveckling och funktion. [Construction and use of labour market areas in Sweden]. Statistics Sweden, In Swedish.

SCB, 2014. På tal om kvinnor och män. Lathund om jämställdhet.[Women and men in Sweden 2014. Facts and figures]. SCB, Örebro.

Sanandaji, T., Leeson, P.T., 2013. Billionaires. Ind. Corp. Change 22 (1), 313–337.

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

Singer, S., Amorós, J.E., Arreola, D.M., 2014. Global Entrepreneurship Monitor 2014 Global Report. Global Entrepreneurship Research Association, pp. 1–116.

Stam, E., 2013. Knowledge and entrepreneurial employees: a country-level analysis. Small Bus. Econ. 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. Am. J. Sociol., 973–1018.

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. World Bank Econ. Rev. 19 (2), 225–261.

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

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

Wennberg, K., Wiklund, J., DeTienne, D., Cardon, M., 2010. Reconceptualizing entrepreneurial exit: divergent exit routes and their drivers. J. Bus. Ventur. 25 (4), 361–375.

Wright, M., Zahra, S., 2011. The other side of paradise: examining the dark side of entrepreneurship. Entrepreneurship Res. J. 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. J. Econ. Behav. Organ. 44 (4), 413–434.