IFAU
Institute for Evaluation of Labour
Market and Education Policy

Top earners: cross-country facts

Alejandro Badel
Moira Daly
Mark Huggett
Martin Nybom

WORKING PAPER 2017:9

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Top earners: cross-country facts[a]

by

Alejandro Badel[b], Moira Daly[c], Mark Huggett[d], and Martin Nybom[e]

June 1, 2017

Abstract

We provide a common set of life-cycle earnings statistics using administrative data from the United States, Canada, Denmark and Sweden. Three qualitative patterns are common across countries: (1) the earnings distribution above the median fans out with age, (2) the extreme right tail of the earnings distribution becomes thicker with age, and (3) the growth rate of earnings over the working lifetime is larger for groups with higher lifetime earnings. Models of top earners should account for these qualitative patterns and, importantly, for how they quantitatively differ across countries.

Keywords: Earnings, inequality, top earners, top incomes.
JEL-codes: D31, D91, H21, J31

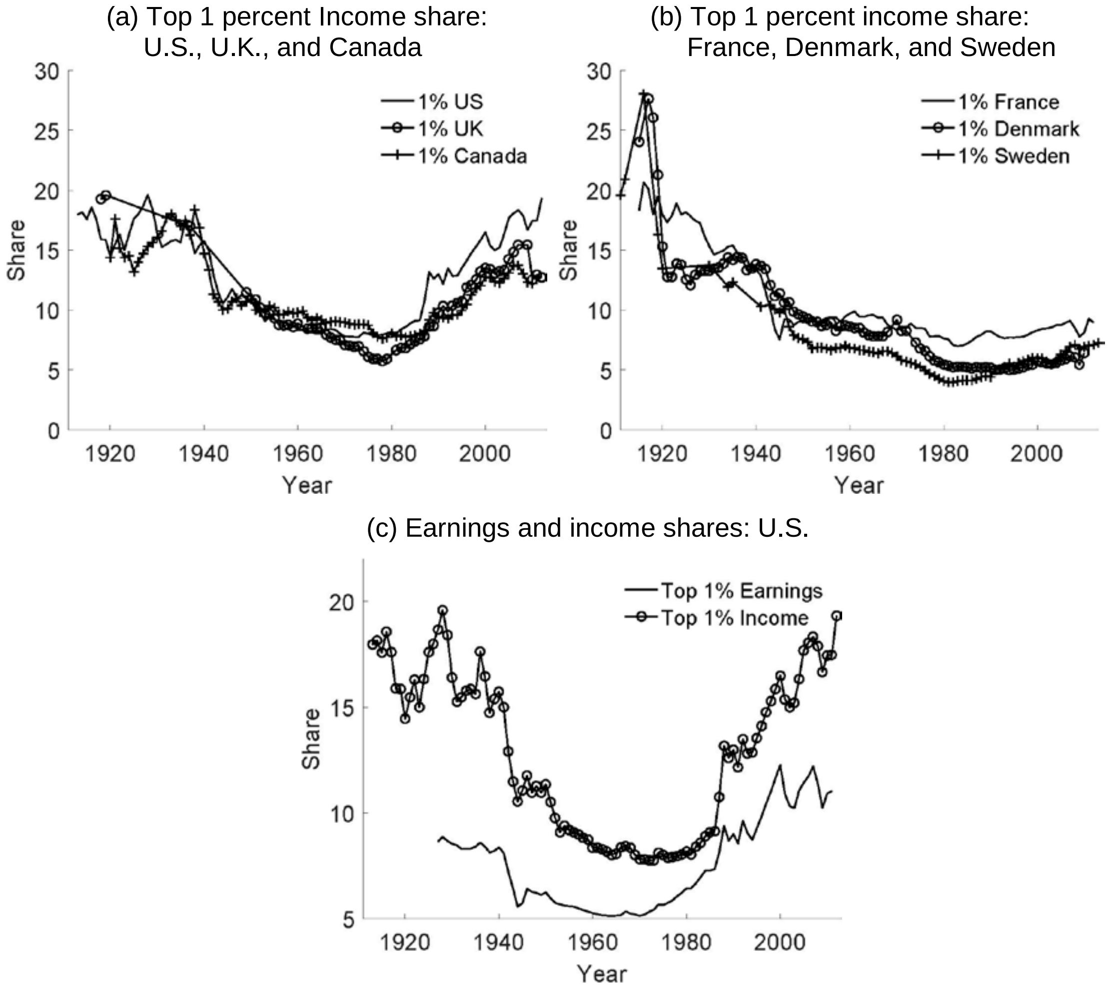
[a] We thank Daniel Waldenström and participants at the 2016 meeting of the Society of Economic Dynamics and the 2017 meeting of the Uppsala Center for Labor Studies for valuable comments.
[b] Georgetown University and Bureau of Labor Statistics; ale.badel@gmail.com
[c] Copenhagen Business School; moda.eco@cbs.dk
[d] Georgetown University; mh5@georgetown.edu
[e] IFAU; martin.nybom@ifau.uu.se

Table of contents

1 Introduction

Over the last century, the inequality of top incomes followed a U-shaped pattern over time in the US, the UK and Canada. For these countries, the recent increase in top-end inequality has become an important topic in academic, policy and media discussions. In other countries, such as Denmark, France and Sweden, income inequality decreased strongly in the first half of the twentieth century, but did not rebound afterwards. Figure 1 plots the top 1 percent income share in these countries.[6]

Figure 1: Basic top-end inequality facts



Note: Incomes come from The World Wealth and Income Database. The earnings measure for the U.S. is from Piketty and Saez (2003 updated). The income measure excludes capital gains and the earnings measure is based on wages and salaries. For the U.K., the sampling unit was changed in 1990 and there is a jump in the series that year.

Wage and salary income play a very important role in shaping these patterns. First, for the US and Canada, wage and salary income is the largest component of top incomes in recent decades (see Piketty and Saez (2003) and Saez and Veall (2005)). Second,

[6] Roine, Vlachos and Waldenström (2009) and Alvaredo, Atkinson, Piketty and Saez (2013), among others, have documented inequality patterns over the last hundred years for many developed countries including those highlighted in Figure 1.

income inequality patterns have mirrored earnings inequality patterns over time. For example, Figure 1 shows that top income and earnings shares in the US have both increased over time starting before 1980. For these reasons, proposed explanations for why top-end income inequality has increased strongly in some countries but not in others have focused on theories of earnings for top earners (see Piketty and Saez (2006)).

The goal of this paper is to document a common set of facts concerning the dynamics of the earnings distribution over the working lifetime. We focus on the US, Canada, Denmark and Sweden. While these countries are all similar in terms of economic and technological development, they are also quite different when it comes to the way their labor markets work. For example, Denmark and Sweden both have fairly high degrees of unionization and relatively compressed wage structures. Canada, and especially the US, are at the other end of the spectrum, which makes a comparison of these four countries interesting. Moreover, for all these four countries, administrative data on earnings are accessible under special arrangements. The datasets have four common features: they are large, earnings are not truncated (i.e. no top coding), they cover several decades and, importantly, they track individuals over time. These features allow us to document the top of the earnings distribution by age or by birth cohort and the earnings profiles of top earners over 30 years of their working lifetime.

We document three common cross-country facts for how male earnings vary with age. First, the earnings distribution above the median fans out with age over the working lifetime. Second, the extreme right tail of the earnings distribution becomes thicker with age over the working lifetime. Third, the growth rate of real earnings over the working lifetime is larger for earnings groups that have higher lifetime earnings and it is especially large for top lifetime earners. We document important differences in the magnitudes of these facts across countries. The patterns that we document provide empirical guidance for quantitative theoretical models aimed at understanding the distribution of earnings, income and wealth within a given country. The cross country facts provide a new challenge for quantitative theoretical work directed at understanding the underlying sources for cross-country differences in cross sectional inequality. A successful quantitative theory should be able to account for cross-country differences in cross-sectional inequality and at the same time account for the substantial quantitative differences in the three age-related earnings facts that we document.

This paper is closest to two literatures. First, there is a large literature that documents the life-cycle evolution of the distribution of earnings, wages and consumption.[7] This literature documents how summary measures of dispersion, such as the variance of log earnings, wages or consumption, vary with age based on survey data, controlling for time or cohort effects. Our work focuses on how the quantiles of the earnings distribution, including the very top of the earnings distribution, evolve with age in administrative data. Quantiles are quite useful as they determine the properties of summary measures (e.g. means, variances and measures of skewness). Much larger sample sizes and the lack of top coding allow us to address the behavior within the top 1 percent of the distribution by age. The behavior of the very top of the distribution is critical for optimal tax theory (see Piketty and Saez (2013) and Badel and Huggett (2015)) as specific statistics of the top of the distribution enter formulae that determine optimal top tax rates. Second, a recent literature uses administrative data to document the characteristics of top earners over time. See, for example, Guvenen, Kaplan and Song (2014), Guvenen, Karahan, Ozkan and Song (2015) or Bloom, Guvenen, Price, Song and Wachter (2016). We differ because we focus on life-cycle facts and document how these life-cycle facts differ across countries.

This paper is organized in four sections. Section 2 describes basic features of each data set. Section 3 documents three facts that characterize the dynamics of earnings over the working lifetime in each country. Section 4 discusses some first steps towards building models that might account for the cross-country differences that we document.

2 Data

This section describes the earnings data and the work time data for each country, the samples and some background facts.

[7] See Creedy and Hart (1979), Deaton and Paxson (1994), Storesletten, Telmer and Yaron (2004), Heathcote, Storesletten and Violante (2005), or Huggett, Ventura and Yaron (2011) among many others. An example of such a study using administrative data is Domeij and Flodén (2010), although their data set is based on a subsample of the population and their focus is on an earlier time period.

2.1 Earnings data

Our earnings data comes from records kept by government agencies for administrative purposes. This data is not publicly available and is only accessible under special arrangements that protect personally identifiable information. Except for the US, we directly access each country's micro data via the relevant statistical agency. For the US we lack access to the micro data, so we use the summary tables provided by Guvenen, Ozkan and Song (2014) and Guvenen, Karahan, Ozkan and Song (2015).

The US summary tables by Guvenen et al. (2014, 2015) are based on data from W-2 forms of wage and salary workers held by the Social Security Administration. Their earnings measure includes wages and salary, bonuses and exercised stock options. The data consists of a 10 percent random sample of males with a social security number in the period 1978-2011. Importantly, their data is not top-coded. The summary tables include minimum, maximum, mean, and various percentiles of the earnings distribution for each year.[8] They also include various percentiles of age-specific earnings distributions for persons age 25 to 55.

The earnings data for Canada comes from the Longitudinal Administrative Databank (LAD) administered by Statistics Canada. LAD is a 20 percent random sample of the Canadian population covering the period 1982 to 2013. The earnings measure we employ is total earnings from T4 slips plus other employment income.[9] T4 slips are issued by employers to the Canadian Revenue Agency and contain employment income and taxes deducted. T4 slips include wages, salaries and commissions and exercised stock option benefits. Other employment income includes tips, gratuities and director's fees not included in T4 slips.

The tax registers for Denmark are provided by Statistics Denmark. The sample period is 1980 to 2013. Over the sample period, the registers provide panel data on earnings for more than 99.9 percent of Danish residents between the ages of 15 and 70. We focus on individuals never classified as immigrants in the data. The earnings measure we employ is the sum of two variables in the registers. The first variable measures taxable wage payments and includes fringe benefits, jubilee and termination benefits and the value of exercised stock options. It excludes contributions to pension plans and ATP (the Danish labor market supplementary pension) contributions. The

[8] In particular, they include percentiles 1, 5, 10, 25, 50, 75, 90, 95, and the 99th percentile.

[9] The earnings measure is based on the code entries *T4E* and *OEI* from the LAD data dictionary.

second variable is ATP contributions which we add to construct our earnings measure. We do not add back regular pension plan contributions because this variable is not available prior to 1995.

Earnings data for Sweden are provided by Statistics Sweden. We have access to earnings data for the years 1980, 1982, and from year 1985 to year 2013. The data cover the entire Swedish population with taxable income in a given year. The earnings measure is based on taxable labor market earnings reported by the individual's employer(s) to the national tax authority.[10]

As we study top earners, realized stock options, bonus programs, share distributions, carried interest, various fringe benefits, etc are potentially important sources of earnings. It is thus important that our measures incorporate such forms of earnings and that they do so in a similar way across countries and over time. A caveat, however, is that our earnings measure for Sweden does not include the value of exercised stock options.

2.2 Work time data

To the best of our knowledge, no administrative measures of work time are available for the US or Canada. This is a shortcoming of administrative data. For Denmark and Sweden, however, administrative records of rough measures of work time are available. For these countries, we employ administrative records of work time for each worker reported by employers and match them to the earnings data described in the previous section to construct rough measures of wage rates.

For Denmark we use data from employer information sheets. The variable contains the number of days the worker was under contract with a particular employer and is used to calculate a worker's pension contribution. Multiple entries for days of work may be available for an individual in a year, with each entry corresponding to a job type.[11] Our work time measure consists of the sum of days worked across all jobs held during the year. Because of the possibility of multiple part-time jobs, our measure of days worked may exceed 365. For this reason, we cap days of work observations at 365.

[10] The earnings measure comes from Statistics Sweden variable ARBINK up to 1985 and from variable LONEINK thereafter. These measures include some labor-related benefits such as parental leave benefits and short-term sick leave benefits. Variable LONEINK includes income from closely held businesses starting in 1994.

[11] For individuals employed in November, the main November job is classified as either Primary (Type=H), assisting self-employed/owner's spouse (Type=M), employer (Type=A) or independent (Type=S). An individual may also have a secondary November job (Type=B) and or a most important not-November job (Type=3). Two additional job types were created in 2004. For comparability over time we do not include these two types in our analysis.

Finally, in 2008, the methodology used by Statistics Denmark for calculating the variable was modified. Therefore, we only employ this measure for 1980 to 2007.

For Sweden, we use data from the Wage Structure Statistics (Lönestrukturstatistiken), collected annually since 1985. These reports, which cover all public sector employees and about 50 percent of employees in the private sector, contain information such as full-time equivalent monthly wages, work time and type of labor contract. Each year, the data are collected during a particular week in September or November. In the surveyed establishments, every worker aged 18 to 65 who were employed for at least one hour during the week is covered. Our work time measure is defined as work time as percent of full time, and distributed between zero and one. A value of one corresponds to full time employment during the measurement week. For comparability with the Danish measure, which is expressed in days of work, we multiply the raw variable by 365. Appendix A2-A3 summarizes the main properties of our work time measures for Denmark and Sweden.[12]

2.3 Sample selection

We employ two types of samples. First, cross-sectional samples are used to produce statistics by year or by age and year. Second, longitudinal samples are used to track a cohort of individuals over their working life. For both types, our samples for Canada, Denmark and Sweden are designed to mimic the sample selection criteria employed in US data by Guvenen et al. (2014, 2015). Thus, we employ harmonized samples that allow cross-country comparisons.

The cross sectional sample used by Guvenen et al. (2014, 2015) includes an individual observation in a given year *t* if (i) the individual is a male age 25 to age 60, (ii) the individual's earnings are greater than a time-varying threshold denoted $\underline{e}_t^{US}$, and (iii) self-employment income does not account for more than 10 percent of the individual's earnings and does not exceed the $\underline{e}_t^{US}$ threshold. The threshold $\underline{e}_t^{US}$ employed by Guvenen et al. (2014, 2015) is defined as half of the hourly minimum hourly wage in year *t* times 520 hours. This equals the earnings of a worker with half the minimum wage working 40 hours per week for 13 weeks.

Our cross-sectional samples for Canada, Denmark and Sweden implement these three criteria. First, each sample includes only males between age 25 and age 60.

[12] The main variables are called ANSDAGE for Denmark and TJOMF for Sweden.

Second, an earnings observation is included for a given country if it exceeds a threshold ($\underline{e}_t^{CA}$, $\underline{e}_t^{DK}$, $\underline{e}_t^{SW}$), which we construct for each country. Third, we implement the self-employment income criteria (iii) described above.[13]

To obtain comparable samples, we define the minimum earnings thresholds for country $i \in \{CA, DK, SW\}$ in year *t* as fixed factors of median earnings, so that earnings observations below a factor of median earnings are excluded. Specifically, we have that

$$\underline{e}_t^i = factor_t * Median_t^i,$$

where variable $factor_t$ is common across countries and is based on the US threshold. We define it as the ratio of the US earnings threshold to US median earnings:[14]

$$factor_t = \frac{\underline{e}_t^{US}}{Median_t^{US}}.$$

For the Danish and Swedish analyses of wage rates described in Section 4, we further restrict the cross-sectional samples to the observations for which (i) a work time observation is available and (2) the work time variable takes a value above 30.

Finally, our longitudinal samples follow all individuals in the cross-sectional sample that were between 24 and 26 years old in the first year of the sample period. Thus, we follow them for more than 30 years. We also impose some additional criteria that mimic those used by Guvenen et al. (2015) and are described in Appendix A1.

2.4 Background facts

We now document a number of earnings facts based on our cross-sectional samples. Figure 2a shows that the share of earnings obtained by the top 1 percent is substantially higher in the US and Canada than in Denmark and Sweden over the sample period. In addition, top earnings shares trended upwards in the US and Canada over the sample period, mirroring the findings for top income shares in Figure 1. Top earnings shares in

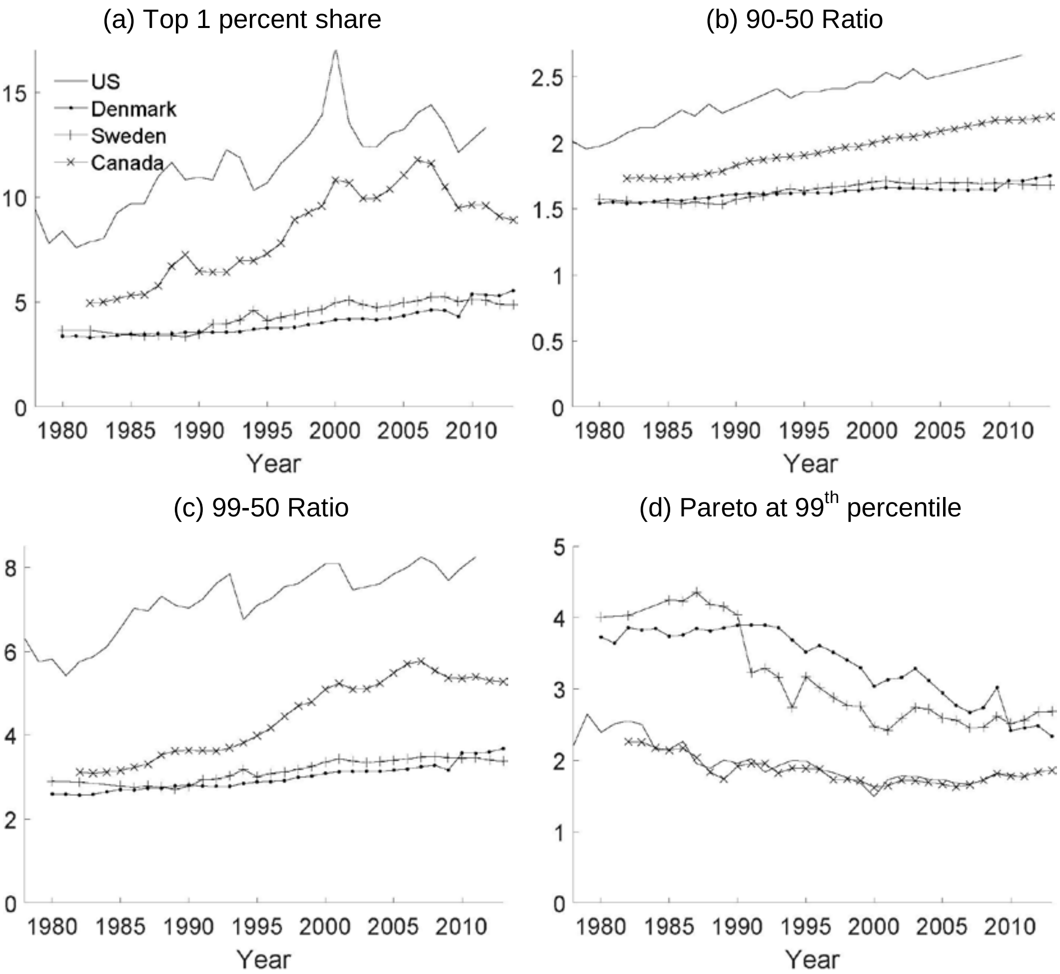
[13] For Canada, self-employment income is measured with the LAD variable SEI, which measures the sum of net income from self-employment. For Denmark, self-employment income is measured with the Statistics Denmark variable NETOVSKUDGL. For Sweden, it is measured with variable FINK which measures net entrepreneurial income.

[14] The median employed in calculating the thresholds for each country and year is calculated after imposing criteria (i) but before imposing criteria (ii) and (iii).

Denmark and Sweden also increased over the sample period but much less than in the US and Canada.

Figure 2b and 2c show that the earnings distribution above the median in Denmark and Sweden is compressed compared to the US. The 90-50 earnings ratio for Denmark and Sweden is about three quarters of the US ratio, whereas the 99-50 ratio for Denmark and Sweden is roughly half of the corresponding value for the US. Thus, compression is more severe above the 90th percentile. Dividing one half by three quarters implies that the 99-90 ratios in Denmark and Sweden have been roughly two thirds of the US 99-90 ratio.

Figure 2: Top-end earnings inequality facts



Note: Authors' calculations based on the cross-sectional samples for each country. For the U.S., the top 1 percent share and the Pareto statistic in each year are based on the assumption of a Pareto distribution within the top 1 percent and tabulated values for the 99th and 99.999th percentiles.

Figure 2b and 2c also show that earnings dispersion above the 50th percentile has increased in all countries over time. Specifically, over the sample period, the 90-50 and

99-50 earnings percentile ratios increased for all countries. Both ratios increased more in the US and Canada compared to Denmark and Sweden.

To gauge earnings inequality within the top 1 percent of the earnings distribution, we document the evolution of the Pareto statistic of earnings at the 99th percentile. This statistic is defined as $\bar{e}_{99}/(\bar{e}_{99} - e_{99})$. That is, mean earnings beyond the 99th percentile, $\bar{e}_{99}$, divided by the difference between itself and the 99th percentile, e_{99}.

The Pareto statistic is particularly important in theories of taxation of top incomes or top earnings. It enters into formulae used to determine welfare or revenue maximizing tax rates (see Piketty and Saez (2013) and Badel and Huggett (2015)). Lower values of the Pareto statistic imply a thicker upper tail in the sense that the conditional mean is a higher multiple of the threshold. Other things equal, lower values also imply a higher revenue maximizing top tax rate.

Figure 2d shows that the Pareto statistic at the 99th percentile has trended downward in all countries over the sample period. So inequality within the top 1 percent has increased in all countries. However, the increase has been more substantial in Denmark and Sweden, which started the period with much higher Pareto statistics.

3 Earnings facts

We document the evolution of the earnings distribution over the working lifetime with a focus on properties of the upper tail of the distribution.

3.1 Fact 1: earnings fan out with age

We first analyze how the earnings distribution above the median evolves with age. We control for either time or cohort effects. To do so, we calculate different earnings percentile ratios at age *j* and year *t* in the data sets described in the last section. For example, we calculate the 99-50 ratio $e_{99,j,t}/e_{50,j,t}$ for all ages *j* and all sample years *t*. We then estimate the time and age effects (α_t, β_j) or, alternatively, the cohort and age effects (γ_c, β_j) in the regressions below. An individual's birth year (i.e., cohort) is denoted *c*. Clearly, the cohort *c*, current age *j* and current year *t* are linearly related: $c = t - j$. The cohort-effects regression controls for cohort-specific effects that impact the 99-50 ratio for a cohort at any age, whereas the time-effects regression controls for time-specific effects that impact the 99-50 ratio for all age groups alive at that time. The variables D_j, D_t, and D_c are dummy variables that take the value 1 when the observation occurs at at age *j*, year *t* or cohort *c*, respectively. We employ a full set of age, year and cohort dummy variables.

$$\text{Time effects: } e_{99,j,t}/e_{50,j,t} = \alpha_t D_t + \beta_j D_j + \epsilon_{j,t}$$
$$\text{Cohort effects: } e_{99,j,t}/e_{50,j,t} = \gamma_c D_c + \beta_j D_j + \epsilon_{j,t}$$

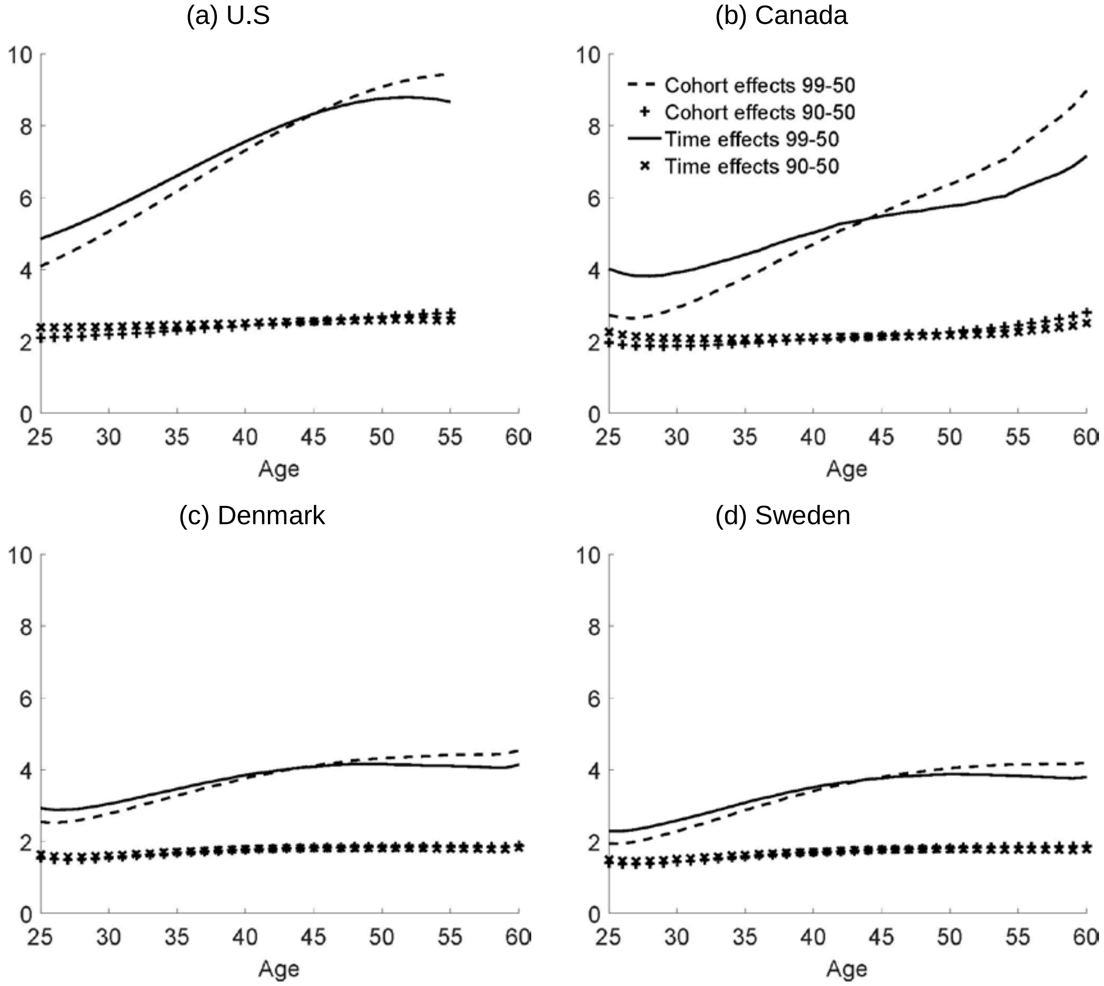
We use the estimated age effects $\hat{\beta}_j$ to describe how the 99-50 earnings percentile ratio evolves with age. We plot the estimated age coefficients adjusted by a constant $\hat{\beta}_j + k$. Constant *k* is chosen so that the height of the age profile at age 45 equals the empirical 99-50 ratio for 45 year olds in 2010 for each country.[15]

Figure 3 presents the results. The main finding is that both the 90-50 and the 99-50 ratios tend to increase with age in all countries. In this sense there is a fanning out in the top half of the distribution with respect to the median in all countries. The cohort effects view produces a more dramatic pattern of fanning out compared to the time effects view. The most striking pattern occurs for the 99-50 ratio. First, the 99-50 ratio is much larger at any age in the US and Canada compared to Denmark and Sweden. Second, the 99-50 ratio roughly doubles from age 25 to age 55 in each country under the cohort effects view. Thus, we conclude that there is growing earnings dispersion with age above the median and that this is driven by earnings beyond the 90th percentile.

Many studies have documented growth in summary measures of earnings or income dispersion with age for individuals or households based on dispersion measures such as the variance of log earnings or the Gini coefficient. The results in Figure 3 indicate that one reason why summary measures display growing dispersion with age is due to the behavior of the very top of the distribution compared to the median.

[15] For the US, the available summary tables contain data for $j \in \{25, 35, \dots, 55\}$ so estimating one age coefficient β_j for each $j = 25, 26, 27, \dots, 60$ is not possible. Therefore, we replace the age effects β_j in the regressions above with a third-order polynomial in age $P(j;\theta) = \theta_0 + \theta_{1j} + \theta_{2j^2} + \theta_{3j^3}$ and set the estimated age effects to $\hat{\beta}_j = P(j;\hat{\theta})$ where $\hat{\theta}$ are the estimated polynomial coefficients.

Figure 3: Percentile ratios: 90-50 and 99-50 ratios by age



Note: The figure plots the estimated age coefficients after adding a vertical shift term so each figure is normalized to equal the data value of the 99-50 ratio or 90-50 ratio at age 45 in year 2010.

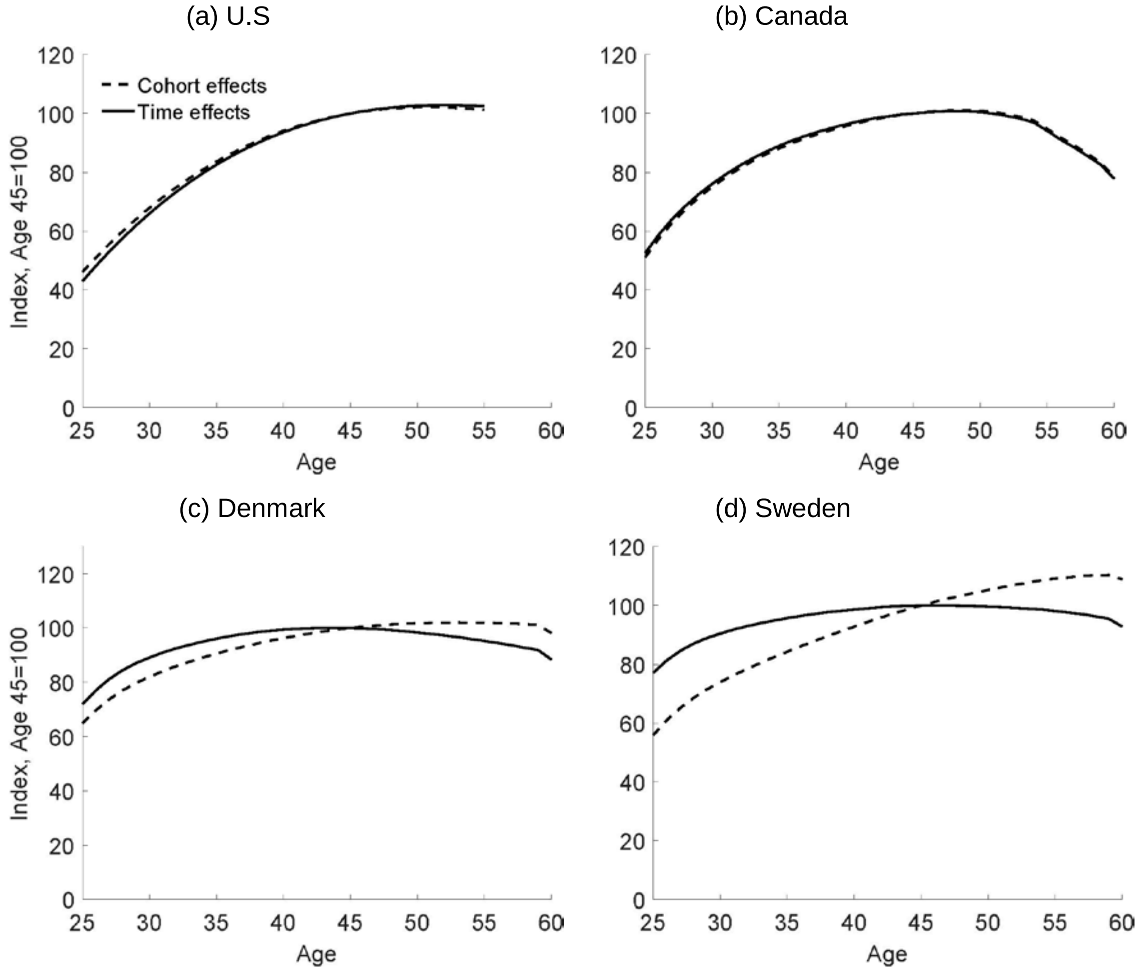
To put these results into perspective, it is useful to characterize how real median earnings evolve with age.[16] Figure 4 provides the results of regressing real median earnings on age and time effects or age and cohort effects. Median earnings display a hump-shaped pattern with age in each country. Based on survey data, many previous studies have documented that profiles of central moments of male earnings or wage rates by age are hump-shaped (or concave) over the working life.[17]

Figure 4 shows that median earnings in the US and Canada approximately double with age from age 25 to age 50. This holds regardless of whether one controls for time or for cohort effects. In contrast, for Denmark and Sweden the time effects view implies

[16] State CPI measures employed in each country are used to calculate real earnings.
[17] For example, the Review of Economic Dynamics special issue on Cross Sectional Facts for Macroeconomists in 2010 covers 9 countries and Lagakos et al. (2016) covers 18 countries.

that the median earnings profile is flatter with less than a doubling of median earnings. Focusing on the time effects view across countries reveals substantial differences in the timing of the peak of the earnings profile. For the US and Canada median earnings peak near age 50, whereas for Denmark and Sweden the peak occurs in the early 40's.

Figure 4: Median earnings by age

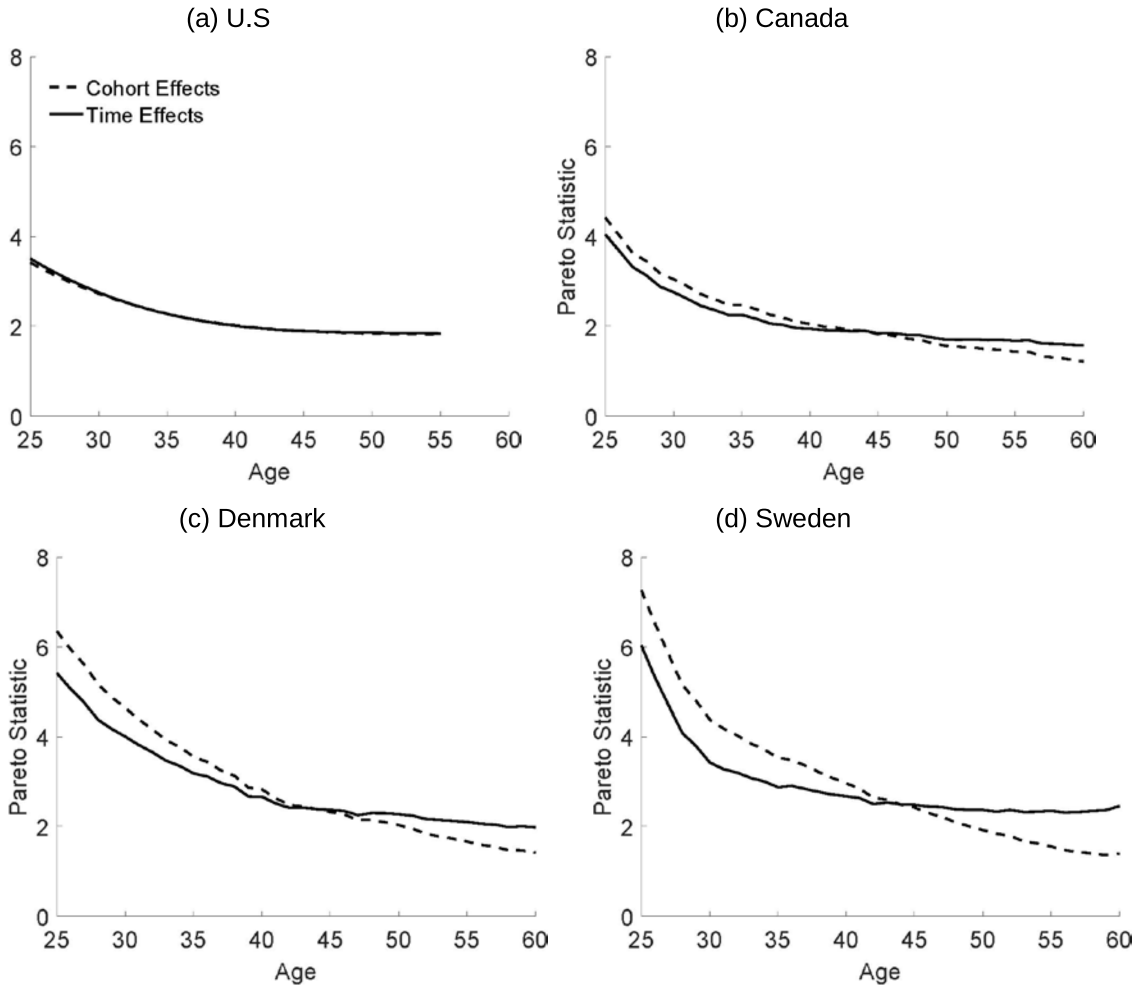


Note: The figure plots the estimated age coefficients after adding a vertical shift term so each figure is normalized to equal 100 at age 45.

3.2 Fact 2: the upper tail becomes thicker with age

Next we analyze how the Pareto statistic at the 99-th percentile evolves with age. This is a way to describe how the thickness of the upper tail of the earnings distribution evolves with age. To do so, we run the two basic regressions from the last section after replacing ratios of earnings percentiles with the Pareto statistic for each age-year pair. Figure 5 shows the results.

Figure 5: Pareto statistic at the 99th percentile by age



Note: The figure plots the estimated age coefficients after adding a vertical shift term so each figure is normalized to equal the data value of the Pareto statistic at age 45 in the year 2010.

We find that the Pareto statistic in the upper tail declines with age in all countries, although the rate of this decline decreases with age. This holds in both the time and cohort effect regressions. Thus, the upper tail of the earnings distribution becomes thicker with age in each country in the sense that mean earnings beyond this threshold is a growing multiple of the threshold with age, in turn reflecting increasing earnings dispersion with age beyond the 99th percentile. To the best of our knowledge, this fact has not been documented in the existing literature for a wide collection of countries.

It is interesting to compare the Pareto statistic in different age groups to the Pareto statistic in cross-sectional data previously documented in Figure 2. For the US, the Pareto statistic at the 99th percentile in cross-sectional data is below two in the last two decades of the sample period. It is below two in the US in Figure 5 for age groups above age 40 while is above two for age groups below age 40. This suggests that the cross-sectional Pareto statistic for the US is largely determined by the earnings distribution for males age 40 and beyond. The same patterns hold in Canadian data. Thus, the cross-sectional Pareto statistic seems to be driven by the tail properties holding for older male earners in both countries.

3.3 Fact 3: high lifetime earners have the largest earnings growth

We now use the longitudinal feature of each data set. For each male in the longitudinal sample, we compute lifetime earnings LE as follows: $LE^i = \sum_{t \in T} \frac{min\{e_t^i, \underline{e}_t\}}{p_t}$, where e_t^i is individual i's nominal earnings in year t, $\underline{e}_t$ the minimum earnings threshold used to construct the cross-section sample, p_t is a country price index in year t, and T is the set of years for which earnings observations are available.[18] We then sort males in the longitudinal sample into 100 bins based on the percentiles of the lifetime earnings distribution. Bin 100 corresponds to males with lifetime earnings above the 99th percentile, whereas bin 1 corresponds to males with lifetime earnings below the 1st percentile of lifetime earnings. The Appendix describes the construction of the longitudinal data sets.

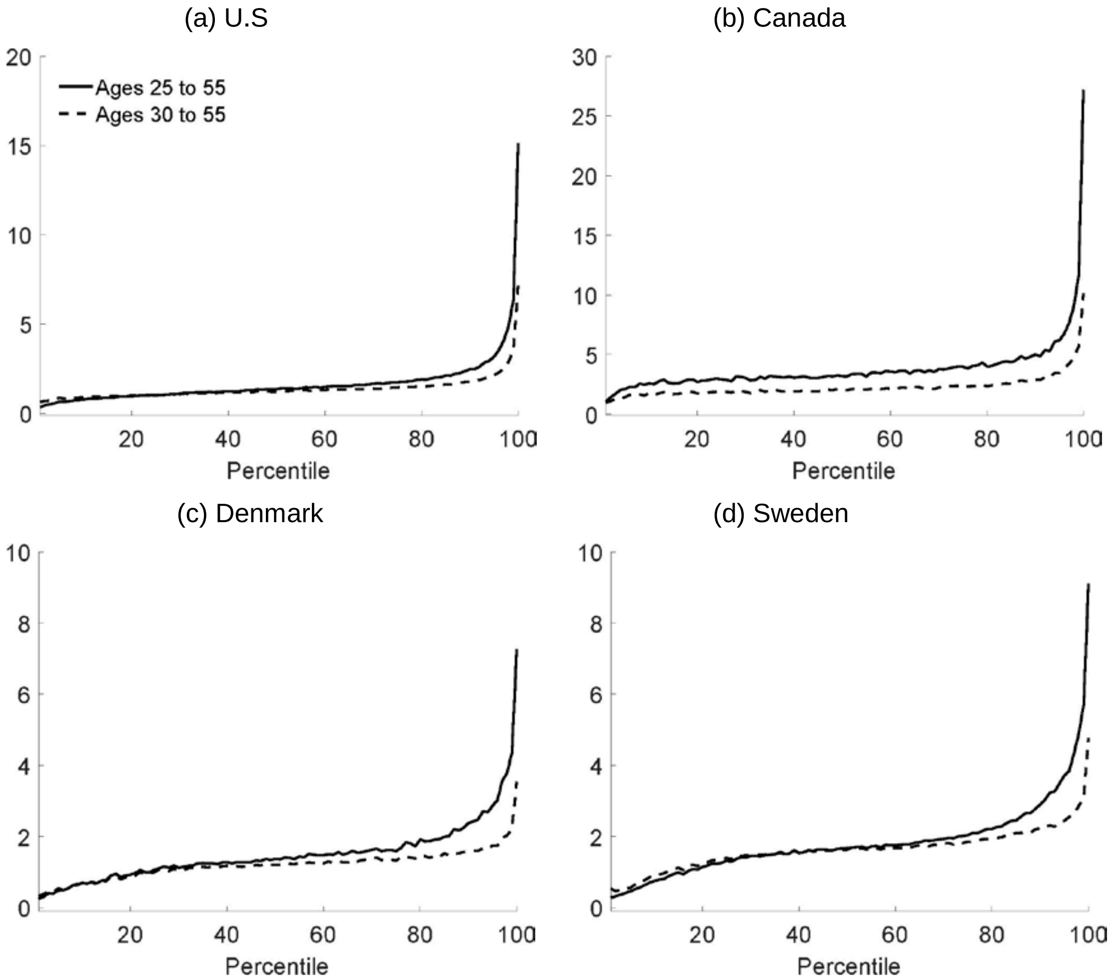
Figure 6 contains two plots for each country. It plots the ratio of mean real earnings at age 55 to mean real earnings at age 25 for individuals sorted by lifetime earnings bin, and the ratio of mean real earnings at age 55 to mean real earnings at age 30. In both plots the grouping of individuals into lifetime earnings bins is unchanged. Thus, for a given country, the two plots differ only insofar as there is growth in real mean earnings for the group from age 25 to age 30.

Figure 6 documents that earnings growth is greater for groups with larger lifetime earnings. While this is hardly an unexpected result in a qualitative sense, the quantitative differences in earnings growth are quite remarkable. The fact that the highest lifetime earnings groups (i.e., groups in lifetime earnings bins 96-100) have a *vastly* higher earnings growth rate than those with lifetime earnings close to the median (i.e., those in bin 50) is noteworthy. The top lifetime earnings bin in the US and Canada have a 13-15 fold increase in earnings from age 25 to 55. The top lifetime earnings bin in Denmark and Sweden have a 7-9 fold increase in earnings from age 25 to 55. Thus, there are large, systematic differences in group earnings growth rates over the working

[18] The set T^{US} is based on years 1979-2011, T^{CA} is based on years 1982-2013, T^{DK} is based on years 1980-2013, and T^{SW} is based on the years 1980, 1982, and 1985-2013.

lifetime particularly at the top. The large differences at the top imply that in each country top earners tend to become top earners late in the working lifetime. We anticipate that Fact 3 will be particularly useful in empirically disciplining quantitative theories of top earners. We conjecture that theories built on purely temporary sources of earnings variation will struggle to produce Fact 3.

Figure 6: Earnings growth by lifetime earnings group



Note: The figure plots the ratio of mean group earnings at age 55 to mean group earnings at age 25 as well as the ratio of mean group earnings at age 55 to mean group earnings at age 30 for groups sorted by percentile of lifetime earnings. US data is taken directly from Guvenen, Karahan, Ozkan and Song (2015). The results for all the other countries are based on our calculations from country longitudinal data.

4 Discussion

This paper documents three life-cycle earnings facts for Canada, Denmark, Sweden, and the United States. A natural question is then what accounts for these qualitative patterns in each country and their quantitative differences across countries? While answers to these big-picture questions are beyond the scope of this paper, we will outline some first steps for how one might begin to address them.

4.1 The policy-difference hypothesis

The finding that the three earnings facts from Section 3 are qualitatively the same across countries suggests that a common theoretical framework may be useful for interpreting them. Within such a framework, one might then hypothesize that policy differences are an important source of the cross-country differences. The policy-difference hypothesis seems attractive since many policies (e.g., labor and capital income taxation, the structure of social security systems, and labor-market regulations) are based on rules and thus subject to quantification.

The policy-difference hypothesis is not new. We are simply suggesting that it can be used to account for differences in Facts 1-3. It is a central hypothesis put forward to explain the patterns in top income shares like those documented in Figure 1 of this paper. For example, Alvaredo, Atkinson, Piketty and Saez (2013, p. 5) state:

"To us, the fact that high-income countries with similar technological and productivity developments have gone through different patterns of income inequality at the very top supports the view that institutional and policy differences play a key role in these transformations. ... The most obvious policy difference - between countries and over time - regards taxation ..."

4.2 Exogenous productivity frameworks

What type of framework is potentially useful for addressing the policy-difference hypothesis? We try to answer this question in a negative way. We provide evidence that indicates severe limitations of one extremely popular framework for interpreting top-end inequality facts.

One popular framework assumes that agents experience exogenous, idiosyncratic variation in labor productivity. To be specific, consider models in which earnings $e = w\theta l$ in a period is the product of a common wage w, individual productivity θ and a work time decision l. At a partial equilibrium level, such models are widely used in labor economics and public economics.[19]

[19] See Mirrlees (1971), Heckman (1976), MaCurdy (1981), and Saez (2001).

Equilibrium models with idiosyncratic productivity shocks have been the work-horse models in modern macroeconomics. See Heathcote, Storesletten and Violante (2009) for a review. They have been applied to interpret the US distribution of earnings, consumption, and labor hours over the life cycle (see Kaplan (2012)) and changes in US earnings, income and wealth distributions over time (see Heathcote, Storesletten and Violante (2010) and Kaymak and Poschke (2016)). They are widely used to analyze tax reforms (see Guner, Lopez-Daneri and Ventura (2015) and Kindermann and Krueger (2015)).

Figure 7 calculates ratios of earnings percentiles in the US, Denmark and Sweden in cross-sectional data. It shows that the 99-50 and 90-50 earnings percentile ratios are much larger in the US in each year compared to Denmark and Sweden. Thus, there is substantial earnings compression above the median in Denmark and Sweden compared to the US. These earnings facts do not rule out this class of models, under the working assumption that idiosyncratic productivity risk is the same across countries, because the behavior of work time is so far unrestricted by theory or by measurement.

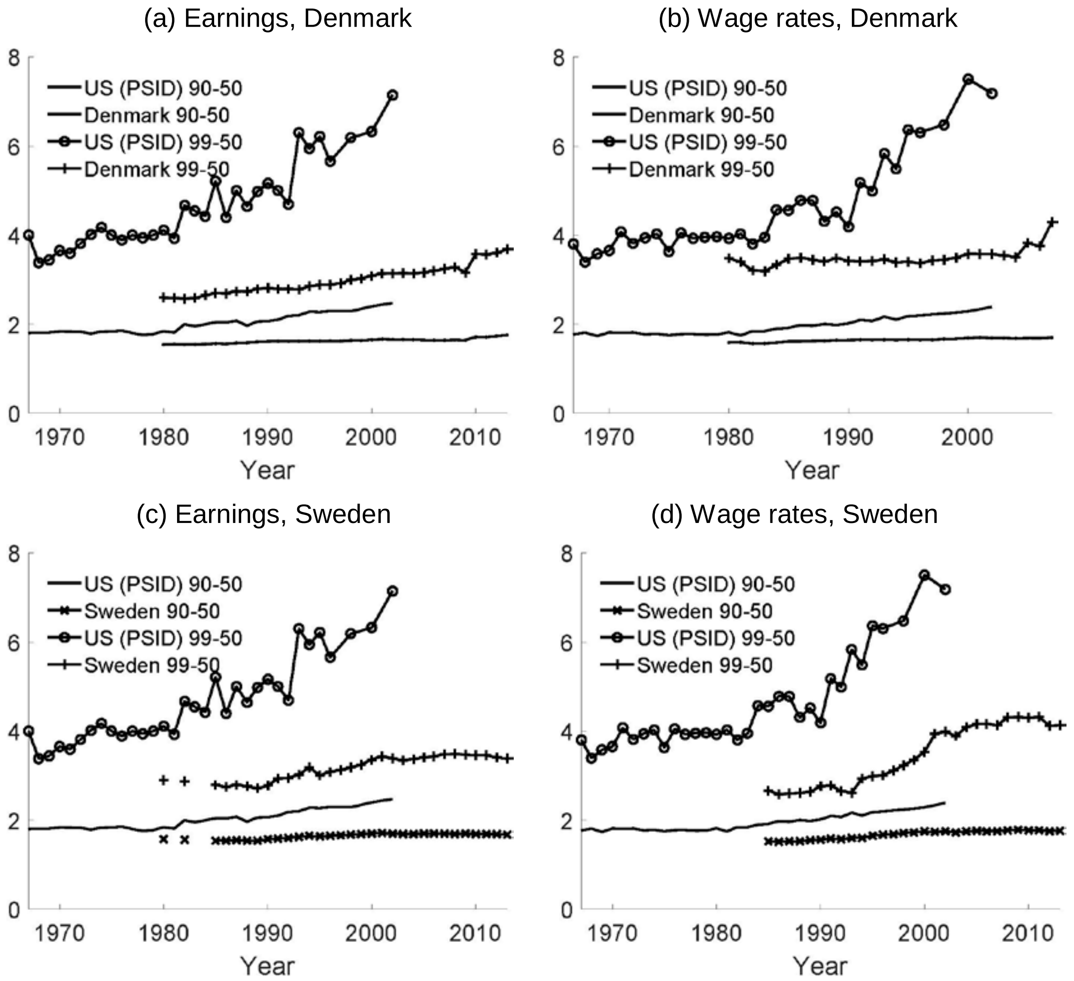
Now assume for the moment that the work time measure in the theory is captured by the data measure and that earnings are measured without error. Wage rates can then be constructed as measured earnings divided by measured work time. These assumptions together with the theory then imply that the wage rate for individual i at time t in a given country is $wage_{it}^{US} = w_t^{US}\theta_{it}^{US}$, $wage_{it}^{DK} = w_t^{DK}\theta_{it}^{DK}$, and $wage_{it}^{SW} = w_t^{SW}\theta_{it}^{SW}$. Ratios of wage rate percentiles in each country are then determined only by the distribution of individual productivity in each country. Thus, under the theory, wage percentile ratios in a cross section can differ only because the exogenous distribution of productivity differs across countries. The policy-difference hypothesis would then have no role to play in endogenously producing cross-country differences in wage percentile ratios.

Figure 7 also calculates ratios of wage rate percentiles in the US, Denmark and Sweden. The wage rate measure for the US is based on PSID data and is male earnings divided by reported work hours.[20] The wage rate measure for Denmark and Sweden is the earnings measure divided by the time measure discussed in Section 2. Figure 7 shows that the 99-50 and 90-50 wage percentile ratios are all larger in the US compared

[20] We employ PSID data because the Social Security data set does not contain a measure of work time.

to Denmark or Sweden in each year. Thus, there is relatively more wage rate compression above the median in Denmark and Sweden compared to the US.

Figure 7: Percentile ratios of earnings and wage rates



Under the theory and related assumptions articulated above, there are only a few possible explanations for the large cross-country differences in wage rate ratios. One possibility is that idiosyncratic productivity follows the same process across countries for individual males but that demographic age weights differ across countries and these demographic differences account for all of the observed cross-country differences. This seems unlikely as the US 99-50 wage ratio nearly doubled from 1990 to 2000 and demographic weights vary gradually across years. Another possibility is that the US ratios of wage rage percentiles in Figure 7 are based on a small sample. Thus, perhaps the US is above Denmark and Sweden due to sampling variability and the small US sample. Figure 7 shows that this is unlikely as the one standard error band for the US in

Figure 7 is for all years above the data values for Denmark and Sweden for both the 90-50 and the 99-50 wage rate percentile ratio.[21]

4.3 Models with endogenous productivity

Going outside of the exogenous productivity theory described previously, there are many possibilities to account for differences in the distribution of wage rates across countries or within countries across time. Fortin and Lemieux (1997) show in US data that minimum wage increases from 1979 to 1988 were associated with a substantial compression of the bottom of the US male wage rate distribution. They also argue that changing unionization may have impacted the middle of the wage rate distribution, given that in the data union workers are found in the middle of the distribution.

We doubt that either of these possibilities is key to understand the movements in the 99-50 and 90-50 wage rate ratios documented in Figure 7. In particular, the US 99-50 ratio almost doubles from 1980 to 2000, while the US 90-50 ratio increases by only about 20 percent. Thus, an explanation cannot be based on forces that only move the median wage or the lower parts of the wage rate distribution. An explanation has to account for differential growth of wage rates within the top 10 percent of the distribution, as documented in Figure 8.

We view the finding of substantial differences in cross-country earnings and wage rate dispersion above the median as a strong argument for considering models where labor productivity is endogenous, particularly when the analysis is concerned with the top of the distribution. At a minimum, what is needed is an endogenous productivity model where (1)-(3) below hold:

(1) The 99th percentile of earnings grows relative to the median for a cohort as the cohort ages.

(2) The earnings distribution beyond the 99th percentile becomes thicker for a cohort as the cohort ages.

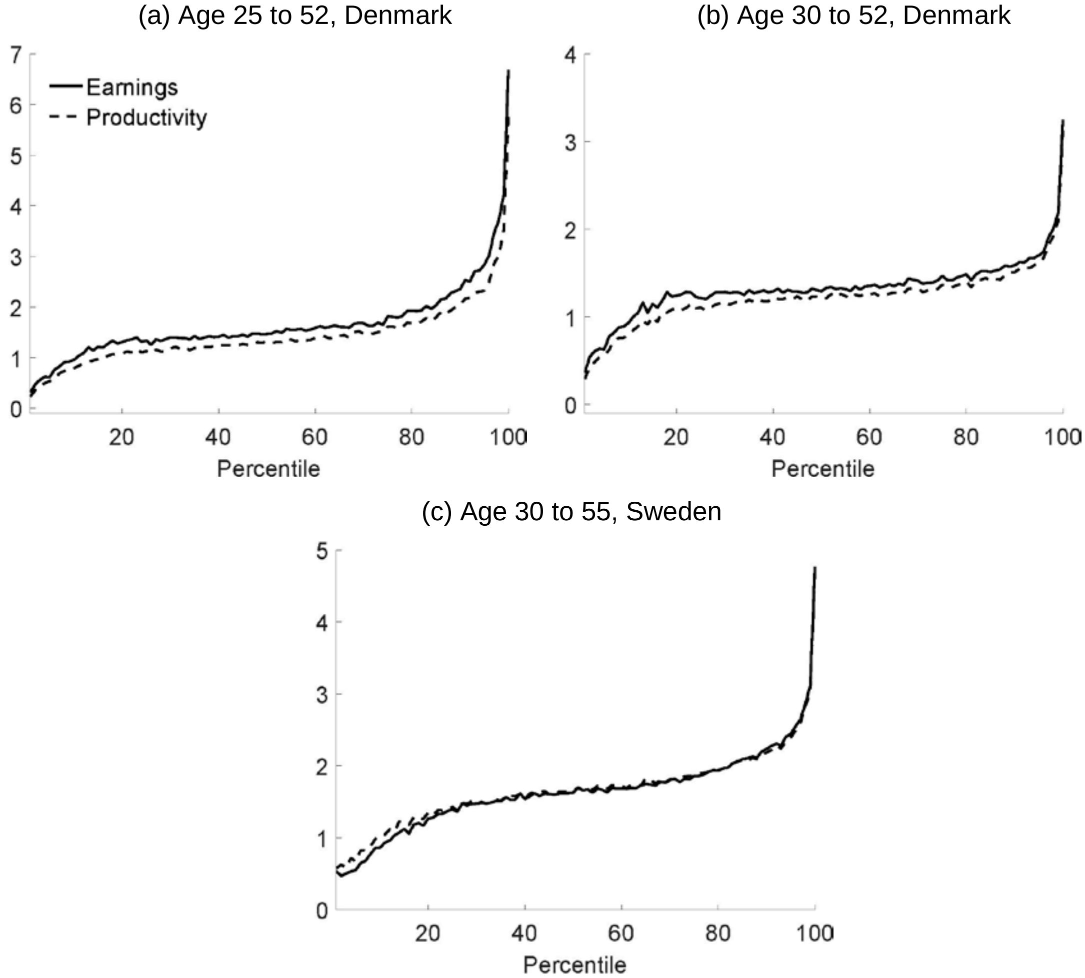
[21] Another potential explanation concerns the technical structure of the tax system. Sweden and Denmark both have a dual system in which labor and capital incomes are taxed separately and at different rates. Moreover, the capital income tax is essentially flat, while taxes on labor income are progressive. For top earners, marginal tax rates on labor income thus tend to be substantially higher, providing incentives to have marginal income increases taxed as capital. In Canada and the United States, on the other hand, labor and capital incomes are typically taxed at the same rate. The quantitative differences across countries could thus to some extent be explained by differences in such incentives. However, since also top shares in total income grew much faster in Canada and the US in recent decades (see Figure 1), this cannot explain more than some of the cross-country differences. We thank Henry Ohlsson for pointing this out.

(3) Top lifetime earners have much higher earnings growth rates than other groups.

For example, models with endogenous human capital accumulation and learning ability differences can produce the three facts above. Badel and Huggett (2014) provide such quantitative model that produces the magnitudes measured in US data.[22] A key ingredient in this model is that learning ability differs across agents. Agents with high learning ability have very steep mean earnings profiles other things equal. The differences in earnings profiles are driven by differences in the accumulation of skills over the lifetime. Such an explanation would be consistent with the results in Figure 6 and also Figure 8 below, which considers wage rates rather than earnings. While such systematic differences in earnings (or wage) growth rates should be a key feature within a model that produces the earnings facts above, other types of mechanisms are of course also possible. Examples include matching models in which match qualities improve heterogeneously over the lifecycle, or models with employer learning.

[22] Huggett, Ventura and Yaron (2011) is the closest precursor to Badel and Huggett (2014). Guvenen, Kuruscu and Ozkan (2014) analyze a quantitative human capital model like that in Huggett, Ventura and Yaron (2011). They show that their model produces cross-country differences in 90-50 earnings percentile ratios when the income tax system and some features of social security systems are set to country-specific values.

Figure 8: Earnings and productivity growth by lifetime earnings group



Note: The figure plots the ratio of mean group earnings across ages and the ratio of productivity across ages for groups sorted by percentile of lifetime earnings. Productivity at a given age is measured as group mean earnings at that age divided by group mean work time at that age.

References

Alvaredo, F., Atkinson, A., Piketty, T. and E. Saez (2013), The Top 1 Percent in International and Historical Perspective, Journal of Economic Perspectives, 27, 3-20.

Alvaredo, F., Atkinson, A., Piketty, T. and E. Saez, The World Wealth and Income Database, http://topincomes.g-mond.parisschoolofeconomics.eu/.

Badel, A. and M. Huggett (2014), Taxing Top Earners: A Human Capital Perspective, Unpublished manuscript.

Badel, A. and M. Huggett (2015), The Sufficient Conditions Approach: Predicting the Top of the Laffer Curve, Unpublished manuscript.

Creedy, J. and P. Hart (1979), Age and the Distribution of Earnings, Economic Journal, 89, 280-93.

Deaton, A. and C. Paxson (1994), Intertemporal Choice and Inequality, Journal of Political Economy, 102, 437-67.

Domeij, D. and M. Flodén (2010), Inequality Trends in Sweden 1978-2004, Review of Economic Dynamics, 13, 178-208.

Guner, N., Lopez-Daneri, M. and G. Ventura (2015), Heterogeneity and Government Revenues: Higher Taxes at the Top?, Unpublished manuscript.

Guvenen, F., Kuruscu, B. and S. Ozkan, Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis, Review of Economic Studies, 81, 818-50.

Guvenen, F., Ozkan, S. and J. Song (2014), The Nature of Countercyclical Income Risk, Journal of Political Economy, 122, 621-60.

Guvenen, F., Kaplan, G. and J. Song (2014), How Risky are Recessions for Top Earners?, American Economic Review Papers and Proceedings, 104, 1-6.

Guvenen, F., Kaplan, G. and J. Song (2015), The Glass Ceiling and The Paper Floor: Gender Differences Among Top Earners, 1981-2012, Unpublished manuscript.

Guvenen, F., Karahan, F., Ozkan, S. and J. Song (2015), What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?, Unpublished manuscript.

Heathcote, J., Storesletten, K. and G. Violante (2005), Two Views of Inequality Over the Life-Cycle, Journal of the European Economic Association, 3, 543-52

Heathcote, J., Storesletten, K. and G. Violante (2009), Quantitative Macroeconomics with Heterogeneous Households, Annual Review of Economics, 1, 319-354.

Heathcote, J., Storesletten, K. and G. Violante (2010) The Macroeconomic Implications of Rising Wage Inequality in the United States, Journal of Political Economy, 118, 681-722.

Heathcote, J, Perri, F. and G. Violante (2010), Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006, Review of Economic Dynamics, 13, 15-51.

Huggett, M., (1996) "Wealth distribution in life-cycle economies," Journal of Monetary Economics, Elsevier, vol. 38(3), pages 469-494, December.

Huggett, M., Ventura, G. and A. Yaron (2011), Sources of Lifetime Inequality, American Economic Review, 101, 2923-54.

Kaplan, G. (2012), Inequality and the Lifecycle, Quantitative Economics, 3, 471-525.

Kaymak, B. and M. Poschke (2016), The Evolution of Weath Inequality Over Half a Century: The Role of Taxes, Transfers and Technology, Journal of Monetary Economics, 77, 1-25.

Kindermann, F. and D. Krueger (2015), High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk, Unpublished manuscript.

Lagakos D., Moll, B., Porzio, T., Qian, N. and T. Schoellman (2016), Life-Cycle Wage Growth Across Countries, Unpublished manuscript.

MaCurdy, T. (1981), An Empirical Model of Labor Supply in a Life-Cycle Setting, Journal of Politial Economy, 89, 1059-85.

Mirrlees, J. (1971), An Exploration into the Theory of Optimum Income Taxation, Review of Economic Studies, 38, 175-208.

Piketty, T., and E. Saez (2003), Income Inequality in the United States, 1913-1998, Quarterly Journal of Economics, 118, 1-39.

Piketty, T., and E. Saez (2006), The Evolution of Top Incomes: A Historical and International Perspective, American Economic Review PP, vol.96, no 2, 2006, p. 200-205.

Piketty, T., and E. Saez (2013), Optimal Labor Income Taxation, Handbook of Public Economics, Volume 5. editors A. Auerbach, R. Chetty, M. Feldstein and E. Saez, Elsevier.

Roine, J., Vlachos, J. and D. Waldenstrom (2009), The Long-run Determinants of Inequality: What Can We Learn From Income Data?, Journal of Public Economics, 93, 974-88.

Saez, E. (2001), Using Elasticities to Derive Optimal Income Tax Rates, Review of Economic Studies, 68, 205-29.

Saez. E. and M. Veall (2005), The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence, American Economic Review, 95, 831-49.

Storesletten, K., Telmer, C. and A. Yaron (2004), Consumption and Risk Sharing Over the Life Cycle, Journal of Monetary Economics, 51, 609-33.

Appendix

A.1 Longitudinal samples

For Canada, our raw data consists of all individuals in the LAD dataset. The LAD is a 20 percent random subsample from the Canadian population that either filed a T1 form or received Canadian child benefits in any year since 1982 and had a social insurance number.[23] For Denmark we use tax registry data kept by Statistics Denmark. In Denmark, all residents are included in the tax registry unless they are not alive. For Denmark we start off by keeping only the population that was never classified as immigrant in 1980-2013. For Sweden we use tax registers kept in the Income and Taxation Register of Statistics Sweden. These data come from the Swedish Tax Agency, which collects information from virtually all persons who are Swedish citizens or hold a residence permit.

We construct longitudinal samples for Canada, Denmark and Sweden. These three samples mimic the construction of the US longitudinal sample described in Guvenen, Karahan, Ozkan and Song (2015). The sample period is 1982-2013 for Canada, 1980-2013 for Denmark and is 1980, 1982 and 1985-2013 for Sweden. Thus, the sample period for each country spans a horizon of more than thirty years.

Our longitudinal sample for each of these three countries contains all individual histories that satisfy conditions 1-4 below. The following notation is employed: e_t^i is individual i's nominal earnings, $\underline{e}_t$ is a minimum earnings threshold, and se_t^i is individual i's self-employment income. The conditions are: (1) the individual is male with age 24, 25 or 26 in the first year of the sample period; (2) the individual has a valid non-missing earnings observation in every year of the sample period; (3) There are more than 15 years for which $e_t^i > \underline{e}_t$; and (4) there are less than 9 years for which $se_t^i > max\{\underline{e}_t, 0.1 * e_t^i\}$.

We now provide a brief discussion of the specifics of imposing conditions 1-4 in the longitudinal samples for each country. Condition 1 is straightforward to implement. Individuals in the first year of the longitudinal sample are viewed to be in the "age 24", "age 25" or "age 26" group in the first sample year. All properties of mean earnings for groups by age are understood to be for the central age within the group. Condition 3 is straightforward to implement in each country. We simply employ the threshold used in

[23] A person who is sampled in a particular reference year is also selected in all other available years.

the construction of each cross-sectional sample. We implement condition 4 in Canada and Denmark by using the self-employment income measure described in section 2 and employed in the construction of the cross-sectional sample.

A.2 PSID data

To document wage facts for the United States we use data from the Panel Study of Income Dynamics. We use data from the Heathcote, Perri and Violante (2010) (HPV) PSID files provided by the Review of Economic Dynamics. The data comes from the PSID 1967 to 1996 annual surveys and from the 1999 to 2003 biennial surveys. We employ earnings and wages from HPV Sample C. Sample C includes male heads of households with age between 25 and 60, at least 260 hours of work, and wage measure above half of minimum wage. The head of household is the oldest working male in the household.[24] The total number of wage observations varies approximately from 1300 to 2800 across sample years. The annual earnings variable provided by HPV includes all income from wages, salaries, commissions, bonuses, overtime and the labor part of self-employment income. Annual hours of work is defined as the sum total of hours worked during the previous year on the main job, on extra jobs and overtime hours. This variable is computed using information on usual hours worked per week times the number of actual weeks worked in the last year.

A.3 Pareto statistic from SSA data

Pareto statistics at the 99th percentile are not provided by Guvenen et al. (2014, 2015). Based on the statistics provided, we estimate the Pareto statistics for the US in two different ways. First, for the Pareto statistics depicted in Figure 2d, we use the 99th and 99.999th percentiles of earnings, provided by Guvenen et al. (2014) for each sample year, to estimate the coefficient of a Type-I Pareto distribution for earnings above the 99th percentile. Such coefficient is the Pareto Statistic. For the Pareto statistic at the 99th percentile by age group and year used to create the life cycle profiles in Figure 5, we employ the method described in Badel and Huggett (2014), which uses the 95th and 99th percentiles, which are provided by age group and year, to estimate a Pareto distribution for earnings above the 95th percentile.

[24] Sample C also excludes households where the spouse has missing age information or a measured wage below half of the minimum legal wage. These exclusions attempt to mitigate measurement error.

A.4 Descriptive statistics

This section presents a brief set of descriptive statistics for Canada, Denmark and Sweden. These are presented in Tables A1-A3.

Table A1. Summary statistics for cross sectional samples: Canada

year	nobs	mean e	e_{50}	e_{99}	$\min_j(obs_{t,j,99})$	obs_{tjk}
1982	886920	24400	23100	72000	220	650
1983	886310	25400	24200	74800	220	650
1984	902625	26800	25500	79600	220	650
1985	914700	28100	26700	84300	215	650
1986	947720	29200	27600	89200	220	650
1987	956945	30700	28800	95300	215	650
1988	983375	32800	30200	106600	220	650
1989	1012465	34700	31600	114500	220	650
1990	1028790	35400	32200	117000	220	650
1991	1022650	36000	32900	119300	220	650
1992	1024415	36700	33600	121700	220	650
1993	1028755	37300	33800	124900	220	650
1994	1033960	38100	34300	130800	215	650
1995	1044510	39100	34900	139000	220	650
1996	1048970	40000	35300	147100	220	650
1997	1058555	41900	36100	160500	220	650
1998	1065610	43600	37100	174300	220	650
1999	1084320	45100	38100	182600	220	650
2000	1101815	47800	39400	200700	220	650
2001	1140225	49000	40200	210500	220	650
2002	1137365	49600	41000	208800	220	650
2003	1149010	50800	42000	214300	220	650
2004	1162555	52700	43100	225900	220	650
2005	1177270	55200	44400	243500	220	650
2006	1186490	57900	45900	261600	220	650
2007	1199525	60000	47400	273200	225	650
2008	1210295	61300	48900	270900	220	650
2009	1201615	59500	48000	257800	220	650
2010	1200940	61400	49400	264300	230	650
2011	1221400	63700	51100	275600	225	650
2012	1233235	65300	52600	279000	240	650
2013	1241750	67000	53900	284200	255	650

Note: Earnings statistics from Statistics Canada are rounded to the nearest 100 for confidentiality.

Table A2. Summary statistics for cross sectional samples: Denmark

Year	nobs	S100	S50	S1	*var*(log(*d*))	mean d	*corr*(*e, d*)	$\min_i(obs_{t,i,99})$	obs_{tik}
1980	871620	0.93	0.96	0.97	0.12	342.33	0.36	165	733
1981	859167	0.92	0.96	0.97	0.13	340.63	0.38	155	733
1982	866315	0.92	0.95	0.96	0.13	339.47	0.42	145	733
1983	879347	0.93	0.96	0.97	0.14	337.53	0.43	150	733
1984	890302	0.94	0.96	0.97	0.13	338.87	0.41	149	733
1985	906252	0.94	0.96	0.97	0.12	339.36	0.38	146	733
1986	917972	0.95	0.97	0.97	0.11	341.45	0.35	147	733
1987	924403	0.95	0.96	0.97	0.11	341.9	0.35	143	733
1988	926431	0.95	0.97	0.98	0.12	341.73	0.37	146	733
1989	927703	0.94	0.97	0.97	0.12	341.01	0.36	137	733
1990	936043	0.95	0.98	0.99	0.12	340.35	0.37	138	733
1991	935039	0.95	0.98	0.98	0.13	339.7	0.39	135	733
1992	943109	0.93	0.96	0.98	0.13	338.82	0.4	132	733
1993	941600	0.93	0.96	0.98	0.15	334.83	0.42	130	733
1994	951024	0.94	0.97	0.99	0.13	337.73	0.4	131	733
1995	962977	0.94	0.97	0.99	0.12	340.43	0.37	131	733
1996	972286	0.94	0.97	0.99	0.11	342.41	0.36	137	733
1997	983871	0.95	0.97	0.98	0.11	342.99	0.34	142	733
1998	998120	0.97	0.98	0.99	0.1	344.22	0.32	150	733
1999	100581	0.97	0.98	0.98	0.1	344.86	0.3	157	733
2000	101132	0.97	0.99	0.99	0.09	345.73	0.26	165	733
2001	101296	0.97	0.99	0.99	0.09	344.72	0.28	169	733
2002	100986	0.97	0.99	0.98	0.09	344.98	0.28	190	733
2003	999303	0.95	0.98	0.98	0.1	344.28	0.31	204	733
2004	993586	0.98	0.99	0.99	0.1	345.08	0.28	225	733
2005	990605	0.98	0.99	0.99	0.1	343.54	0.28	241	733
2006	989524	0.98	0.99	0.99	0.09	345.04	0.26	229	733
2007	984137	0.99	0.99	0.99	0.1	345.5	0.22	231	733
2008	969799	0	0	0	0	0	0	218	733
2009	942820	0	0	0	0	0	0	206	733
2010	923739	0	0	0	0	0	0	202	733
2011	918254	0	0	0	0	0	0	203	733
2012	913586	0	0	0	0	0	0	203	733
2013	911549	0	0	0	0	0	0	203	733

Note: Columns S100, S50, and S1 display the shares of earnings observations for which there is a valid days of work observation in the full sample, the top 50 percent of the earnings sample, and the top 1 percent of the earnings sample.

Table A3. Summary statistics for cross sectional samples: Sweden

Year	nobs	S100	S50	S1	*var*(log(e))	*var*(log(*d*))	*corr*(log(*e*), log(*d*))	Mean *d*	min$_j$($obs_{t,j,99}$)	obs_{tjk}
1980	1845140	-	-	-	0.30	-	-	-	403	1434
1981	-	-	-	-	-	-	-	-	-	-
1982	1830333	-	-	-	0.34	-	-	-	395	1434
1983	-	-	-	-	-	-	-	-	-	-
1984	-	-	-	-	-	-	-	-	-	-
1985	1615820	0.4	0.53	0.28	0.26	0.04	0.34	289.44	309	1434
1986	1627315	0.41	0.54	0.27	0.27	0.04	0.36	286.09	300	1434
1987	1644682	0.41	0.53	0.29	0.27	0.04	0.36	292.59	290	1434
1988	1665408	0.41	0.52	0.28	0.27	0.04	0.35	296.69	293	1434
1989	1691587	0.4	0.5	0.3	0.27	0.04	0.34	291.56	280	1434
1990	1871002	0.36	0.49	0.35	0.38	0.04	0.36	277.79	323	1434
1991	1898011	0.31	0.41	0.24	0.39	0.04	0.4	297.79	325	1434
1992	1875173	0.38	0.47	0.27	0.45	0.05	0.4	288.07	317	1434
1993	1840234	0.39	0.5	0.3	0.52	0.05	0.41	279.06	293	1434
1994	1838130	0.3	0.36	0.18	0.54	0.06	0.41	284.36	289	1434
1995	1856135	0.28	0.39	0.32	0.52	0.04	0.36	283.45	295	1434
1996	1857699	0.28	0.41	0.34	0.53	0.04	0.38	280.98	305	1434
1997	1860797	0.29	0.42	0.36	0.52	0.04	0.39	279.59	311	1434
1998	1883857	0.29	0.4	0.36	0.5	0.04	0.38	283.59	329	1434
1999	1914785	0.29	0.41	0.39	0.5	0.04	0.37	281.29	348	1434
2000	1945461	0.29	0.41	0.4	0.48	0.04	0.34	280.42	352	1434
2001	1962558	0.44	0.55	0.4	0.47	0.08	0.35	269.75	377	1434
2002	1963068	0.44	0.55	0.44	0.47	0.09	0.36	266.47	435	1434
2003	1945148	0.44	0.56	0.44	0.48	0.09	0.37	270.73	439	1434
2004	1928007	0.44	0.56	0.47	0.5	0.09	0.37	265.31	445	1434
2005	1914243	0.44	0.55	0.45	0.5	0.1	0.37	264.3	452	1434
2006	1917082	0.45	0.55	0.47	0.49	0.09	0.36	267.88	454	1434
2007	1913805	0.44	0.54	0.44	0.48	0.09	0.35	270.25	465	1434
2008	1906596	0.44	0.54	0.48	0.45	0.09	0.33	267.27	462	1434
2009	1875741	0.44	0.54	0.49	0.47	0.1	0.35	263.46	450	1434
2010	1871732	0.45	0.54	0.5	0.46	0.09	0.35	269.57	452	1434
2011	1885636	0.43	0.51	0.47	0.43	0.09	0.34	272.1	446	1434
2012	1886082	0.42	0.51	0.47	0.43	0.09	0.36	271.01	455	1434
2013	1886746	0.42	0.51	0.47	0.44	0.09	0.36	268.76	456	1434

Note: Columns S100, S50, and S1 display the shares of earnings observations for which there is a valid days of work observation in the full sample, the top 50 percent of the earnings sample, and the top 1 percent of the earnings sample.