CESifo WORKING PAPERS

9605
2022

February 2022

Inequality and Income Dynamics in Germany

Moritz Drechsel-Grau, Andreas Peichl, Johannes Schmieder, Kai D. Schmid, Hannes Walz, Stefanie Wolter

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9605*

Inequality and Income Dynamics in Germany

Abstract

We provide a comprehensive analysis of income inequality and income dynamics for Germany over the last two decades. Combining personal income tax and social security data allows us — for the first time — to offer a complete picture of the distribution of annual earnings in Germany. We find that cross-sectional inequality rose until 2009 for men and women. After the Great Recession inequality continued to rise at a slower rate for men and fell slightly for women due to compression at the lower tail. We further document substantial gender differences in average earnings and inequality over the life-cycle. While for men earnings rise and inequality falls as they grow older, many women reduce working hours when starting a family such that average earnings fall and inequality increases. Men's earnings changes are on average smaller than women's but are substantially more affected by the business cycle. During the Great Recession, men's earnings losses become magnified and gains are attenuated. Apart from recession years, earnings changes are significantly right-skewed reflecting the good overall state of the German labor market and increasing labor supply. In the second part of the paper, we study the distribution of total income including incomes of self-employed, business owners, and landlords. We find that total inequality increased significantly more than earnings inequality. Regarding income dynamics, entrepreneurs' income changes are more dispersed, less skewed, less leptokurtic and less dependent on average past income than workers' income changes. Finally, we find that top income earners have become less likely to fall out of the top 1 and 0.1 percent.

JEL-Codes: D310, E240, E310, J310.

Keywords: inequality, income dynamics, mobility, non-labor income.

Moritz Drechsel-Grau
University of Zurich / Switzerland
moritz.drechsel-grau@econ.uzh.ch

Andreas Peichl
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Munich / Germany
peichl@ifo.de

*Johannes Schmieder**
Boston University / MA / USA
johannes@bu.edu

Kai D. Schmid
Heilbronn University of Applied Sciences
Heilbronn / Germany
kdschmid@gmail.com

Hannes Walz
Institute for Employment Research (IAB)
Nürnberg / Germany
Hannes.Walz2@iab.de

Stefanie Wolter
Institute for Employment Research (IAB)
Nürnberg / Germany
Stefanie.Wolter@iab.de

*corresponding author

February 26, 2022

We thank the Research Data Lab of the German Federal Statistical Agency, especially Maurice Brand and Hannes Fauser, for steady support in accessing the data. Sophie Friebe provided excellent research assistance. We also thank Andreas Ganzer and Stefan Seth for providing imputed wages. Peichl gratefully acknowledges financial support from the German Science Foundation (DFG) via CRC TRR 190 (project number 280092119) and Walz from the joint doctoral program of the Institute of Employment Research (IAB) and Friedrich-Alexander University Erlangen-Nürnberg, GradAB.

1 Introduction

Since the beginning of the 21st century we have seen a renewed interest among economists to study and understand the structure and evolution of income inequality as well as the forces that shape it. Focusing solely on cross-sectional inequality leaves out many aspects of the distribution of welfare. A young worker entering the labor market might be perfectly content with a low starting salary at the bottom of the wage distribution if she may expect rapid wage growth and upward mobility over the coming years. Similarly, a worker with middle class income might experience significant uncertainty about his income in subsequent years, for example, due to uncertainty about bonus pay and hours or due to the risk of job displacement and subsequent earnings losses. This has motivated recent work in studying income inequality jointly with income dynamics to provide a more complete picture of the distribution of economic well-being in an economy (e.g. Guvenen et al., 2021b).

This paper provides a comprehensive analysis of income inequality and income dynamics for Germany over the last two decades. We combine two high quality administrative data sources for this analysis: personal income tax records from the Taxpayer Panel (TPP) as well as social security data from the Institute for Employment Research (IAB). Since each of these datasets has distinct advantages and weaknesses, combining the two allows us, for the first time, to offer a complete picture of the German income structure ranging from workers in marginal employment at the very bottom all the way to the highest paid CEOs at the top of the distribution. We begin by focusing on labor earnings, which is the main source of income for the vast majority and most easily compared across datasets and countries. We first describe the evolution of cross-sectional earnings inequality. We then show patterns in the distribution of earnings changes for workers at different stages of their lives and different parts of the earnings distribution. Finally, we relate to the recent literature on capitalists (see, e.g., Smith et al., 2019; Autor et al., 2020) and add self-employed, business owners and landlords to analyze in detail how this affects the picture of inequality and income dynamics.

Germany is a particularly interesting setting with a number of considerable changes in its labor market throughout the last two decades. After a period of high unemployment rates throughout the 1990s and early 2000s, the German labor market experienced a remarkable turnaround starting around 2005 with unemployment rates falling steadily from over 13% to less than 6% in 2018.[1] With this considerable decline in unemployment, the attention of policymakers and the public increasingly focused on income inequality which had risen dramatically during the 1990s and in the early 2000s (see, e.g., Fuchs-Schündeln et al., 2010). In the following years, pushes for higher wages, for example through more aggressive collective bargaining negotiations, could be observed.[2] And finally, for the first time in Germany, in 2015 a nation-wide minimum wage was introduced.

[1] There are several reasons for this improvement: Some economists credit the 2003-2006 labor market reforms ("Hartz reforms") and the short-time work program (Gehrke et al., 2019), while others claim that the combination of flexible collective bargaining institutions and restructuring of supply chains were the driving forces (e.g. Dustmann et al. 2014, Hoffmann and Lemieux 2016). (Macro)economic analyses of the Hartz reforms (e.g., Krebs and Scheffel, 2013; Launov and Wälde, 2013; Hartung et al., 2018; Bradley and Kügler, 2019; Hochmuth et al., 2021) show that the reforms did play an important role for the decline in unemployment, but other factors were also important.

[2] For example, nominal union wages in manufacturing increased by 2% on average annually from 1998–2006 and by 2.8% from 2007–2018 (3.1% from 2007–2009, i.e. before the Great Recession) – see Statistisches Bundesamt (2021b).

Our paper adds to the literature on the German wage structure and labor income inequality in several important ways. Prior to the 1990s, studies find stable (cross-sectional) wage inequality in West Germany (e.g., Biewen, 2000). Using survey data, Fuchs-Schündeln et al. (2010) show that inequality trended upwards for wages and market incomes since the mid-1990s (until about 2005, where their analysis stops). Using IAB data Antonczyk et al. (2018) show that the increase in earnings inequality was mostly restricted to the right tail of the distribution.[3] More recently papers have looked at the 2010s when labor market conditions were improving (e.g., Lochner et al., 2020). Using survey data, Biewen et al. (2019) show that the general rise in wage inequality became less steep (but did not stop) after 2005, while inequality in annual labor incomes did not increase further after 2005 as employment rates increased. Brüll and Gathmann (2020) show that wage inequality decreased already after the introduction of sectoral minimum wages while Bossler and Schank (2020) focus particularly on the consequences of the general minimum wage introduction in 2015. This literature focuses almost exclusively on wage inequality, in particular daily wages of full-time male workers.[4] Focusing on full-time employees is a natural choice for understanding the supply and demand forces shaping the wage structure as the full-time wage is most easily interpreted as a 'price'. However, focusing on full-time wages is less sensible if one is interested in the distribution of individual welfare and how it is affected by income mobility and risk. This is particularly true for women who have a high propensity for working part-time.[5]

We focus on annual earnings (labor income) as our main outcome variable. In contrast to the existing literature, we analyze both men and women and include individuals in marginal or part-time employment as well as people who work partially during the year. This broad sample is much better suited to analyze and compare the income distributions of women and men. This also allows us to capture the unique phenomenon of marginal employment in so called mini-jobs in Germany (see section 2 for more details).

Another difference is that the vast majority of previous work for Germany has used either survey data – which suffer from all the typical problems of selection bias, measurement error, and small sample sizes – or solely IAB data. While the administrative IAB data is of high quality and provides complete coverage of all employees who are in marginal employment or in jobs liable to social security (that is all jobs excluding civil servants and self-employed), the earnings are top coded at the social security contribution limit (corresponding approximately to the 90th percentile for men and the 96th percentile for women). Hence, these studies cannot address the upper tail

[3] Several further papers have studied the evolution of wage inequality during the 1990s and 2000s (e.g., Dustmann et al. 2009, Card et al. 2013, Goldschmidt and Schmieder 2017 and Bartels 2019). The reasons for the steep increase are manifold. For instance, Bayer and Juessen (2012), Antonczyk et al. (2018) and Biewen et al. (2018) find cohort and composition effects to be especially important while Peichl et al. (2012) document that the increase in inequality is strongly related to changes in household structure and employment behavior. Fuest et al. (2018) find that corporate tax hikes increase wage inequality as low-skilled, young, and female employees bear a larger share of the tax burden.

[4] Biewen et al. (2018) show the importance of part-time and employment interruptions for the increase in income inequality. Bönke et al. (2015) analyze lifetime earnings inequality and mobility of yearly earnings for 35 cohorts of West German men. Moreover, several studies analyze inequality in (equivalized) disposable household incomes – see, e.g., Hufe et al. (2018) for a survey and Stockhausen and Maiworm (2021) or Grabka (2021) for recent studies.

[5] For example, in 2019, 58% of women in Germany worked part-time (Wanger, 2020) and women held 81% of all part-time jobs (Statistisches Bundesamt, 2021a).

of the distribution. In contrast, the TPP data does not have any top-coding and hence includes the very top of the distribution.[6] However, many low income workers (and especially workers in mini-jobs) do not file a tax return and thus are not included in the TPP. Combining IAB and TPP data, allows us, for the first time, to study the top and the bottom of the earnings distribution in Germany simultaneously.

This paper is also the first to combine the study of income inequality with a detailed analysis of income dynamics in Germany based on administrative data.[7] We build on the work by Busch et al. (forthcoming) who study income risk over the business cycle for Germany (and other countries) up to 2010 using survey data, documenting highly pro-cyclical skewness in short-term income growth. We extend their work as our combined tax and social security data allows us to study income dynamics for high earners and to include various forms of income. By extending the time horizon we are also able to study how the prolonged period of a strong labor market has affected inequality, mobility and the distribution of earnings changes.

Our main analysis focuses on the time period from 2001 to 2016. We start in 2001 since this is the first year the TPP data is available and when the IAB data provides high-quality information on mini-jobs. As of 2018 there are about 7.5 million mini-jobs and almost 5 million workers work only in a mini-job, thus including these jobs is crucial for getting a complete picture of the German earnings distribution at the lower end. The downside of focusing on the last two decades is, that this is less comparable to earlier papers for Germany that have focused on time series for wages excluding mini-jobs. For this reason, we provide an analysis in the Appendix based on two alternative samples from the IAB data where we exclude mini-jobs: Germany from 1993 to 2018 (starting right when high quality data for East Germany becomes available) and West Germany from 1985 to 2018.

This paper is part of the Global Income Dynamics (GID) project and as such, we follow the comprehensive guidelines of the project to provide a consistent set of core results presenting the evolution of inequality and dynamics in labor income over time. Apart from focusing on annual earnings, the GID project specifies a number of sample restrictions, such as age range and a minimum annual earnings threshold as well as various key measures for the outcomes of interest.

In the first part of the paper, we present several new key results for the structure and evolution of the earnings distribution in Germany by gender: The first result is that for men, continuing the trend of rising inequality in the 1990s and early 2000s, income inequality kept increasing until the Great Recession both at the top and at the bottom of the distribution. After the Great Recession, incomes throughout the distribution, including the lower half, began to increase, slowing the rise in inequality. For women the picture is more complex: While inequality for most women increased until 2009 as well, the 10th percentile, which is mostly composed of mini-jobs, actually rose throughout the entire sample period. After the Great Recession, women's incomes rose quickly, particularly at the 25th

[6] Following Piketty and Saez (2003), researchers have used tax data to study the upper tail of the distribution. For Germany, Bartels (2019) provides a long-run picture of the evolution of top income shares using tabulated tax data. Moreover, Bach et al. (2009) combine survey data and repeated cross-sections of tax data to document inequality trends from 1992 to 2001. Jenderny (2016) uses a sub-sample of the TPP to study top income mobility from 2001–2006.

[7] Using survey data, Fuchs-Schündeln et al. (2010) estimate stochastic wage and earnings processes for Germany from 1984–2003 and decompose both the levels and the growth rates in transitory and permanent components

percentile, which catches up with higher income levels leading to a substantial decrease in lower tail inequality. In contrast to the lower tail, earnings inequality at the very top increased substantially for both men and women, but the increase is almost twice as large for women. For example, across both genders, the percentile with the highest earnings growth is the 99.99th percentile for women.

A second key result is that earnings inequality is substantially larger for women than for men but converges throughout our sample period due to the different trends by gender. Interestingly, initial conditions (inequality at age 25, i.e. around labor market entry) are virtually identical, but while inequality is falling for men within cohorts, it is rising for women. This is driven by many women opting for part-time work or mini-jobs later in life after having children (Kleven et al., 2019), which leads to large variations in working hours within cohorts over time.

A third key result is that volatility (measured as the dispersion of 1 year log earnings changes) exhibits opposite cyclicality at the bottom and at the top of the distribution with shocks in the lower half of the distribution increasing during downturns, while shocks at the top become more muted. The overall volatility is relatively constant, but the skewness of the shocks becomes markedly more negative during downturns. This holds for men and women. Volatility is also significantly higher for women than for men, especially for younger women and at higher income levels.[8]

In the second part of the paper, we present novel results regarding total income inequality and dynamics. First, we show that non-labor income is a major source of total income especially at the bottom and at the top of the distribution. Taking total income, i.e. the sum of labor, rental, self-employed and business income, as the main outcome measure, we find much higher levels of income inequality. In addition, over time, the top percentiles of the total income distribution increased significantly more than the corresponding percentiles in the earnings distribution.

Second, we compare income dynamics between workers and entrepreneurs (individuals with non-labor income as their main income source) and find that entrepreneurs' income changes are more dispersed, less skewed, and much less leptokurtic. In addition, income changes of entrepreneurs are much less state-dependent in the sense that we do not find significant heterogeneity between low-income and high-income entrepreneurs. Third, we document that income mobility at the top has declined significantly between 2001 and 2016. That is, the probabilities of dropping out of the top 1% and top 0.1% of the income distribution have declined both for 1- and 5-year time intervals.

In the next section, we discuss the institutional and macroeconomic setting in Germany over our analysis period, present our data sources and explain the sample construction. Section 3 presents the core results following the GID framework showing the evolution of inequality and income dynamics for labor earnings. Section 4 expands the analysis to total income. Finally, Section 5 concludes.

[8] In the Appendix we also document results on income mobility. Mobility is fairly high at lower ages and then decreases with worker age. Furthermore, mobility is quite a bit larger for women than for men, perhaps again reflecting the impact of hours reductions after childbirth and increases in hours after children grow older (Kleven et al., 2019).

2 Background and Data

2.1 Institutional and Macroeconomic Background

In this subsection, we give a brief overview of the relevant institutions and the macroeconomic situation in Germany for the period from 1993 to 2018 – see Appendix A for more details. While our main analysis focuses on 2001 to 2016, we provide additional results for this longer time period in Appendix F. Furthermore, starting slightly earlier than our main sample window helps to better understand the economic environment during that period.

Personal Income Tax. Germany applies a comprehensive income tax on income from all sources. Married couples file their tax return jointly. Both features are important when constructing our sample from tax return data – as discussed below.

Marginal Employment ("Mini-Jobs"). Marginal employment contracts, called mini-jobs, are jobs with earnings below a time-varying threshold (see Panel C of Figure A.1). Jobs below this threshold, which currently amounts to 450 Euro per month, are exempted from social security contributions and income tax.[9] Two reforms during our sample period increased the monthly earnings threshold from 325 Euro to 400 Euro (in 2003) and then to 450 Euro (in 2013). Over our sample period in each year around 4.5-5 million workers hold only a mini-job. Another 2.7 million workers have a secondary mini-job next to a regular contract.

Minimum Wage. Germany introduced a statutory national minimum wage of 8.50 Euro in 2015. After that, the minimum wage was gradually increased (see Figure A.1, Panel C). Before 2015 different wage floors existed in 12 industries. Furthermore, some of the larger industries have binding collective agreements that set minimum wages. The impact of the wage floor on wages varied by region and affected about 15% of all employees (Dustmann et al., 2022).

Collective Bargaining. Agreements between unions and employer representatives often have a binding character for most firms (above a certain size) in a specific industry (with some possibilities for firms to opt-out) in Germany. The worker coverage of industry-level collective bargaining agreements varies between former West and East Germany and decreases over time (see Figure A.1, Panel B). Especially start-ups and smaller firms are less likely to be part of a collective agreement.

Macroeconomic Background. The macroeconomic development in Germany from 1993-2018 can be broadly split into two periods: before and after 2005 (see Figure A.2). The first part was characterized by low growth and high unemployment (above 12%) and Germany was often referred to as "the sick man of Europe" (Dustmann et al., 2014). This changed in the mid-2000s after a series of (labor market and tax) reforms[10] The Great Recession did not affect the labor market severely.

[9] A person can hold multiple mini-jobs but then only the first 450 Euro are tax exempt.

[10] The causal effect of these reforms and the exact mechanisms are still discussed in the literature – see Footnote 1.

Moreover, labor force participation rates increased steadily after 2004 and the unemployment rate fell below 6% in 2018. Especially the large increase in labor force participation of women, from around 55% to more than 70%, is notable (see Figure A.2, Panel E). However, unlike in countries such as the US, this increase was almost exclusively driven by women entering the labor market in part-time and marginal employment, so that the full-time share over this period fell from 75% to around 50% for women. For men, labor force participation and the part-time share also increased substantially since 2003, though nowhere near as pronounced as for women.

2.2 Data

For our analysis, we combine two high quality administrative data sources: social security data (IAB) and personal income tax records (TPP). Each of these datasets has distinct advantages but also some weaknesses. The combination of both data sources offers a unique possibility for the analysis of inequality and income dynamics along the whole distribution. We describe the two datasets in the next sub-sections before explaining our sample selection, comparing the income distributions in both datasets and describing how we merge them. Appendices B (IAB data), C (TPP data) and D (combined IAB-TPP data) contain further details.

2.2.1 The IAB Social Security Data

The first source of data, which we refer to as the IAB data, is the Integrated Employment Biographies (IEB, version 13.01) supplied by the Institute for Employment Research (*"Institut für Arbeitsmarkt- und Berufsforschung", IAB*). The data contains information on employment and earnings as well as worker and firm characteristics such as gender, education, year of birth, occupation or industry code.[11] The information is spell based, i.e. accurate to the date and especially with respect to earnings very reliable. However, there are two important limitations to this widely used dataset (see, e.g., Dustmann et al., 2009; Card et al., 2013). First, labor earnings are reported including bonuses and extra pay but only up to the social security contribution limit (see Figure A.1 for its real values over time). This censoring affects men and women in West and East Germany differently with West German men being affected the most as here the top 10% are subject to censoring (see Appendix B and especially Figure B.1 for details).[12] Second, the IAB data does not include self-employed individuals (around 4 million) and civil servants (around 1.9 million individuals).

We use a 10% random sample of the IEB for the years 1993-2018, which gives us 87,012,649 observations. For our main results we focus on the period 2001-2016 (65,900,481 observations on 6,250,877 individuals). We show results for the period 1993-2018 in Appendix F.1 and, restricting to West Germany, for the period 1985-2018 in Appendix F.2.

[11] Note that the education information contains some missing values which we impute using the procedure suggested by the IAB. Moreover, throughout 2011, the reporting procedure for full-time and part-time employment in the data changed leading to a small fraction of workers being falsely classified before 2012. We correct the full-time indicator using a cell-wise reclassification approach. See Appendix B for details on both imputations.

[12] We use the algorithm suggested by Card et al. (2013) to impute daily wages which we then aggregate to annual incomes for our analysis. Note, however, that we do not use these imputed wages for our baseline analysis as we will draw on tax data to complement the IAB data at the top as we explain below.

2.2.2 German Taxpayer Panel (TPP)

The second source of data is the German Taxpayer Panel (TPP), which is an administrative dataset based on the universe of personal income tax returns (Kriete-Dodds and Vorgrimler, 2007).[13] The dataset covers all tax units filing tax returns in Germany in the period 2001-2016. The panel has a total of 58,808,899 unique records for which information is available for at least two years. We use a 25% random sample of these records. The unit of observation is the taxpayer, i.e., either a single individual or a couple filing jointly. In the latter case, incomes are measured on the individual level. Moreover, while we observe both spouses before and after a marriage (or divorce), we can only track the husband over these events. The reason is that the wife is assigned to the husband's tax unit in the event of marriage and withdrawn from it in the event of divorce.[14] As less than 10% of the women in our data could be affected once over the period of analysis, we are confident that our income dynamics results for women are mostly unaffected by this.

The dataset contains all information necessary to calculate a taxpayer's annual income tax. This includes basic socio-demographic characteristics such as year of birth, gender, family status, number of children as well as detailed information on gross income (differentiated by seven different sources) and tax-specific parameters such as deductions. As the data are not top-coded, they are especially suited for the analysis of inequality in the upper tail of the distribution. They are, however, missing the very bottom of the income distribution as incomes below the marginal employment threshold are except from the income tax and hence not included in the data (in the case of the mini-job being the only source of income). Note, however, that information on mini-jobs of secondary earners as well as recipients of income from other sources are included in the data. Furthermore, there is a structural break in the dataset in 2011 for the classification of workers which are subject to social security contributions (which are represented in the IAB data) affecting about 4% of the observations. We describe in Appendix C how we correct the data using reweighting techniques.

2.2.3 Sample and Variable Construction

For comparability with other countries covered in the GID project, we focus our analysis on individuals who are between 25 and 55 years old. Following the GID guidelines, the first part of our analysis (section 3) focuses on labor earnings excluding self-employment. Throughout our analysis, we examine both men and women as well as not only full-time but also part-time and marginally employed workers. The definition of gross annual labor earnings is the same for both IAB and TPP data: annual earnings is broadly defined and include, among others, overtime pay, bonuses, 13^{th} month pay, paid sick leave, severance pay, and vacation allowance. We exclude workers with weak

[13] See Appendix C for more information. The TPP has been, for example, used by Doerrenberg et al. (2017) and Dolls et al. (2018) who also provide additional information on the data.

[14] While we see the wife in all years, we could have 3 independent spells for her: before, during and after marriage. On average about 0.45–0.5% of individuals get married each year while roughly 0.2% file for divorce (https://www.destatis.de/EN/Themes/Society-Environment/Population/Marriages-Divorces-Life-Partnerships/Tables/lrbev06.html). The average duration of a marriage is about 15 years. As we analyze a period of 16 years we would expect that about 8% of all women in the data marry while less than 2% file for divorce.

attachment to the labor force by trimming annual earnings below a threshold $\underline{y_t}$, which corresponds to working part-time for one quarter at the national minimum wage (2,300 Euro in 2018). As we combine IAB and TPP data, and since the IAB data only covers social-security-liable earnings, we restrict our analysis to this group which accounts for more than 93% of all workers. In the second part of our analysis, we investigate the distribution of total income and study total income dynamics for individuals with different main income sources (see Section 4 which also includes some further details on the total income sample). All incomes are deflated using the CPI and Euro figures in the text, tables and figures refer to 2018 Euro.

For both labor earnings and total income, we follow the GID project's conventions and refer to three samples. In the "CS sample" (cross-sectional sample), we only impose the restrictions on age and minimum (labor) income. For the analyses that involve dynamics, we impose additional restrictions on the data and focus on two subsamples. To study trends, we use the "LS sample" (longitudinal sample) which only includes observations with non-missing one-year or five-year income changes. When studying heterogeneity in income dynamics by age and income, we use the "H sample" (heterogeneity sample) where we drop observations for which we cannot compute our measure of permanent (labor) income based on observations of the past three years.

2.2.4 Comparison and Combination of IAB and TPP Data

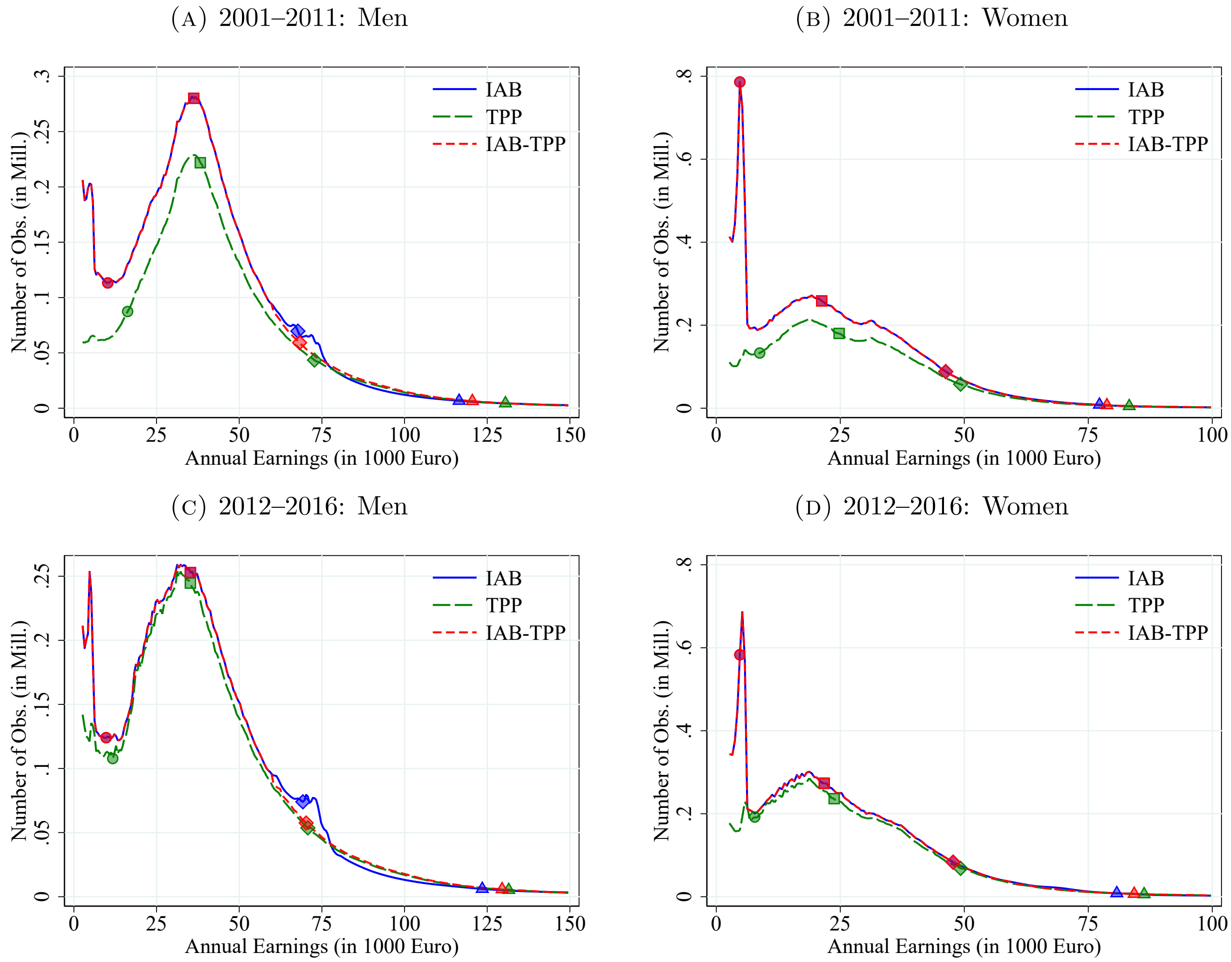
The four key differences between the IAB and TPP with respect to labor income are: several (but not all) missing mini-jobs and the incomplete coverage of regular social security liable jobs due to non-filing (before 2012) in the TPP data as well as the top-coding and the omission of civil servants and self-employed individuals in the IAB data. Table D.1 in the Appendix displays descriptive statistics for the IAB and TPP datasets for the year 2008 separately for men and women who are between 25 and 55 years old. Unsurprisingly, the TPP has fewer observations due to missing non-filers and mini-jobs. As the TPP data contains only very limited demographic information, we can only compare both datasets in terms of age. The TPP population is slightly older which again can be attributed to missing observations representing persons who are more likely to be at the beginning of their career. There are more men than women in both datasets due to their higher labor force participation rate (see Figure A.2, Panel E).

Figure 1 provides a comparison of the earnings distributions in the two raw datasets separately by gender for the periods 2001–2011 and 2012–2016 (as the TPP data includes information on non-filers starting from 2012).[15] For this graph we focus on jobs subject to social security contributions also in the tax data such that all lines refer to the same population. Since only workers who submit a tax return were covered in the TPP until 2012 (Panels A and B) it is not surprising that the green line for the TPP data consistently lies underneath the blue line for the IAB data until about the social security contribution limit at around 70,000 Euro is reached (note that above the contribution limit the IAB data is imputed). From 2012 onward (Panels C and D), both IAB and TPP data

[15] Figure D.1 in the Appendix shows the same information for the full population. Tables D.2 – D.4 show selected earnings percentiles across the different datasets for men, women and in the population respectively for all years.

are much closer together. The biggest difference is still at the bottom of the distribution where the IAB data show a large mass-point stemming from mini-jobs while still about half of the mini-jobs covered in the IAB data are missing in the TPP.

Figure 1: Annual Earnings Distribution in IAB, TPP and Combined Data



(a) 2001–2011: Men

(b) 2001–2011: Women

(c) 2012–2016: Men

(d) 2012–2016: Women

Notes: This figure shows the number of observations in real earnings bins for the IAB, the TPP and the combined data (IAB-TPP) by gender (see Figure D.1 in the Appendix for the combined distribution). Panels A and B show averages across the years 2001 to 2011 where non-filing workers (Lohnsteuerfälle) are not included in the TPP and Panels C and D show averages across the years 2012 to 2016 where the TPP data include these workers. We exclude earnings from the TPP that are not subject to social security contributions (e.g. salaries of civil servants) which are not covered in the IAB. The circular, square, diamond and triangle-shaped markers depict the 10^{th}, 50^{th}, 90^{th} and 99^{th} earnings percentile in the respective datasets. We use 500 Euro bins below 80,000 Euro and 1,000 Euro bins above 80,000 Euro but always plot the number of observations per 1,000 Euro bins. The IAB data are imputed above the social security contribution limit. Tables D.2, D.3 and D.4 show selected earnings percentiles across the different datasets for men, women and in the population respectively.

The figure also displays symbols on each line to indicate certain percentiles of each distribution. Interestingly, while both the median (squares) and the 90th percentile (diamonds) values lie relatively close to each other in both datasets, there are larger differences at the 10th percentile (circles), which for women actually lies to the left of the largest mini-job mass point, as well as at the 99th percentile (triangles). Overall, the IAB data is slightly shifted to the left compared to the

TPP data. Note also that the imputation procedure for the IAB data does a fairly good job at approximating the top tail compared to the TPP data but is not fully able to overcome the problem of top-coding and the resulting biases. The earnings distribution of men is more affected by this top coding than the distribution for women, which in turn is more seriously affected by the omission of most mini-jobs at the bottom of the distribution in the TPP data.

For our analysis, we combine both datasets. Due to data protection legislation in Germany, we are not allowed to directly link the individual micro data. Hence, we employ non-parametric matching techniques as described in Appendix D. To do so, we first reweight the TPP data (conditional on gender, age and earnings bin) such that we match the total number of workers liable to social security contributions (from the IAB data). Second, we combine the results from both datasets. For the core analysis in Section 3, we use the (true) earnings distribution from the IAB data below the top-coding threshold. Above the cutoff, we use the conditional earnings distribution from the (reweighted) TPP. For the analysis of total incomes in Section 4, we use the reweighted TPP data.

Figure 1 also shows the combined data which roughly corresponds to the IAB data at the bottom and in the middle of the distribution and to the TPP data at the top. Tables D.2 – D.4 in the Appendix show selected percentiles of the earnings distribution in the combined IAB-TPP data as well as in the IAB and TPP data to confirm this observation.

3 Earnings Inequality and Dynamics in Germany

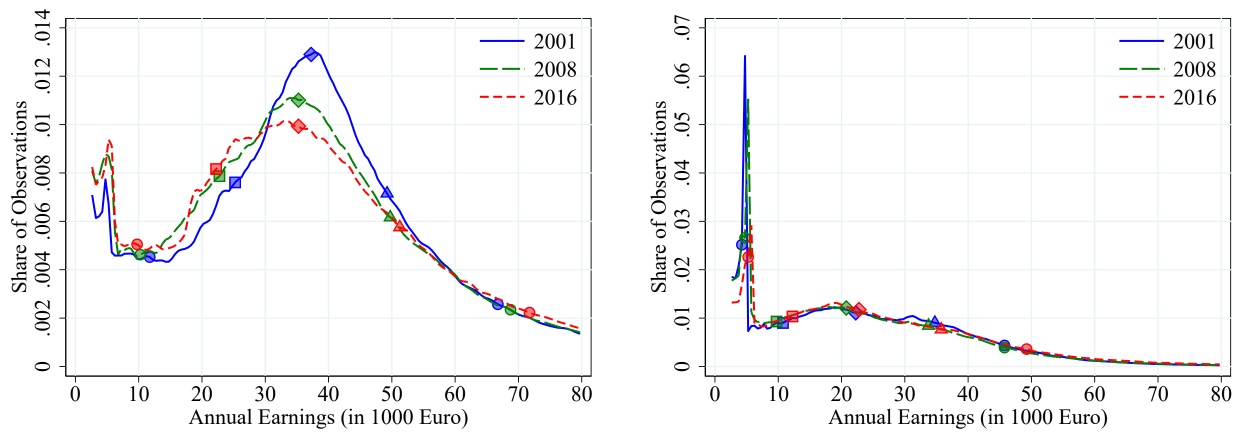
3.1 Earnings Inequality

We start our analysis by documenting the evolution of income inequality over the past 2 decades. Figure 2 shows the density of (real) earnings for men and women for 3 selected years over our sample period. Panel A shows the distribution for men. The mode and median of the distribution in 2001 are just below 40,000 Euro. The distribution also shows a sizable mass (about 6%) at about 5,000 Euro, which corresponds to the earnings of someone working in a mini-job for the full year (325 Euro times 12 adjusted for inflation). Panel B shows the earnings distribution for women: The mini-job spike is much more pronounced for women and the mode of the distribution above the mini-job threshold is close to 20,000 Euro, much lower than for men. The markers on the density correspond to various percentiles of the distribution. Thus, we can see that for men about half of the workers' earnings are between 25,000 and 50,000 Euro per year, while for women the inter-quartile range lies between about 10,000 and 35,000 Euro. Table 1 complements this visual presentation of the percentiles with the exact numbers for selected years. It is striking how much lower the respective percentiles for women are compared to those for men. For example, the median for women is below the 25th percentile for men, while the 75th percentile for women is below the median for men.

Comparing the densities across years in Figure 2 gives a first impression of how the distribution changed over our sample period. For men, there is a clear reduction in mass in the middle of the distribution; e.g. at 40,000 Euro, the density decreased from 0.013 (that is 1.3% of all workers make between 40,000 and 41,000 Euro) in 2001, to 0.01 in 2008, and further to 0.009 in 2016. Instead,

Figure 2: Selected Real Earnings Distributions

(a) Men (b) Women



Notes: This figure shows the distribution of real annual earnings (in 2018 Euro) for selected years in the combined IAB-TPP data (CS sample) by gender. The data is smoothed (by year and gender) using a three-bin moving average for bins above 10,000 Euro. The markers indicate the 10th (circle), 25th (square), 50th (i.e. median; diamond), 75th (triangle) and 90th (circle again) percentiles of the respective distributions.

we have a significant increase in the density in the range of 20,000 to 30,000 Euro, as well as in the mini-job range and at the very top of the distribution. For women, the differences are harder to make out due to the compressed axis which is necessary to show the huge mini-job share (see Figure E.1 for a zoomed-in version). The basic pattern of hollowing out of the middle of the earnings range and increases in the 20,000 to 30,000 Euro range holds for women as well, though is somewhat less pronounced.

Table 1: Percentiles of Real Annual Earnings in Combined IAB-TPP Data

Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
Men												
2001	13.113	42,479	6,331	11,893	25,563	37,599	50,524	69,852	86,892	141,844	331,421	939,818
2004	12.419	42,455	5,406	10,088	24,398	37,434	51,274	71,866	89,516	145,790	335,445	896,172
2008	12.430	42,459	5,409	10,282	23,013	36,029	50,985	72,925	92,556	158,500	397,202	1,117,735
2012	12.535	42,433	5,121	10,104	22,479	35,528	51,330	74,097	94,212	160,328	387,358	1,025,836
2016	13.096	43,665	5,445	9,991	22,831	36,206	52,817	76,755	97,788	168,152	415,058	1,188,906
Women												
2001	11.476	24,628	3,671	4,673	10,881	22,162	34,723	46,032	54,306	78,226	137,340	285,996
2004	11.101	25,281	3,599	4,686	9,905	21,778	34,760	46,552	55,572	81,110	141,506	281,579
2008	11.228	25,002	3,695	4,736	9,811	20,845	33,761	46,100	55,701	84,335	158,265	335,160
2012	11.510	25,610	3,832	4,861	10,664	21,107	34,085	47,149	57,410	88,128	167,849	363,598
2016	11.799	27,562	4,178	5,366	12,387	22,805	36,241	49,859	61,077	95,403	185,926	414,783

Notes: This table shows the number of observations (in millions) and selected percentiles of real annual earnings (in 2018 Euro) in the combined IAB-TPP data (CS sample) by gender for selected years. Tables D.2 and D.3 in the Appendix show the percentiles in the underlying IAB and TPP data.

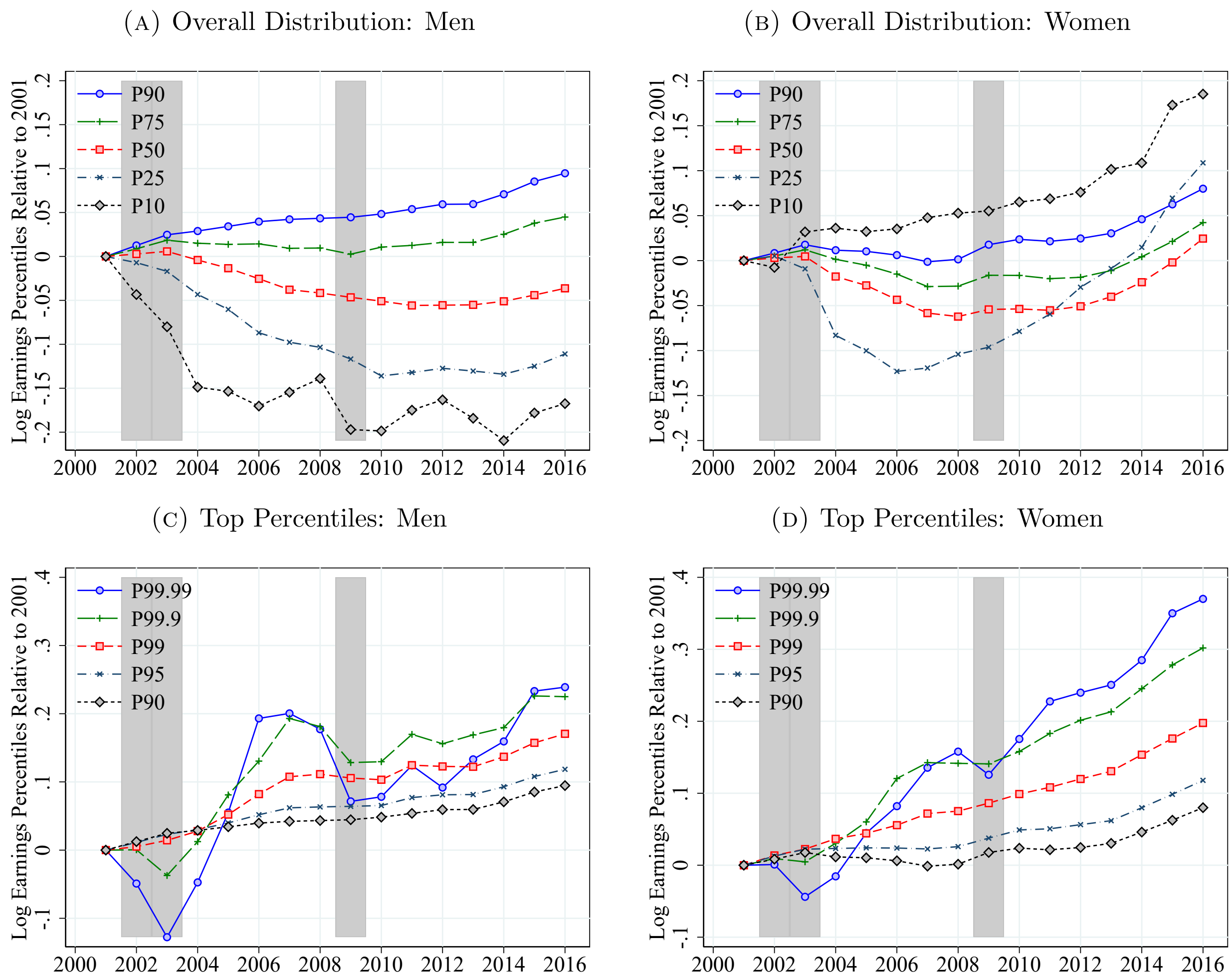
To get a clearer picture of how earnings inequality has evolved, Figure 3 shows a range of percentiles of the log earnings distribution relative to 2001 separately by gender (for the full population - men and women combined - see Figure E.5). Panels A and B illustrate the percentiles up to the 90th for both men and women. Men experienced a pronounced increase in earnings inequality until 2009. At the lower tail of the distribution, the 10th percentile dropped by about 20 log points from 2001 to the Great Recession. Since median earnings also declined, but by only around 5 log points, the log earnings gap between the 50th and 10th percentile increased by around 15 log points. The 90th percentile grew slowly until the Great Recession by around 5 log points, increasing the gap relative to the median. For women inequality also rose in the middle of the distribution as can be seen by the fall of the 25th percentile relative to the 90th percentile. However, the 10th percentile showed a gradual increase after 2001.

After the Great Recession the economy recovered, unemployment rates fell and earnings levels throughout the distribution began to rise slowly for men. Women on the other hand show relatively faster growth in earnings after 2011 throughout the full distribution with the fastest growth at the lower percentiles, in particular the 25th. The faster growth at lower percentiles implies that for women's earnings inequality fell substantially over the second half of our sample period.

Turning to the evolution of incomes at the upper end of the earnings distribution, we can see that for men (Panel C), the top percentiles grow at a faster rate than lower incomes. For example earnings at the 99.9th percentile grew almost twice as much as at the 90th percentile (20 vs. 10 log points). This observation is, for example, consistent with the increase in CEO compensation in Germany during that time (see, e.g., Prinz and Schwalbach, 2020). The fanning out of top incomes is even more pronounced for women (Panel D), where the very top of the earnings distribution (the 99.99th percentile) grew by almost 40 log points or 4 times as much as the 90th percentile. The faster earnings growth for women at the very top also suggests that women at the top end of the income distribution are catching up with the highest-earning men in the economy – although even in 2016 the 99.99th percentile of earnings for women is still only equal to the 99.9th percentile for men (see Table 1). It is also noteworthy that the highest percentiles for both men and women show much more cyclicality than the lower percentiles with marked drops during recessions. The cyclicality is likely due to bonus pay for top managers and highly qualified workers, which in its nature fluctuates over the business cycle. Female top incomes are less cyclical than male top incomes. This might be because female top earners are more often employed in industries that are less prone to business cyclicality. Another potential explanation for the differences between men and women is the increase of the female share in management positions – both at the very top of firms (Kirsch and Wrohlich, 2020) but also at lower management levels (Kohaut and Möller, 2017).

To provide a concise picture of the evolution of inequality, Figure 4 shows different log percentile differentials. These figures support the impression from before: For men, inequality is rising moderately until around the Great Recession, followed by a period of slow growth until 2016. Comparing these differentials with women, we see that the top half of the distribution are very similar between men and women (P90-P50), while the lower two gaps are much larger for women. Regarding

Figure 3: Evolution of Log Earnings Percentiles



Notes: This figure shows the evolution of selected percentiles of log real annual earnings (relative to 2001) in the combined IAB-TPP data (CS sample) by gender. Shaded areas indicate recessions.

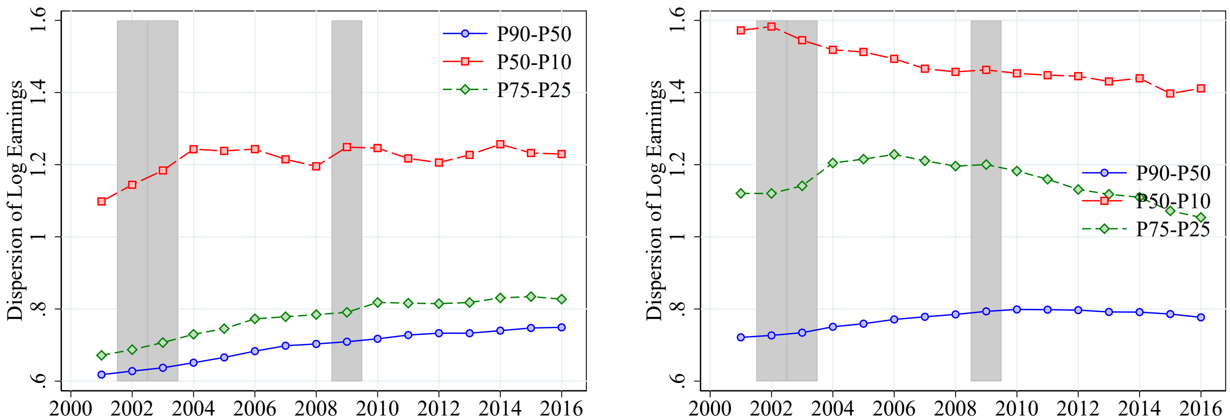
changes over time for women, inequality in the middle and the top half of the distribution (P90-P50 and P75-P25) is increasing slightly from 2001 to 2009, but then decreases until 2016. Inequality at the bottom (measured as P50-P10) by contrast falls throughout the entire sample period, driven by the rising 10th percentile as shown before. As was highlighted by the density figures before, the 10th percentile for women is somewhat special in that it is in the range of mini-job incomes and roughly traces the increases of the mini-job threshold over time. The inverted U-shape of the P75-P25 differential where inequality is rising up to the Great Recession and falling afterward, is probably a better measure for the overall picture of inequality and also in line with measures such as the Gini coefficient or standard deviation (see Figures E.3 and E.14).[16]

What explains the different patterns for men and women? As we saw in Table 1, a key difference between men and women is that the gender-specific percentiles are located at different absolute

[16]In Appendix E, we also report labor income shares of various parts of the income distribution (Figure E.12 and Tables E.1 and E.2) and estimated Pareto coefficients for the top 5% and top 1% (Figure E.13).

Figure 4: Evolution of Earnings Inequality: Log Percentile Differentials

(a) Upper and Lower Inequality: Men (b) Upper and Lower Inequality: Women



Notes: This figure shows the evolution of different log percentile differentials over time in the combined IAB-TPP data (CS sample) by gender. Shaded areas indicate recessions.

income levels. Taking this into account, it appears that at specific income levels men and women are somewhat more similar. For example, the 75th percentile for men is at a similar level as the 90th percentile for women and the two lines look fairly similar in Figure 3. Nevertheless, important differences remain, in particular at the bottom and at the very top of the income distribution.

Turning to the bottom of the distribution, the 10th percentile for women is at about 5,000 Euro, right at the level of someone who works only a mini-job for the full year. Thus, the steady increase of this percentile is essentially earnings growth for mini-job workers. Indeed, the jumps in the 10th percentile line in 2003 and 2013 are driven by the increases in the mini-job threshold in those years (see Figure A.1). As shown by Gudgeon and Trenkle (2020), the adjustment process took several years with many workers initially remaining at the pre-reform levels. This is likely a factor in the rise in the 10th percentile in the years between the reforms. The other marked jump is from 2014 to 2015 when the Federal minimum wage was introduced pushing some mini-jobbers above the earnings threshold (and thus out of the mini-job range). For men, however, the 10th percentile lies substantially above the mini-job threshold (only 2.4% men have only a mini-job in 2008 - see Table D.1) corresponding to individuals working in low-paying part-time jobs. The minimum wage introduction increased the 10th percentile slightly for men, but much less than for women.[17]

There are several potential drivers for these developments. Both men and women are more likely to work part-time in recent years (see Figure A.2). For men, the part-time share increased from around 3% to almost 10%, while for women it increased from around 30% to around 45%. While for men mini-jobs (as a primary job) never played a big role (only around 2-3%), the share for women is substantial (around 12-13% in 2001) and falling to around 6% towards the end of our sample

[17] Dustmann et al. (2022) also report that about two thirds of workers affected by the minimum wage were women and Caliendo and Wittbrodt (2021) find that this reform indeed decreases the gender wage gap.

period. Furthermore, there has been an increase in the college share among workers in our sample over the time period, an increase in non-German immigrants, and average age (see Figure E.15).

Reweighting Analysis. To gauge the importance of these observables for the evolution of the earnings distribution, we reweight our sample such that observable dimensions are held constant at the 2001 level (see Appendix E.2 for details). Figure 5 shows the counterfactual evolution of several percentiles for men and women when we use this reweighting procedure.

For men, job type (i.e., full-time / part-time / mini-job) plays a key role in explaining the fall in real earnings at the 10th percentile. This suggests that much of the increase in part-time occurred at these very low earnings levels pushing them down. Holding job type constant, the 10th percentile would have fallen less in the early 2000s and fully recovered to the 2001 level by 2016. Job type is also important in explaining the decline of the median up to the Great Recession, and the line is shifted upwards after controlling for this via reweighting. In contrast, for women holding job type constant does not affect the 10th percentile much, since those are always mini-jobbers, but has a huge impact on the median. If the part-time share had not increased, the median would have grown by almost 10 log points compared with the observed growth of around 3 log points. Even at the 90th percentile, the growth in part-time employment led to a significant slowdown in earnings growth, and after controlling for job type composition the P90 grows by an additional 5 log points.

Mean age first increases, peaks between 2010 and 2012 before slightly decreasing again, but staying above initial levels (see Figure E.15). However these modest changes are not large enough to meaningfully affect the earnings structure. Education levels rose gradually over our time period (see Figure E.15). Since education is positively associated with earnings, it is not surprising that after reweighting for this dimension, median earnings and the 90th percentile grow slightly slower.

A different shock to the German labor market was the opening of the market to workers from the new EU member countries in Eastern Europe that occurred on May 1st, 2011. As can be seen in Figure E.15, the share of non-German workers increased sharply from 2010 to 2016 (for men from 9 to 15% and for women from 6 to 10.5%). These workers typically work in low-wage jobs. Indeed, based on the reweighting analysis, they appear to significantly push down the 10th and 50th percentile for men and to a lesser extent the median for women.

Finally, annual earnings are of course in part determined by the number of days worked in a year. For example, if workers move between jobs with intermittent unemployment spells, this will decrease annual earnings for that year, even if wages remain the same. In Figure E.16, we show the average number of days worked for all workers as well as for above and below median earnings individuals. On average, days in employment increased slightly at the beginning of our sample period and decreased again towards the end, dipping during the Great Recession for men. This is consistent with the fall in unemployment after 2005. These changes in days in employment are driven almost solely by workers with below-median earnings. Indeed, the reweighting analysis in Figure 5 shows that this increase did substantially contribute to earnings growth for both genders in the lower half of the distribution.

Figure 5: Counterfactual Evolution of Log Earnings Percentiles (Reweighting)

(a) P10: Men

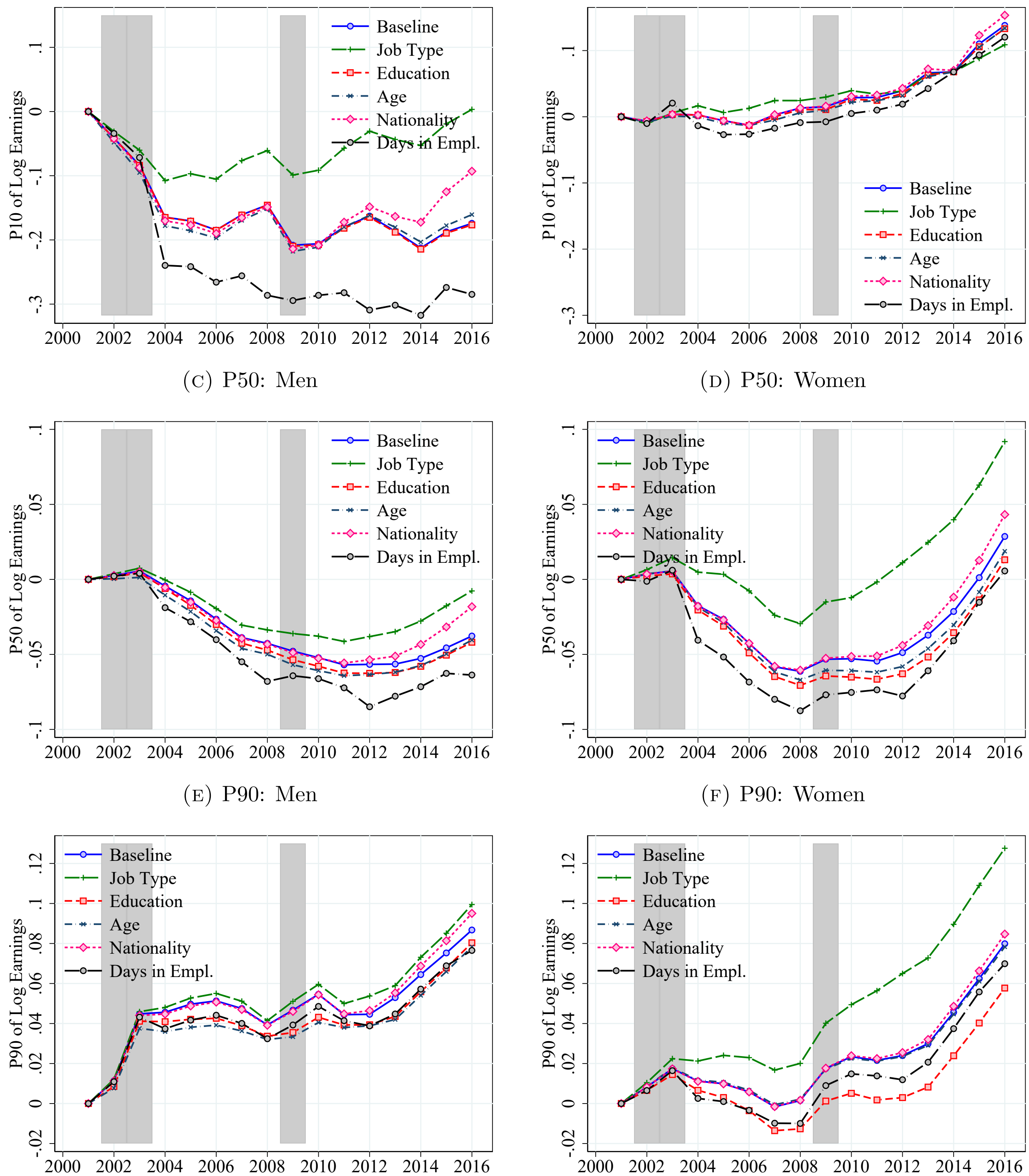
(b) P10: Women

(c) P50: Men

(d) P50: Women

(e) P90: Men

(f) P90: Women



Notes: This figure shows the evolution of different counterfactual percentiles of the log real annual earnings distribution over time in the IAB data (CS sample) by gender. The P90 for men is imputed in the IAB data as it lies above the social security contribution limit. The counterfactual percentiles are constructed by reweighting the data such that observable dimensions are held constant at the 2001 level (see Appendix E.2 for details). For example, the green line shows how different percentiles would have evolved over time had the job type distribution stayed as it was in 2001. A value of this counterfactual percentile above (below) the baseline value (blue lines) thus means that absent any change in the specific variable, earnings (at the given percentile level) would have been higher (lower) than what was actually observed. Thus, the observed change in the specific variable led to lower (higher) real earnings. Shaded areas indicate recessions. Figures E.17 and E.18 show the evolution of different counterfactual percentiles for each reweighted observable, while Figure E.19 shows counterfactual percentile differentials.

Inequality over the Lifecycle. The previous results showed markedly different developments for men and women. However, women's careers diverge from men's especially after childbirth (Kleven et al., 2019). We, therefore, turn to investigating inequality over the lifecycle. Since we are mostly interested in somewhat younger workers with fewer observations above the top coding threshold (which is always above the 90th percentile in the subsequent analysis), we use only the IAB data and therefore can show results up to 2018 as well as include further demographics.

Figure 6 shows the evolution of labor income over time for the four cohorts who are at the age of 25 in 2001, 2005, 2009, and 2013, respectively. Panel A shows median earnings for men. For all cohorts, earnings grow fast for young workers: around 60 log points in the first 10 years. While earnings at 25 are slightly higher for the 2001 cohort, earnings growth is faster for the later cohorts who make up the initial disadvantage. The decline in initial earnings may be due to the weak labor market from 2001 to 2005, leading to depressed wages, while the later cohorts enter during a much stronger labor market (see Figure A.2). Panel C shows within cohort inequality (P90-P10 gap) over the lifecycle among men for each cohort-by-year cell. It reveals that within cohort inequality falls rather fast over the first 10 years and then flattens out. In Figure E.9 we show that while the share of men working part-time (but more than a mini-job) is fairly constant over the lifecycle (at around 10%); young men are much more likely to work in mini-jobs compared to older men. This decline over the lifecycle is an important factor in reducing inequality. Another factor is that men who attain university degrees enter the labor market relatively late.[18] Moreover, workers at age 25 with college degrees have lower mean earnings than non-college workers but then quickly catch up and overtake the non-college group. This suggests that the decline in within cohort inequality is due to the decline in mini-jobs and entry of college workers in the middle of the income distribution.[19]

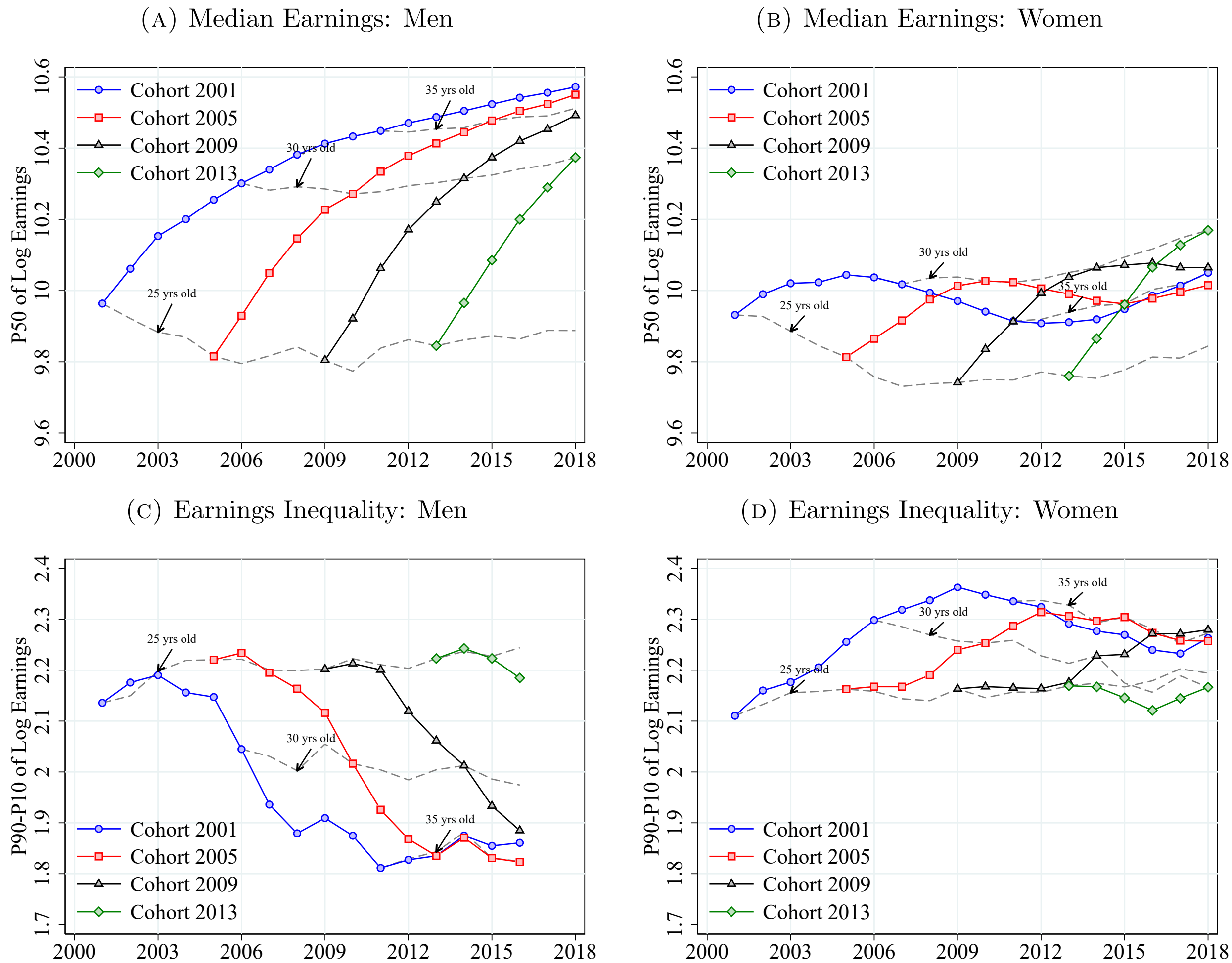
For women, the pattern is markedly different. While median earnings and inequality are almost the same as for men at age 25, median earnings rise much slower. Furthermore, unlike for men, earnings inequality rises continuously from age 25 to 35. This is likely since many women have children in their late 20s and early 30s (Bundesinstitut für Bevölkerungsforschung, 2021) and then transition to working part-time afterward (Schrenker and Zucco, 2020).[20] Figure E.9 shows that while only around 15% of women work part-time at age 25, this increases to over 50% by age 40. In addition, the mini-job share falls much less strongly for women than for men (for the 2001 cohort it even increases somewhat initially). Additional analysis (Figures E.10 and E.11) reveals that this is the combination of a rising mini-job share with age for non-college women and a sharp fall in mini-jobs with age for women with college degrees. Furthermore, while part-time increases with age

[18] During our sample period University-qualifying high school degrees (Abitur) took 13 years and men were regularly spending a year in civil or military service, followed by typically 5-6 years of University. In fact, Figure E.9 shows that the share of workers with college degrees is only around 18% at age 25 and increases to 25% by age 30.

[19] Consistent with this we show in the Appendix, that the decline and flattening of the P90-P10 log earnings differential with age is the result of offsetting trends in lower and upper inequality: Figure E.8 shows that while the P50-P10 declines, the P90-P50 increases as workers pass the age of around 30.

[20] The age for having the first child for women increases over time (Institut Arbeit und Qualifikation, 2021). Kleven et al. (2019) and Bönke et al. (2022) find that child penalties in female earnings are especially pronounced in Germany. Interestingly, the latter paper finds that participation and earnings gaps for women after birth increased until the 1990s but then this trend was reversed. This fits our findings of somewhat larger median earnings profiles and lower/declining inequality profiles of younger female cohorts in Figure 6.

Figure 6: Earnings Profiles and Inequality by Cohort



Notes: This figure shows the evolution of the median as well as the P90-P10 differential of the log real annual earnings distribution over time in the combined IAB-TPP data (CS sample) by gender. As the P90 of men is imputed and the TPP data end in 2016, Panel C also ends in 2016. Grey dashed lines correspond to earnings of 25, 30 and 35 year olds in each year as indicated by arrows. Each colored line corresponds to an individual cohort, where "cohort t" represents the cohort aged 25 in year t.

for all women, the increase is stronger and earlier for women without a college degree (likely because college educated women have children later). Thus, at young ages non-college educated women tend to move to part-time just as more college educated women enter the labor market working full-time, thus pushing up inequality. Then around age 30 college educated women start working part-time (or drop out of the labor force) leading to a mild decline in within cohort inequality.

Long-term Evolution. Our main analysis focused on the period from 2001 to 2016. To put these results into context with the long-term development, we replicate the key figures for two longer samples in the IAB data. The downside of the longer time frame is that prior to 1999 the IAB data does not contain mini-jobs, so we can only capture the earnings distribution above the mini-job threshold. In addition, we have to rely on imputed values for the 90th percentile for men (for women it is still below the top coding threshold).

The first long sample starts right after Germany's reunification covering all workers from 1993 to 2018 (see Appendix F.1). Figures F.1 and F.2 show that the increase in earnings inequality in the 2000s, was preceded by even faster increases in inequality in the 1990s. The overall growth in inequality was also much larger for men than for women, both in the 1990s and 2000s. Despite the data differences, the figure is fairly consistent with the post-2001 sample, except for the 10th percentile for women which is by construction very different. The lifecycle plots for this sample (Figure F.4) show very similar patterns for average earnings growth (much faster for men than for women) for the full period. A difference is that, due to the lack of data on mini-jobs, the fall in within cohort inequality is less pronounced for men, while the increase is even stronger for women.

Focusing on West Germany only, we can show results starting in 1985 (see Appendix F.2).[21] This sample shows that the increase in inequality started in the late 1980s. Even over this very long horizon, there are marked gender differences. Women at the lower percentiles fared much better than men, while the median and 90th percentile are closer to each other. As a result, men experienced a much larger increase in earnings inequality, especially at the bottom of the distribution. The lifecycle plots are similar as in the 1993-2018 sample, showing that the stark differences in lifecycle profiles by gender have been a feature of the German labor market for a long time.

3.2 Earnings Dynamics

From the vantage point of an individual worker, the earnings dynamics over time are arguably just as important as cross-sectional earnings differences as earnings risk affects key economic decisions such as consumption and savings. Therefore, we analyze the distribution of earnings growth, g^1_{it}, over time. It is defined as 1-year changes in residualized log earnings where we take out year- and gender-specific age effects.[22] To limit the effect of extremely large changes, we rely on percentile-based measures and report percentile differentials, Kelley skewness, and excess Crow-Siddiqui kurtosis instead of standard deviation and higher moments of the earnings growth distribution.[23]

Similar to the distribution of earnings, neither IAB nor TPP data alone allow us to estimate the true distribution of earnings growth.[24] To obtain this distribution, we combine IAB and (reweighted)

[21] For a detailed analysis of full-time wage inequality by gender from 1975 to 2004, see Dustmann et al. (2009).

[22] 1-year earnings growth in year t is defined as $g^1_{it} = \epsilon_{i,t+1} - \epsilon_{it}$ where $\epsilon_{it} = y_{it} - x'_{it}\hat{\beta}$ is the residual of a regression of log earnings on gender and year specific age dummies. Figure E.22 shows the density of g^1_{it} for the year 2005. The corresponding results for 5-year earnings growth are reported in Appendix E.3.

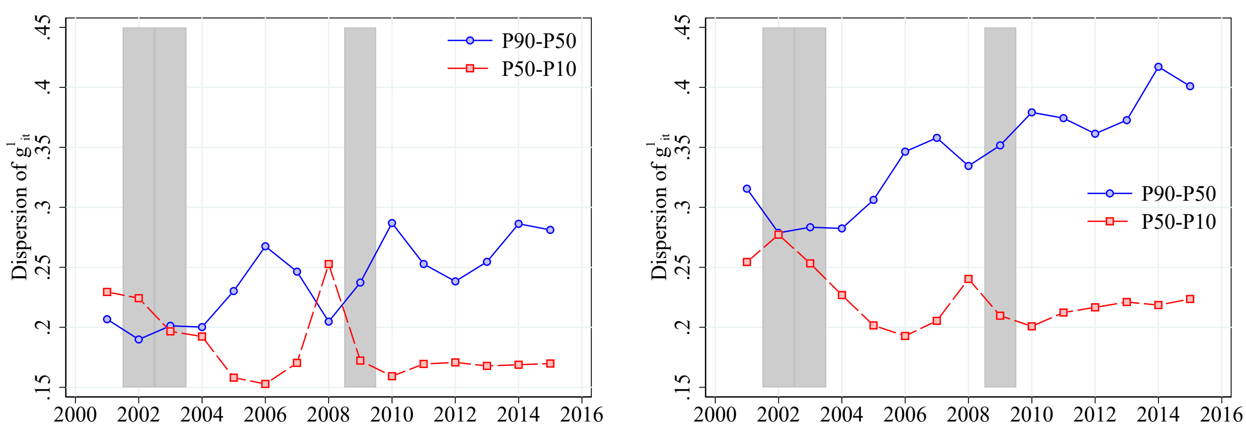
[23] Kelley skewness is defined as $\frac{P90-2P50+P10}{P90-P10}$. A positive (negative) value implies that the right tail of the distribution is longer (shorter) than the left one. Excess Crow-Siddiqui kurtosis is defined as $\frac{P97.5-P2.5}{P75-P25} - 2.91$ (i.e. Crow-Siddiqui kurtosis minus 2.91 – its value for a Gaussian distribution). Larger (positive) values indicate a leptokurtic distribution having a lot of mass around zero and relatively fat tails. In the context of earnings growth, this implies that workers experience fewer medium-sized shocks but more extreme positive or negative shocks.

[24] In the IAB data, top-coding implies that a substantial share of growth rates is calculated from imputed values. As the imputation does not take dynamics into account, this inflates growth rates. In the TPP data, most transitions in and out of mini-jobs are not observed. In addition, the absence of non-filers prior to 2012 distorts the distribution of growth rates. Non-filers are likely to have less volatile earnings, while, for example, workers who switch jobs or receive unemployment insurance are obliged to file a tax return and also experience relatively large earnings changes.

TPP data following a three-step procedure. First, we estimate growth distributions conditional on gender and current earnings separately using IAB and TPP data. Second, we construct a combined conditional distribution of earnings growth. That is, we use the conditional growth distribution from the IAB data for low levels of current earnings where only very few (less than 2%) growth rates are affected by top-coding, and the conditional growth distribution from the TPP data above. This cutoff lies between 45,000 and 50,000 Euro depending on gender and year. Finally, we integrate the conditional distribution with respect to the combined IAB-TPP earnings distribution in the LS sample to obtain the unconditional distribution of earnings growth by year and gender.[25]

Figure 7: Dispersion of 1-Year Log Earnings Changes

(A) Men (B) Women



Notes: This figure shows the P90-P50 and P50-P10 differentials of the distribution of 1-year changes in residualized log real annual earnings (from t to $t+1$) in the combined IAB-TPP data (LS sample) for men and women. See Appendix D.2.2 for details on the construction of the log earnings growth distribution from IAB and TPP data. Shaded areas indicate recessions.

Figure 7 shows the P90-P50 and P50-P10 differentials of 1-year log earnings changes over time. The distribution of men's earnings is more stable in the sense that large negative and (especially) positive changes are less likely.[26] Yet, men's earnings are more strongly affected by the business cycle, in particular by the Great Recession: Large positive (negative) changes become less (more) likely. The fact that the right tail shrinks and the left tail expands in the Great Recession directly explains the drop in the Kelley skewness for men shown in Panel A of Figure 8. Quantitatively, this drop is small compared to other GID countries such as Italy, Spain or the US (Hoffmann et al., 2021; Arellano et al., 2021; McKinney et al., 2021) and of similar magnitude as in the UK and Sweden (Bell et al., 2021; Friedrich et al., 2021). While the qualitative pattern is the same for women, the

[25] Appendix D.2.2 describes this procedure in detail. Figures D.3 and D.4 compare key statistics of the earnings growth distributions conditional on current earnings in IAB and TPP data. Importantly, for intermediate levels of earnings, both datasets deliver highly consistent results. Figure D.5 shows the share of growth rates affected by top-coding in the IAB data. Figure D.2 shows the combined IAB-TPP distribution of earnings in the LS sample that is constructed analogously to the CS sample: Below the top-coding threshold we use the IAB distribution while above we use the TPP distribution rescaled to match the total number of workers above the threshold in the IAB data.

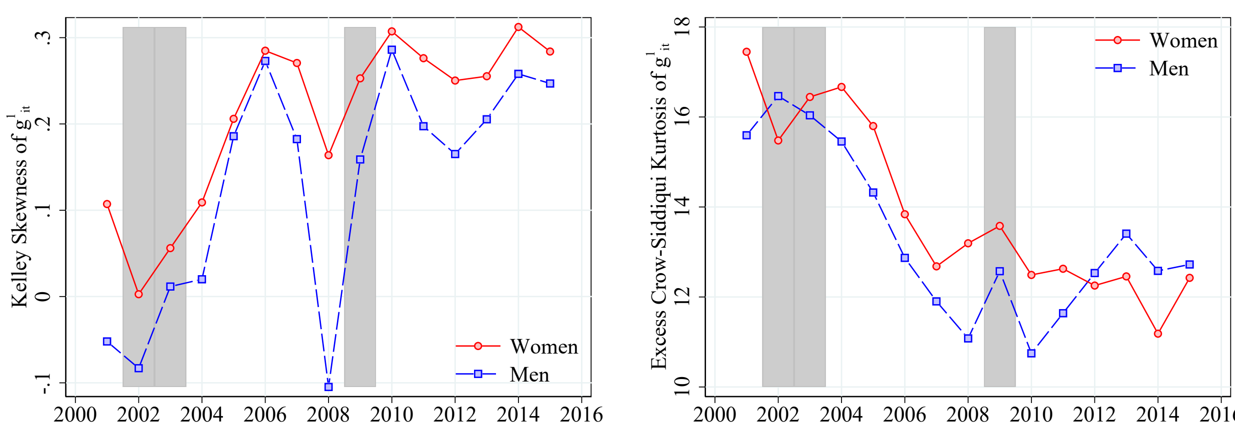
[26] As shown in Figure E.20, the median 1-year log-earnings change is approximately zero for all years and for both men and women. Hence, below-median earnings changes are negative and above-median earnings changes are positive.

cyclicality of earnings changes is much lower. This is consistent with Doepke and Tertilt (2016) and Alon et al. (2021) who show that recessions (prior to COVID-19) tend to affect men more severely.

Besides its pro-cyclicality, which has been documented before (e.g. Guvenen et al., 2014; Busch et al., forthcoming), it is noteworthy that after 2005 the Kelley skewness of earnings growth is more positive than before as well as compared to other countries (e.g. Italy, Spain, Sweden, UK, US – see Hoffmann et al. (2021); Arellano et al. (2021); Friedrich et al. (2021); Bell et al. (2021); McKinney et al. (2021)).[27] This reflects the good overall labor market conditions in Germany over this time period (Figure A.2). The low risk of becoming unemployed limits large earnings losses, compresses the P50-P10 differential, and increases the Kelley skewness. In addition, increasing labor supply may explain the relatively high share of large positive changes.[28] At the extensive margin, a previously unemployed person who starts working has significantly lower annual earnings in the first (incomplete) year than in the second (complete) year of an employment spell. At the intensive margin, workers increasing working hours experience a substantial increase in earnings.

Figure 8: Skewness and Excess Kurtosis of 1-Year Log Earnings Changes

(a) Kelley Skewness (b) Excess Crow-Siddiqui Kurtosis



Notes: This figure shows the evolution of Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real annual earnings (from t to $t+1$) in the combined IAB-TPP data (LS sample) by gender. See Footnote 23 for definitions and interpretations of Kelley skewness and excess Crow-Siddiqui kurtosis. See Appendix D.2.2 for details on how we construct the distribution of log earnings growth from IAB and TPP data. Shaded areas indicate recessions.

Figure 8 (B) shows that earnings growth is substantially more leptokurtic than a Gaussian distribution with the same variance. Both the level and evolution of excess Crow-Siddiqui kurtosis are similar for men and women. For both, the kurtosis of earnings growth declined from about 16 to 12 between 2004 and 2007. Hence, medium-sized earnings changes became more important relative to large changes. For women, the decline is mainly driven by the strong increase in the P75-P25 differential of earnings growth that dominates the widening of the tails. For men, the decline is driven both by moderate increases in the P75-P25 and decreases in the P97.5-P2.5 differential.[29]

[27] The Kelley skewness of earnings growth is also positive when looking at 5-year changes (see Figure E.27).

[28] Figure A.2 shows a substantial increase in labor force participation after 2005 (especially for women). Moreover, transition probabilities from mini-job to part- or full-time increase by 5–10 percentage points (see Figure E.24).

[29] See Figure E.21 for a decomposition of the evolution of excess Crow-Siddiqui kurtosis.

3.3 Heterogeneity in Earnings Dynamics by Age and Permanent Earnings

We now study heterogeneity in earnings dynamics by age and workers' position in the permanent earnings distribution using the H sample. Permanent earnings P_{it} are defined as the residual of the log of average earnings between $t-2$ and t.[30] To avoid mechanical mean reversion, we analyze the distribution of (residual) log earnings changes between t and $t+1$ conditional on past permanent earnings $P_{i,t-1}$. As we are now interested in conditional earnings growth distributions, we use those from the IAB below and those from the TPP above the cutoff of 45,000 Euro.[31]

Figure 9 shows how different moments of the 1-year earnings growth distribution depend on age and the permanent earnings distribution. Panels A and B reveal that young workers' earnings are more volatile. However, while for men this is only relevant in the bottom half of the distribution, this holds across the entire distribution for women. The P90-P10 differential for young women is 30–50 log points higher compared to women above 35. Volatility is U-shaped in permanent earnings. Male workers in the bottom half and at the very top experience substantially larger changes than workers between the median and the 95th percentile of the distribution. For women, the P90-P10 decreases steadily until the 80th percentile and spikes back up above the 90th percentile.[32]

The Kelley skewness of earnings growth is decreasing in permanent earnings for both men and women (Panels C and D) albeit this gradient is larger for women. The fact that young women's Kelley skewness is smaller than for the other groups and even negative for the top 70% (Panel D) suggests that earnings volatility is higher for young women because they experience disproportionately many large earnings losses. This mainly reflects reduced female labor supply following childbirth. As children get older and mothers rejoin the labor force or switch from marginal to part-time or from part- to full-time jobs, they experience substantial earnings gains. Hence, the Kelley skewness of earnings changes for women between 35 and 55 is positive (except for the top 30% where it is close to zero). For men, the Kelley skewness measure does not vary as much over the lifecycle (Panel C). Consistent with men's concave lifecycle profile for median earnings (Figure 6), large positive changes become relatively less likely and the Kelley skewness drops as men get older.[33]

Panels E and F show that the relationship between excess Crow-Siddiqui kurtosis and permanent earnings differs between men and women. While earnings growth is more leptokurtic in the bottom half of the male distribution, the opposite is true for women. In terms of lifecycle profiles, older women and younger men differ relative to the other age groups. Especially for older women the excess kurtosis profile is flat across the earnings distribution.

[30] We compute permanent earnings P_{it} by regressing the log of average past earnings on a full set of gender and year-specific age dummies, taking the residual and adding the raw sample mean.

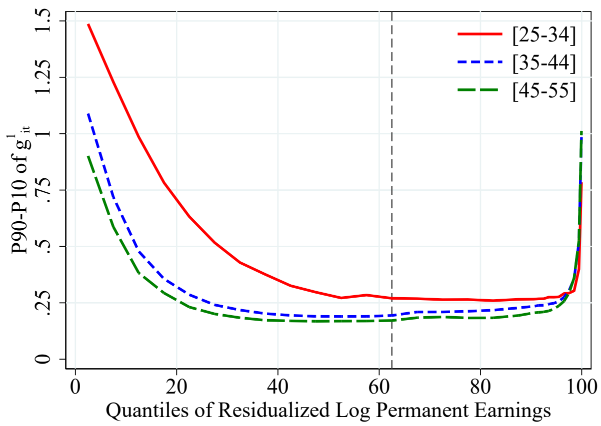
[31] See Appendix D.3 for further details. In particular, we show that the distributions of permanent earnings in the IAB and TPP data are almost identical in the middle of the distribution and differ only at the bottom and very top. We also show the conditional earnings growth distribution by permanent earnings in both datasets separately.

[32] This tilted U-shape is consistent with findings for Sweden, Denmark, Norway, France, and the US (Friedrich et al., 2021; Leth-Petersen and Sæverud, 2021; Halvorsen et al., 2021; Kramarz et al., 2021; McKinney et al., 2021).

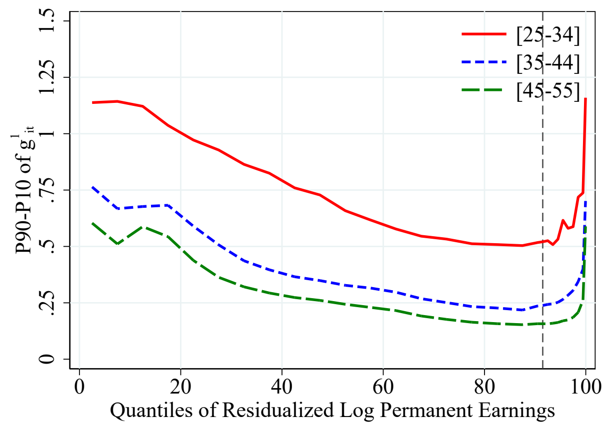
[33] Figure E.28 shows that this also holds for 5-year log earnings changes. The main difference is that earnings growth of male workers between 45 and 55 is negatively skewed.

Figure 9: Heterogeneity in Dispersion, Skewness and Kurtosis of 1-Year Log Earnings Changes

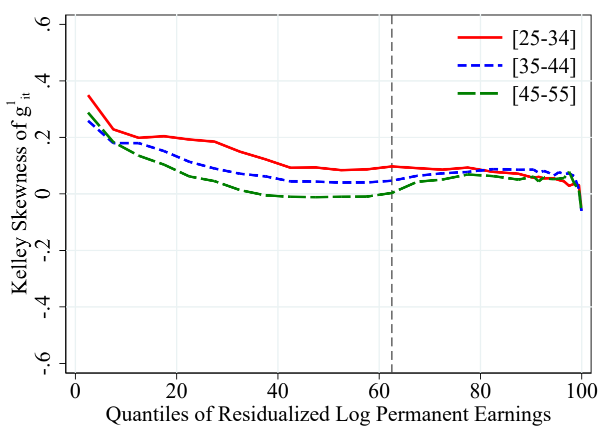
(a) P90-P10: Men



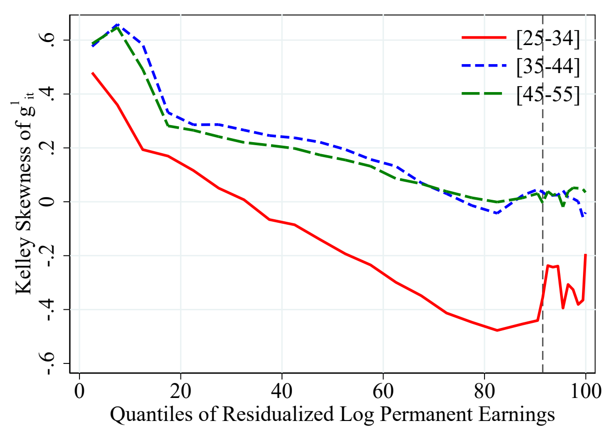
(b) P90-P10: Women



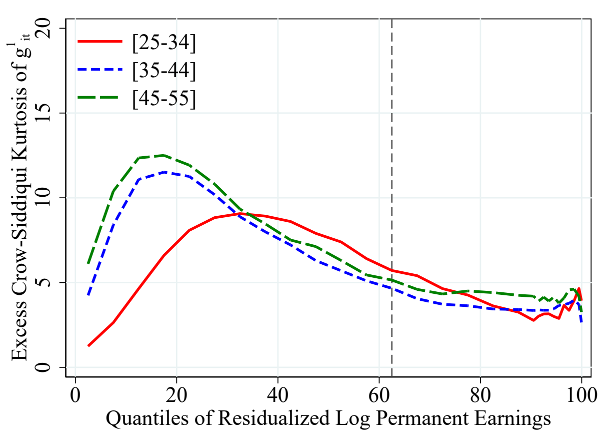
(c) Kelley Skewness: Men



(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



(f) Excess Crow-Siddiqui Kurtosis: Women

Notes: This figure shows the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real annual earnings (from t to $t+1$) by quantiles of residualized permanent earnings and age groups in the combined IAB-TPP data (H sample) as averages from 2004 to 2011 by gender. Permanent earnings $P_{i,t-1}$ are defined as the residual (net of a full set of gender and year specific age dummies) of the log of average earnings between $t-3$ and $t-1$. See Footnote 23 for definitions and interpretations of Kelley skewness and excess Crow-Siddiqui kurtosis. See Figures D.7, D.8 and D.9 for a comparison of the underlying data in both data sources.

4 Inequality and Dynamics of Total Income

In this section, we move beyond labor earnings by adding entrepreneurs and non-labor income to the picture in order to study overall income inequality and income dynamics of all taxpayers.

Total Income and its Components. Total income is defined as the sum of labor and non-labor (i.e. self-employment, business or rental) income. While total income can be negative (less than 2% of the observations in our data), we still impose the same minimum income threshold of 2,300 Euro as for labor income. This has two reasons: First, the results are comparable to those for earnings. Second, we avoid computing growth rates between negative and positive values.[34]

Table 2 shows descriptive statistics for our analysis sample (for the year 2008). We exclude capital income (interest, dividends and capital gains from low-stake investments) from the total income analysis as there are several changes in capital income taxation that make the information on capital income unreliable after 2008.[35] Note, however, that capital income from high-stake investments (ownership share of at least 1%, "wesentliche Beteiligung") are included in the data as they count as business income. Table 2 (B) shows that capital income amounts to only 1.3% of total income for men and even less for women (0.5%) in 2008. As a comparison, other non-labor income accounts for 16.4% of men's and 10.6% of women's total income. Thus, omitting capital income should only have small effects on the total income distribution.[36]

Panels A and B of Figure 10 show the share of each component in total income conditional on total income. As expected, labor income accounts for the lion's share of total income except for very high incomes (above 200,000 Euro). Self-employment income is especially relevant for individuals with income between 100,000 to 1 million Euro. The share of business income rises continuously after 50,000 Euro and it becomes the dominant source after around 1 million Euro. While it is not surprising that the very rich receive mostly business income, the share of non-labor income is not monotonically increasing in total income. At the bottom, non-labor income accounts for up to 20% of total income. This is because, compared to labor income, we observe relatively more individuals with very high but also very low business income.[37] In other words, the distribution of non-labor income is much wider than that of labor income (see also Panels C and D of Figure 10).

Furthermore, compared to Section 3, the sample additionally includes non-social-security workers and individuals without labor but with non-labor income. We refer to individuals who receive most of their income from labor as "workers" and to those who receive at least half of their income

[34]In Figure D.14 we show the share of non-zero and negative values for each income component over time. The fact that there are both trends and breaks in the time series is one reason why we stay away from negative values. Table D.11 shows how our analysis sample compares to the data without the threshold.

[35]After the introduction of a dual income tax in 2009 ("Abgeltungsteuer-Reform") capital income largely disappeared from personal tax records.

[36]While capital income is most important at the top, its share never rises above 9% (see Figure 10, A and B). On average, the share is below 0.9% for men and 0.4% for women (averaged over 2001–2008; the values for 2008 in Table 2 (B) are the highest in this period) – independent of whether we exclude capital income from total income or whether we impose the minimum income threshold.

[37]For women, we observe many low-earnings observations due to the high share of mini-jobs.

Table 2: Summary Statistics for Total Income Data

	Men	Women
Observations (in mill.)	14.667	12.351
A. Income Distribution		
Mean	45,810	26,163
P50	36,620	21,698
P90	78,113	48,393
P99.9	696,521	270,828
P99.99	2,919,253	931,065
B. Share of Total Income		
Labor	0.836	0.894
Non-Labor	0.164	0.106
Self-Empl.	0.054	0.043
Business	0.109	0.053
Rental	0.001	0.011
Capital*	0.013	0.005
C. Main Income Source		
Workers	0.882	0.918
Entrepreneurs	0.118	0.082
Self-Employed	0.026	0.024
Business Owners	0.082	0.036
Landlords	0.010	0.021
D. Non-Zero Income		
Labor	0.895	0.934
Non-Labor	0.300	0.207
Self-Empl.	0.052	0.049
Business	0.165	0.080
Rental	0.144	0.102
Capital*	0.103	0.042

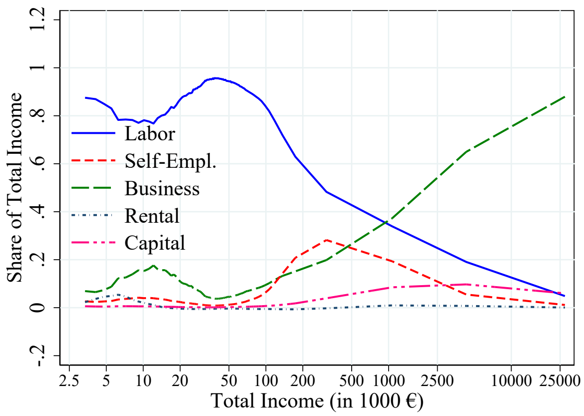
Notes: This table shows descriptive statistics for the combined IAB-TPP data by gender for the year 2008. The data includes all individuals with labor and non-labor income. See Appendix D for details on how we construct this combined dataset. Our analysis sample is restricted to individuals with total income (excluding capital income) above the minimum threshold of 2,300 Euro (2018 prices) and between 25 and 55 years of age. Panel A shows the mean and selected percentiles of the total income distribution (in 2018 Euro). Panel B shows the share of each income source in total income. Panel C reports the share of observations by main income source. Panel D shows the share of observations with non-zero income from different sources.

from non-labor sources as "entrepreneurs". Among the latter, a person is called "self-employed" if the main income source is self-employment, "business owner" if it is business income or "landlord" if it is rental income.[38] Panel C of Table 2 shows that 11.8% of men are entrepreneurs with the majority of them classified as business owners (8.2%) or self-employed (2.6%) and few landlords (1.0%). While the share of self-employed is similar for women (2.4%), fewer women are business owners (3.6%) and more are workers (91.8%) or landlords (2.1%). While this classification is intuitive, we

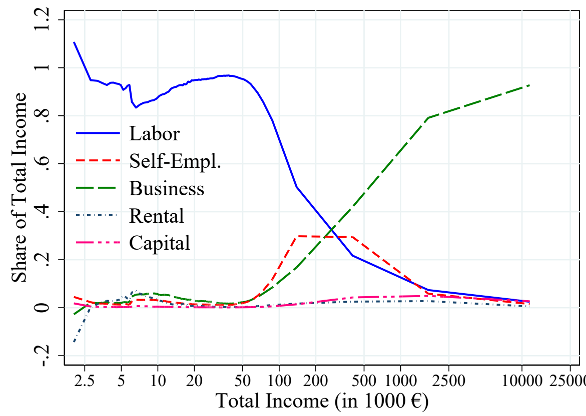
[38]The distinction between self-employed and business owners is somewhat special in the German tax code as both own their business. Self-employed are high-skilled independent professionals with special qualifications ("Freiberufler") such as doctors, lawyers, journalists or tax consultants. Any other self-employed individuals (e.g., delivery drivers) or owners of firms that produce or sell products are referred to as business owners ("Gewerbetreibende").

Figure 10: Income Components

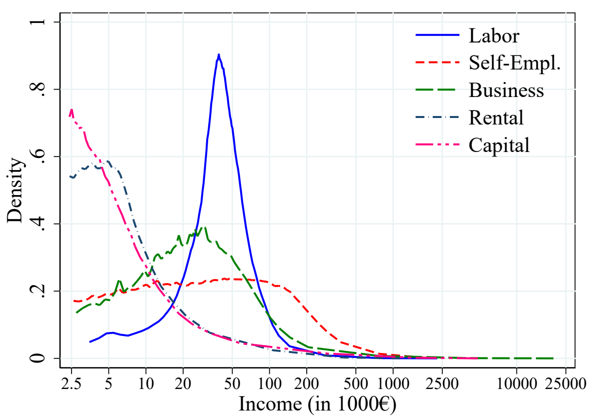
(A) Component Shares by Income: Men



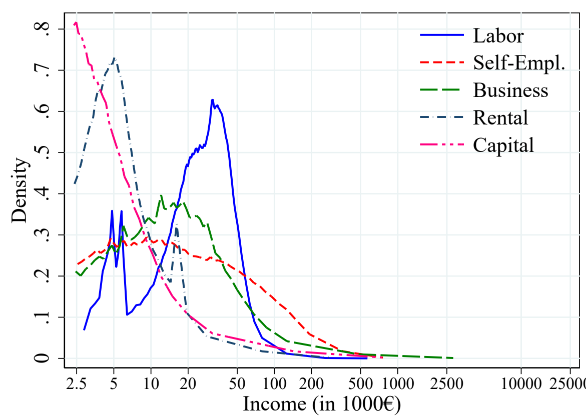
(B) Component Shares by Income: Women



(C) Densities: Men



(D) Densities: Women



Notes: This figure shows different statistics for the components of total income in the combined IAB-TPP data by gender (averages from 2001 to 2008). Panels A and B show how total income (including capital income) is split into labor, self-employment, business, rental and capital income across the total income distribution. Panels C and D show the densities of each income sources for all individuals with income from the respective source above 2,300 Euro (in 2018 Euro).

emphasize that non-labor income is not only relevant for entrepreneurs but also for workers. Indeed, Table 2 (D) shows that 30% of men and 20.7% of women have some non-labor income.

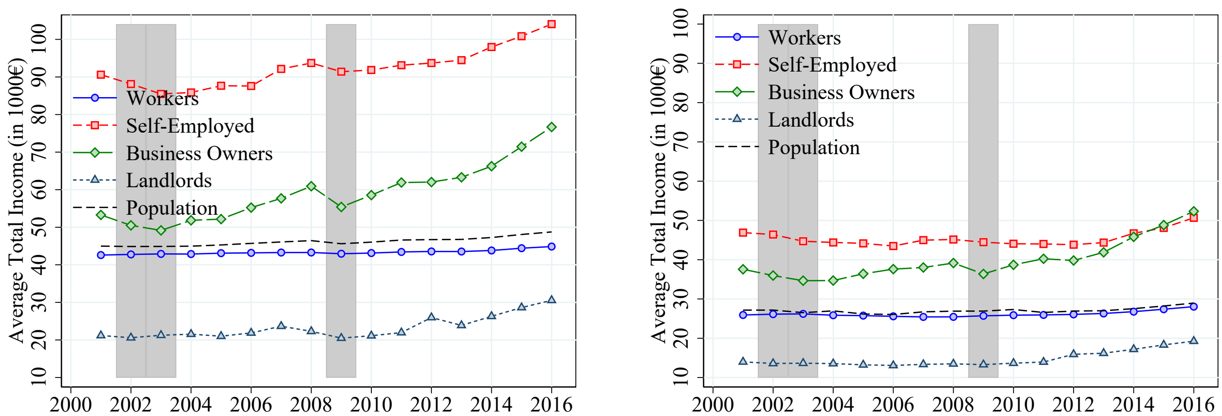
4.1 The Distribution of Total Income

We now turn to the evolution of the total income distribution over time. Total income inequality is driven by income differences among workers (as discussed in Section 3) or entrepreneurs, but also by income differences between those groups.

Figure 11 shows that average incomes differ substantially between workers and entrepreneurs. While male workers (Panel A) receive average annual income of around 43,000 Euro in 2001, business owners and self-employed receive 53,000 and 90,000 Euro respectively. Landlords, on the other hand, earn only 20,000 Euro in 2001. Panel B reveals a similar pattern for women. However, there is a substantial gender gap in average income not just among workers but also among self-employed (53.3%), business owners (34.8%) and landlords (40.3%). Strikingly, the overall gender gap increased from 39.6% in 2001 to 40.6% in 2016 despite the fact that the gender gap among workers decreased from 39.1% to 37.5%.[39] Hence, the gender gap among non-workers has increased significantly.

Figure 11: Average Total Income by Main Component

(A) Men (B) Women



Notes: This figure shows average total income (excluding capital income) for different sub-populations defined by the main income source as well as the full population (gray dashed line) over time in the combined IAB-TPP data by gender. Shaded areas indicate recessions. Figure G.2 shows the evolution in terms of log differences relative to 2001.

Over time, we find that average incomes of workers are remarkably flat compared to entrepreneurs.[40] For men, average incomes of self-employed and business owners decrease in the recession of 2002 and 2003, but then increase significantly. While workers' incomes only rise from around 43,000 to 45,000 Euro between 2003 and 2016, business owners' incomes grow from around 50,000 to 75,000 Euro and incomes of self-employed men go from roughly 85,000 to 105,000 Euro over that period. Average incomes of landlords also increase by much more than workers' incomes but the increase takes place only after the Great Recession (possibly due to a combination of lower mortgage interest payments and rising real estate prices). Panel B shows that female business owners and landlords exhibit a similarly strong increase in average income, while average income of self-employed women is essentially flat between 2003 and 2013 before it starts to increase. Overall, these differences between entrepreneurs and workers underline the importance of looking at the distribution of total income to get a more complete picture of income inequality over time.

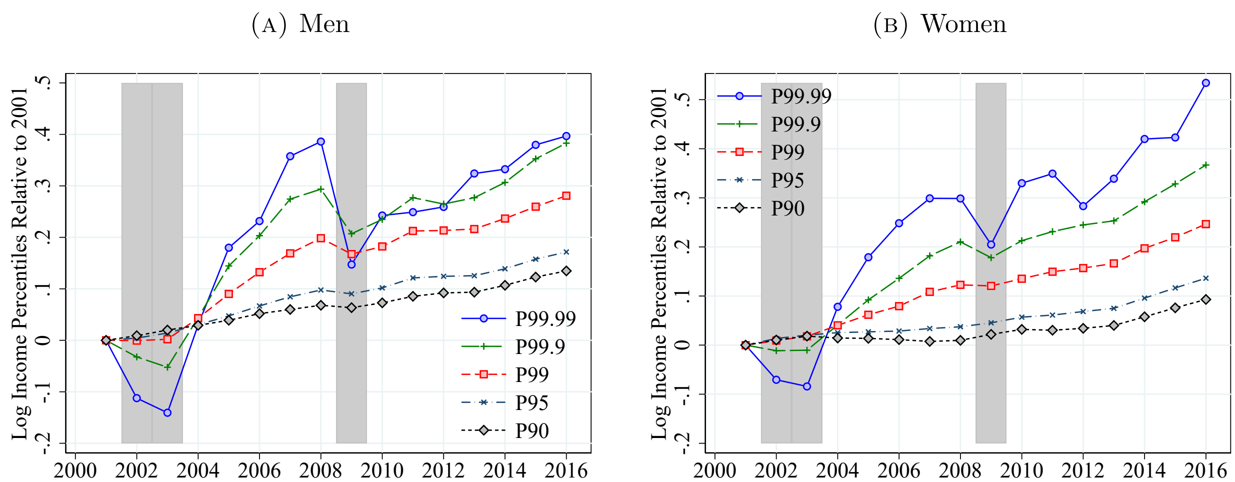
Table D.12 in the Appendix shows selected percentiles of total income. While percentiles below the 90th percentile are very similar to those of labor income (Table 1), the right tail of the distribution is much longer for total incomes. In 2016, for example, the 90th percentile of total income for men is about 9% higher than that of labor income (84 vs. 77 Euro). This gap grows substantially as we move to the very top of the distribution. While the 99.99th percentile of labor income is around

[39] In the data on social security workers in Section 3, the gender gap in average earnings decreased even more strongly from 42% to 37% between 2001 and 2016 (see Table D.1).

[40] Figure G.2 shows the evolution of average income by main income source as log differences relative to 2001.

1.2 million Euro, it is close to 3 million Euro for total income – a factor of 2.5. For women, this factor is even larger (2.8) although the levels are considerably lower (400,000 vs. about 1.2 million Euro). These findings again highlight the large gender differences throughout the distribution and especially at the top: While men earn more than women at all percentiles, at the very top men's incomes are almost three times as large as women's.

Figure 12: Evolution of Top Log Total Income Percentiles

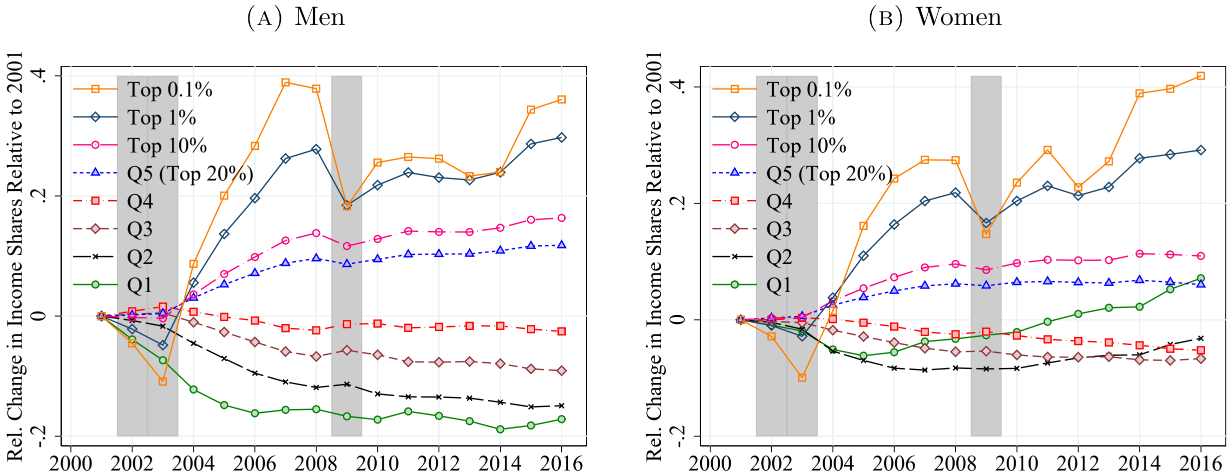


Notes: This figure shows the evolution of selected top percentiles of log real annual total income (relative to 2001) in the combined IAB-TPP data (CS sample) by gender. Shaded areas indicate recessions. See Figure 3 for the same analysis of only labor earnings (albeit for a slightly different sample as discussed in the text) and Figure G.3 for more percentiles of the total income distribution.

To analyze of the changes in total income inequality over time, Figure 12 shows the evolution of selected top percentiles relative to 2001. Not surprisingly, lower percentiles of the overall distribution appear very similar to those for labor income (compare Panels A and B in Figures G.3 and 3). However, at the top of the distribution (Figure 12) total income percentiles grow much faster. For example, while at the 99.99th percentile men's labor incomes grew by 20 log points, total income grew by around 40 log points. For women the growth at the top of total income is also stronger than for labor income but the difference is not as pronounced as for men. Indeed while top earning women seemed to clearly catch up with men based on Figure 3 (C) and (D), in Figure 12 (A) and (B) only the 99.99th percentile grows faster than for men, while all other top percentiles grow less.

Taking a closer look at the very top of the income distribution, it becomes apparent that total incomes show larger fluctuations over the business cycle than labor earnings. Especially the P99.99 (i.e. the top 0.01%) of men are hit by large negative shocks of 15 (in 2003) and 25 (in 2009) log points in the two recessions in our period of analysis. While top income men recover very quickly after the 2003 recession, the recovery is more muted after 2009 and only reaches the 2008 level again in 2016. The top 0.1% also show a similar pattern albeit of lesser magnitude while the top 1% grew relatively steadily over time. Business income at the high end is of course dependent on the profitability of the respective businesses which may fluctuate much more over the cycle than wages. Business income may also be affected by deductions of negative incomes in subsequent years.

Figure 13: Changes in Total Income Shares Relative to 2001



Notes: This figure shows the evolution of selected income shares of real annual total income (relative to 2001) in the combined IAB-TPP data (CS sample) by gender. The absolute shares for all groups can be found in Appendix Tables G.2, G.3 and G.4 for men, women and the entire population. Shaded areas indicate recessions.

As an alternative way to highlight changes in the income distribution, Figure 13 shows the relative changes in income shares of various groups of the total income distribution.[41] For men (Panel A), the bottom 4 quintiles show substantial losses, especially in the 2001-2008 window. For example, the bottom quintile share fell from 5.6 to just 4.6% (or by around 10%). By contrast, the income share of the top quintile rose from 42 to 47% (or by around 11% relative to 2001). For women (Panel B) the increase at the top is somewhat more muted from 42 to 45%. The relative increase of the lowest quintile (Q1) is striking, however this is from a low level (4 to 4.2%). Zooming in at the very top shows a similar strong (cyclical) increase as before both for men and women.

4.2 Income Dynamics for Workers and Entrepreneurs

We now study the distribution of total income growth, measured as 1-year changes in residualized log total income using the IAB-reweighted TPP data (as the IAB data itself does not contain information on total income)[42] for workers and entrepreneurs.[43] For both groups we analyze 1-year

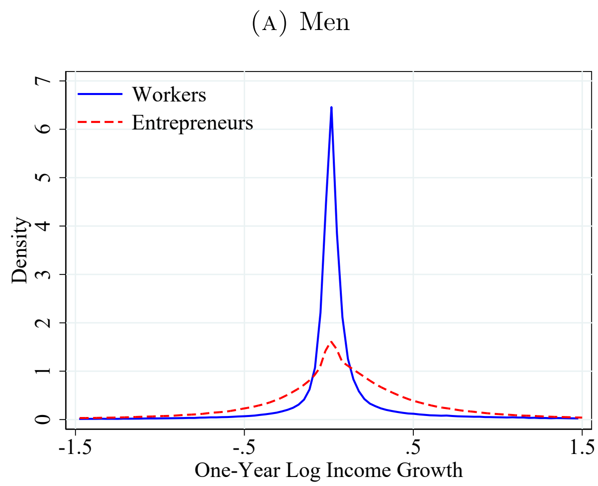
[41] Our results are consistent with those of Bartels (2019) or the World Inequality Database, though our top shares are a bit lower. Differences are a narrower age window (25–55 vs 20–60), analyzing individuals instead of tax units as well as not imputing capital incomes. Jenderny and Bartels (2015) analyze how the absence of capital incomes from the post-2009 tax data affects top income shares.

[42] This means that we cannot use the procedure used in Section 3.2 to correct for non-random attrition in the TPP before 2012. However, this only affects pure labor income earners as observations with any non-labor income always have to file a tax return. In addition, attrition will be more problematic for women than for men as most mini-jobs are missing in the TPP. However, the evolution (and levels) of the statistics of total income growth for workers reported in this section are consistent with the results for earnings growth in Section 3.2. In addition, there are no clear breaks from 2011 to 2012 when most of the attrition problem vanishes with the inclusion of non-filers in the TPP.

[43] The distribution of 1-year income changes in the population is very similar to that of workers as 88% of men and 92% of women are classified as workers. In Appendix G, we show that income changes of self-employed are slightly more similar to those of workers but still much more similar to those of other business owners.

changes in *total* income. The densities in Figure 14 give a first indication that the distribution of 1-year income growth differs substantially between workers and entrepreneurs.

Figure 14: Density of 1-Year Income Growth by Main Income Source (Year 2005)



Notes: This figure shows the density of 1-year changes of residualized log total income separately for workers (labor income as main income source) and entrepreneurs (non-labor income as main income source) for men and women in the combined IAB-TPP data (LS sample) for the year 2005.

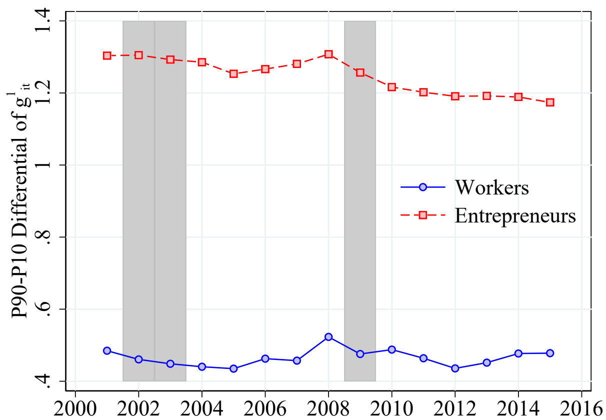
Figure 15 compares the evolution of the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year income growth between workers and entrepreneurs. Focusing on men, we find that entrepreneurs' income are substantially more volatile (Panel A). That is, the P90-P10 differential is three times as large as that of male workers (see Panel A of Figure 15). This means that about 80% of workers' 1-year income changes are smaller than 25 log points while about 20% of entrepreneurs' 1-year income changes exceed 70 log points (in absolute value). Panel C shows that while workers' log income changes are (mostly) positively skewed, Kelley skewness of entrepreneurs' income changes is essentially zero. Perhaps surprisingly, male entrepreneurs' log income changes are far less cyclical than male workers' income changes. In particular, during the Great Recession Kelley skewness dropped sharply for workers but only mildly for entrepreneurs. Panel E documents that entrepreneurs' income changes are much less leptokurtic than workers' as excess Crow-Siddiqui kurtosis of 1-year income growth is around 3 throughout the sample period.[44] Strikingly and in contrast to workers, we find that gender differences play almost no role for the dispersion, skewness and excess kurtosis of entrepreneurs' income growth. Note that this is true despite the fact that we observe substantial gender gaps in non-labor income (Figure 11).

In Figure 16, we show how our percentile-based measures of dispersion, skewness, and excess kurtosis of total income growth depend on permanent income. As in the core analysis of earnings changes, permanent income is defined as the residual of the log of average past income where we take out gender and year specific age effects. Again, we compare entrepreneurs to workers. The

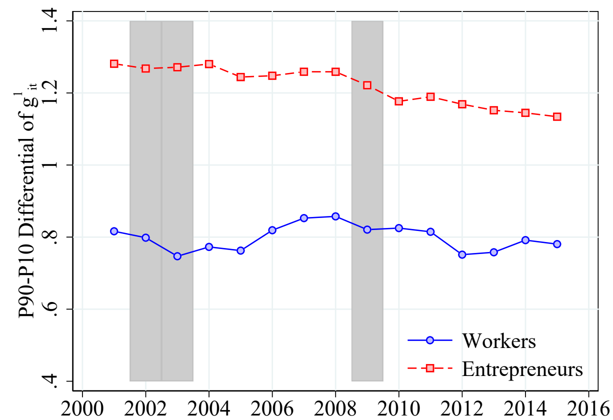
[44] Figure G.8 shows that this difference is mostly driven by the shoulders instead of the tails of the distribution. This means that entrepreneurs experience intermediate income changes more frequently than workers.

Figure 15: Dispersion, Skewness and Kurtosis of 1-Year Log Income Changes

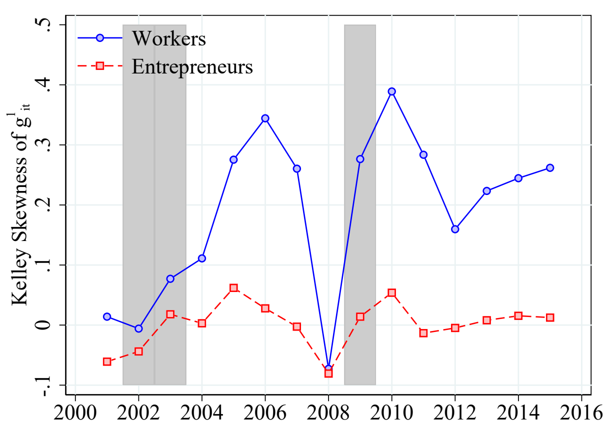
(a) P90–P10: Men



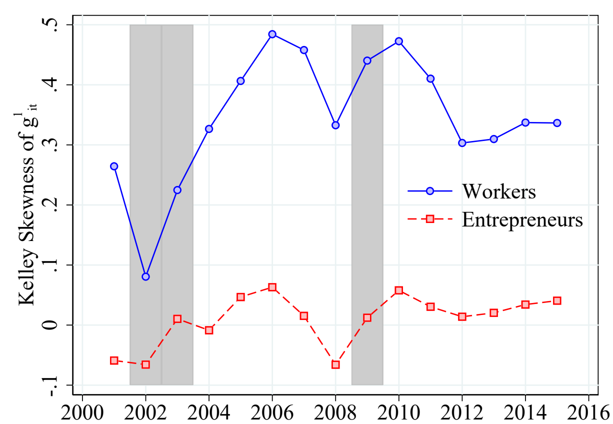
(b) P90–P10: Women



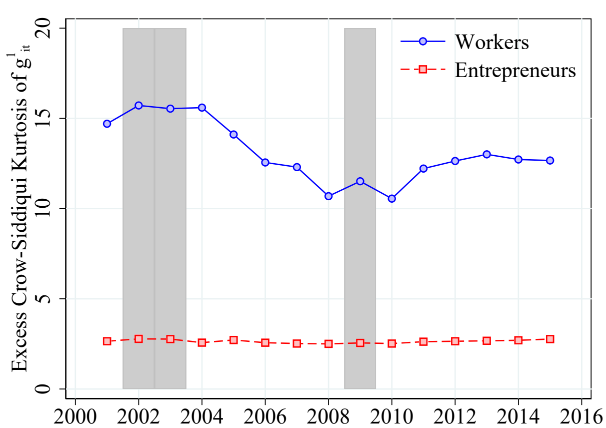
(c) Kelley Skewness: Men



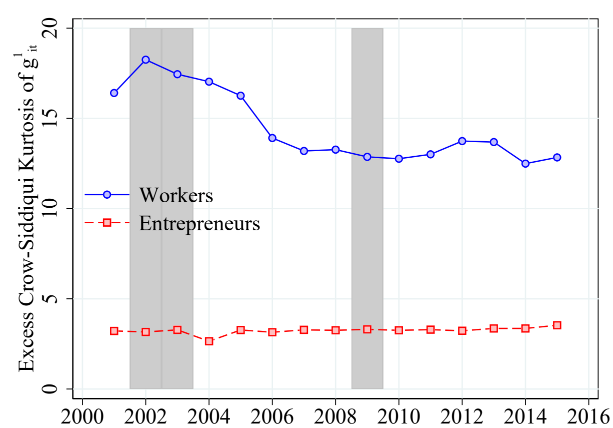
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



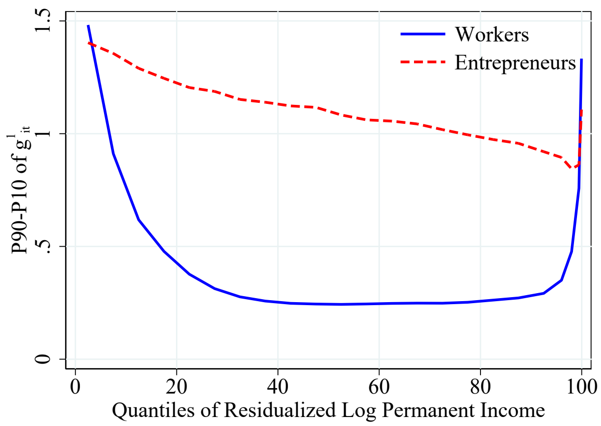
(f) Excess Crow-Siddiqui Kurtosis: Women



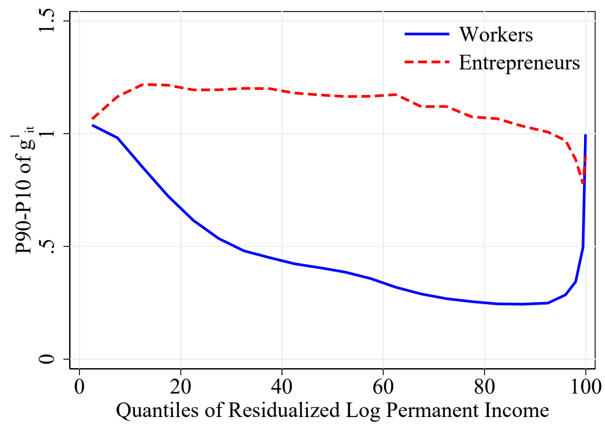
Notes: This figure shows the evolution of P90-P10 differentials, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real annual total income (from t to $t+1$) in the combined IAB-TPP data (LS sample) by main income source (labor, non-labor) and gender. A person's main income source is labor whenever more than 50% of her income comes from (dependent) employment instead of self-employment, corporate business income or rental income. See Footnote 23 for definitions and interpretation of Kelley skewness and excess Crow-Siddiqui kurtosis. Shaded areas indicate recessions.

Figure 16: Heterogeneity in Dispersion, Skewness and Kurtosis of 1-Year Log Income Growth by Main Income Source

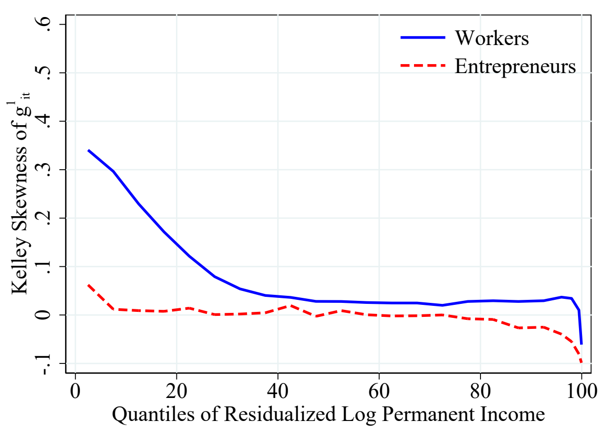
(a) P90–P10: Men



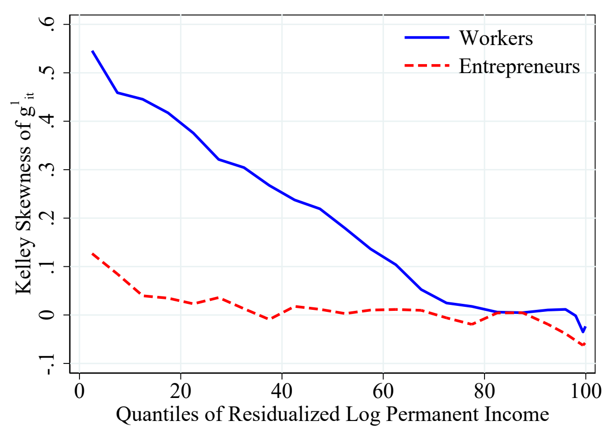
(b) P90–P10: Women



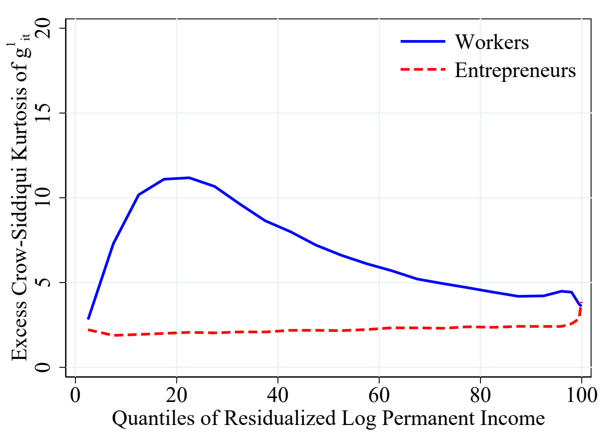
(c) Kelley Skewness: Men



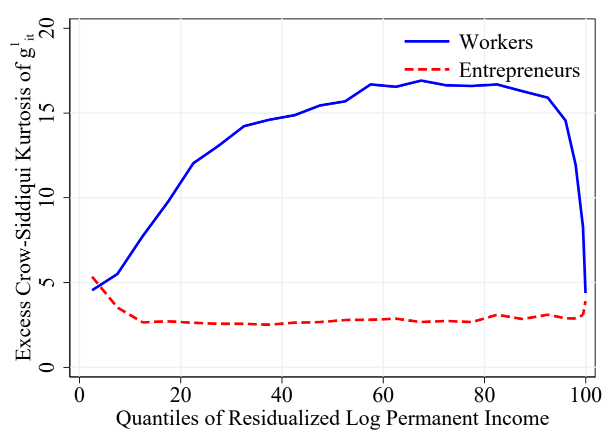
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



(f) Excess Crow-Siddiqui Kurtosis: Women



Notes: This figure shows the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real total income by quantiles of the distribution of permanent total income (from t to $t+1$) in the combined IAB-TPP data (H Sample) as averages from 2004 to 2011 by main income source (labor, non-labor) and gender. The (gender-specific) ranking of permanent income is based on the distribution of total income of both workers and entrepreneurs. A person's main income source is labor (worker) whenever more than 50% of her income comes from (dependent) employment instead of self-employment, corporate business income or rental income. See Footnote 23 for definitions and interpretation of Kelley skewness and excess Crow-Siddiqui kurtosis. Shaded areas indicate recessions.

P90-P10 differential of male entrepreneurs' income growth decreases steadily in permanent income but spikes up for the top 1%. While heterogeneity for workers is more pronounced, the difference in the P90-P10 between entrepreneurs in the bottom 10% and the top 10% (without the top 1%) still differs by about 50 log points. For female entrepreneurs, this relationship is hump-shaped. However, this difference compared to male entrepreneurs is largely driven by differences in the distribution of permanent income (see Figure G.11). In stark contrast to workers, there is almost no dependence on permanent income for Kelley skewness and excess Crow-Siddiqui kurtosis of entrepreneurs' income growth. Overall, the fact that excess Crow-Siddiqui kurtosis is much closer to zero and that Kelley skewness is close to zero implies that entrepreneurs' incomes change are more similar to a Gaussian distribution than those of workers. In addition, the moments of entrepreneurs' income growth exhibit much less state-dependence with respect to (permanent) income than those of workers.

4.3 Top Income Mobility

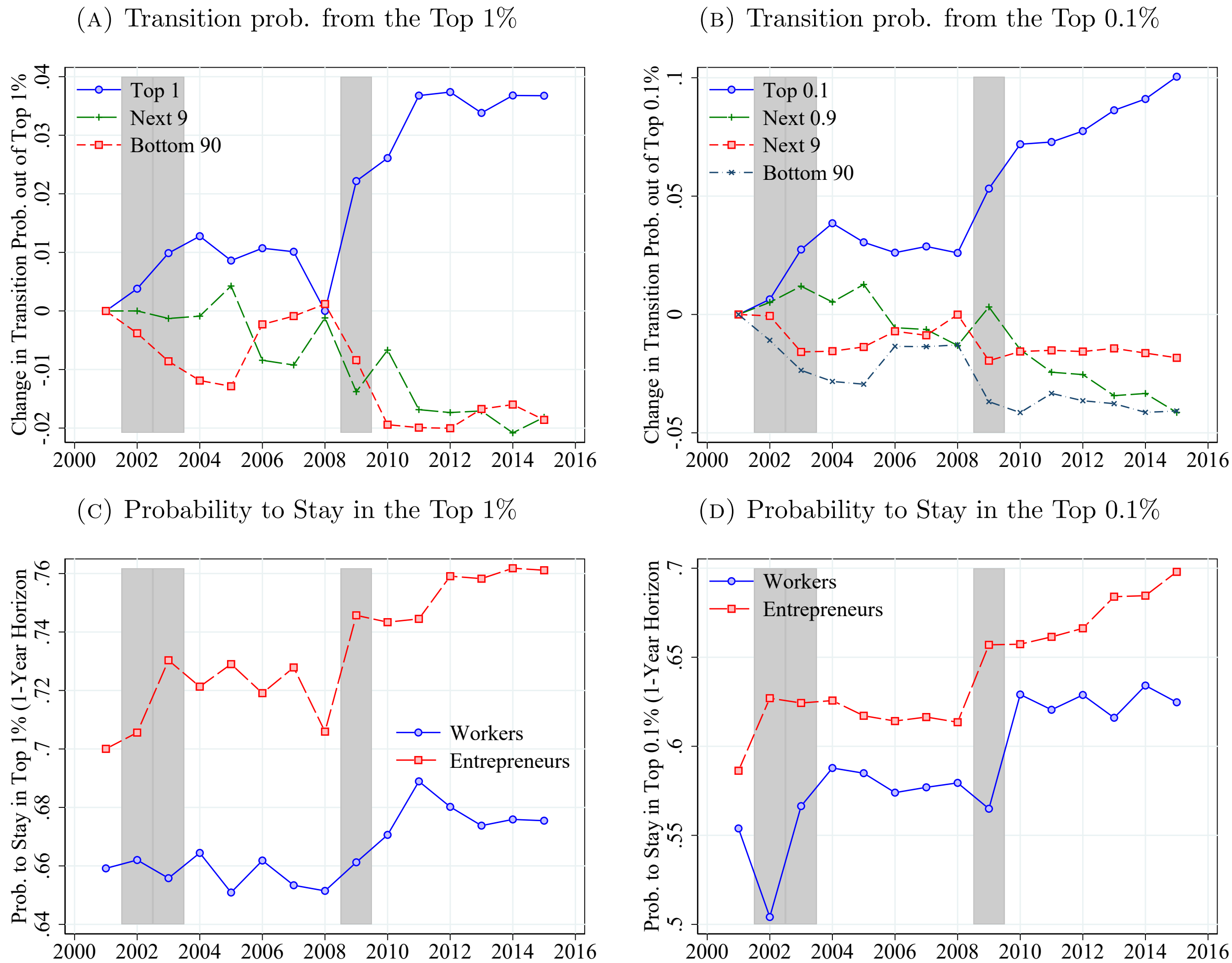
In this final part of our analysis, we ask how the probability of someone someone dropping out the top 1% or 0.1% of the (age-specific) income distribution has evolved over time. Panel A of Figure 17 shows that the probability of staying in the top 1% of the income distribution has increased by 4 percentage points between 2001 and 2016.[45] Both the probability of falling into the next 9%, i.e. between the 90th and 99th percentile of the income distribution, as well as the probability of falling out of the top 10% have decreased. This decrease in top income mobility seems to have occurred mainly around the two recessions in our sample.

This becomes even more evident in Panel B which shows the evolution of transition probabilities of income earners in the top 0.1% whose probability to stay at the very top has increased by 10 percentage points. After the Great Recession, the continued increase in the probability to stay is mirrored by a decline in the probability to fall into the next 9% instead of the bottom 90%.

Panels C and D show the probabilities to stay in the top 1% and top 0.1% separately for workers and entrepreneurs. Workers are more likely to fall out of the top than entrepreneurs. In addition, this gap has widened after 2010 when entrepreneurs' likelihood to stay at the top kept increasing while workers' did not. While we focus on 1-year transition rates here, we show in Figure G.13 that the same patterns (both qualitatively and quantitatively) hold also for 5-year transition probabilities. Both magnitudes and trends reported here are very similar to the findings for US workers studied in Guvenen et al. (2021a). If anything, German top income earners are slightly more likely to stay at the top.

[45] In 2001, the probability of staying in the top 1% is equal to 0.676 and the corresponding probabilities to fall into the next 9% and bottom 90% are 0.257 and 0.067 respectively.

Figure 17: Top Income Mobility – 1-Year Transition Probabilities



Notes: This figure shows statistics on top income mobility in the combined IAB-TPP data (LS sample but including negative incomes). Panels A and B show the evolution of 1-year transition probabilities from the top 1% and top 0.1% of the income distribution into selected parts of the income distribution from one year to the next. The rankings are computed conditional on age. The "Top 1" ("Top 0.1%")is the probability of staying in the Top 1% (0.1%). The "Next 9" is the part of the distribution between the P90 and P99 and the "Next 0.9" is the part between the P99 and the P99.9. The lines sum to zero. Panels C and D show the 1-year probability of staying in the top 1% or top 0.1% for workers and entrepreneurs. The ranking is based on the total income distribution and not conditional on the main income source. Shaded areas indicate recessions. Figure G.13 shows the same statistics for 5-year transition probabilities.

5 Conclusion

This paper provides a comprehensive analysis of inequality and income dynamics for Germany over the last two decades. By combining two high quality administrative data sources – personal income tax and social security records – this is the first paper to offer a complete picture of the German income distribution ranging from the very bottom to the very top.

The first part of the analysis focuses on labor earnings, which is the main source of income for the vast majority of individuals and is most easily compared across datasets and countries. We find that earnings inequality among men has been increasing over the entire sample period from 2001 to 2016 and in particular before the Great Recession – a period where only the top 25% experienced real earnings growth and the bottom half realized real earnings losses of 5 to 20 % and more. After the Great Recession, earnings below the median stabilized while those above the median continued to grow. For women, the evolution of earnings inequality is a tale of two halves. While bottom inequality has been falling due to strong earnings growth at the bottom, top inequality has been rising.

A striking finding is that female top earners (above the 90th percentile) have seen the strongest growth in real earnings. In fact, women's earnings have been catching up with male earnings throughout most of the distribution. This happened even though the share of women working full-time has been declining. While we provide some evidence on how changes in observable job and worker characteristics have contributed to this, a fruitful avenue for future research could be to disentangle the economic forces behind this closing of the gender earnings gap.

Our analysis of individual earnings changes reveals that the earnings growth distribution has been significantly skewed to the right for both men and women since 2005 (apart from the Great Recession). This is in contrast to various other countries (e.g. Italy, Spain, Sweden, UK, US). The fact that large positive gains are more likely than large negative shocks reflects the low risk of job-loss and increasing labor force participation.

In the second part of the paper, we investigate how taking the incomes of self-employed, business owners and landlords into account enriches the overall picture on total income inequality. While workers' incomes are more stable over the business cycle, non-labor incomes have increased substantially relative to labor incomes. Between 2001 and 2016, average incomes of workers grew by around 5% while average incomes of entrepreneurs increased by around 25%. Hence, total income inequality is higher and increased more strongly than labor income inequality.

Our analysis also shows that there exist large gender differences in non-labor incomes. While we find some convergence over time, we document large gaps between men and women at the very top of the total income distribution driven by women being less likely to have high business incomes.

Finally, we contrast income dynamics of workers with those of entrepreneurs and find that the latter face significantly more volatility. From a modeling point of view, non-labor income risk is much better approximated by a normal distribution and exhibits less dependence on age or permanent income.

References

Alon, T., S. Coskun, M. Doepke, D. Koll, and M. Tertilt (2021): "From Mancession to Shecession: Women's Employment in Regular and Pandemic Recessions," NBER Working Papers 28632.

Antonczyk, D., T. DeLeire, and B. Fitzenberger (2018): "Polarization and Rising Wage Inequality: Comparing the U.S. and Germany," *Econometrics*, 6.

Arellano, M., S. Bonhomme, M. D. Vera, L. Hospido, and S. Wei (2021): "Income Risk Inequality: Evidence from Spanish Administrative Record," GID Working Paper.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The Fall of the Labor Share and the Rise of Superstar Firms*," *The Quarterly Journal of Economics*, 135, 645–709.

Bach, S., G. Corneo, and V. Steiner (2009): "From bottom to top: The entire distribution of market income in Germany, 1992–2003," *Review of Income and Wealth*, 55, 303–330.

Bartels, C. (2019): "Top incomes in Germany, 1871–2014," *The Journal of Economic History*, 79, 669–707.

Bayer, C. and F. Juessen (2012): "The life-cycle and the business-cycle of wage risk — Cross-country comparisons," *Economics Letters*, 117, 831–833.

Bell, B., N. Bloom, and J. Blundell (2021): "Income Dynamics in the United Kingdom 1975-2020," GID Working Paper.

Biewen, M. (2000): "Income Inequality in Germany during the 1980s and 1990s," *Review of Income and Wealth*, 46, 1–19.

Biewen, M., B. Fitzenberger, and J. de Lazzer (2018): "The role of employment interruptions and part-time work for the rise in wage inequality," *IZA Journal of Labor Economics*, 7.

Biewen, M., M. Ungerer, and M. Löffler (2019): "Why Did Income Inequality in Germany Not Increase Further After 2005?" *German Economic Review*, 20, 471–504.

Bossler, M. and T. Schank (2020): "Wage inequality in Germany after the minimum wage introduction," LASER Discussion Papers 117.

Bradley, J. and A. Kügler (2019): "Labor market reforms: An evaluation of the Hartz policies in Germany," *European Economic Review*, 113, 108–135.

Brenke, K., U. Rinne, and K. F. Zimmermann (2013): "Short-time work: The German answer to the Great Recession," *International Labour Review*, 152, 287–305.

Brüll, E. and C. Gathmann (2020): "Evolution of the East German Wage Structure," ZEW Discussion Paper 20-081.

Bundesinstitut für Bevölkerungsforschung (2021): "Nach Altersgruppen zusammengefasste Geburtenziffern deutscher Frauen in Deutschland (1991-2019)," https://www.bib.bund.de/Permalink.html?id=10244004, accessed: 2022-01-04.

Busch, C., D. Domeij, F. Guvenen, R. Madera, et al. (forthcoming): "Skewed Idiosyncratic Income Risk over the Business Cycle: Sources and Insurance," *American Economic Journal: Macroeconomics*.

Bönke, T., G. Corneo, and H. Lüthen (2015): "Lifetime Earnings Inequality in Germany," *Journal of Labor Economics*, 33, 171–208.

Bönke, T., U. Glogowosky, E. Hansen, H. Lüthen, and D. Sachs (2022): "Family Policies and Child-Related Earnings Gaps in Germany," Mimeo.

Caliendo, M. and L. Wittbrodt (2021): "Did the Minimum Wage Reduce the Gender Wage Gap in Germany?" workingpaper 40, Center for Economic Policy Analysis (CEPA).

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128, 967–1015.

Dauth, W. and J. Eppelsheimer (2020): "Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide," *Journal for Labour Market Research*, 54.

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor Market Institution and the Distribution of Wages," *Econometrica*, 1001–1044.

Doepke, M. and M. Tertilt (2016): "Families in Macroeconomics," Elsevier, vol. 2, chap. Chapter 23, 1789–1891.

Doerrenberg, P., A. Peichl, and S. Siegloch (2017): "The elasticity of taxable income in the presence of deduction possibilities," *Journal of Public Economics*, 151, 41–55.

Dolls, M., P. Doerrenberg, A. Peichl, and H. Stichnoth (2018): "Do retirement savings increase in response to information about retirement and expected pensions?" *Journal of Public Economics*, 158, 168–179.

Dustmann, C., B. Fitzenberger, U. Schönberg, and A. Spitz-Oener (2014): "From Sick Man of Europe to Economic Superstar: Germany's Resurgent Economy," *Journal of Economic Perspectives*, 28, 167–88.

Dustmann, C., A. Lindner, U. Schönberg, M. Umkehrer, and P. vom Berge (2022): "Reallocation Effects of the Minimum Wage," *The Quarterly Journal of Economics*, 137, 267–328.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009): "Revisiting the German wage structure," *The Quarterly Journal of Economics*, 124, 843–881.

Ellguth, P. and S. Kohaut (2019): "Tarifbindung und betriebliche Interessenvertretung. Aktuelle Ergebnisse aus dem IAB-Betriebspanel 2018," *WSI-Mitteilungen*, 72, 290–297.

——— (2020): "Tarifbindung und betriebliche Interessenvertretung: Aktuelle Ergebnisse aus dem IAB-Betriebspanel 2019," *WSI-Mitteilungen*, 73.

Fitzenberger, B. and A. Seidlitz (2020): "The 2011 break in the part-time indicator and the evolution of wage inequality in Germany," *Journal for Labour Market Research*, 54.

Friedrich, B., L. Laun, and C. Meghir (2021): "Income Dynamics in Sweden 1985-2016," GID Working Paper.

Frodermann, C., A. Schmucker, S. Seth, and P. vom Berge (2021): "Sample of Integrated Labour Market Biographies (SIAB) 1975-2019," FDZ-Datenreport 01, Nuremberg.

Fuchs-Schündeln, N., D. Krueger, and M. Sommer (2010): "Inequality trends for Germany in the last two decades: A tale of two countries," *Review of Economic Dynamics*, 13, 103–132, special issue: Cross-Sectional Facts for Macroeconomists.

Fuest, C., A. Peichl, and S. Siegloch (2018): "Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany," *American Economic Review*, 108, 393–418.

Gehrke, B., W. Lechthaler, and C. Merkl (2019): "The German Labor Market during the Great Recession: Shocks and Institutions," IZA Discussion Papers 11858.

Goldschmidt, D. and J. F. Schmieder (2017): "The rise of domestic outsourcing and the evolution of the German wage structure," *The Quarterly Journal of Economics*, 132, 1165–1217.

Grabka, M. M. (2021): "Income Inequality in Germany Stagnating over the Long Term, But Decreasing Slightly during the Coronavirus Pandemic," DIW Weekly Report 17/18, IW.

Gudgeon, M. and S. Trenkle (2020): "The speed of earnings responses to taxation and the role of firm labor demand," IZA Discussion Paper 13931.

Guvenen, F., G. Kaplan, and J. Song (2021a): "The Glass Ceiling and the Paper Floor: Changing Gender Composition of Top Earners since the 1980s," *NBER Macroeconomics Annual*, 35, 309–373.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021b): "What do data on millions of US workers reveal about life-cycle earnings risk?" *Econometrica*, 89, 2303–2339.

Guvenen, F., S. Ozkan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122, 621–660.

Halvorsen, E., S. Ozkan, and S. Salgado (2021): "Earnings Dynamics and Its Intergenerational Transmission: Evidence from Norway," GID Working Paper.

Hartung, B., P. Jung, and M. Kuhn (2018): "What Hides Behind the German Labor Market Miracle? Unemployment Insurance Reforms and Labor Market Dynamics," CESifo Working Paper Series 7379, CESifo.

Hochmuth, B., B. Kohlbrecher, C. Merkl, and H. Gartner (2021): "Hartz IV and the decline of German unemployment: A macroeconomic evaluation," *Journal of Economic Dynamics and Control*, 127, 104114.

Hoffmann, E. B., D. Malacrino, and L. Pistaferri (2021): "Labor Reforms and Earnings Dynamics: The Italian Case," GID Working Paper.

Hoffmann, F. and T. Lemieux (2016): "Unemployment in the Great Recession: A Comparison of Germany, Canada, and the United States," *Journal of Labor Economics*, 34, S95–S139.

Hohendanner, C. and J. Stegmaier (2012): "Geringfügig Beschäftigte in deutschen Betrieben: Umstrittene Minijobs," *IAB-Kurzbericht*.

Hufe, P., A. Peichl, and M. Stöckli (2018): "Ökonomische Ungleichheit in Deutschland – ein Überblick," *Perspektiven der Wirtschaftspolitik*, 19, 185–199.

Institut Arbeit und Qualifikation (2021): "Durchschnittsalter der Frauen bei der Geburt des ersten Kindes, 1965 – 2020," https://www.sozialpolitik-aktuell.de/files/sozialpolitik-aktuell/_Politikfelder/Bevoelkerung/Datensammlung/PDF-Dateien/abbVII4.pdf, accessed: 2022-01-04.

Jenderny, K. (2016): "Mobility of Top Incomes in Germany," *Review of Income and Wealth*, 62, 245–265.

Jenderny, K. and C. Bartels (2015): "The Role of Capital Income for Top Income Shares in Germany," World Top Incomes Database (WTID) Working Paper.

Kirsch, A. and K. Wrohlich (2020): "Frauenanteile in Spitzengremien großer Unternehmen steigen - abgesehen von Aufsichtsräten im Finanzsektor," *DIW Wochenbericht*, 87, 38–49.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019): "Child Penalties across Countries: Evidence and Explanations," *AEA Papers and Proceedings*, 109, 122–26.

Kohaut, S. and I. Möller (2017): "Führungspositionen in Betrieben und Verwaltungen: Oberste Chefetage bleibt Männerdomäne," IAB-Kurzbericht 24/2017, Nürnberg.

Kramarz, F., E. Nimier-David, and T. Delemotte (2021): "Inequality and Earnings Dynamics in France: National Policies and Local Consequences," GID Working Paper.

Krebs, T. and M. Scheffel (2013): "Macroeconomic Evaluation of Labor Market Reform in Germany," *IMF Economic Review*, 61, 664–701.

——— (2017): "Labor Market Institutions and the Cost of Recessions," IMF Working Papers 2017/087, International Monetary Fund.

Kriete-Dodds, S. and D. Vorgrimler (2007): "The German Taxpayer-Panel," *Schmollers Jahrbuch*, 127, 497–509.

Krolage, C., A. Peichl, and D. Waldenström (forthcoming): "Long Run Trends in Top Income Shares: the Role of Income and Population Growth," *Journal of Economic Inequality*.

Launov, A. and K. Wälde (2013): "Estimating Incentive and Welfare Effects of Nonstationary Unemployment Benefits," *International Economic Review*, 54, 1159–1198.

Leth-Petersen, S. and J. Sæverud (2021): "Trends in Income Risk in Denmark 1987-2016," GID Working Paper.

Lochner, B., S. Seth, and S. Wolter (2020): "Decomposing the large firm wage premium in Germany," *Economics Letters*, 194, 109368.

McKinney, K. L., J. M. Abowd, and H. P. Janicki (2021): "U.S. Long-Term Earnings Outcomes by Sex, Race, Ethnicity, and Place of Birth," GID Working Paper.

Peichl, A., N. Pestel, and H. Schneider (2012): "Does Size Matter? The Impact of Changes in Household Structure on Income Distribution in Germany," *Review of Income and Wealth*, 58, 118–141.

Piketty, T. and E. Saez (2003): "Income Inequality in the United States, 1913–1998," *The Quarterly Journal of Economics*, 118, 1–41.

Prinz, E. and J. Schwalbach (2020): "Vergütung der DAX-Vorstände in den 30-Jahren seit 1987 verzehnfacht," *Forschungsbericht, Humboldt-Univerität zu Berlin.*

Riebesell, P. (1922): *Steuer-Mathematik. Die Fehler in den Reichssteuertarifen.*, Hamburg: Henri Grand.

Schrenker, A. and A. Zucco (2020): "Gender Pay Gap steigt ab dem Alter von 30 Jahren stark an," *DIW Wochenbericht*, 87, 137–145.

Smith, M., D. Yagan, O. Zidar, and E. Zwick (2019): "Capitalists in the Twenty-First Century*," *The Quarterly Journal of Economics*, 134, 1675–1745.

Statistisches Bundesamt (2021a): "Abhängig Erwerbstätige: Deutschland, Jahre, Beschäftigungsumfang, Geschlecht," Tabellen.

——— (2021b): "Verdienste und Arbeitskosten," Index der Tarifverdienste und Arbeitszeiten – Lange Reihe, Destatis.

Stockhausen, M. and K. Maiworm (2021): "IW-Verteilungsreport 2021. Wohlstand für alle? Inklusives Einkommenswachstum vor Corona," IW-Report 36, IW.

Thomsen, U., J. Ludsteck, and A. Schmucker (2018): "Skilled or unskilled – Improving the information on qualification for employee data in the IAB Employee Biography," FDZ-Methodenreport 09, Nuremberg.

Wanger, S. (2020): "Entwicklung von Erwerbstätigkeit, Arbeitszeit und Arbeitsvolumen nach Geschlecht," *IAB-Forschungsbericht*, 16.

Online Appendix for

Inequality and Income Dynamics in Germany

by

Moritz Drechsel-Grau. Andreas Peichl, Johannes Schmieder, Kai D. Schmid, Hannes Walz and Stefanie Wolter

Appendices

Contents

A Institutional and Macroeconomic Background: Details

In this section we extend the discussion of subsection 2.1 on relevant institutions and the macroeconomic situation in Germany for the period 1993 to 2018. We complement further elaborations by figures of institutional and trend patterns over this period.

A.1 Institutions

The Personal Income Tax. Germany applies a comprehensive income tax on income from all sources. The German tax law distinguishes seven different types of income: (i) income from agriculture and forestry, (ii) (non-corporate) business income (this includes dividends and capital gains from closely held corporations, i.e. with an ownership share of at least 1%), (iii) entrepreneurial income, (iv) salaries and wages from employment, (v) investment income (i.e. interest payments and dividends from "normal" stock holdings), (vi) rental income, and (vii) "miscellaneous income" (including, for example, taxable (private) pensions, annuities and capital gains).[46] For each type of income, all expenses that are necessary to obtain, maintain or preserve the income from a given source are deductible. The same holds for education costs, child care costs and donations to charity.

In contrast to most other countries, which use a bracket system with constant marginal tax rates within a bracket, Germany uses a formula (which is quadratic in income) to compute the tax liability. As a consequence, marginal tax rates increase linearly in taxable income from 14% up to 42% (for taxable income above 52,151 Euro in 2008). At the very top, an additional tax bracket with a marginal tax rate of 45% was introduced in 2007 for taxable income above 250,000 Euro.[47]

Between 2000 and 2005, a major reform of the German personal income tax took place. The basic tax allowance was increased in several steps from 6,902 Euro in 2000 to 7,664 Euro (2004-2008). The lowest marginal tax rate decreased from 22.9% in 2000 to 15% (2005-2008) and 14% (since 2009) – see Figure A.1 (A). The top marginal tax rate was reduced from 51% in 2000 to 42% in 2005. The threshold for application of the top marginal tax rate was reduced from 58,643 Euro in 2000 to 52,151 Euro in 2004. In 2007, an additional tax bracket (for taxable income above 250,000

[46] The following types of income are tax exempt: payments from health insurance, accident insurance and insurance for disability and old age, welfare benefits and scholarships.

[47] The reasoning behind using such a formula instead of tax brackets was "to avoid bunching at kink points" (see, e.g., Riebesell, 1922, Chapter 5). The formula for the year 2008 (the last year of a major change) is defined as follows:

$$T = \begin{cases} 0 & \text{if } TI \leq 7,664 \\ (883.74\frac{TI-7,664}{10,000} + 1,500)\frac{TI-7,664}{10,000} & \text{if } 7,664 < TI \leq 12,739 \\ (228.74\frac{TI-12,739}{10,000} + 2,397)\frac{TI-12,739}{10,000} + 989 & \text{if } 12,739 < TI \leq 52,151 \\ 0.42TI - 7,914 & \text{if } 52,151 < TI \leq 250,000 \\ 0.45TI - 15,414 & \text{if } TI > 250,000. \end{cases}$$

For married taxpayers filing jointly, the tax is twice the amount of applying the formula to half of the married couple's joint taxable income: $T_m(TI_1, TI_2) = 2 * T\left(\frac{TI_1+TI_2}{2}\right)$. In addition to the personal income tax, households pay the "Solidaritätszuschlag", a tax supplement originally introduced to finance the German reunification. During the period of interest, 2000-2018, the supplement amounts to 5.5% of the income tax liability. See Doerrenberg et al. (2017) for an overview of the German personal income tax and its deduction possibilities.

Euro) was introduced with a top marginal tax rate of 45%. All nominal start and end points have been adjusted multiple times since 2008 to correct for inflation.

Marginal Employment ("Mini-Jobs"). Marginal employment contracts, called mini-jobs, are jobs with earnings below a time-varying threshold as pictured in Panel C of Figure A.1. The maximum income for marginal employment currently amounts to 450 Euro per month. Jobs below this threshold are exempted from social security contributions and income tax.[48] The so-called mini-jobs were part of the Agenda 2010 labor market reforms (also called Hartz reforms) to lower entry barriers to the labor market. Over our sample period in each year around 4.5-5 million workers hold only a mini-job, while another 2.7 million workers use marginal employment as a form of secondary jobs. Mini-jobs are common among benefit recipients, students and pensioners to increase their monthly income. As a result of the tax incentives for married couples, that rewards unequal labor incomes in marriages, there are also many married women who take up mini-jobs. While, in principle, marginal employment is not limited to certain industries, the share of marginal employees is highest in hospitality, services, retail and agriculture (Hohendanner and Stegmaier, 2012). There were two reforms during our sample period (see Gudgeon and Trenkle 2020 for details). In April of 2003, the monthly earnings threshold for mini-jobs was raised from 325 Euro to 400 Euro. It also abolished an weekly working hours limit of 15 hours for mini-jobs, a constraint that was likely not binding. Probably most importantly, the reform also allowed workers to hold a (tax exempt) mini-job as a secondary job at a different employer. A second reform in 2013 raised the earnings threshold from 400 to 450 Euro. Note that apart from these reforms the earnings thresholds have remained constant at the nominal values and thus were gradually falling in real terms.

Minimum Wage. Germany introduced a statutory national minimum wage of 8.50 Euro in 2015. It was gradually increased to 8.84 Euro (January 2017), 9.19 Euro (January 2019) and after several more steps is currently at 9.82 Euro (January 2022). Real term values for our sample period are displayed in Figure A.1 (C). Before 2015 different wage floors existed in 12 industries: construction, roofing, cleaning, and nursing among others. Furthermore, some of the larger industries have binding collective agreements that set minimum wages. The impact of the wage floor on wages varied by region and affected about 15 percent of all employees (Dustmann et al., 2022).

Collective bargaining and union density. An effective instrument in Germany to set wages are tariff agreements between union and employer representatives that often have a binding character for all firms in a certain industry. The worker coverage of industry-level collective bargaining agreements varies between former West and East Germany and decreases over time (see Panel B of Figure A.1). Especially start-ups and smaller firms are less likely to be part of a collective agreement. Less common firm-level collective bargaining agreements cover an additional 2% of firms

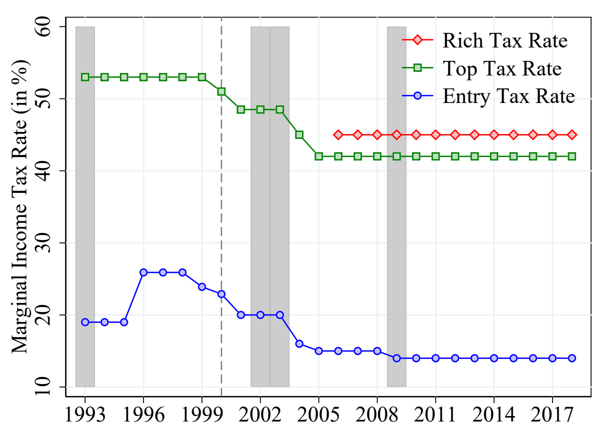
[48] A person can hold multiple mini-jobs but then only the first 450 Euro are tax exempt.

and 8% of employees in 2018 (Ellguth and Kohaut, 2019). The union density (union members out of all employees) declined steadily at the same time.

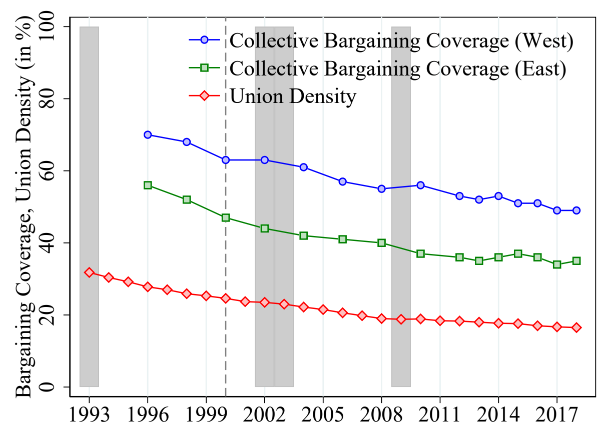
Social Security Contribution Limits. The contributions to the pension system and unemployment insurance are capped. The limit differs between East and West Germany and increases over time, roughly following inflation. Figure A.1 (D) shows the limits for East and West Germany from 1993 to 2018 in real terms (2018 Euro).

Figure A.1: Institutional Background

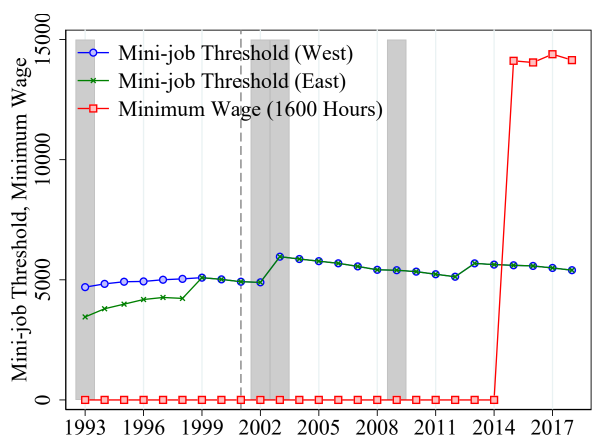
(A) Entry & Top Tax Rates



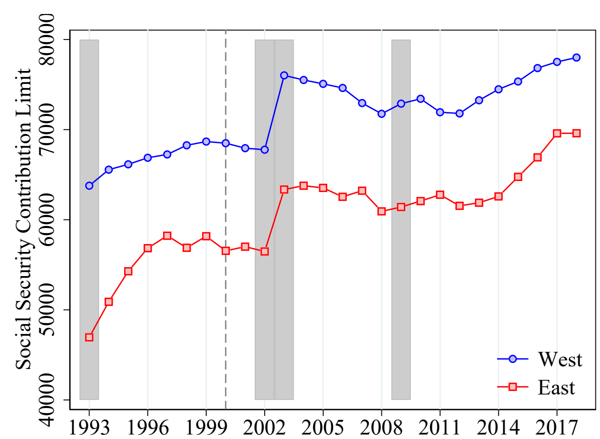
(B) Collective Bargaining Coverage and Union Density



(C) Minimum Wage and Mini-job Threshold



(D) Social Security Contribution Limit



Notes: This figure shows key institutional parameters for our period of analysis including tax rates (Panel A, source: Federal Ministry of Finance), collective bargaining (Panel B, source: Ellguth and Kohaut (2020) and OCED), mini-job thresholds in 2018 Euro (Panel C, source: Deutsche Rentenversicherung) and the social security contribution limit in 2018 Euro (Panel D, source: Deutsche Rentenversicherung), which is relevant for the top coding in the IAB data. Shaded areas indicate recessions.

A.2 Macroeconomic Background

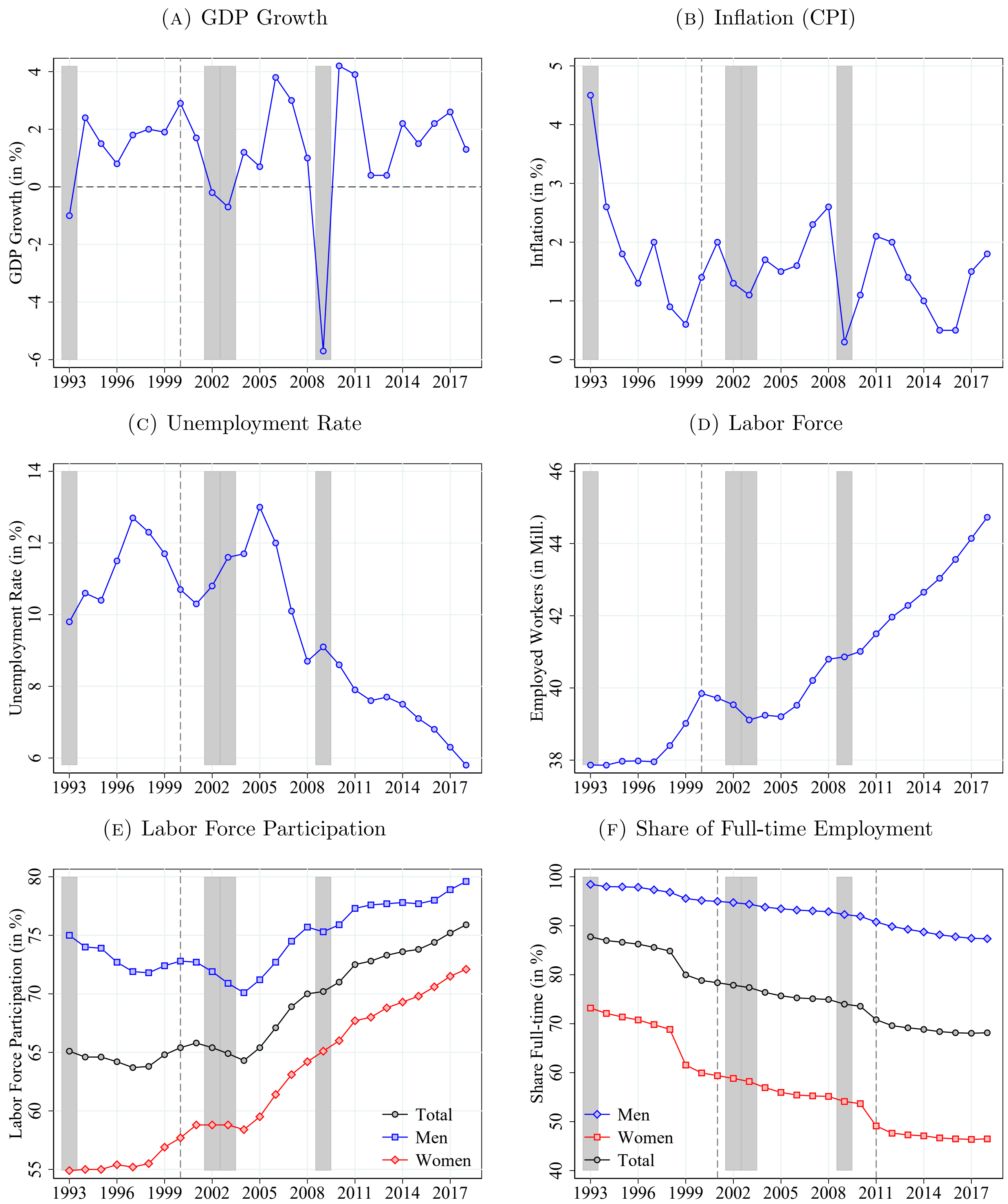
The macroeconomic development in Germany from 1993-2018 can be broadly split into two periods: before and after 2005 (see Figure A.2). The first time span was characterized by low growth and high unemployment (above 10%) and Germany was often referred to as "the sick man of Europe" (Dustmann et al., 2014). This changed in the mid-2000s after a series of labor market and tax reforms were implemented. While the causal effect of these reforms (called "Hartz reforms") on the labor market development and the exact mechanisms are still discussed in the literature, it is undisputed that these reforms "worked" - somehow. How and whether the effects were as desired is sometimes the subject of controversial debate. Critics complain, for example, that the new system is unfair and fosters the growth of the low-wage sector in Germany. Supporters of the existing system counter that the reforms have made it possible to reduce unemployment in Germany since 2005 in the first place, and that abolishing them would jeopardize this success. Critics, in turn, doubt the thesis of the positive labor market effects of the reforms and cite other reasons for the reduction in unemployment. (Macro)economic analyses of the reforms (e.g., Krebs and Scheffel, 2013, 2017; Launov and Wälde, 2013; Hartung et al., 2018; Bradley and Kügler, 2019; Hochmuth et al., 2021) show that the reforms indeed played an important role for the decline in (structural) unemployment, but they are not the only explanatory factor for the positive labor market development.

Nevertheless, neither the Great Recession nor the Euro Crisis affected the German labor market severely. In contrast to the United States and most other EU countries, Germany experienced almost no increase in unemployment, despite a sharp decline in GDP in 2008 and 2009.[49] Moreover, labor force participation rates of both women and men increased steadily after 2004 and the unemployment rate fell below 6% in 2018.

A notable feature over this time period was a large increase in labor force participation of women, from around 55 percent to more than 70 percent as shown in Figure A.2 (E). However, unlike in countries such as the US, this increase was almost exclusively driven by women entering the labor market in part-time and marginal employment, so that the full-time share over this period fell from 75 to around 50 percent for women. For men, labor force participation and the part-time share also increased substantially since 2003, though nowhere near as dramatic as for women.

[49] The system of short-time work buffered the potential increase in unemployment in Germany as at the height of the economic crisis in mid-2009, the number of short-time workers peaked at 1.5 million helping to cushion the labour market impact of the crisis (Brenke et al., 2013).

Figure A.2: Macroeconomic Background



Notes: This figure shows key macroeconomic variables for Germany from 1993-2018 (source: Federal Statistical Office for Panels A - E). The data on share of full-time employment (Panel F) is taken from the IAB data and the reporting procedure for full-time status changed in 2011, leading to a structural break indicated by the dashed line, which is not corrected here. Shaded areas indicate recessions.

B IAB: Social Security Data

The first source of data, which we refer to as the IAB data, is the Integrated Employment Biographies (IEB) supplied by the Institute for Employment Research (*"Institut für Arbeitsmarkt- und Berufsforschung (IAB)"* in German). The IEB are administrative data covering all individuals subject to social security contributions and marginal employment. Moreover, unemployment spells and episodes in active labor market policies are included as well. The IEB allows to follow individuals from labor market entry to retirement. We use 10% random sample of individuals that are either in employed or unemployed, i.e. we exclude persons in active labor market policies.

Employers have to file employment records at least annually or whenever information changes that impacts unemployment benefit or pension calculation. Labor earnings are reported including bonuses and extra pay but only up to the social security contribution limit, which is at an annual labor income of 78,000 Euro in West Germany and 69,600 Euro in East Germany in 2018 (see Figure A.1 for real values over time). All earnings above that limit are censored. We describe below how we impute wages for some of the analyses. Besides to the top-coding, another limitation of the IAB data is that it does not include self-employed individuals (around 4 million) and civil servants (around 1.9 million individuals).

The data contains information on the exact dates of employment and earnings as well as a variety of worker and firm characteristics such as gender, education, year of birth, occupation or industry code. The information is spell based, i.e. accurate to the date and especially with respect to earnings trustworthy. Note, however, that the education information contains some missing values which we impute (described below) using the procedure suggested by the IAB. Moreover, throughout 2011, the reporting procedure for full-time and part-time employment in the social security data changed. This results in a small fraction of workers being falsely classified as working full-time before 2012. We are able to partially correct the full-time indicator in the years prior to 2012 using a cell-wise reclassification approach (see below).

B.1 Top-Coding and Imputation of Wages

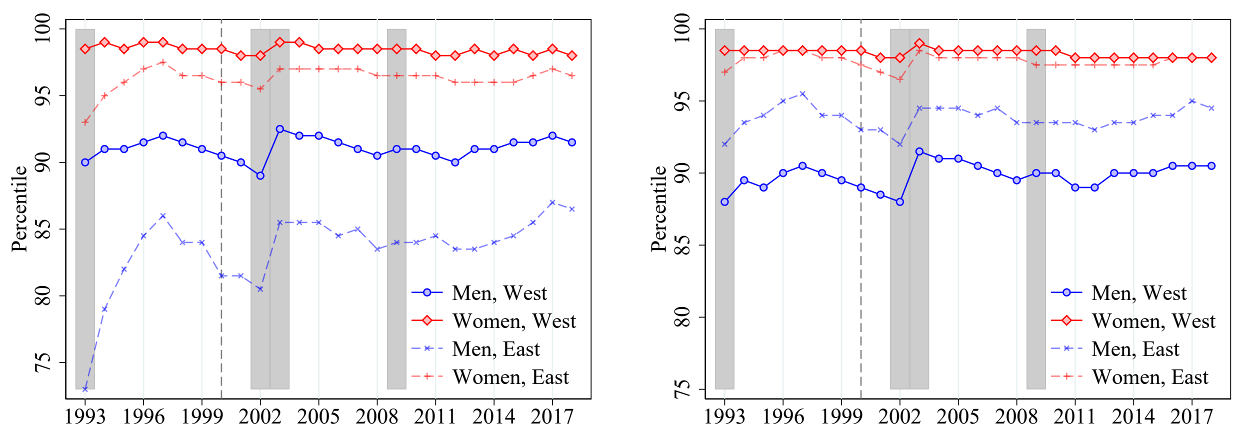
Figure B.1 (A) shows that in the overall labor income distribution for women, the West German social security contribution limit is binding for women at roughly the 99th percentile, while the East German limit is binding at around the 96th percentile. For the earnings distribution of men the limits are much more binding, with the West German threshold binding at roughly the 90th percentile and the East German threshold at roughly the 85th percentile when applied to the whole distribution.

In Figure B.1 (B) we ask the related but different question, where within the earnings distribution of East and West German workers the social security contribution limits fall in a given year by gender. Since East German incomes are still much lower than in West Germany, this pushes up at which percentile in the East/West-specific distribution the social security limit actually becomes binding. The figure highlights clearly that only the highest earning women in East and West are affected by their respective thresholds and, thus, have censored earnings reported. West German men are the most likely to be affected (threshold around the 90th percentile), while the East German limit now lies close to the 95th percentile of the East German earnings distribution. This means that for West German men the highest 10% of earnings are subject to censoring while for East German men this is only the case for the highest 5-7%. Another interpretation of Figure B.1 (A) is hence that it shows where the same 5-7 percent of male workers top coded in the East according to Figure B.1 (B) rank in the overall male income distribution of Germany. The IAB data does not contain information on incomes above the social security contribution limit. Several imputation algorithms have been proposed for wages above the top-coding limit. We use the algorithm suggested by Card et al. (2013) and implemented by Dauth and Eppelsheimer (2020) for SIAB to impute daily wages which we then aggregate to annual incomes for our analysis.

Figure B.1: Share of Uncensored Observations in the IAB data

(A) Share by Year and Gender below top coding for East/West

(B) Share by Year, Gender and East/West below respective top coding



Notes: This figure shows the percentile of the labor earnings distribution at which the top-coding due to the social security contribution limit becomes binding. This corresponds to the share of uncensored observations. In Panel A, the percentiles are calculated by year based on the earnings distributions of men and women separately. In Panel B, four different distributions are calculated for men and women in East and West Germany separately for each year. Shaded areas indicate recessions.

B.2 Imputation of the Education Indicator

The education information in the IEB contains missing values predominantly for workers holding a mini-job. The number of missings increases over time and amounts to 22 percent for regular employees and 60 percent for marginal part-time employees in our data. To cope with the missing information we use the imputed education variable provided by the IAB, which adds missing information by forward and backward writing. The procedure is described in Thomsen et al. (2018).

B.3 Correction of the Full-Time Indicator

In 2011, the reporting procedure for full-time and part-time employment in the social security data changed. This results in a enhanced number of classification updates of workers that have been misclassified as full-time beforehand, but in fact were working part-time, leading to an artificial drop in full-time share and jump in part-time share. The procedure changed throughout the whole year of 2011, which leads to a structural break between 2010 and 2012 with an intermediate update in 2011. Fitzenberger and Seidlitz (2020) document the consequences of this break for analyses of wage inequality and provide an reweighting procedure to correct for misclassifications before 2012.

We use a non-parametric correction approach instead of estimating weights, reclassifying full-time to part-time in 2011 and before if potentially misclassified. This allows us to use the IAB sample consistently without inducing potential bias to other (correct) variables when applying weights to the sample.

First, we restrict our sample to potentially affected individuals in the relevant time period and age group (25 to 55). We apply our correction only to full-time and part-time workers, marginal employment should be unaffected. Following Fitzenberger and Seidlitz (2020), we exclude individuals with wages above a certain threshold. We similarly exclude all observations in the years 2001 to 2011 from the correction when the respective real earnings are above the 80$^{\text{th}}$ percentile of earnings in 2012 for women and 25$^{\text{th}}$ percentile of earnings in 2012 for men. We further calculate a distance measure θ to the percentile threshold, normalized to 0 to 100.

Second, we use gender, an indicator for former West or East Germany, 11 age groups, educational attainment (6 groups) and days in employment (4 groups) to divide our sample into cell-groups. We then cell-wise calculate the share of full-time employment separately for 2009 to 2013. Using this full-time share, we apply a smoothed correction to cell-wise full-time shares for the years 2011 and 2010, based on the full-time share differences as well as (smoothed) pre- and post-trends. For the years 2001 to 2009, we cell-wise deduct the full-time share difference 2010 to 2012 and smoothed trend from the original full-time share. This results in (at least partially) corrected full-time shares for each cell for 2001 - 2011.

Third, we cell-wise reclassify full-time workers to part-time until the share of full-time workers is decreased to the corrected full-time share of the cell. We do not purely pick observations at random for this but sort according to θ, adding a small amount of noise to the latter. This means the probability to be reclassified increases with lower real earnings but not fully depends on those. We do this separately by year for all observations, because workers frequently change cells between years. Thus, we do not carry forward any reclassification from 2010 and 2011 to earlier years. This means workers' classification of full-time or part-time may switch repeatedly because of the correction. This provides us with more reliable (repeated) cross-section aggregates but may result in higher 1-year transition probabilities from full-time to part-time and vice versa in 2001 - 2011.

This procedure resolves most of the structural break for most of the cells in (repeated) cross-section. However, our approach does not necessary fully correct the structural break in attempt to not over-correct. This means there still occur some artifacts in the data around 2011 but to a much smaller degree than without the correction.

C TPP: Tax Data

C.1 General Description

The second source of data is the German Taxpayer Panel (TPP) (Kriete-Dodds and Vorgrimler, 2007), which is an administrative data set based on the universe of personal income tax returns in Germany.[50]

The data set covers all tax units filing tax returns in the period 2001-2016 in Germany. The 2001 to 2016 panel has a total of 58,808,899 unique records for which information is available for at least two waves of years. We work with a 25% random sample of these records. The unit of observation is the taxpayer, i.e., either a single individual or a couple filing jointly. In the latter case, income from all different sources (such as labor or business income) are measured on the individual level before the income is aggregated at the couple's level. The same is true for many deductions and allowances which are available on the individual level.

The data set contains all information necessary to calculate a taxpayer's annual income tax. This includes basic socio-demographic characteristics such as year of birth, gender, family status, number of children as well as detailed information on gross income (differentiated by seven different sources) and basic tax-specific parameters such as work-related expenses and deductions. A list of the variables - differentiated by assessment year - is included in the dataset description available for download.

The data set is not top-coded. Therefore, this data set is especially suited for the analysis of inequality in the upper tail of the income distribution. It is, however, missing the very bottom of the income distribution as incomes below the marginal employment threshold are except from the personal income tax and hence not included in the data.

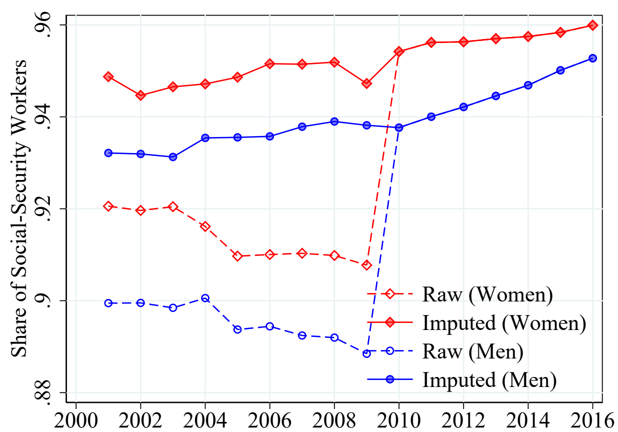
The 2001 to 2011 waves of the Taxpayer Panel (TPP) were compiled on the basis of annual income tax statistics (*Geschäftsstatistik*) of each of the 16 federal states which were then combined into one dataset for Germany. These cross-sectional data contain information from the income tax returns of around 27 million German taxpayers that filed a tax return and were linked to form a panel via the tax numbers and indirect identifiers. Starting with the 2012 assessment year, there was a change in the procedure. Instead of the annual income tax statistics, the federal wage and income tax statistics (*Bundesstatistik zur Lohn- und Einkommensteuer*), which had been collected every three years until then, was collected annually and formed the new basis for the TPP from 2013 onwards. In addition to taxpayers filing a tax return, the federal statistics also include around 12 million non-assessed taxpayers who did not file a tax return and paid the income tax withholding tax (*Lohnsteuer*). We describe how we deal with this structural break below in Appendix D.

[50]See `https://www.forschungsdatenzentrum.de/de/steuern/tpp` for more information (albeit only available in German) on the TPP data. This data has been, for example, used by Doerrenberg et al. (2017) and Dolls et al. (2018) who also provide additional information on the data. More detailed information on the construction and use of the TPP is presented in the usage concept available for download (in German only) here: `https://www.forschungsdatenzentrum.de/de/10-21242-73111-2016-00-01-1-1-0`.

C.2 Imputation of Social Security Indicator in Pre-2010 TPP Data

The definition of non-social-security workers ($C_{it} = 1$) in the TPP is imprecise prior to 2010 resulting in too many non-social-security workers (compared to official IAB data). Figure C.1 shows that the share of social-security workers is too low prior to 2010 (dashed lines). The differences is roughly 4 percentage points for both men an women Panel A of Figure C.2 shows the share of social-security workers by year and earnings bin. Again, this share is lower at almost all income levels for all years before 2010 compared to the years after.

FIGURE C.1: SHARE OF SOCIAL-SECURITY WORKERS IN THE TPP



Notes: This figure shows the average share of social-security liable workers in the raw TPP data over time before (dashed lines) and after the structural break in 2010 for men and women separately. It also shows the corrected share after the application of the imputation procedure described in this Appendix.

Backward-Imputation Procedure. In order to correct the data, we use the following backward imputation procedure:

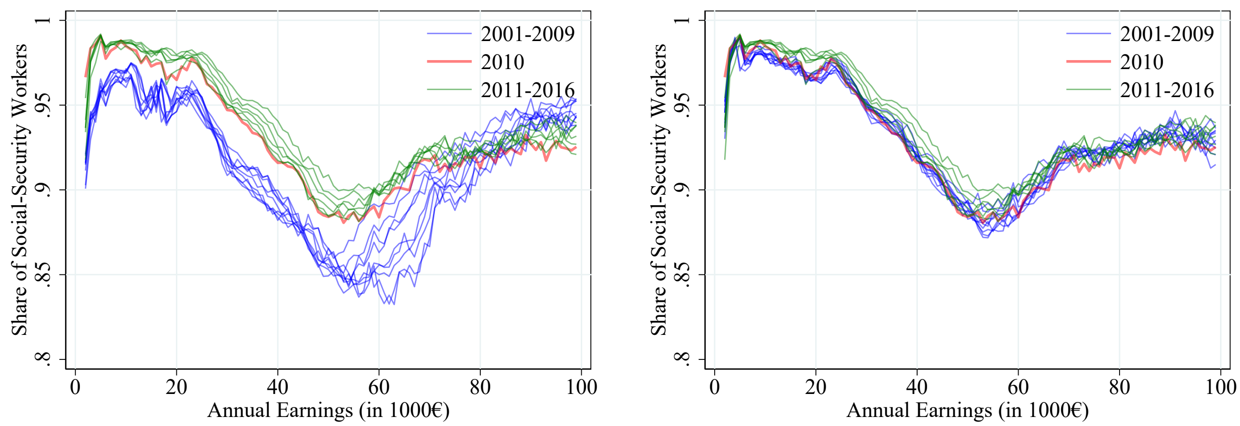
- Let G capture all combinations of the observables gender, age group and binned earnings.
- Let c_{gt} be the share of social-security workers observed in the raw TPP data in group $g \in G$ and year t and c^*_{gt} the corresponding true share.
- The observed share is correct for $t \geq 2010$ ($c = c^*$) and incorrect for $t < 2010$ ($c \neq c^*$).
- We assume that the true share in each gender-age-earnings group is roughly time-invariant.

$$c^*_{gt} = c^*_g + \varepsilon_{gt} \text{ with } \varepsilon_{gt} \sim N(0, \sigma_g) \tag{C.1}$$

- We use the years $t \geq 2010$ to estimate c^*_g and the standard deviation of ε_{gt} (could also be done via regression of C_{it} on a set of group dummies).
- We then approximate the true shares for $t < 2010$ using these estimates. Denote these by $\hat{c}_{gt}$.
- This allows us to predict the share of mis-coded observations, $\eta_{gt} = c_{gt} - \hat{c}_{gt}$ for $t < 2010$.

Figure C.2: Share of Social-Security Workers by Earnings Bins in the TPP

(A) Raw Data (B) Imputed Data



Notes: This figure shows in Panel A the share of social-security liable workers in the raw TPP data over time before (blue lines) and after (green lines) the structural break in 2010 (red line) across the earnings distribution. Panel B shows the same information after the application of the imputation procedure described in this Appendix.

- Using data from 2010 onward, we also estimate the transition probabilities for the social-security indicator conditional on gender, age and earnings bin.

$$\pi^0_{gt} = \Pr(C_t = 0 | C_{t+1} = 1, G_t = g) \tag{C.2}$$

$$\pi^1_{gt} = \Pr(C_t = 1 | C_{t+1} = 0, G_t = g) \tag{C.3}$$

- For the years 2001 to 2009, we re-code the social-security indicator C_{it} as follows:
  (1) Define τ as the first year where the indicator is not (yet) correctly coded (or imputed). Set $\tau = 2009$.
  (2) Set the imputation flag F_i to zero for all workers.
  (3) For workers who are observed in year $\tau + 1$, we impute C_{it} for $t \leq \tau$ using the transition probabilities and their value of $C_{i,\tau+1}$ as a starting point.[51]
  (4) Re-compute $c_{g\tau}$ and update the share of mis-coded observations in year τ and group g, $\eta_{g\tau}$.[52]
  (5) If, as expected, $\eta_{g\tau} \geq 0$, set $x = 0$, otherwise set $x = 1$.
  (6) Randomly choose a fraction η_{gt} of the subset of workers with $G_{it} = 1$ and $C_{i,2009} = x$, and re-code their civil servant indicator $C_{i\tau}$ accordingly.
  (7) If $\tau = 2001$, stop. Otherwise, set τ to $\tau - 1$ and return to step (3).

The results of this imputation procedure are shown in Panel B of Figure C.2. Now the share of social-security workers is similar for all years across the income distribution.

[51] In the initial step with $\tau = 2009$, if a worker is not observed in 2010 but is observed in some later period t' (starting in 2012, the TPP has full coverage), we use $C_{i,t'}$ as a starting point for the imputation for year 2009.

[52] There should still be too few social security workers as some workers who exit the data before 2010 are still mis-coded.

D Combined IAB-TPP Data

This Appendix describes how we combine the IAB and TPP data for our main analysis. As we are not allowed to directly link the micro data of IAB and TPP due to data protection legislation in Germany, we need to rely on non-parametric matching techniques to construct earnings/income distributions as well as distributions of income changes.

Before combining the data, we show descriptive statistics for the IAB and TPP data sets for the year 2008 separately for men and women who are between 25 and 55 years old in Table D.1. Unsurprisingly, the TPP has fewer observations due to missing non-filers and mini-jobs. As the TPP data contains only very limited demographic information, we can only compare both datasets in terms of age. The TPP population is slightly older which again can be attributed to missing observations who are more likely to be at the beginning of their career.

TABLE D.1: DESCRIPTIVE STATISTICS FOR EARNINGS DATA (YEAR 2008)

	Men		Women	
	IAB (1)	TPP (2)	IAB (3)	TPP (4)
Observations (in mill.)	12.430	9.058	11.228	7.409
Mean Earnings (in 2018-Euro)	40,562	46,406	24,010	28,089
A. Age and Nationality				
Share Age 25–34	0.275	0.230	0.261	0.233
Share Age 35–44	0.346	0.352	0.334	0.331
Share Age 45–55	0.378	0.418	0.405	0.436
Non-German	0.090	–	0.066	–
B. Education				
Schooling ($\leq$ 10 years)	0.050	–	0.057	–
Vocational training	0.621	–	0.596	–
Abitur (& voc. training)	0.111	–	0.157	–
College Degree	0.206	–	0.177	–
No Education Data	0.011	–	0.014	–
C. Employment Level				
Full-Time	0.919	–	0.521	–
Part-Time	0.057	–	0.347	–
Mini-Job	0.024	–	0.132	–
Days in Employment	342.3	–	342.2	–

Notes: This table shows descriptive statistics for the IAB and TPP data sets for the year 2008 separately for men and women who are between 25 and 55 years old.

D.1 Reweighting the TPP Data to Match the IAB Data

While we have access to the 'population' version of the available taxpayer data, the TPP still does not cover the *entire* population of income taxpayers. In particular, there are two deviations. First, the TPP only includes tax units that appear in at least two waves [D1]. Second, for the years 2001 to 2011 the TPP only includes information of taxpayers who file a tax return statement [D2]. Hence, around 12 million non-filers are missing per year. Importantly, only workers who do not receive any non-labor income (above an exemption level of roughly 400 Euro) have the option not to file a tax return.[53]

We correct these two deviations by reweighting the TPP data. Thereby, we distinguish between workers whose earnings are subject to social security contributions and who are included in the IAB, and workers whose earnings are not subject to social-security contributions (e.g. civil servants). Note that for our core analysis in Section 3 we only consider the former. The latter are only part of the total income sample in Section 4.

D.1.1 Reweighting the Pre-2012 TPP Data to Account for Missing Non-Filers

For social security workers, we use information from the IAB (headcounts by gender, age group and 1,000 Euro earnings bin) as well as post-2012 TPP data to reweight observations. The reweighted data match the joint distribution of gender, age group and earnings below the social security contribution limit and the number of workers above this limit as well as the share of non-filers by gender, age group and earnings (above the top-coding threshold in the IAB) observed in the post-2012 TPP data.

In particular, we compute from the IAB the number of workers in each (real) annual earnings bin by gender and age group (25-29, 30-34, ..., 50-55). We use bins of 1,000 Euro each up until 60,000 Euro, above which the IAB is top-coded. Hence, we only know the total number of workers above 60,000 Euro. To reweight workers above this cutoff, we additionally compute from the post 2011 TPP data the average share of non-filers in 20 time-invariant earnings vingtiles above the cutoff (by gender and age group). The TPP data further allows us to distinguish between mandatory filers and voluntary filers. Loosely speaking, filing a tax return is mandatory when a worker files jointly with his/her married spouse, received non-labor income (including transfers) above 410 Euro or received other labor income for which the employer did not deduct (enough) income taxes.

In the following, we describe the reweighting procedure in more detail.

Notation:

- G stratification group (combination of gender, age group and earnings bin)
- N^*_{gt} target number of workers in group $g \in G$, computed using IAB data
- N_{gt} observed number of workers in group g in the TPP ($N_{gt} = N^v_{gt} + N^m_{gt}$)
- N^m_{gt} observed number of mandatory filers in group g in the TPP
- N^v_{gt} observed number of voluntary filers in group g in the TPP
- N^n_{gt} observed number of non-filers in group g in the TPP (equals zero before 2012, $N^n < N^v$ after 2012)

[53] The earnings distributions (headcounts by bins) in Figures 1 and D.1 visualize this difference.

w_{gt}^m constructed weight of mandatory filers in group g

w_{gt}^v constructed weight of voluntary filers in group g

Procedure for Workers Below the IAB Top-Coding Cutoff.

(i) Compute the average ratio between target and observed headcounts for the years 2012 to 2016:

$$\delta_g = E_t\left[N_{gt}^*/N_{gt}|t \geq 2012\right] \tag{D.1}$$

(ii) Construct target headcounts net of D2 for the years 2001 to 2011 as $N_{gt}^{1*} = \frac{N_{gt}^*}{\delta_g}$

(iii) Compute the weights for voluntary and mandatory workers as

$$w_{gt}^v = \begin{cases} \frac{N_{gt}^v + (N_{gt} - N_{gt}^{1*})}{N_{gt}^v}\delta_g & \text{if } t < 2012 \\ \frac{N_{gt}^*}{N_{gt}} & \text{if } t \geq 2012 \end{cases} \tag{D.2}$$

$$w_{gt}^m = \begin{cases} \delta_g & \text{if } t < 2012 \\ \frac{N_{gt}^*}{N_{gt}} & \text{if } t \geq 2012 \end{cases} \tag{D.3}$$

Procedure for Workers Above the IAB Top-Coding Cutoff. We partition the top earnings bin (above 60,000) into 20 fractiles by gender and age group. Let H be the combination of gender, age group and this partition. We use the same notation as for below-cutoff workers but replace G and g by H and h. The key assumption is that D2 is constant over time and that the share of non-filers is time-invariant within each combination of gender, age group and earnings fractile.

(i) Compute the average share of non-filers in each group h for the years 2012 to 2016

$$\eta_h = E_t\left[N_{ht}^n/N_{ht}|h, t \geq 2012\right] \tag{D.4}$$

(ii) Compute the number of missing non-filers in $t < 2012$ as

$$\hat{N}_{ht}^n = \frac{N_{ht}}{1 - \eta_h} - N_{ht} \tag{D.5}$$

(iii) To correct for D2, compute the auxiliary weights for voluntary and mandatory workers as

$$\tilde{w}_{ht}^v = \begin{cases} \frac{N_{ht}^v + \hat{N}_{ht}^n}{N_{ht}^v} & \text{if } t < 2012 \\ 1 & \text{if } t \geq 2012 \end{cases} \tag{D.6}$$

$$\tilde{w}_{ht}^m = 1 \qquad \text{for all } t \tag{D.7}$$

(iv) Compute the total headcount implied by the auxiliary weights in the original top earnings bin (by gender and age group):

$$\tilde{N}_t = \sum_h \left(\tilde{w}^v_{ht} N^v_{ht} + \tilde{w}^m_{ht} N^m_{ht} \right) \tag{D.8}$$

(v) To correct for D1, we rescale the auxiliary weights to match the target headcount in the top earnings bin. This gives:

$$w^x_{ht} = \tilde{w}^x_{ht} \frac{N^*_t}{\tilde{N}_t} \qquad \text{for } x \in \{v, m\} \tag{D.9}$$

D.1.2 Reweighting Non-Social-Security Workers in the TPP

For non-social-security workers, we only use post-2012 TPP data for reweighting as these workers are not included in the IAB data. Hence, the reweighted data match the share of non-filers by gender, age group and earnings observed in the post-2012 TPP data. For brevity, we sometimes refer to social-security workers as regular workers and to non-social-security workers as other workers.

The reweighting procedure to account for non-filing non-social-security workers is very similar to the one used for social security workers above the cutoff. The main difference is that we have no data to correct for D2 as we cannot rely on IAB data for non-social-security workers. We first group civil servants based on gender, age group and (time-invariant) earnings fractiles.[54] We use the same notation as above.

(i) Compute the average share of non-filers for the years 2012 to 2016

$$\eta_g = E_t \left[N^n_{gt} / N_{gt} | t \geq 2012 \right] \tag{D.10}$$

(ii) Compute the number of missing non-filers in $t < 2012$ as

$$\hat{N}^n_{gt} = \frac{N_{gt}}{1 - \eta_g} - N_{gt} \tag{D.11}$$

(iii) Correcting for D2, compute the weights for voluntary and mandatory workers as

$$w^v_{gt} = \begin{cases} \frac{N^v_{gt} + \hat{N}^n_{gt}}{N^v_{gt}} & \text{if } t < 2012 \\ 1 & \text{if } t \geq 2012 \end{cases} \tag{D.12}$$

$$w^m_{gt} = 1 \qquad \text{for all } t \tag{D.13}$$

[54] Age groups are the four quartiles and earnings bins are defined by the gender and age group specific P5, P10, P20, . . . , P90, P95 of the real earnings distribution pooled over the entire sample period. This gives $2 \times 4 \times 12 = 96$ groups in each year.

D.2 Combined IAB-TPP Data in Earnings Analysis (Section 3)

D.2.1 Combined Earnings Distribution

For the core analysis of labor earnings, we focus exclusively on social-security workers as we do not have IAB data for non-social-security workers. The main idea in constructing the combined distribution of earnings is the following: Below the top-coding threshold of 60,000 Euro, we use the (true) earnings distribution from the IAB data. Above the cutoff, we use the conditional earnings distribution from the (reweighted) TPP along with the (true) number of workers above the cutoff in the IAB.

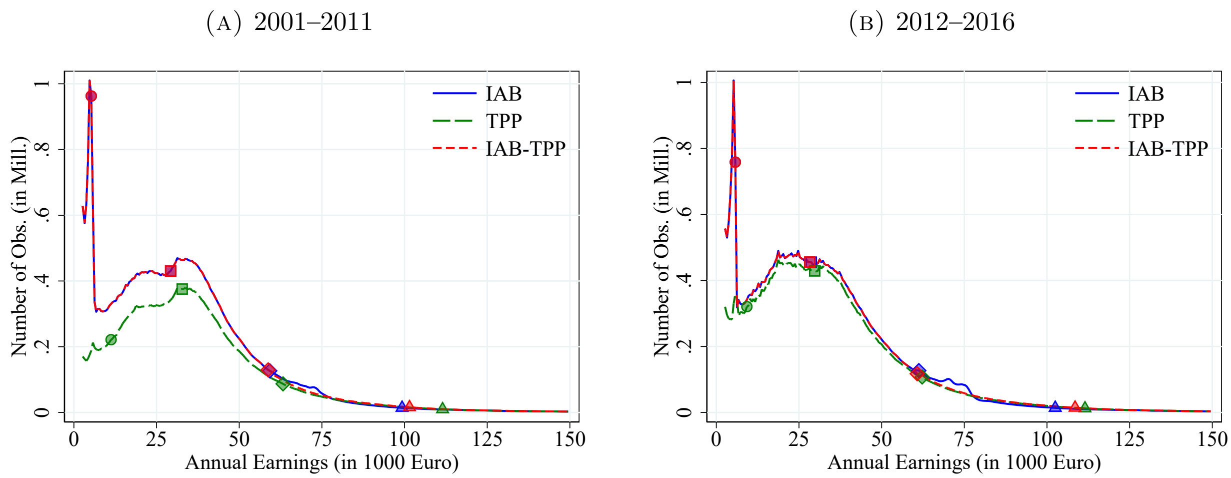
Technically, we (i) estimate the CDF of earnings in both data sources by monotonically interpolating a wide range of quantiles, and (ii) construct the combined CDF as:

$$F(y) = \begin{cases} F^{IAB}(y) & \text{if } y \leq \bar{y} \\ F^{IAB}(\bar{y}) + F^{TPP}(y|y > \bar{y})\Big(1 - F^{IAB}(\bar{y})\Big) & \text{if } y > \bar{y} \end{cases} \tag{D.14}$$

Figure 1 in the main text, Figure D.1 as well as Tables D.2, D.3 and D.4 show selected percentiles of the earnings distribution in the combined IAB-TPP data (CS sample) as well as in the IAB and TPP data for men, women and in the population respectively. Percentiles below 60,000 Euro (P75 and below) are practically identical in the IAB-TPP and IAB data, while higher percentiles are closer to the TPP data.[55]

[55]The small deviations below 60,000 Euro are the result of how we combine the IAB and TPP data. After interpolating the quantiles, we discretize the respective distributions on a very fine grid and then combine the discrete distributions. The deviations for higher percentiles are mostly driven by the fact that the number of observations in the TPP is smaller than in the combined IAB-TPP data. Adding individuals mostly at the bottom of the distribution moves the higher percentiles to different (lower) points in the income distribution (Krolage et al., forthcoming).

Figure D.1: Annual Earnings Distribution in IAB, TPP and Combined Data – Population



Notes: This figure shows the number of observations in real earnings bins for the IAB, the TPP and the combined data (IAB-TPP) in the full population (men and women). A complementing figure by gender can be found in Figure 1. Panel A shows averages across the years 2001 to 2011 where non-filing workers (Lohnsteuerfälle) are not included in the TPP and Panel B shows averages across the years 2012 to 2016 where the TPP data include these workers. We exclude earnings from the TPP that are not subject to social security contributions (e.g. salaries of civil servants) which are not covered in the IAB. The circular, square, diamond and triangle-shaped markers depict the 10^{th}, 50^{th}, 90^{th} and 99^{th} earnings percentile in the respective data sets. We use 500 Euro bins below 80,000 Euro and 1,000 Euro bins above 80,000 Euro but always plot the number of observations per 1,000 Euro bins. The IAB data are imputed above the social security contribution limit. Table D.4 shows selected earnings of these distributions percentiles across the different datasets.

Table D.2: Earnings Percentiles in IAB, TPP and Combined IAB-TPP Data – Men

Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
IAB-TPP Data												
2001	13.113	42,479	6,331	11,893	25,563	37,599	50,524	69,852	86,892	141,844	331,421	939,818
2002	12.822	42,568	6,028	11,410	25,382	37,700	50,959	70,717	87,871	142,546	331,502	894,797
2003	12.593	42,728	5,677	10,939	25,104	37,809	51,455	71,592	88,929	143,893	319,357	827,246
2004	12.419	42,455	5,406	10,088	24,398	37,434	51,274	71,866	89,516	145,790	335,445	896,172
2005	12.165	42,486	5,300	10,029	23,991	37,066	51,193	72,267	90,411	149,392	359,312	992,910
2006	12.214	42,524	5,308	9,886	23,364	36,615	51,229	72,641	91,491	153,904	377,530	1,135,252
2007	12.373	42,485	5,434	10,126	23,142	36,169	50,978	72,833	92,430	157,890	401,891	1,143,906
2008	12.430	42,459	5,409	10,282	23,013	36,029	50,985	72,925	92,556	158,500	397,202	1,117,735
2009	12.223	42,138	5,274	9,657	22,696	35,843	50,622	72,998	92,614	157,588	376,740	1,004,920
2010	12.275	42,061	5,293	9,678	22,230	35,687	51,048	73,269	92,728	157,180	377,176	1,011,613
2011	12.464	42,355	5,226	9,934	22,373	35,518	51,155	73,682	93,818	160,584	392,686	1,059,602
2012	12.535	42,433	5,121	10,104	22,479	35,528	51,330	74,097	94,212	160,328	387,358	1,025,836
2013	12.638	42,398	5,153	9,871	22,385	35,536	51,306	74,104	94,235	160,215	392,385	1,069,089
2014	12.796	42,714	5,172	9,615	22,332	35,669	51,782	74,946	95,322	162,603	396,617	1,097,696
2015	12.958	43,249	5,355	9,858	22,519	35,922	52,422	76,035	96,760	165,946	415,441	1,182,006
2016	13.096	43,665	5,445	9,991	22,831	36,206	52,817	76,755	97,788	168,152	415,058	1,188,906
IAB Data												
2001	13.113	40,781	6,336	11,898	25,568	37,604	50,529	67,943	83,324	136,482	241,005	385,336
2002	12.822	41,012	6,033	11,415	25,387	37,705	50,964	68,722	85,047	138,889	243,050	387,108
2003	12.593	40,952	5,682	10,944	25,109	37,814	51,460	71,056	84,389	133,374	229,630	343,458
2004	12.419	40,696	5,411	10,093	24,403	37,439	51,279	71,116	85,040	136,700	233,129	358,618
2005	12.165	40,658	5,305	10,034	23,996	37,071	51,197	71,404	86,047	139,593	243,628	371,340
2006	12.214	40,559	5,313	9,891	23,369	36,620	51,234	71,520	86,821	144,560	253,764	388,987
2007	12.373	40,516	5,439	10,131	23,147	36,174	50,983	71,249	87,457	148,149	268,884	418,559
2008	12.430	40,562	5,414	10,287	23,018	36,034	50,990	70,667	88,129	151,311	272,001	422,230
2009	12.223	40,260	5,279	9,662	22,701	35,848	50,627	71,190	87,712	149,965	267,141	424,189
2010	12.275	40,370	5,298	9,683	22,235	35,692	51,053	71,757	88,669	152,613	275,987	435,400
2011	12.464	40,438	5,231	9,939	22,378	35,523	51,160	71,030	89,298	152,928	274,180	428,103
2012	12.535	40,499	5,126	10,109	22,484	35,533	51,335	71,043	89,484	152,143	266,911	406,601
2013	12.638	40,348	5,158	9,876	22,390	35,541	51,311	71,638	88,450	151,273	262,699	391,128
2014	12.796	40,620	5,177	9,620	22,337	35,674	51,787	72,468	89,232	152,698	271,648	412,144
2015	12.958	41,009	5,360	9,863	22,524	35,927	52,427	73,256	89,845	153,676	271,108	407,527
2016	13.096	41,347	5,450	9,996	22,836	36,211	52,822	74,101	90,180	154,300	274,245	415,386
TPP Data												
2001	10.570	45,203	10,396	17,269	28,853	39,482	53,752	75,043	93,188	154,641	370,435	1,065,239
2002	10.722	44,927	9,601	16,295	28,308	39,302	53,898	75,336	93,294	153,864	366,047	1,005,616
2003	10.396	45,073	9,396	16,037	28,262	39,495	54,283	76,071	94,201	154,617	349,718	914,305
2004	10.057	45,240	9,230	15,884	28,032	39,409	54,413	76,692	95,174	157,625	371,057	1,007,321
2005	9.662	46,041	9,486	16,089	27,826	39,232	54,519	77,338	96,326	162,219	397,859	1,135,101
2006	9.404	45,889	9,923	16,383	27,648	39,156	54,767	78,085	97,752	168,418	420,974	1,307,196
2007	9.308	46,664	10,493	16,930	27,686	39,013	54,829	78,846	99,425	173,991	454,312	1,330,240
2008	9.058	46,406	10,875	17,191	27,582	38,890	54,923	79,042	99,794	174,872	448,526	1,301,801
2009	8.851	46,016	10,128	16,533	26,994	38,473	54,675	79,374	100,190	173,638	423,921	1,166,691
2010	8.634	46,250	10,366	16,576	27,049	38,965	55,360	79,845	100,482	173,391	427,305	1,161,216
2011	8.688	47,027	11,069	17,385	27,464	39,235	56,177	80,965	102,446	179,416	448,392	1,232,365
2012	11.392	42,685	7,099	12,690	23,776	36,071	52,039	75,327	95,646	163,445	395,580	1,046,754
2013	11.782	42,345	6,684	11,869	23,239	35,774	51,846	75,096	95,361	162,571	399,861	1,089,920
2014	11.973	42,762	6,571	11,663	23,260	35,987	52,275	75,876	96,417	164,851	403,828	1,117,001
2015	12.150	43,114	6,680	11,789	23,339	36,139	52,830	76,832	97,749	168,104	421,757	1,193,978
2016	12.034	44,278	7,535	13,218	24,439	36,942	53,713	78,095	99,400	171,702	425,666	1,219,048

Notes: This table shows selected earnings percentiles for men in the combined IAB-TPP, the (imputed) IAB and TPP data. CS sample.

Table D.3: Earnings Percentiles in IAB, TPP and Combined IAB-TPP Data – Women

Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
IAB-TPP Data												
2001	11.476	24,628	3,671	4,673	10,881	22,162	34,723	46,032	54,306	78,226	137,340	285,996
2002	11.363	24,794	3,664	4,642	10,950	22,240	34,917	46,413	55,011	79,345	138,699	286,513
2003	11.166	24,898	3,630	4,692	10,759	22,289	35,129	46,840	55,509	79,988	137,961	273,802
2004	11.101	25,281	3,599	4,686	9,905	21,778	34,760	46,552	55,572	81,110	141,506	281,579
2005	10.965	25,360	3,596	4,646	9,736	21,580	34,540	46,490	55,603	81,729	145,781	299,090
2006	11.012	25,182	3,601	4,614	9,520	21,237	34,199	46,300	55,609	82,697	154,891	310,675
2007	11.146	25,025	3,653	4,686	9,602	20,907	33,722	45,957	55,521	84,004	158,305	327,633
2008	11.228	25,002	3,695	4,736	9,811	20,845	33,761	46,100	55,701	84,335	158,265	335,160
2009	11.223	25,309	3,706	4,745	9,897	21,013	34,147	46,845	56,361	85,345	158,294	324,855
2010	11.304	25,248	3,749	4,811	10,105	21,020	34,148	47,123	57,013	86,429	161,010	341,354
2011	11.439	25,443	3,791	4,812	10,376	20,987	34,034	47,034	57,110	87,228	165,093	359,576
2012	11.510	25,610	3,832	4,861	10,664	21,107	34,085	47,149	57,410	88,128	167,849	363,598
2013	11.585	25,855	3,868	4,994	10,913	21,353	34,345	47,430	57,738	89,057	169,827	367,494
2014	11.667	26,327	3,937	5,002	11,219	21,693	34,896	48,180	58,802	91,216	175,577	380,719
2015	11.756	26,915	4,090	5,218	11,846	22,188	35,487	48,998	59,914	93,360	181,589	406,595
2016	11.799	27,562	4,178	5,366	12,387	22,805	36,241	49,859	61,077	95,403	185,926	414,783
IAB Data												
2001	11.476	24,558	3,676	4,678	10,886	22,167	34,728	46,037	54,311	77,237	126,590	185,188
2002	11.363	24,751	3,669	4,647	10,955	22,245	34,922	46,418	55,016	79,544	130,643	194,996
2003	11.166	24,823	3,635	4,697	10,764	22,293	35,134	46,845	55,514	77,428	126,331	187,049
2004	11.101	24,455	3,604	4,691	9,910	21,783	34,765	46,557	55,577	78,231	128,578	194,310
2005	10.965	24,334	3,601	4,651	9,741	21,585	34,545	46,495	55,608	78,982	132,035	197,873
2006	11.012	24,135	3,606	4,619	9,525	21,242	34,204	46,305	55,614	79,686	136,445	205,417
2007	11.146	23,954	3,658	4,691	9,607	20,912	33,727	45,962	55,526	80,683	140,641	222,722
2008	11.228	24,010	3,700	4,741	9,816	20,850	33,766	46,105	55,706	81,403	144,091	225,600
2009	11.223	24,248	3,711	4,750	9,902	21,018	34,152	46,850	56,366	82,021	144,067	226,262
2010	11.304	24,394	3,754	4,816	10,110	21,025	34,153	47,128	57,018	83,606	149,191	242,831
2011	11.439	24,380	3,796	4,817	10,381	20,992	34,039	47,039	57,115	83,009	145,068	224,170
2012	11.510	24,522	3,837	4,866	10,669	21,112	34,090	47,154	57,415	84,084	143,421	220,381
2013	11.585	24,744	3,873	4,999	10,918	21,358	34,350	47,435	57,743	83,587	143,554	220,791
2014	11.667	25,182	3,942	5,007	11,224	21,698	34,901	48,185	58,807	84,805	146,710	221,356
2015	11.756	25,758	4,095	5,223	11,851	22,193	35,492	49,003	59,919	85,858	149,172	219,736
2016	11.799	26,362	4,183	5,371	12,392	22,811	36,246	49,864	61,022	86,977	148,851	226,320
TPP Data												
2001	7.979	28,866	5,536	8,757	15,945	25,605	37,069	48,836	57,770	83,738	152,169	325,486
2002	8.316	28,430	5,548	8,769	15,966	25,674	37,366	49,393	58,579	84,761	153,326	314,386
2003	8.040	28,607	5,744	8,885	16,071	25,829	37,775	49,892	59,066	85,574	153,277	307,138
2004	7.870	28,541	5,826	8,709	15,878	25,622	37,624	49,906	59,557	87,223	156,735	319,596
2005	7.624	29,001	5,848	8,708	15,852	25,511	37,520	49,852	59,595	88,426	163,486	351,511
2006	7.468	28,433	5,828	8,671	15,703	25,267	37,282	49,710	59,685	89,720	175,841	360,206
2007	7.480	28,201	5,817	8,594	15,493	24,901	36,816	49,290	59,671	91,234	177,866	378,985
2008	7.409	28,089	5,752	8,514	15,279	24,618	36,667	49,015	59,452	91,185	175,199	384,082
2009	7.363	28,338	5,736	8,492	15,297	24,753	37,174	50,068	60,632	92,857	177,928	369,099
2010	7.330	28,402	5,731	8,489	15,242	24,746	37,242	50,487	61,074	93,594	178,933	391,947
2011	7.409	28,402	5,737	8,501	15,224	24,658	37,103	50,513	61,299	94,831	184,320	414,964
2012	9.528	27,466	5,186	7,555	13,978	23,270	35,739	48,838	59,463	91,641	177,184	386,224
2013	9.883	27,137	5,155	7,524	13,957	23,319	35,879	49,124	59,839	92,625	178,170	390,722
2014	10.043	28,046	5,139	7,531	14,166	23,667	36,405	49,889	60,877	94,553	183,884	398,124
2015	10.225	28,419	5,327	7,788	14,531	23,923	36,819	50,539	61,826	96,454	188,698	428,613
2016	10.108	28,836	5,678	8,374	15,194	24,784	37,796	51,689	63,149	98,685	193,986	441,093

Notes: This table shows selected earnings percentiles for women in the combined IAB-TPP, the (imputed) IAB and TPP data. CS sample.

Table D.4: Earnings Percentiles in IAB, TPP and Combined IAB-TPP Data – Population

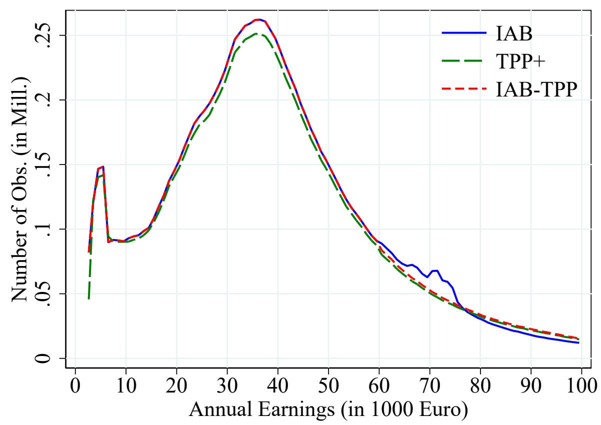
Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
IAB-TPP Data												
2001	24.588	33,650	4,291	5,698	16,757	31,007	43,605	59,612	74,334	119,461	265,812	701,658
2002	24.185	34,680	4,253	5,613	16,587	30,994	43,835	60,288	75,170	119,931	263,841	683,320
2003	23.759	34,733	4,208	5,668	16,439	31,012	44,169	60,886	76,026	120,941	258,531	620,545
2004	23.520	34,419	4,127	5,652	15,746	30,440	43,860	60,903	76,436	122,193	267,605	669,387
2005	23.130	34,329	4,086	5,594	15,559	30,030	43,671	60,960	76,885	124,538	282,627	758,668
2006	23.226	34,228	4,105	5,585	15,249	29,517	43,438	61,120	77,455	127,077	297,428	838,679
2007	23.519	34,212	4,179	5,445	15,173	29,081	43,079	61,052	77,876	129,878	309,054	878,039
2008	23.657	34,194	4,226	5,409	15,179	28,901	43,053	61,113	77,962	130,154	311,270	847,518
2009	23.445	34,060	4,188	5,394	14,975	28,794	43,109	61,053	78,047	130,264	298,007	772,972
2010	23.579	33,168	4,210	5,335	14,946	28,566	43,243	61,482	78,343	130,072	297,598	779,249
2011	23.903	33,337	4,212	5,235	15,195	28,489	43,136	61,849	79,027	132,557	309,794	836,579
2012	24.044	33,462	4,258	5,419	15,378	28,525	43,165	62,223	79,501	133,152	308,835	806,156
2013	24.224	34,473	4,332	5,677	15,420	28,600	43,327	62,344	79,680	133,260	312,106	811,316
2014	24.463	34,905	4,391	5,638	15,588	28,808	43,806	63,206	80,862	135,377	314,114	833,196
2015	24.715	35,500	4,533	6,059	16,169	29,148	44,362	64,193	82,172	138,339	325,499	898,856
2016	24.895	36,046	4,629	6,371	16,599	29,688	44,908	64,990	83,174	140,168	330,090	905,756
IAB Data												
2001	24.588	33,210	4,296	5,703	16,762	31,012	43,610	59,617	71,786	116,085	211,082	344,387
2002	24.185	33,372	4,258	5,618	16,592	30,999	43,840	60,285	73,015	118,395	212,391	341,753
2003	23.759	33,372	4,213	5,673	16,444	31,017	44,174	60,873	74,597	114,596	201,486	311,873
2004	23.520	33,030	4,132	5,657	15,751	30,445	43,865	60,885	74,394	116,687	207,185	323,429
2005	23.130	32,920	4,091	5,599	15,564	30,035	43,676	60,944	74,251	119,296	212,988	337,394
2006	23.226	32,772	4,110	5,590	15,254	29,522	43,443	61,079	74,469	121,980	221,301	352,315
2007	23.519	32,667	4,184	5,558	15,178	29,086	43,084	61,013	74,345	124,768	232,408	369,988
2008	23.657	32,706	4,231	5,414	15,184	28,906	43,058	61,037	74,385	126,856	238,196	379,414
2009	23.445	32,595	4,193	5,398	14,980	28,799	43,114	60,976	74,260	125,714	233,515	376,841
2010	23.579	32,711	4,215	5,340	14,951	28,571	43,248	61,399	74,985	128,004	240,624	388,516
2011	23.903	32,753	4,217	5,232	15,200	28,494	43,141	61,746	75,041	128,224	239,371	382,640
2012	24.044	32,851	4,263	5,424	15,383	28,530	43,170	62,025	75,396	128,269	234,306	365,760
2013	24.224	32,885	4,337	5,682	15,425	28,605	43,332	62,090	74,753	127,146	230,019	353,749
2014	24.463	33,257	4,396	5,642	15,593	28,813	43,811	62,851	75,566	128,792	236,544	369,246
2015	24.715	33,754	4,538	6,064	16,174	29,153	44,367	63,866	76,383	129,541	236,768	371,329
2016	24.895	34,245	4,634	6,376	16,604	29,693	44,913	64,684	77,146	130,054	238,211	374,076
TPP Data												
2001	18.549	37,871	7,288	11,741	21,331	33,960	46,672	64,861	80,946	131,421	302,132	821,165
2002	19.039	37,687	7,029	11,465	20,961	33,768	46,693	65,048	81,171	130,897	297,634	788,232
2003	18.436	37,897	7,094	11,442	21,000	33,939	47,116	65,627	81,964	131,644	288,229	706,906
2004	17.927	37,932	6,964	11,249	20,738	33,722	47,070	65,987	82,740	133,523	302,414	772,816
2005	17.286	38,013	7,039	11,344	20,663	33,440	46,989	66,238	83,462	136,557	318,894	875,066
2006	16.872	38,800	7,107	11,371	20,529	33,227	47,001	66,601	84,488	140,408	336,207	972,140
2007	16.788	38,227	7,158	11,445	20,451	32,925	46,776	66,902	85,430	144,138	356,608	1,015,749
2008	16.467	38,078	7,177	11,406	20,309	32,724	46,565	66,887	85,427	144,384	354,247	997,556
2009	16.214	37,918	7,013	11,148	20,108	32,524	46,607	67,244	85,755	144,617	339,340	867,527
2010	15.964	38,053	7,027	11,138	19,993	32,586	47,056	67,588	86,058	144,271	340,943	910,341
2011	16.097	38,448	7,167	11,353	20,171	32,691	47,327	68,238	87,254	147,770	358,185	944,551
2012	20.921	36,136	5,769	9,301	17,949	30,315	44,495	64,207	81,958	137,468	321,052	831,895
2013	21.665	35,403	5,693	9,082	17,716	30,077	44,438	64,153	81,904	136,935	323,332	844,378
2014	22.016	36,293	5,713	9,076	17,846	30,300	44,870	64,871	82,915	138,998	325,341	869,572
2015	22.375	36,187	5,822	9,286	18,255	30,458	45,305	65,710	84,056	141,618	335,844	921,056
2016	22.142	37,560	6,263	10,069	19,098	31,344	46,237	66,867	85,536	144,316	342,562	946,374

Notes: This table shows selected earnings percentiles for men and women in the combined IAB-TPP, the (imputed) IAB and TPP data. CS sample.

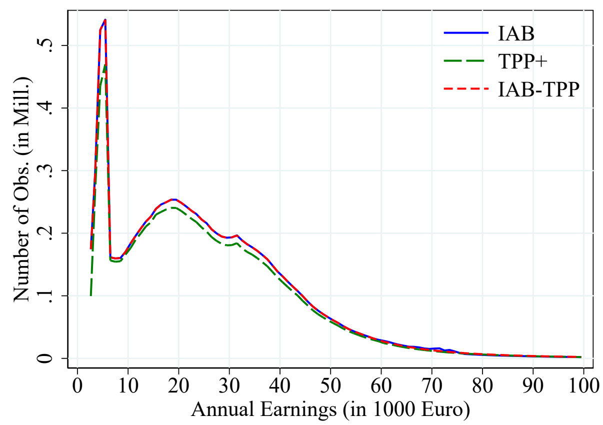
For the LS samples, we follow the same procedure. While the cross-sectional earnings distribution in the reweighted TPP data matches the IAB data (by construction of the weights), this is no longer the case for the LS and H samples due to attrition in the TPP.[56] The LS samples differ from the CS sample in that workers have to be in the data in year t and $t+1$ or $t+5$. Figure D.2 shows the earnings distribution in these samples in the IAB, the reweighted TPP and the combined IAB-TPP data. The attrition in the reweighted TPP data becomes particularly visible in Panels C and D which plots the earnings distribution in the LS sample for 5-year earnings changes.

Figure D.2: IAB vs. TPP: Earnings Distribution in Longitudinal Samples

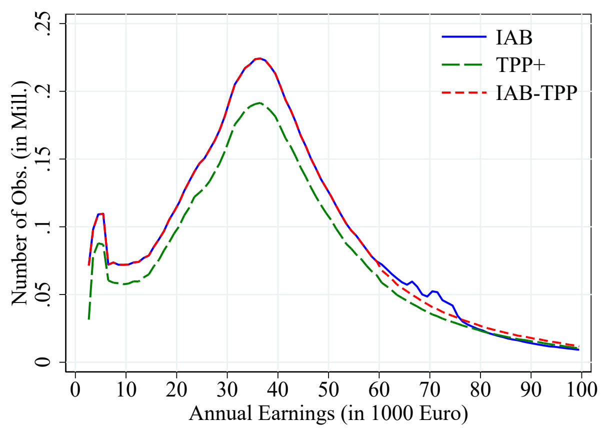
(a) One-Year Changes: Men



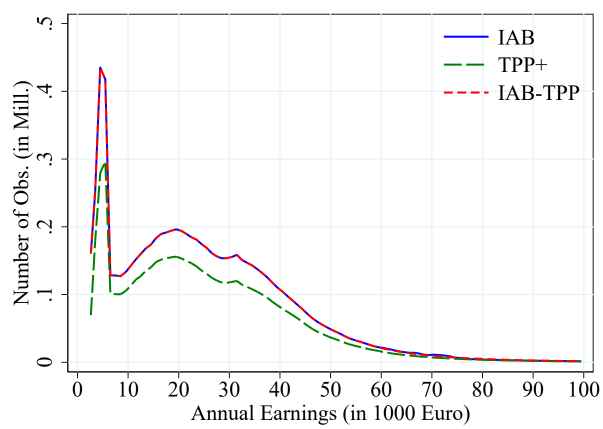
(b) One-Year Changes: Women



(c) Five-Year Changes: Men



(d) Five-Year Changes: Women



Notes: LS sample. Annual earnings. Averaged over years 2001-2015 for one-year changes and 2001-2011 for five-year changes. Source: IAB and TPP.

D.2.2 Combined Earnings Growth Distribution

For the analysis of earnings dynamics, we are interested in the distribution of earnings growth, i.e. the distribution of earnings *changes* in addition to the earnings distribution shown in Figure D.2

[56]Recall, that the reweighting does not target moments of earnings changes over time. For example, many workers who switch from a regular job to a mini-job will drop out of the TPP.

for the different samples. To construct this distribution of changes, we proceed as follows. For simplicity, we drop time subscripts for all variables and use the following notation:

- earnings y (continuous)
- earnings bins Y (discrete and finite support)
- earnings growth $g = \log(y_{t+k}) - log(y)$ (continuous)
- earnings growth bins G (discrete and finite support)

Available Data. For each data source (IAB and reweighted TPP) and by year and gender, we have

- the share of workers in each earnings bin: $\Pr(Y)$
- summary statistics (e.g. mean, standard deviation, skewness, kurtosis) and selected quantiles[57] of earnings growth by earnings bin:

$$q^p(g|Y) \equiv F^{-1}_{g|Y}(p/100|Y) \quad \text{for selected values of } q \in (0,1) \tag{D.15}$$

Conditional Growth Rate Distributions by Earnings Bins. In a first step, we approximate the conditional CDF of earnings growth, $F_{g|Y}$ in both the IAB and reweighted TPP data using a continuous interpolation of its quantiles.[58]

Figures D.3 shows the P90-P10 differential, Kelley Skewness and Excess Crow-Siddiqui kurtosis of 1-year earnings growth by current earnings in the IAB and reweighted TPP data. In the middle of the earnings distribution, the conditional earnings growth distributions are very similar in the IAB and reweighted TPP data. However, there are stark differences at the bottom (where the TPP has a lot of attrition because of missing mini-jobs) and even more so above the top-coding threshold where imputed earnings in the IAB are essentially iid. Figure D.4 shows the corresponding statistics for 5-year earnings growth. While the IAB and (reweighted) TPP are again remarkably similar in the middle of the male earnings distribution, the fit becomes slightly worse for women.

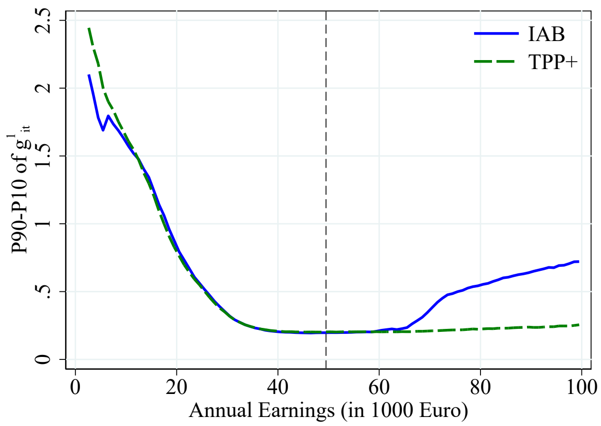
In order to construct the combined IAB-TPP data set for earnings changes, we proceed as follows. First, for low initial earnings bins, we use the conditional earnings growth distribution from the IAB as the probability of having top-coded earnings in $t + k$ is very low. Second, as soon as more than 2% of the IAB earnings growth distribution is affected by top-coding, we switch to the conditional earnings growth distribution in the TPP. Figure D.5 plots the share of 1- and 5-year log earnings growth rates affected by top-coding in the IAB. The vertical lines indicate the chosen thresholds for switching from IAB to TPP data.

[57] We have the following percentiles: 0.1, 0.5, 1, 2, . . . , 10, 15, . . . , 90, 91, . . . 99, 99.5, 99.9.

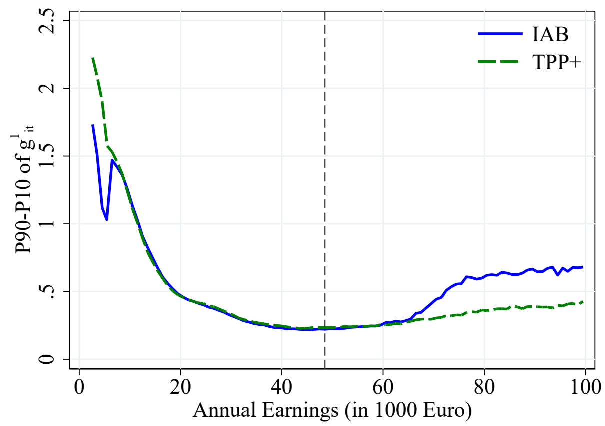
[58] In order to approximate the CDF using monotonic spline interpolation of quantiles in a given dataset, we have to impose a minimum and maximum for g, i.e. q^0 and q^{100}. Let $\hat{F}_{g|Y}$ denote the resulting approximation of the CDF of earnings growth. We set the minimum and the maximum such that the standard deviation and skewness of $\hat{F}_{g|Y}$ equal the values that we observe in the data.

Figure D.3: IAB vs. TPP: 1-Year Log Earnings Changes by Current Earnings

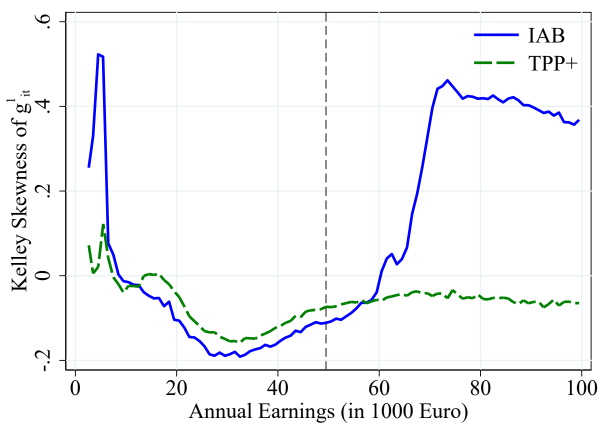
(a) P90-P10: Men



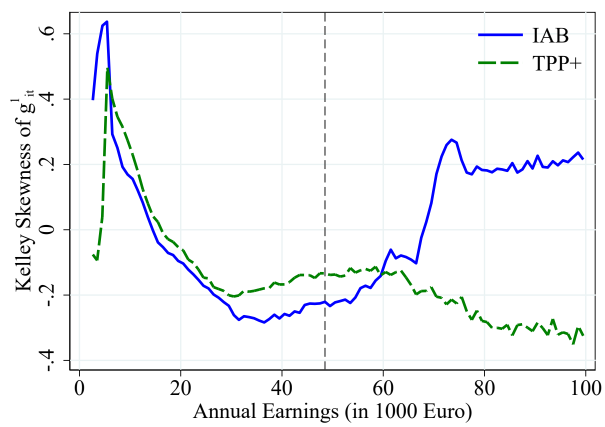
(b) P90-P10: Women



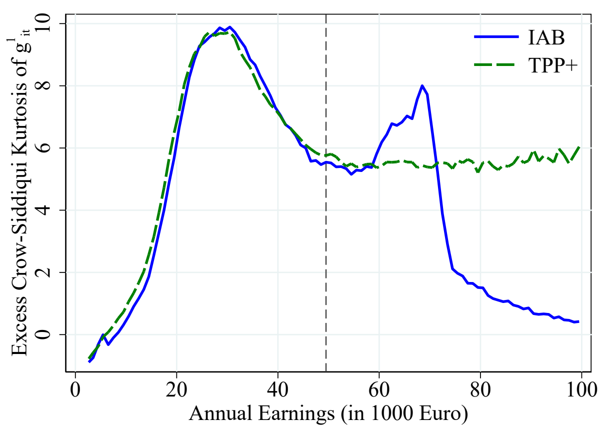
(c) Kelley Skewness: Men



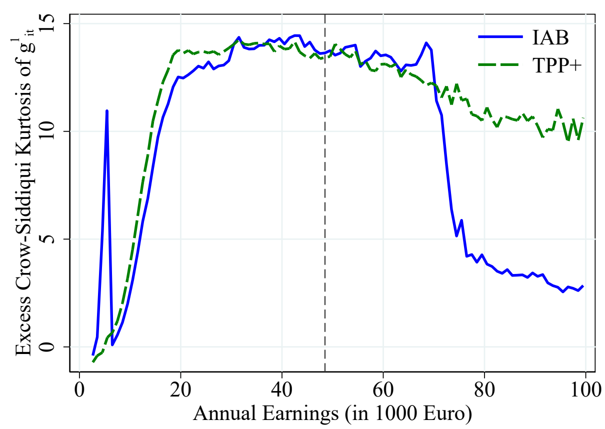
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



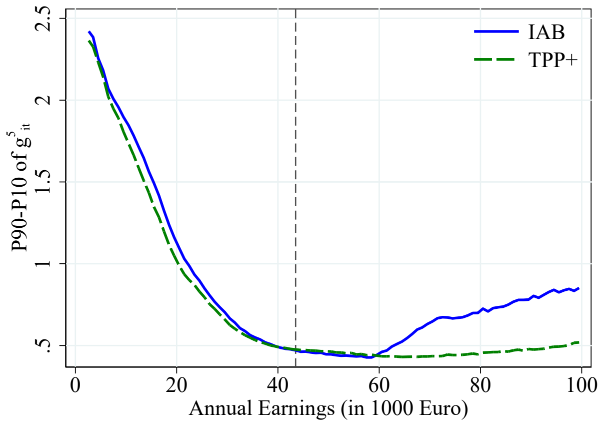
(f) Excess Crow-Siddiqui Kurtosis: Women



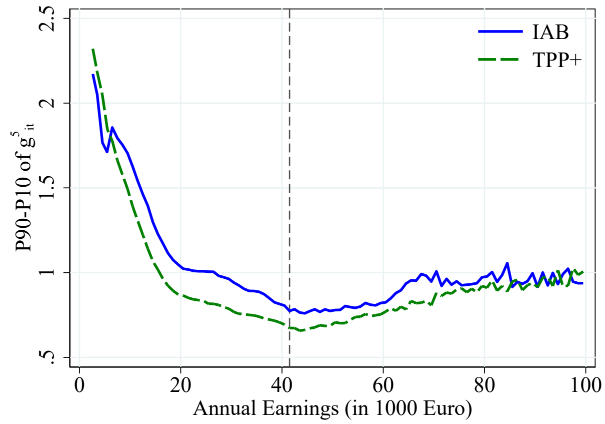
Notes: LS sample. One-year residualized log earnings growth. Averaged over years 2001-2015. Source: IAB and TPP.

Figure D.4: IAB vs. TPP: 5-Year Log Earnings Changes by Current Earnings

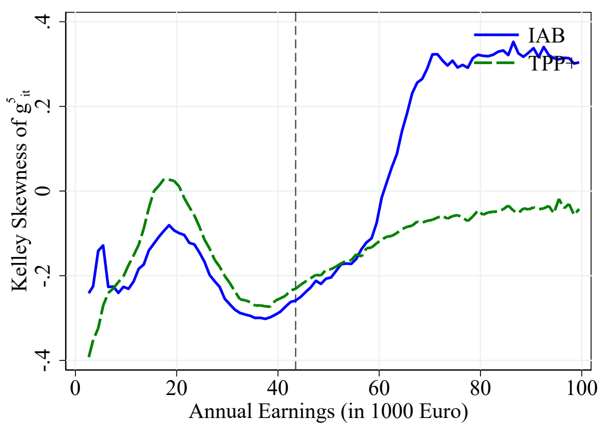
(a) P90-P10: Men



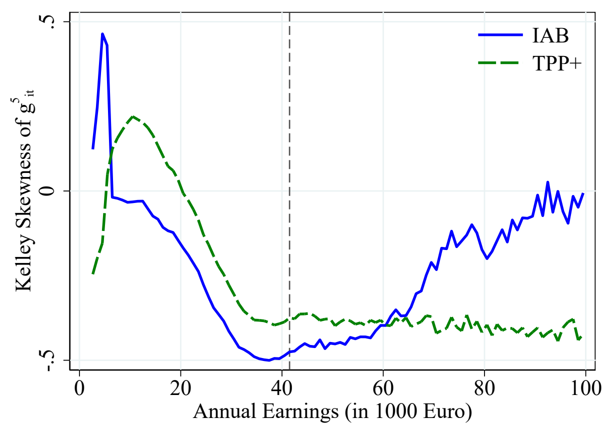
(b) P90-P10: Women



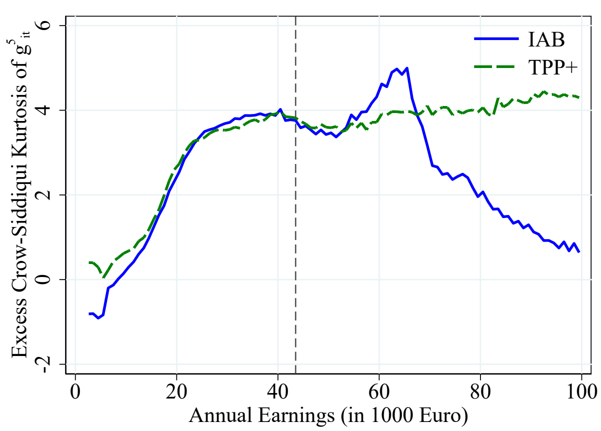
(c) Kelley Skewness: Men



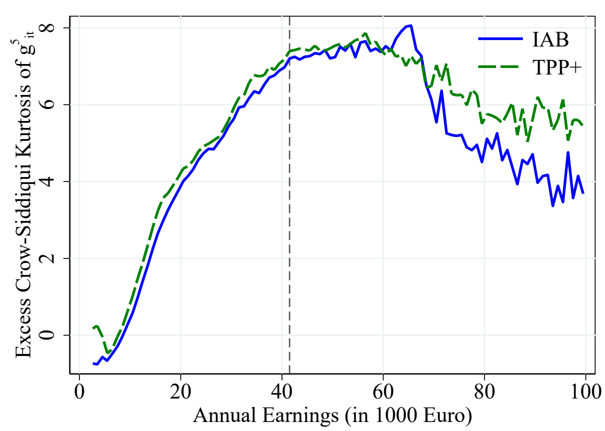
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men

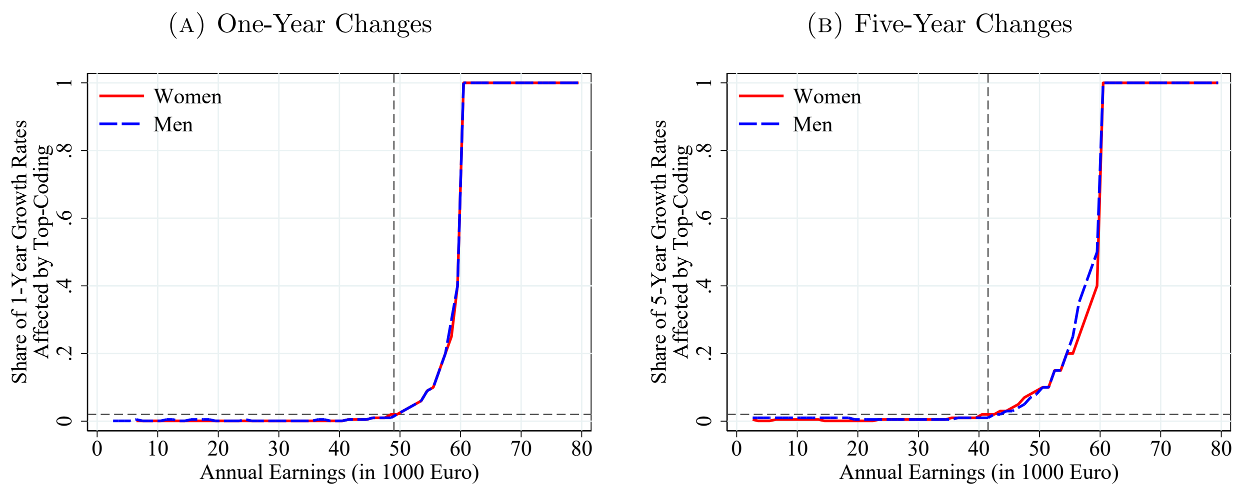


(f) Excess Crow-Siddiqui Kurtosis: Women



Notes: LS sample. Five-year residualized log earnings growth. Averaged over years 2001-2011. Source: IAB and TPP.

FIGURE D.5: SHARE OF 1-YEAR LOG EARNINGS CHANGES AFFECTED BY TOP-CODING IN THE IAB DATA



Notes: LS sample. Averages over years. Earnings changes are affected by top-coding if current or future earnings are above 60,000 Euro. The dashed vertical depicts the point where 2% of earnings changes are affected by top-coding. Source: IAB.

In the next step, we discretize the continuous conditional earnings growth distributions. To do so, we set up a fine grid for g ranging from the global minimum to the global maximum of the support of $\hat{F}_{g|Y}$. The grid defines earnings growth bins G with upper and lower bounds denoted by G^+ and G^- respectively. Using those, we discretize the continuous conditional distributions to obtain $\Pr(G|Y)$ for all G and Y:

$$\Pr(G|Y) = \Pr(G^- \leq g \leq G^+|Y) = \hat{F}_{g|Y}(G^+|Y) - \hat{F}_{g|Y}(G^-|Y) \tag{D.16}$$

Unconditional Growth Rate Distribution. Finally, this discretized conditional growth distribution allows us to recover the (unconditional) marginal probability mass function of earnings growth (discretized) defined by the probabilities

$$\Pr(G) = \sum_Y \Pr(G, Y) = \sum_Y \Pr(G|Y) \Pr(Y) \tag{D.17}$$

where $\Pr(Y)$ is the discretized combined IAB-TPP earnings distribution in the corresponding LS sample (see Figure D.2). As the bins are very fine, we simply use their midpoints along with the above probabilities to compute summary statistics and selected percentiles of the unconditional distribution of earnings growth. Tables D.6 and D.7 show selected percentiles of the 1-year earnings growth distribution using the combined IAB-TPP data as well as the IAB and (reweighted) TPP data. Tables D.9 and D.10 show the corresponding statistics for 5-year earnings growth.

Table D.5: Percentiles of Real Annual Earnings – LS Sample with 1-Year Changes

Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
Men												
2001	11.912	43,185	8,791	15,719	27,982	38,659	51,532	70,997	88,082	142,873	330,815	913,507
2005	11.179	43,018	6,707	13,234	26,104	38,036	52,112	73,355	91,463	149,478	355,985	980,908
2010	11.379	42,175	6,183	11,844	23,742	36,483	51,748	74,028	93,335	156,230	369,829	1,000,144
Women												
2001	10.214	25,856	4,056	4,890	12,765	23,666	35,769	46,942	55,245	79,505	138,651	287,446
2005	9.874	25,434	3,951	4,992	11,545	22,757	35,423	47,268	56,446	82,735	145,376	294,698
2010	10.253	25,286	4,080	5,095	11,241	21,930	34,863	47,791	57,719	87,362	159,988	335,696

Notes: This table shows the number of observations (in millions) and selected percentiles of real annual earnings (in 2018 Euro) in the combined IAB-TPP data. LS sample with non-missing 1-year log earnings changes (from t to $t+1$). Sources: IAB and TPP.

Table D.6: Percentiles of 1-Year Earnings Growth in Combined IAB-TPP Data – Men

Year	N	P1	P2.5	P10	P25	P50	P75	P90	P95	P97.5	P99
IAB-TPP Data											
2001	11.912	-1.576	-1.039	-0.234	-0.058	-0.005	0.046	0.202	0.497	0.873	1.362
2005	11.179	-1.379	-0.813	-0.161	-0.049	-0.003	0.052	0.227	0.551	0.939	1.391
2010	11.379	-1.250	-0.722	-0.164	-0.054	-0.005	0.071	0.282	0.616	0.985	1.427
IAB Data											
2001	11.912	-1.571	-1.052	-0.294	-0.067	-0.006	0.053	0.270	0.563	0.901	1.370
2005	11.179	-1.365	-0.813	-0.204	-0.057	-0.004	0.059	0.280	0.593	0.946	1.396
2010	11.379	-1.259	-0.764	-0.212	-0.064	-0.006	0.076	0.342	0.660	1.001	1.433
TPP+ Data											
2001	9.694	-1.678	-1.032	-0.215	-0.041	0.013	0.064	0.244	0.583	1.006	1.484
2005	8.413	-1.455	-0.801	-0.135	-0.031	0.012	0.069	0.270	0.657	1.088	1.556
2010	7.791	-1.257	-0.661	-0.128	-0.033	0.014	0.095	0.349	0.772	1.205	1.653

Notes: This table shows the number of observations (in millions) and selected percentiles of the combined IAB-TPP distribution of 1-year changes in residualized log earnings (from t to $t+1$) for men and selected years. LS sample. Sources: IAB and TPP.

Table D.7: Percentiles of 1-Year Earnings Growth in Combined IAB-TPP Data – Women

Year	N	P1	P2.5	P10	P25	P50	P75	P90	P95	P97.5	P99
IAB-TPP Data											
2001	11.912	-1.576	-1.039	-0.234	-0.058	-0.005	0.046	0.202	0.497	0.873	1.362
2005	11.179	-1.379	-0.813	-0.161	-0.049	-0.003	0.052	0.227	0.551	0.939	1.391
2010	11.379	-1.250	-0.722	-0.164	-0.054	-0.005	0.071	0.282	0.616	0.985	1.427
IAB Data											
2001	11.912	-1.571	-1.052	-0.294	-0.067	-0.006	0.053	0.270	0.563	0.901	1.370
2005	11.179	-1.365	-0.813	-0.204	-0.057	-0.004	0.059	0.280	0.593	0.946	1.396
2010	11.379	-1.259	-0.764	-0.212	-0.064	-0.006	0.076	0.342	0.660	1.001	1.433
TPP+ Data											
2001	9.694	-1.678	-1.032	-0.215	-0.041	0.013	0.064	0.244	0.583	1.006	1.484
2005	8.413	-1.455	-0.801	-0.135	-0.031	0.012	0.069	0.270	0.657	1.088	1.556
2010	7.791	-1.257	-0.661	-0.128	-0.033	0.014	0.095	0.349	0.772	1.205	1.653

Notes: This table shows the number of observations (in millions) and selected percentiles of the combined IAB-TPP distribution of 1-year changes in residualized log earnings (from t to $t+1$) for women and selected years. LS sample. Sources: IAB and TPP.

Table D.8: Percentiles of Real Annual Earnings – LS Sample with 5-Year Changes

Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
Men												
2001	9.599	43,174	9,561	16,990	28,791	38,951	51,505	70,357	86,747	135,360	297,389	834,582
2005	9.189	42,397	6,782	13,277	26,222	37,984	51,743	72,294	89,458	139,913	310,479	842,905
2010	9.253	41,387	6,169	11,788	23,680	36,211	51,149	72,504	90,854	147,325	329,631	872,714
Women												
2001	7.982	25,973	4,039	4,903	12,973	23,888	35,891	46,956	55,198	79,289	135,779	273,620
2005	7.899	25,205	3,845	4,918	11,282	22,560	35,182	46,965	56,099	82,314	140,779	277,829
2010	8.119	25,059	3,999	5,042	11,015	21,751	34,595	47,442	57,398	86,954	156,671	320,050

Notes: This table shows the number of observations (in millions) and selected percentiles of real annual earnings (in 2018 Euro) in the combined IAB-TPP data. LS sample with non-missing 5-year log earnings changes (from t to $t+5$). Sources: IAB and TPP.

Table D.9: Percentiles of 5-Year Earnings Growth in Combined IAB-TPP Data – Men

Year	N	P1	P2.5	P10	P25	P50	P75	P90	P95	P97.5	P99
IAB-TPP Data											
2001	9.599	-1.990	-1.396	-0.436	-0.147	-0.008	0.110	0.343	0.719	1.205	1.788
2005	9.189	-1.807	-1.178	-0.380	-0.127	0.009	0.137	0.462	0.926	1.414	1.935
2010	9.253	-1.650	-1.013	-0.336	-0.106	0.024	0.175	0.548	1.013	1.473	1.973
IAB Data											
2001	9.599	-1.974	-1.378	-0.452	-0.160	-0.012	0.115	0.386	0.752	1.203	1.794
2005	9.189	-1.777	-1.155	-0.391	-0.139	0.001	0.139	0.512	0.950	1.420	1.941
2010	9.253	-1.631	-1.018	-0.369	-0.128	0.012	0.172	0.582	1.024	1.479	1.977
TPP+ Data											
2001	6.531	-2.026	-1.276	-0.372	-0.123	0.007	0.123	0.379	0.811	1.331	1.898
2005	5.708	-1.835	-1.090	-0.340	-0.106	0.023	0.148	0.478	0.976	1.488	1.992
2010	6.383	-1.713	-0.975	-0.299	-0.085	0.040	0.195	0.616	1.161	1.659	2.130

Notes: This table shows the number of observations (in millions) and selected percentiles of the combined IAB-TPP distribution of 5-year changes in residualized log earnings (from t to $t+5$) for men and selected years. LS sample. Sources: IAB and TPP.

Table D.10: Percentiles of 5-Year Earnings Growth in Combined IAB-TPP Data – Women

Year	N	P1	P2.5	P10	P25	P50	P75	P90	P95	P97.5	P99
IAB-TPP Data											
2001	11.912	-1.576	-1.039	-0.234	-0.058	-0.005	0.046	0.202	0.497	0.873	1.362
2005	11.179	-1.379	-0.813	-0.161	-0.049	-0.003	0.052	0.227	0.551	0.939	1.391
2010	11.379	-1.250	-0.722	-0.164	-0.054	-0.005	0.071	0.282	0.616	0.985	1.427
IAB Data											
2001	11.912	-1.571	-1.052	-0.294	-0.067	-0.006	0.053	0.270	0.563	0.901	1.370
2005	11.179	-1.365	-0.813	-0.204	-0.057	-0.004	0.059	0.280	0.593	0.946	1.396
2010	11.379	-1.259	-0.764	-0.212	-0.064	-0.006	0.076	0.342	0.660	1.001	1.433
TPP+ Data											
2001	9.694	-1.678	-1.032	-0.215	-0.041	0.013	0.064	0.244	0.583	1.006	1.484
2005	8.413	-1.455	-0.801	-0.135	-0.031	0.012	0.069	0.270	0.657	1.088	1.556
2010	7.791	-1.257	-0.661	-0.128	-0.033	0.014	0.095	0.349	0.772	1.205	1.653

Notes: This table shows the number of observations (in millions) and selected percentiles of the combined IAB-TPP distribution of 5-year changes in residualized log earnings (from t to $t+5$) for women and selected years. LS sample. Sources: IAB and TPP.

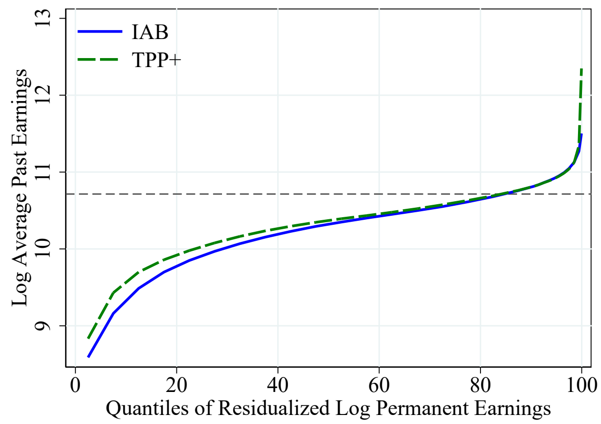
D.3 Earnings Growth by Permanent Earnings

For the heterogeneity analysis by permanent earnings, we use a simple cut-off rule to combine IAB and TPP data. For 1-year growth rates, this cutoff is equal to 45,000 Euro. Hence, for all quantiles of the residualized permanent earnings distribution above this cutoff, we use the conditional growth rate distribution computed from the IAB data. Above this cutoff, we use the corresponding conditional statistics from the reweighted TPP data. There are two reasons for the choice of 45,000 Euro as the cutoff. First, Figure D.6 shows that both residualized permanent earnings and raw average past earnings converge in the middle of the distribution and are almost identical at the cutoff of 45,000 Euro. Second, we argue it is reasonable to assume that average past earnings below the cutoff are mostly unaffected by the top-coding threshold of 60,000 Euro such that the IAB data is reliable. Figures D.7, D.8 and D.9 show the P90-P10 differential, Kelley Skewness and Excess Crow-Siddiqui kurtosis by permanent earnings quantiles in the IAB and reweighted TPP data.

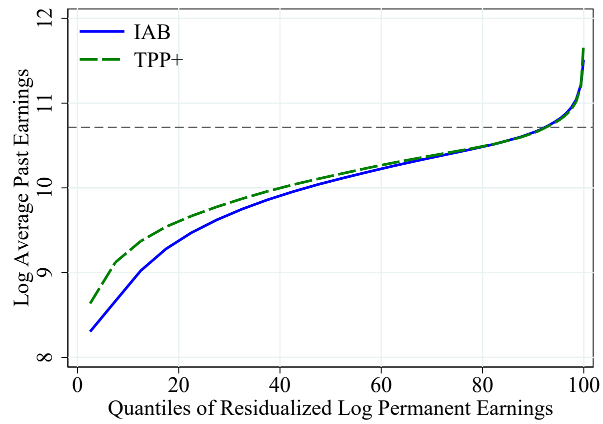
For 5-year earnings changes we proceed analogously, but use a cutoff of 40,000 Euro as jumping into the top-coded range is more likely over a period of five years. Figures D.10, D.11 and D.12 show the P90-P10 differential, Kelley Skewness and Excess Crow-Siddiqui kurtosis by permanent earnings quantiles in the IAB and reweighted TPP data.

Figure D.6: IAB vs. TPP: Permanent Earnings (H Sample)

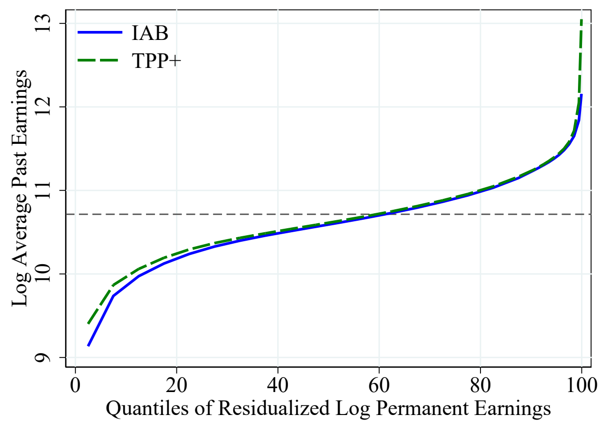
(a) Age Group 25–34: Men



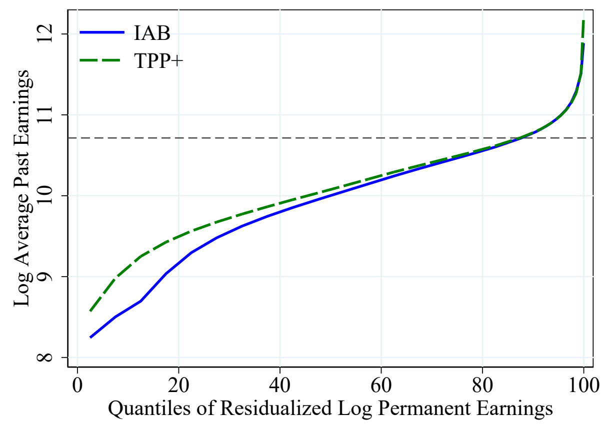
(b) Age Group 25–34: Women



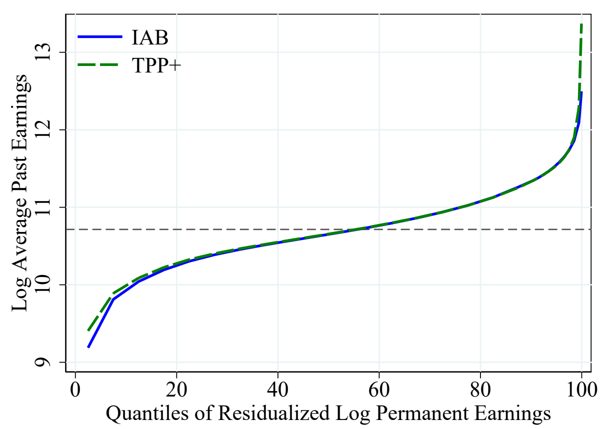
(c) Age Group 35–44: Men



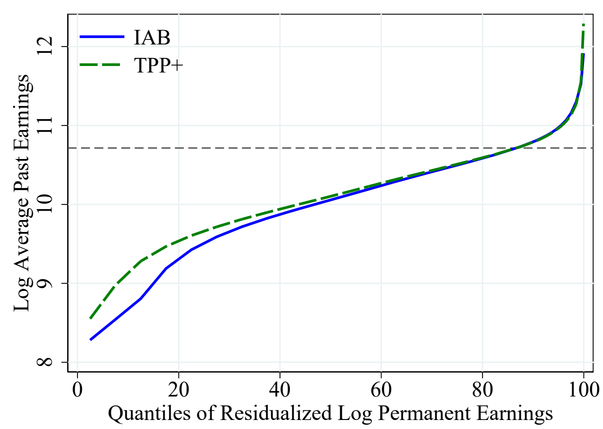
(d) Age Group 35–44: Women



(e) Age Group 45–55: Men



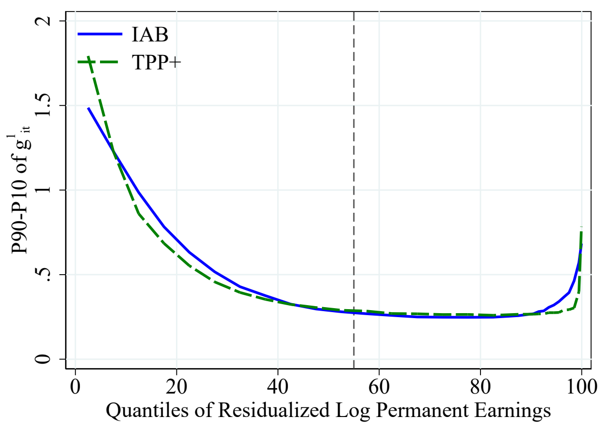
(f) Age Group 45–55: Women



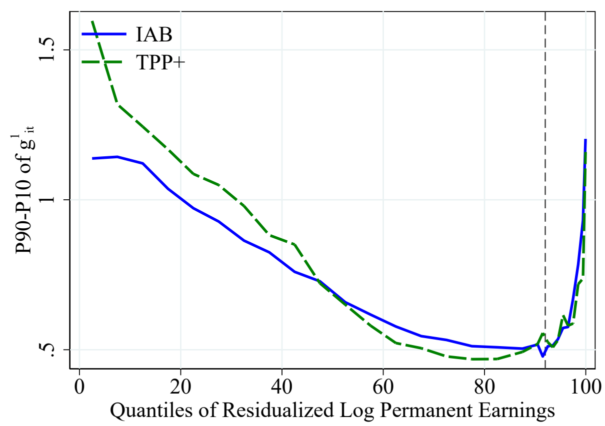
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 45,000 Euro, i.e. the point where the lines in Figure 9 in the main text switch from IAB to TPP data. Source: IAB and TPP.

Figure D.7: IAB vs. TPP: 1-Year Log Earnings Changes by Permanent Earnings, Age Group 25–34

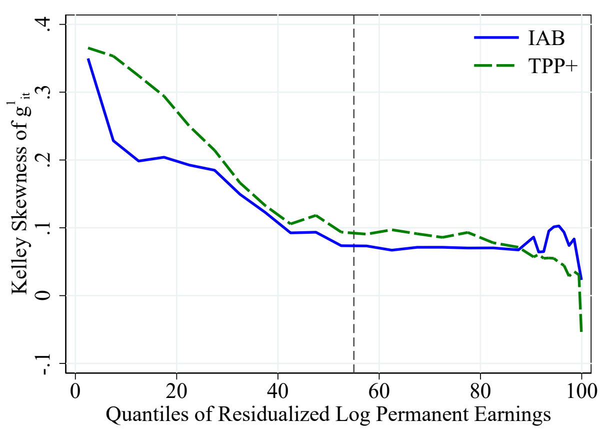
(a) P90-P10: Men



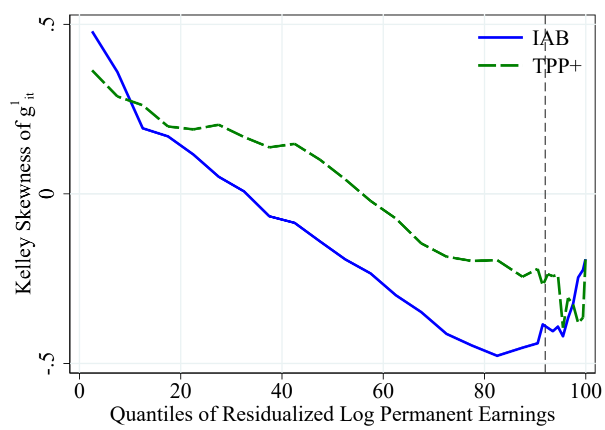
(b) P90-P10: Women



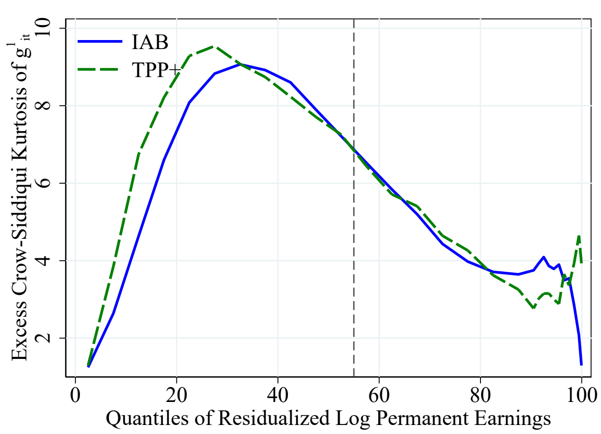
(c) Kelley Skewness: Men



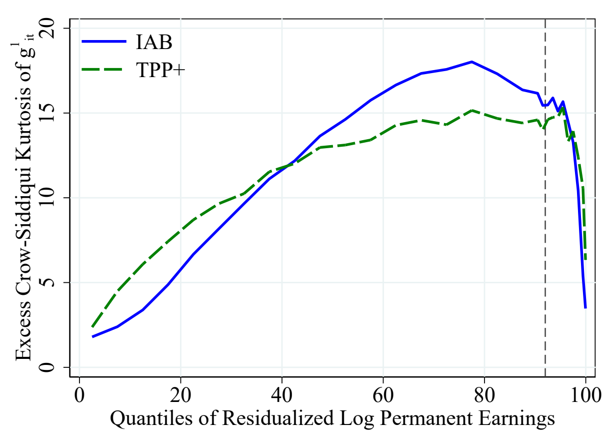
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



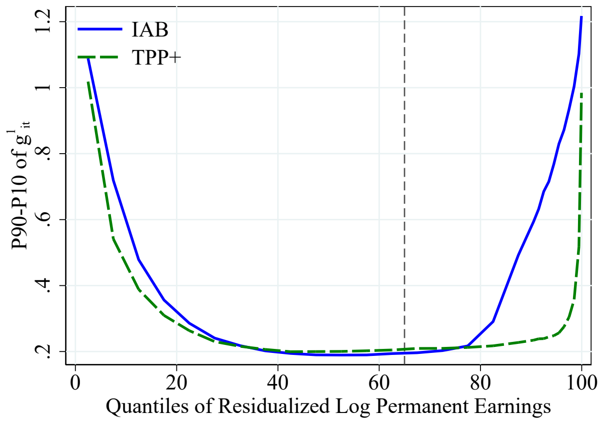
(f) Excess Crow-Siddiqui Kurtosis: Women



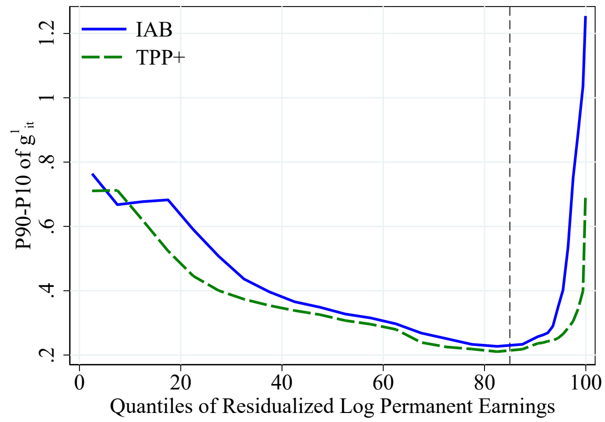
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 45,000 Euro, i.e. the point where the lines in Figure 9 in the main text switch from IAB to TPP data. Source: IAB and TPP.

Figure D.8: IAB vs. TPP: 1-Year Log Earnings Changes by Permanent Earnings, Age Group 35–44

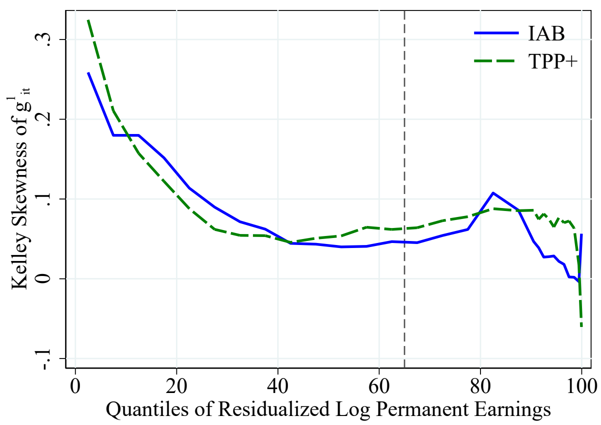
(a) P90-P10: Men



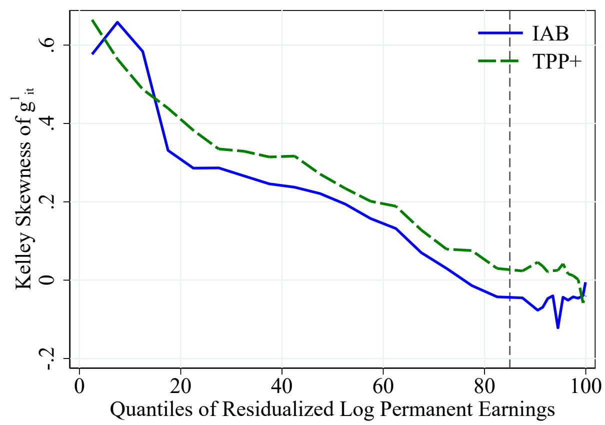
(b) P90-P10: Women



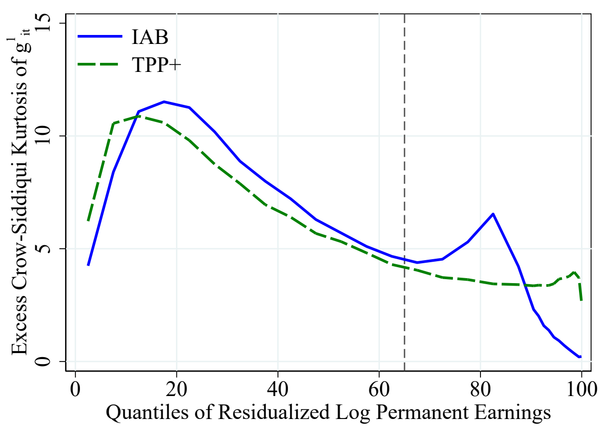
(c) Kelley Skewness: Men



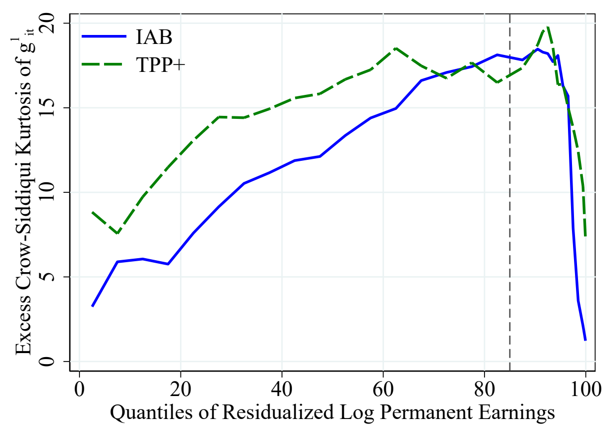
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



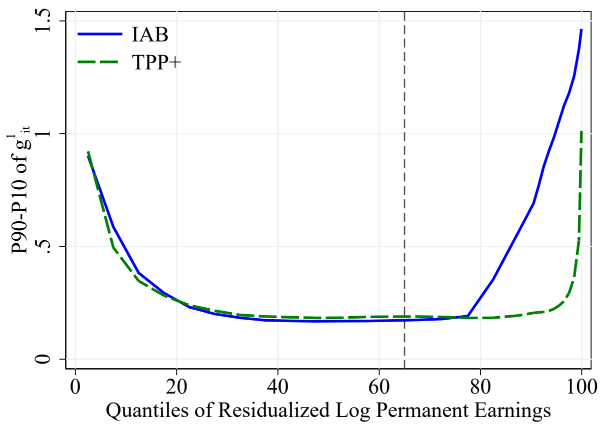
(f) Excess Crow-Siddiqui Kurtosis: Women



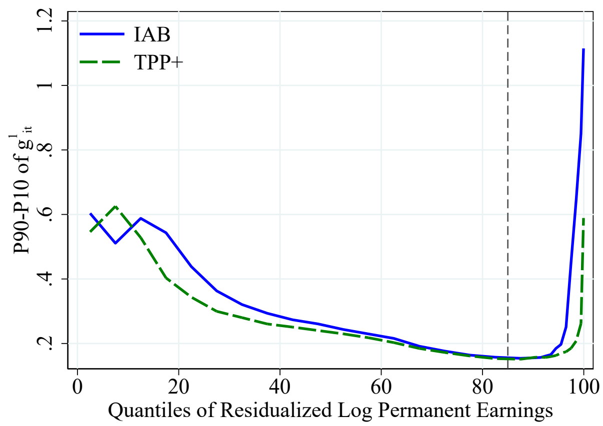
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 45,000 Euro, i.e. the point where the lines in Figure 9 in the main text switch from IAB to TPP data. Source: IAB and TPP.

FIGURE D.9: IAB VS. TPP: 1-YEAR LOG EARNINGS CHANGES BY PERMANENT EARNINGS, AGE GROUP 45–55

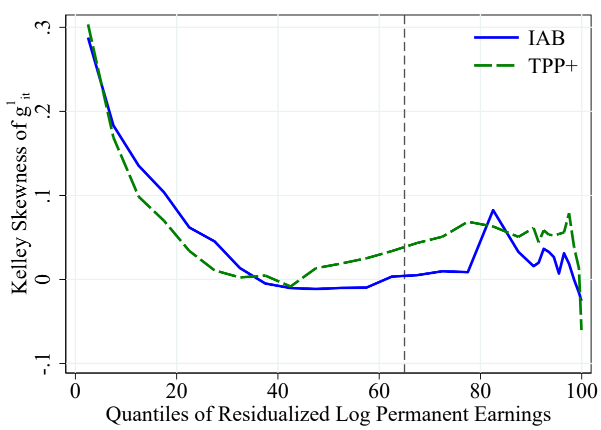
(A) P90-P10: Men



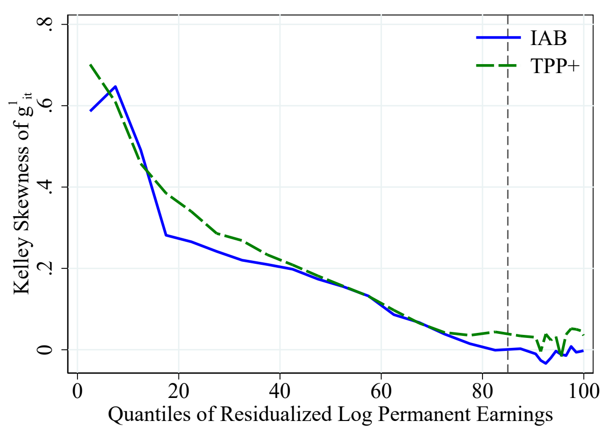
(B) P90-P10: Women



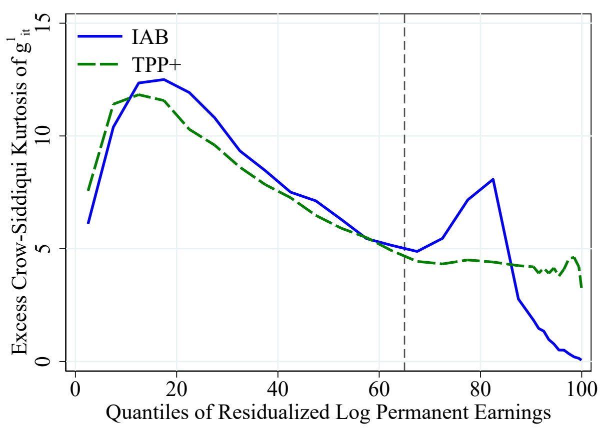
(C) Kelley Skewness: Men



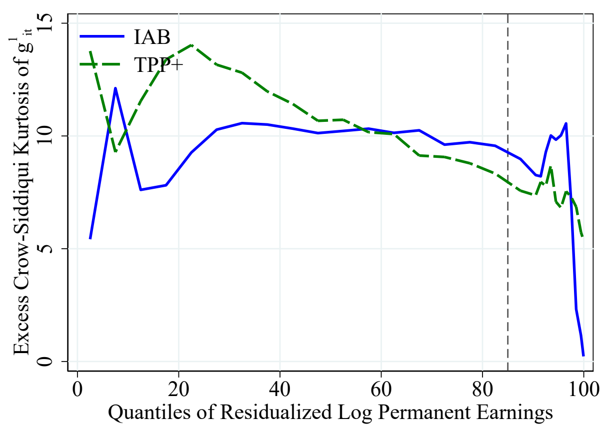
(D) Kelley Skewness: Women



(E) Excess Crow-Siddiqui Kurtosis: Men



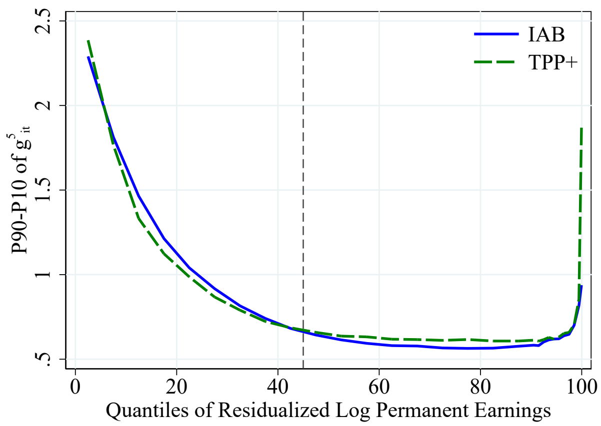
(F) Excess Crow-Siddiqui Kurtosis: Women



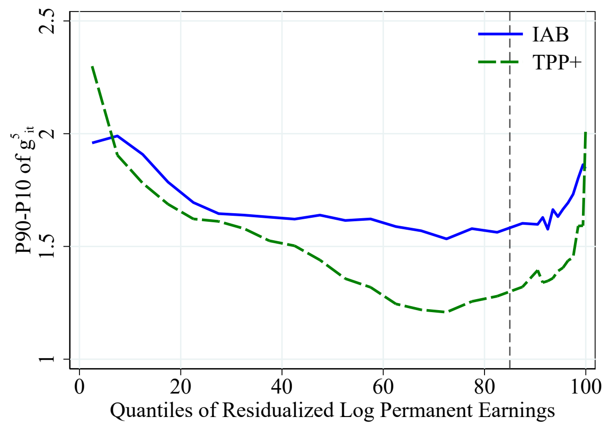
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 45,000 Euro, i.e. the point where the lines in Figure 9 in the main text switch from IAB to TPP data. Source: IAB and TPP.

Figure D.10: IAB vs. TPP: 5-Year Log Earnings Changes by Permanent Earnings, Age Group 25–34

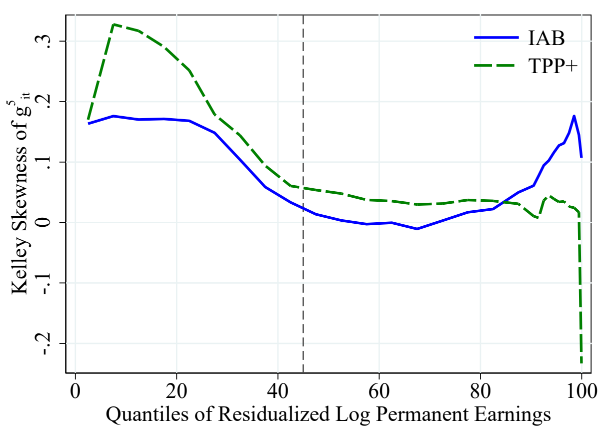
(a) P90-P10: Men



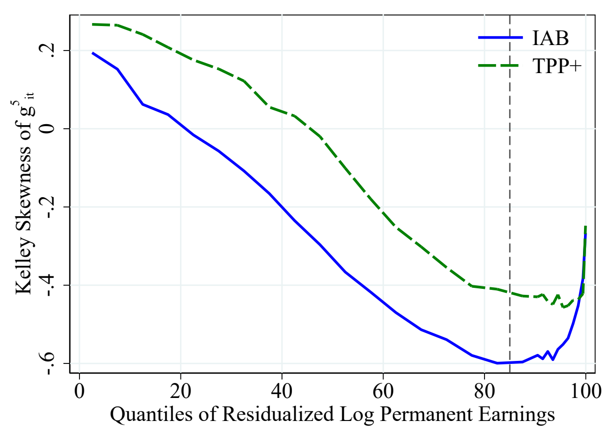
(b) P90-P10: Women



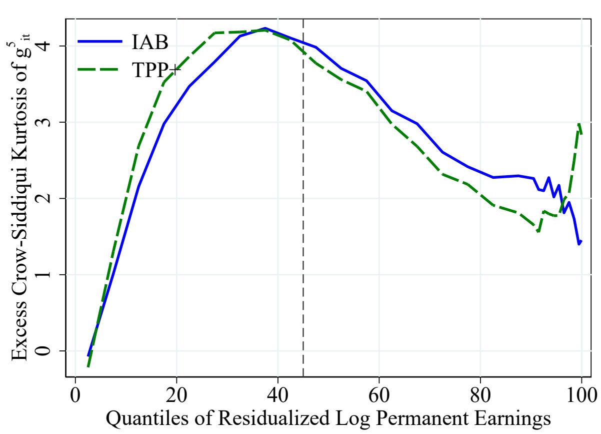
(c) Kelley Skewness: Men



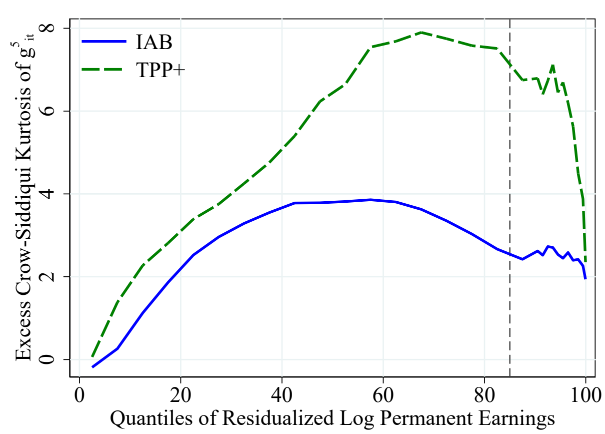
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



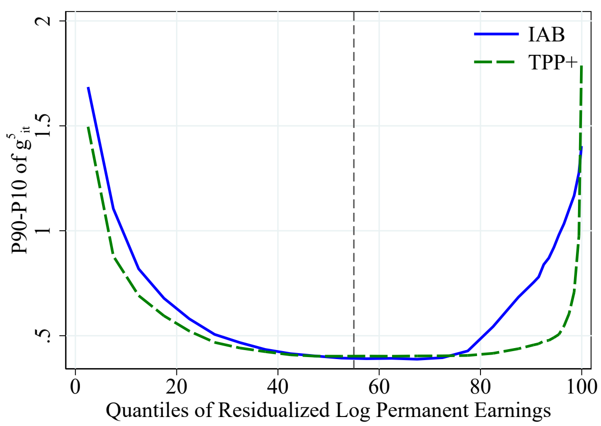
(f) Excess Crow-Siddiqui Kurtosis: Women



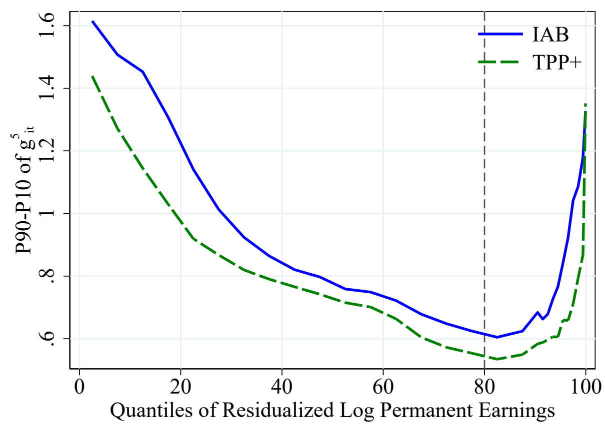
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 40,000 Euro, i.e. the point where the lines in Figure E.28 in the main text switch from IAB to TPP data. Source: IAB and TPP.

Figure D.11: IAB vs. TPP: 5-Year Log Earnings Changes by Permanent Earnings, Age Group 35–44

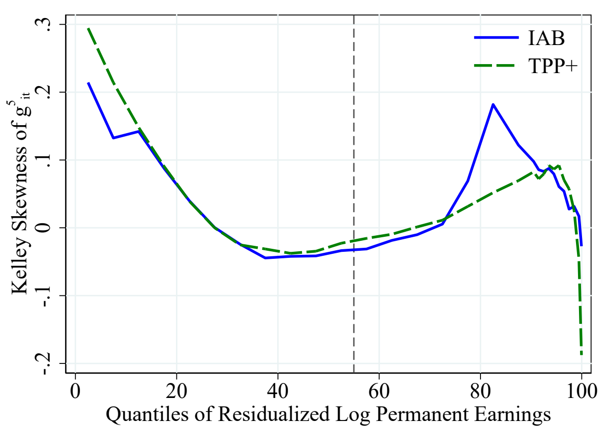
(a) P90-P10: Men



(b) P90-P10: Women

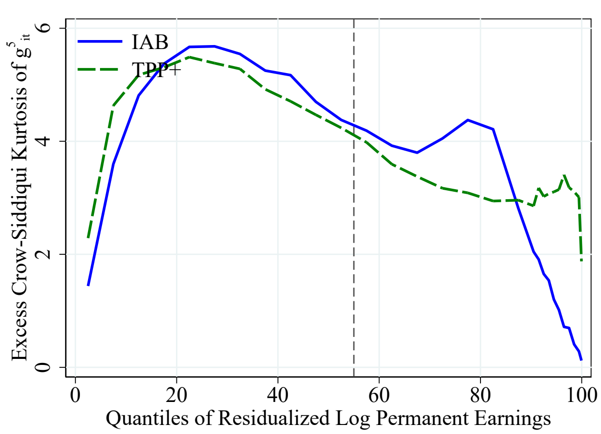


(c) Kelley Skewness: Men

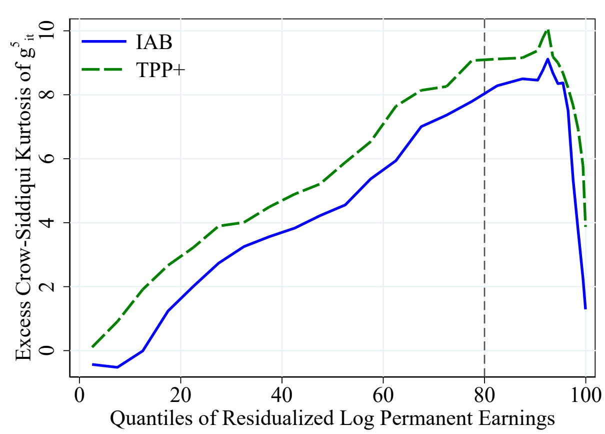


(d) Kelley Skewness: Women

(e) Excess Crow-Siddiqui Kurtosis: Men



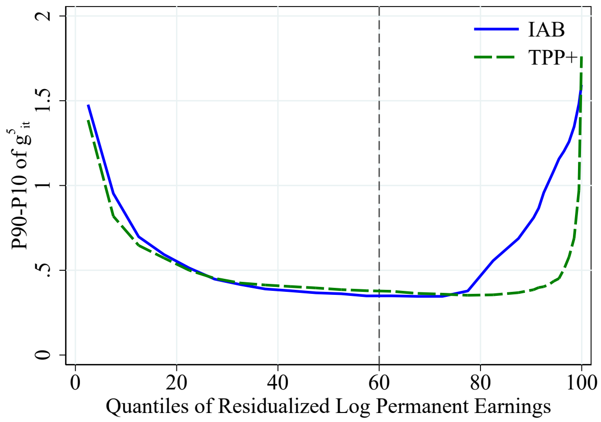
(f) Excess Crow-Siddiqui Kurtosis: Women



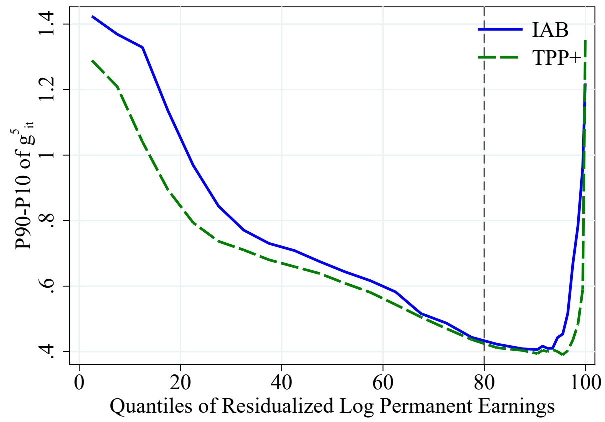
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 40,000 Euro, i.e. the point where the lines in Figure E.28 in the main text switch from IAB to TPP data. Source: IAB and TPP.

Figure D.12: IAB vs. TPP: 5-Year Log Earnings Changes by Permanent Earnings, Age Group 45–55

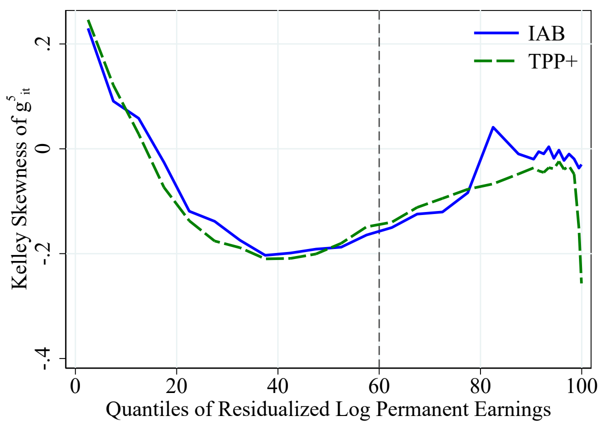
(a) P90-P10: Men



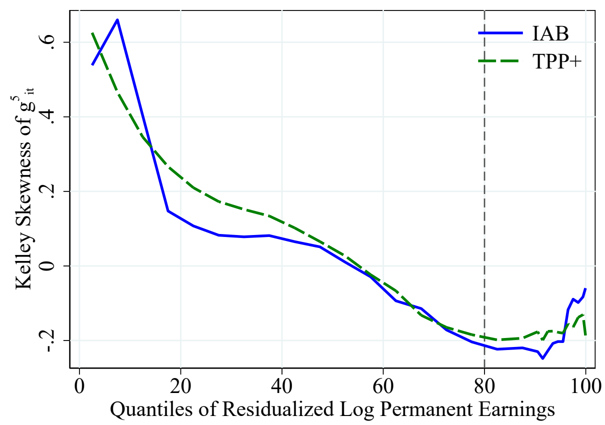
(b) P90-P10: Women



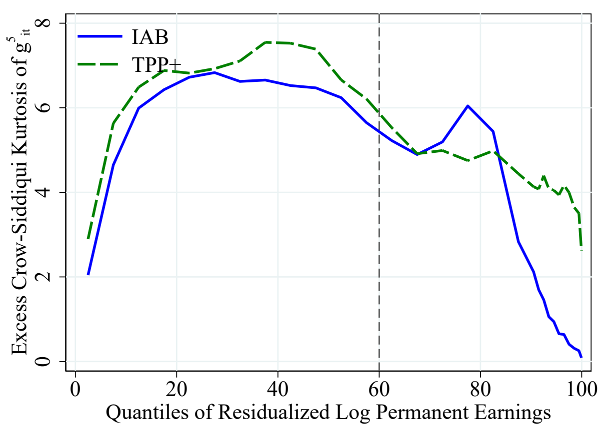
(c) Kelley Skewness: Men



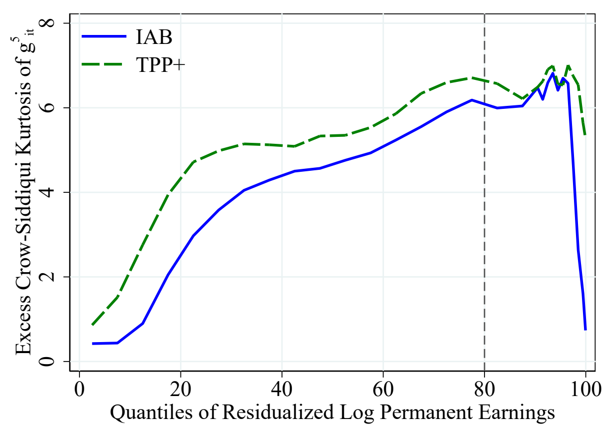
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



(f) Excess Crow-Siddiqui Kurtosis: Women



Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 40,000 Euro, i.e. the point where the lines in Figure E.28 in the main text switch from IAB to TPP data. Source: IAB and TPP.

D.4 Combined IAB-TPP Data in Total Income Analysis (Section 4)

For the analysis of total income, we use the reweighted TPP data. Recall that the distribution of earnings in the subsample of social-security workers in the reweighted data matches the earnings distribution of the combined IAB-TPP data in the earnings analysis (see Appendix D.1). Note that the reweighting procedure does not distort the distribution of non-labor income as only workers who were not obliged to file a tax return are assigned a weight larger than one. The key point is that voluntary filers must not have annual non-labor income above 410 Euro.

The total income analysis sample additionally includes non-social-security workers (e.g. civil servants) and taxpayers who do not receive labor income (self-employed, business owners, landlords). Table D.11 shows how we arrive at the analysis sample starting from the unweighted TPP data (columns 1 and 4). Columns 2 and 5 show the reweighted TPP data before imposing the minimum income threshold of 2,300 Euro and columns 3 and 6 refer to the analysis sample used in Section 4 (see Table 2). In particular, Panel E shows that 1.1% of men and 1.8% of women have negative total income in 2008. While those observations are excluded from the analysis sample, there are still observations with above-threshold total income but negative non-labor income.

In Tables D.13 and D.14 we show pairwise correlations between the different income components. As expected, labor income is negatively correlated with business and self-employment income, and all income components are positively correlated with total income. The surprisingly low correlation of labor and total income is due to the presence of outliers, i.e. entrepreneurs (mostly business owners) who have no labor income but business and hence total income of more than 1 million Euro (up to 25 million Euro).

Table D.12 shows selected percentiles of the earnings distribution in the combined IAB-TPP data (CS sample). As mentioned above, percentiles below 60,000 Euro (P75 and below) are practically identical in the IAB-TPP and IAB data, while higher percentiles are closer to the TPP data.

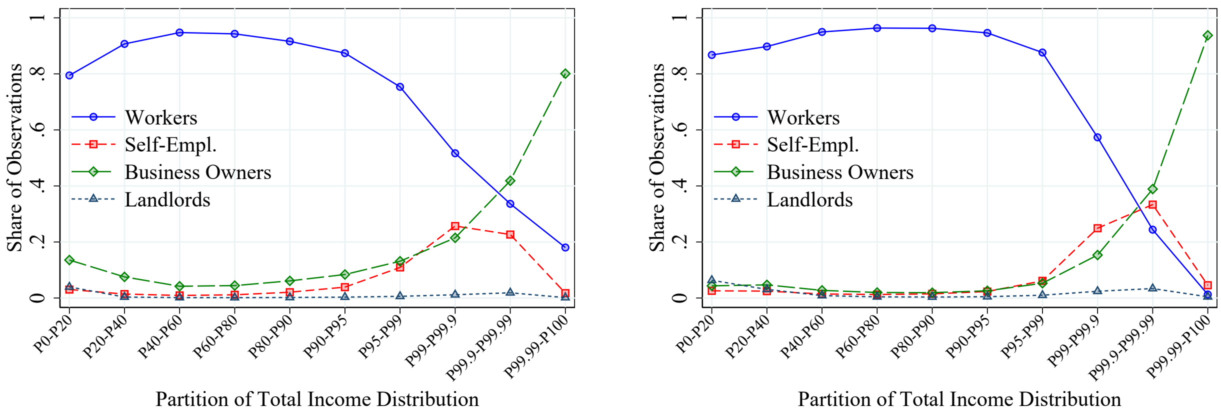
Table D.11: Summary Statistics for Total Income Data

	Men			Women		
	TPP	IAB-TPP	IAB-TPP (analysis)	TPP	IAB-TPP	IAB-TPP (analysis)
	(1)	(2)	(3)	(4)	(5)	(6)
Observations (in mill.)	11.584	14.907	14.667	8.986	12.701	12.351
A. Income Distribution						
Mean	49,323	44,845	45,810	28,406	25,099	26,163
P50	38,664	36,213	36,620	24,227	21,080	21,698
P90	83,998	77,595	78,113	50,888	47,978	48,393
P99.9	786,435	690,267	696,521	312,805	267,797	270,828
P99.99	3,313,004	2,910,567	2,919,253	1,154,808	898,215	931,065
B. Share of Total Income						
Labor	0.813	0.841	0.836	0.886	0.908	0.894
Non-Labor	0.187	0.159	0.164	0.114	0.092	0.106
Self-Empl.	0.063	0.054	0.054	0.052	0.043	0.043
Business	0.123	0.104	0.109	0.053	0.041	0.053
Rental	0.001	0.000	0.001	0.009	0.008	0.011
Capital*	0.016	0.014	0.013	0.006	0.006	0.005
C. Main Income Source						
Workers	0.835	0.869	0.882	0.866	0.897	0.918
Entrepreneurs	0.165	0.131	0.118	0.134	0.103	0.082
Self-Employed	0.034	0.027	0.026	0.037	0.027	0.024
Business Owners	0.115	0.091	0.082	0.060	0.046	0.036
Landlords	0.015	0.012	0.010	0.037	0.030	0.021
D. Non-Zero Income						
Labor	0.851	0.884	0.895	0.884	0.918	0.934
Non-Labor	0.364	0.311	0.300	0.272	0.229	0.207
Self-Empl.	0.064	0.053	0.052	0.065	0.053	0.049
Business	0.206	0.175	0.165	0.107	0.093	0.080
Rental	0.170	0.147	0.144	0.134	0.111	0.102
Capital*	0.123	0.103	0.103	0.055	0.044	0.042
E. Negative Income (if $\neq 0$)						
Total	0.012	0.011	0.000	0.020	0.018	0.000
Non-Labor	0.101	0.098	0.088	0.074	0.071	0.051
Self-Empl.	0.007	0.006	0.005	0.009	0.009	0.006
Business	0.043	0.043	0.034	0.027	0.029	0.019
Rental	0.084	0.076	0.074	0.051	0.043	0.035
Capital*	0.007	0.006	0.006	0.003	0.003	0.002

Notes: This table shows descriptive statistics for the full TPP and IAB-TPP data by gender for the year 2008. The data includes all workers independent of their social-security status and individuals with non-labor income. Columns 1 and 4 refer to the raw TPP data (earnings not reweighted using IAB data). Columns 2 and 5 refer to the combined IAB-TPP data where observations with earnings are reweighted using IAB data (see Appendix D). Columns 3 and 6 refer to the analysis sample of the combined IAB-TPP where we require total income to be above the minimum income threshold of 2,300 Euro (2018 prices). Panel A shows the mean and selected percentiles of the total income distribution in 2018 Euro (excluding capital income). Panel B shows the share of each income source in total income (excluding capital income). Hence, the capital share is not part of the non-labor income share. Panel C reports the share of observations whose most important source of income is labor, non-labor (and sub-categories of non-labor income). Panel D shows the share of observations with non-zero income from different sources. Panel E shows the share of observations with negative income from different sources provided that the person has non-zero income from this source.

Figure D.13: Main Income Sources Across the Income Distribution

(a) Men (b) Women



Notes: This figure shows the share of observations classified as workers, self-employed, business owners and landlords in different parts of the total income distribution in the combined IAB-TPP data for the year 2008. A person is classified as a worker if her labor income is positive and (pairwise) larger than incomes from other sources. Source: IAB and TPP.

Table D.12: Total Income Percentiles in the Combined IAB-TPP Data

Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
Men												
2001	15.275	44,979	7,245	12,568	25,608	37,924	51,695	72,973	93,531	176,591	519,341	1,984,431
2002	15.029	44,878	6,815	11,925	25,230	37,982	52,064	73,634	93,941	176,495	502,857	1,773,707
2003	14.763	44,900	6,382	11,432	24,872	38,065	52,541	74,445	94,863	176,950	492,898	1,724,135
2004	14.635	44,969	6,013	10,774	24,204	37,717	52,500	75,128	96,410	184,374	542,735	2,044,026
2005	14.451	45,293	5,925	10,546	23,655	37,332	52,487	75,872	98,073	193,222	600,038	2,375,527
2006	14.509	45,707	5,923	10,529	23,184	36,989	52,734	76,852	99,989	201,573	636,386	2,501,665
2007	14.654	46,088	6,017	10,749	23,081	36,700	52,622	77,480	101,779	209,085	683,272	2,837,472
2008	14.667	46,438	6,077	10,898	22,996	36,620	52,804	78,113	103,133	215,338	696,522	2,919,252
2009	14.393	45,611	5,873	10,415	22,633	36,345	52,289	77,753	102,367	208,794	638,749	2,299,090
2010	14.528	46,041	5,968	10,553	22,386	36,368	52,918	78,483	103,556	211,896	656,857	2,528,981
2011	14.697	46,602	6,105	10,898	22,646	36,338	53,245	79,487	105,586	218,373	685,103	2,545,453
2012	14.739	46,710	5,978	10,746	22,710	36,391	53,517	80,013	105,916	218,586	676,732	2,571,310
2013	14.774	46,763	5,935	10,586	22,681	36,457	53,583	80,124	106,038	219,216	685,103	2,743,865
2014	14.853	47,267	5,883	10,445	22,698	36,695	54,179	81,205	107,470	223,726	705,679	2,766,737
2015	14.931	48,079	6,036	10,787	22,984	37,035	54,936	82,514	109,496	228,932	738,823	2,900,730
2016	14.955	48,756	6,237	11,160	23,472	37,435	55,525	83,507	111,041	233,868	761,639	2,950,836
Women												
2001	12.389	27,165	3,959	4,801	12,037	23,199	35,910	47,924	56,978	89,493	219,459	690,697
2002	12.365	27,177	3,946	4,774	12,032	23,251	36,131	48,424	57,756	90,323	216,999	643,630
2003	12.193	26,518	3,918	4,939	11,850	23,205	36,310	48,812	58,150	91,108	217,184	634,949
2004	12.176	26,978	3,890	4,937	11,117	22,686	35,983	48,619	58,446	93,163	228,673	746,738
2005	12.116	26,173	3,861	4,933	10,860	22,385	35,740	48,588	58,541	95,198	240,643	826,045
2006	12.152	26,057	3,879	4,975	10,661	22,061	35,413	48,463	58,635	96,886	251,420	885,390
2007	12.307	26,724	3,976	5,044	10,656	21,798	35,021	48,286	58,936	99,744	263,245	931,383
2008	12.351	26,904	4,011	5,072	10,791	21,698	35,022	48,393	59,147	101,178	270,828	931,066
2009	12.370	26,924	4,034	5,077	10,832	21,852	35,425	48,985	59,626	100,961	262,240	847,725
2010	12.469	27,321	4,036	5,152	10,999	21,892	35,487	49,484	60,318	102,453	271,484	960,557
2011	12.603	26,571	4,063	5,209	11,237	21,901	35,414	49,399	60,573	103,927	276,565	979,455
2012	12.717	26,929	4,092	5,211	11,472	21,987	35,453	49,590	61,010	104,702	280,337	916,708
2013	12.777	27,007	4,174	5,305	11,680	22,231	35,733	49,883	61,411	105,697	282,672	969,343
2014	12.839	27,547	4,222	5,424	12,005	22,600	36,360	50,770	62,705	108,991	293,866	1,051,058
2015	12.902	28,220	4,376	5,689	12,614	23,120	37,006	51,705	64,019	111,466	304,776	1,054,335
2016	12.881	28,961	4,481	5,787	13,195	23,794	37,791	52,596	65,294	114,521	316,734	1,178,313
Population												
2001	27.664	36,145	4,568	7,263	17,701	31,829	44,927	62,387	79,529	144,849	410,730	1,508,540
2002	27.394	36,609	4,513	7,062	17,455	31,776	45,161	62,912	80,102	144,686	398,651	1,381,143
2003	26.956	36,115	4,533	6,657	17,198	31,738	45,465	63,461	80,911	145,553	393,548	1,331,338
2004	26.810	36,049	4,493	6,036	16,534	31,230	45,208	63,737	81,904	150,171	425,888	1,535,035
2005	26.568	36,174	4,462	5,940	16,219	30,766	45,058	64,007	82,858	155,886	466,683	1,777,965
2006	26.661	36,643	4,481	5,953	15,975	30,335	44,984	64,513	84,192	161,680	494,922	1,937,037
2007	26.961	36,483	4,561	6,063	15,918	29,989	44,748	64,767	85,245	167,071	523,091	2,110,520
2008	27.018	36,669	4,569	6,200	15,920	29,845	44,765	65,127	86,028	171,488	536,687	2,156,058
2009	26.763	36,202	4,555	6,115	15,725	29,703	44,681	64,881	85,409	167,327	498,942	1,748,128
2010	26.997	36,535	4,589	6,304	15,758	29,625	45,039	65,579	86,369	169,905	512,422	1,899,162
2011	27.300	37,621	4,628	6,548	16,019	29,600	45,090	66,096	87,563	174,503	532,829	1,969,698
2012	27.457	37,603	4,622	6,657	16,047	29,617	45,157	66,567	88,123	174,675	524,838	1,902,555
2013	27.551	37,129	4,711	6,718	16,145	29,699	45,325	66,733	88,336	175,007	529,973	1,983,229
2014	27.692	37,648	4,743	6,858	16,335	29,988	45,892	67,687	89,658	178,723	545,531	2,056,196
2015	27.833	39,026	4,943	7,293	16,932	30,356	46,554	68,856	91,266	183,152	566,431	2,207,568
2016	27.837	39,960	5,073	7,716	17,463	30,972	47,181	69,869	92,599	186,681	584,788	2,300,887

Notes: This table shows selected total income percentiles for men, women and in the population in the combined IAB-TPP. Capital income is not included in total or non-labor income. Note that total incomes in the analysis sample must exceed the minimum income threshold of 2,300 Euro (in 2018 prices). CS sample. Source: IAB and TPP.

Table D.13: Correlations Between Income Components – Men

	Total	Labor	Non-Labor	Business	Self-Empl.	Capital	Rental
Total	1.0000	0.3158	0.9432	0.9237	0.1871	0.0297	0.0525
Labor	0.3158	1.0000	-0.0172	-0.0060	-0.0445	0.0237	-0.0300
Non-Labor	0.9432	-0.0172	1.0000	0.9755	0.2128	0.0230	0.0658
Business	0.9237	-0.0060	0.9755	1.0000	0.0017	0.0215	0.0105
Self-Empl.	0.1871	-0.0445	0.2128	0.0017	1.0000	0.0070	-0.0318
Capital	0.0297	0.0237	0.0230	0.0215	0.0070	1.0000	0.0103
Rental	0.0525	-0.0300	0.0658	0.0105	-0.0318	0.0103	1.0000

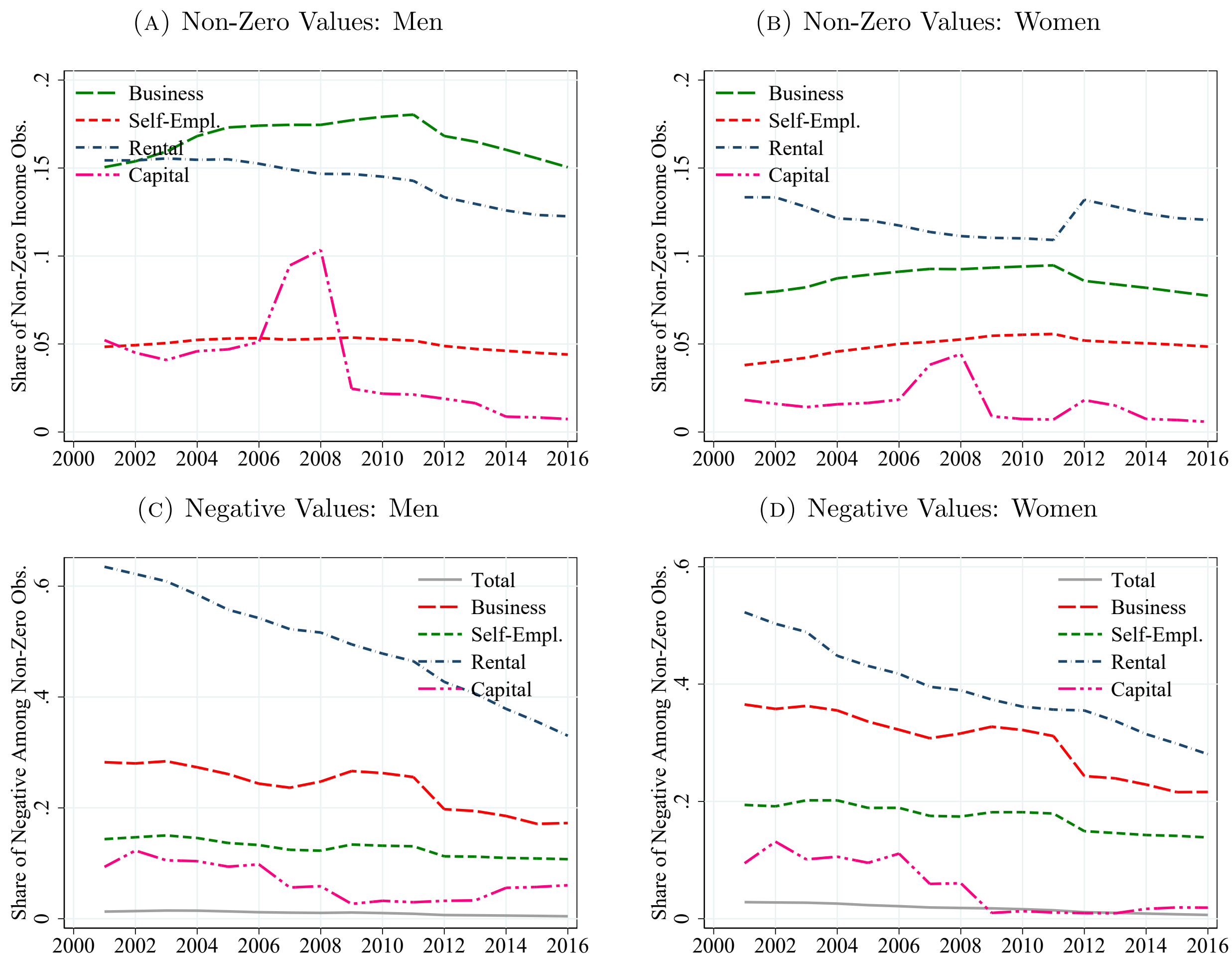
Notes: This table shows correlations between different income components in the combined IAB-TPP analysis sample for men in 2008. Capital income is not included in total or non-labor income. Note that total incomes in the analysis sample must exceed the minimum income threshold of 2,300 Euro (in 2018 prices). CS sample. Source: IAB and TPP.

Table D.14: Correlations Between Income Components – Women

	Total	Labor	Non-Labor	Business	Self-Empl.	Capital	Rental
Total	1.0000	0.3054	0.9394	0.9088	0.1747	0.0801	0.1554
Labor	0.3054	1.0000	-0.0395	-0.0174	-0.0875	0.0403	-0.0220
Non-Labor	0.9394	-0.0395	1.0000	0.9599	0.2148	0.0695	0.1710
Business	0.9088	-0.0174	0.9599	1.0000	0.0004	0.0592	-0.0084
Self-Empl.	0.1747	-0.0875	0.2148	0.0004	1.0000	0.0195	-0.0080
Capital	0.0801	0.0403	0.0695	0.0592	0.0195	1.0000	0.0467
Rental	0.1554	-0.0220	0.1710	-0.0084	-0.0080	0.0467	1.0000

Notes: This table shows correlations between different income components in the combined IAB-TPP analysis sample for women in 2008. CS sample. Source: IAB and TPP.

Figure D.14: Non-Zero and Negative Values for Non-Labor Income



Notes: Panels A and B show the share of total income for different non-labor income components. Panels C and D show the share of observations with non-zero income from these components. Panels E and F show the share of observations out of all non-zero observations with negative income. Total income includes capital income. Source: TPP re-weighted using IAB data.

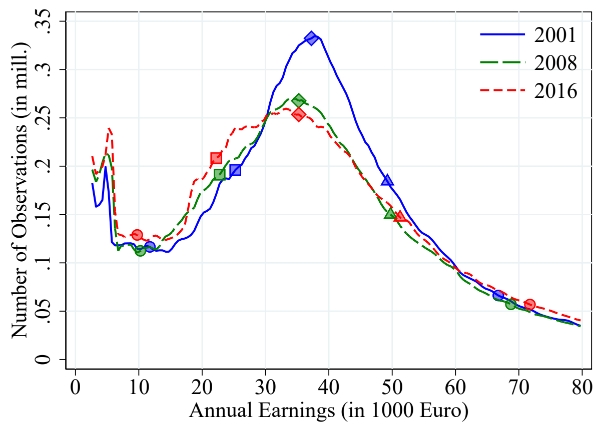
E Core Analysis: Additional Results for Combined IAB-TPP Data 2001–2016

In this Appendix we present additional results for the core analysis in Section 3.

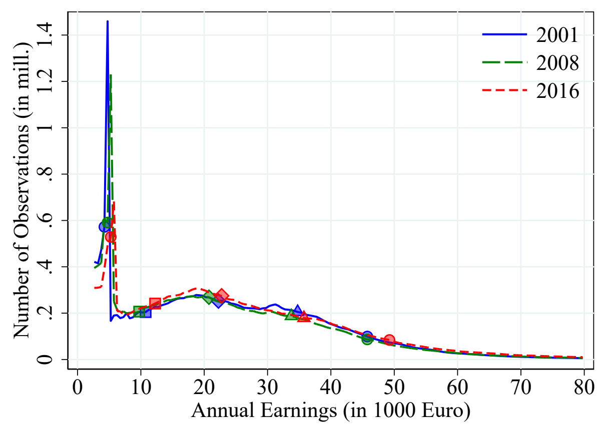
E.1 Additional Results for Earnings Inequality (Section 3.1)

Figure E.1: Earnings Distribution

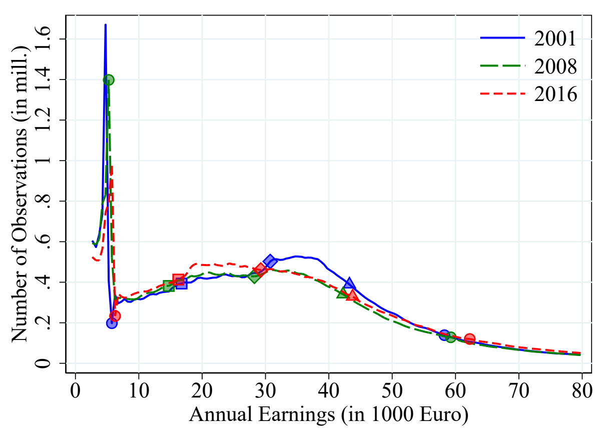
(a) Number of Observations: Men



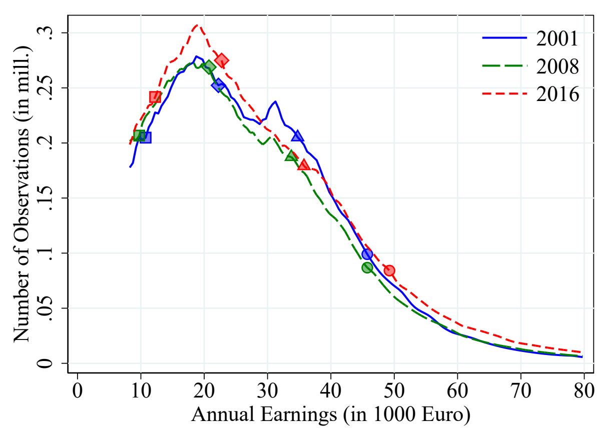
(b) Number of Observations: Women



(c) Number of Observations: Population



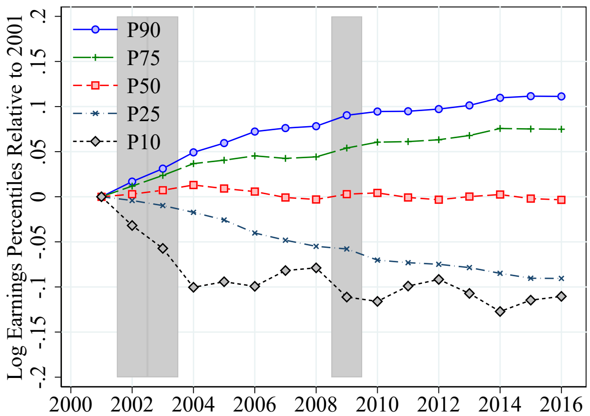
(d) Number of Observations: Women (Zoom)



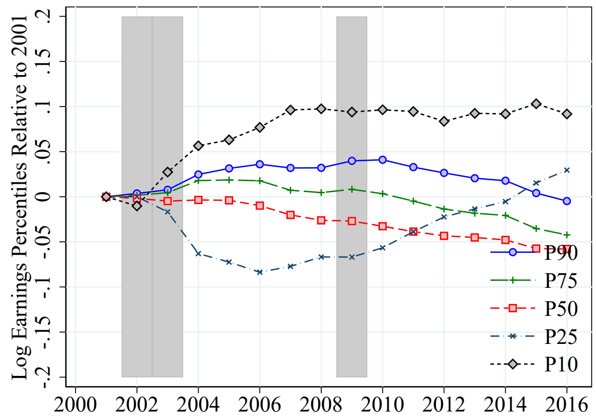
Notes: This figure shows the number of observations per 1,000 Euro earnings bins of real annual earnings for selected years in the combined IAB-TPP data (CS sample) separately for men and women. Panel A and B are depicted as shares in Figure 2. The data is smoothed (by year and gender) using a three-bin moving average for bins above 10,000 Euro. The markers indicate the 10th (circle), 25th (square), 50th (i.e. median; diamond), 75th (triangle) and 90th (circle again) percentiles of the respective distributions.

Figure E.2: Evolution of Residual Log Earnings Percentiles (Controlling for Age)

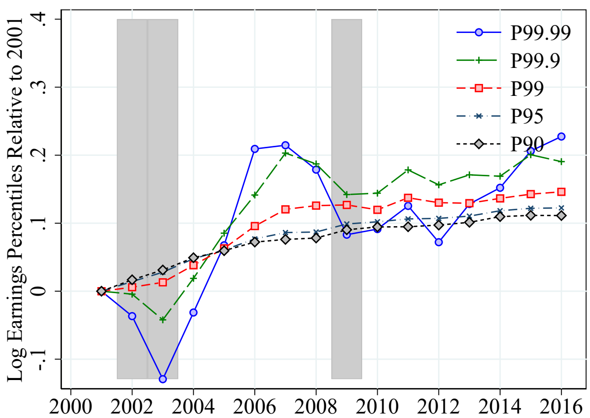
(a) Core Percentiles: Men



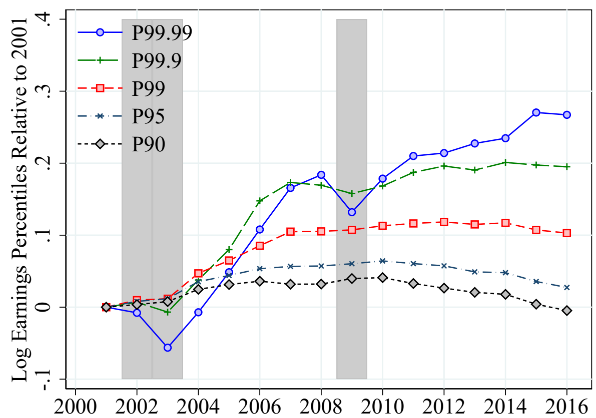
(b) Core Percentiles: Women



(c) Top Percentiles: Men



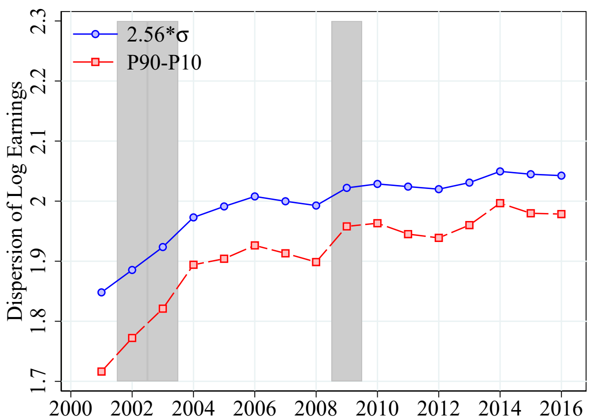
(d) Top Percentiles: Women



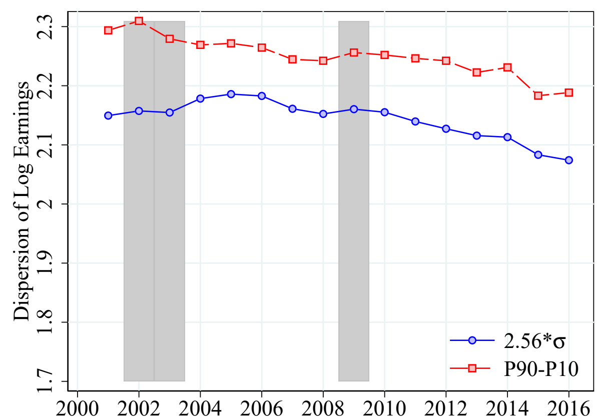
Notes: This figure shows the evolution of residualized log real annual earnings (controlling for age, for unconditioned percentiles, see Figure 3) in the combined IAB-TPP data (CS sample). Shaded areas indicate recessions.

Figure E.3: Evolution of Earnings Inequality: Standard Deviation and Log Percentile Differentials

(a) Men



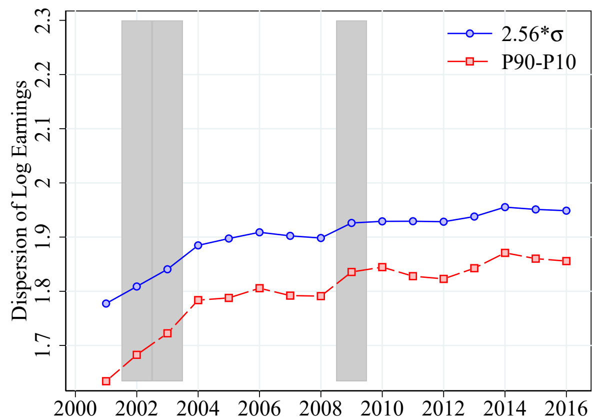
(b) Women



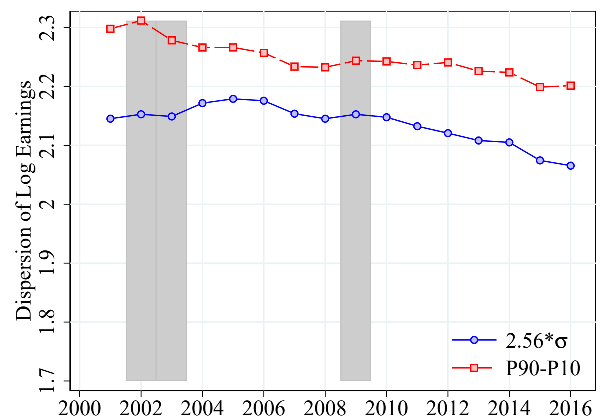
Notes: This figure shows the evolution of different log percentile differentials as well as the (rescaled) standard deviation of the log real annual earnings distribution over time in the combined IAB-TPP data (CS sample) separately for men and women. The standard deviation σ is rescaled as $2.56 * \sigma$ corresponds to P90-P10 differential for a Gaussian distribution. Shaded areas indicate recessions.

Figure E.4: Residual Earnings Inequality (Controlling for Age)

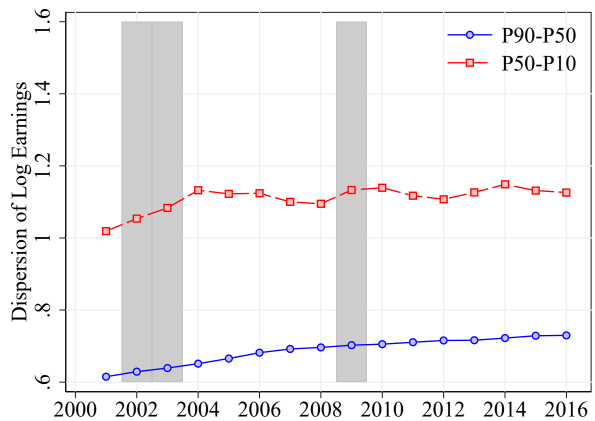
(a) Inequality: Men



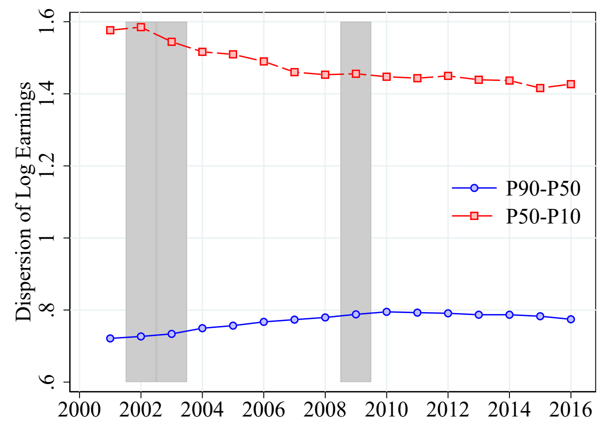
(b) Inequality: Women



(c) Upper and Lower Inequality: Men

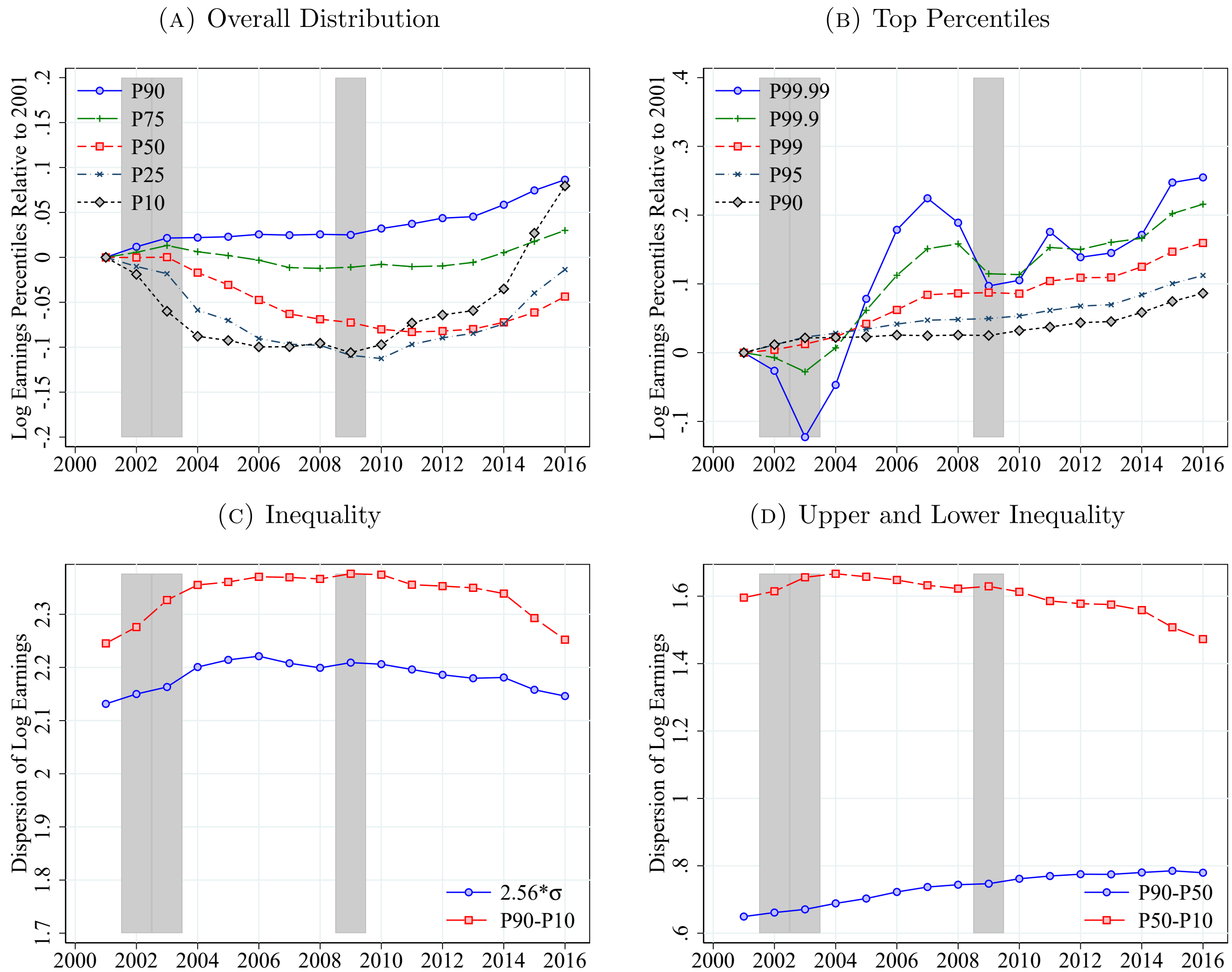


(d) Upper and Lower Inequality: Women



Notes: This figure shows the evolution of residualized log real annual earnings (controlling for age, unconditioned results can be found in Figure 4) in the combined IAB-TPP data (CS sample). Shaded areas indicate recessions.

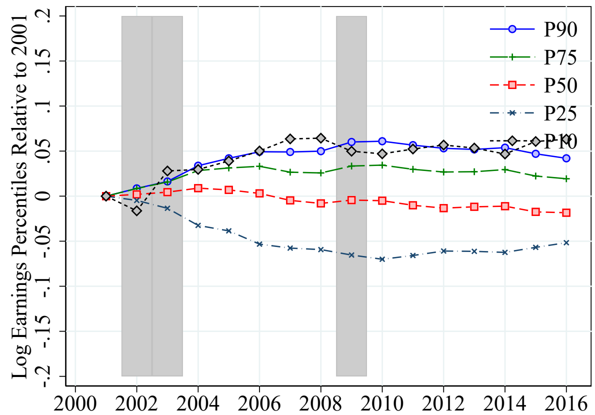
Figure E.5: Evolution of Log Earnings Percentiles in the Population



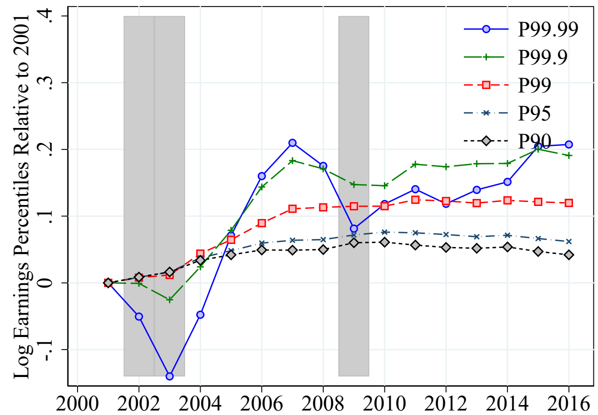
Notes: This figure shows the evolution of selected percentiles of log real annual earnings (relative to 2001) in the combined IAB-TPP data (CS sample) in the join data of men and women. Shaded areas indicate recessions.

Figure E.6: Residual Log Earnings Inequality in the Population (Controlling for Gender and Age)

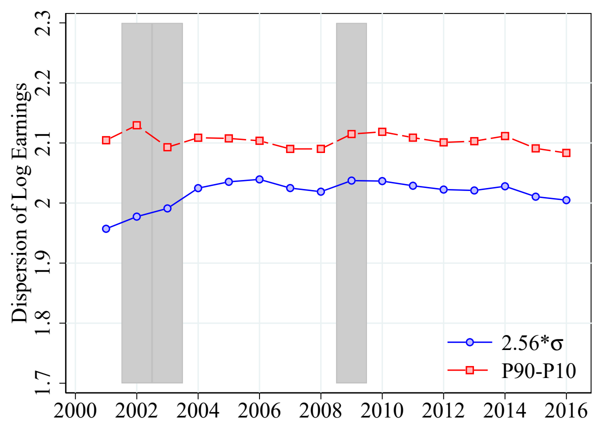
(a) Core Percentiles



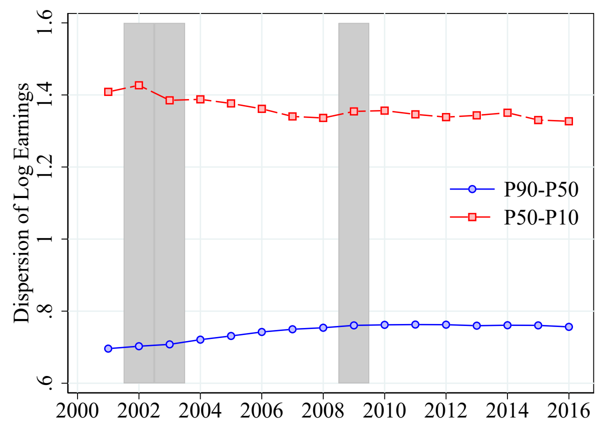
(b) Top Percentiles



(c) Inequality



(d) Upper and Lower Inequality

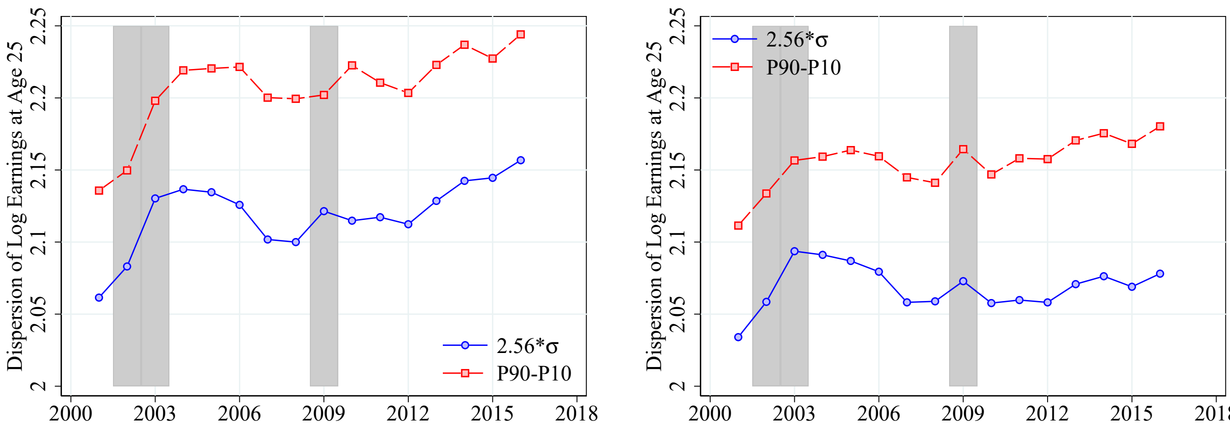


Notes: This figure shows the evolution of residualized log real annual earnings (controlling for gender and age, unconditioned results can be found in Figures 3 and 4.) in the combined IAB-TPP data (CS sample). Shaded areas indicate recessions.

Figure E.7: Initial Income Inequality (at age 25)

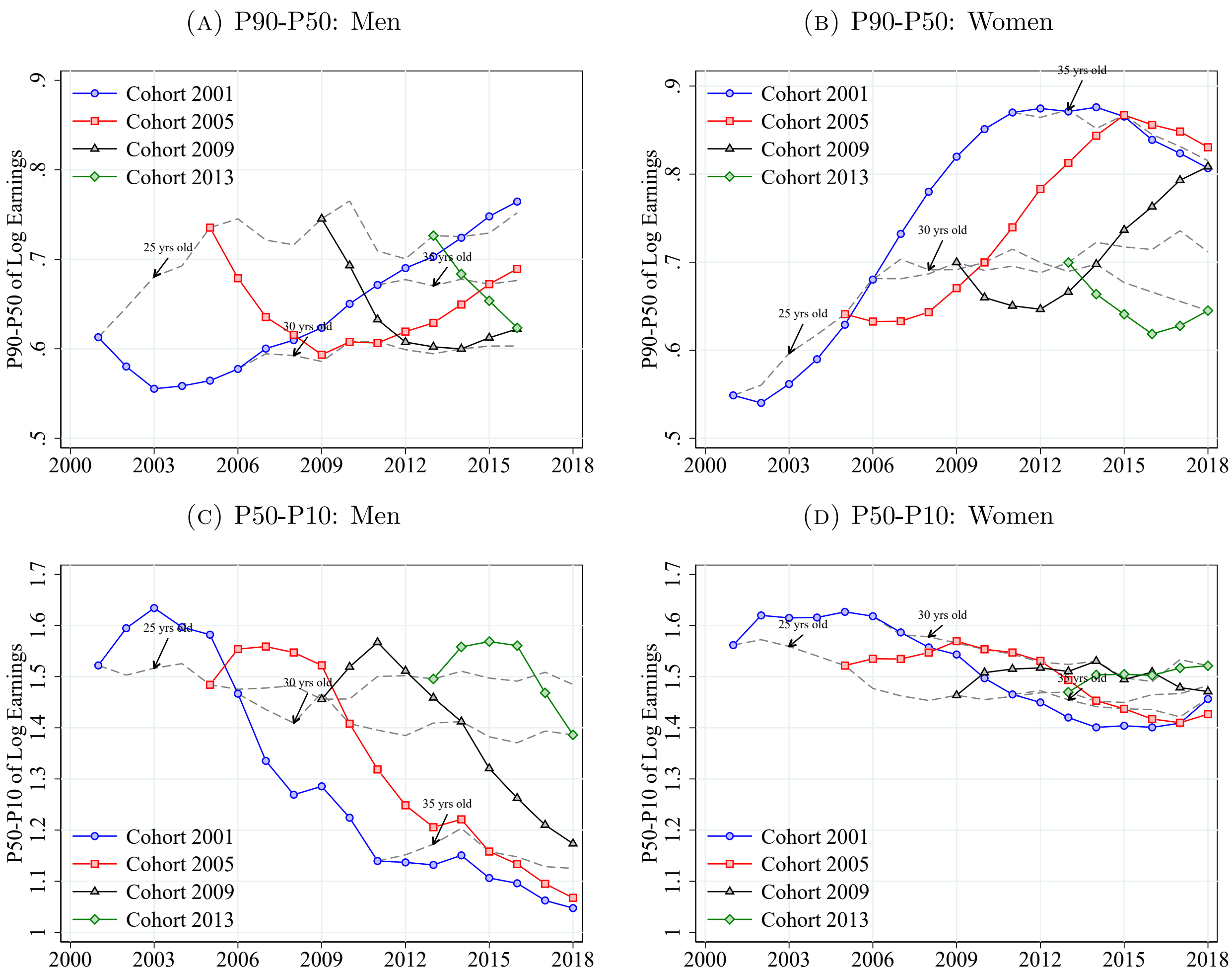
(a) Inequality: Men

(b) Inequality: Women



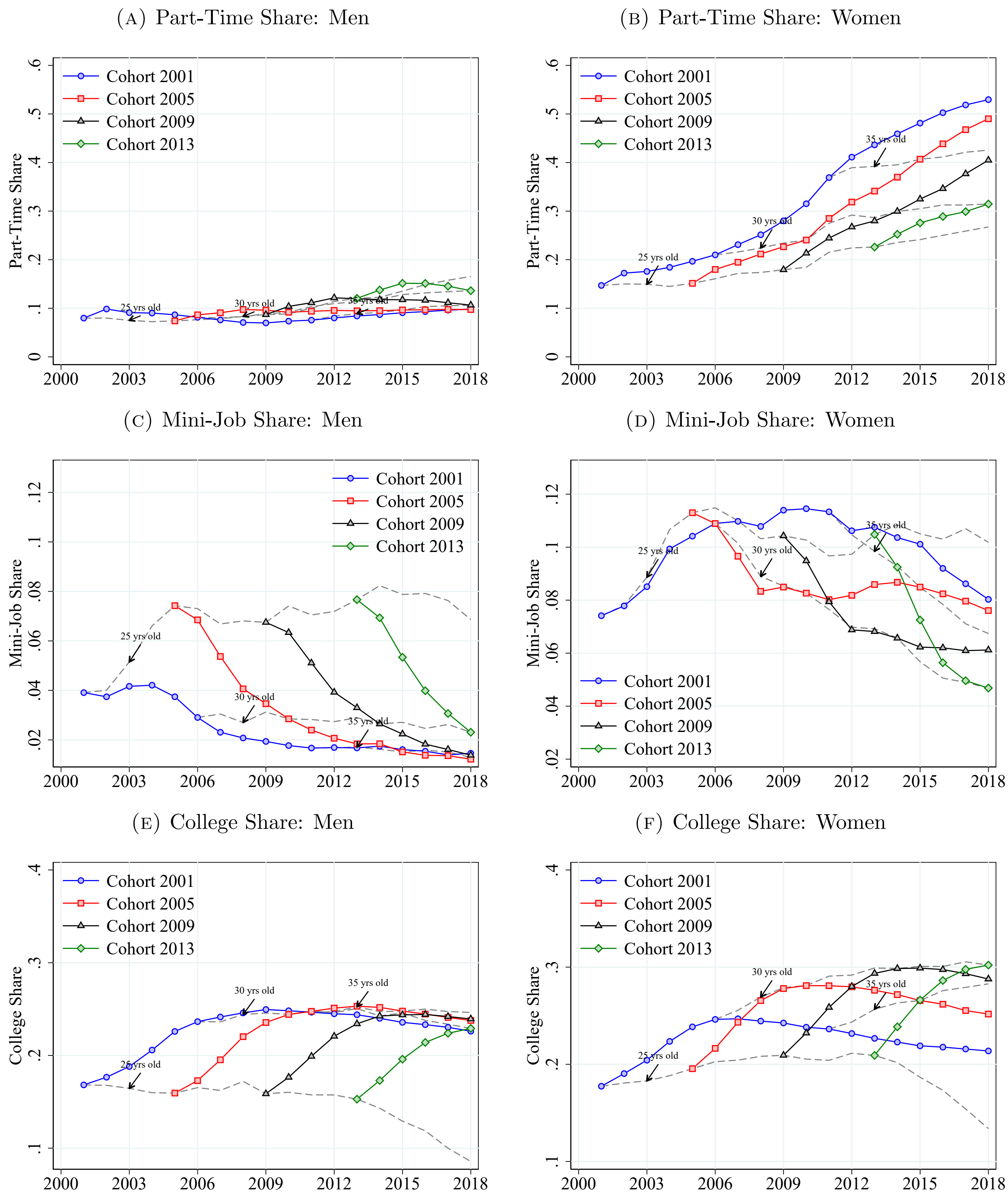
Notes: This figure shows the evolution of the P90-P10 log percentile differential as well as the (rescaled) standard deviation of the log real annual earnings distribution over time in the IAB data (CS sample) separately for men and women at the age of 25 in each year. The standard deviation σ is rescaled as $2.56 * \sigma$ corresponds to P90-P10 differential for a Gaussian distribution. Shaded areas indicate recessions.

Figure E.8: Upper and Lower Earnings Inequality by Cohort



Notes: This figure shows the evolution of the P90-P50 and the P50-P10 differentials of the log real annual earnings distribution over time in the combined IAB-TPP data (CS sample) separately for men and women. As the P90 of men is imputed and the TPP data end in 2016, Panel A also ends in 2016. Grey dashed lines correspond to earnings inequality of 25, 30 and 35 year olds in each year as indicated by arrows. Each colored line corresponds to an individual cohort, where "cohort t" represents the cohort aged 25 in year t.

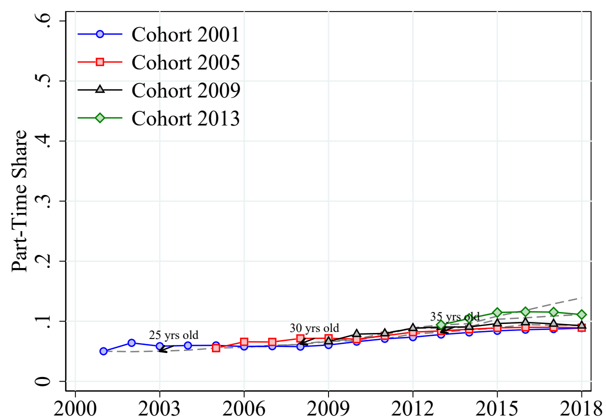
Figure E.9: Employment Levels and Education over the Lifecycle



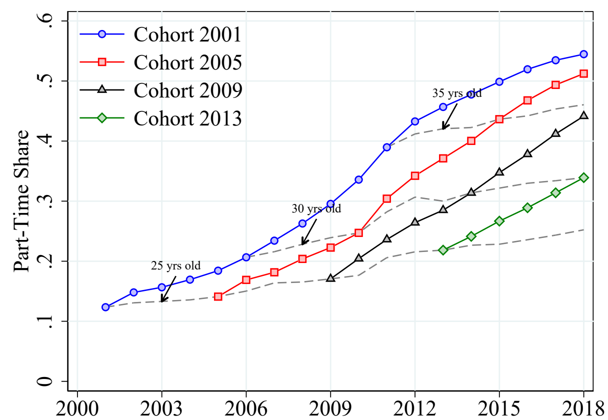
Notes: This figure shows selected emplyoment and education shares in the IAB data (CS sample). Panels A and B show the part-time share over the lifecycle of selected cohorts. Panels C and D show the mini-job share. Panels E and F show the share of college graduates.

Figure E.10: Employment Levels and Average Earnings over the Lifecycle – Non-College Workers

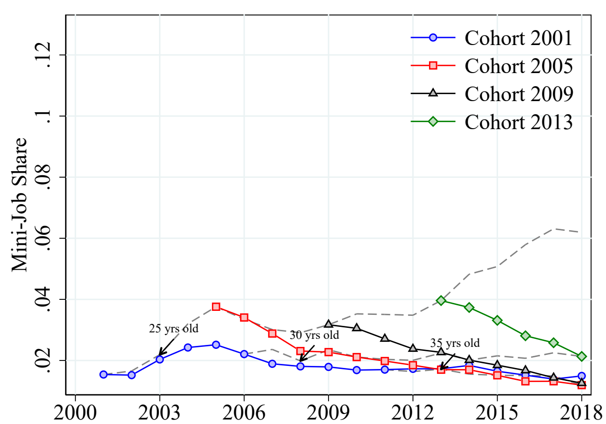
(A) Part-Time Share: Men



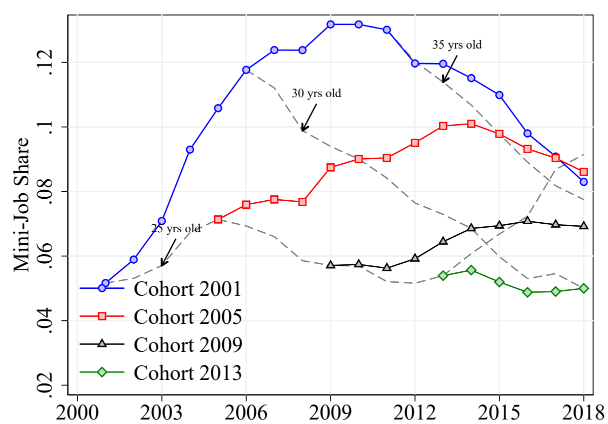
(B) Part-Time Share: Women



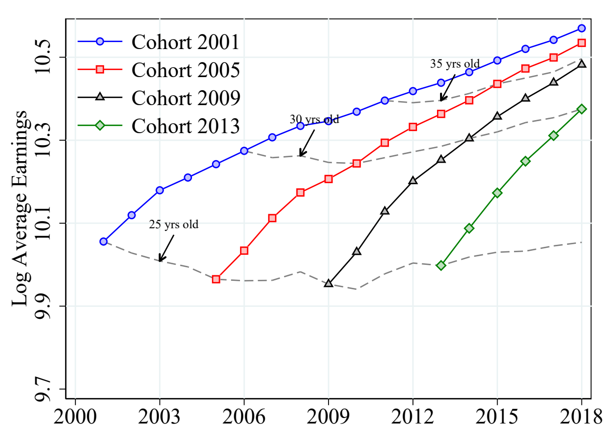
(C) Mini-Job Share: Men



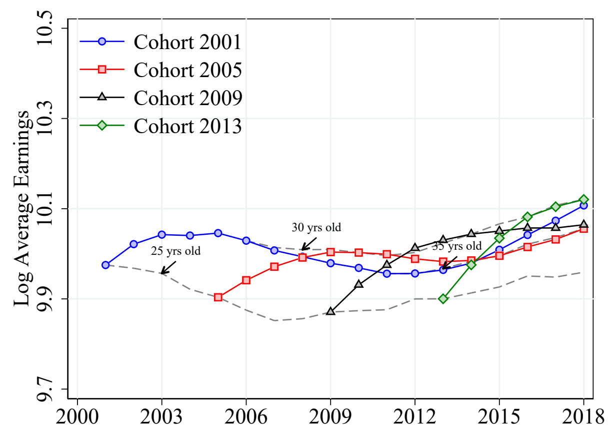
(D) Mini-Job Share: Women



(E) Average Earnings: Men



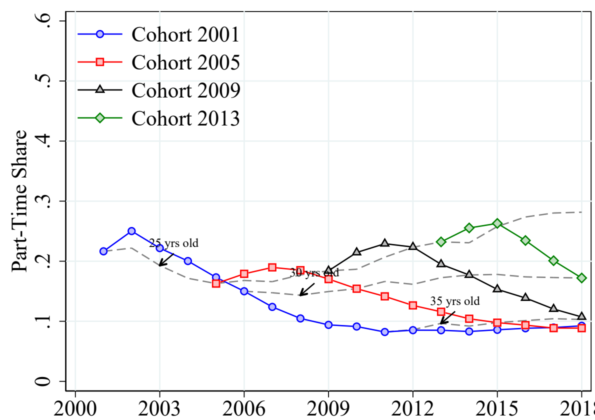
(F) Average Earnings: Women



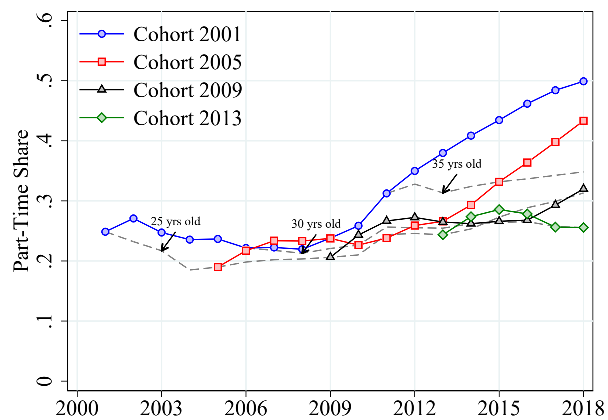
Notes: This figure shows employment levels and average earnings for workers without college degree by cohort in the IAB data (CS sample). Panels A and B show the part-time share over the lifecycle of selected cohorts for non-college workers. Panels C and D show the mini-job share. Panels E and F show average earnings.

Figure E.11: Employment Levels and Average Earnings over the Lifecycle – College Workers

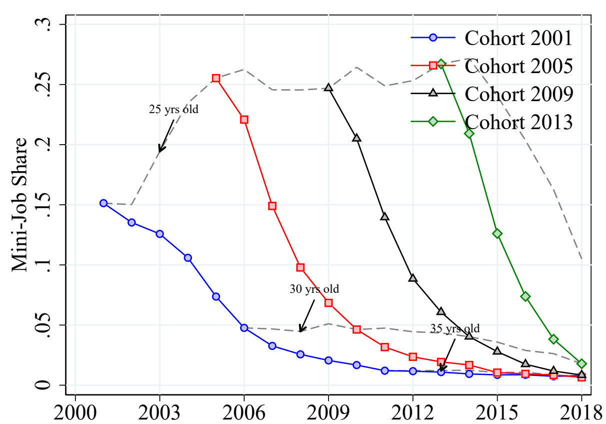
(a) Part-Time Share: Men



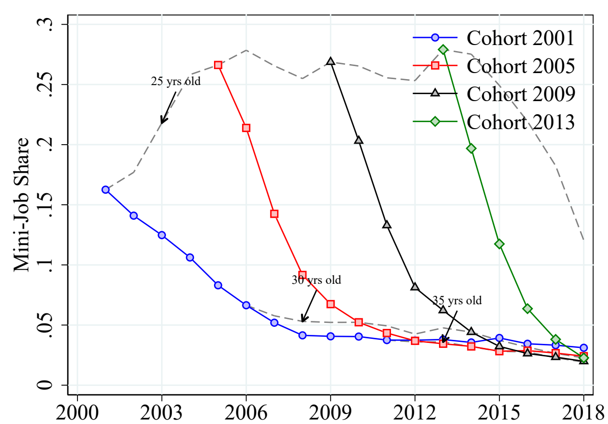
(b) Part-Time Share: Women



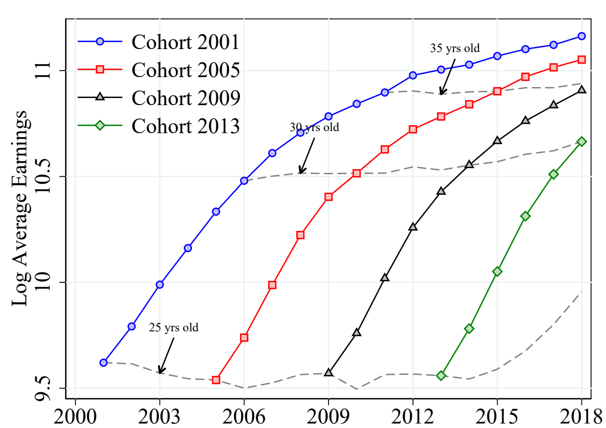
(c) Mini-Job Share: Men



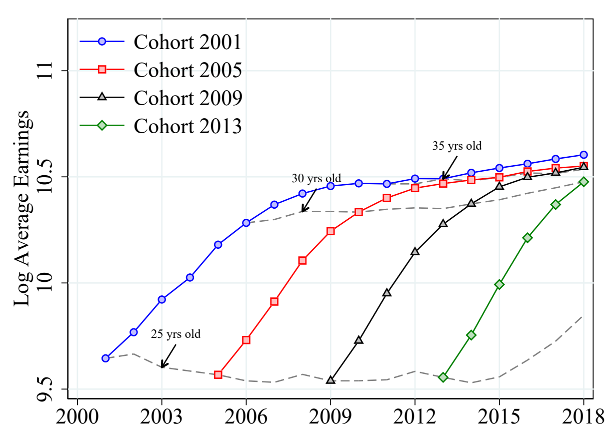
(d) Mini-Job Share: Women



(e) Average Earnings: Men

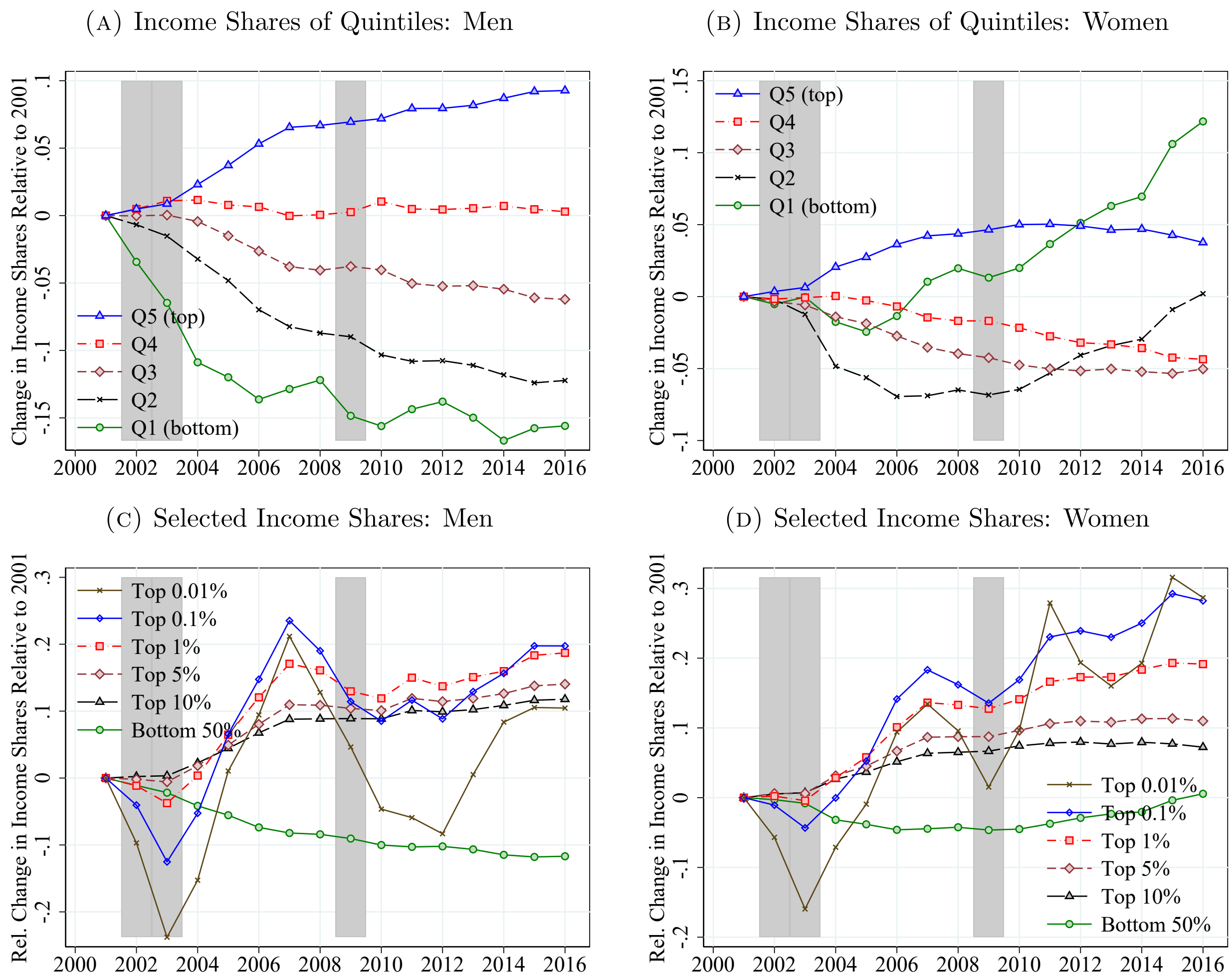


(f) Average Earnings: Women



Notes: This figure shows employment levels and average earnings for workers with college degree by cohort in the IAB data (CS sample). Panels A and B show the part-time share over the lifecycle of selected cohorts for college workers. Panels C and D show the mini-job share. Panels E and F show average earnings.

Figure E.12: Changes in Labor Income Shares Relative to 2001



Notes: This figure shows the evolution of selected income shares of real annual earnings (relative to 2001) in the combined IAB-TPP data (CS sample) separately for men and women. The relative change in income shares of each group relative to 2001 is the differences of the income share in year t minus the income share in 2001 divided by the income share in 2001. Shaded areas indicate recessions. See Tables E.1 and E.2 for more details.

Table E.1: Labor Income Shares – Men

Year	Q1	Q2	Q3	Q4	Q5	Bot50	Bot90	Mid40	Top10	Top5	Top1	Top0.1	Top0.01
2001	6.01	13.69	18.04	22.75	39.52	28.24	75.17	46.93	24.83	15.58	5.55	1.47	0.45
2002	5.80	13.59	18.04	22.86	39.71	27.93	75.11	47.18	24.89	15.56	5.48	1.41	0.41
2003	5.62	13.48	18.05	22.99	39.86	27.62	75.08	47.46	24.92	15.50	5.34	1.29	0.35
2004	5.35	13.24	17.96	23.01	40.43	27.06	74.60	47.54	25.40	15.87	5.57	1.39	0.39
2005	5.29	13.03	17.77	22.93	40.99	26.67	74.06	47.39	25.94	16.35	5.90	1.57	0.46
2006	5.19	12.73	17.56	22.89	41.62	26.15	73.49	47.33	26.51	16.84	6.21	1.69	0.50
2007	5.24	12.56	17.36	22.74	42.11	25.92	72.98	47.06	27.02	17.29	6.49	1.82	0.55
2008	5.27	12.49	17.31	22.76	42.16	25.86	72.97	47.10	27.03	17.28	6.44	1.75	0.51
2009	5.12	12.46	17.36	22.80	42.27	25.68	72.96	47.27	27.04	17.21	6.26	1.64	0.48
2010	5.07	12.27	17.31	22.98	42.36	25.41	72.97	47.56	27.03	17.16	6.21	1.60	0.43
2011	5.15	12.21	17.13	22.86	42.66	25.33	72.66	47.33	27.34	17.44	6.38	1.64	0.43
2012	5.18	12.21	17.09	22.85	42.67	25.35	72.70	47.35	27.30	17.37	6.31	1.60	0.42
2013	5.11	12.17	17.10	22.87	42.75	25.23	72.62	47.40	27.38	17.44	6.38	1.66	0.46
2014	5.01	12.07	17.06	22.91	42.96	25.00	72.47	47.47	27.53	17.55	6.43	1.70	0.49
2015	5.06	11.99	16.94	22.85	43.16	24.91	72.28	47.37	27.72	17.73	6.56	1.76	0.50
2016	5.07	12.01	16.92	22.81	43.18	24.94	72.24	47.30	27.76	17.77	6.58	1.76	0.50

Notes: This table shows the share of earnings that goes to selected parts of the earnings distribution of men in the combined IAB-TPP data (CS sample). Q1 to Q5 refer to the five quintiles where Q1 (Q5) stands for the bottom (top) 20% of the earnings distribution. The quintile shares sum to one. Bot 50, Bot 90 and Mid 40 refer to observations in the bottom 50%, the bottom 90% and between the median and the 90th percentile of the earnings distribution. Top x refers to the top x% of the earnings distribution.

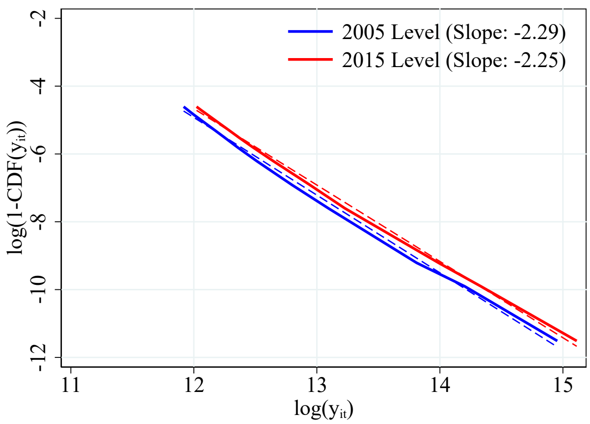
Table E.2: Labor Income Shares – Women

Year	Q1	Q2	Q3	Q4	Q5	Bot50	Bot90	Mid40	Top10	Top5	Top1	Top0.1	Top0.01
2001	3.91	11.00	18.09	25.99	41.02	23.06	75.68	52.63	24.32	14.31	4.24	0.82	0.19
2002	3.89	10.97	18.03	25.94	41.16	22.99	75.55	52.57	24.45	14.39	4.25	0.81	0.18
2003	3.91	10.86	17.99	25.97	41.28	22.87	75.51	52.64	24.49	14.39	4.22	0.79	0.16
2004	3.84	10.46	17.84	26.00	41.86	22.32	75.03	52.71	24.97	14.76	4.36	0.82	0.17
2005	3.81	10.38	17.76	25.92	42.14	22.17	74.79	52.61	25.21	14.96	4.49	0.87	0.19
2006	3.85	10.23	17.60	25.81	42.50	21.99	74.43	52.44	25.57	15.26	4.67	0.94	0.21
2007	3.95	10.24	17.46	25.61	42.75	22.03	74.13	52.11	25.87	15.55	4.82	0.97	0.21
2008	3.98	10.28	17.38	25.55	42.81	22.08	74.10	52.02	25.90	15.56	4.80	0.96	0.21
2009	3.96	10.24	17.33	25.55	42.92	21.99	74.06	52.08	25.94	15.56	4.78	0.93	0.19
2010	3.99	10.29	17.23	25.42	43.07	22.02	73.88	51.86	26.12	15.69	4.84	0.96	0.21
2011	4.05	10.41	17.19	25.27	43.08	22.19	73.78	51.59	26.22	15.83	4.95	1.01	0.24
2012	4.11	10.55	17.16	25.15	43.03	22.39	73.74	51.35	26.26	15.88	4.97	1.02	0.22
2013	4.15	10.62	17.19	25.12	42.92	22.52	73.81	51.29	26.19	15.86	4.97	1.01	0.22
2014	4.18	10.67	17.15	25.06	42.94	22.58	73.75	51.17	26.25	15.93	5.02	1.03	0.22
2015	4.32	10.90	17.13	24.89	42.77	22.97	73.81	50.84	26.19	15.93	5.06	1.06	0.25
2016	4.38	11.02	17.18	24.85	42.56	23.18	73.92	50.74	26.08	15.88	5.05	1.06	0.24

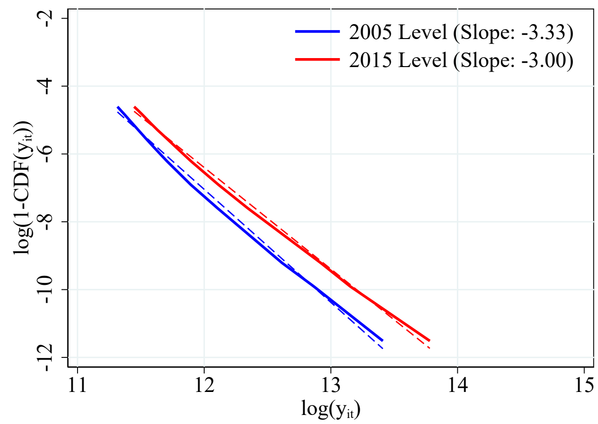
Notes: This table shows the share of earnings that goes to selected parts of the earnings distribution of women in the combined IAB-TPP data (CS sample). Q1 to Q5 refer to the five quintiles where Q1 (Q5) stands for the bottom (top) 20% of the earnings distribution. The quintile shares sum to one. Bot 50, Bot 90 and Mid 40 refer to observations in the bottom 50%, the bottom 90% and between the median and the 90th percentile of the earnings distribution. Top x refers to the top x% of the earnings distribution.

Figure E.13: Top Earnings Inequality: Pareto Tail at Top 1% and Top 5%

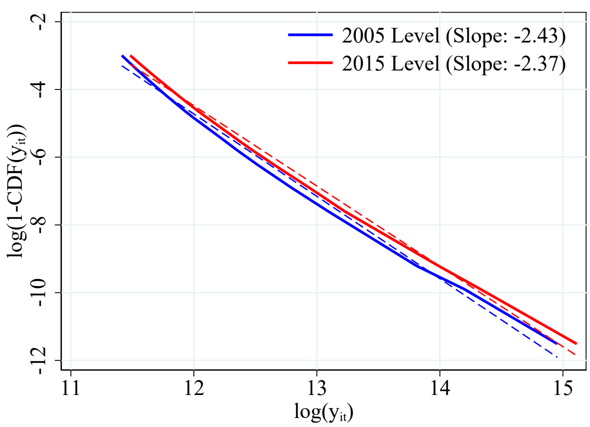
(a) Top 1%: Men



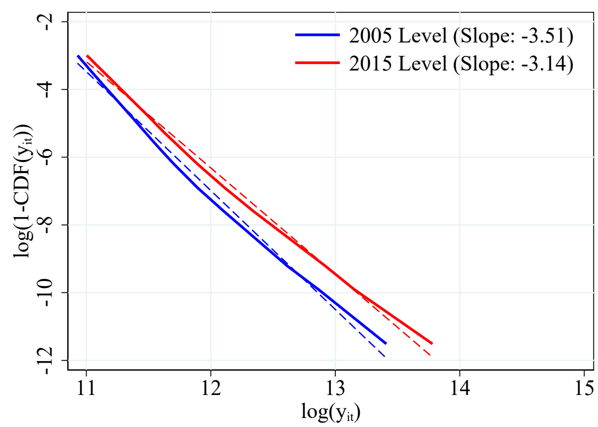
(b) Top 1%: Women



(c) Top 5%: Men



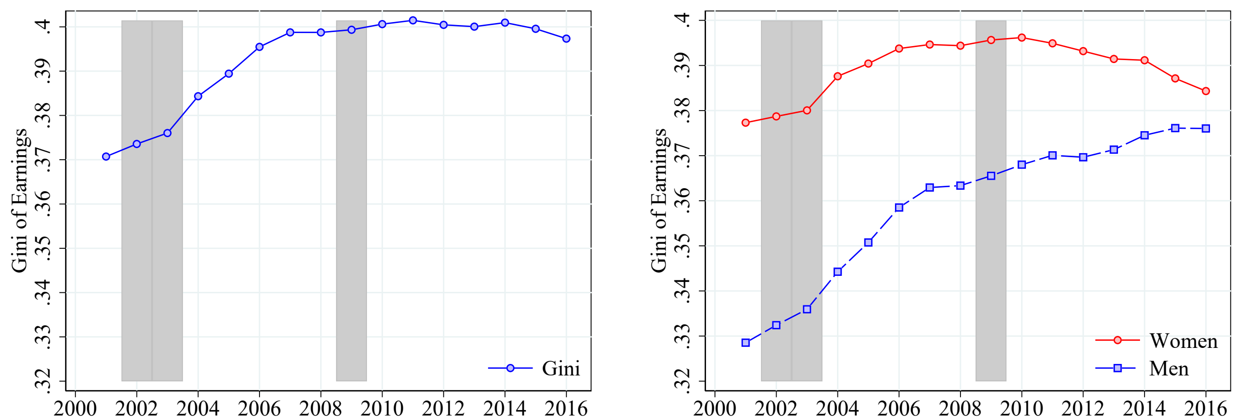
(d) Top 5%: Women



Notes: This figure shows the log of the inverse empirical CDF of log earnings and a fitted linear regression line for observations with earnings in the top 1% and top 5% in the combined IAB-TPP data (CS sample). The absolute value of the slope of the regression line is the Pareto parameter above the respective cutoff.

Figure E.14: Gini Coefficient of Labor Income

(A) Population (B) By Gender



Notes: This figure shows the Gini coefficient of labor income in the population and by gender in the combined IAB-TPP data (CS sample). Shaded areas indicate recessions. .

E.2 Details on Reweighting Analysis (Section 3.1)

To shed light on the different development of the percentiles in more detail and reveal underlying drivers we use the reweighting proposed by DiNardo et al. (1996), henceforth DFL, to analyze the income distribution. We employ the reweighting function keeping different observable characteristics fixed at their 2001 value. For e.g. the year 2015, we can now observe the wage density that would have prevailed if employees were still equipped with their 2001 characteristics and received wages of 2015. The reweighting function is given by:

$$\psi_z(z) = \frac{dF(z|t_z = 2001)}{dF(z|t_z = 2015)}, \tag{E.1}$$

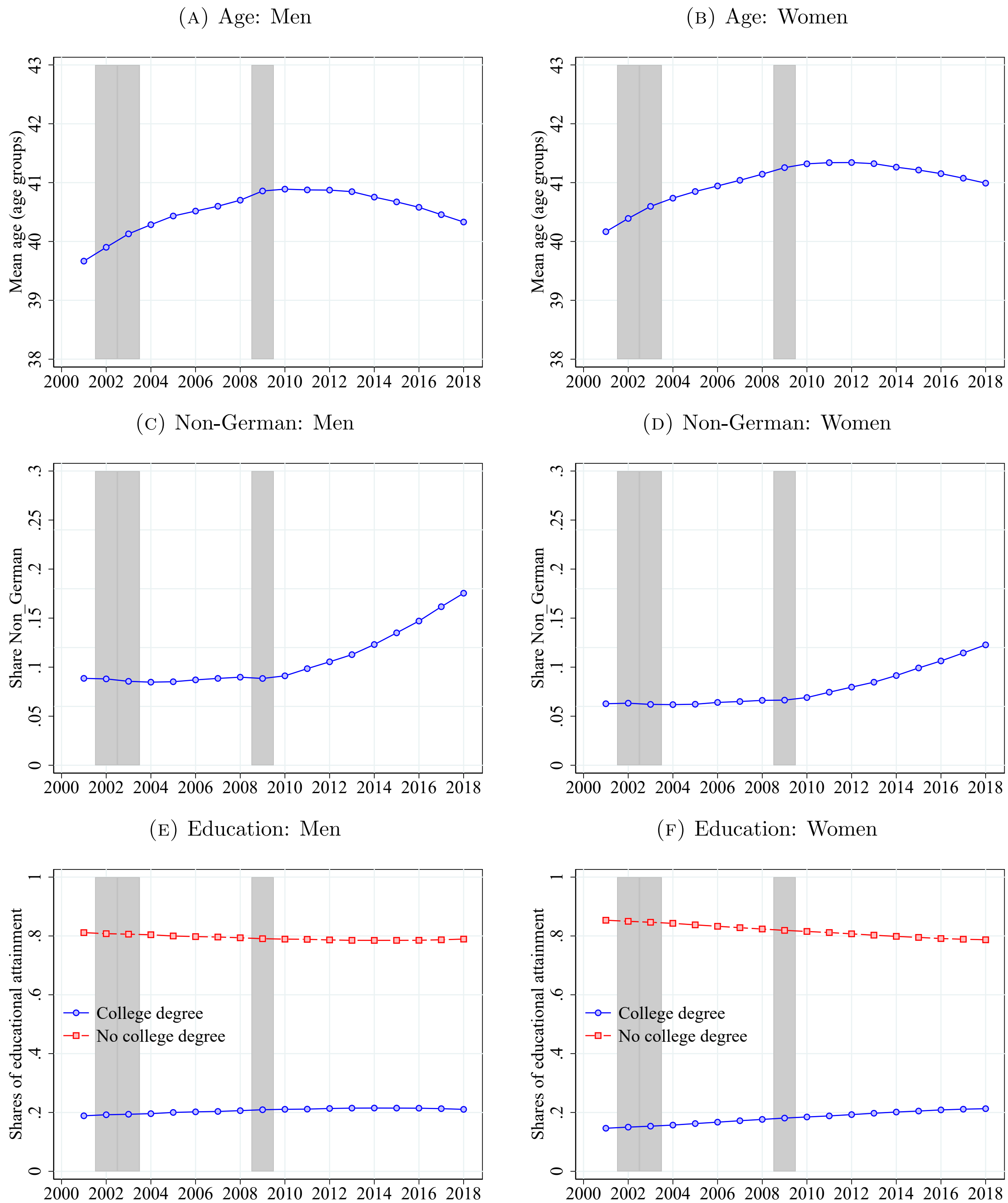
where z denotes the respective attribute to be held constant and $F(z|t_z)$ the respective individual distribution of z in year t.

Figure E.15 displays the evolution of the demographic observables age, non-German nationality and educational attainment (2 groups) before reweighting separately for men and women. Mean age increases in the sample until about 2010 before slightly decreasing until 2018 as displayed in Panels A and B. It starts at 39.6 for men and 40.2 for women in 2010, peaks at 40.9 (men) and 41.3 (women) and ends at 40.3 (men) and 41 (women) in 2018.[59] Panels C and D show that the share of non-German citizens is almost constant until 2010 and then almost doubles from 2010 to 2018 for both men and women. It is constantly higher for men (9 to 17.5 percent) than for women (6.5 to 12.5 percent). The share of workers with college degree plotted in Panels E and F, slightly

[59] This only holds for our sample with the restriction to prime age workers. The average age of the total population and the age of the workforce constantly increases during this time. The decrease in our sample tends to reflect larger birth cohorts leaving the sample whn passing age 55.

increases from 2001 to 2018. For men it increases from 19% to 21% and for women from from 15% to 21%.

Figure E.15: Weighting Variables: Demographics

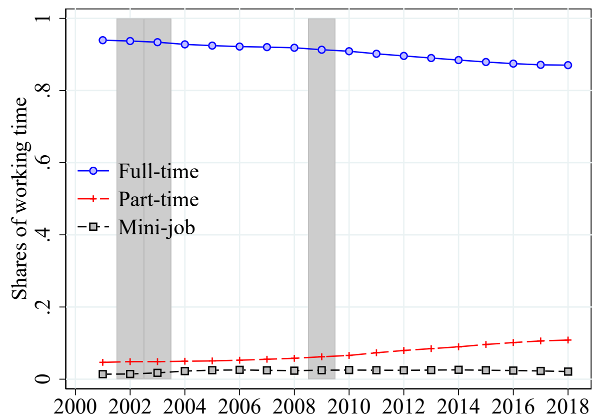


Notes: This figure plots the evolution of demographic observables in the IAB data (CS sample) before and after reweighting for men. Shaded areas indicate recessions.

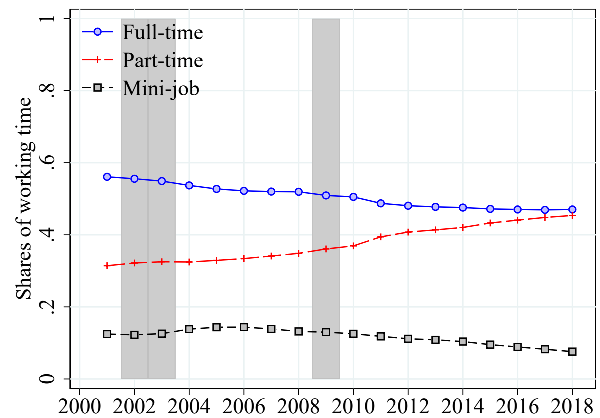
Figure E.16 plots the evolution of work characteristics before applying the DFL weights separately for men and women. In Panels A and B, we show the evolution of full-time, part-time and mini-job shares in our sample before reweighting. The share of full-time workers decreased for men and women. While decreasing, it is consistently higher for men (94% to 87%) than for women (56% to 47%). The share of part-time workers increases over time, from 4.5% (men) and 31.5% (women) in 2001 to 11% (men) and 45.5% (women) in 2018. The share of mini-jobbers is comparatively small (men: 1.5-2.5%, women: 7.5-14.5%). In Panels C and D we depict mean days in employment for men and women for all workers as well as split by median earnings. For men, mean days in employment increase from 337 in 2001 to 342 in 2012 before decreasing again to 339 in 2018. Similarly, days in employment for women increase from 338 in 2001 to 342.5 in 2006 before decreasing again to 338 in 2018. For both genders this changes are almost purely driven by below median earnings workers. For above median earning men, days in employment even decrease slightly while below median earning men experience a notable overall increase from 264 in 2002 to 290 in 2008, 293 in 2012 and then slightly decreasing to 284 in 2018.

Figure E.16: Weighting Variables - Work Characteristics

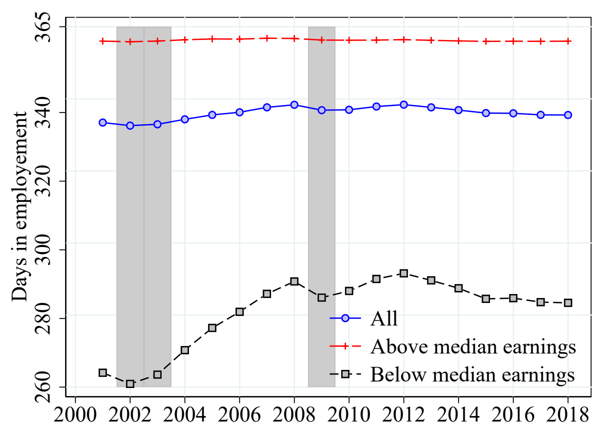
(a) Job type share: men



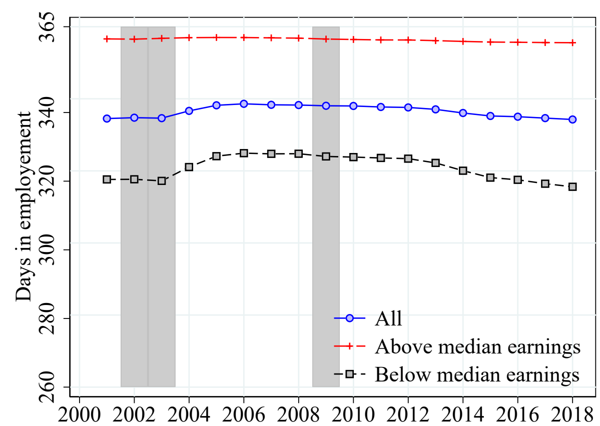
(b) Job type shares: Women



(c) Days in employment: Men



(d) Days in employment: Women



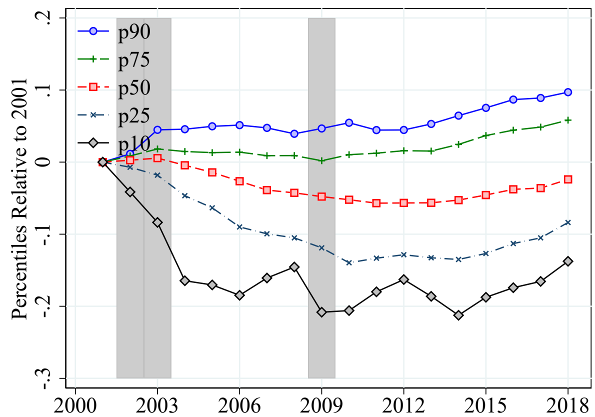
Notes: This figure plots the evolution of work characteristics in the IAB data (CS sample) before and after reweighting for men. For days in employment above and below median earnings, the earnings are weighted by $w = \frac{365}{daysinemployment}$ to account for the positive correlation of earnings and days in employment. Thereby, the median is applied to earnings as if every worker would have worked all days. Shaded areas indicate recessions.

In Figures E.17 (for men) and E.18 (for women), we show the evolution of log earnings percentiles before and after reweighting separately by certain demographic and work characteristics. Counterfactual percentiles are constructed by applying the weights obtained using the DFL approach as described above. These figures complement Figure 5 by plotting several percentiles for each reweighted observable in one single graph similarly to Figure 3. Holding age or education constant at their 2001 values appear not to affect percentile evolution patterns much. Keeping non-German nationality constant at initial values moves lower percentile patterns upwards in later years. Thus, earnings inequality would be lower if share of non-Germans would have stayed constant. This is in line with the share of non-Germans being almost constant until 2010 and increasing after 2010 (see Figure E.15). When holding job type (full-time, part-time or mini-job) or days in employment constant over time, we observe more notable changes to percentile evolution patterns.

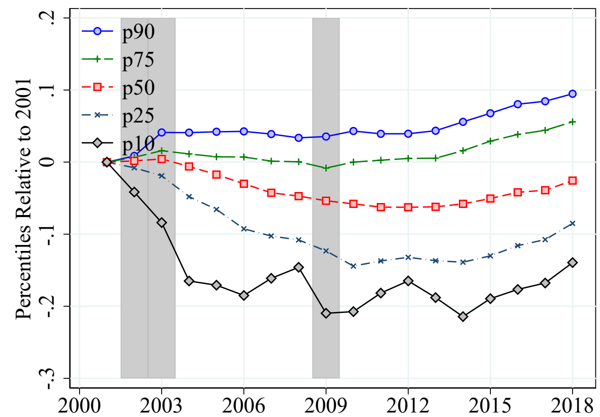
Those tend to affect lower percentiles more. For both, men and women, holding share of full-time, part-time and mini-job workers constant would have compressed the distribution such that percentile evolution appears more compressed. This would have resulted in a more constant evolution of real earnings inequality. The opposite is true for days in employment but almost solely for men. If days in employment would have been remained on (lower) 2001 values (see E.16), this would have resulted in a more spread evolution of real earnings percentiles and thus higher inequality. The result is in line with days in employment increasing by 15 days between 2001 and 2018 for men earning below-median but slightly decreased by 1 day for above-median earning men. The detailed percentile-wise results of the reweighting analyses are discussed in section 3.1.

Figure E.17: Percentiles of the log real annual earnings before and after reweighting – Men

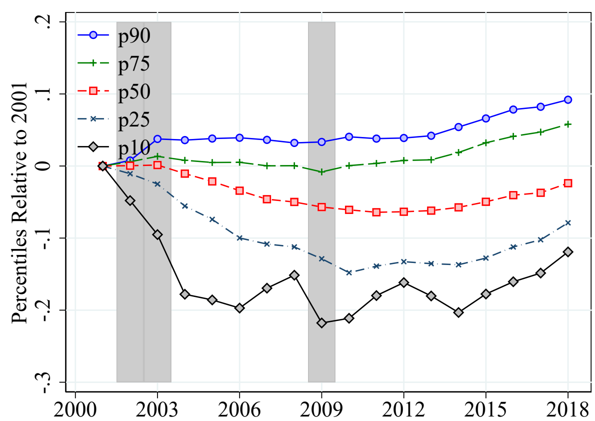
(a) Baseline (unweighted)



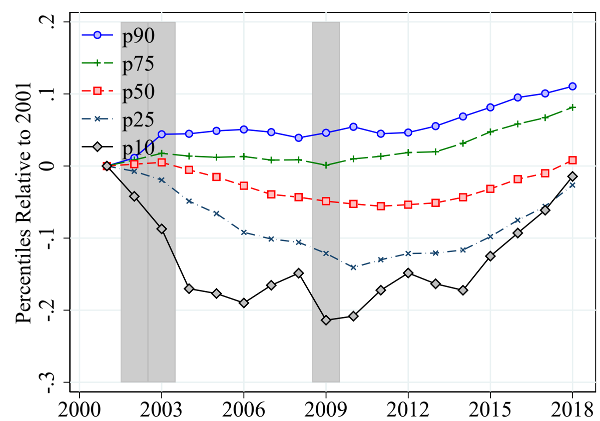
(b) Reweighted for Education



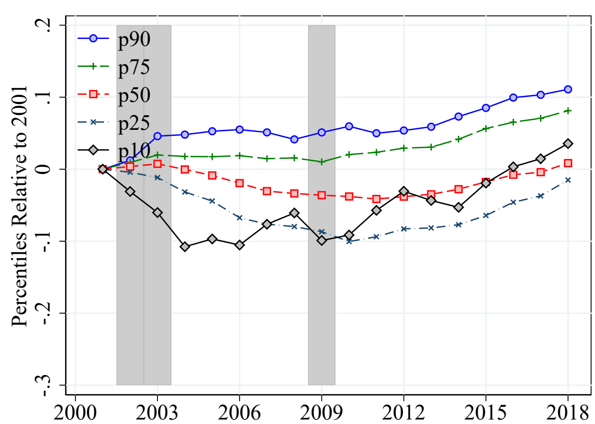
(c) Reweighted for Age



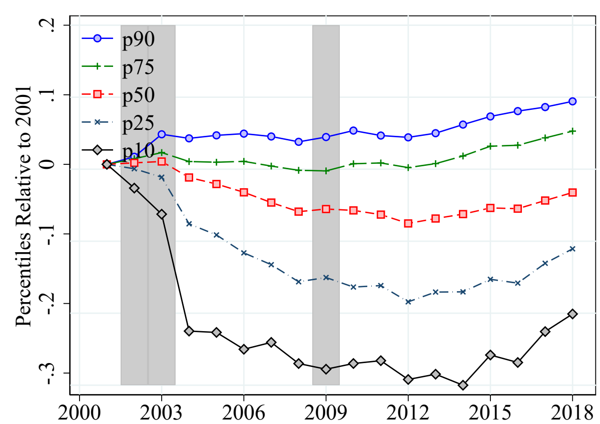
(d) Reweighted for Non-German nationality



(e) Reweighted for Job type



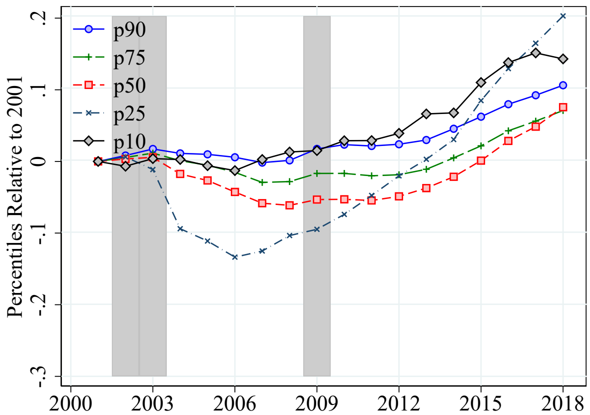
(f) Reweighted for Days in Employment



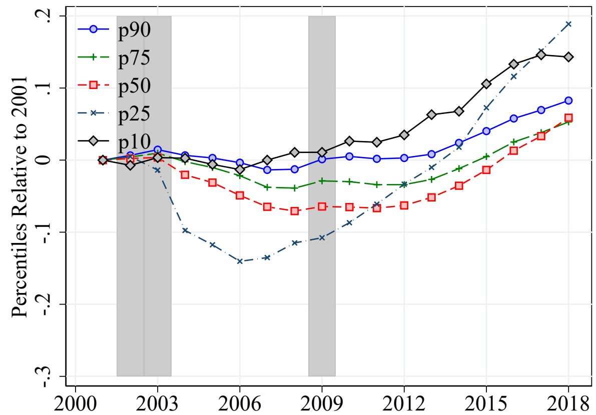
Notes: This figure shows the evolution of different counterfactual log real annual earnings percentiles in the IAB data (CS sample) for men. The counterfactual percentiles are constructed by reweighting the data such that observable dimensions are held constant at the 2001 level. Figure 5 in the main text includes the 10th, 50th and 90th percentile. Shaded areas indicate recessions.

Figure E.18: Percentiles of the log real annual earnings before and after reweighting – Women

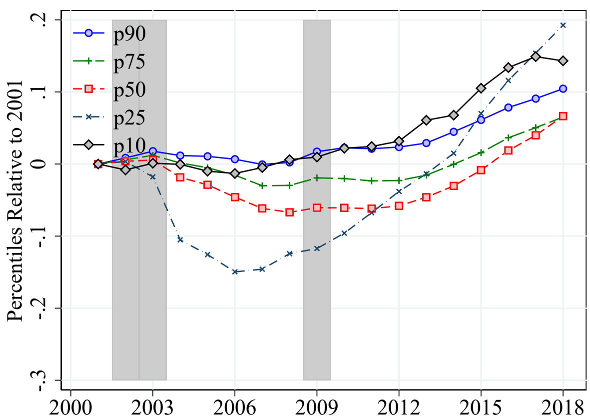
(a) Baseline (unweighted)



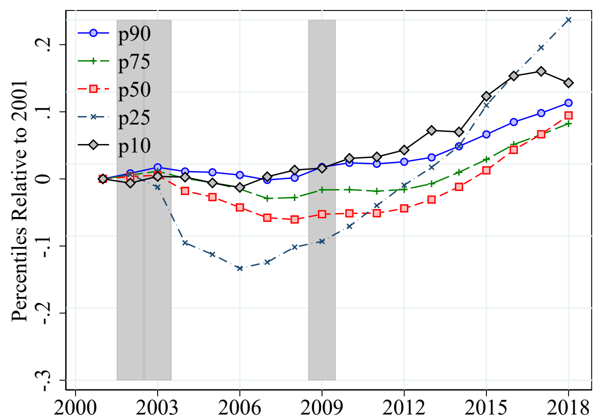
(b) Reweighted for Education



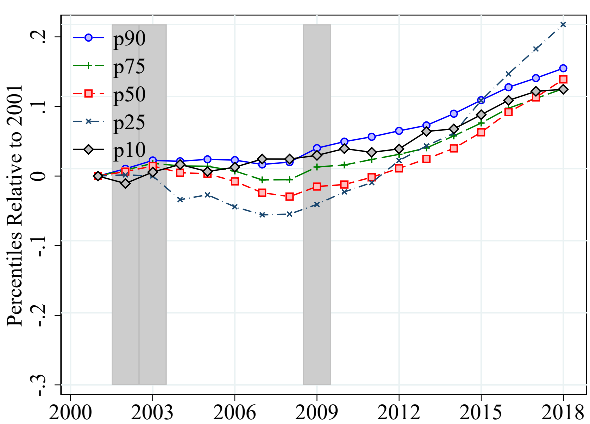
(c) Reweighted for Age



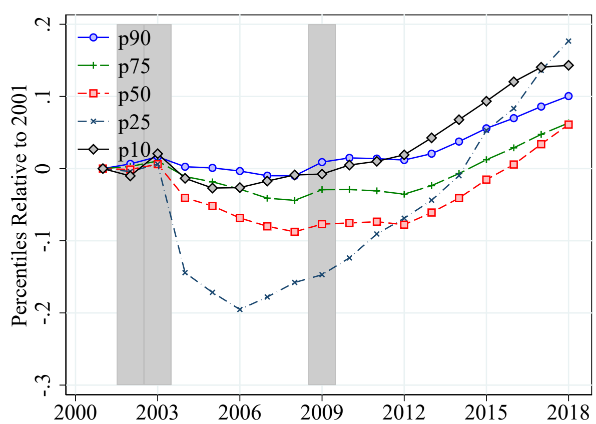
(d) Reweighted for Non-German nationality



(e) Reweighted for Job type



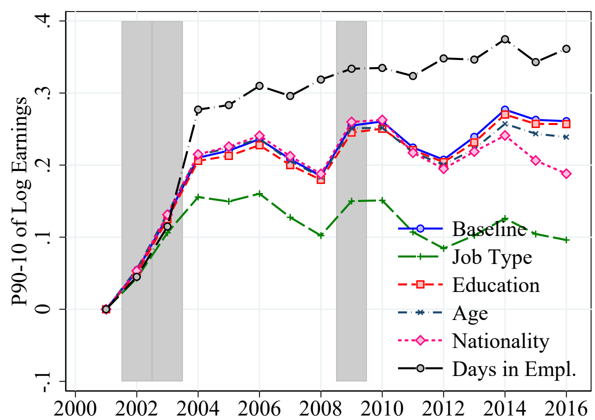
(f) Reweighted for Days in Employment



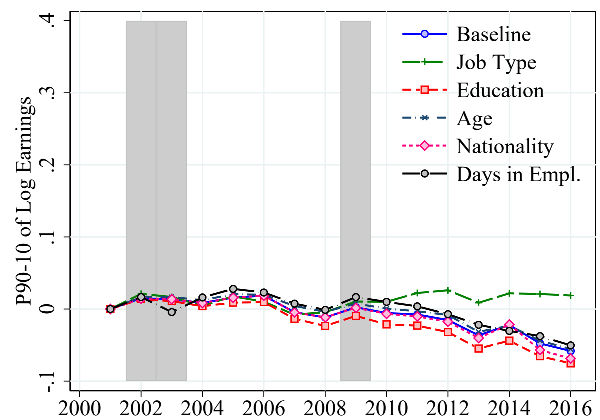
Notes: This figure shows the evolution of different counterfactual log real annual earnings percentiles in the IAB data (CS sample) for women. The counterfactual percentiles are constructed by reweighting the data such that observable dimensions are held constant at the 2001 level. Figure 5 in the main text includes the 10th, 50th and 90th percentile. Shaded areas indicate recessions.

Figure E.19: Counterfactual Evolution of Log Earnings Percentile Differentials (Reweighting)

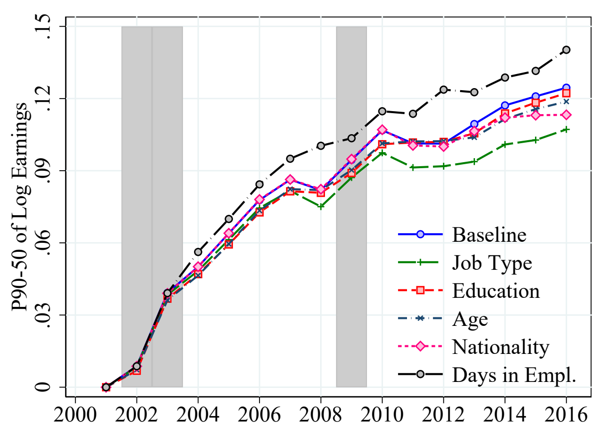
(a) P90-P10: Men



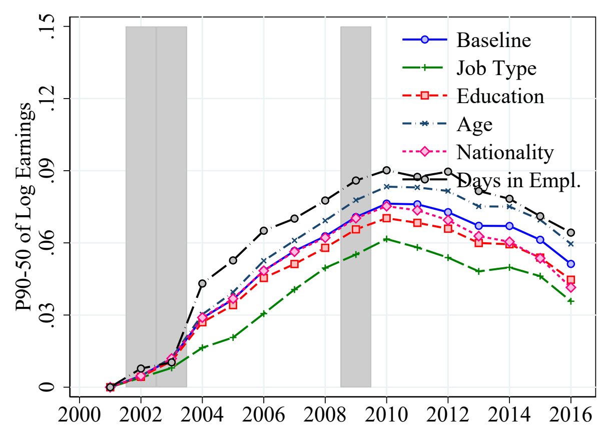
(b) P90-P10: Women



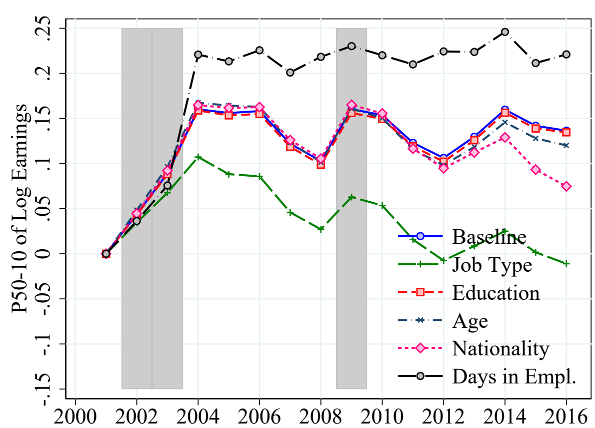
(c) P90-P50: Men



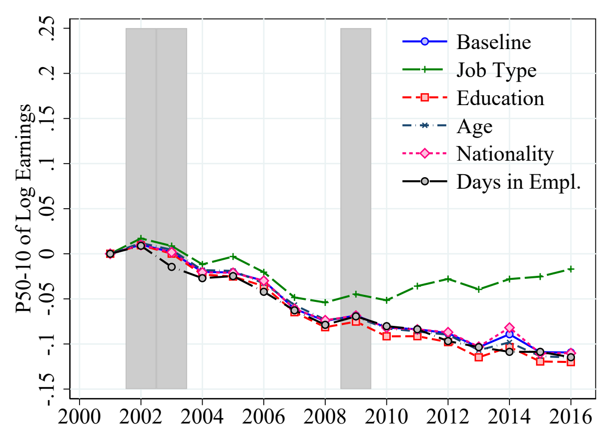
(d) P90-P50: Women



(e) P50-P10: Men



(f) P50-P10: Women

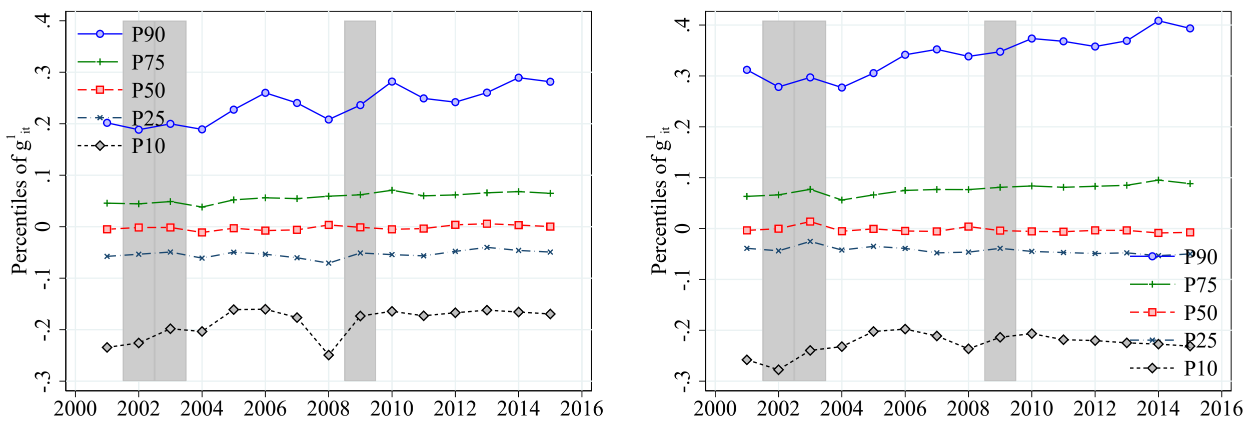


Notes: This figure shows the evolution of different counterfactual percentile differences of the log real annual earnings distribution over time in the IAB data (CS sample) separately for men and women. The counterfactual percentiles are constructed by reweighting the data such that observable dimensions are held constant at the 2001 level. Figure 5 in the main text includes the 10th, 50th and 90th percentile. Shaded areas indicate recessions.

E.3 Additional Results for Earnings Dynamics (Section 3.2)

FIGURE E.20: PERCENTILES OF 1-YEAR LOG EARNINGS CHANGES

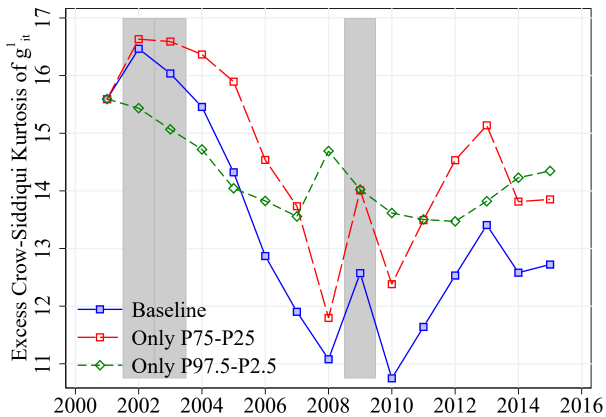
(A) Men (B) Women



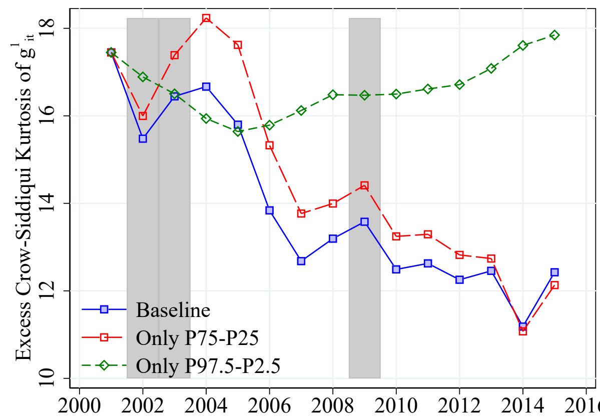
Notes: This figure shows selected percentiles of the distribution of 1-year changes in residualized log real annual earnings (from t to $t+1$) in the combined IAB-TPP data (LS sample) separately for men and women. Shaded areas indicate recessions. See Appendix D.2.2 for details on how we construct the distribution of log earnings growth from IAB and TPP data.

Figure E.21: Decomposition of Excess Crow-Siddiqui Kurtosis of 1-Year Log Earnings Changes

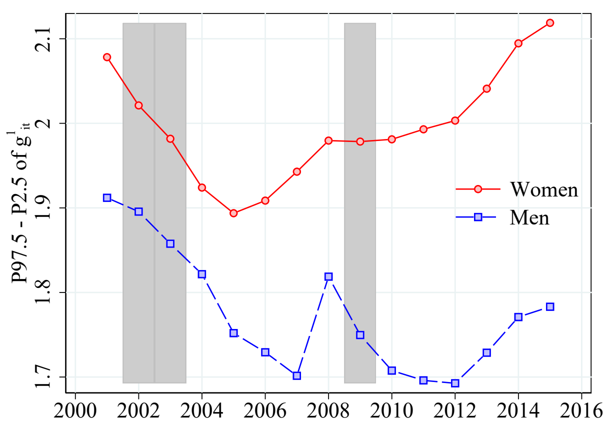
(a) Counterfactuals: Men



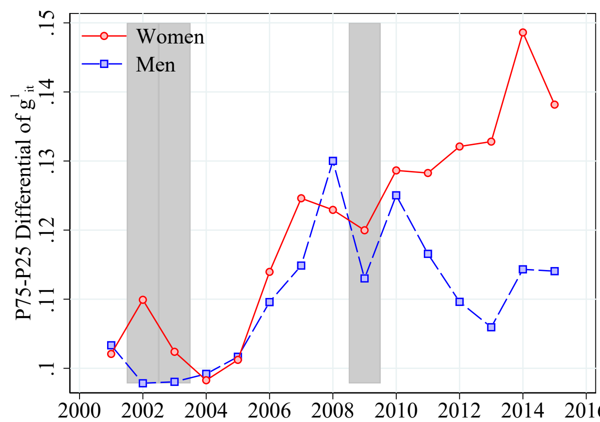
(b) Counterfactuals: Women



(c) P97.5 - P2.5

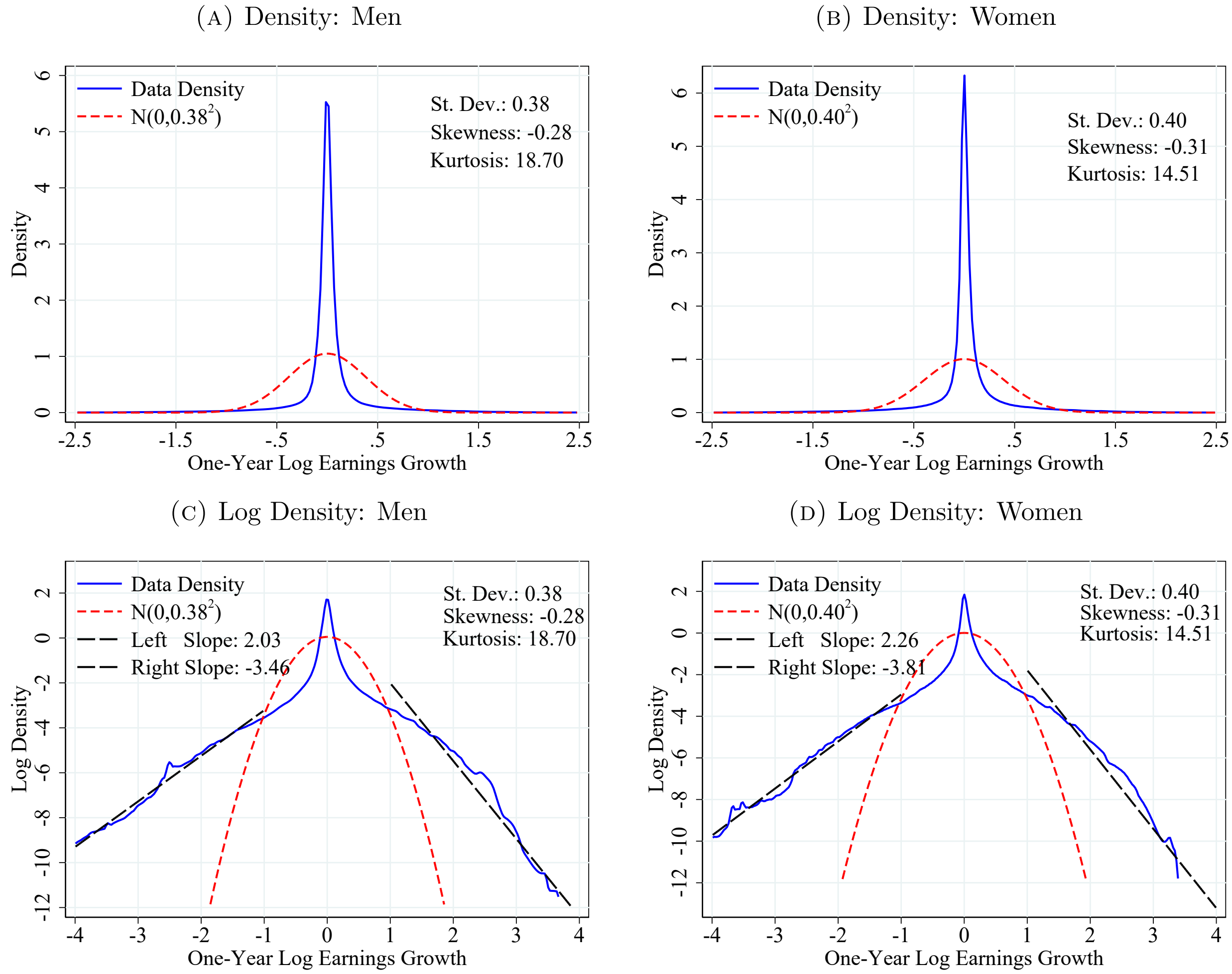


(d) P75-P25



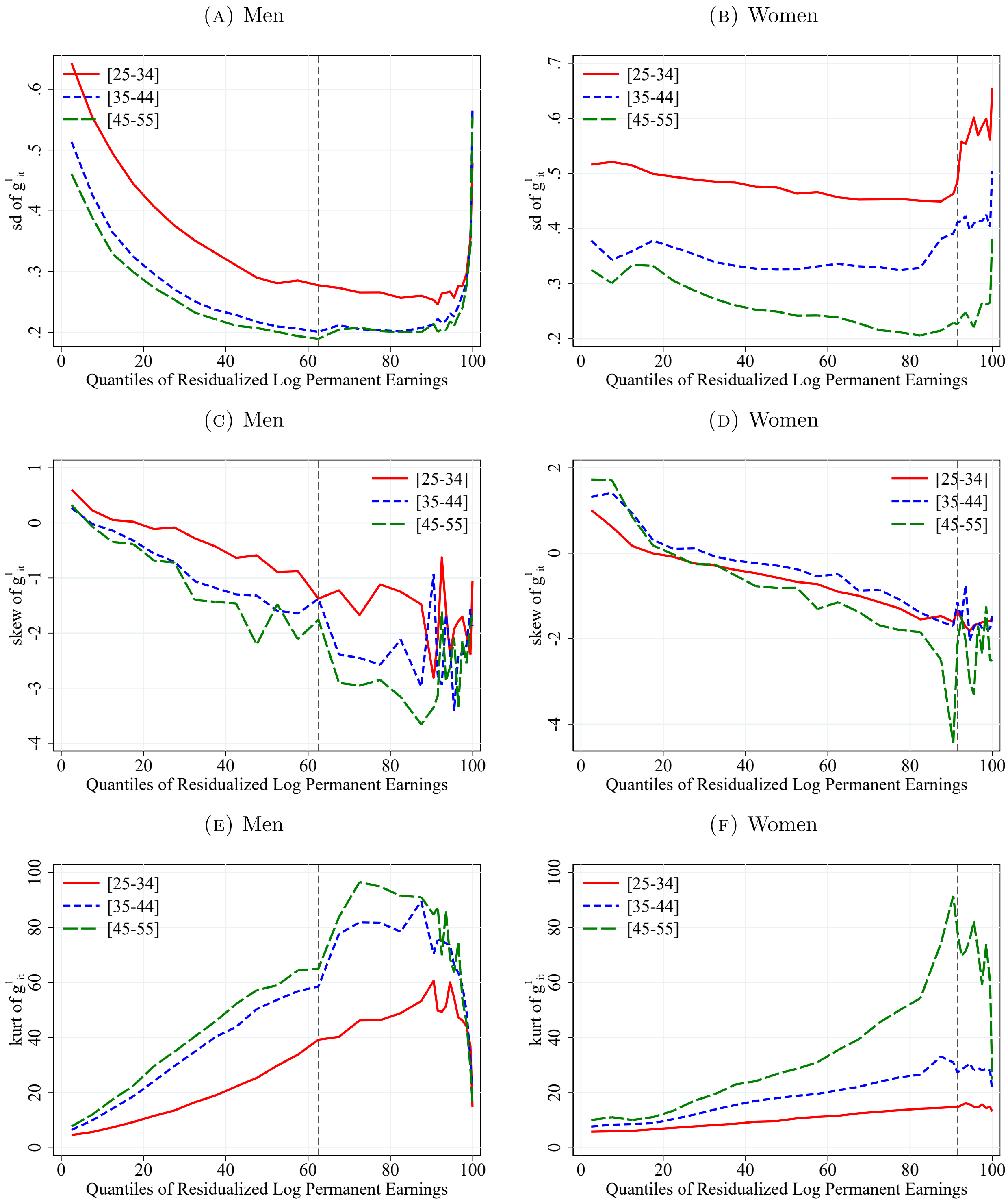
Notes: This figure shows decomposition analyses of the excess Crow-Siddiqui kurtosis in the combined IAB-TPP data (LS sample). Panels A and B show how the excess Crow-Siddiqui kurtosis of 1-year residualized log earnings changes (from t to $t+1$) would have evolved if only the numerator (P97.5–P2.5) or only the denominator (P75-P25) of the excess Crow-Siddiqui kurtosis would have changed over time. Panels C and D show the evolution of these components. Excess Crow-Siddiqui kurtosis is calculated as $\frac{P97.5-P2.5}{P75-P25} - 2.91$ where the first term is the Crow-Siddiqui measure of kurtosis and 2.91 corresponds to the value of this measure for Normal distribution. Shaded areas indicate recessions. See Appendix D.2.2 for details on how we construct the distribution of log earnings growth from IAB and TPP data.

Figure E.22: Densities of 1-Year Log Earnings Changes (Year 2005)



Notes: This figure shows Kernel density estimates of 1-year changes in residualized log earnings for the year 2005 and the respective density of a Normal distribution with zero mean and the same standard deviation as in the combined IAB-TPP data (LS sample).

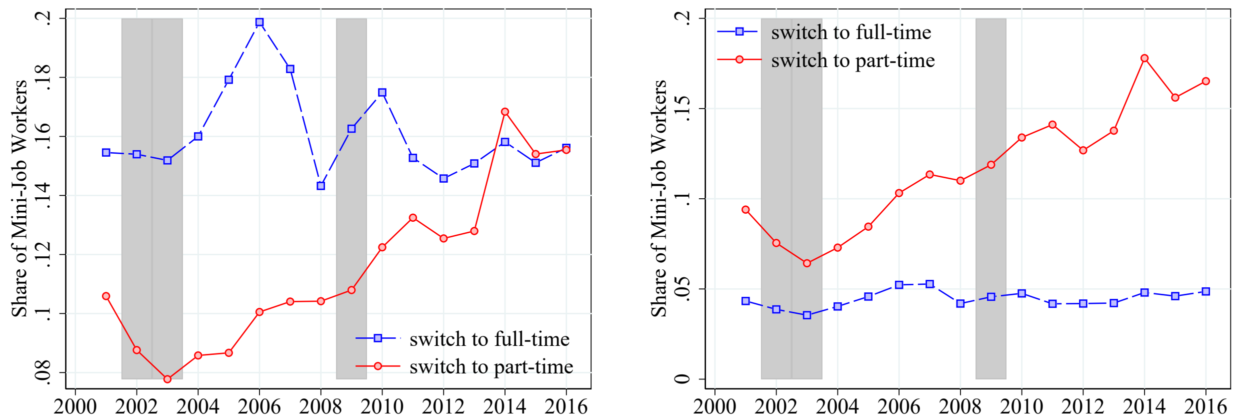
Figure E.23: Heterogeneity in Standardized Moments of 1-Year Log Earnings Changes



Notes: This figure shows the standard deviation, skewness and excess kurtosis (third and fourth standardized moments) of 1-year changes in residualized log real total income by quantiles of residualized permanent earnings and age groups in the combined IAB-TPP data (H sample) as averages from 2004 to 2011 and separately for men and women. Permanent earnings $P_{i,t-1}$ are defined as the residual (net of a full set of gender and year specific age dummies) of the log of average earnings between $t-3$ and $t-1$. See Footnote 23 definitions and interpretation of Kelley skewness and excess Crow-Siddiqui kurtosis. See Appendix Figures D.10, D.11 and D.12 for a comparison of the underlying data in both data sources.

Figure E.24: Transitions out of Mini-Jobs

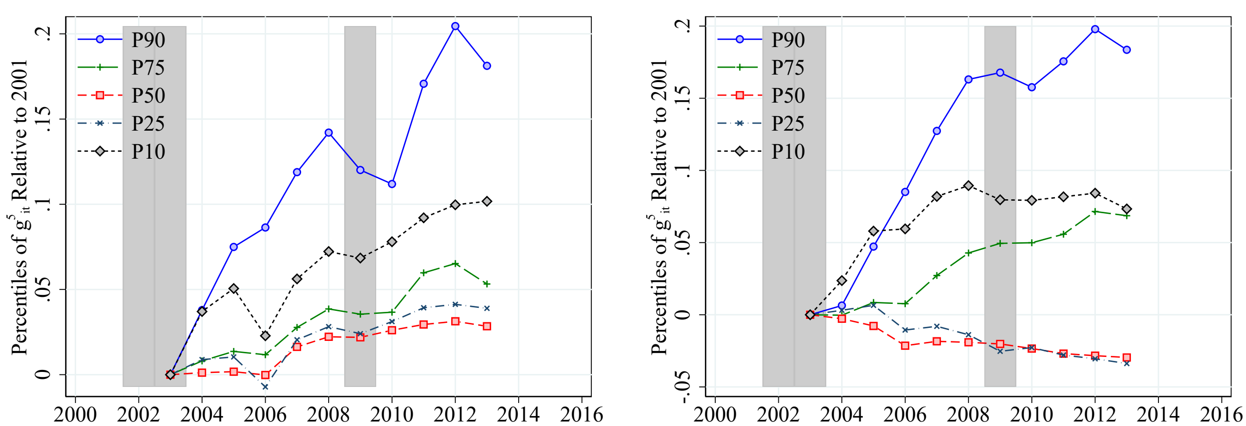
(A) Men (B) Women



Notes: This figure shows the share of workers who transition from a mini-job to part-time and full-time employment (from t to $t+1$) in the IAB data (CS sample).

Figure E.25: Percentiles of 5-Year Log Earnings Changes

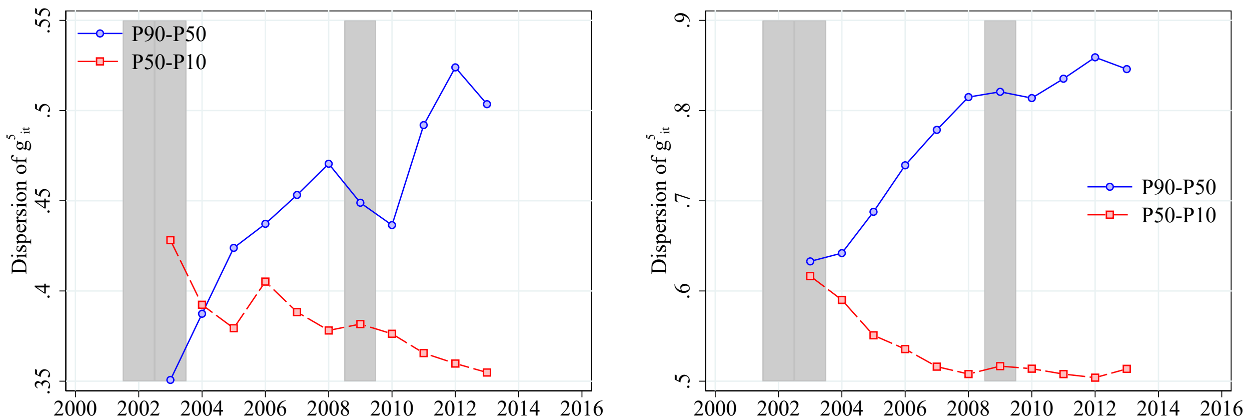
(A) Men (B) Women



Notes: This figure shows selected percentiles of the distribution of 5-year changes in residualized log real annual earnings (from t to $t+1$) in the combined IAB-TPP data (LS sample) separately for men and women. Shaded areas indicate recessions. See Appendix D.2.2 for details on how we construct the distribution of log earnings growth from IAB and TPP data.

Figure E.26: Dispersion of 5-Year Log Earnings Changes

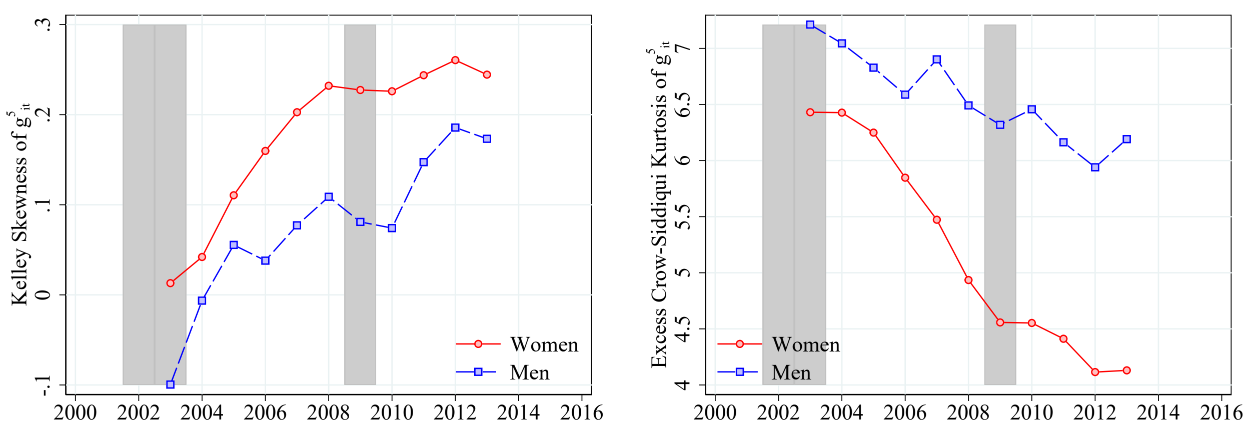
(a) Men (b) Women



Notes: This figure shows 5-year changes in residualized log earnings (from $t-2$ to $t+3$) in the combined IAB-TPP data (LS sample). Shaded areas indicate recessions.

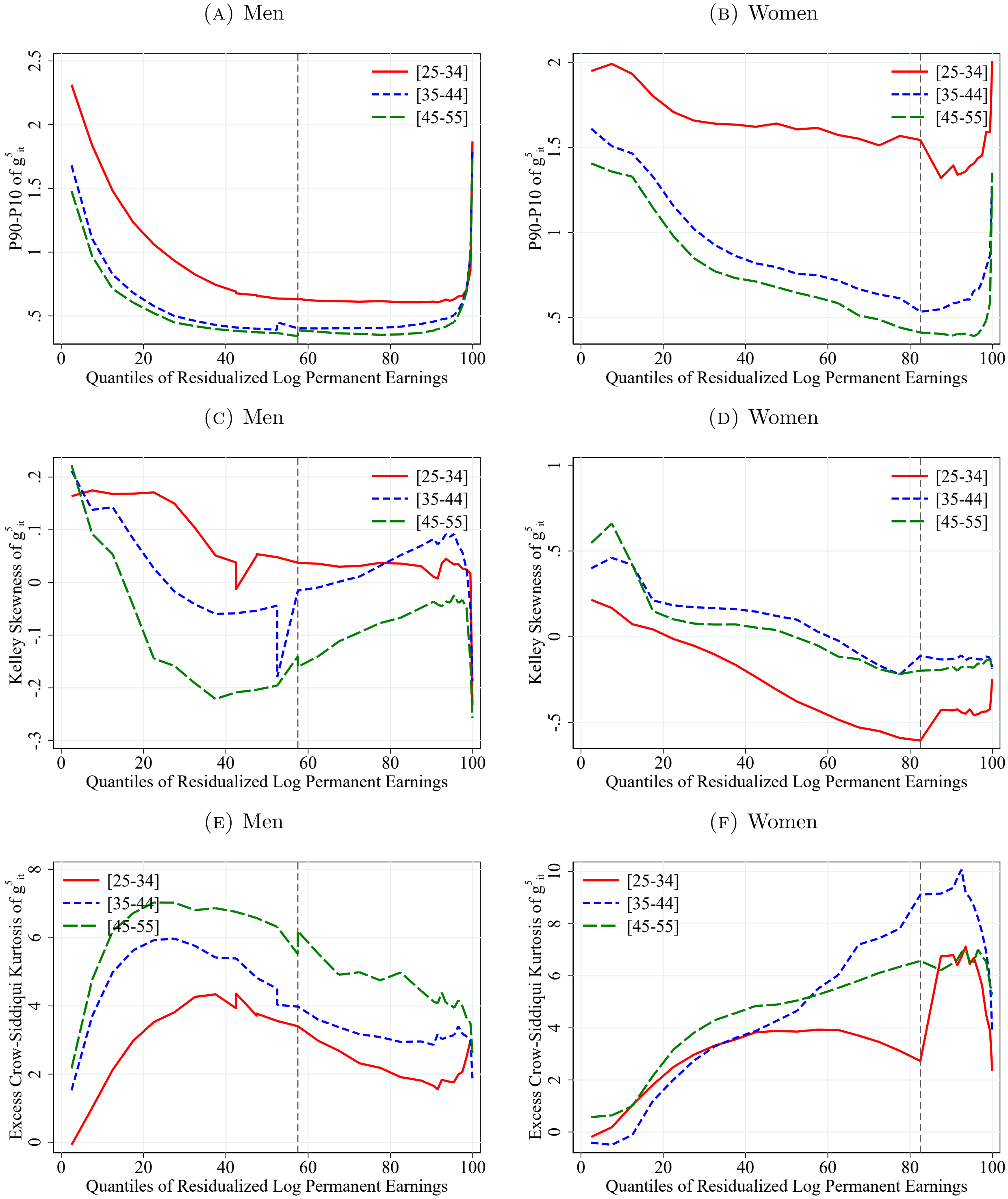
Figure E.27: Skewness and Kurtosis of 5-Year Log Earnings Changes

(a) Kelley Skewness (b) Excess Crow-Siddiqui Kurtosis



Notes: This figure shows 5-year changes in residualized log earnings (from $t-2$ to $t+3$) in the combined IAB-TPP data (LS sample). Kelley skewness is $\frac{P90-2P50+P10}{P90-P10}$. Excess Crow-Siddiqui kurtosis is calculated as $\frac{P97.5-P2.5}{P75-P25}-2.91$ where the first term is the Crow-Siddiqui measure of kurtosis and 2.91 corresponds to the value of this measure for Normal distribution. Shaded areas indicate recessions.

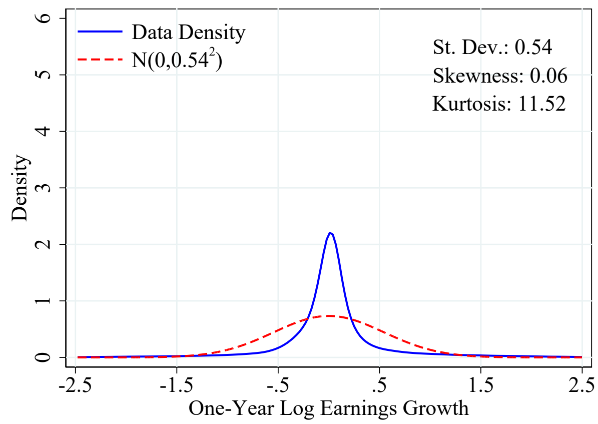
FIGURE E.28: HETEROGENEITY IN DISPERSION, SKEWNESS AND KURTOSIS OF 5-YEAR LOG EARNINGS CHANGES



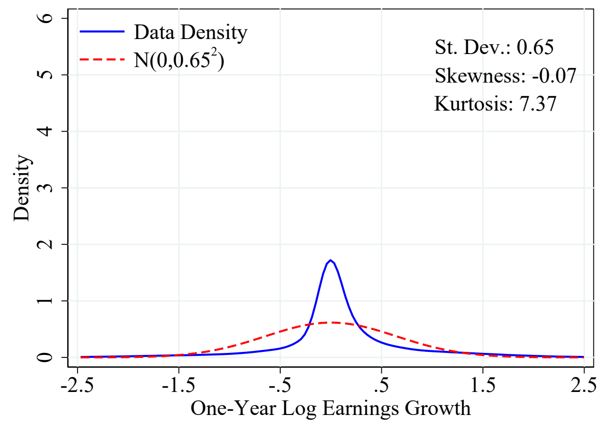
This figure shows the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 5-year changes in residualized log earnings (from $t-2$ to $t+3$) in the combined IAB-TPP data (H sample) as averages from 2004 to 2011 by quantiles of residualized permanent earnings and age groups. Kelley skewness is $\frac{P90-2P50+P10}{P90-P10}$. Excess Crow-Siddiqui kurtosis is calculated as $\frac{P97.5-P2.5}{P75-P25} - 2.91$ where the first term is the Crow-Siddiqui measure of kurtosis and 2.91 corresponds to the value of this measure for Normal distribution. Shaded areas indicate recessions.

Figure E.29: Densities of 5-Year Log Earnings Changes (Year 2005)

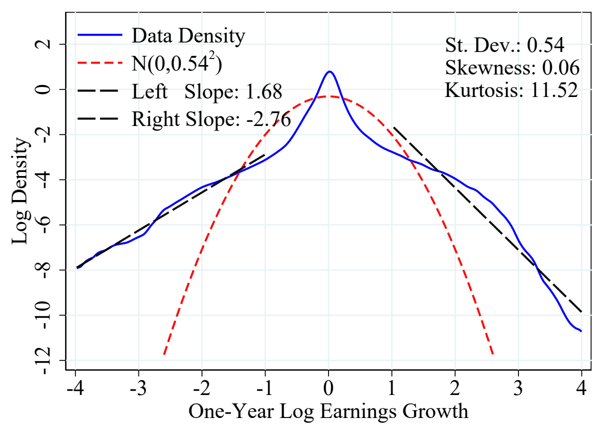
(a) Density: Men



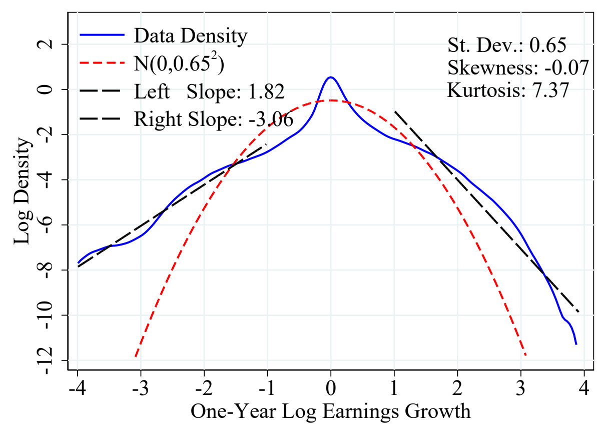
(b) Density: Women



(c) Log Density: Men



(d) Log Density: Women

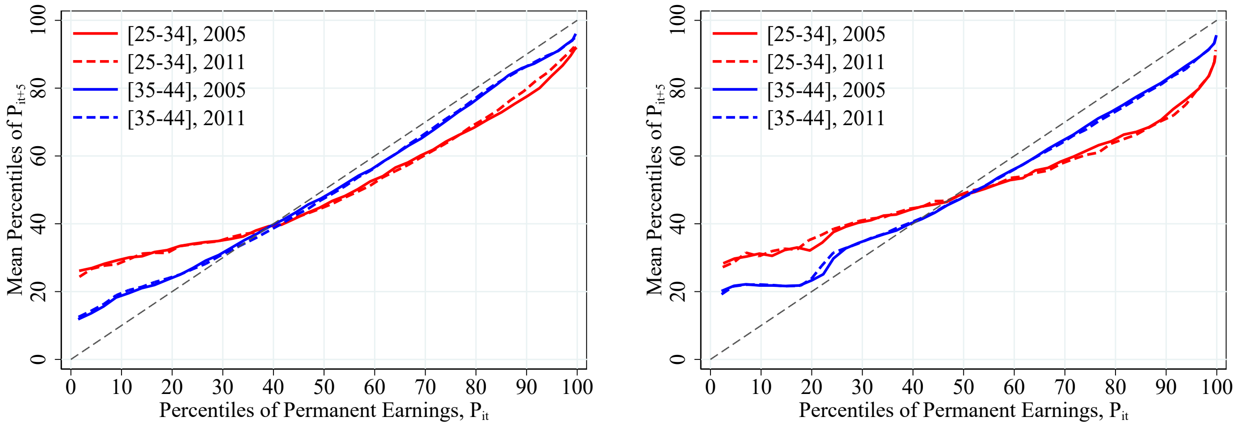


Notes: This figure shows Kernel density estimates of 5-year changes in residualized log earnings (from $t-2$ to $t+3$) for the year 2005 and the respective density of a Normal distribution with zero mean and the same standard deviation as in the combined IAB-TPP data (LS sample).

Figure E.30: Evolution of 5-Year Permanent Earnings Mobility

(a) Rank-Rank Mobility: Men

(b) Rank-Rank Mobility: Women



Notes: This figure shows the evolution of average 5-year rank-rank mobility of permanent earnings in the combined IAB-TPP data (H sample) as averages from 2004 to 2011, separately for men and women and two different age groups. Permanent income calculated using earnings from $t-1$, $t-2$ and $t-3$.

F Core Analysis of Earnings with Longer Samples

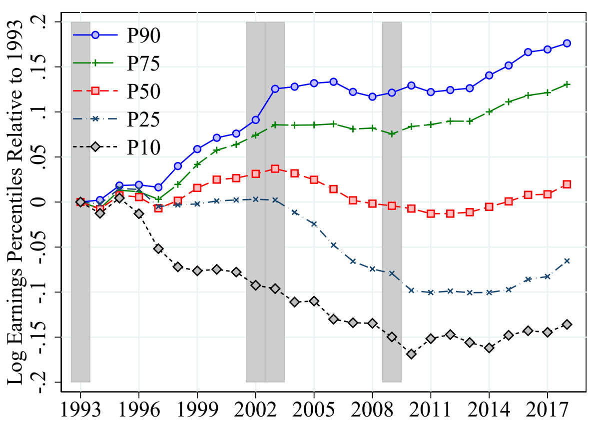
In this section, we present figures similar to those of the core analysis of this paper in Section 3 for longer samples based on IAB data only.

F.1 IAB Data 1993–2018

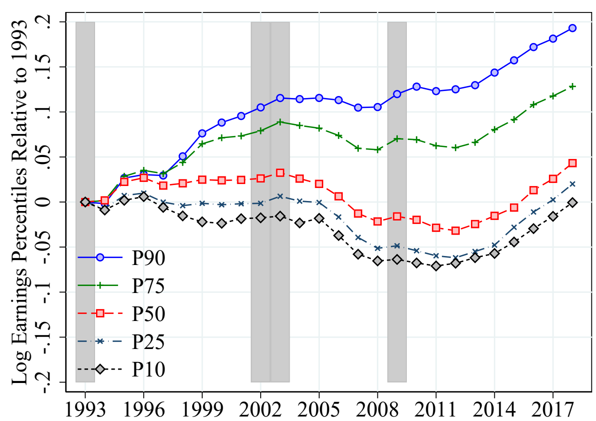
Using data from the IAB for the years 1993 to 2018, we extended the analysis by including several years prior to the sample used in the main section of this paper. To account for changes in mini-job regulations and workforce composition changes due to measurement changes in 1999, the minimum earnings threshold is set to 6,250 Euro annual earnings in 2018 to obtain a consistent sample over the whole time span, i.e. mini-jobs are not included in the longer sample.

Figure F.1: Evolution of Log Earnings Percentiles

(a) Men



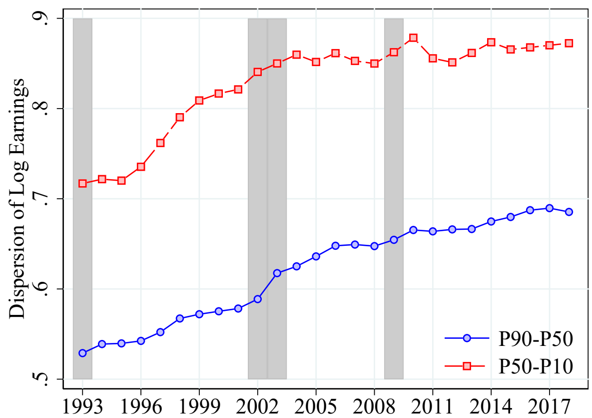
(b) Women



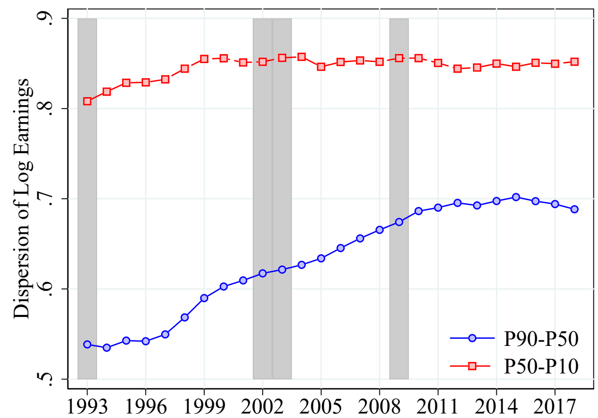
Notes: This figure shows the evolution of selected percentiles of log real earnings from 1993 to 2018 in the IAB data (CS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. All percentiles are normalized to 0 in 1993. Shaded areas indicate recessions. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. Shaded areas indicate recessions. The analysis for the core sample is in Figure 3.

Figure F.2: Earnings Inequality: Log Percentile Differentials

(a) Upper and Lower Inequality: Men



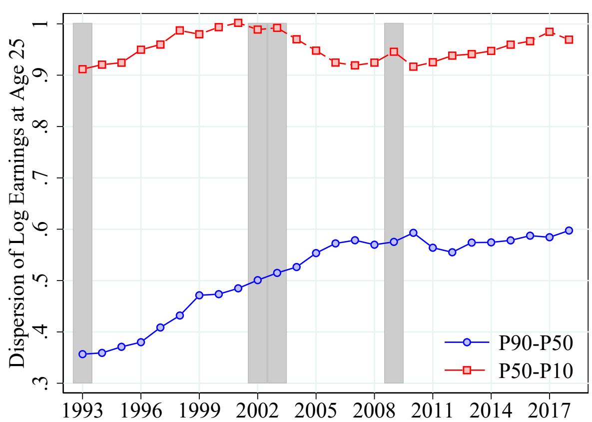
(b) Upper and Lower Inequality: Women



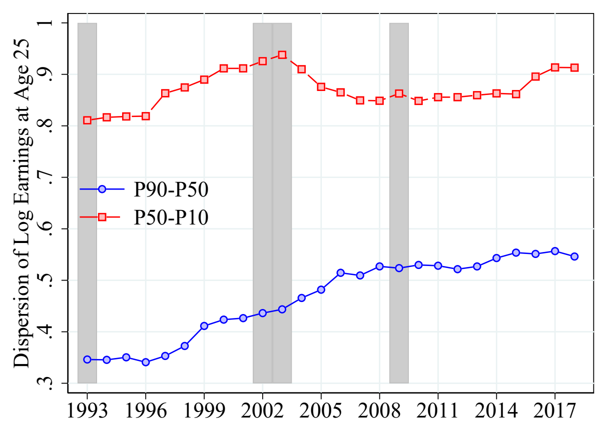
Notes: This figure shows percentile differentials of log real annual earnings in the IAB data (CS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. Shaded areas indicate recessions. The results of our main sample can be found in Figure 4.

Figure F.3: Initial Income Inequality (at age 25): Log Percentile Differentials

(a) Upper & Lower Inequality: Men

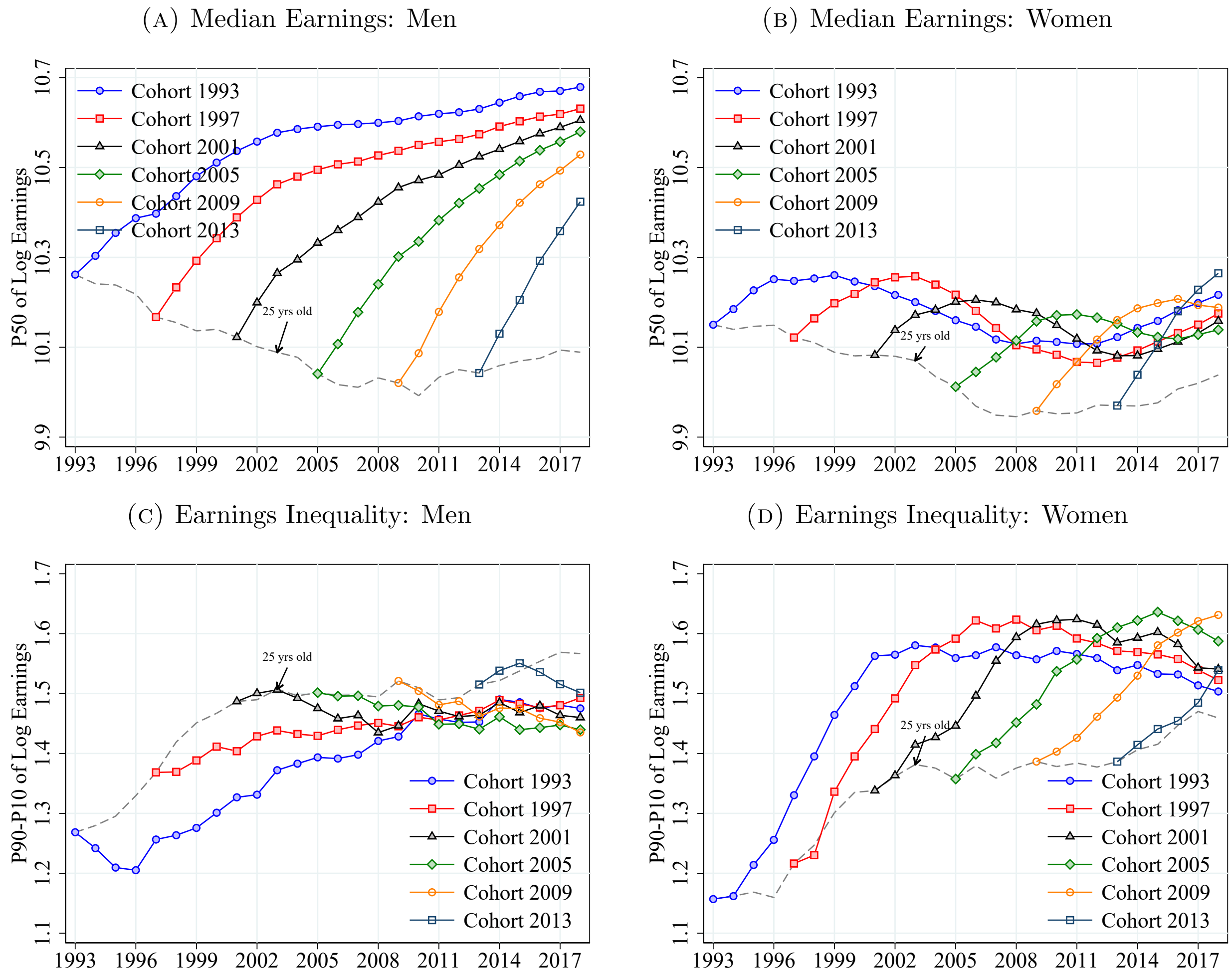


(b) Upper & Lower Inequality: Women



Notes: This figure shows initial inequality at age 25 in the IAB data (CS sample, truncated as stated below). CS sample with a minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. The IAB data is top-coded and imputed above about 60,000 Euro, which is above the P90 here. Shaded areas indicate recessions.

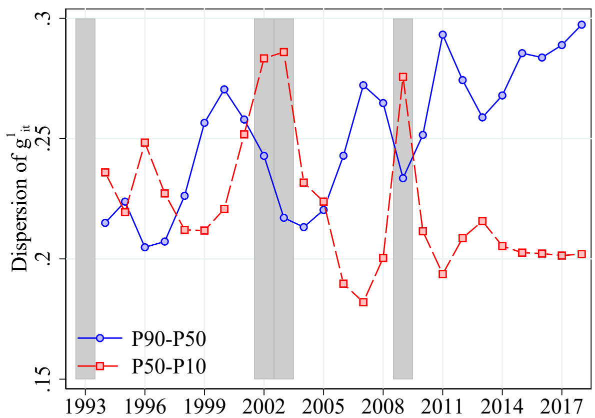
Figure F.4: Earnings Profiles and Inequality by Cohort



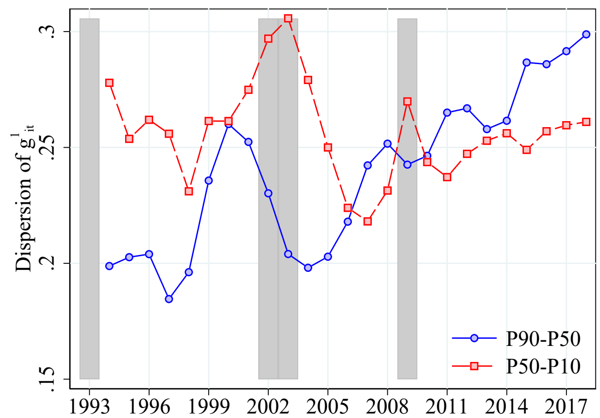
Notes: This figure shows the evolution of the median (P50) as well as the P90-P10 differential of the log real annual earnings distribution over time in the IAB data (CS sample, truncated as stated below) separately for men and women. Each colored line corresponds to an individual cohort, where "cohort t" represents the cohort aged 25 in year t. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. Shaded areas indicate recessions. Results for our main sample can be found in Figure 6.

Figure F.5: Dispersion of 1-Year Log Earnings Changes

(a) Men



(b) Women



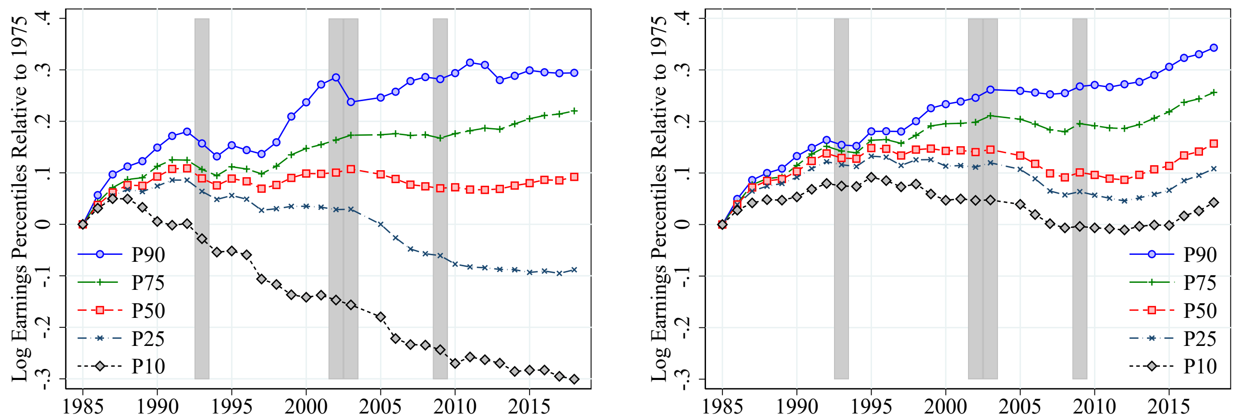
Notes: This figure shows the the P90-P50 and P50-P10 differentials of the distribution of 1-year changes in residualized log earnings (from $t-1$ to t) in the IAB data (LS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. LS sample with minimum income threshold of 6,250 Euro (2018 prices). The LS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. Shaded areas indicate recessions. The results for our core sample can be found in Figure 7.

F.2 IAB Data 1985–2018 (West Germany)

For our longest sample we use the SIAB 1975-2019 (Frodermann et al., 2021) for the years 1985 to 2018. We start in 1985 due to a structural break in the data in 1984. We apply the same minimum earnings threshold of 6,250 in 2018 Euro to exclude mini-jobs from the data. Furthermore, as data for East Germany is available from 1992 onward, we show the earnings development for West Germany only to avoid a structural break in the time series. The

Figure F.6: Evolution of Log Earnings Percentiles

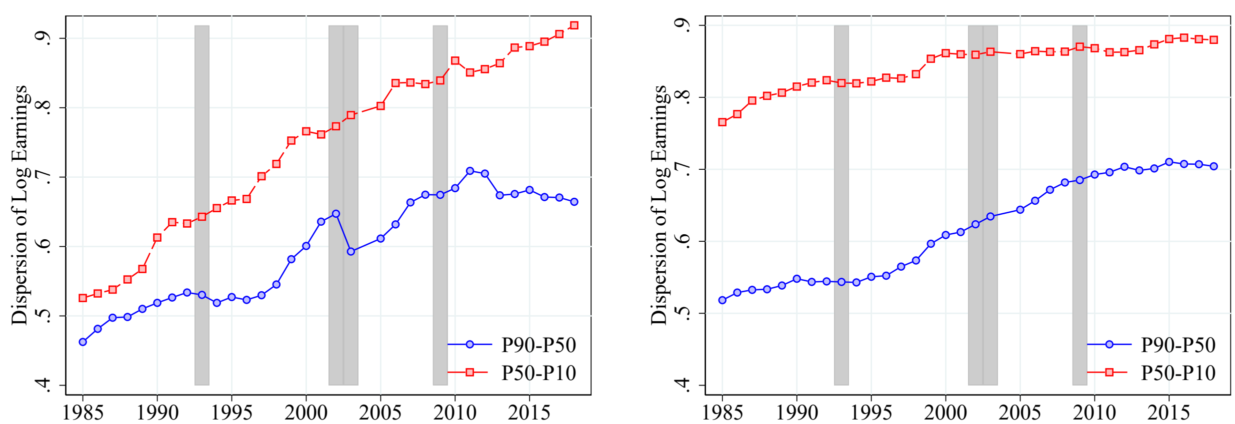
(a) Men (b) Women



Notes: This figure shows the evolution of selected percentiles of log real earnings from 1985 to 2018 in the IAB data (CS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. All percentiles are normalized to 0 in 1985. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. Shaded areas indicate recessions. The results for our core sample can be found in Figure 7.

Figure F.7: Earnings Inequality: Log Percentile Differentials

(a) Upper and Lower Inequality: Men (b) Upper and Lower Inequality: Women

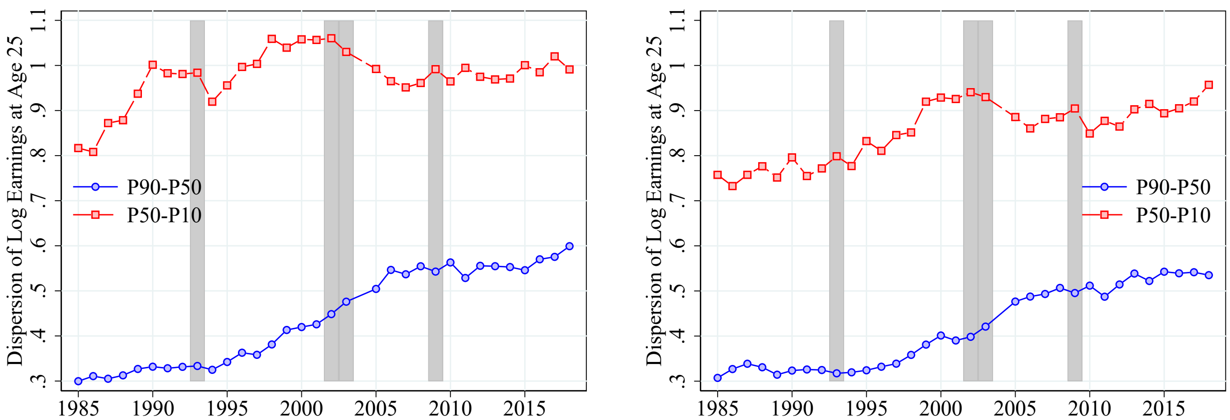


Notes: This figure shows percentile differentials of log real annual earnings in the IAB data (CS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. Shaded areas indicate recessions. The results for our core sample can be found in Figure 3.

Figure F.8: Initial Income Inequality (at age 25): Log Percentile Differentials

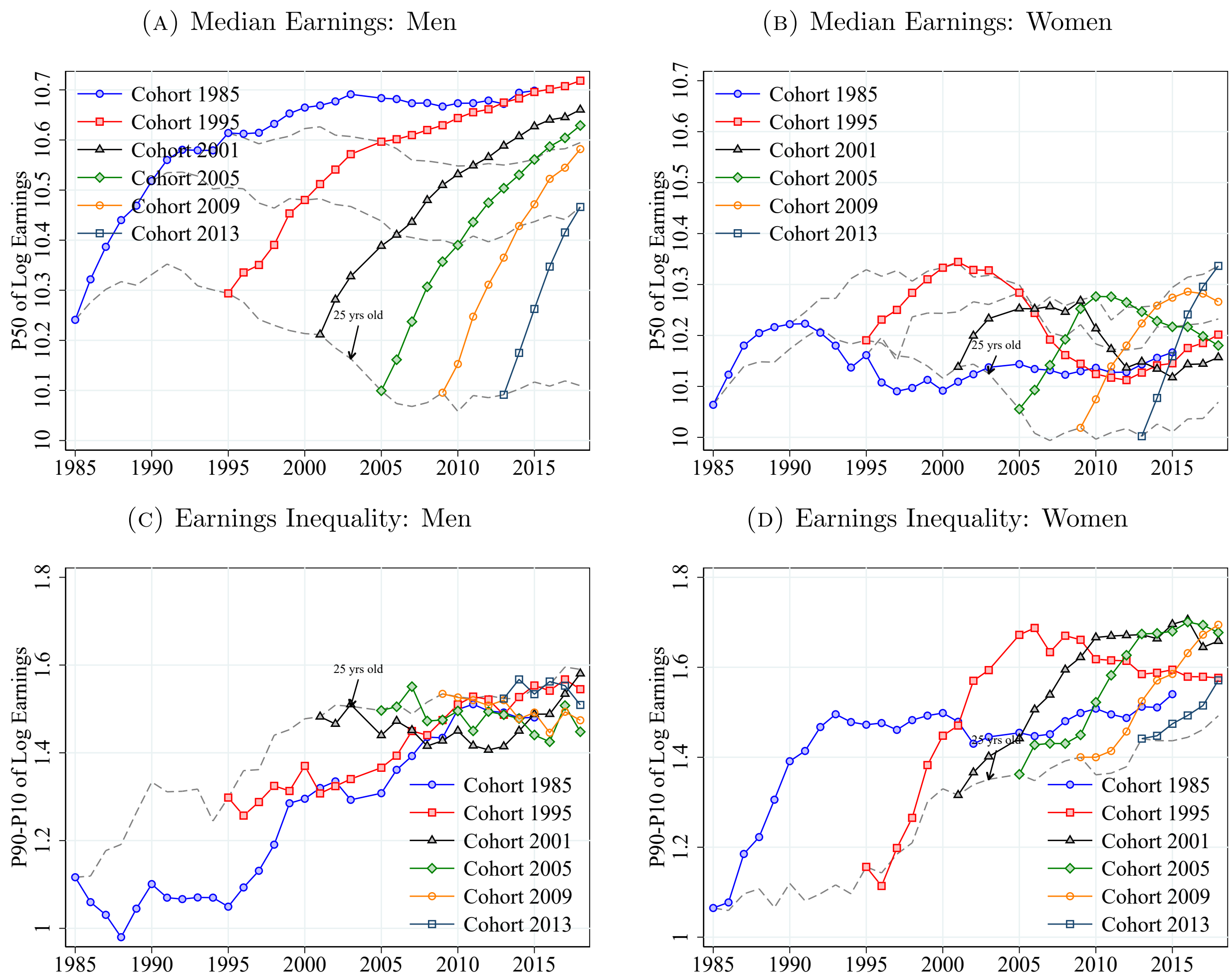
(a) Upper & Lower Inequality: Men

(b) Upper & Lower Inequality: Women



Notes: Shaded areas indicate recessions. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. The IAB data is top-coded and imputed above about 60,000 Euro, which is above the P90 here. The results for our main sample can be found in Figure 4.

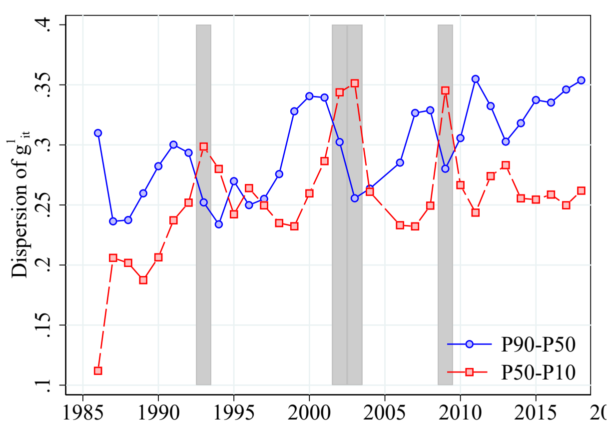
Figure F.9: Earnings Profiles and Inequality by Cohort



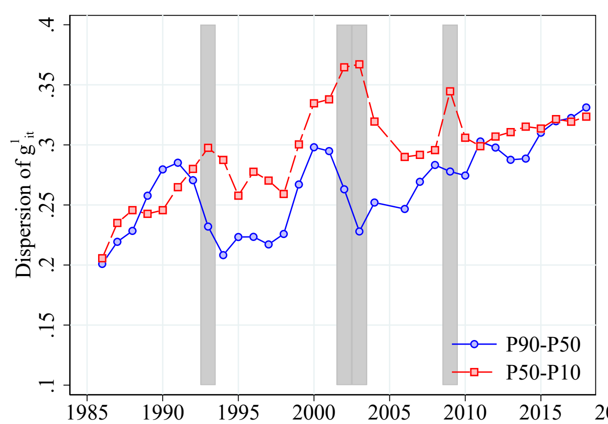
Notes: This figure shows the evolution of the median (P50) as well as the P90-P10 differential of the log real annual earnings distribution over time in the IAB data (CS sample, truncated as stated below) separately for men and women. Each colored line corresponds to an individual cohort, where "cohort t" represents the cohort aged 25 in year t. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. Shaded areas indicate recessions. Results for our main sample can be found in Figure 6.

Figure F.10: Dispersion of 1-Year Log Earnings Changes

(a) Men



(b) Women

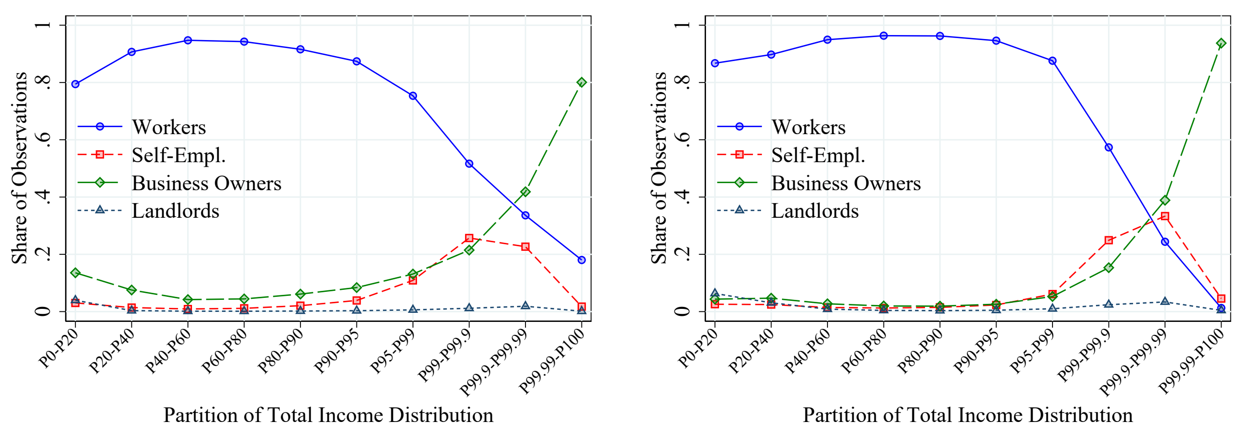


Notes: This figure shows the the P90-P50 and P50-P10 differentials of the distribution of 1-year changes in residualized log earnings (from $t-1$ to t) in the IAB data (LS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. LS sample with minimum income threshold of 6,250 Euro (2018 prices). The LS sample in the main text uses 2,300 Euro as cutoff to include mini-jobs. Shaded areas indicate recessions. The results for our core sample can be found in Figure 7.

G Specific Analysis: Additional Figures and Tables on Total Income Inequality and Dynamics (Section 4)

Figure G.1: Main Income Sources Across the Income Distribution

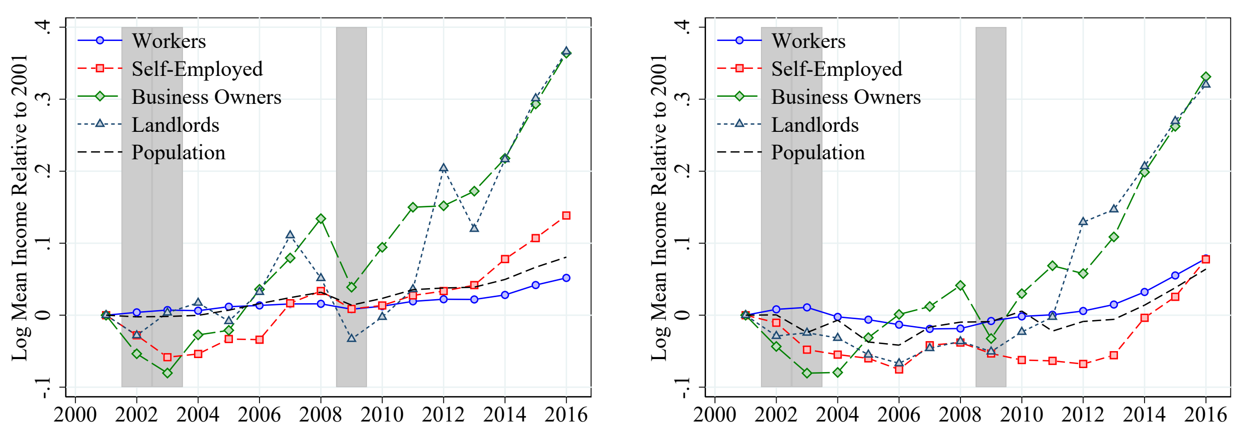
(a) Men (b) Women



Notes: This figure shows the share of observations with different main income source for different groups of the total income distribution in the combined IAB-TPP data (CS analysis sample). The figure shows averages from 2001 to 2016.

Figure G.2: Evolution of Log Average Income by Main Income Source

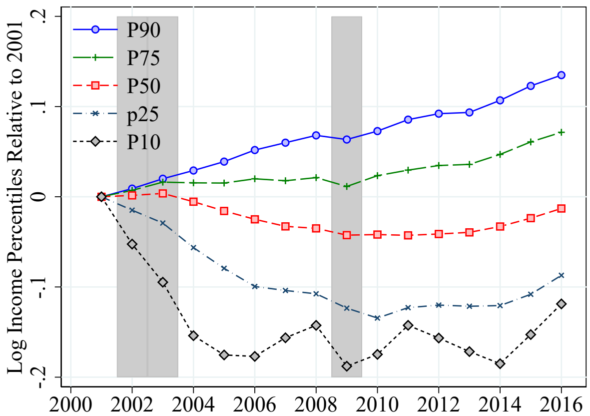
(a) Men (b) Women



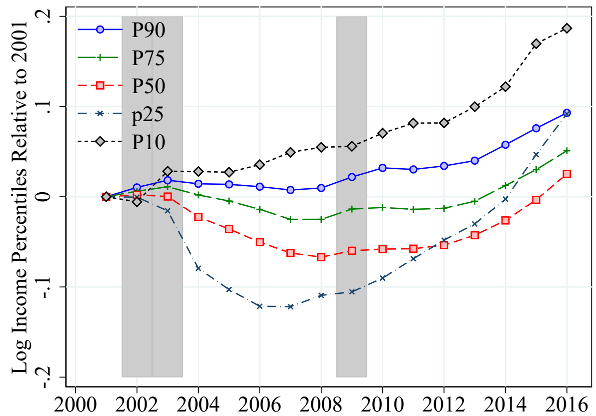
Notes: This figure shows the evolution the log of average real annual total income (relative to 2001) in the combined IAB-TPP data (CS analysis sample) by main income source separately for men and women. Shaded areas indicate recessions. See Figure 11 for corresponding levels.

Figure G.3: Evolution of Log Total Income Percentiles

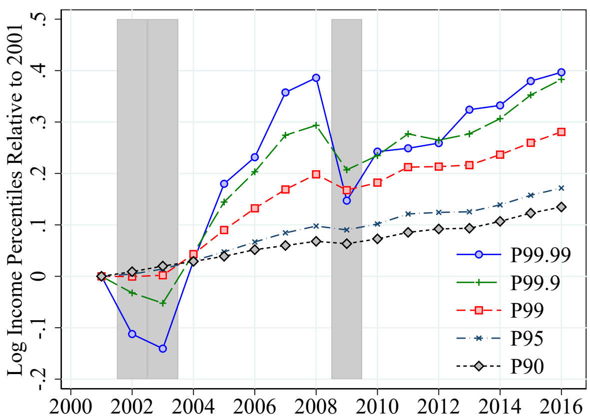
(a) Overall Distribution: Men



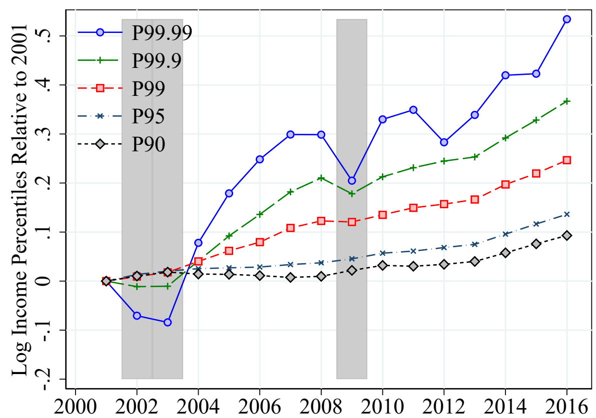
(b) Overall Distribution: Women



(c) Top Percentiles: Men



(d) Top Percentiles: Women



Notes: This figure shows the evolution of selected percentiles of log real annual total income (relative to 2001) in the combined IAB-TPP data (CS analysis sample) separately for men and women. Shaded areas indicate recessions. See Figure 3 for the same analysis of only labor earnings (albeit for a slightly different sample as discussed in the text).

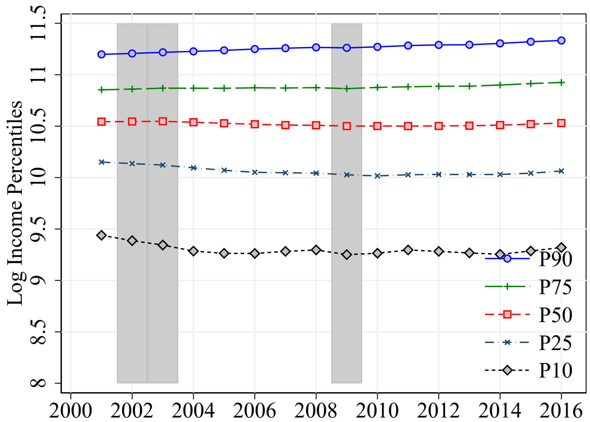
Table G.1: Percentiles of Real Annual Total Income (Analysis Sample)

Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
Men												
2001	15.275	44,979	7,245	12,568	25,608	37,924	51,695	72,973	93,531	176,591	519,341	1,984,431
2002	15.029	44,878	6,815	11,925	25,230	37,982	52,064	73,634	93,941	176,495	502,857	1,773,707
2003	14.763	44,900	6,382	11,432	24,872	38,065	52,541	74,445	94,863	176,950	492,898	1,724,135
2004	14.635	44,969	6,013	10,774	24,204	37,717	52,500	75,128	96,410	184,374	542,735	2,044,026
2005	14.451	45,293	5,925	10,546	23,655	37,332	52,487	75,872	98,073	193,222	600,038	2,375,527
2006	14.509	45,707	5,923	10,529	23,184	36,989	52,734	76,852	99,989	201,573	636,386	2,501,665
2007	14.654	46,088	6,017	10,749	23,081	36,700	52,622	77,480	101,779	209,085	683,272	2,837,472
2008	14.667	46,438	6,077	10,898	22,996	36,620	52,804	78,113	103,133	215,338	696,522	2,919,252
2009	14.393	45,611	5,873	10,415	22,633	36,345	52,289	77,753	102,367	208,794	638,749	2,299,090
2010	14.528	46,041	5,968	10,553	22,386	36,368	52,918	78,483	103,556	211,896	656,857	2,528,981
2011	14.697	46,602	6,105	10,898	22,646	36,338	53,245	79,487	105,586	218,373	685,103	2,545,453
2012	14.739	46,710	5,978	10,746	22,710	36,391	53,517	80,013	105,916	218,586	676,732	2,571,310
2013	14.774	46,763	5,935	10,586	22,681	36,457	53,583	80,124	106,038	219,216	685,103	2,743,865
2014	14.853	47,267	5,883	10,445	22,698	36,695	54,179	81,205	107,470	223,726	705,679	2,766,737
2015	14.931	48,079	6,036	10,787	22,984	37,035	54,936	82,514	109,496	228,932	738,823	2,900,730
2016	14.955	48,756	6,237	11,160	23,472	37,435	55,525	83,507	111,041	233,868	761,639	2,950,836
Women												
2001	12.389	27,165	3,959	4,801	12,037	23,199	35,910	47,924	56,978	89,493	219,459	690,697
2002	12.365	27,177	3,946	4,774	12,032	23,251	36,131	48,424	57,756	90,323	216,999	643,630
2003	12.193	26,518	3,918	4,939	11,850	23,205	36,310	48,812	58,150	91,108	217,184	634,949
2004	12.176	26,978	3,890	4,937	11,117	22,686	35,983	48,619	58,446	93,163	228,673	746,738
2005	12.116	26,173	3,861	4,933	10,860	22,385	35,740	48,588	58,541	95,198	240,643	826,045
2006	12.152	26,057	3,879	4,975	10,661	22,061	35,413	48,463	58,635	96,886	251,420	885,390
2007	12.307	26,724	3,976	5,044	10,656	21,798	35,021	48,286	58,936	99,744	263,245	931,383
2008	12.351	26,904	4,011	5,072	10,791	21,698	35,022	48,393	59,147	101,178	270,828	931,066
2009	12.370	26,924	4,034	5,077	10,832	21,852	35,425	48,985	59,626	100,961	262,240	847,725
2010	12.469	27,321	4,036	5,152	10,999	21,892	35,487	49,484	60,318	102,453	271,484	960,557
2011	12.603	26,571	4,063	5,209	11,237	21,901	35,414	49,399	60,573	103,927	276,565	979,455
2012	12.717	26,929	4,092	5,211	11,472	21,987	35,453	49,590	61,010	104,702	280,337	916,708
2013	12.777	27,007	4,174	5,305	11,680	22,231	35,733	49,883	61,411	105,697	282,672	969,343
2014	12.839	27,547	4,222	5,424	12,005	22,600	36,360	50,770	62,705	108,991	293,866	1,051,058
2015	12.902	28,220	4,376	5,689	12,614	23,120	37,006	51,705	64,019	111,466	304,776	1,054,335
2016	12.881	28,961	4,481	5,787	13,195	23,794	37,791	52,596	65,294	114,521	316,734	1,178,313
Population												
2001	27.664	36,145	4,568	7,263	17,701	31,829	44,927	62,387	79,529	144,849	410,730	1,508,540
2002	27.394	36,609	4,513	7,062	17,455	31,776	45,161	62,912	80,102	144,686	398,651	1,381,143
2003	26.956	36,115	4,533	6,657	17,198	31,738	45,465	63,461	80,911	145,553	393,548	1,331,338
2004	26.810	36,049	4,493	6,036	16,534	31,230	45,208	63,737	81,904	150,171	425,888	1,535,035
2005	26.568	36,174	4,462	5,940	16,219	30,766	45,058	64,007	82,858	155,886	466,683	1,777,965
2006	26.661	36,643	4,481	5,953	15,975	30,335	44,984	64,513	84,192	161,680	494,922	1,937,037
2007	26.961	36,483	4,561	6,063	15,918	29,989	44,748	64,767	85,245	167,071	523,091	2,110,520
2008	27.018	36,669	4,569	6,200	15,920	29,845	44,765	65,127	86,028	171,488	536,687	2,156,058
2009	26.763	36,202	4,555	6,115	15,725	29,703	44,681	64,881	85,409	167,327	498,942	1,748,128
2010	26.997	36,535	4,589	6,304	15,758	29,625	45,039	65,579	86,369	169,905	512,422	1,899,162
2011	27.300	37,621	4,628	6,548	16,019	29,600	45,090	66,096	87,563	174,503	532,829	1,969,698
2012	27.457	37,603	4,622	6,657	16,047	29,617	45,157	66,567	88,123	174,675	524,838	1,902,555
2013	27.551	37,129	4,711	6,718	16,145	29,699	45,325	66,733	88,336	175,007	529,973	1,983,229
2014	27.692	37,648	4,743	6,858	16,335	29,988	45,892	67,687	89,658	178,723	545,531	2,056,196
2015	27.833	39,026	4,943	7,293	16,932	30,356	46,554	68,856	91,266	183,152	566,431	2,207,568
2016	27.837	39,960	5,073	7,716	17,463	30,972	47,181	69,869	92,599	186,681	584,788	2,300,887

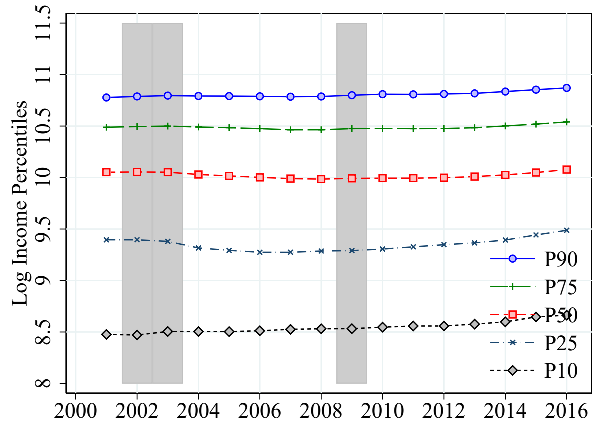
Notes: This table shows the number of observations (in millions) and selected percentiles of real annual total income (in 2018 Euro) in the combined IAB-TPP data (CS analysis sample) separately for men and women and in the population. See Table 1 for the percentiles of labor earnings (albeit for a slightly different sample, as discussed in the text).

Figure G.4: Percentiles of Log Income

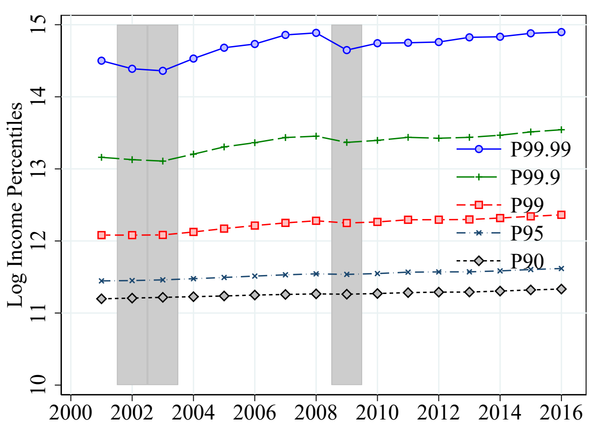
(a) Overall Distribution: Men



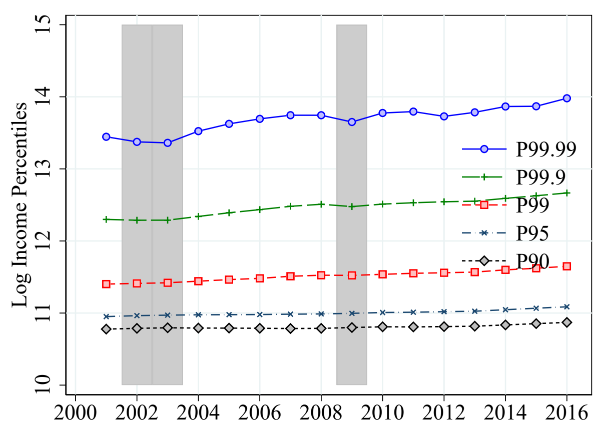
(b) Overall Distribution: Women



(c) Top Percentiles: Men

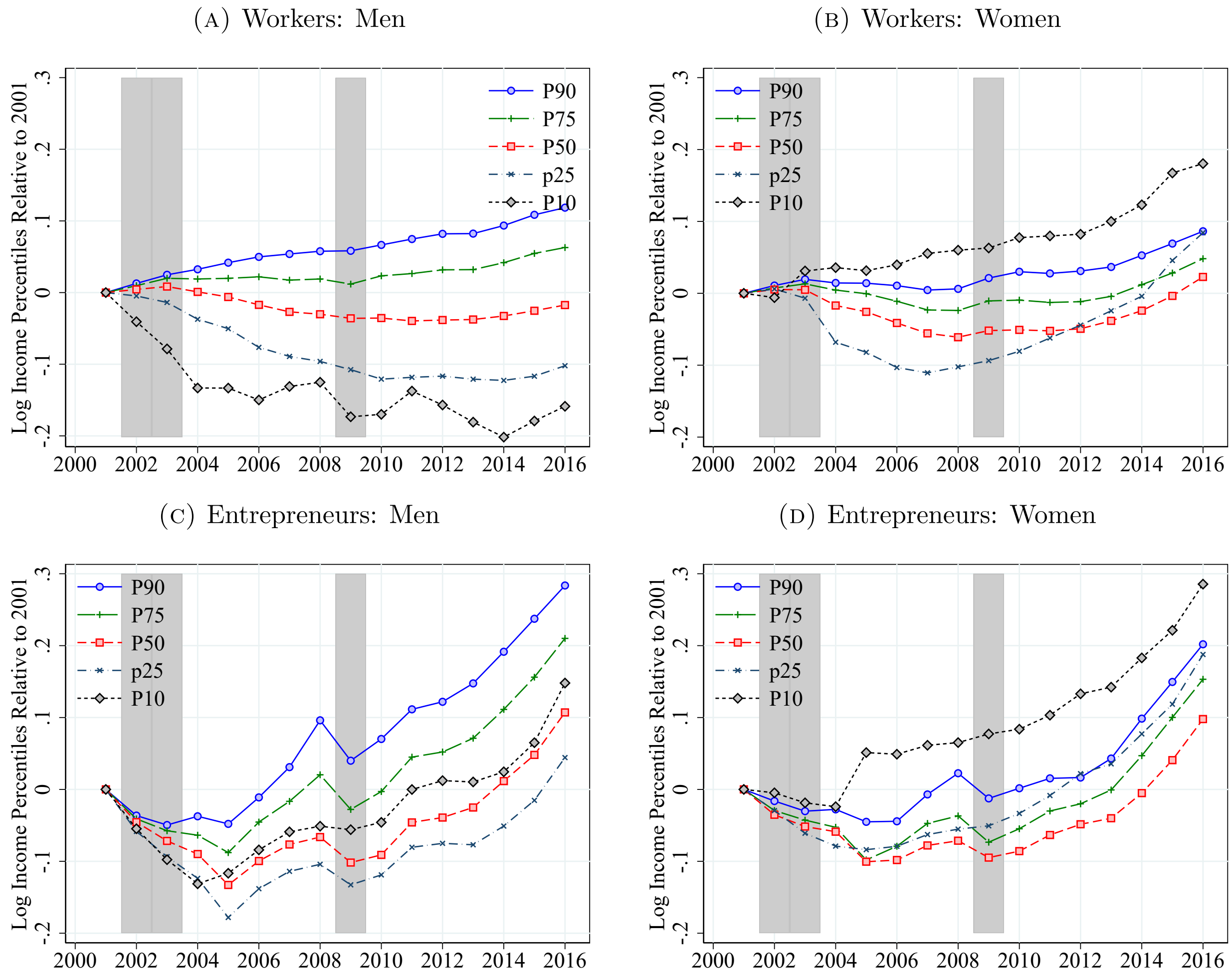


(d) Top Percentiles: Women



Notes: This figure shows the evolution of absolute log real annual total income percentiles in the combined IAB-TPP data (CS sample) separately for men and women. Shaded areas indicate recessions.

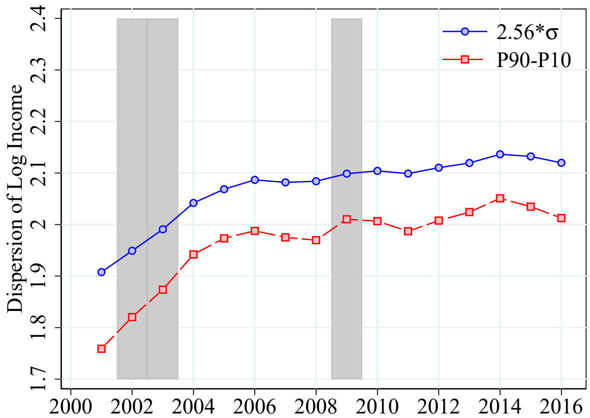
Figure G.5: Evolution of Log Income Percentiles by Main Income Source



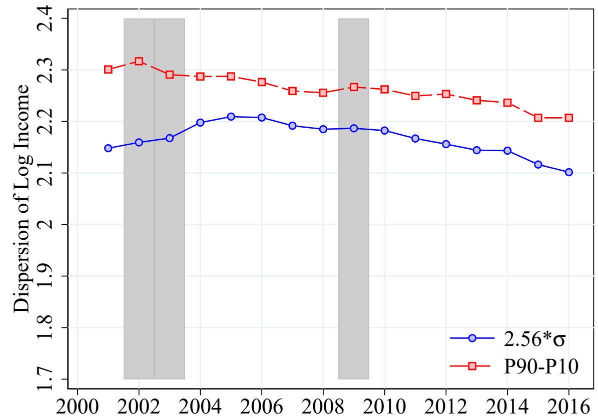
Notes: This figure shows the evolution of different percentiles of log total income among workers and entrepreneurs in the combined IAB-TPP data (CS sample). Workers receive at least half of their income from labor earnings.The jump in the P10 for entrepreneurs (while it is obvious for women, it is hidden for men) is related to a similar jump in the number of observations classified as landlords from 2004 to 2005 which is plausibly related to a reform in the taxation of pensions. In line with this, the jump is entirely driven by landlords (as opposed to self-employed or business owners). Shaded areas indicate recessions.

Figure G.6: Dispersion of Log Real Income Distribution

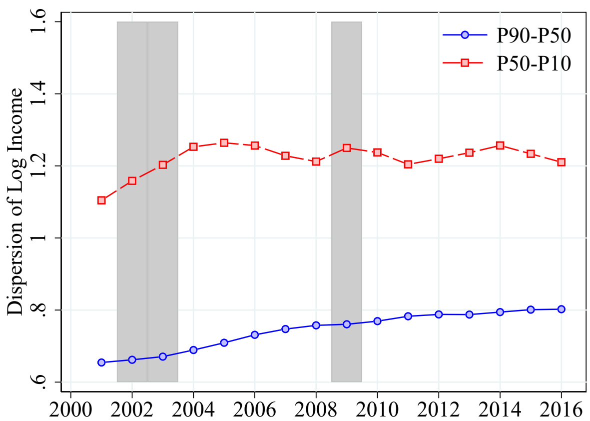
(a) Overall Inequality: Men



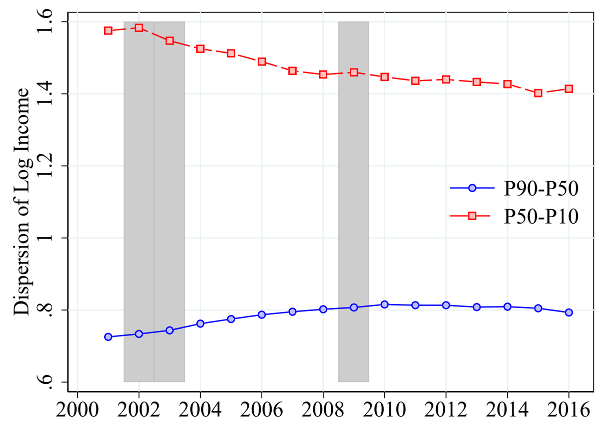
(b) Overall Inequality: Women



(c) Upper and & Lower Inequality: Men



(d) Upper and & Lower Inequality: Women



Notes: This figure shows the evolution of different log percentile differentials as well as the (rescaled) standard deviation of the log real annual total income distribution over time in the combined IAB-TPP data (CS sample) separately for men and women. The standard deviation σ is rescaled as $2.56 * \sigma$ corresponds to P90-P10 differential for a Gaussian distribution. Shaded areas indicate recessions.

Table G.2: Income Shares – Men

Year	Q1	Q2	Q3	Q4	Q5	Bot 50	Bot 90	Mid 40	Top 10	Top 5	Top 1	Top 0.1	Top 0.01
2001	5.62	12.86	17.16	21.92	42.44	26.58	71.91	45.32	28.09	18.86	8.09	2.85	1.14
2002	5.40	12.76	17.22	22.09	42.52	26.29	71.97	45.69	28.03	18.71	7.91	2.72	1.11
2003	5.21	12.64	17.26	22.26	42.63	25.97	72.01	46.03	27.99	18.57	7.70	2.54	0.98
2004	4.94	12.28	16.99	22.07	43.73	25.18	70.89	45.71	29.11	19.63	8.54	3.09	1.36
2005	4.79	11.95	16.71	21.89	44.67	24.56	69.94	45.38	30.06	20.52	9.19	3.42	1.46
2006	4.71	11.64	16.42	21.75	45.47	24.01	69.14	45.13	30.86	21.24	9.67	3.65	1.55
2007	4.74	11.45	16.15	21.48	46.18	23.72	68.37	44.65	31.63	21.98	10.21	3.95	1.70
2008	4.75	11.33	16.01	21.40	46.51	23.53	68.02	44.49	31.98	22.29	10.33	3.92	1.65
2009	4.68	11.40	16.19	21.62	46.11	23.62	68.63	45.01	31.37	21.56	9.58	3.37	1.36
2010	4.65	11.19	16.06	21.65	46.45	23.30	68.30	45.00	31.70	21.88	9.85	3.57	1.47
2011	4.73	11.13	15.86	21.49	46.79	23.22	67.93	44.71	32.07	22.21	10.02	3.60	1.45
2012	4.69	11.13	15.85	21.52	46.82	23.16	67.97	44.80	32.03	22.14	9.95	3.59	1.48
2013	4.64	11.10	15.87	21.56	46.84	23.09	67.97	44.88	32.03	22.12	9.92	3.51	1.35
2014	4.56	11.01	15.81	21.56	47.06	22.89	67.78	44.88	32.22	22.29	10.02	3.53	1.33
2015	4.60	10.91	15.66	21.44	47.39	22.75	67.40	44.65	32.60	22.69	10.41	3.82	1.54
2016	4.66	10.94	15.60	21.36	47.44	22.81	67.31	44.50	32.69	22.79	10.49	3.87	1.54

Notes: This table shows the share of (total) income that goes to selected parts of the income distribution of men in the combined IAB-TPP data (CS sample). Q1 to Q5 refer to the five quintiles where Q1 (Q5) stands for the bottom (top) 20% of the income distribution. The quintile shares sum to one. Bot 50, Bot 90 and Mid 40 refer to observations in the bottom 50%, the bottom 90% and between the median and the 90th percentile of the income distribution. Top x refers to the top x% of the income distribution.

Table G.3: Income Shares – Women

Year	Q1	Q2	Q3	Q4	Q5	Bot 50	Bot 90	Mid 40	Top 10	Top 5	Top 1	Top 0.1	Top 0.01
2001	3.96	10.81	17.63	25.25	42.34	22.71	73.94	51.24	26.06	16.22	6.03	1.93	0.82
2002	3.92	10.77	17.58	25.25	42.48	22.60	73.86	51.26	26.14	16.23	5.97	1.87	0.82
2003	3.89	10.64	17.54	25.33	42.60	22.41	73.84	51.43	26.16	16.19	5.86	1.74	0.68
2004	3.76	10.23	17.31	25.28	43.41	21.76	73.10	51.34	26.90	16.83	6.26	1.96	0.80
2005	3.72	10.06	17.12	25.13	43.98	21.45	72.53	51.08	27.47	17.37	6.69	2.24	0.98
2006	3.74	9.92	16.94	24.96	44.44	21.24	72.04	50.80	27.96	17.83	7.01	2.39	1.07
2007	3.81	9.88	16.77	24.72	44.81	21.20	71.60	50.40	28.40	18.26	7.26	2.46	1.05
2008	3.83	9.92	16.66	24.62	44.96	21.22	71.44	50.22	28.56	18.40	7.34	2.45	1.03
2009	3.86	9.90	16.68	24.74	44.82	21.23	71.71	50.48	28.29	18.08	7.03	2.21	0.88
2010	3.88	9.91	16.56	24.57	45.08	21.21	71.41	50.20	28.59	18.36	7.26	2.38	0.96
2011	3.95	10.01	16.50	24.40	45.13	21.37	71.25	49.89	28.75	18.55	7.41	2.49	1.05
2012	4.00	10.11	16.50	24.33	45.07	21.52	71.28	49.76	28.72	18.51	7.31	2.37	0.96
2013	4.04	10.16	16.51	24.27	45.02	21.62	71.27	49.65	28.73	18.57	7.40	2.45	1.03
2014	4.05	10.16	16.42	24.15	45.22	21.60	70.98	49.38	29.02	18.89	7.70	2.68	1.22
2015	4.17	10.36	16.40	24.00	45.08	21.92	71.02	49.10	28.98	18.90	7.74	2.69	1.21
2016	4.24	10.47	16.45	23.93	44.90	22.14	71.08	48.94	28.92	18.91	7.79	2.73	1.21

Notes: This table shows the share of (total) income that goes to selected parts of the income distribution of women in the combined IAB-TPP data (CS sample). Q1 to Q5 refer to the five quintiles where Q1 (Q5) stands for the bottom (top) 20% of the income distribution. The quintile shares sum to one. Bot 50, Bot 90 and Mid 40 refer to observations in the bottom 50%, the bottom 90% and between the median and the 90th percentile of the income distribution. Top x refers to the top x% of the income distribution.

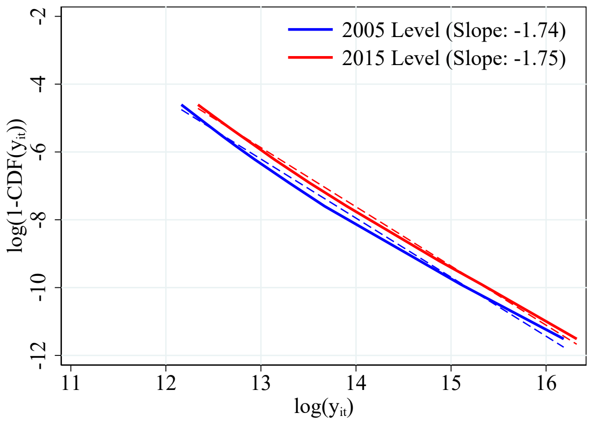
Table G.4: Income Shares – Population

Year	Q1	Q2	Q3	Q4	Q5	Bot 50	Bot 90	Mid 40	Top 10	Top 5	Top 1	Top 0.1	Top 0.01
2001	4.27	11.37	17.52	23.23	43.61	23.70	71.48	47.77	28.52	18.91	7.87	2.71	1.10
2002	4.18	11.25	17.50	23.35	43.72	23.47	71.50	48.03	28.50	18.79	7.72	2.60	1.07
2003	4.07	11.12	17.48	23.48	43.85	23.20	71.50	48.30	28.50	18.71	7.54	2.42	0.93
2004	3.84	10.77	17.19	23.33	44.88	22.46	70.49	48.03	29.51	19.63	8.24	2.87	1.22
2005	3.75	10.52	16.89	23.13	45.72	21.96	69.61	47.64	30.39	20.47	8.89	3.22	1.38
2006	3.72	10.30	16.58	22.93	46.47	21.56	68.85	47.29	31.15	21.16	9.35	3.45	1.47
2007	3.75	10.19	16.32	22.67	47.08	21.35	68.16	46.81	31.84	21.82	9.81	3.69	1.58
2008	3.77	10.13	16.17	22.55	47.37	21.24	67.86	46.61	32.14	22.09	9.95	3.68	1.54
2009	3.78	10.16	16.32	22.81	46.93	21.34	68.43	47.09	31.57	21.44	9.28	3.19	1.27
2010	3.80	10.06	16.14	22.73	47.28	21.16	68.11	46.95	31.89	21.74	9.54	3.38	1.37
2011	3.86	10.08	15.98	22.52	47.57	21.18	67.77	46.59	32.23	22.07	9.74	3.45	1.38
2012	3.88	10.08	15.96	22.50	47.59	21.19	67.80	46.61	32.20	22.00	9.65	3.39	1.39
2013	3.90	10.10	15.96	22.51	47.54	21.23	67.83	46.60	32.17	21.97	9.64	3.36	1.32
2014	3.89	10.06	15.88	22.45	47.72	21.15	67.64	46.49	32.36	22.16	9.79	3.45	1.36
2015	3.98	10.11	15.76	22.30	47.86	21.24	67.43	46.19	32.57	22.40	10.05	3.65	1.50
2016	4.06	10.20	15.76	22.20	47.78	21.42	67.44	46.01	32.56	22.44	10.11	3.69	1.50

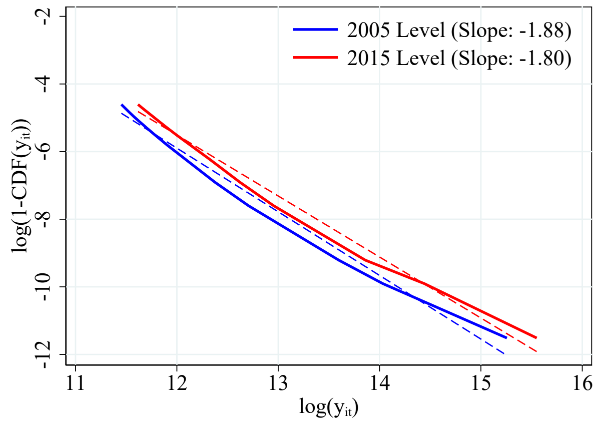
Notes: This table shows the share of (total) income that goes to selected parts of the income distribution in the combined IAB-TPP data (CS sample). Q1 to Q5 refer to the five quintiles where Q1 (Q5) stands for the bottom (top) 20% of the income distribution. The quintile shares sum to one. Bot 50, Bot 90 and Mid 40 refer to observations in the bottom 50%, the bottom 90% and between the median and the 90th percentile of the income distribution. Top x refers to the top x% of the income distribution.

Figure G.7: Top Income Inequality: Pareto Tail at Top 1% and Top 5%

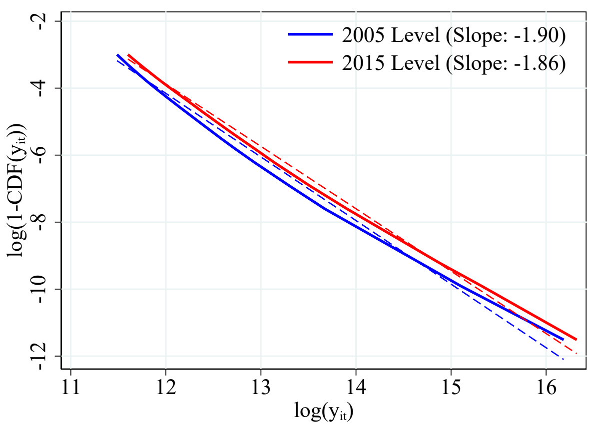
(a) Top 1%: Men



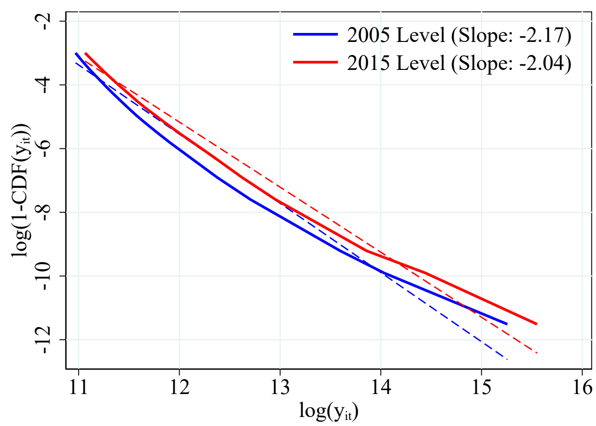
(b) Top 1%: Women



(c) Top 5%: Men



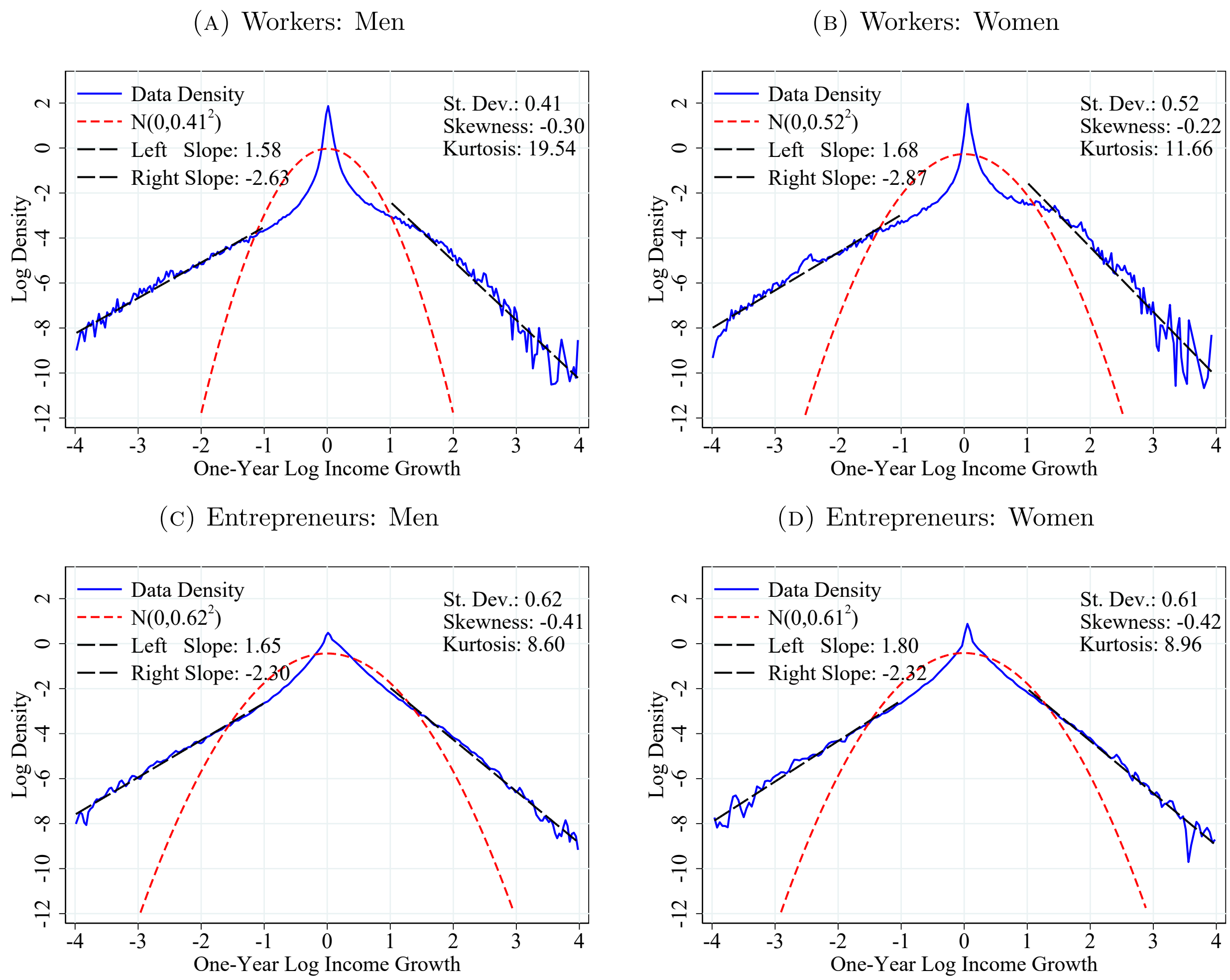
(d) Top 5%: Women



Notes: This figure shows the log of the inverse empirical CDF of log total income and a fitted linear regression line for observations with income in the top 1% and top 5% in the combined IAB-TPP data (CS sample). The absolute value of the slope of the regression line is the Pareto parameter above the respective cutoff.

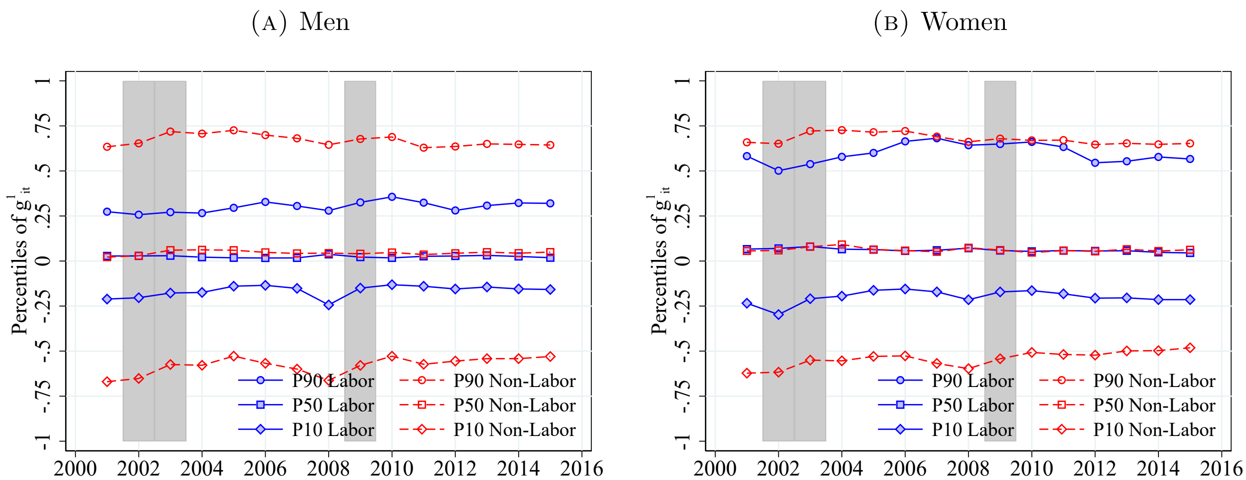
Figure G.8: Log Density of 1-Year Income Growth by Main Income Source (Year 2005)

(a) Workers: Men — (b) Workers: Women — (c) Entrepreneurs: Men — (d) Entrepreneurs: Women



Notes: This figure shows the log density of 1-year changes of residualized log total income separately for workers (labor income as main income source) and entrepreneurs (non-labor income as main income source) and for men and women in the year 2005. LS sample of the combined IAB-TPP data. The dashed line corresponds to the log density of a Normal distribution with the same variance.

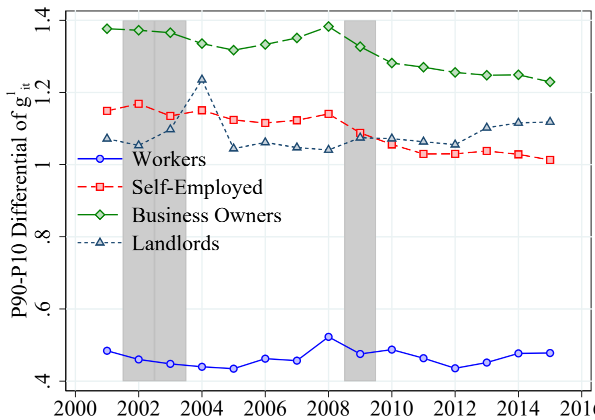
FIGURE G.9: PERCENTILES OF 1-YEAR INCOME GROWTH BY MAIN INCOME SOURCE



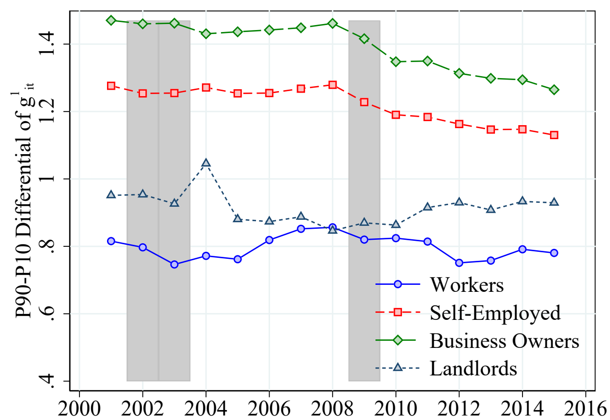
Notes: This figure the 90th, 50th and 10th percentiles of the distribution of 1-year changes in residualized log income (from $t-1$ to t) by main income source (workers vs. entrepreneurs) using the combined IAB-TPP data (LS sample).

Figure G.10: Dispersion, Skewness and Kurtosis of 1-Year Log Income Changes

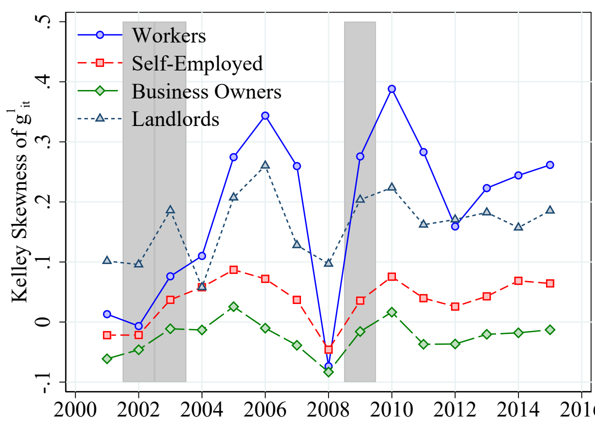
(a) P90–P10: Men



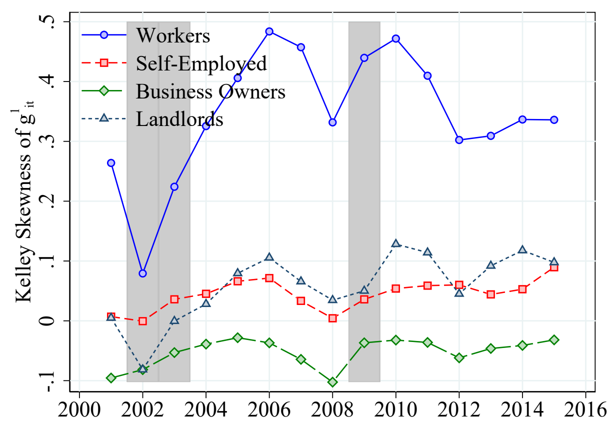
(b) P90–P10: Women



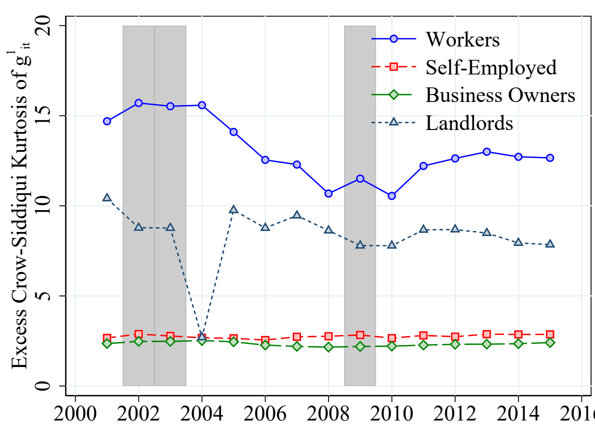
(c) Kelley Skewness: Men



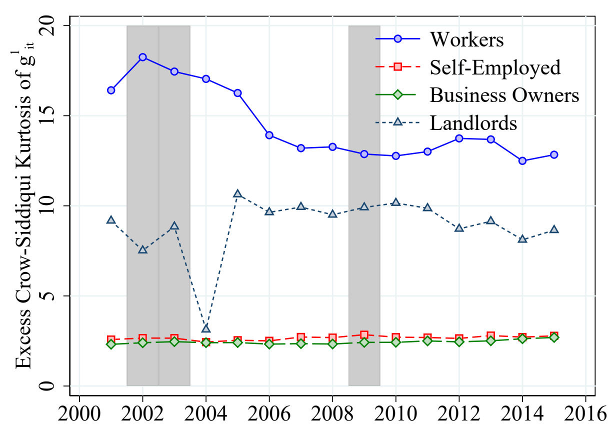
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



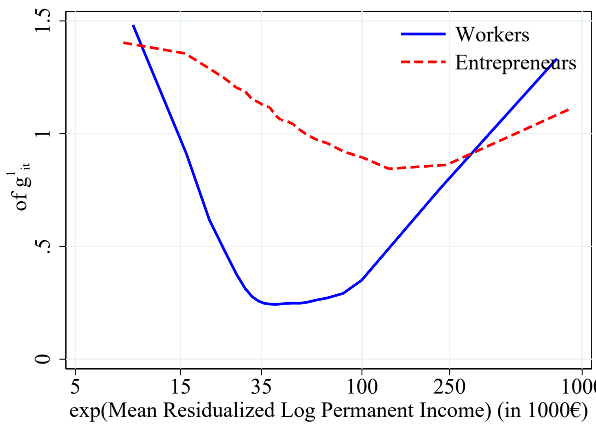
(f) Excess Crow-Siddiqui Kurtosis: Women



Notes: This figure shows the evolution of the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real annual total income (from $t-1$ to t) in the combined IAB-TPP data (LS sample) separately for men and women by main income source (workers, self-employment, business owners, landlords). See Footnote 23 definitions and interpretation of Kelley skewness and excess Crow-Siddiqui kurtosis. Shaded areas indicate recessions.

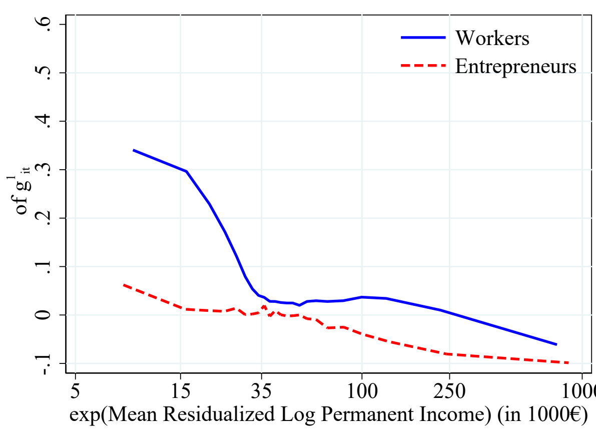
Figure G.11: Heterogeneity in Dispersion, Skewness and Kurtosis of 1-Year Log Income Growth by Main Income Source

(a) P90-P10: Men

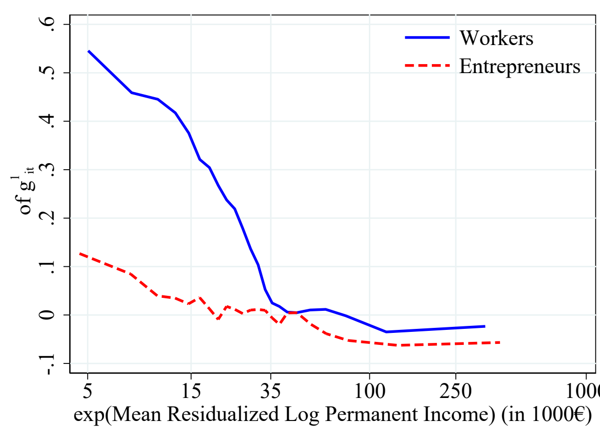


(b) P90-P10: Women

(c) Kelley Skewness: Men



(d) Kelley Skewness: Women



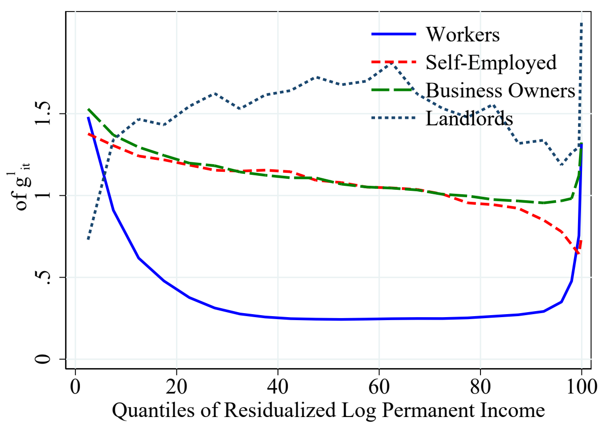
(e) Excess Crow-Siddiqui Kurtosis: Men

(f) Excess Crow-Siddiqui Kurtosis: Women

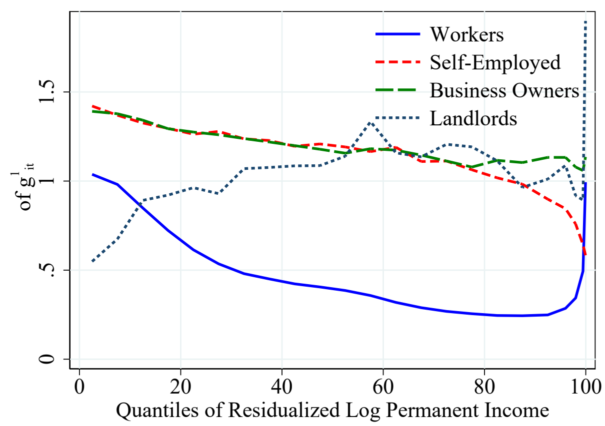
Notes: This figure shows the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real total income by permanent total income (from $t-1$ to t) in the combined IAB-TPP data (H Sample) as averages from 2004 to 2011 and separately for men and women by main income source (workers, self-employment, business owners, landlords). The horizontal axis plots the exponential of mean permanent income in 1,000 Euro. See Footnote 23 definitions and interpretation of Kelley skewness and excess Crow-Siddiqui kurtosis. Shaded areas indicate recessions.

Figure G.12: Heterogeneity in Dispersion, Skewness and Kurtosis of 1-Year Log Income Growth by Main Income Source

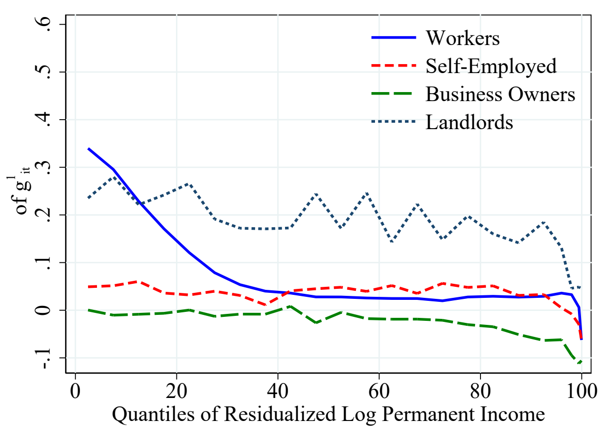
(a) P90-P10: Men



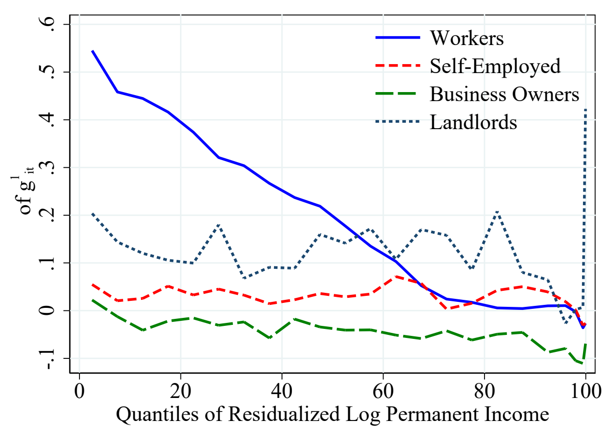
(b) P90-P10: Women



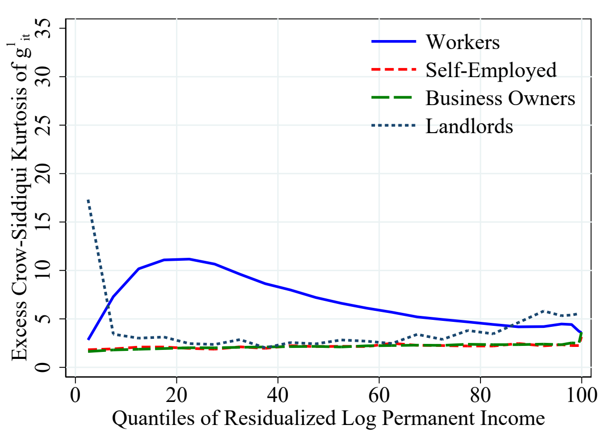
(c) Kelley Skewness: Men



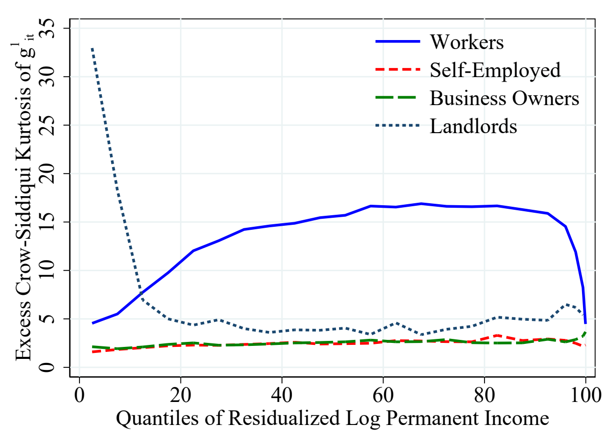
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



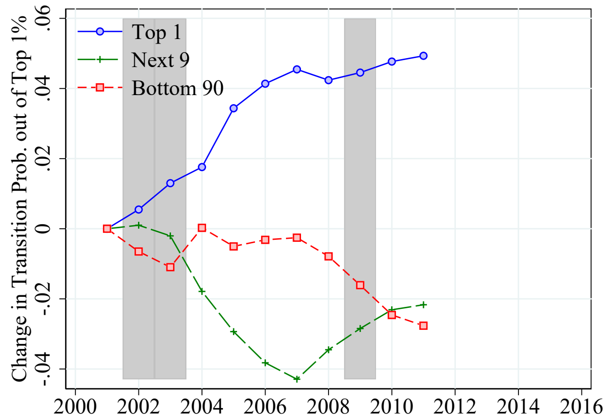
(f) Excess Crow-Siddiqui Kurtosis: Women



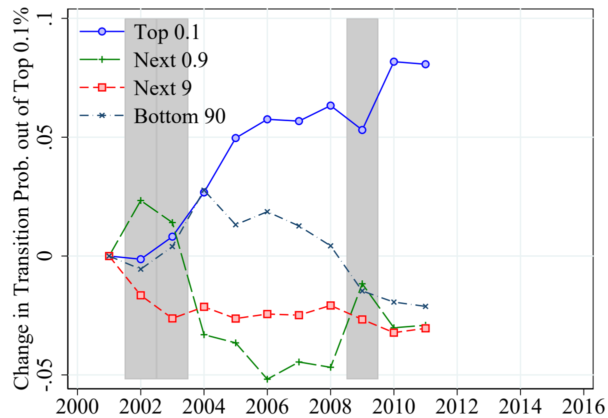
Notes: This figure shows the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real total income by quantiles of the distribution of permanent total income (from $t-1$ to t) in the combined IAB-TPP data (H Sample) as averages from 2004 to 2011 and separately for men and women by main income source (workers, self-employment, business owners, landlords). The (gender-specific) ranking of permanent income is based on the distribution of total income of all taxpayers. See Footnote 23 definitions and interpretation of Kelley skewness and excess Crow-Siddiqui kurtosis. Shaded areas indicate recessions.

Figure G.13: Top Income Mobility – 5-Year Transition Probabilities

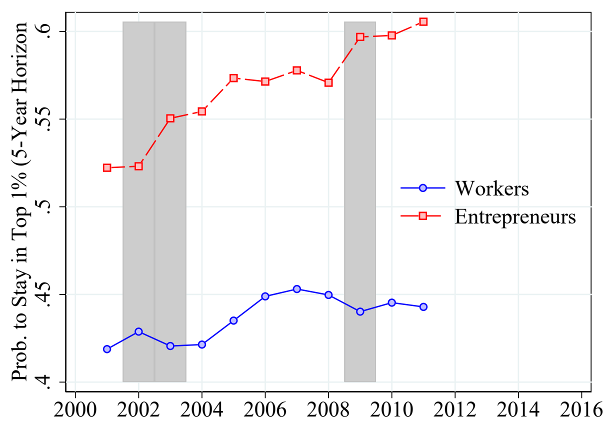
(a) Transitions out of the Top 1%



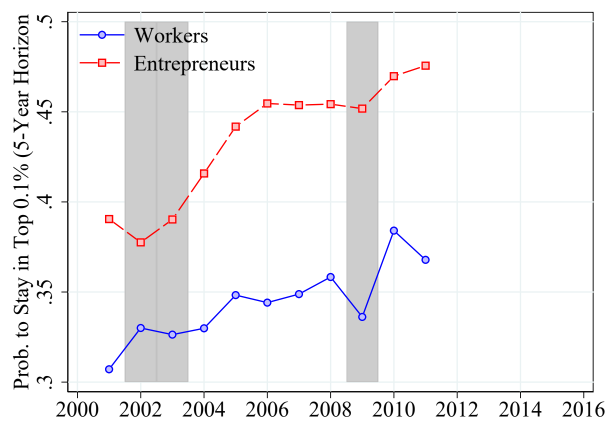
(b) Transitions out of the Top 0.1%



(c) Probability to Stay in the Top 1%



(d) Probability to Stay in the Top 0.1%



Notes: This figure plots transition probabilities from top income using the combined IAB-TPP data (LS sample). Panels A and B show the evolution of 5-year transition probabilities out of the top 1% and top 0.1% of the income distribution into selected parts of the income distribution from one year to the next. The "Next 9" is the part of the distribution between the P90 and P99 and the "Next 0.9" is the part between the P99 and the P99.9. The lines sum to zero. Panels C and D show the 5-year probability of staying in the top 1% or top 0.1% for workers and entrepreneurs. The ranking is based on the total income distribution and not conditional on the main income source. Shaded areas indicate recessions.