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 **44**
A.1 Institutions . . . 44
A.2 Macroeconomic Background . . . 47

B IAB: Social Security Data **49**
B.1 Top-Coding and Imputation of Wages . . . 49
B.2 Imputation of the Education Indicator . . . 50
B.3 Correction of the Full-Time Indicator . . . 50

C TPP: Tax Data **52**
C.1 General Description . . . 52
C.2 Imputation of Social Security Indicator in Pre-2010 TPP Data . . . 53

D Combined IAB-TPP Data **55**
D.1 Reweighting the TPP Data to Match the IAB Data . . . 55
D.2 Combined IAB-TPP Data in Earnings Analysis (Section 3) . . . 59
D.3 Earnings Growth by Permanent Earnings . . . 71
D.4 Combined IAB-TPP Data in Total Income Analysis (Section 4) . . . 80

E Core Analysis: Additional Results for Combined IAB-TPP Data 2001–2016 **87**
E.1 Additional Results for Earnings Inequality (Section 3.1) . . . 87
E.2 Details on Reweighting Analysis (Section 3.1) . . . 101
E.3 Additional Results for Earnings Dynamics (Section 3.2) . . . 109

F Core Analysis of Earnings with Longer Samples **118**
F.1 IAB Data 1993–2018 . . . 118
F.2 IAB Data 1985–2018 (West Germany) . . . 121

G Specific Analysis: Additional Figures and Tables on Total Income Inequality and Dynamics (Section 4) **126**

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

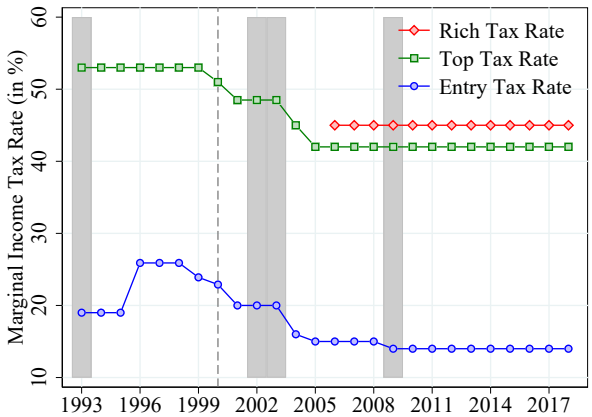
[48] A person can hold multiple mini-jobs but then only the first 450 Euro are tax exempt.

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.

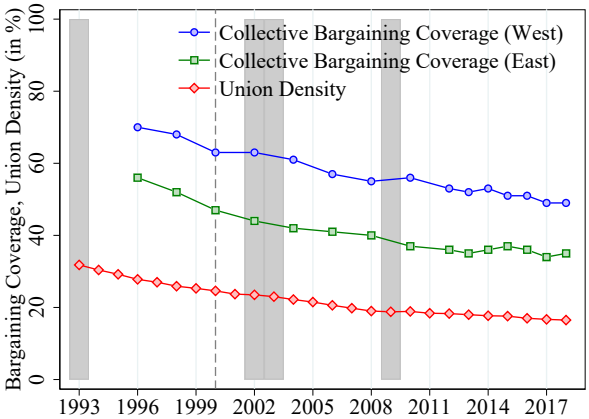
Social Security Contribution Limits. The contributions to the pension system and unemployment insurance are capped. The limit differs between East and West Germany and increases over time, roughly following inflation. Figure A.1 (D) shows the limits for East and West Germany from 1993 to 2018 in real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

Figure A.1: Institutional Background

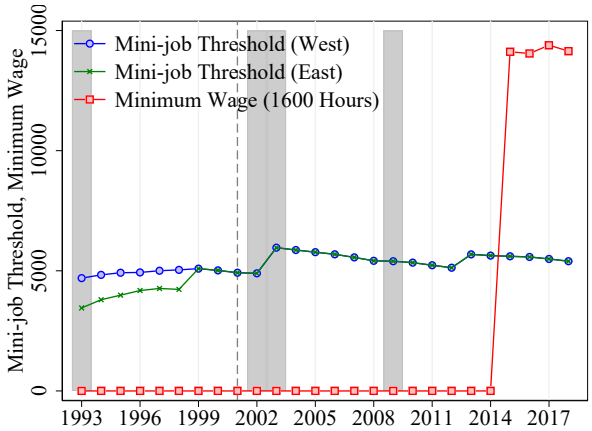
(a) Entry & Top Tax Rates



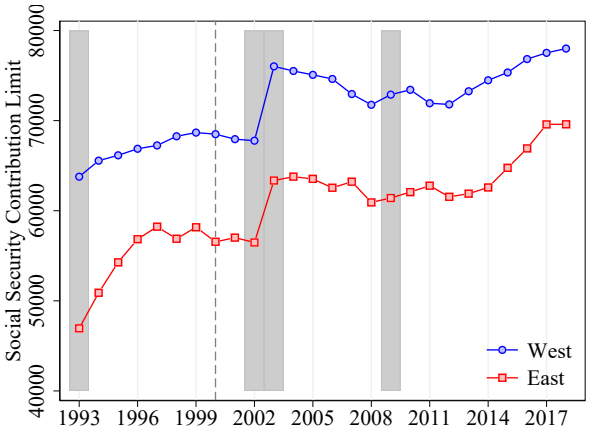
(b) Collective Bargaining Coverage and Union Density



(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

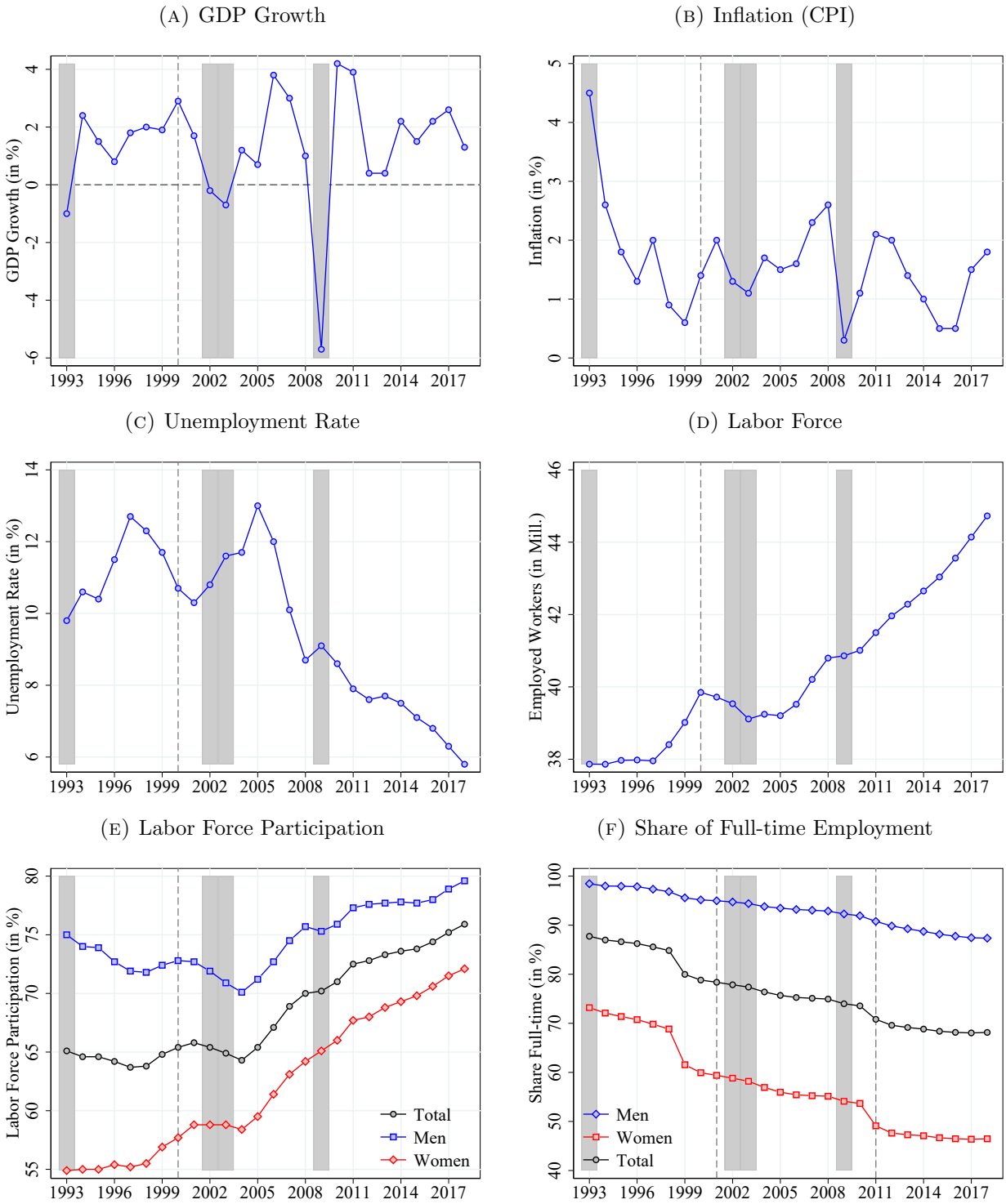
The macroeconomic development in Germany from 1993-2018 can be broadly split into two periods: before and after 2005 (see Figure A.2). The first time span was characterized by low growth and high unemployment (above 10%) and Germany was often referred to as "the sick man of Europe" (Dustmann et al., 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.

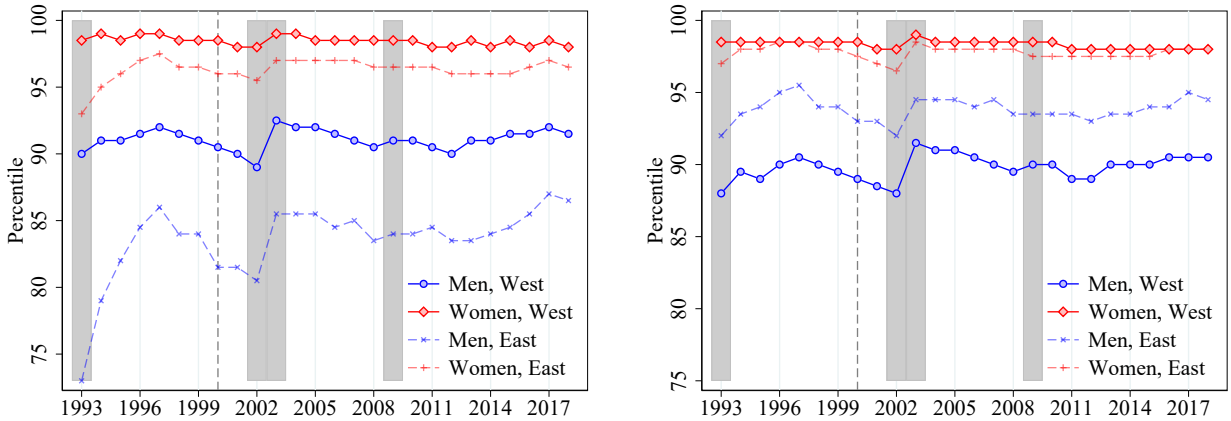
In Figure B.1 (B) we ask the related but different question, where within the earnings distribution of East and West German workers the social security contribution limits fall in a given year by gender. Since East German incomes are still much lower than in West Germany, this pushes up at which percentile 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 80th percentile of earnings in 2012 for women and 25th 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.

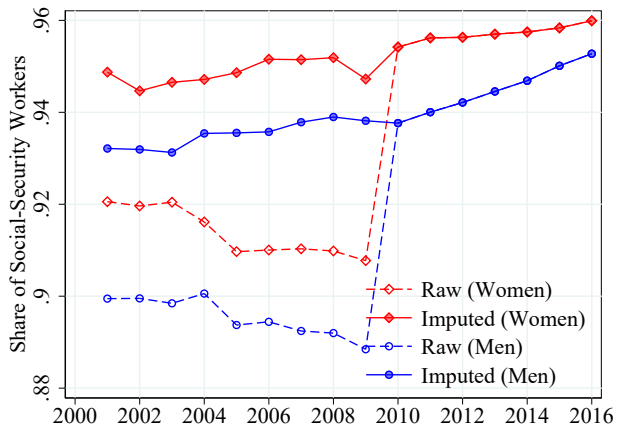
The 2001 to 2011 waves of the Taxpayer Panel (TPP) were compiled on the basis of annual income tax statistics (*Geschäftsstatistik*) of each of the 16 federal states which were then combined into one dataset for Germany. These cross-sectional data contain information from the income tax returns of around 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:

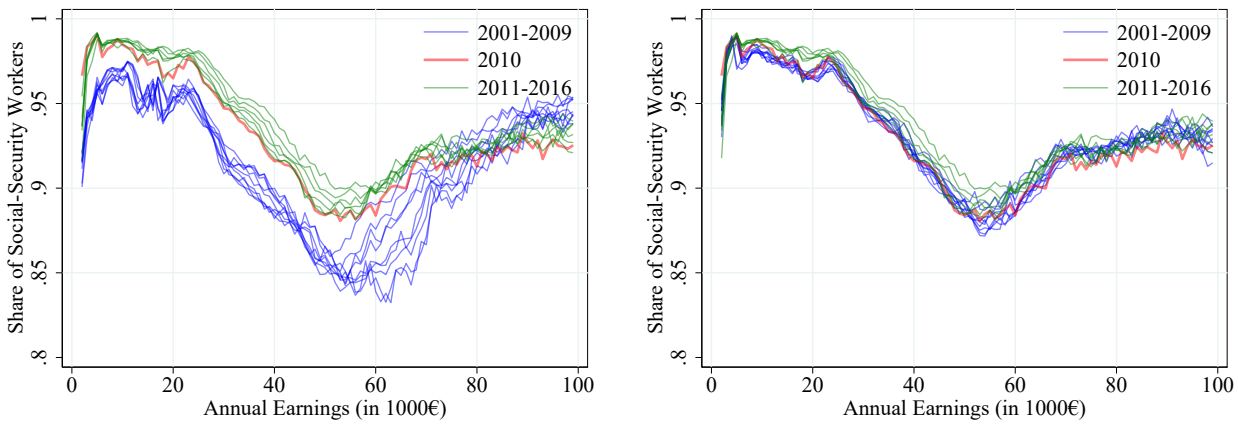
- Let G capture all combinations of the observables gender, age group and binned earnings.
- Let c_{gt} be the share of social-security workers observed in the raw TPP data in group $g \in G$ and year t and c^*_{gt} the corresponding true share.
- The observed share is correct for $t \geq 2010$ ($c = c^*$) 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^m_{gt} constructed weight of mandatory filers in group g

w^v_{gt} 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:

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

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

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

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

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

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

$$\eta_h = E_t\big[N^n_{ht}/N_{ht}|h, t \geq 2012\big] \tag{D.4}$$

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

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

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

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

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

(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:

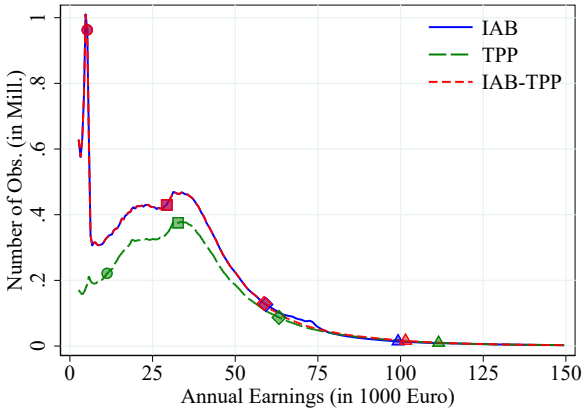
$$F(y) = \begin{cases} F^{IAB}(y) & \text{if } y \leq \bar{y} \\ F^{IAB}(\bar{y}) + F^{TPP}(y|y > \bar{y})\Big(1 - F^{IAB}(\bar{y})\Big) & \text{if } y > \bar{y} \end{cases} \tag{D.14}$$

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]

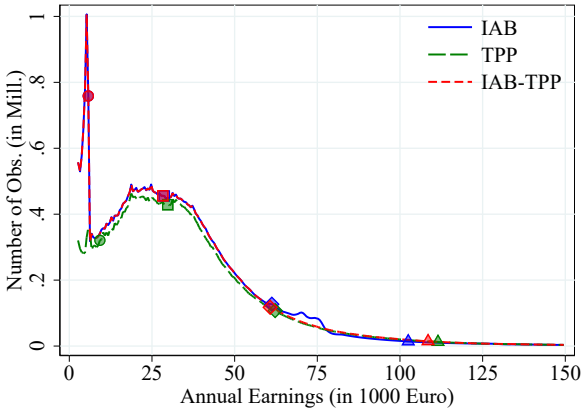
[55] The small deviations below 60,000 Euro are the result of how we combine the IAB and TPP data. After interpolating the quantiles, we discretize the respective distributions on a very fine grid and then combine the discrete distributions. The deviations for higher percentiles are mostly driven by 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

(a) 2001–2011



(b) 2012–2016



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 10th, 50th, 90th and 99th 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

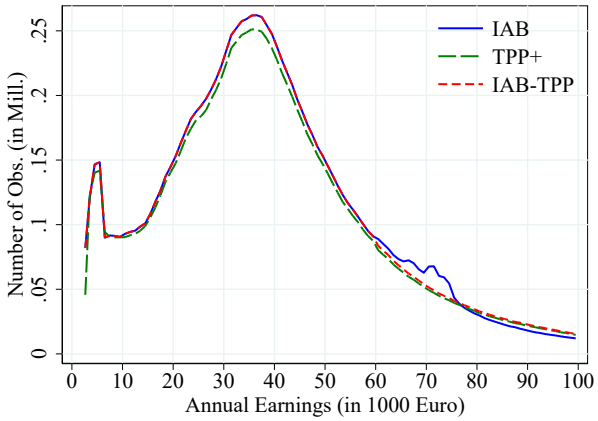
Year	N	Mean	P5	P10	P25	P50	P75	P90	P95	P99	P99.9	P99.99
IAB-TPP Data												
2001	24.588	33,650	4,291	5,698	16,757	31,007	43,605	59,612	74,334	119,461	265,812	701,658
2002	24.185	34,680	4,253	5,613	16,587	30,994	43,835	60,288	75,170	119,931	263,841	683,320
2003	23.759	34,733	4,208	5,668	16,439	31,012	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.

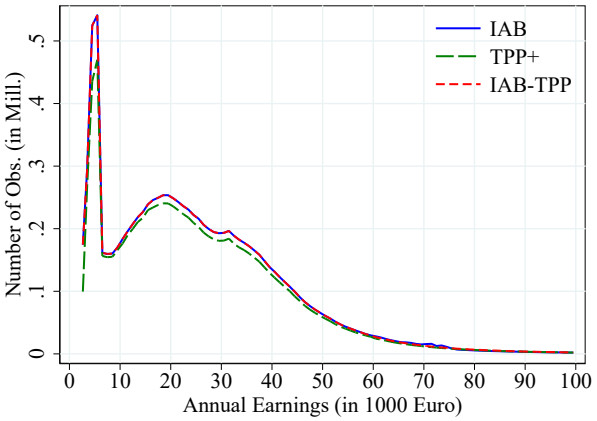
For the LS samples, we follow the same procedure. While the cross-sectional earnings distribution in the reweighted TPP data matches the IAB data (by construction of the weights), this is no longer the case for the LS and H samples due to attrition in the TPP.[56] The LS samples differ from the CS sample 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

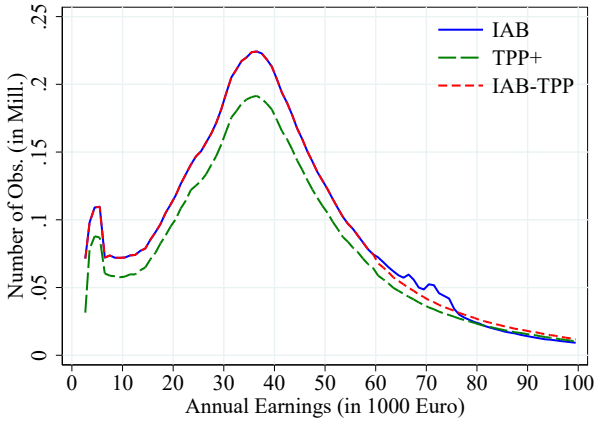
(a) One-Year Changes: Men



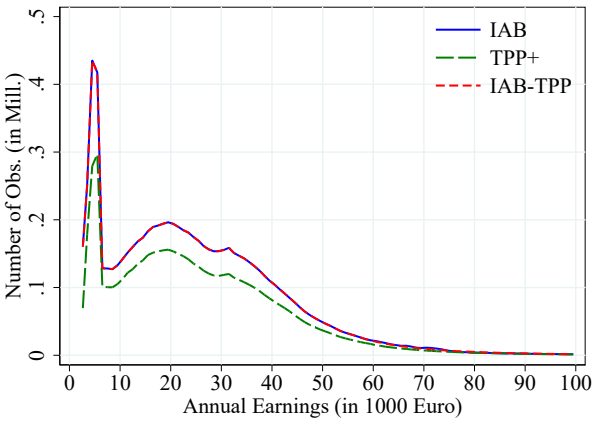
(b) One-Year Changes: Women



(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.

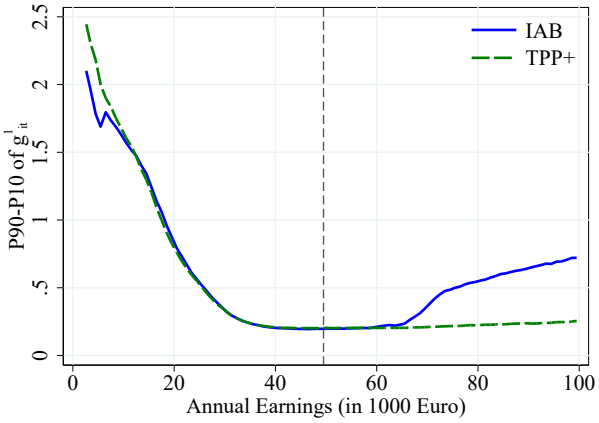
In order to construct the combined IAB-TPP data set for earnings changes, we proceed as follows. First, for low initial earnings bins, we use the conditional earnings growth distribution from the IAB as the probability of having top-coded earnings in $t + k$ is very low. Second, as soon as more than 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.

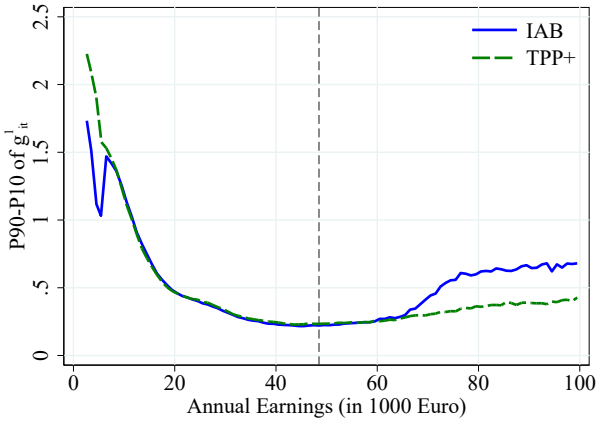
[58] In order to approximate the CDF using monotonic spline interpolation of quantiles in a given dataset, we have to impose a minimum and maximum for g, i.e. q^0 and q^{100}. Let $\hat{F}_{g|Y}$ denote the resulting approximation of the CDF of earnings growth. We set the minimum and the maximum such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

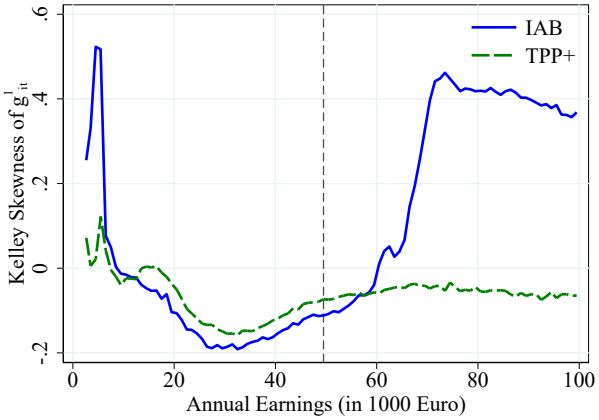
(a) P90-P10: Men



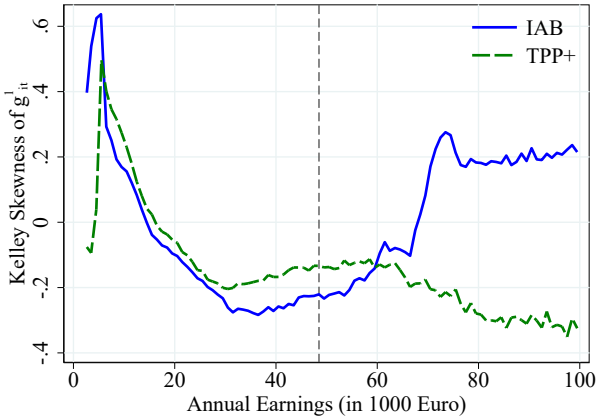
(b) P90-P10: Women



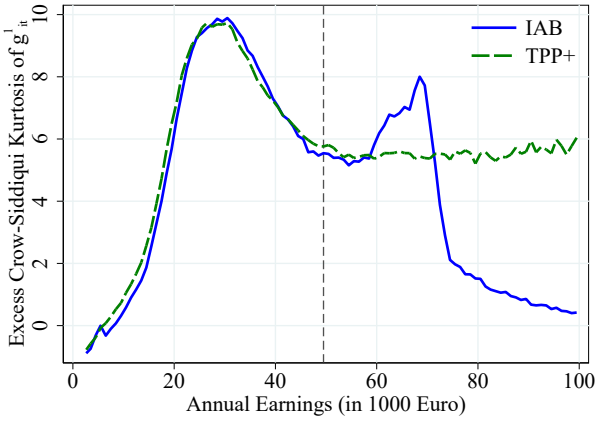
(c) Kelley Skewness: Men



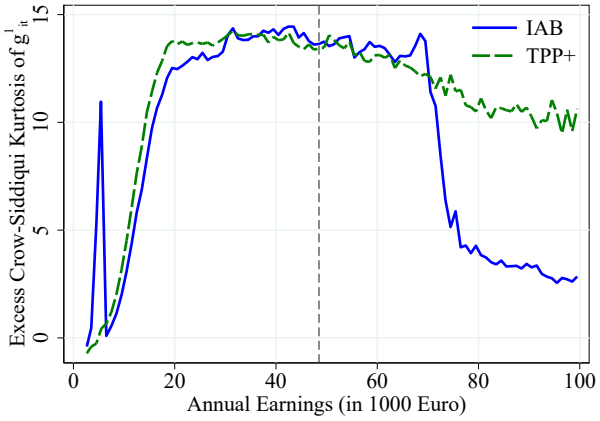
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



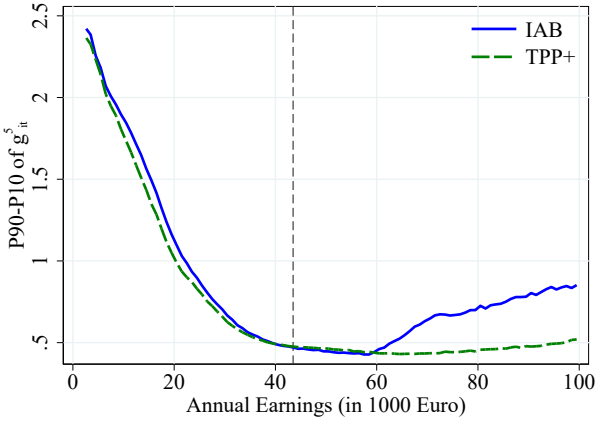
(f) Excess Crow-Siddiqui Kurtosis: Women



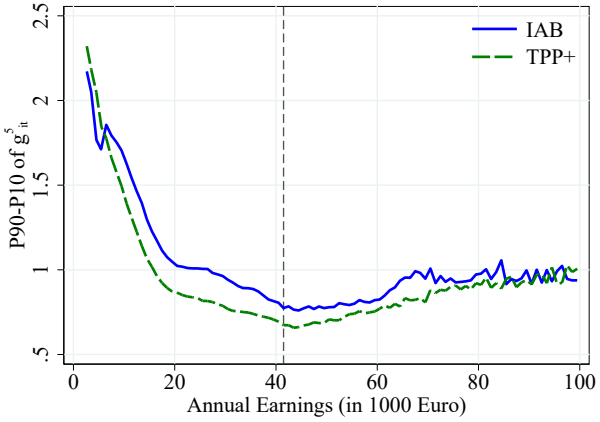
Notes: LS sample. One-year residualized log earnings growth. Averaged over years 2001-2015. Source: IAB and TPP.

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

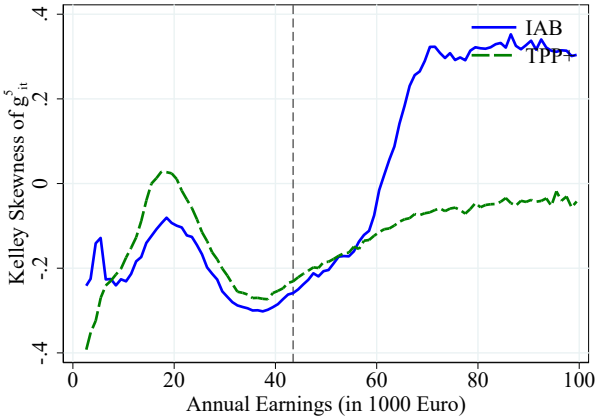
(a) P90-P10: Men



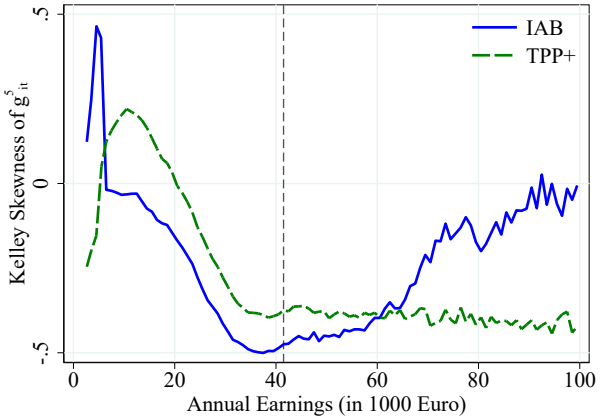
(b) P90-P10: Women



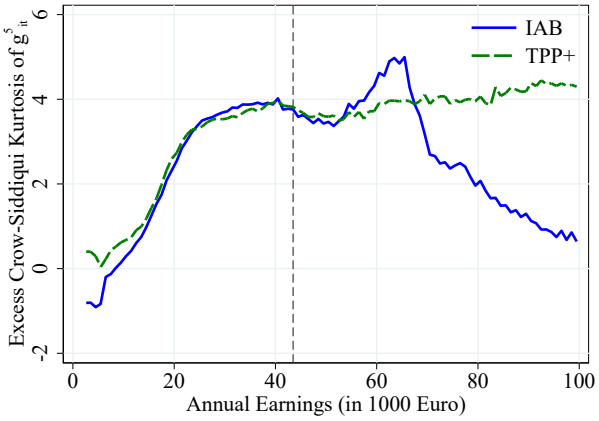
(c) Kelley Skewness: Men



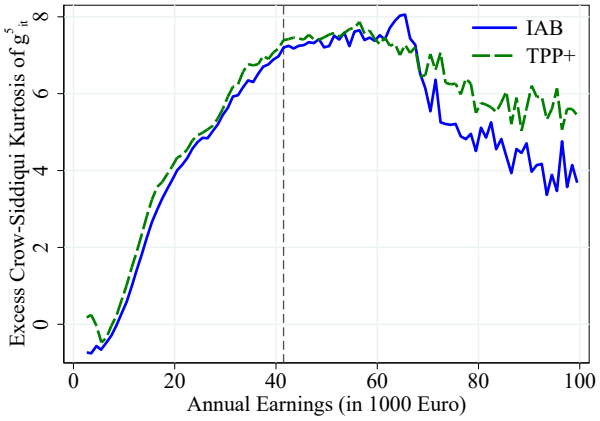
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



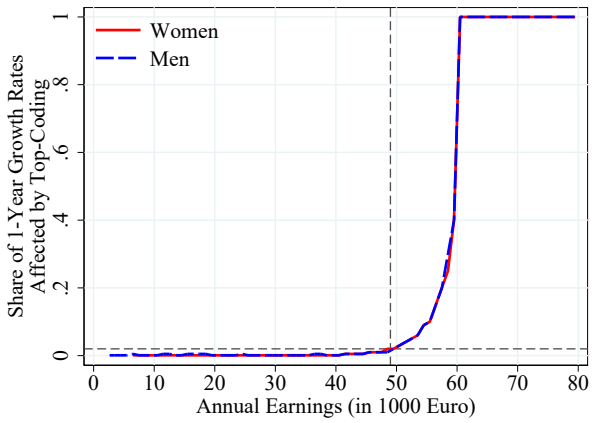
(f) Excess Crow-Siddiqui Kurtosis: Women



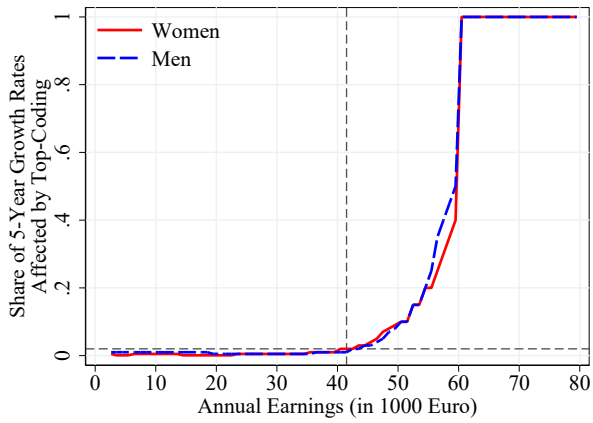
Notes: LS sample. Five-year residualized log earnings growth. Averaged over years 2001-2011. Source: IAB and TPP.

Figure D.5: Share of 1-Year Log Earnings Changes Affected by Top-Coding in the IAB Data

(a) One-Year Changes



(b) Five-Year Changes



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

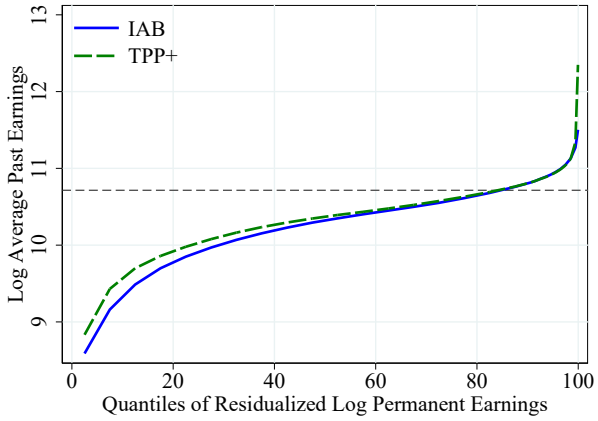
For the heterogeneity analysis by permanent earnings, we use a simple cut-off rule to combine IAB and TPP data. For 1-year growth rates, this cutoff is equal to 45,000 Euro. Hence, for all quantiles of the residualized permanent earnings distribution above this cutoff, we use the conditional growth rate 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.

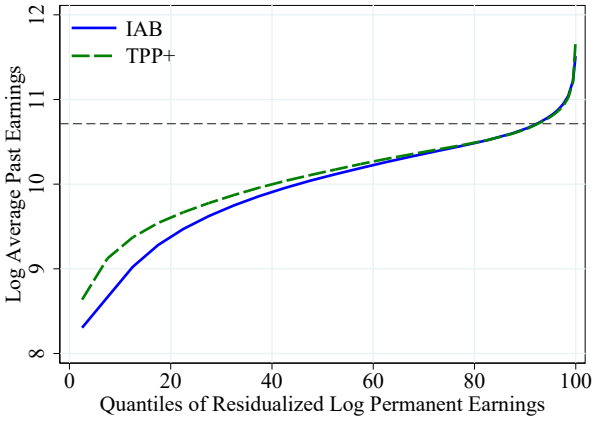
For 5-year earnings changes we proceed analogously, but use a cutoff of 40,000 Euro as jumping into the top-coded range is more likely over a period of five years. Figures D.10, D.11 and D.12 show the P90-P10 differential, Kelley Skewness and Excess Crow-Siddiqui kurtosis by permanent earnings quantiles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

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

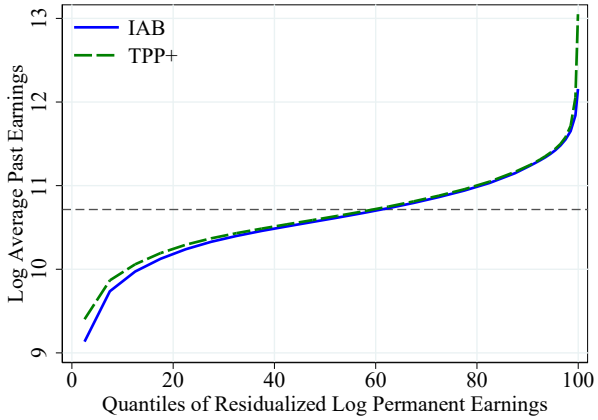
(a) Age Group 25–34: Men



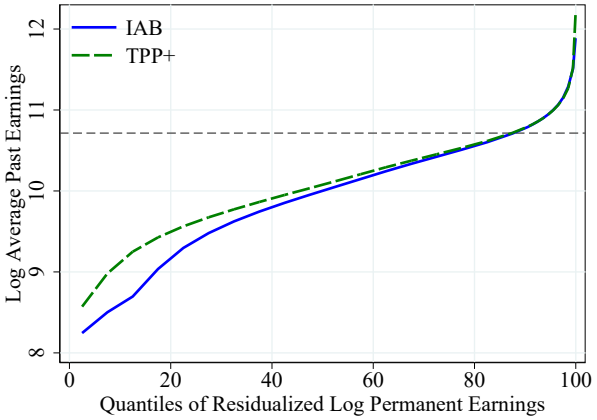
(b) Age Group 25–34: Women



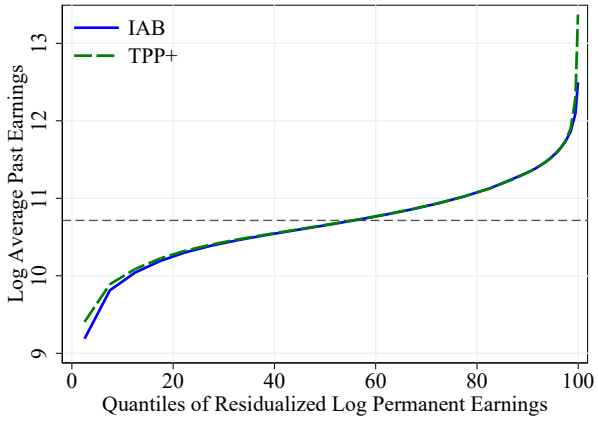
(c) Age Group 35–44: Men



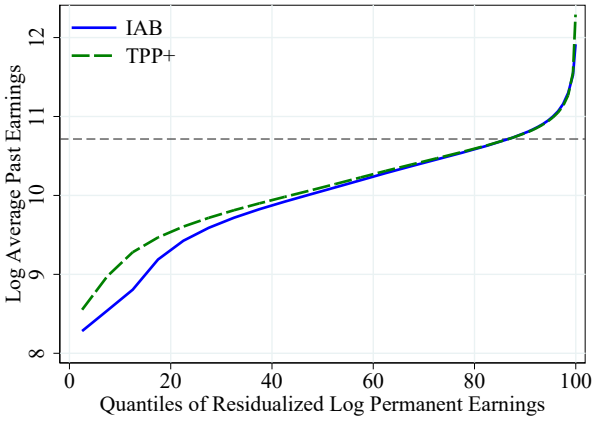
(d) Age Group 35–44: Women



(e) Age Group 45–55: Men



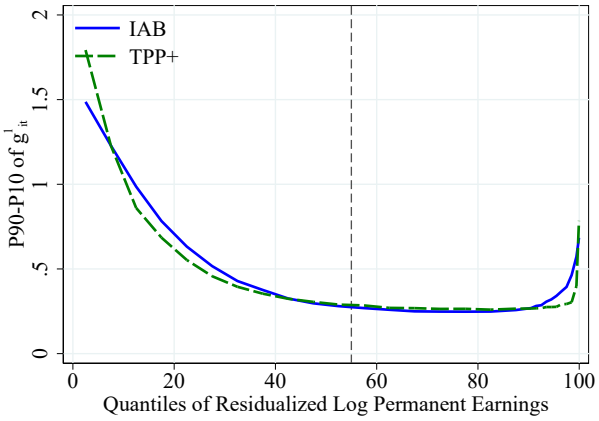
(f) Age Group 45–55: Women



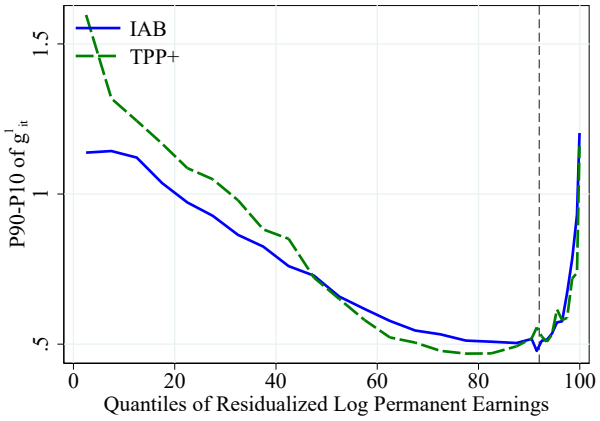
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 45,000 Euro, i.e. the point where the lines in Figure 9 in the main text switch from IAB to TPP data. Source: IAB and TPP.

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

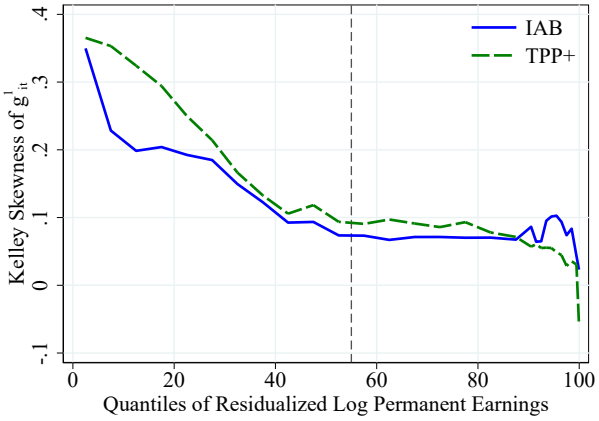
(a) P90-P10: Men



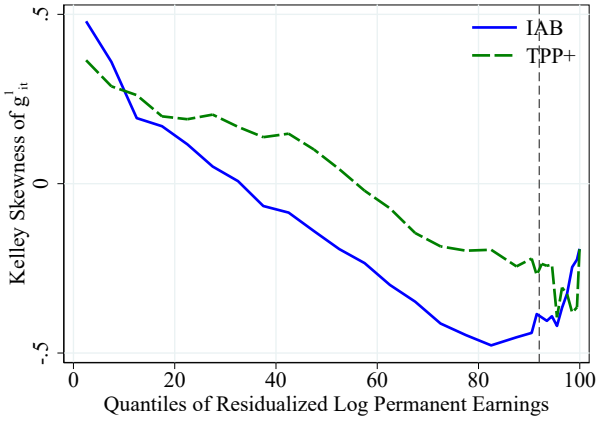
(b) P90-P10: Women



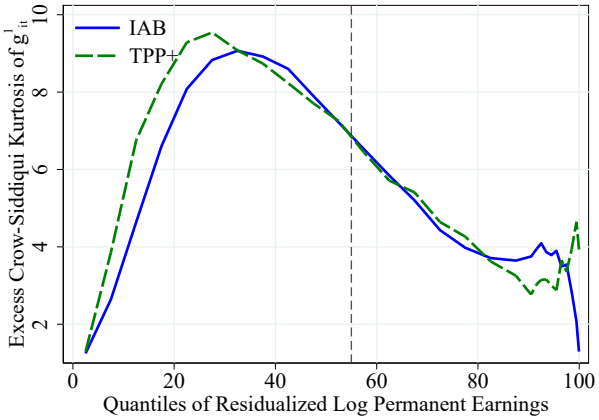
(c) Kelley Skewness: Men



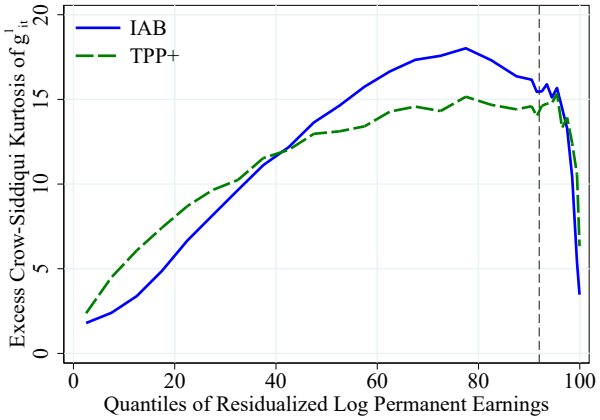
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



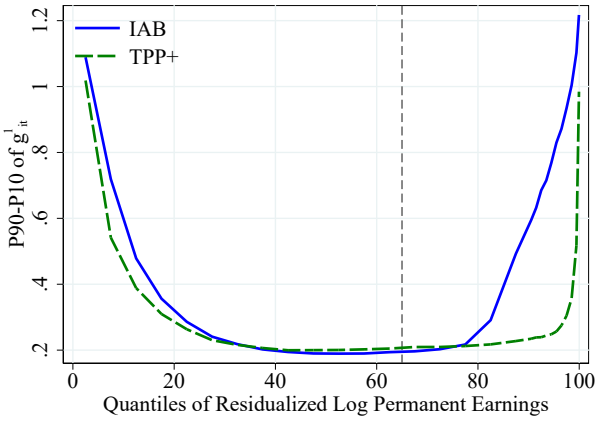
(f) Excess Crow-Siddiqui Kurtosis: Women



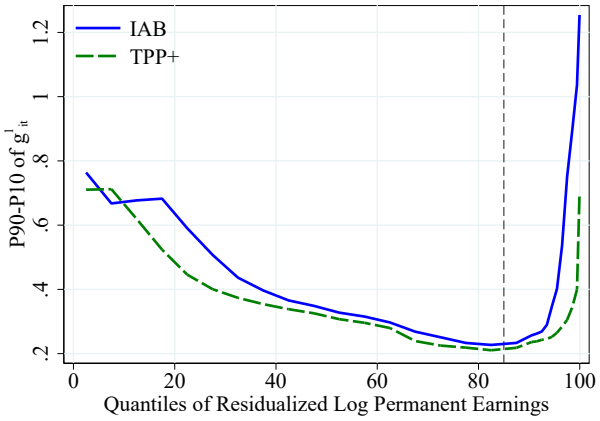
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 45,000 Euro, i.e. the point where the lines in Figure 9 in the main text switch from IAB to TPP data. Source: IAB and TPP.

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

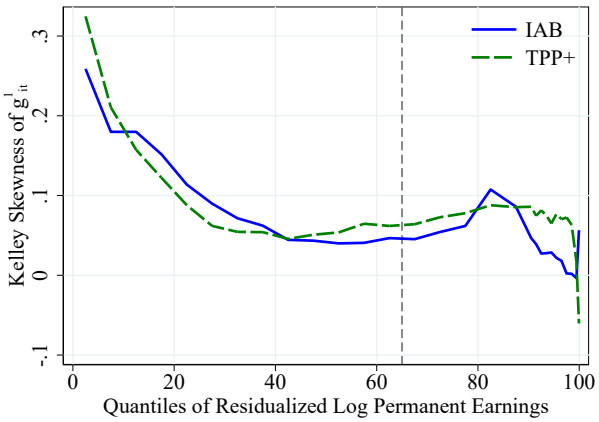
(a) P90-P10: Men



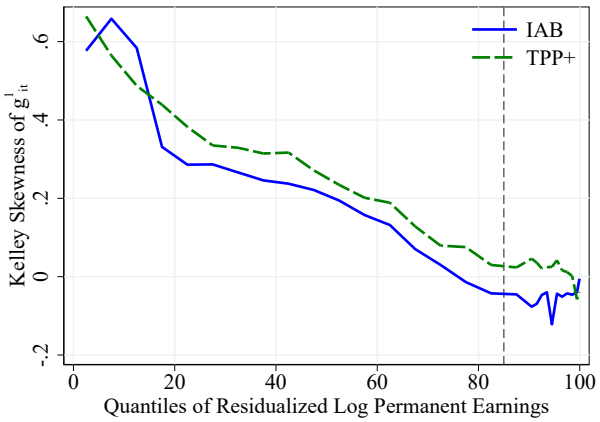
(b) P90-P10: Women



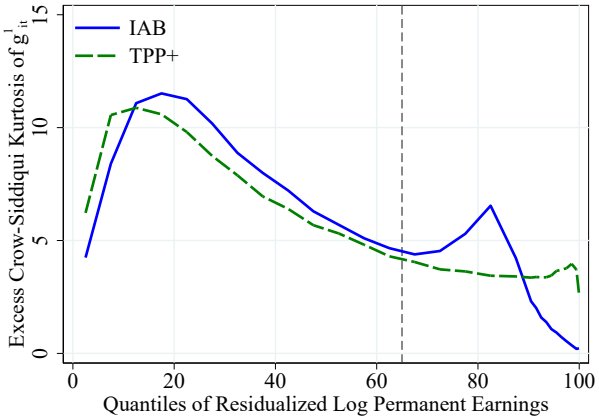
(c) Kelley Skewness: Men



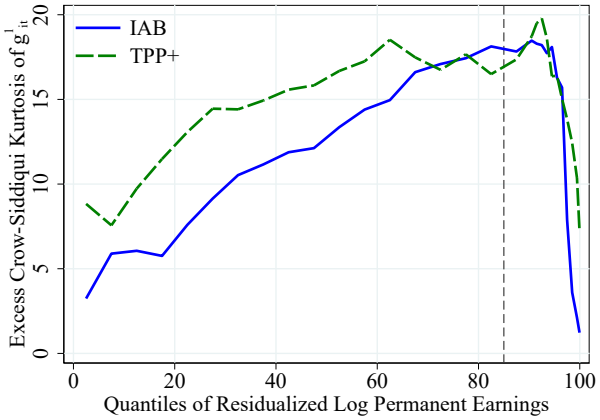
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



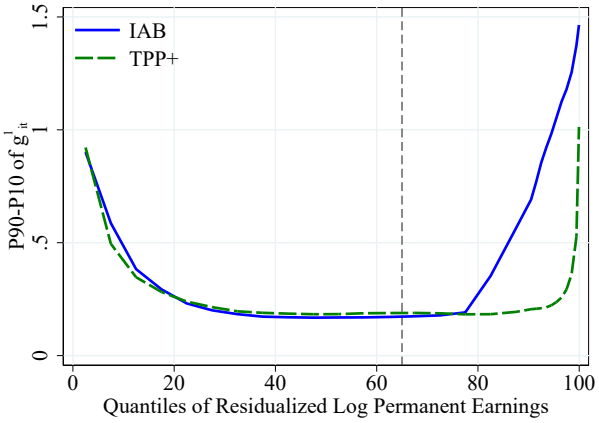
(f) Excess Crow-Siddiqui Kurtosis: Women



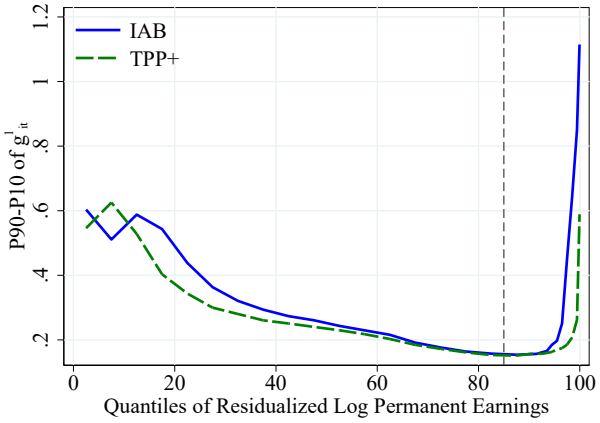
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 45,000 Euro, i.e. the point where the lines in Figure 9 in the main text switch from IAB to TPP data. Source: IAB and TPP.

Figure D.9: IAB vs. TPP: 1-Year Log Earnings Changes by Permanent Earnings, Age Group 45–55

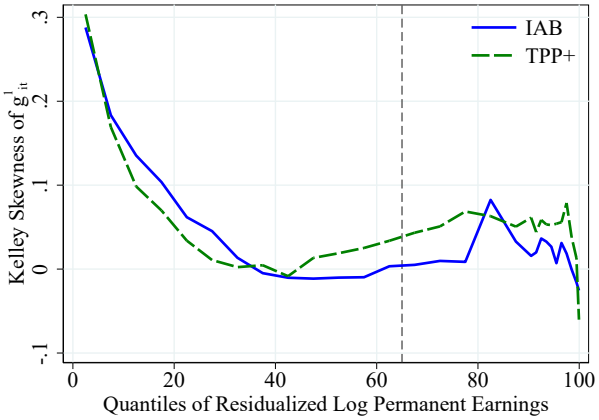
(a) P90-P10: Men



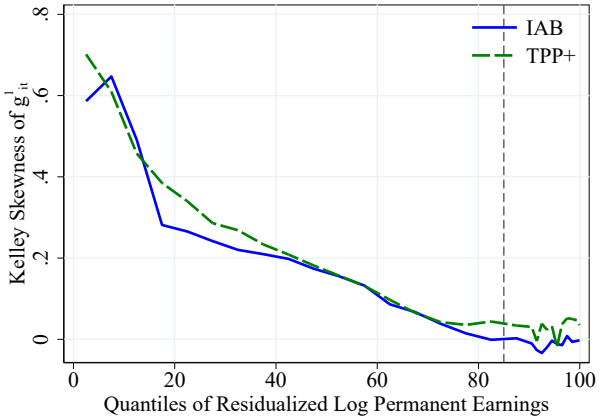
(b) P90-P10: Women



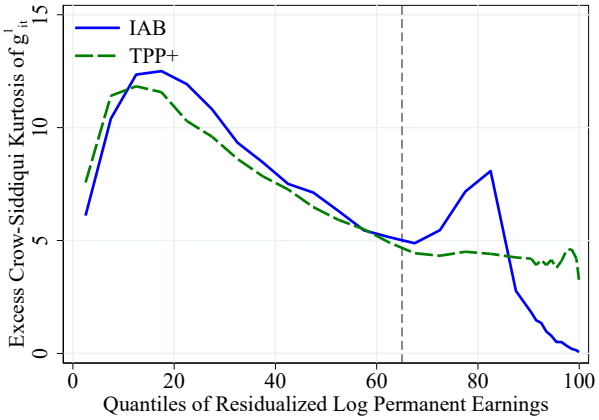
(c) Kelley Skewness: Men



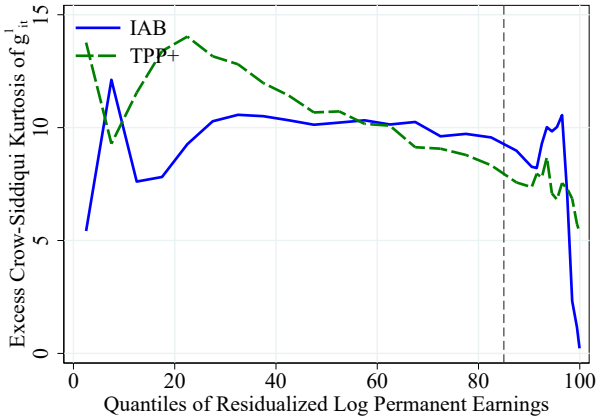
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



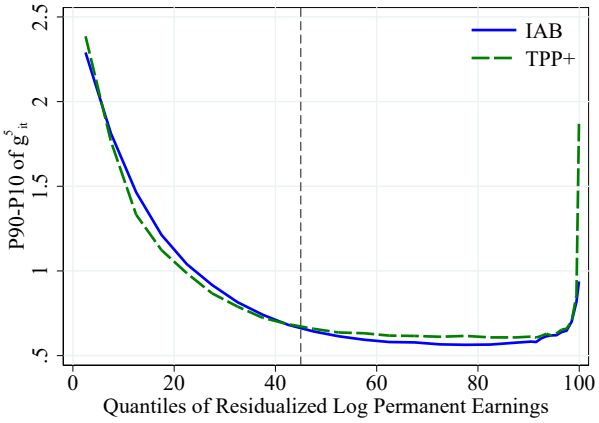
(f) Excess Crow-Siddiqui Kurtosis: Women



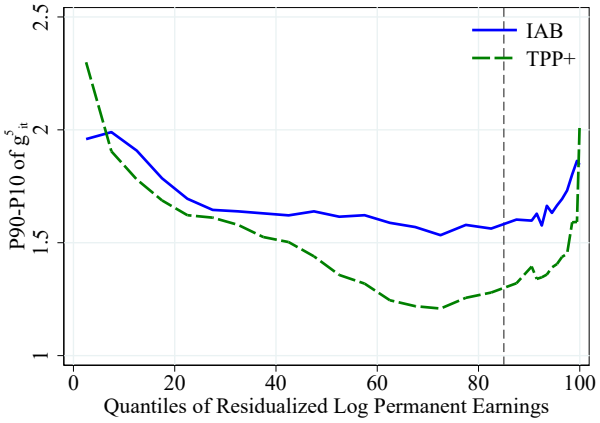
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 45,000 Euro, i.e. the point where the lines in Figure 9 in the main text switch from IAB to TPP data. Source: IAB and TPP.

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

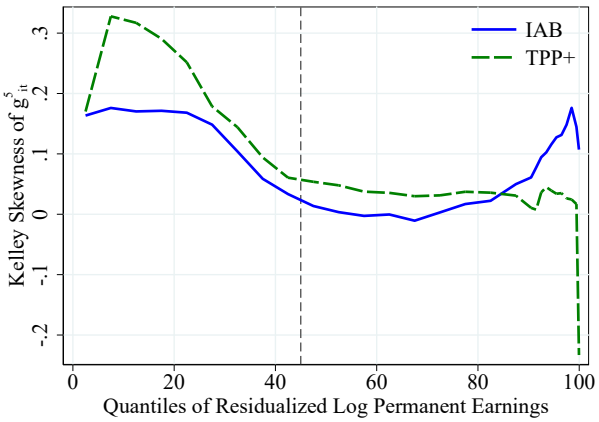
(a) P90-P10: Men



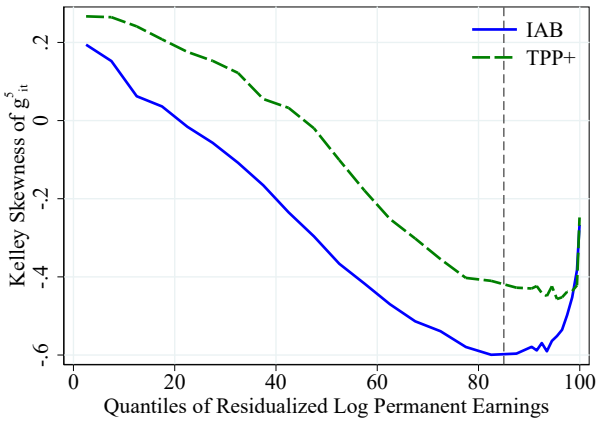
(b) P90-P10: Women



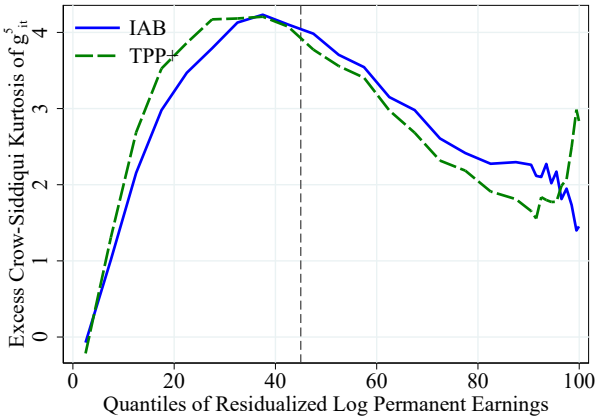
(c) Kelley Skewness: Men



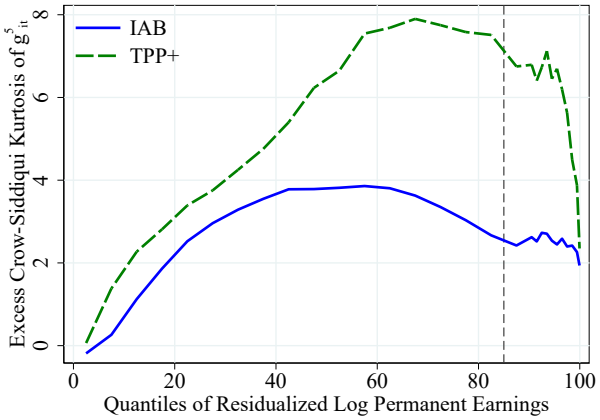
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



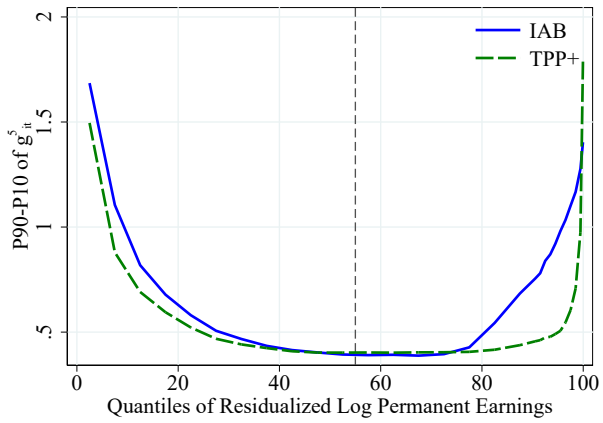
(f) Excess Crow-Siddiqui Kurtosis: Women



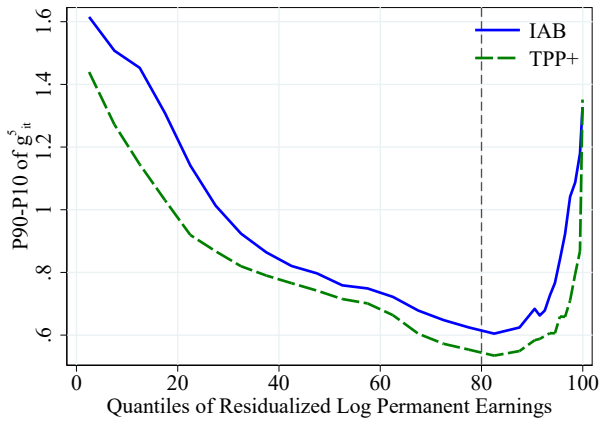
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 40,000 Euro, i.e. the point where the lines in Figure E.28 in the main text switch from IAB to TPP data. Source: IAB and TPP.

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

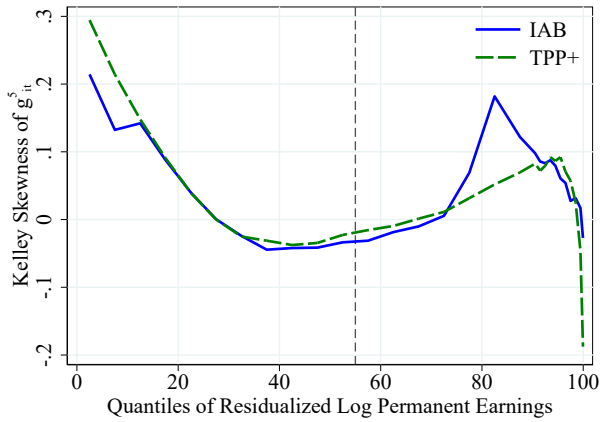
(a) P90-P10: Men



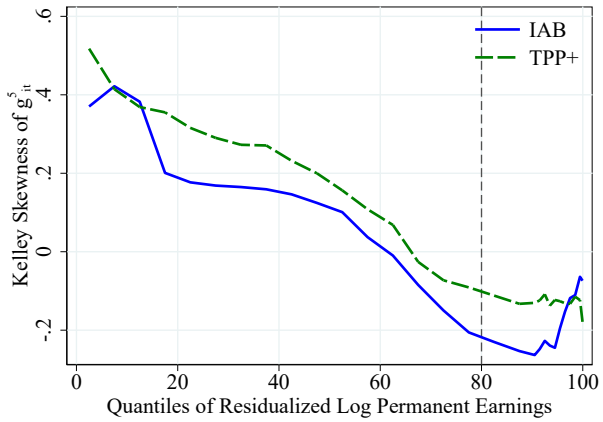
(b) P90-P10: Women



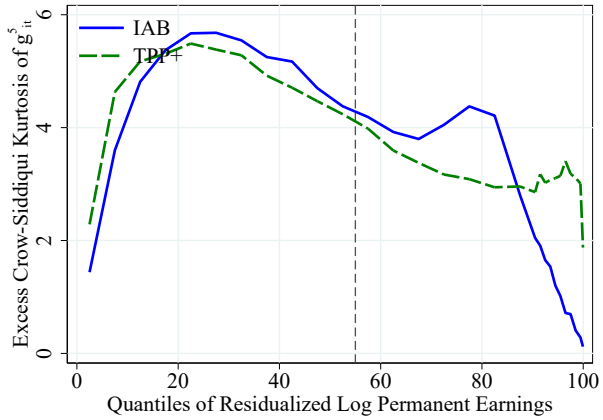
(c) Kelley Skewness: Men



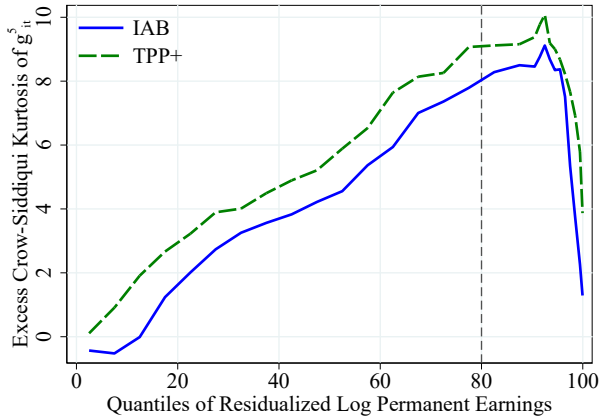
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



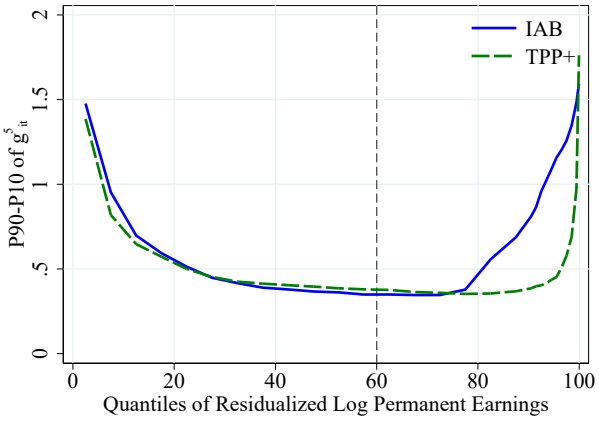
(f) Excess Crow-Siddiqui Kurtosis: Women



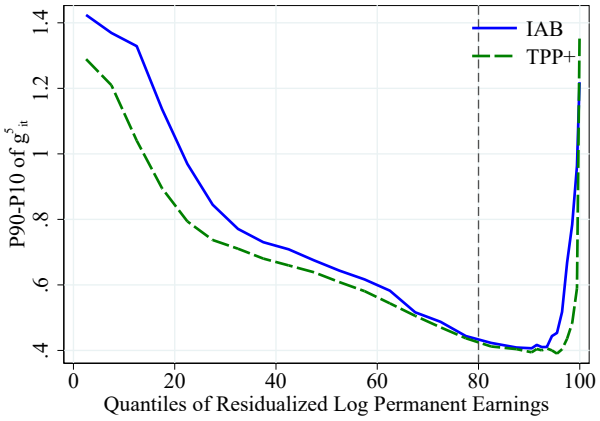
Notes: H sample, averages from 2004 to 2011. The dashed vertical depicts the point where permanent earnings is equal to 40,000 Euro, i.e. the point where the lines in Figure E.28 in the main text switch from IAB to TPP data. Source: IAB and TPP.

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

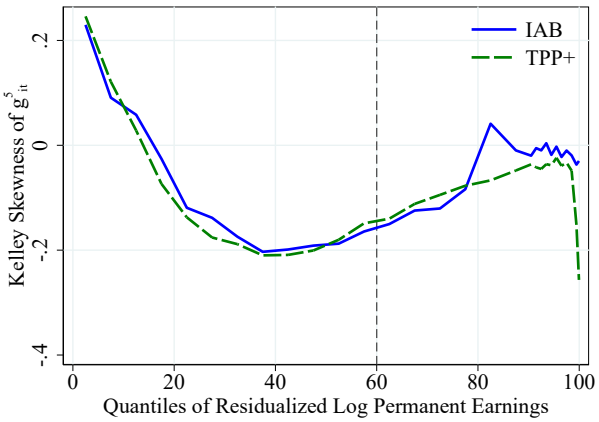
(a) P90-P10: Men



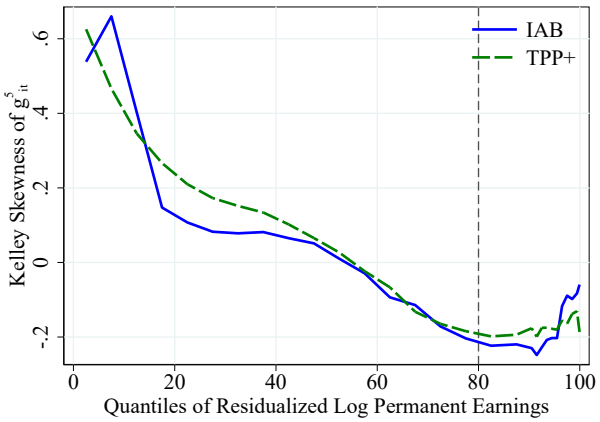
(b) P90-P10: Women



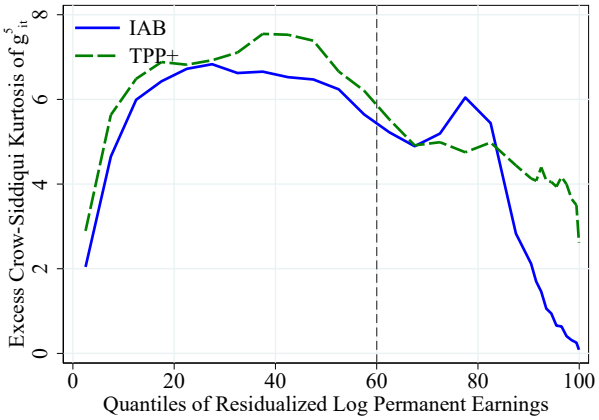
(c) Kelley Skewness: Men



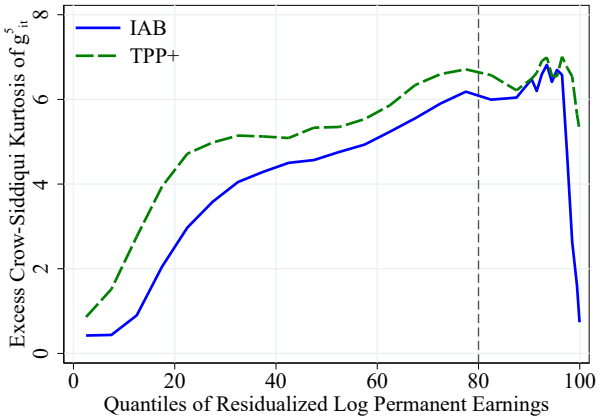
(d) Kelley Skewness: Women



(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.

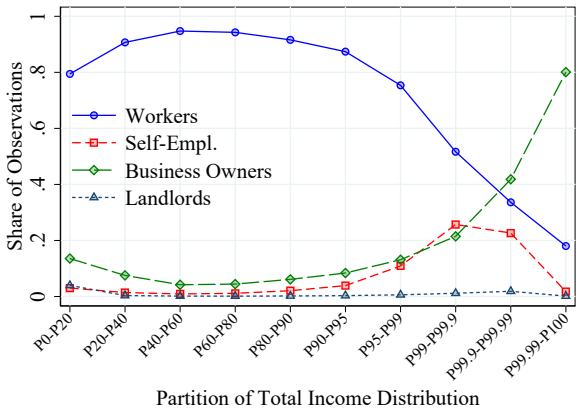
Table D.11: Summary Statistics for Total Income Data

	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

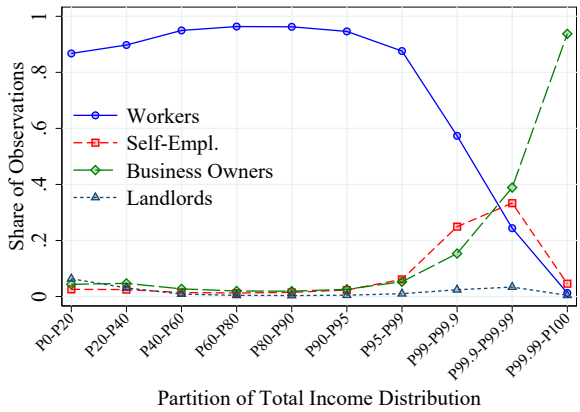
Notes: This table shows descriptive statistics for the full TPP and IAB-TPP data by gender for the year 2008. The data includes all workers independent of their social-security status and individuals with non-labor income. Columns 1 and 4 refer to the raw TPP data (earnings not reweighted using IAB data). 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.

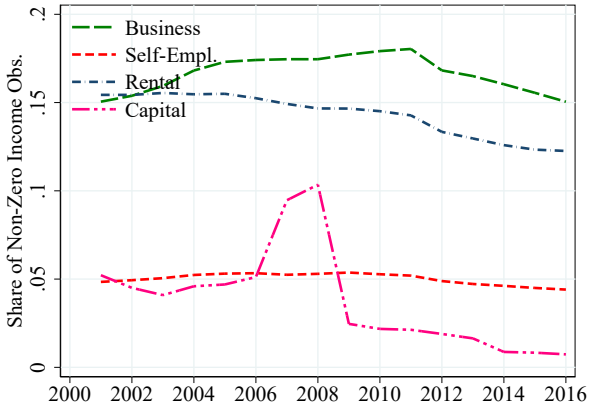
Table D.14: Correlations Between Income Components – Women

	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

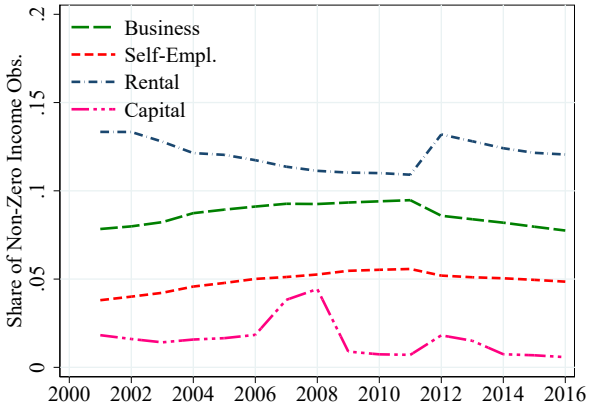
Notes: This table shows correlations between different income components in the combined IAB-TPP analysis sample for women in 2008. CS sample. Source: IAB and TPP.

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

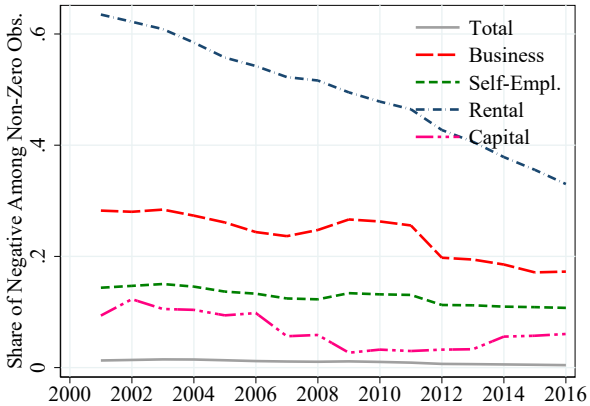
(a) Non-Zero Values: Men



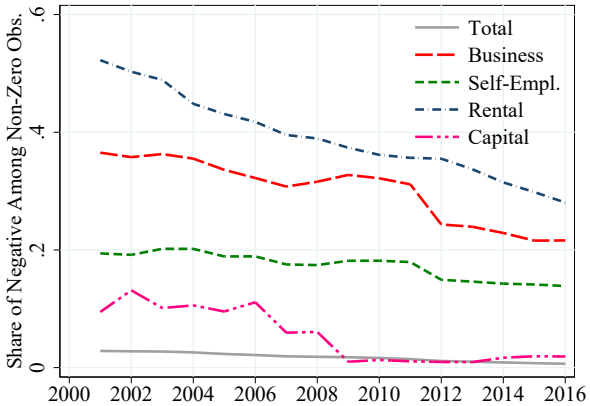
(b) Non-Zero Values: Women



(c) Negative Values: Men



(d) Negative Values: Women



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.

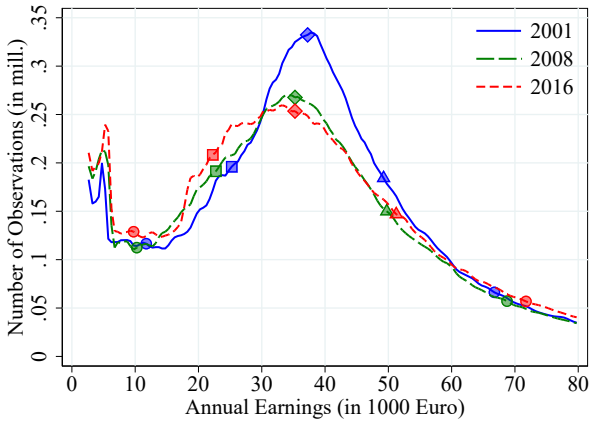
E Core Analysis: Additional Results for Combined IAB-TPP Data 2001–2016

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

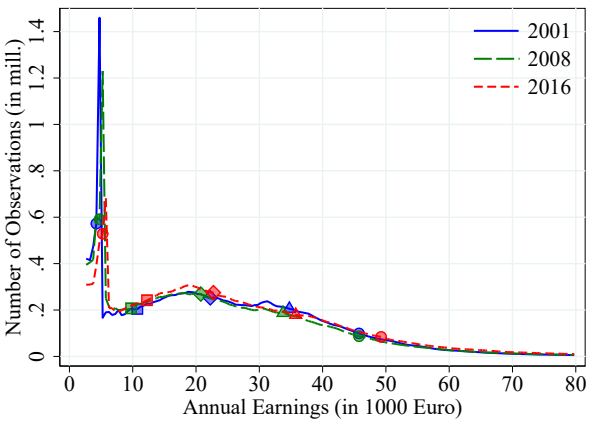
E.1 Additional Results for Earnings Inequality (Section 3.1)

Figure E.1: Earnings Distribution

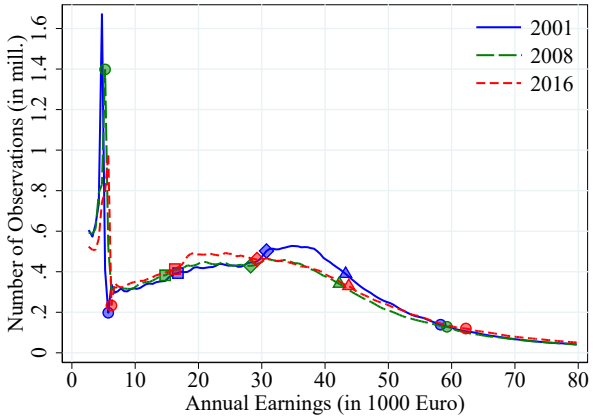
(a) Number of Observations: Men



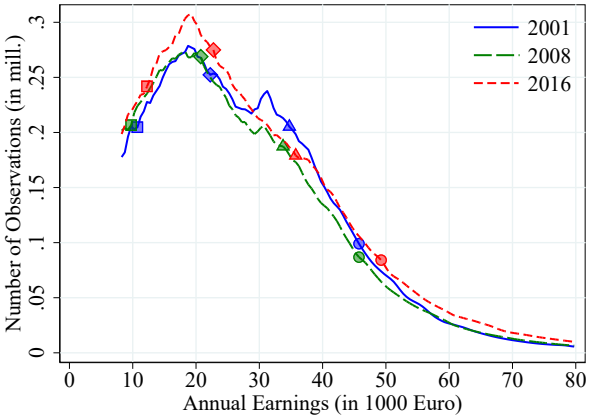
(b) Number of Observations: Women



(c) Number of Observations: Population



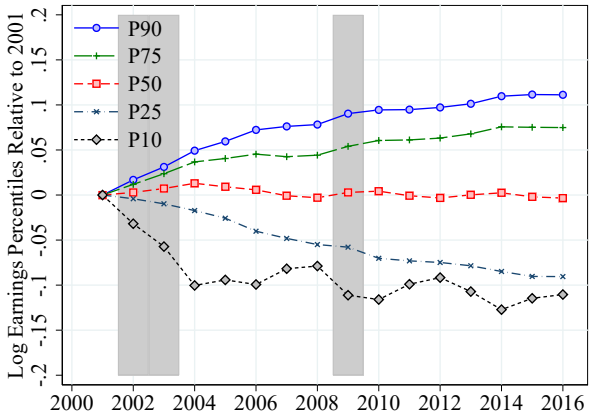
(d) Number of Observations: Women (Zoom)



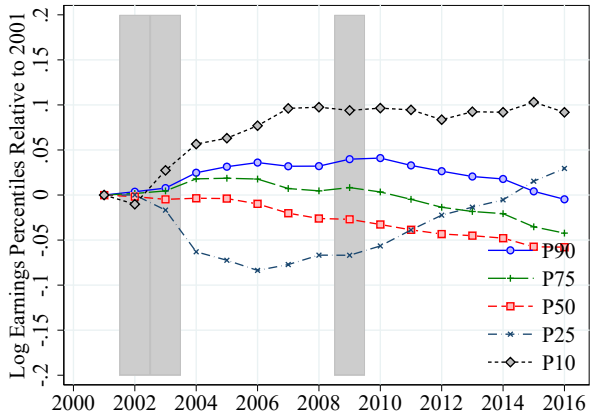
Notes: This figure shows the number of observations per 1,000 Euro earnings bins of real annual earnings for selected years in the combined IAB-TPP data (CS sample) separately for men and women. Panel A and B are depicted as shares in Figure 2. The data is smoothed (by year and gender) using a three-bin 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)

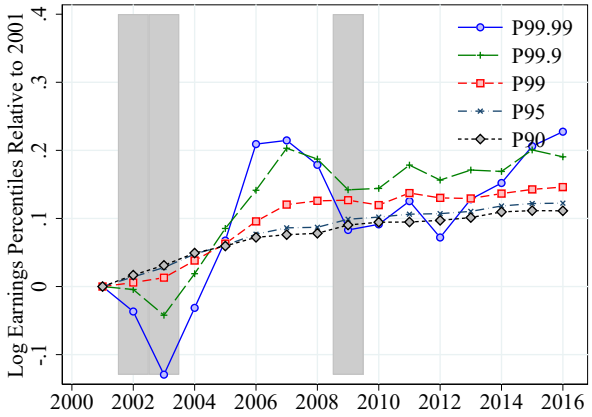
(a) Core Percentiles: Men



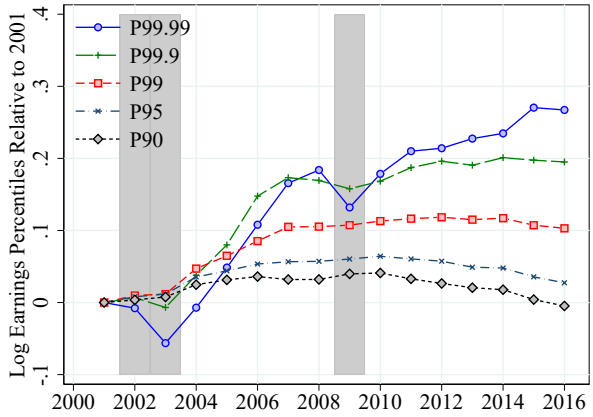
(b) Core Percentiles: Women



(c) Top Percentiles: Men



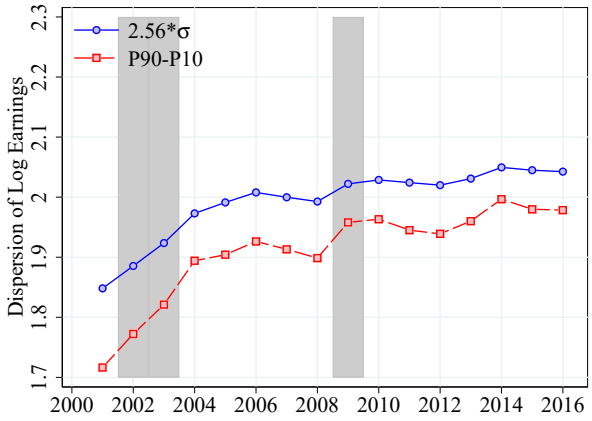
(d) Top Percentiles: Women



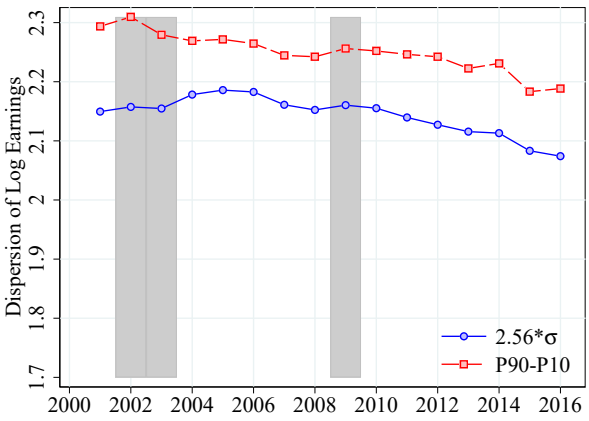
Notes: This figure shows the evolution of residualized log real annual earnings (controlling for age, for unconditioned percentiles, see Figure 3) in the combined IAB-TPP data (CS sample). Shaded areas indicate recessions.

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

(a) Men



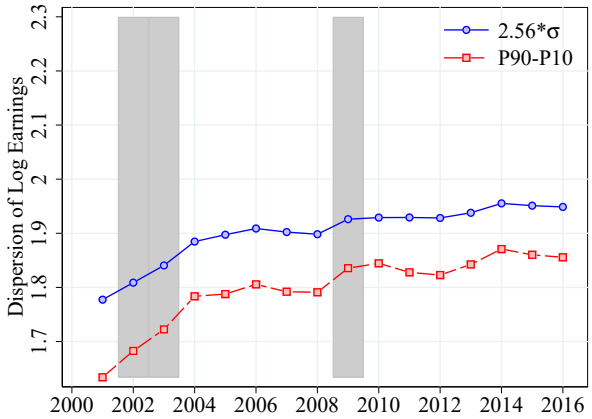
(b) Women



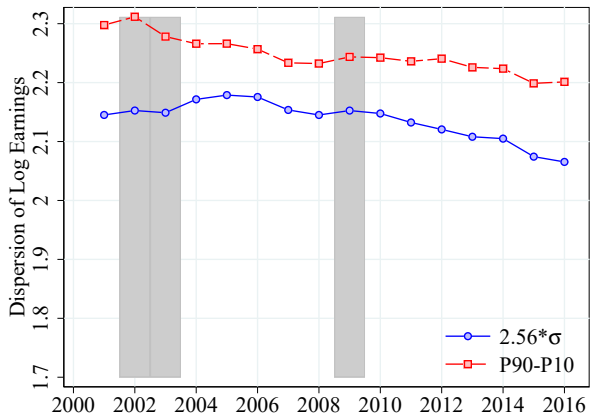
Notes: This figure shows the evolution of different log percentile differentials as well as the (rescaled) standard deviation of the log real annual earnings distribution over time in the combined IAB-TPP data (CS sample) separately for men and women. The standard deviation σ is rescaled as $2.56 * \sigma$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

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

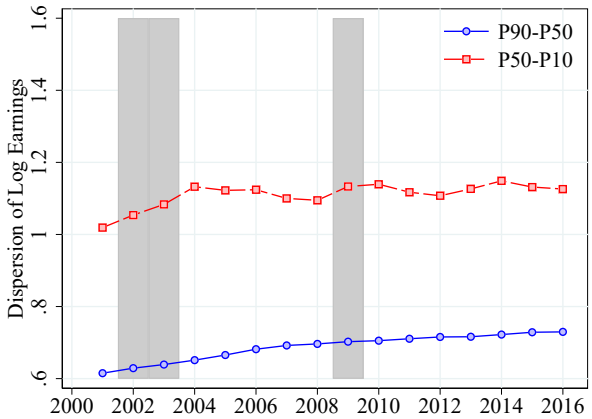
(a) Inequality: Men



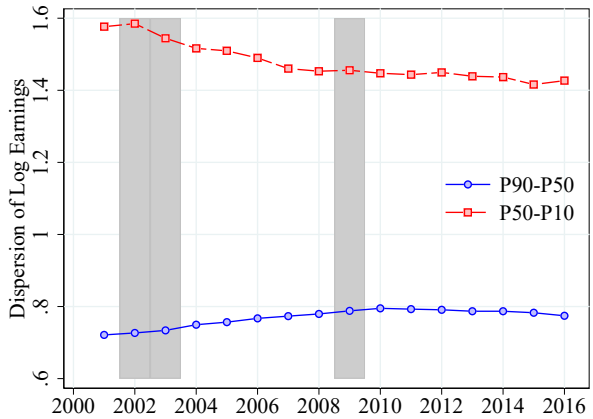
(b) Inequality: Women



(c) Upper and Lower Inequality: Men

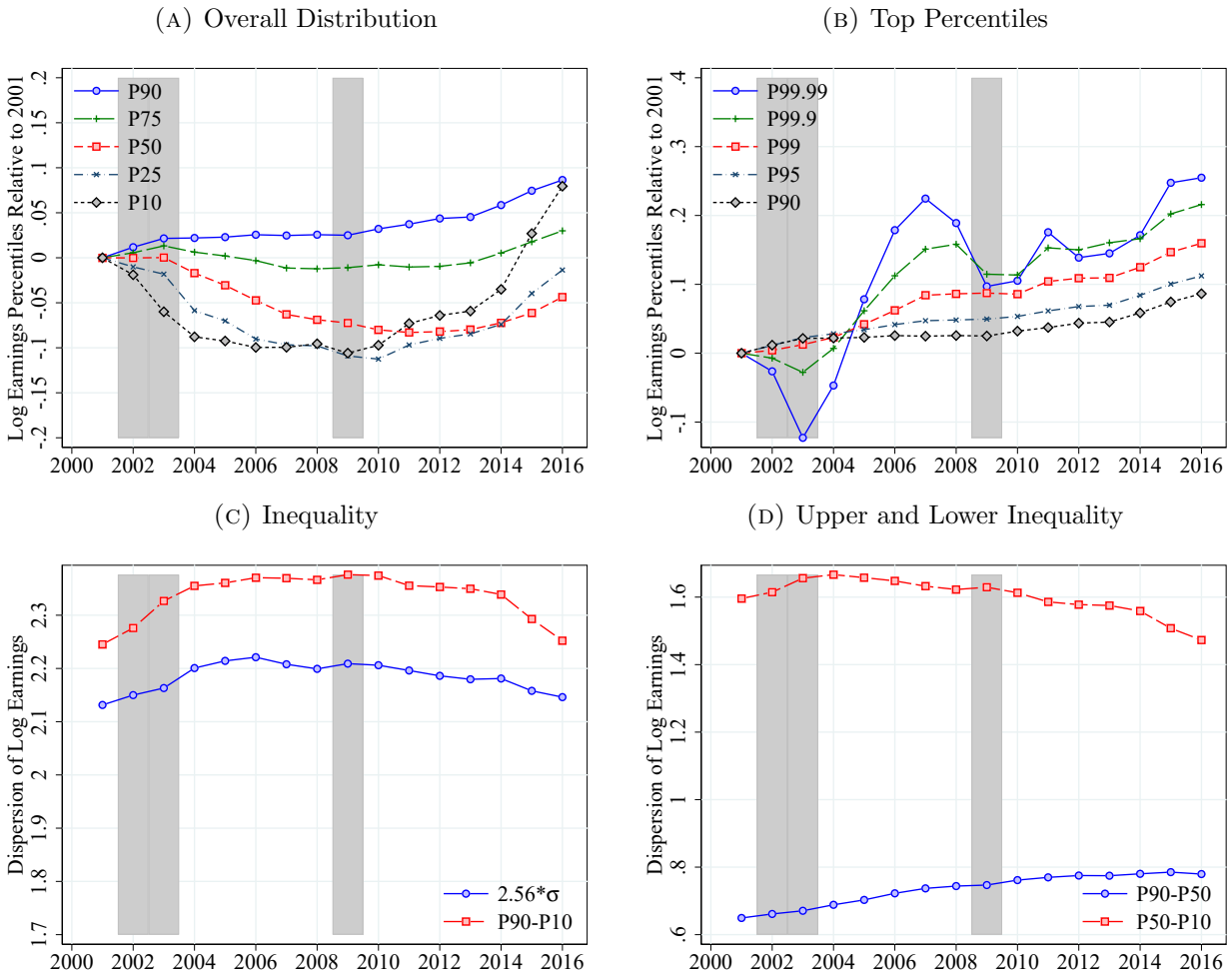


(d) Upper and Lower Inequality: Women



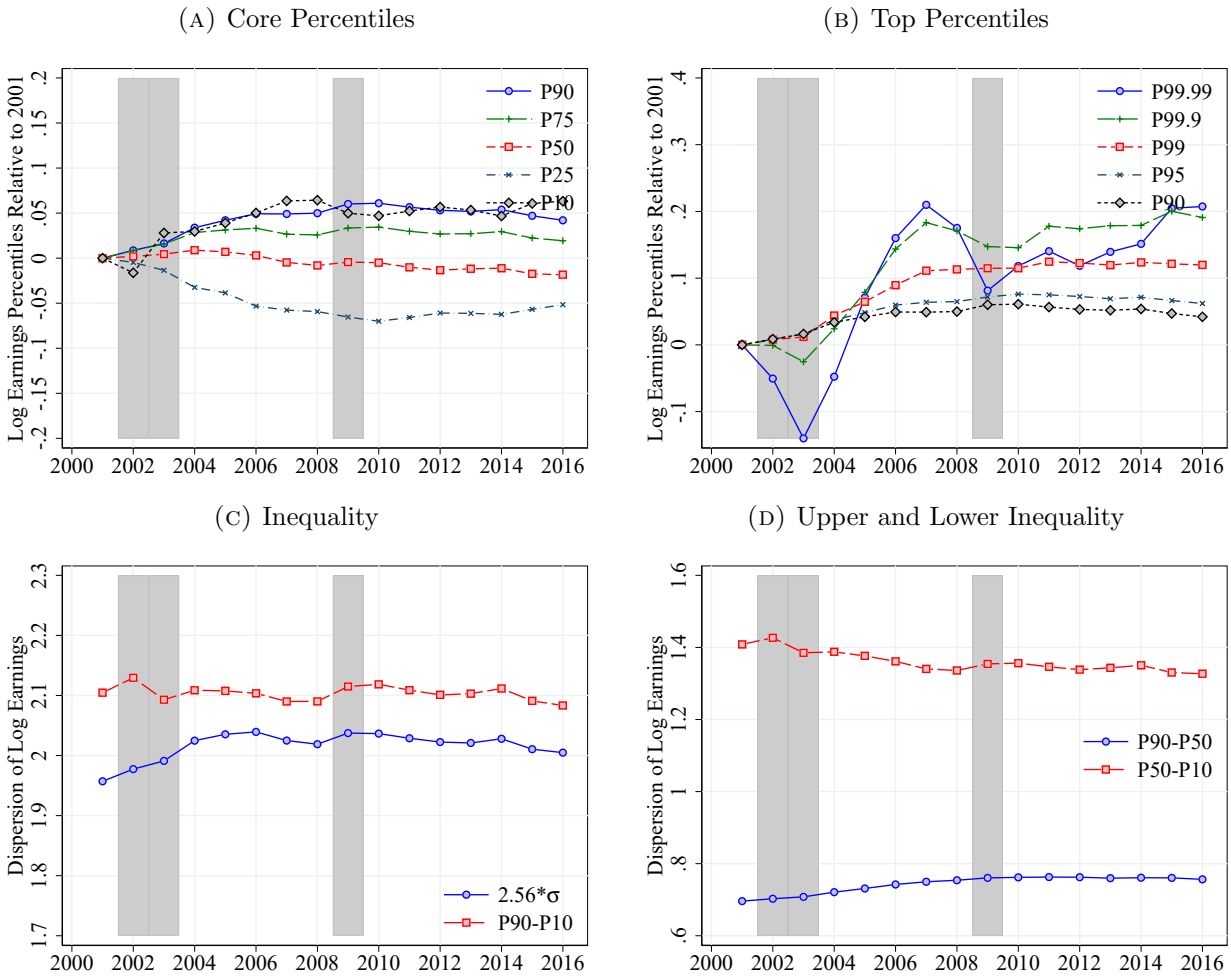
Notes: This figure shows the evolution of residualized log real annual earnings (controlling for age, unconditioned results can be found in Figure 4) in the combined IAB-TPP data (CS sample). Shaded areas indicate recessions.

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.

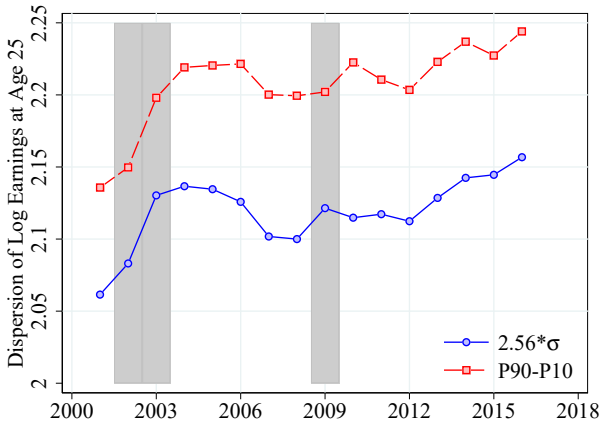
Figure E.6: Residual Log Earnings Inequality in the Population (Controlling for Gender and Age)



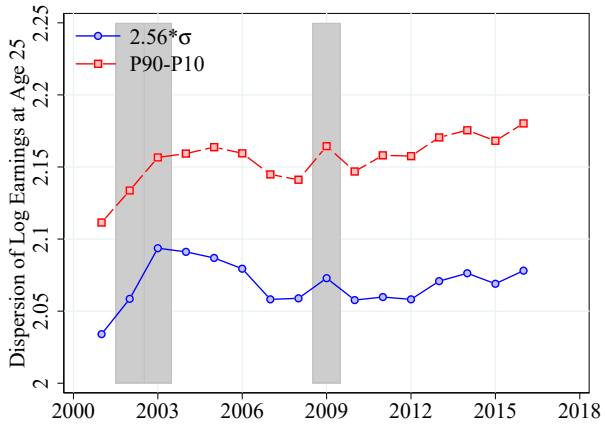
Notes: This figure shows the evolution of residualized log real annual earnings (controlling for gender and age, unconditioned results can be found in Figures 3 and 4.) in the combined IAB-TPP data (CS sample). Shaded areas indicate recessions.

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

(a) Inequality: Men

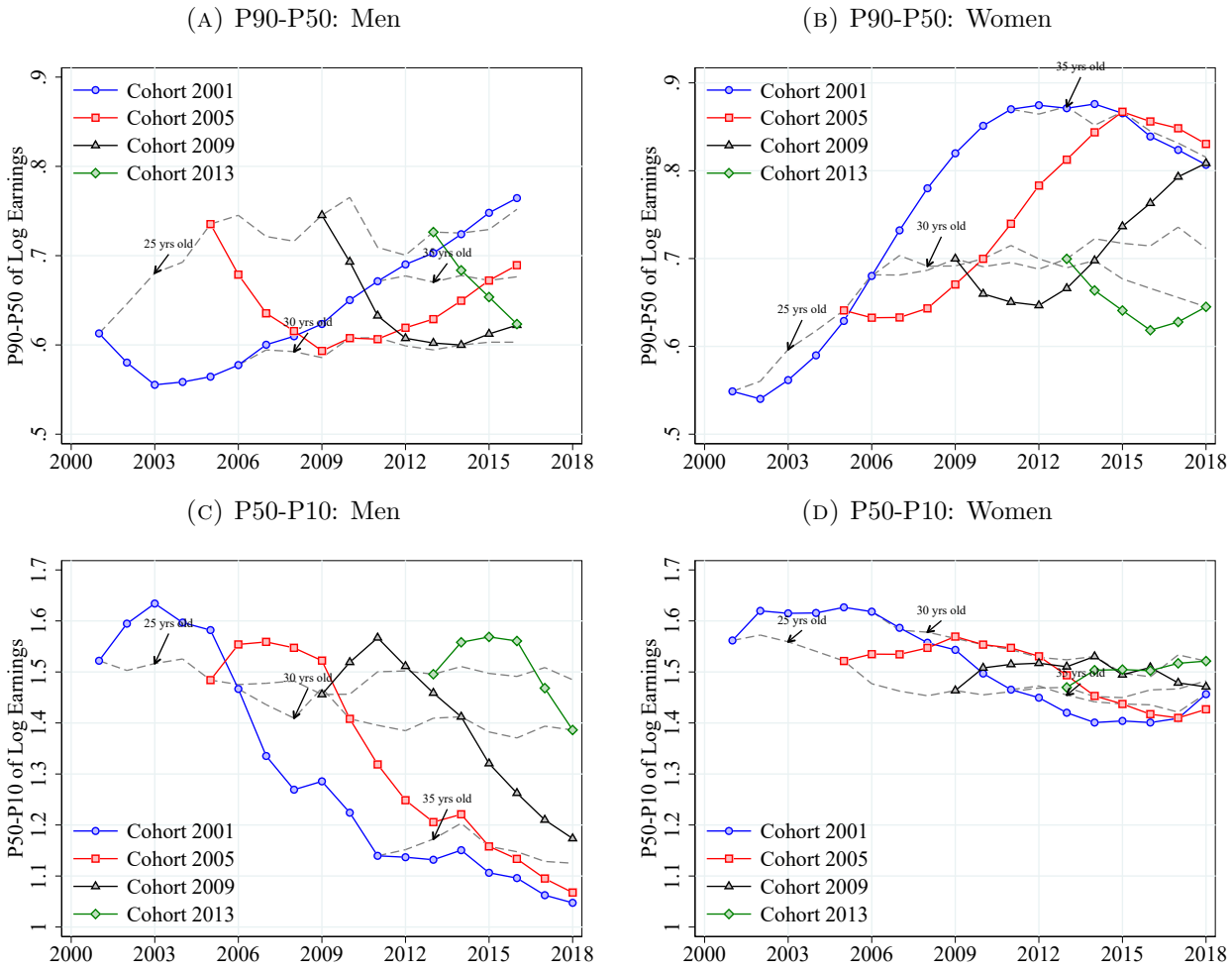


(b) Inequality: Women



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.

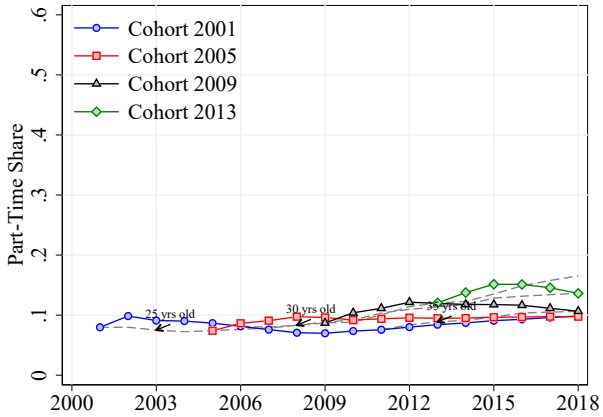
Figure E.8: Upper and Lower Earnings Inequality by Cohort



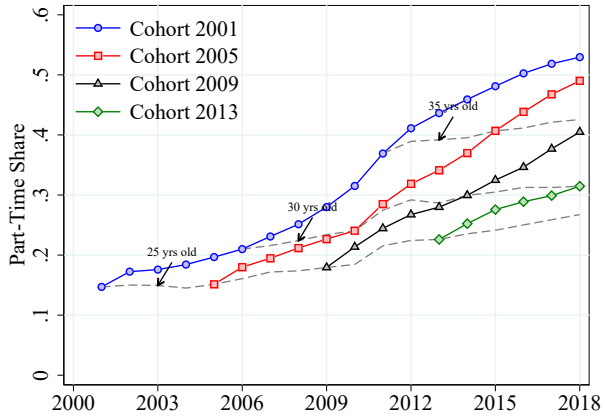
Notes: This figure shows the evolution of the P90-P50 and the P50-P10 differentials of the log real annual earnings distribution over time in the combined IAB-TPP data (CS sample) separately for men and women. As the P90 of men is imputed and the TPP data end in 2016, Panel A also ends in 2016. Grey 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

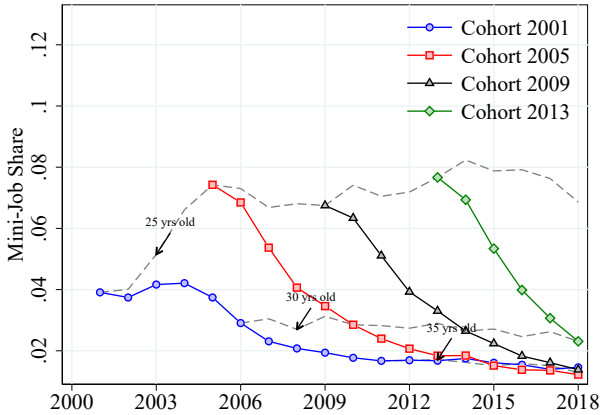
(a) Part-Time Share: Men



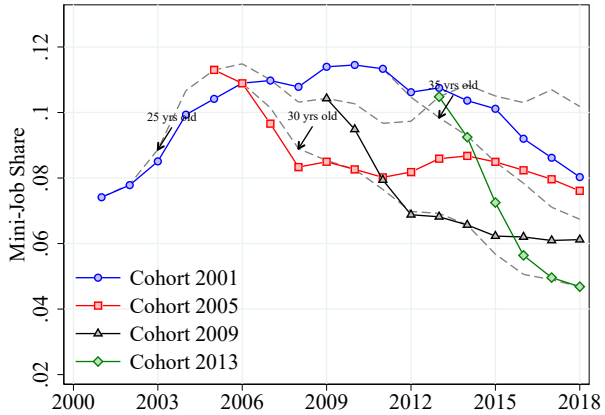
(b) Part-Time Share: Women



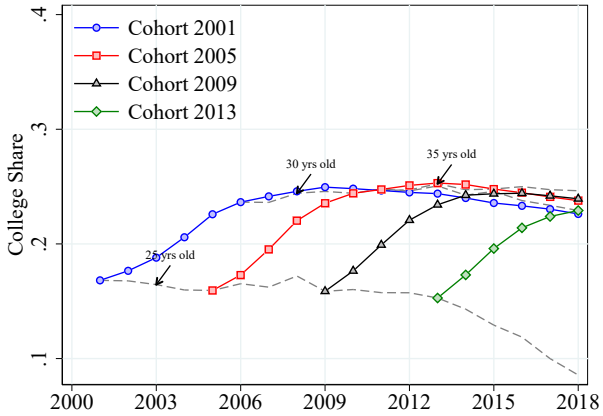
(c) Mini-Job Share: Men



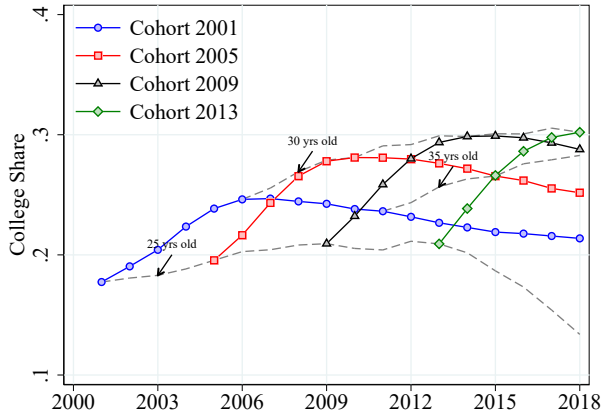
(d) Mini-Job Share: Women



(e) College Share: Men



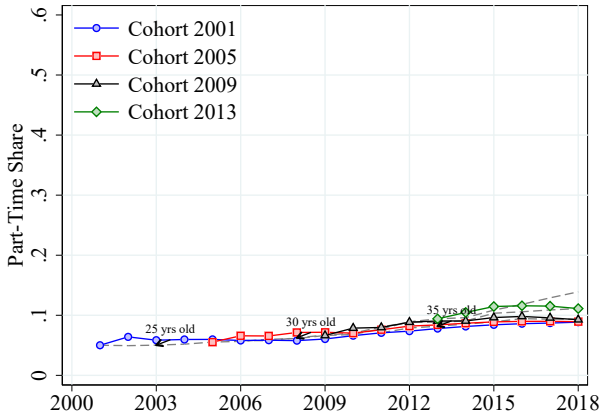
(f) College Share: Women



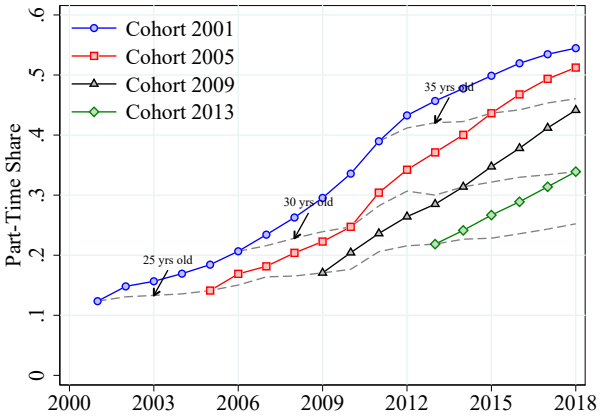
Notes: This figure shows selected emplyoment and education shares in the IAB data (CS sample). Panels A and B show the part-time share over the lifecycle of selected cohorts. Panels C and D show the mini-job share. Panels E and F show the share of college graduates.

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

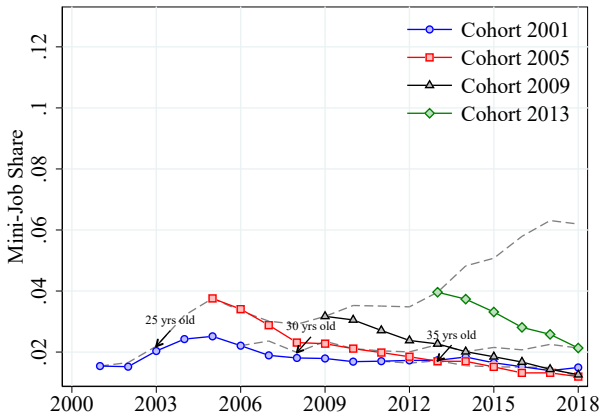
(a) Part-Time Share: Men



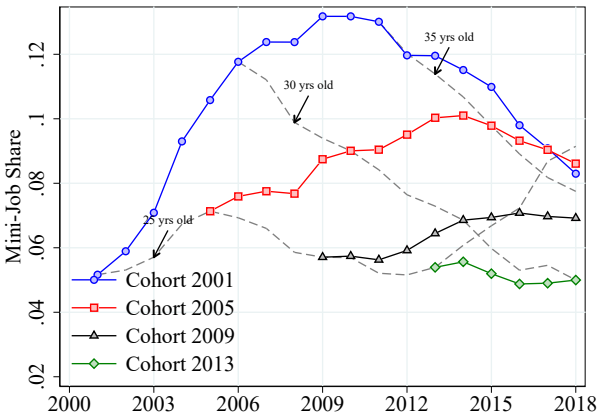
(b) Part-Time Share: Women



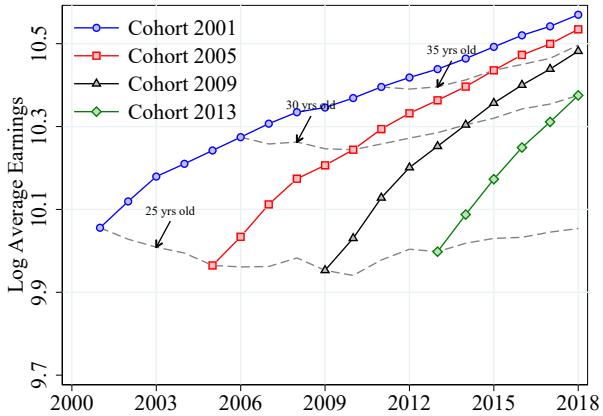
(c) Mini-Job Share: Men



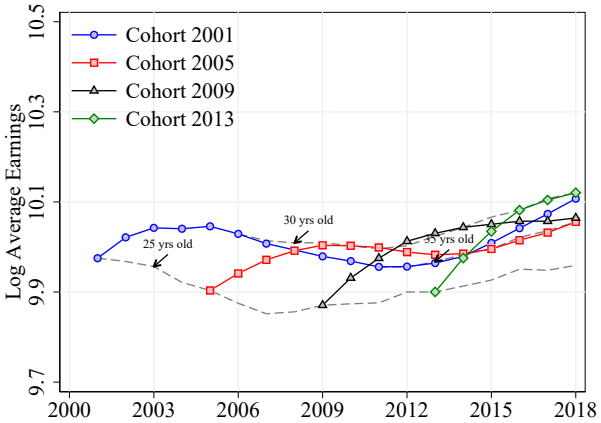
(d) Mini-Job Share: Women



(e) Average Earnings: Men



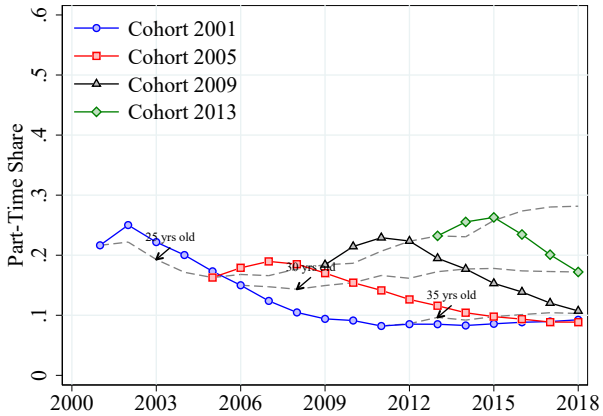
(f) Average Earnings: Women



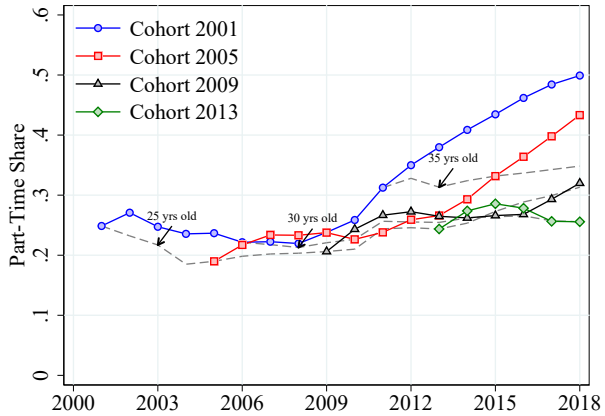
Notes: This figure shows employment levels and average earnings for workers without college degree by cohort in the IAB data (CS sample). Panels A and B show the part-time share over the lifecycle of selected cohorts for non-college workers. Panels C and D show the mini-job share. Panels E and F show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

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

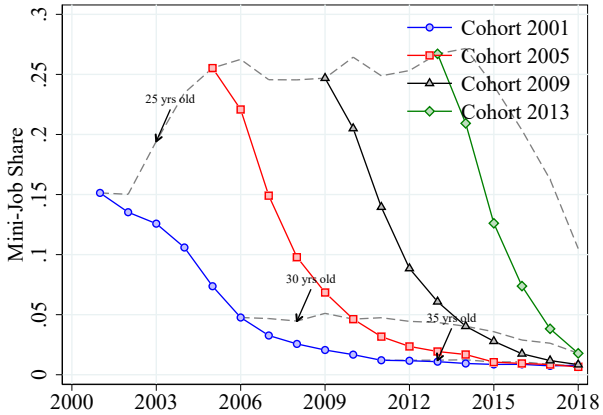
(a) Part-Time Share: Men



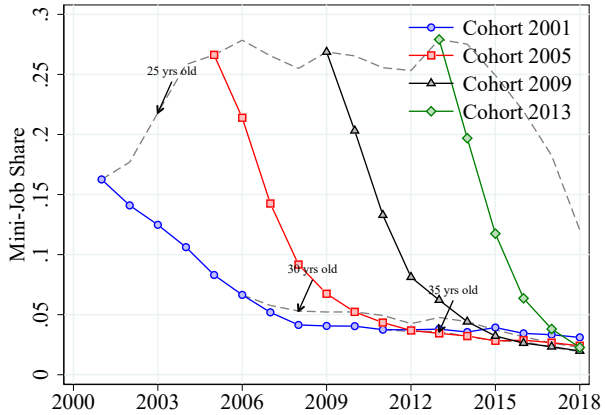
(b) Part-Time Share: Women



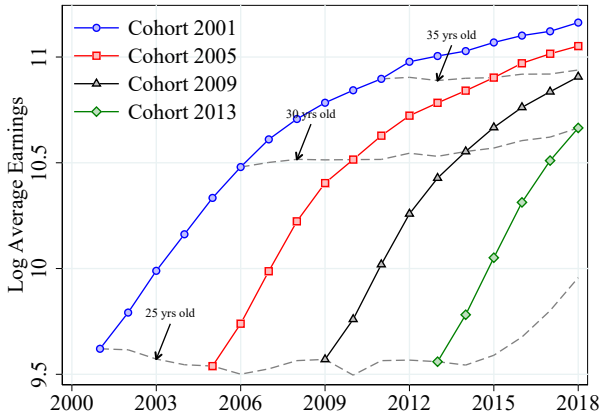
(c) Mini-Job Share: Men



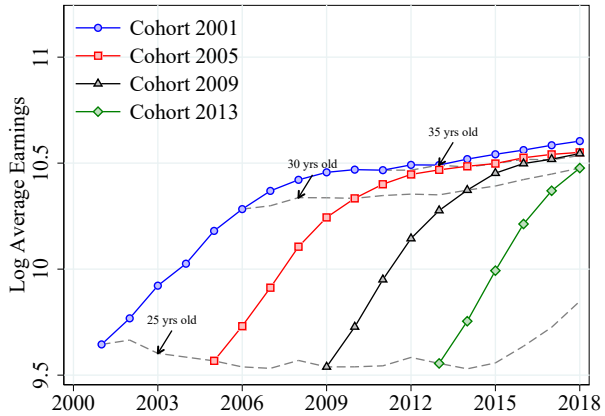
(d) Mini-Job Share: Women



(e) Average Earnings: Men



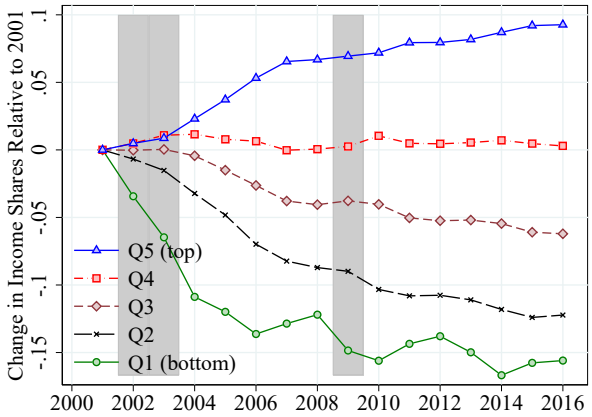
(f) Average Earnings: Women



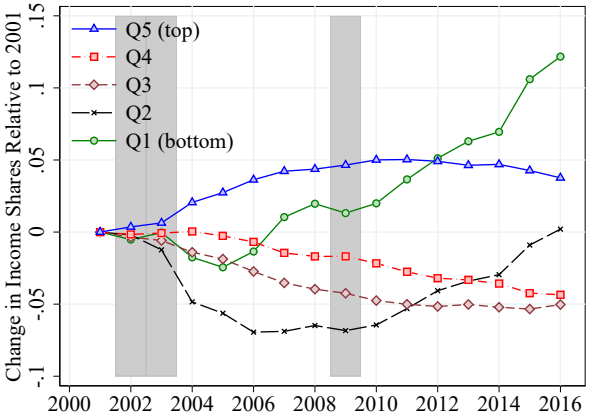
Notes: This figure shows employment levels and average earnings for workers with college degree by cohort in the IAB data (CS sample). Panels A and B show the part-time share over the lifecycle of selected cohorts for college workers. Panels C and D show the mini-job share. Panels E and F show average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

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

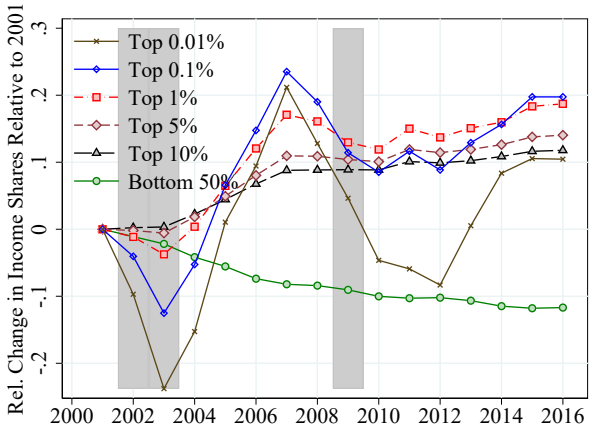
(a) Income Shares of Quintiles: Men



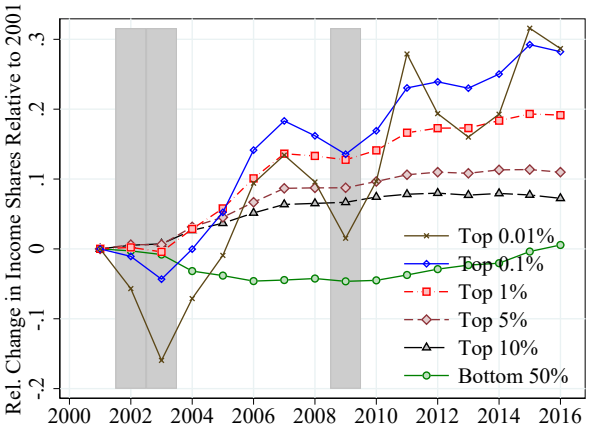
(b) Income Shares of Quintiles: Women



(c) Selected Income Shares: Men



(d) Selected Income Shares: Women



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.

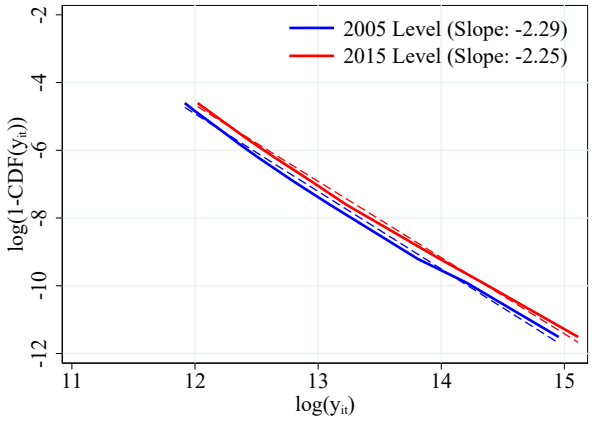
Table E.2: Labor Income Shares – Women

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

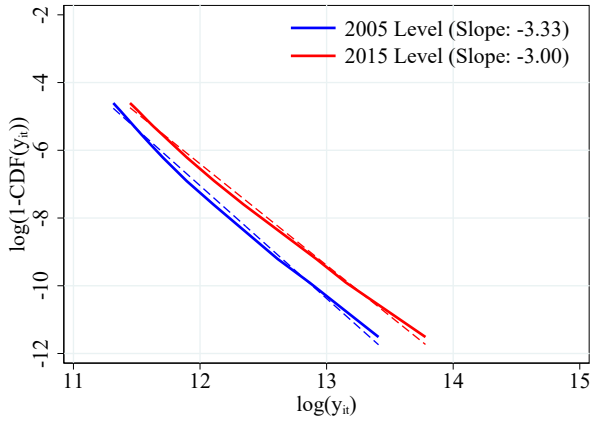
Notes: This table shows the share of earnings that goes to selected parts of the earnings distribution of women in the combined IAB-TPP data (CS sample). Q1 to Q5 refer to the five quintiles where Q1 (Q5) stands for the bottom (top) 20% of the earnings distribution. The quintile shares sum to one. Bot 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%

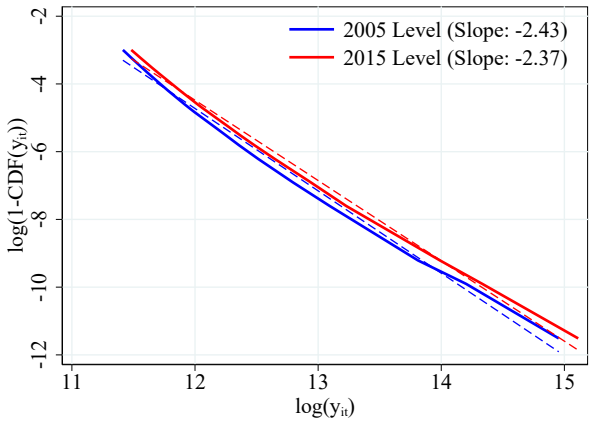
(a) Top 1%: Men



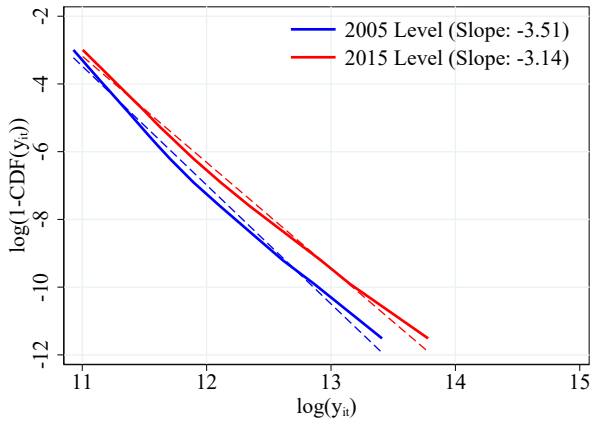
(b) Top 1%: Women



(c) Top 5%: Men



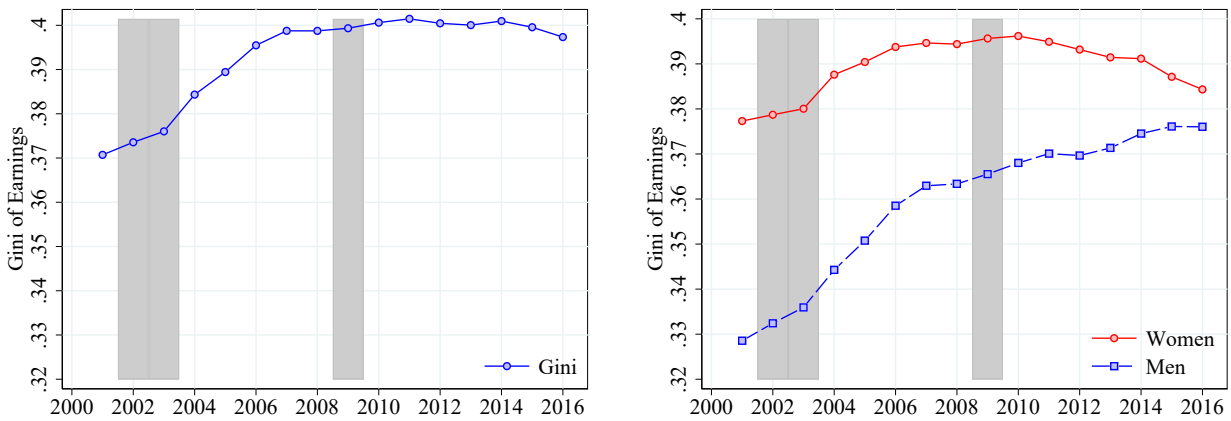
(d) Top 5%: Women



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.

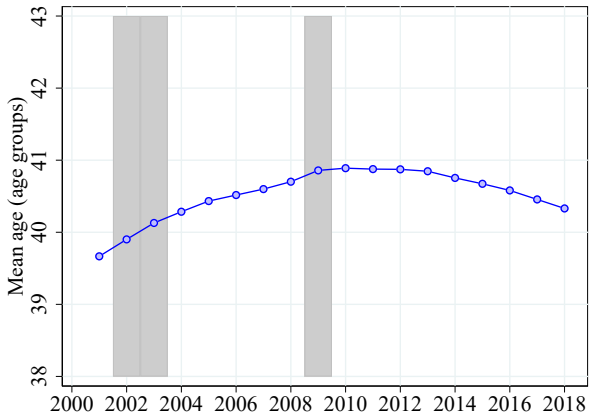
Figure E.15 displays the evolution of the demographic observables age, non-German nationality and educational attainment (2 groups) before reweighting separately for men and women. Mean age increases in the sample until about 2010 before slightly decreasing until 2018 as displayed in Panels A and B. 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.

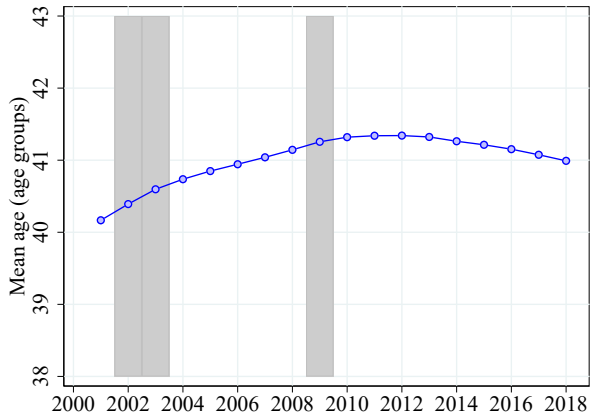
increases from 2001 to 2018. For men it increases from 19% to 21% and for women from from 15% to 21%.

Figure E.15: Weighting Variables: Demographics

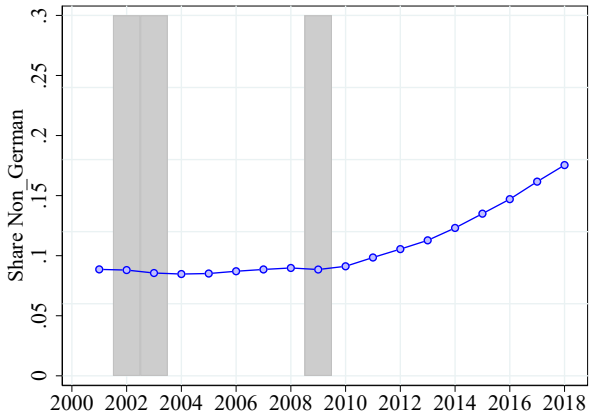
(a) Age: Men



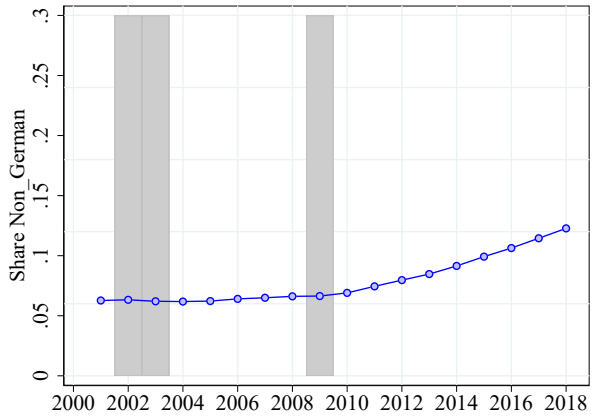
(b) Age: Women



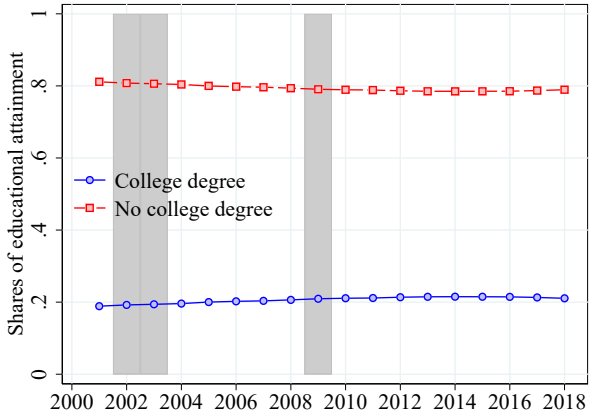
(c) Non-German: Men



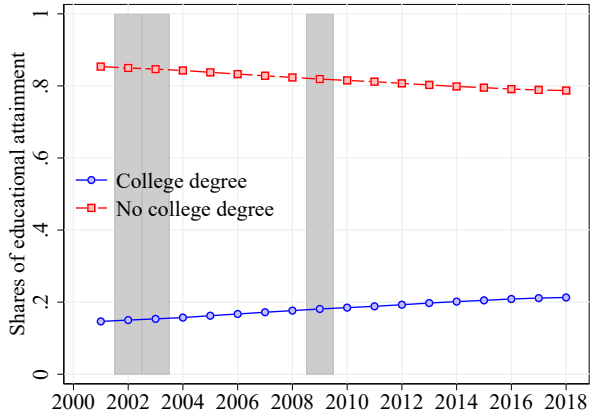
(d) Non-German: Women



(e) Education: Men



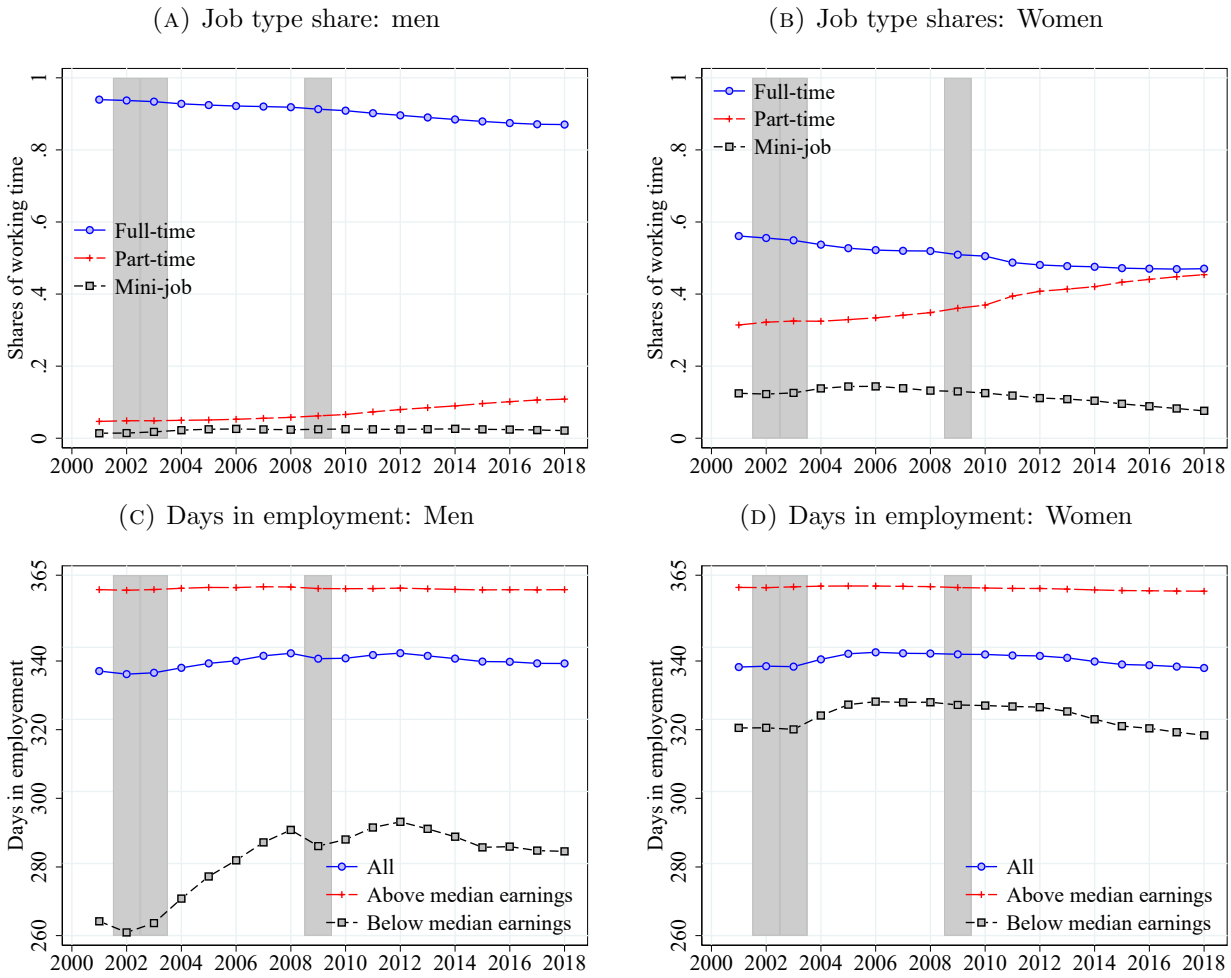
(f) Education: Women



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



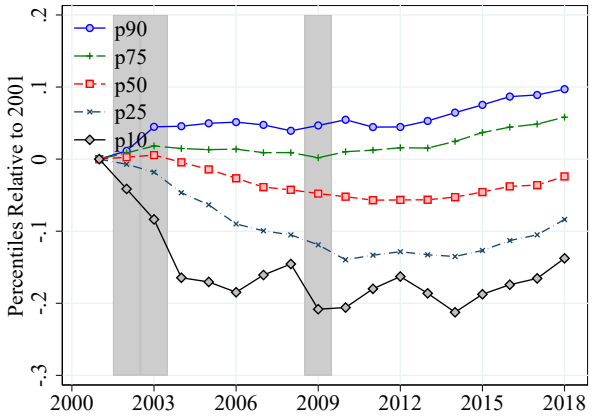
Notes: This figure plots the evolution of work characteristics in the IAB data (CS sample) before and after reweighting for men. For days in employment above and below median earnings, the earnings are weighted by $w = \frac{365}{daysinemployment}$ to account for the positive correlation of earnings 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.

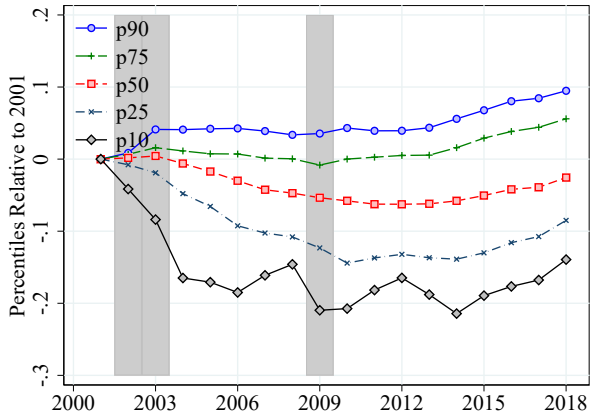
Those tend to affect lower percentiles more. For both, men and women, holding share of full-time, part-time and mini-job workers constant would have compressed the distribution such that percentile evolution appears more compressed. This would have resulted in a more constant evolution of real earnings 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

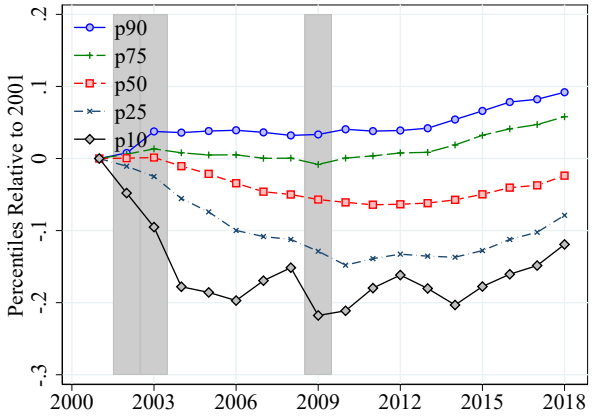
(a) Baseline (unweighted)



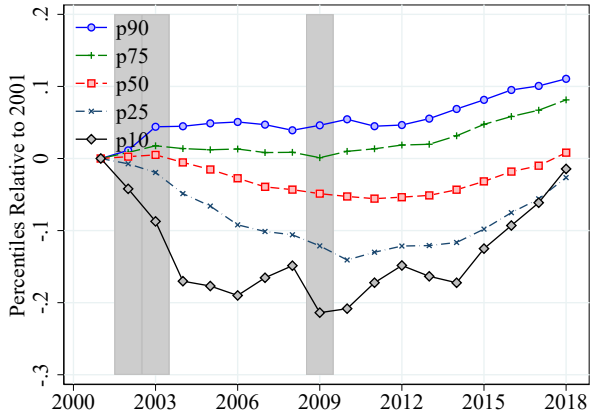
(b) Reweighted for Education



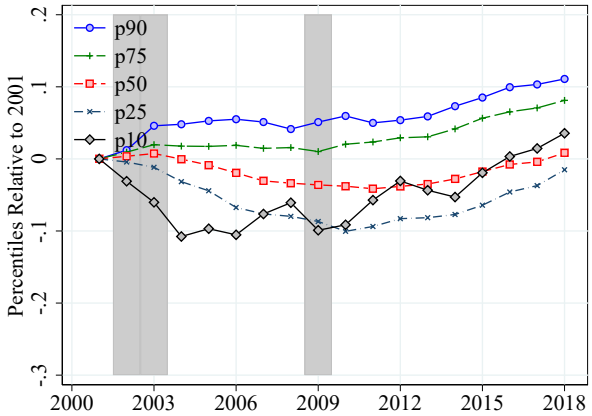
(c) Reweighted for Age



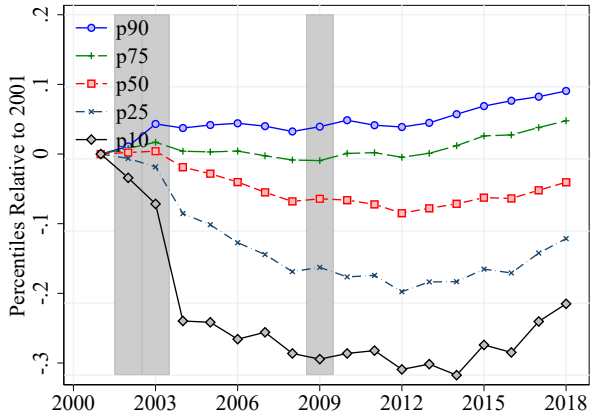
(d) Reweighted for Non-German nationality



(e) Reweighted for Job type



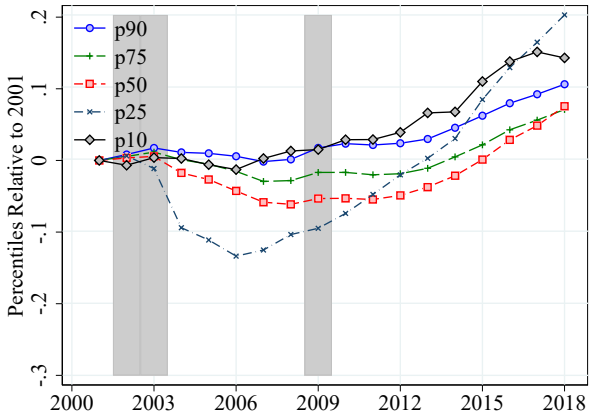
(f) Reweighted for Days in Employment



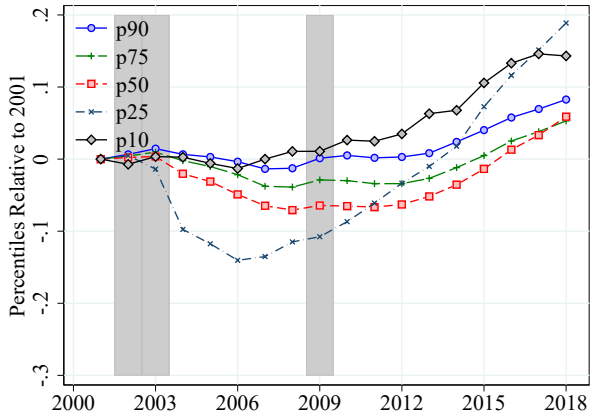
Notes: This figure shows the evolution of different counterfactual log real annual earnings percentiles in the IAB data (CS sample) for men. The counterfactual percentiles are constructed by reweighting the data such that observable dimensions are held constant at the 2001 level. Figure 5 in the main 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

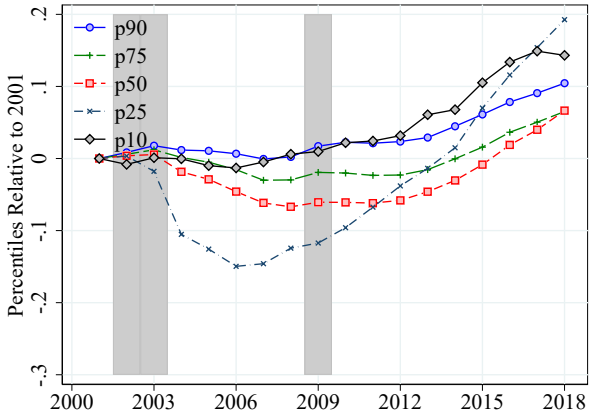
(a) Baseline (unweighted)



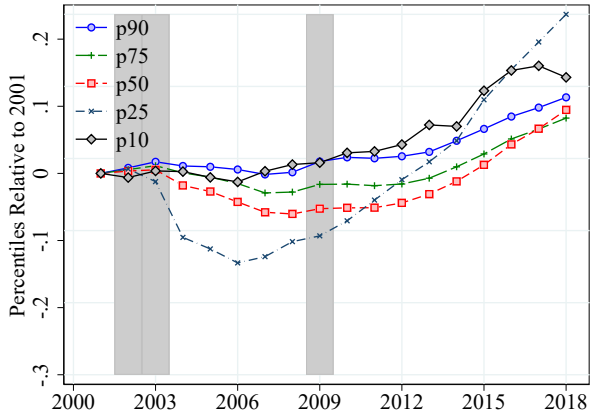
(b) Reweighted for Education



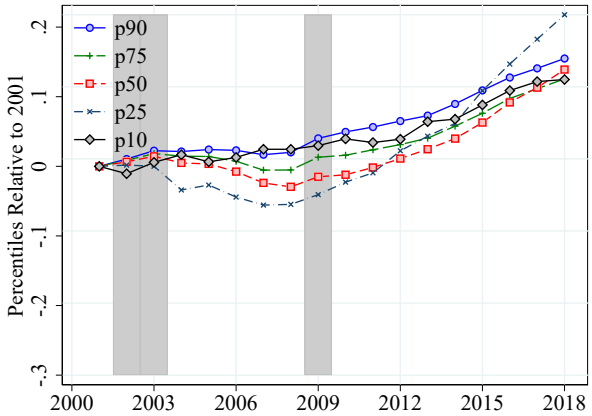
(c) Reweighted for Age



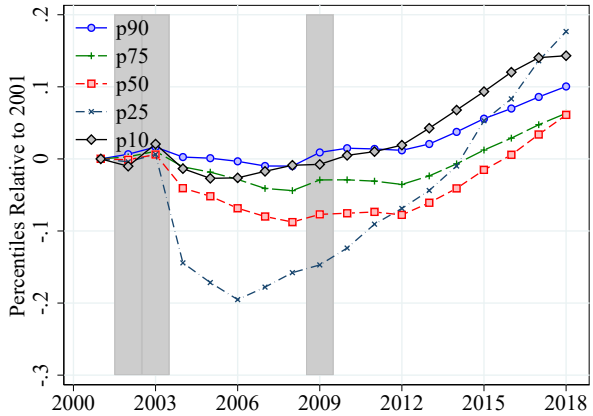
(d) Reweighted for Non-German nationality



(e) Reweighted for Job type



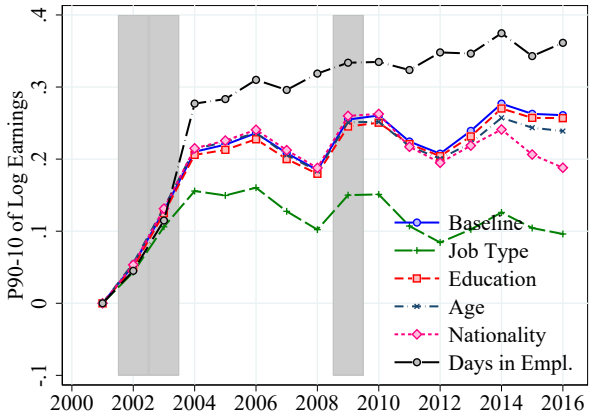
(f) Reweighted for Days in Employment



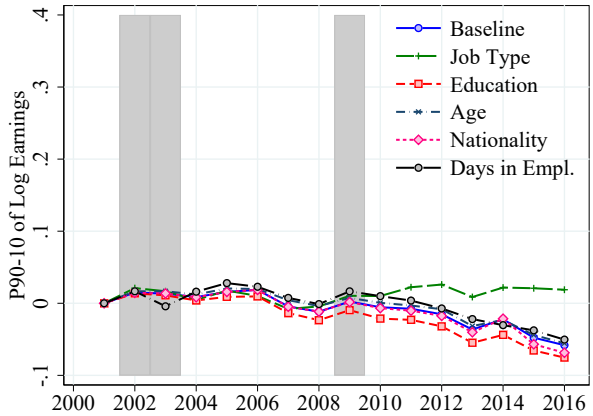
Notes: This figure shows the evolution of different counterfactual log real annual earnings percentiles in the IAB data (CS sample) for women. The counterfactual percentiles are constructed by reweighting the data such that observable dimensions are held constant at the 2001 level. Figure 5 in the main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

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

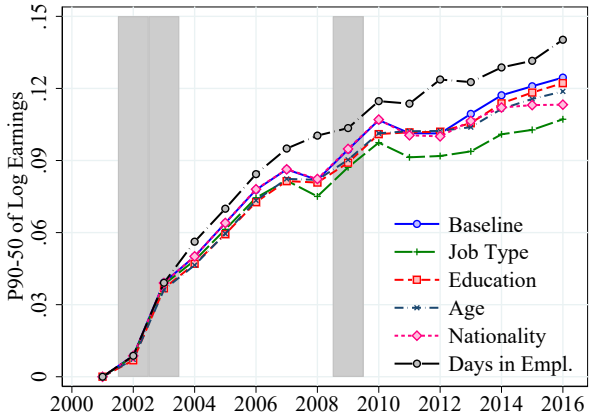
(a) P90-P10: Men



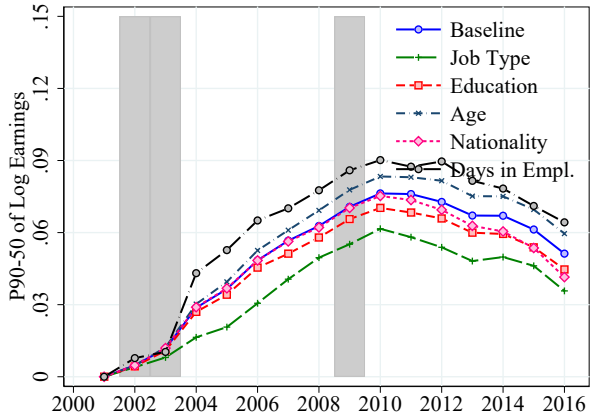
(b) P90-P10: Women



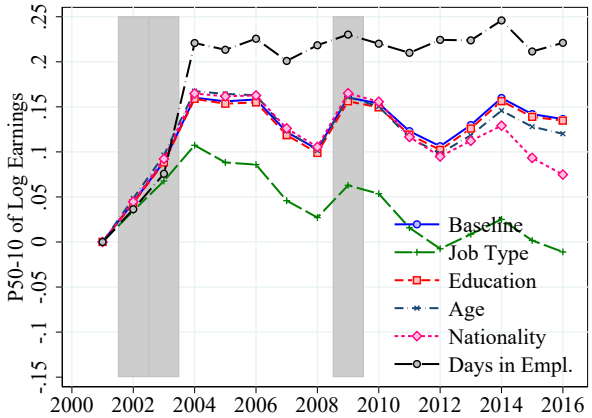
(c) P90-P50: Men



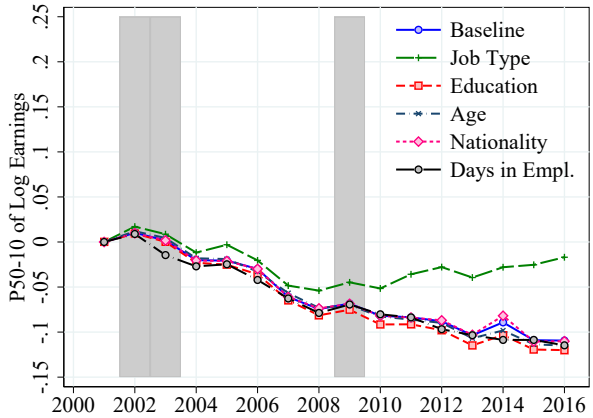
(d) P90-P50: Women



(e) P50-P10: Men



(f) P50-P10: Women

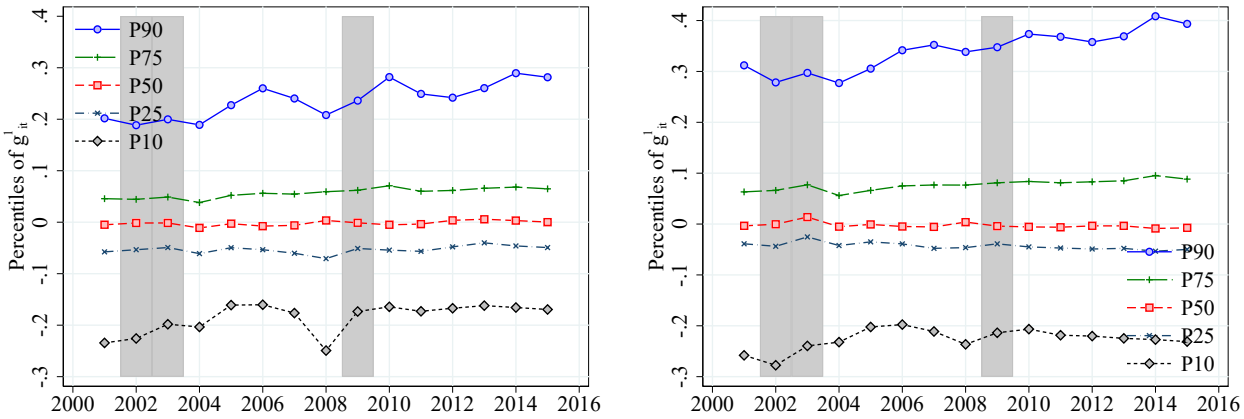


Notes: This figure shows the evolution of different counterfactual percentile differences of the log real annual earnings distribution over time in the IAB data (CS sample) separately for men and women. The counterfactual percentiles are constructed by reweighting the data such that observable dimensions 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)

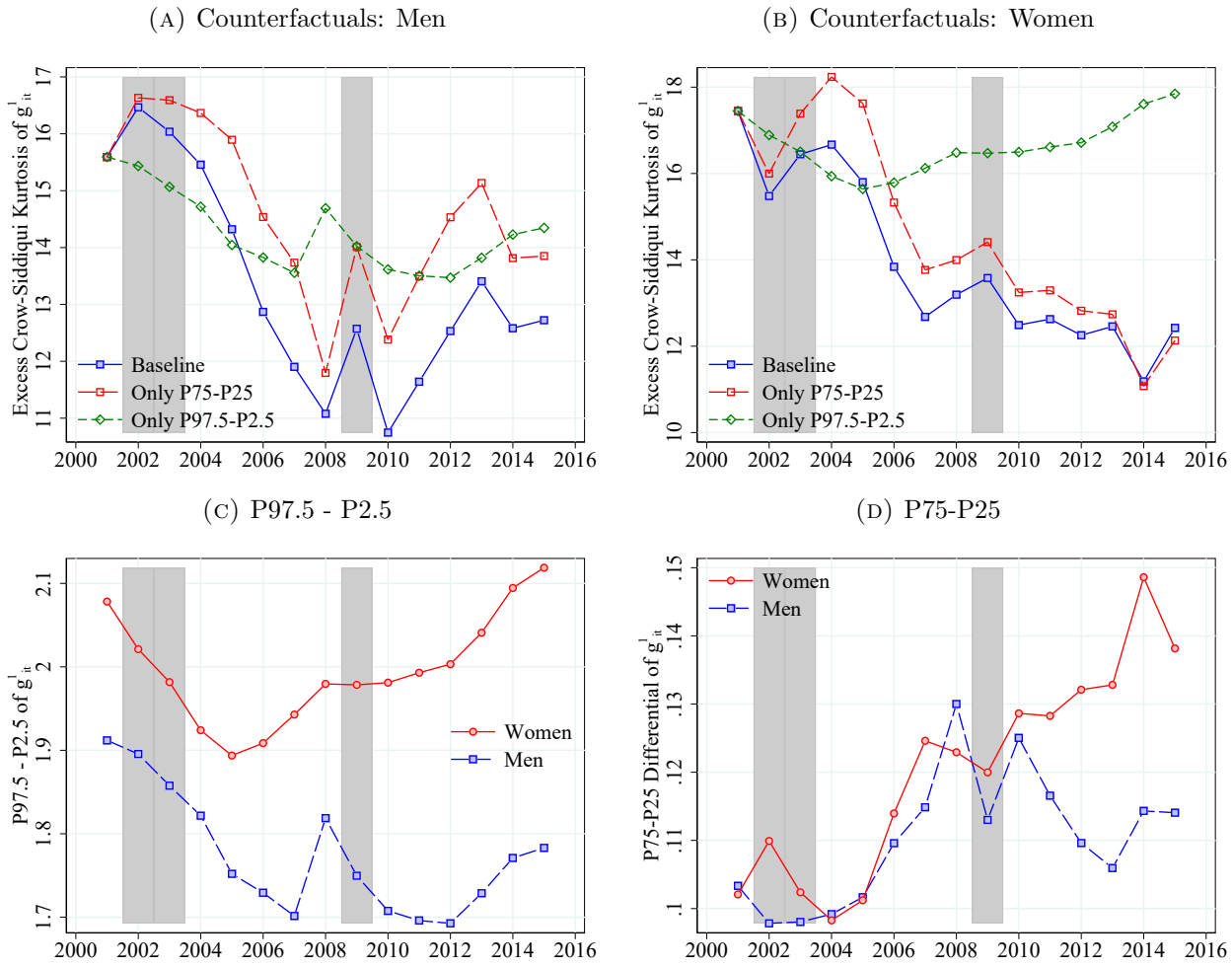
Figure E.20: Percentiles of 1-Year Log Earnings Changes

(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.

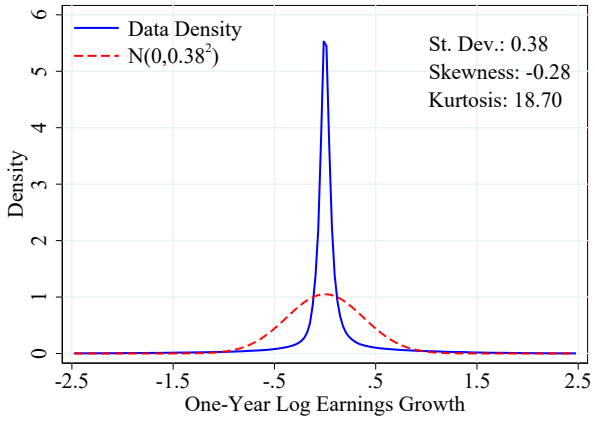
Figure E.21: Decomposition of Excess Crow-Siddiqui Kurtosis of 1-Year Log Earnings Changes



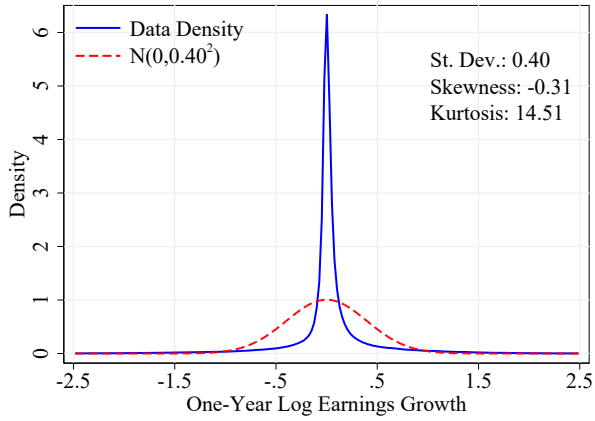
Notes: This figure shows decomposition analyses of the excess Crow-Siddiqui kurtosis in the combined IAB-TPP data (LS sample). Panels A and B show how the excess Crow-Siddiqui kurtosis of 1-year residualized log earnings changes (from t to $t+1$) would have evolved if only the numerator (P97.5–P2.5) 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)

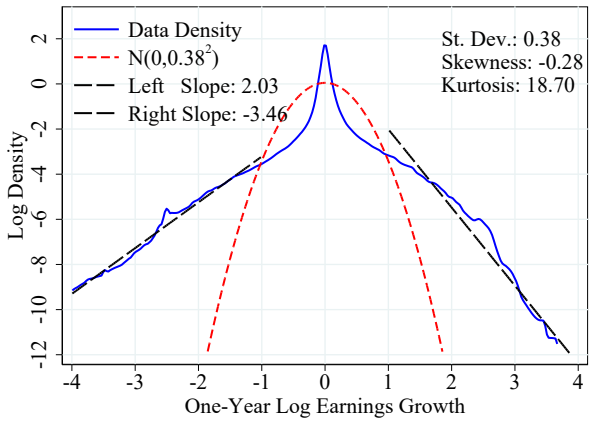
(a) Density: Men



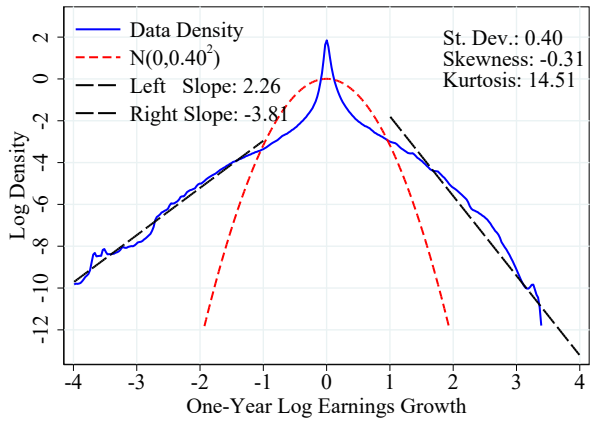
(b) Density: Women



(c) Log Density: Men

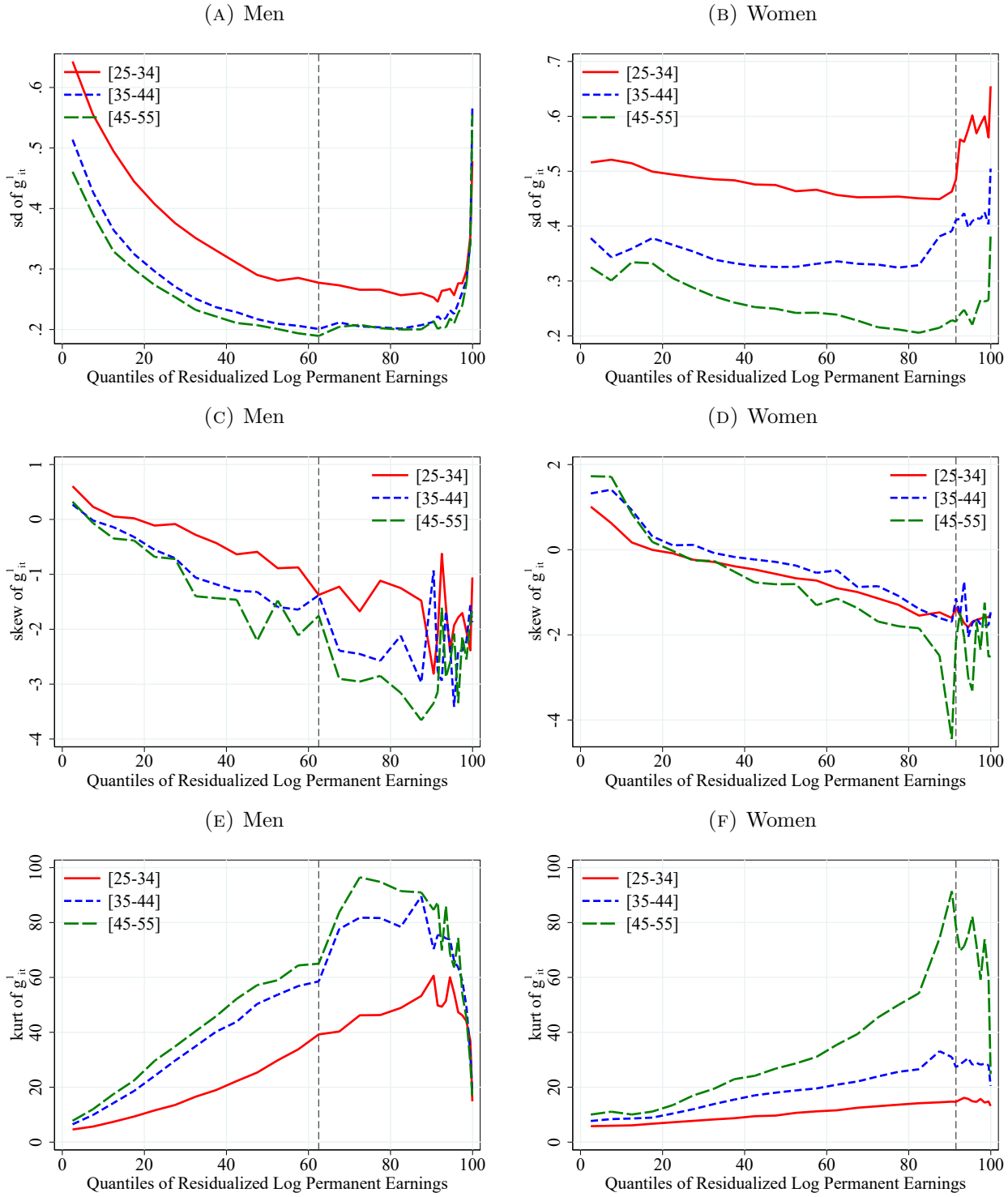


(d) Log Density: Women



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).

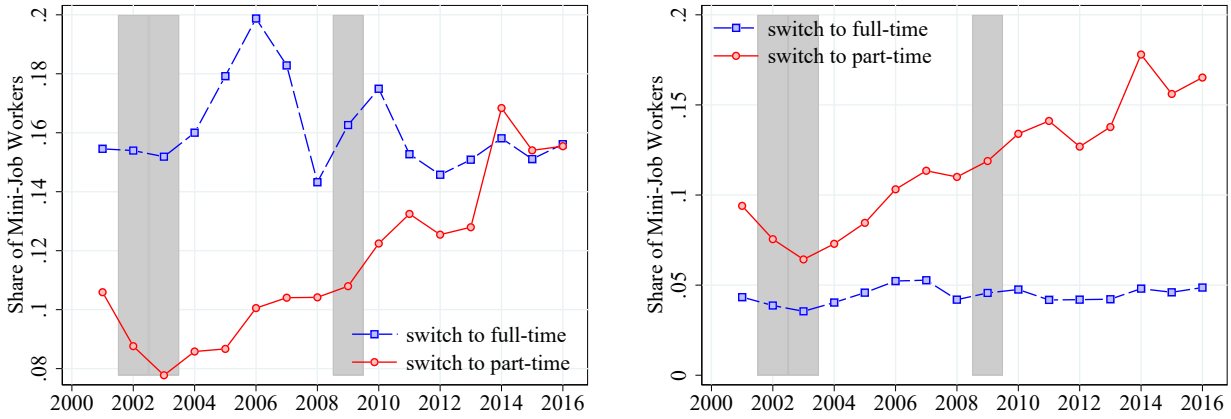
Figure E.23: Heterogeneity in Standardized Moments of 1-Year Log Earnings Changes



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

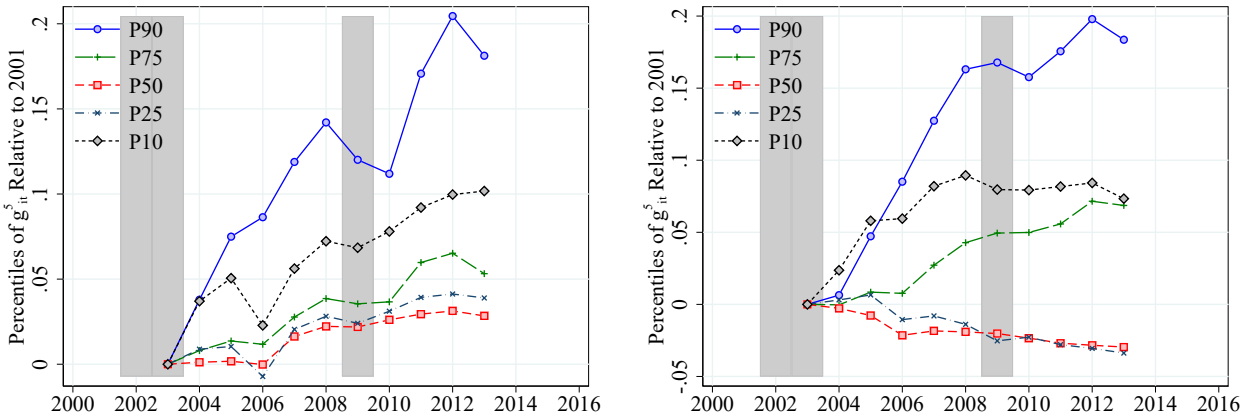
(a) Men (b) Women



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

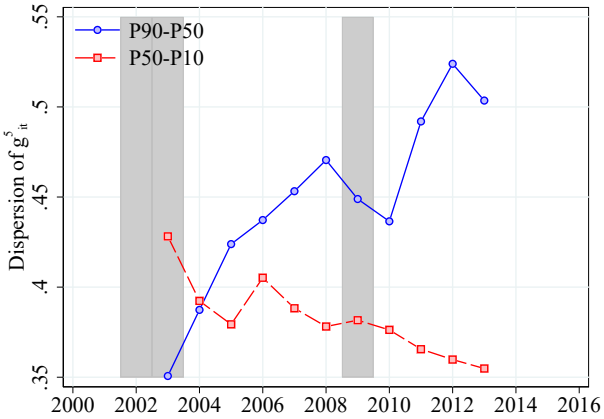
(a) Men (b) Women



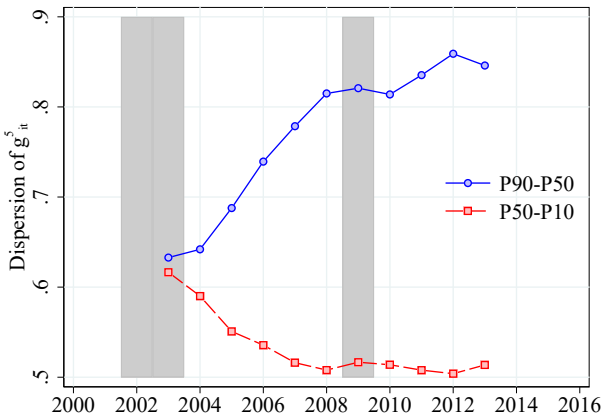
Notes: This figure shows selected percentiles of the distribution of 5-year changes in residualized log real annual earnings (from t to $t+1$) in the combined IAB-TPP data (LS sample) separately for men and women. Shaded areas indicate recessions. See Appendix D.2.2 for details on how we construct the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

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

(a) Men



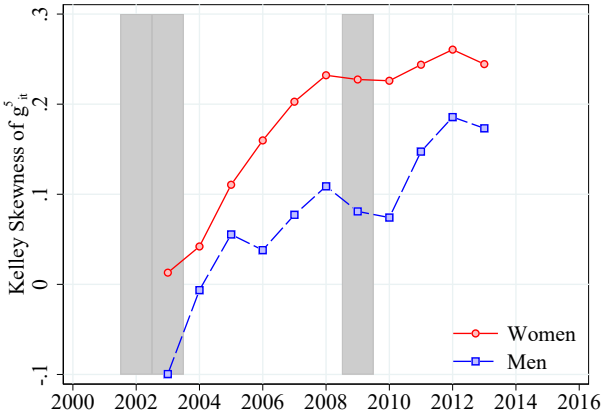
(b) Women



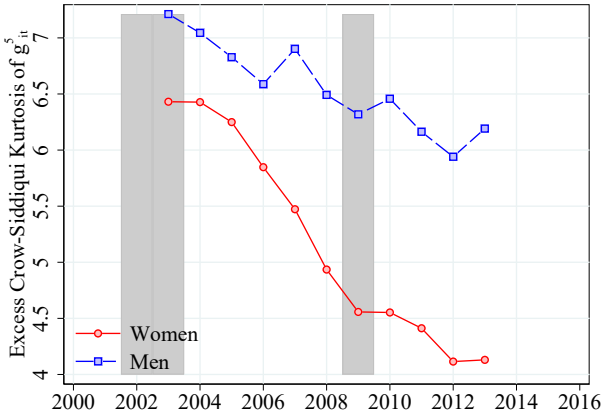
Notes: This figure shows 5-year changes in residualized log earnings (from $t-2$ to $t+3$) in the combined IAB-TPP data (LS sample). Shaded areas indicate recessions.

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

(a) Kelley Skewness

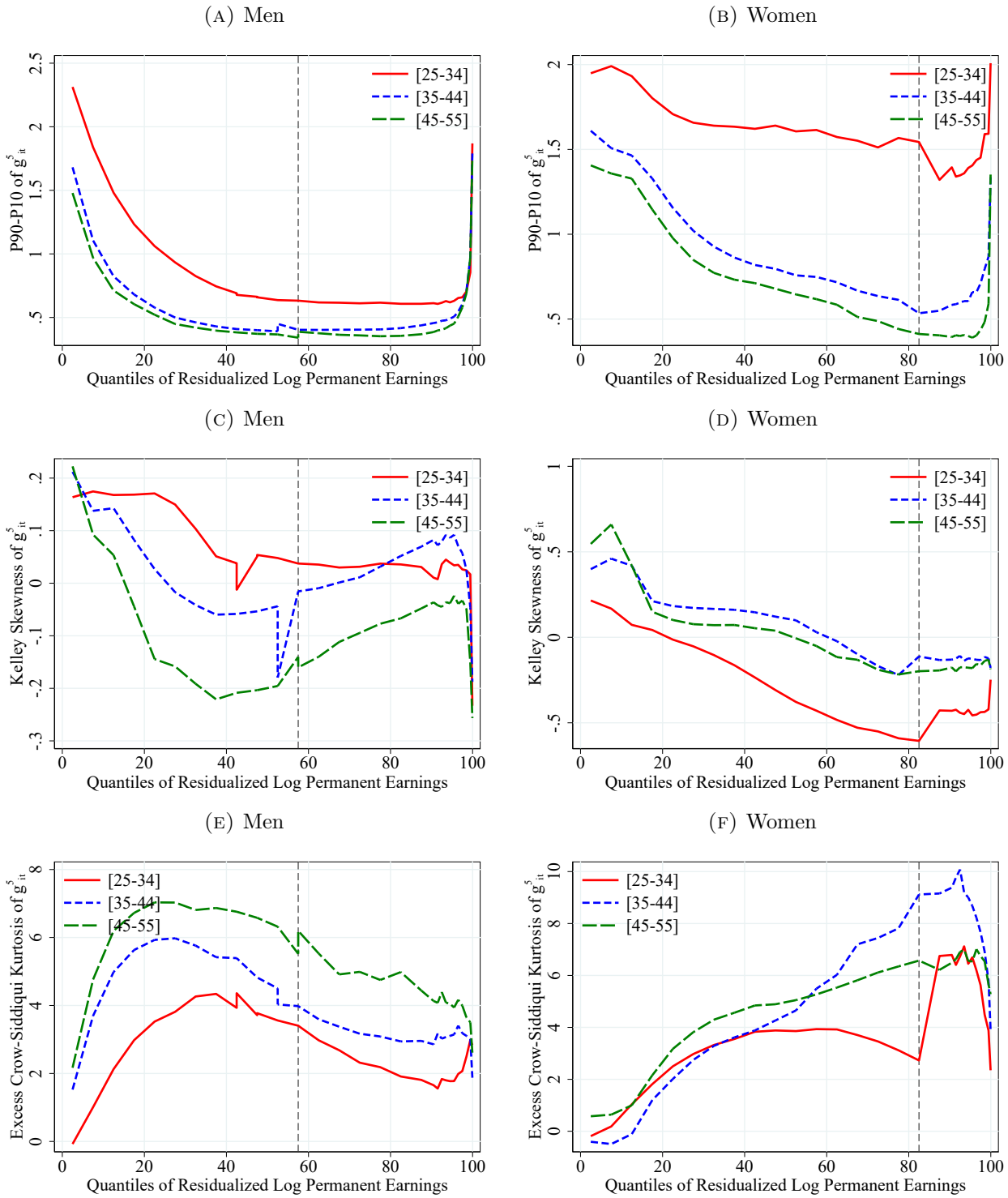


(b) Excess Crow-Siddiqui Kurtosis



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.

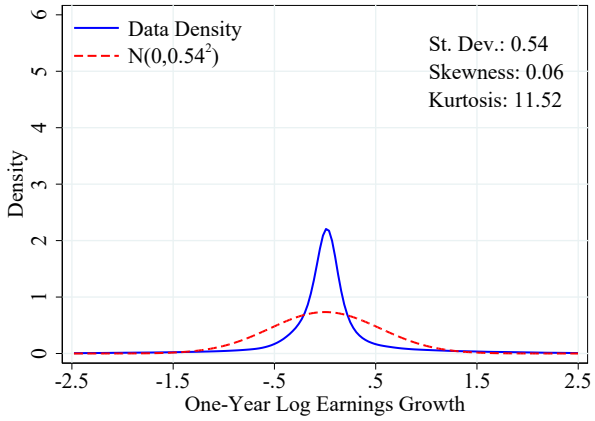
Figure E.28: Heterogeneity in Dispersion, Skewness and Kurtosis of 5-Year Log Earnings Changes



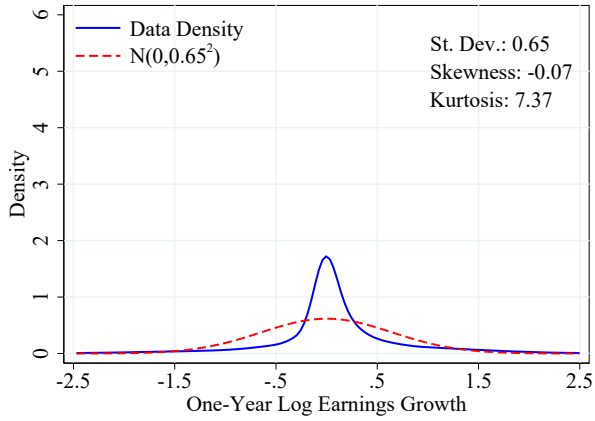
This figure shows the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 5-year changes in residualized log earnings (from $t-2$ to $t+3$) in the combined IAB-TPP data (H sample) as averages from 2004 to 2011 by quantiles of residualized permanent earnings and age groups. Kelley 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)

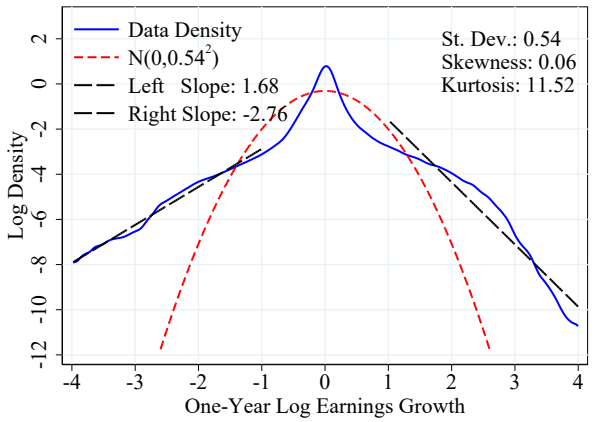
(a) Density: Men



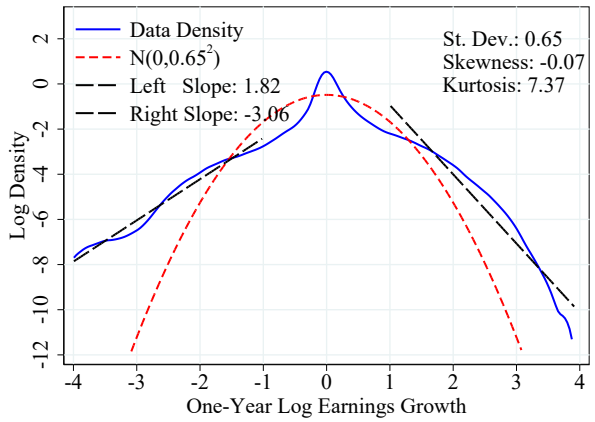
(b) Density: Women



(c) Log Density: Men



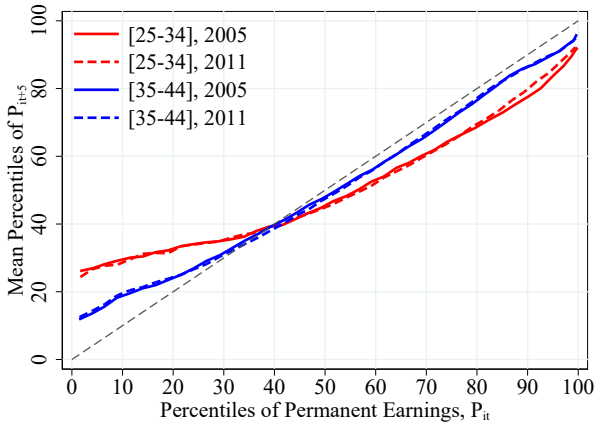
(d) Log Density: Women



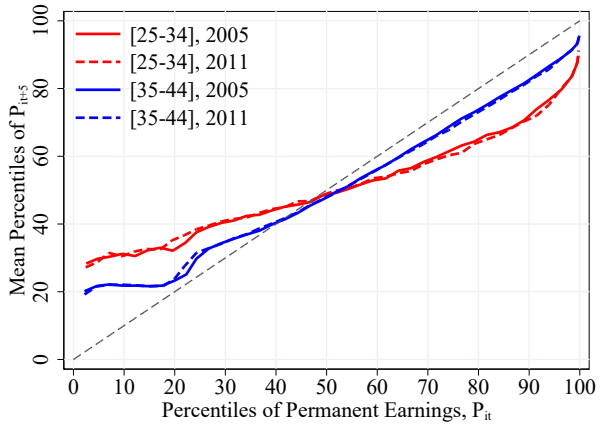
Notes: This figure shows Kernel density estimates of 5-year changes in residualized log earnings (from $t-2$ to $t+3$) for the year 2005 and the respective density of a Normal distribution with zero mean and the same standard deviation as in the combined IAB-TPP data (LS sample).

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

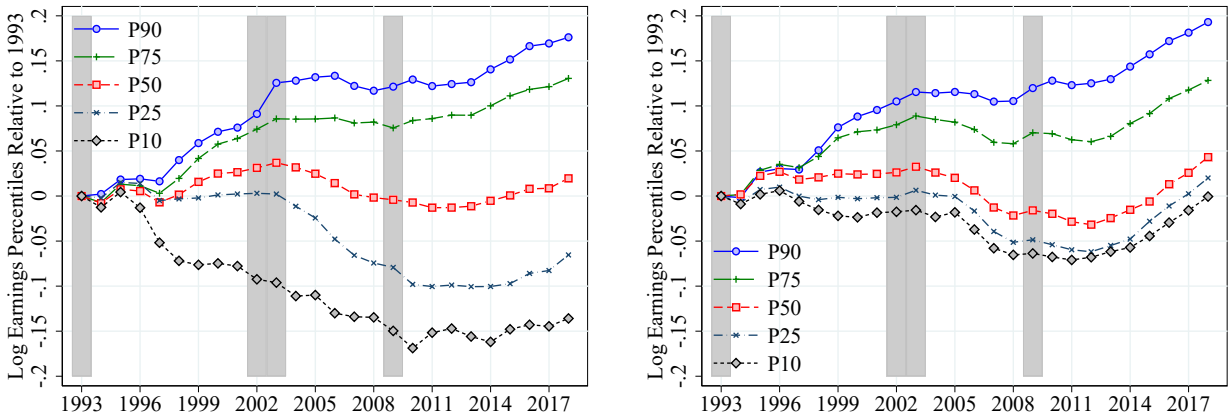
In this section, we present figures similar to those of the core analysis of this paper in Section 3 for longer samples based on IAB data only.

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

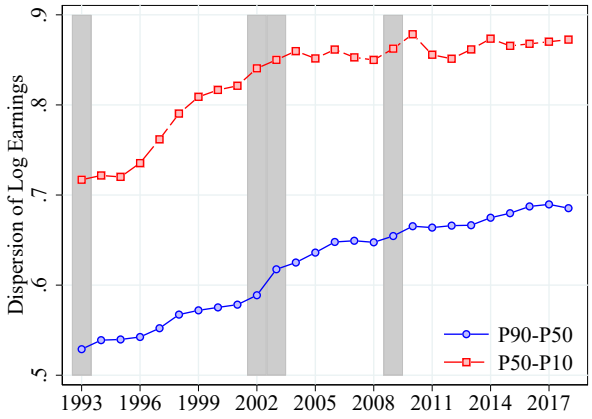
(a) Men (b) Women



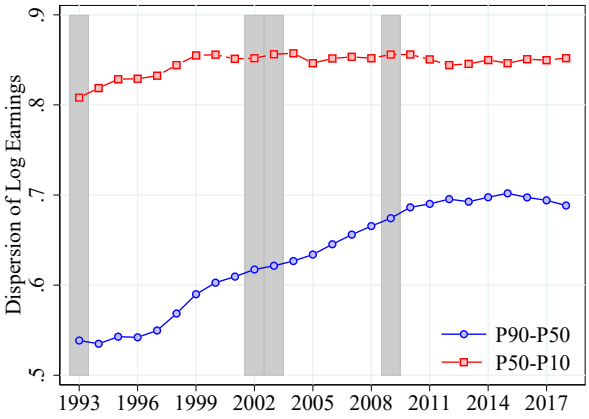
Notes: This figure shows the evolution of selected percentiles of log real earnings from 1993 to 2018 in the IAB data (CS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. All percentiles are normalized to 0 in 1993. Shaded areas indicate recessions. 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



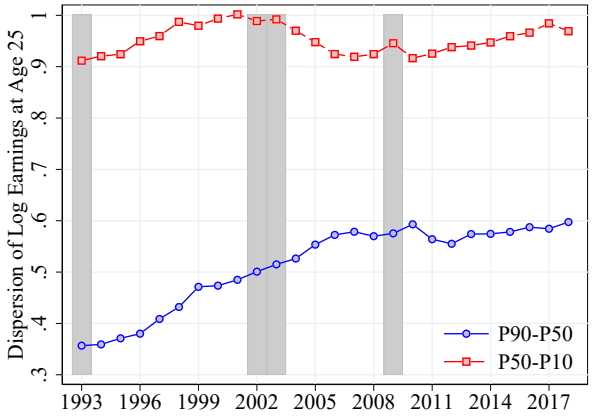
(b) Upper and Lower Inequality: Women



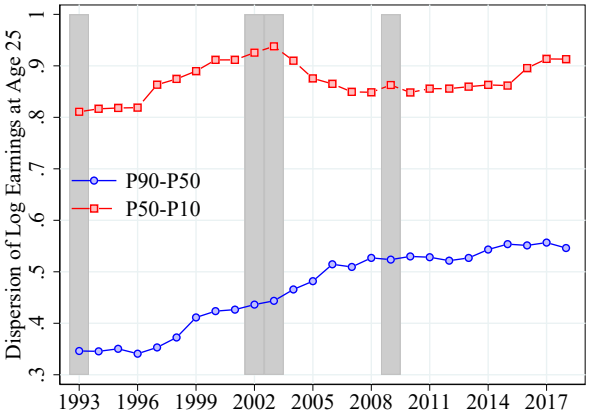
Notes: This figure shows percentile differentials of log real annual earnings in the IAB data (CS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main 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

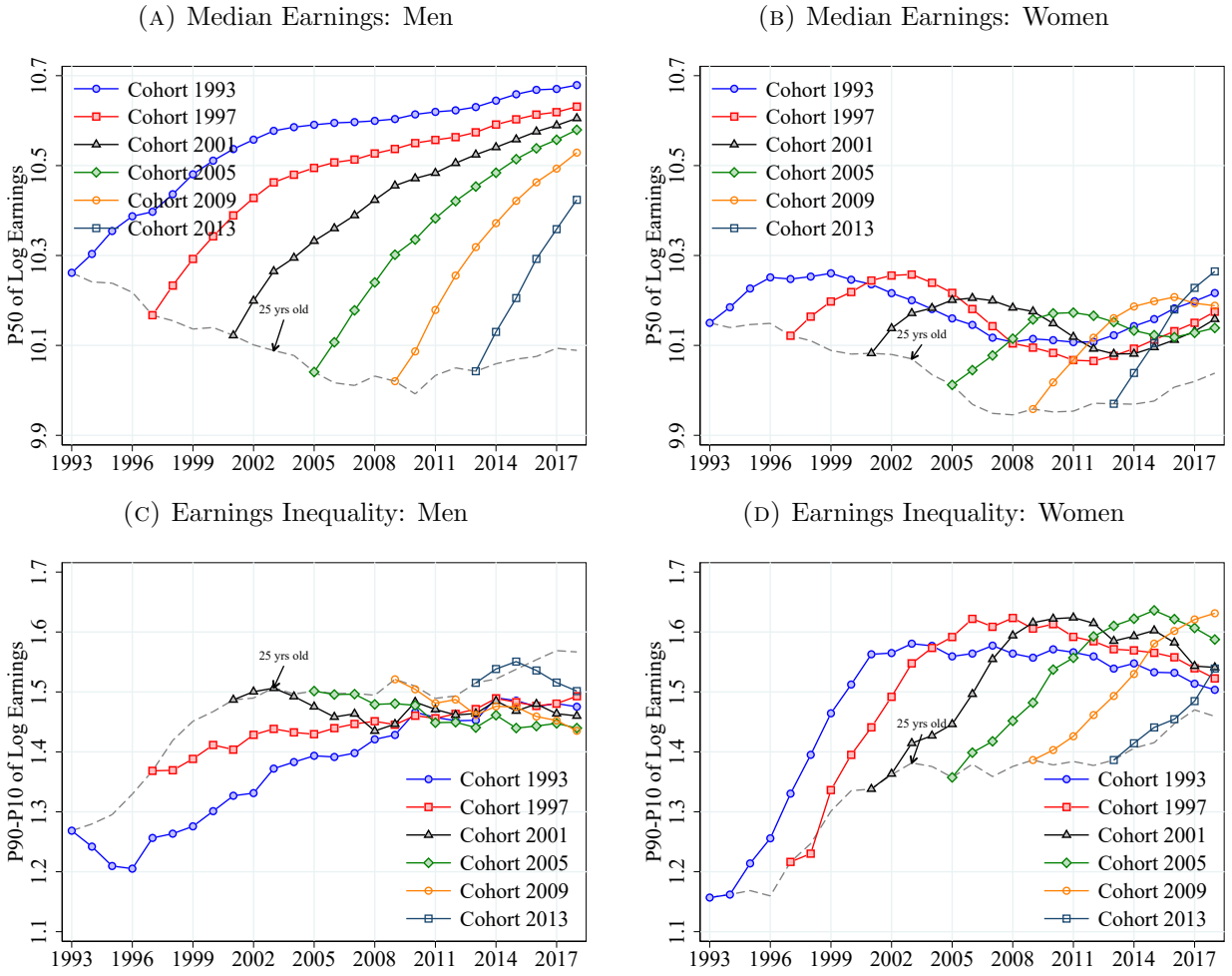


(b) Upper & Lower Inequality: Women



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.

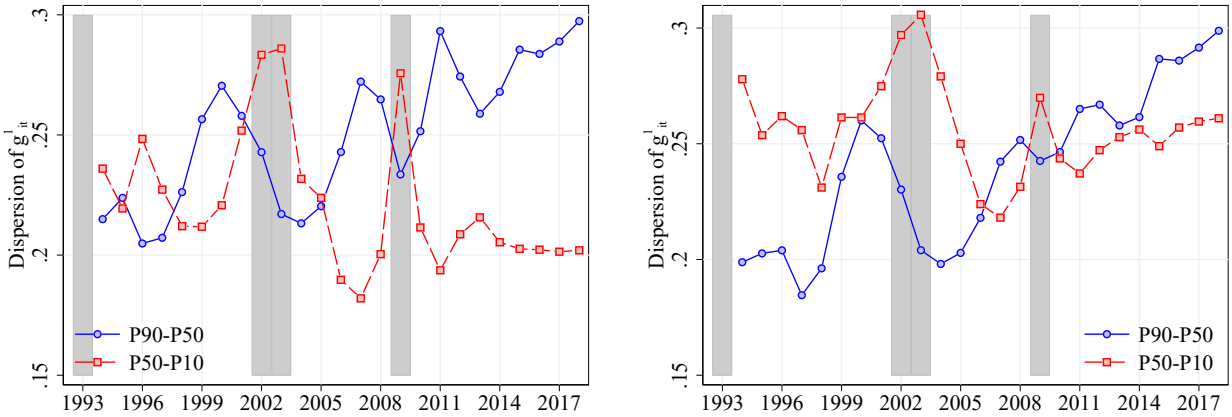
Figure F.4: Earnings Profiles and Inequality by Cohort



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



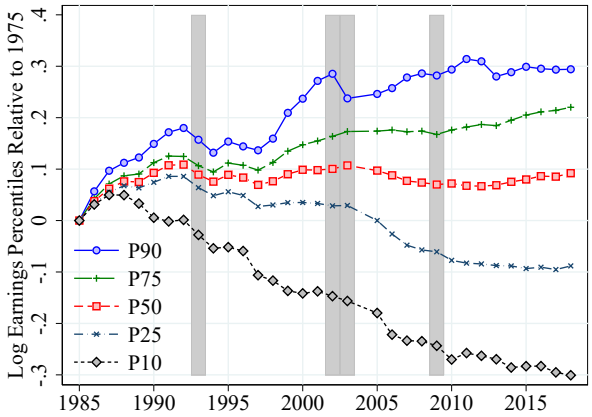
Notes: This figure shows the the P90-P50 and P50-P10 differentials of the distribution of 1-year changes in residualized log earnings (from $t-1$ to t) in the IAB data (LS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. LS sample with minimum 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)

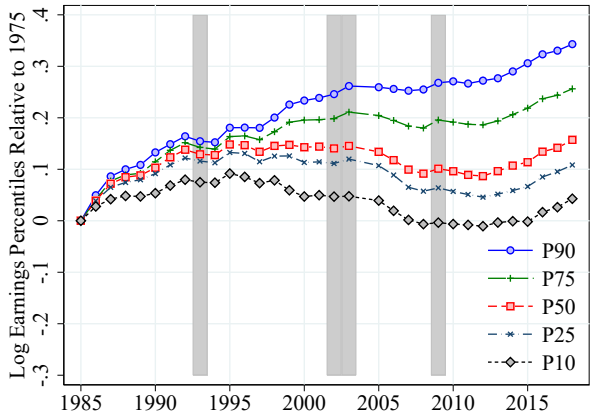
For our longest sample we use the SIAB 1975-2019 (Frodermann et al., 2021) for the years 1985 to 2018. We start in 1985 due to a structural break in the data in 1984. We apply the same minimum earnings threshold of 6,250 in 2018 Euro to exclude mini-jobs from the data. Furthermore, as data for East Germany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



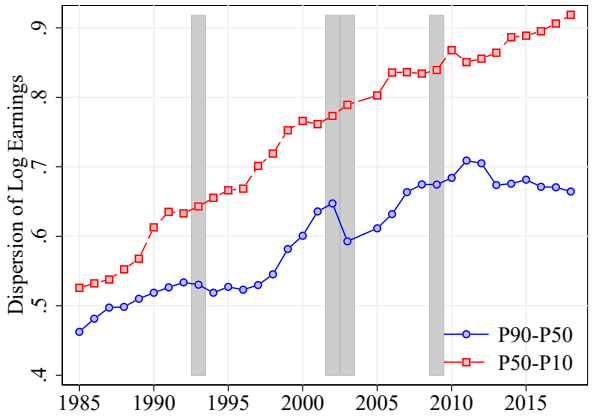
(b) Women



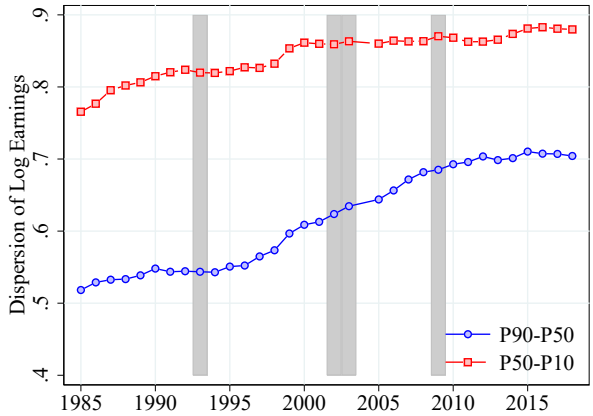
Notes: This figure shows the evolution of selected percentiles of log real earnings from 1985 to 2018 in the IAB data (CS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. All percentiles are normalized to 0 in 1985. CS sample with minimum income 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



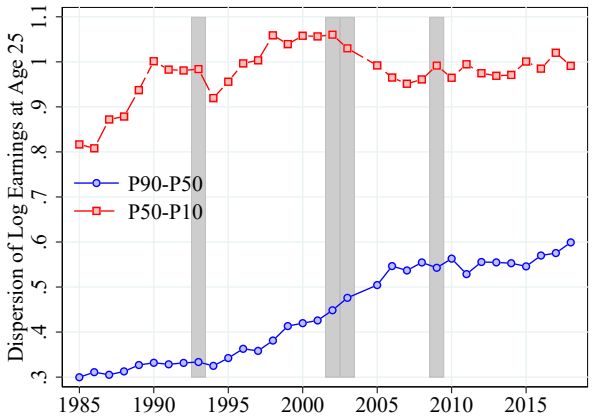
(b) Upper and Lower Inequality: Women



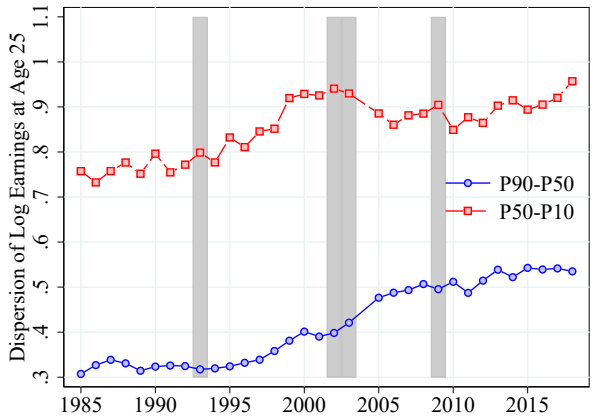
Notes: This figure shows percentile differentials of log real annual earnings in the IAB data (CS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. CS sample with minimum income threshold of 6,250 Euro (2018 prices). The CS sample in the main 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

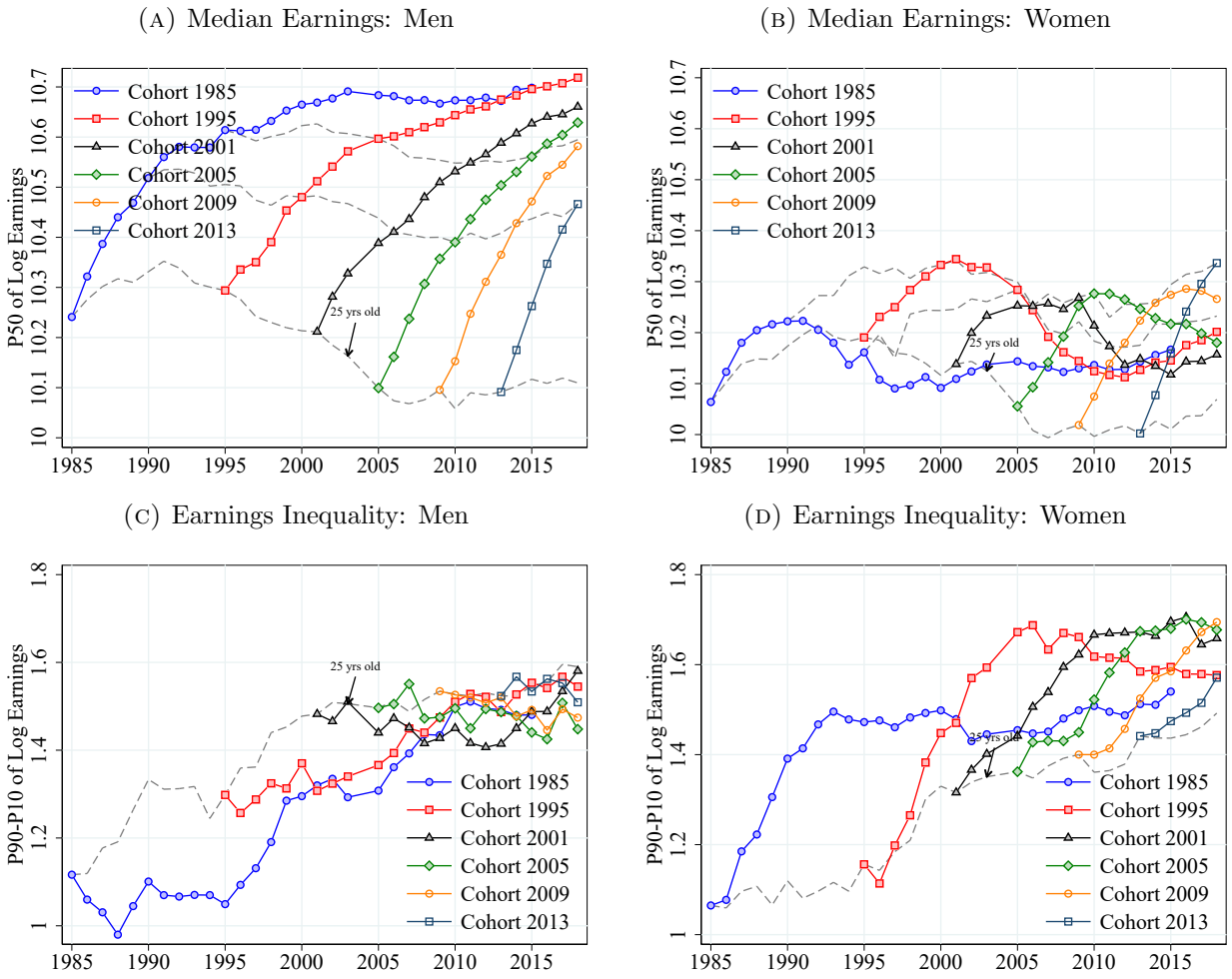


(b) Upper & Lower Inequality: Women



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.

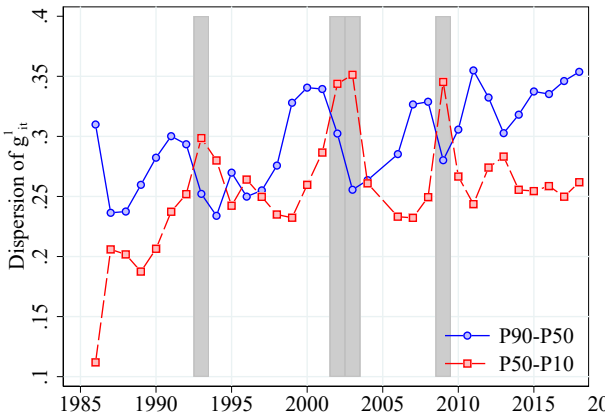
Figure F.9: Earnings Profiles and Inequality by Cohort



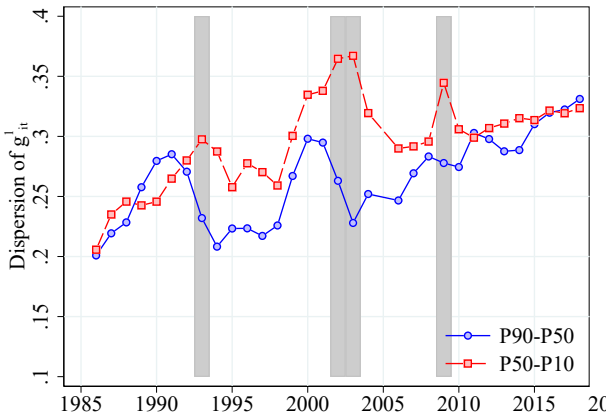
Notes: This figure shows the evolution of the median (P50) as well as the P90-P10 differential of the log real annual earnings distribution over time in the IAB data (CS sample, truncated as stated below) separately for men and women. Each colored line corresponds to an individual cohort, where "cohort 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

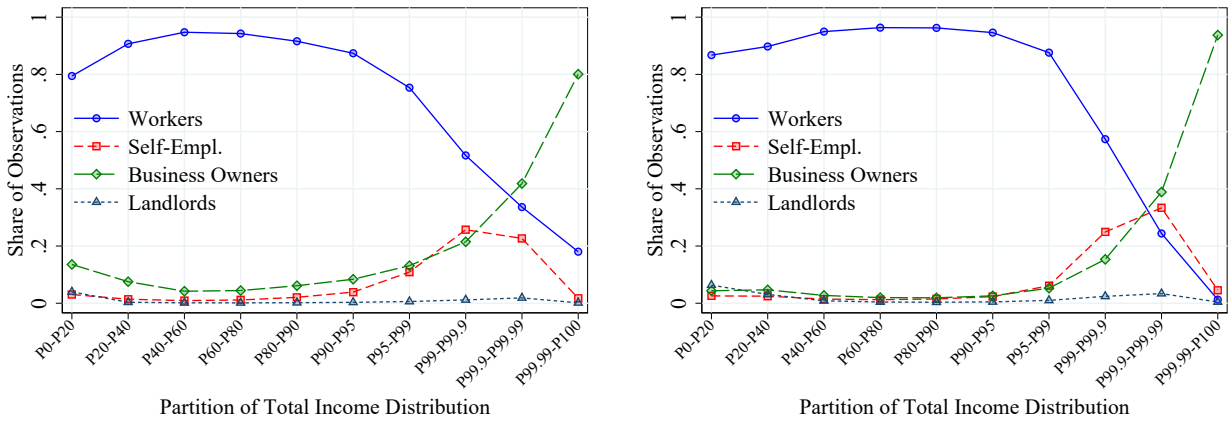


Notes: This figure shows the the P90-P50 and P50-P10 differentials of the distribution of 1-year changes in residualized log earnings (from $t-1$ to t) in the IAB data (LS sample, truncated as stated below). The P90 for men is above the top-coding threshold and therefore imputed. LS sample with minimum 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

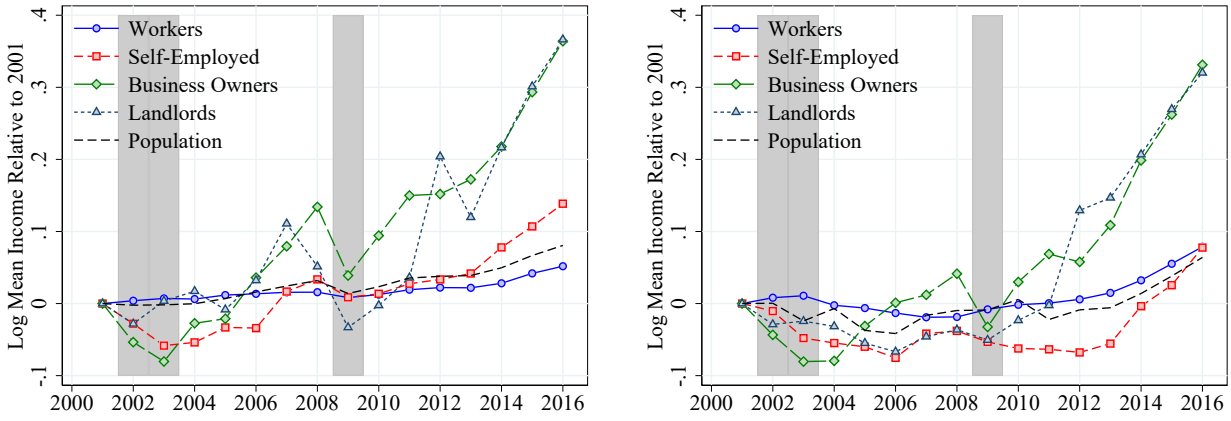
(a) Men (b) Women



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

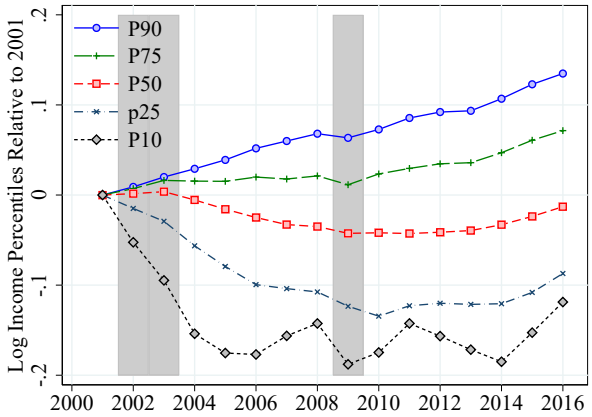
(a) Men (b) Women



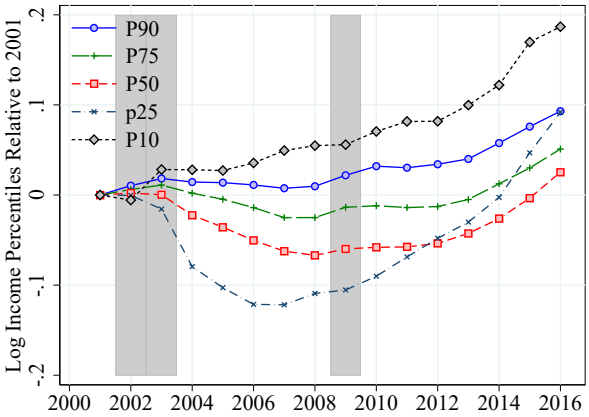
Notes: This figure shows the evolution the log of average real annual total income (relative to 2001) in the combined IAB-TPP data (CS analysis sample) by main income source separately for men and women. Shaded areas indicate recessions. See Figure 11 for corresponding levels.

Figure G.3: Evolution of Log Total Income Percentiles

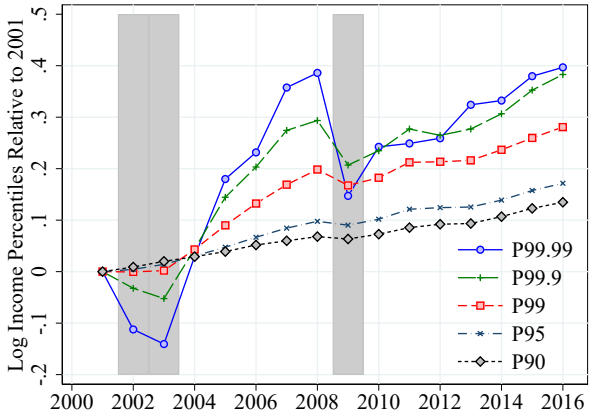
(a) Overall Distribution: Men



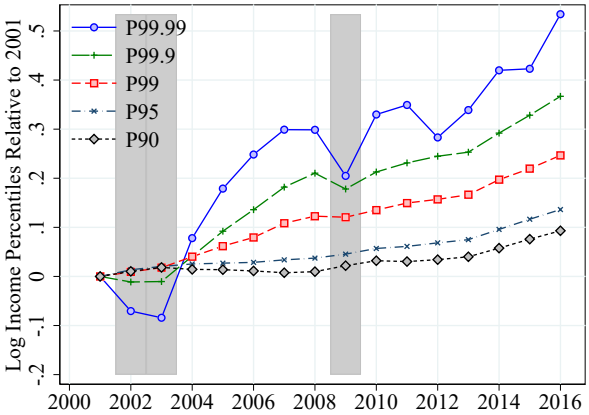
(b) Overall Distribution: Women



(c) Top Percentiles: Men



(d) Top Percentiles: Women



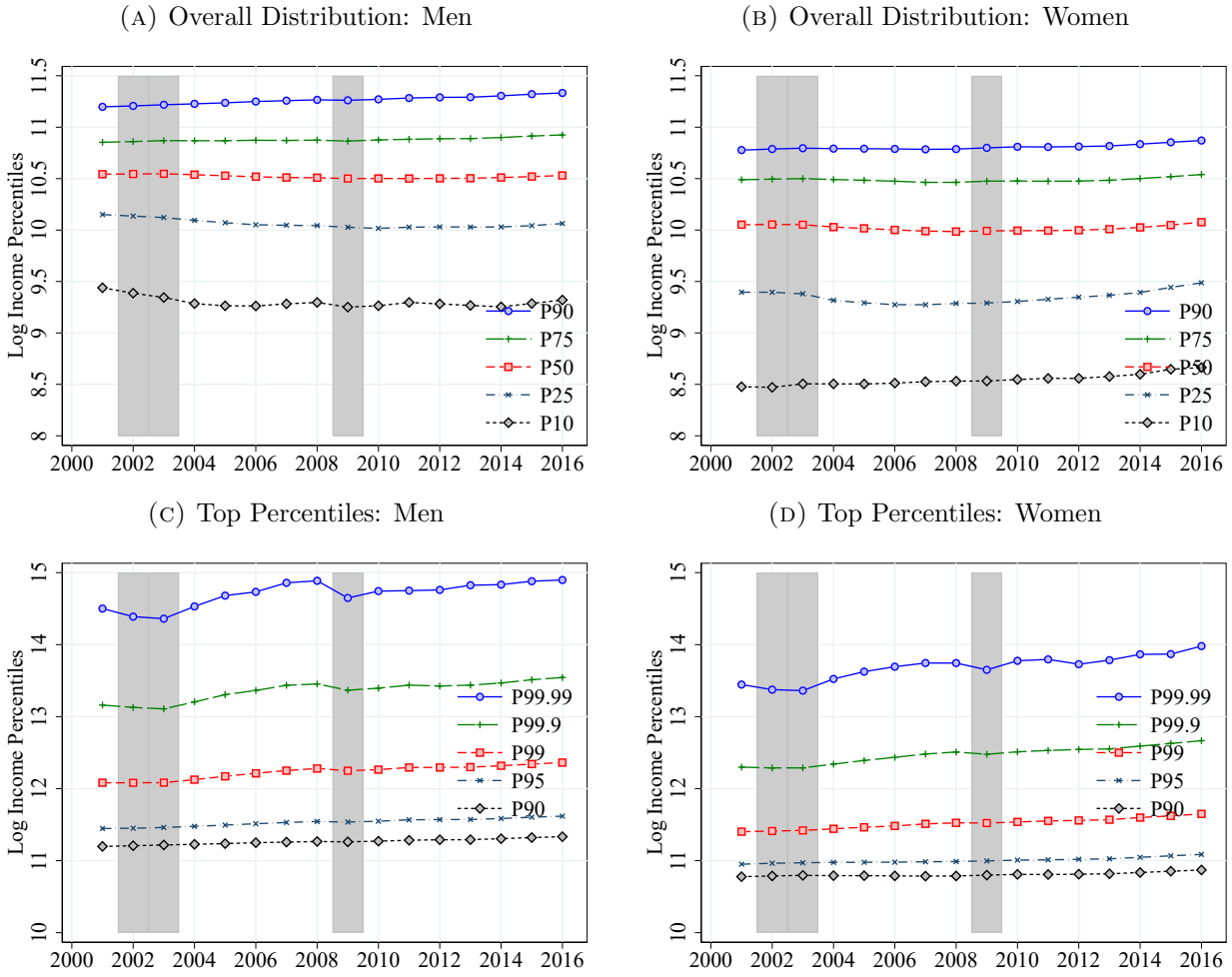
Notes: This figure shows the evolution of selected percentiles of log real annual total income (relative to 2001) in the combined IAB-TPP data (CS analysis sample) separately for men and women. Shaded areas indicate recessions. See Figure 3 for the same analysis of only labor earnings (albeit for a slightly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

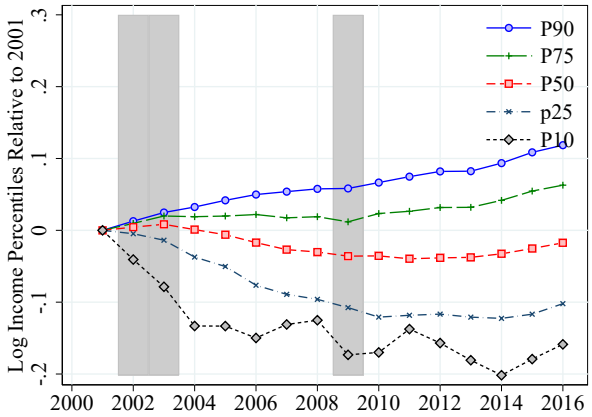
Figure G.4: Percentiles of Log Income



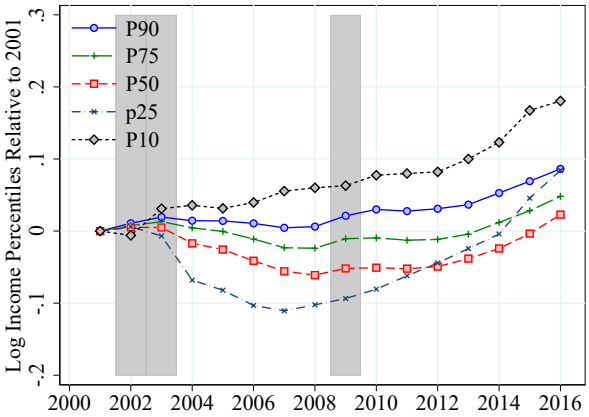
Notes: This figure shows the evolution of absolute log real annual total income percentiles in the combined IAB-TPP data (CS sample) separately for men and women. Shaded areas indicate recessions.

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

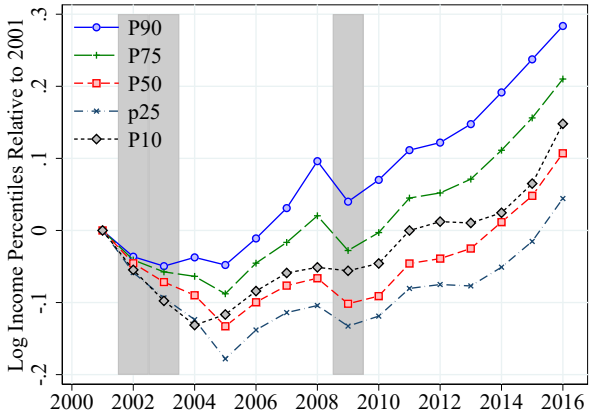
(a) Workers: Men



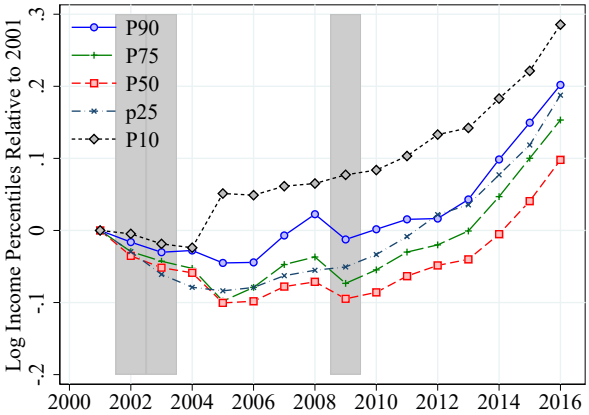
(b) Workers: Women



(c) Entrepreneurs: Men



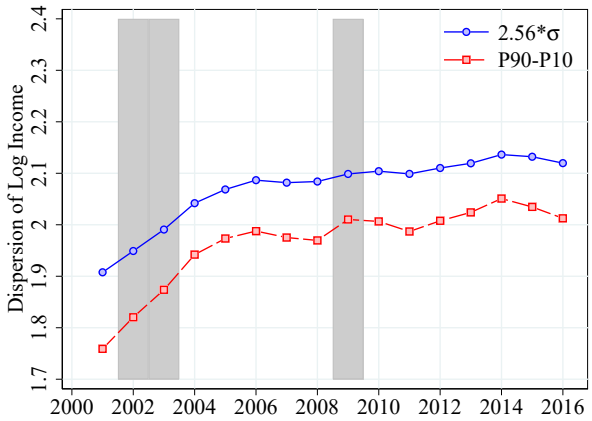
(d) Entrepreneurs: Women



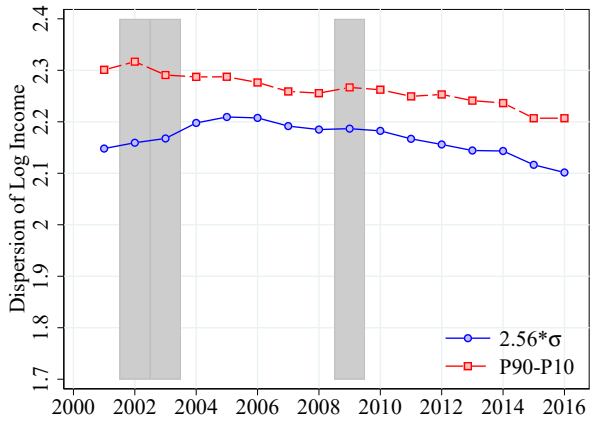
Notes: This figure shows the evolution of different percentiles of log total income among workers and entrepreneurs in the combined IAB-TPP data (CS sample). Workers receive at least half of their income from labor earnings.The jump in the P10 for entrepreneurs (while it is obvious for women, it is hidden 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

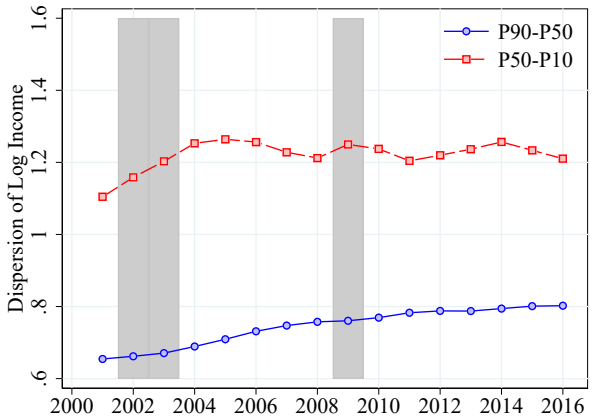
(a) Overall Inequality: Men



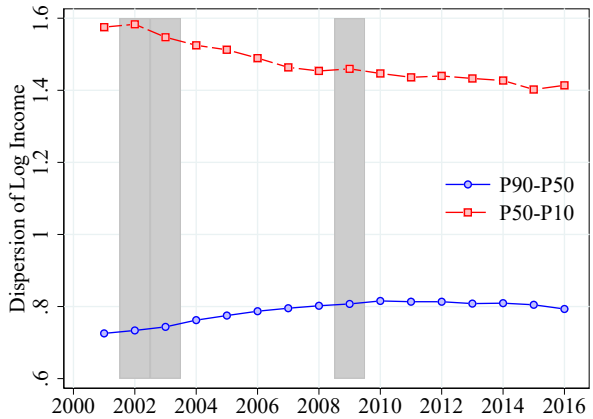
(b) Overall Inequality: Women



(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.

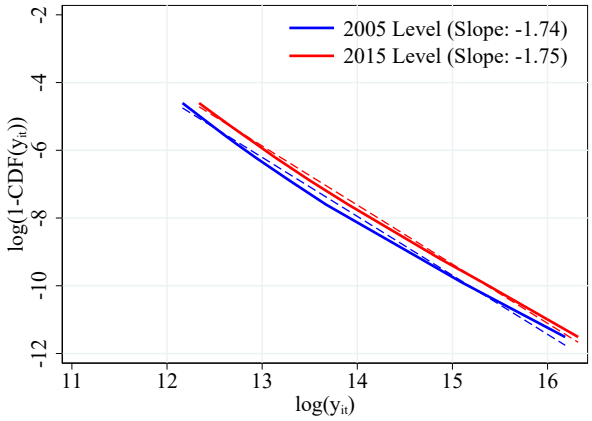
Table G.4: Income Shares – Population

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

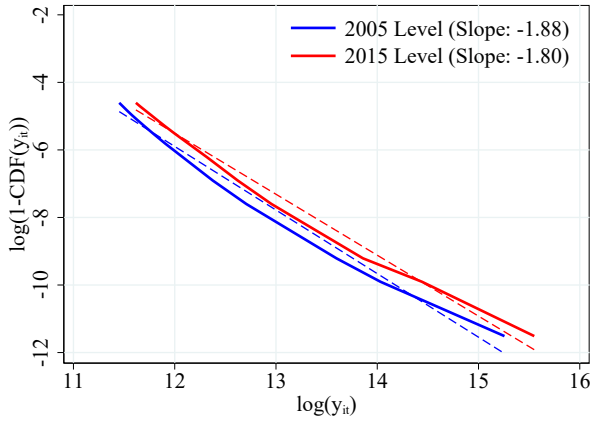
Notes: This table shows the share of (total) income that goes to selected parts of the income distribution in the combined IAB-TPP data (CS sample). Q1 to Q5 refer to the five quintiles where Q1 (Q5) stands for the bottom (top) 20% of the income distribution. The quintile shares sum to one. Bot 50, Bot 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%

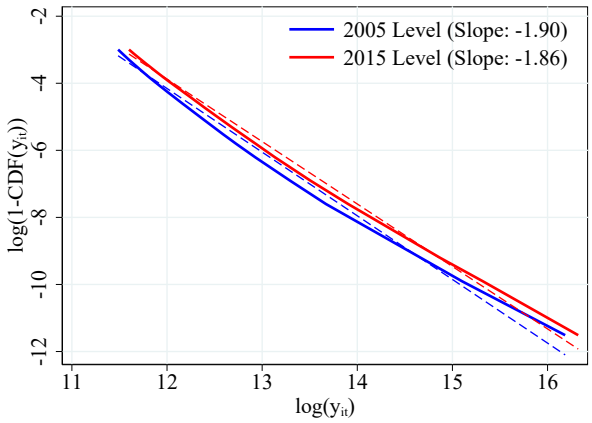
(a) Top 1%: Men



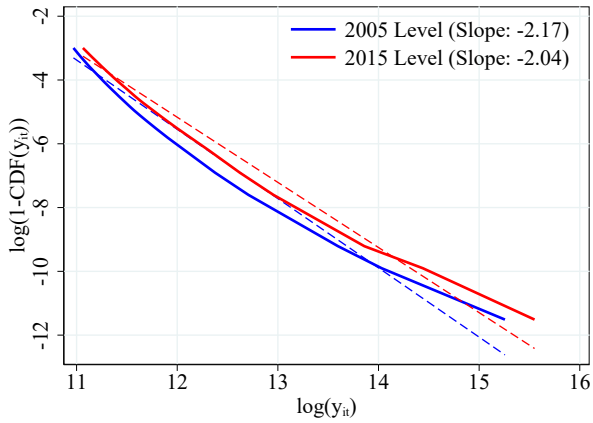
(b) Top 1%: Women



(c) Top 5%: Men



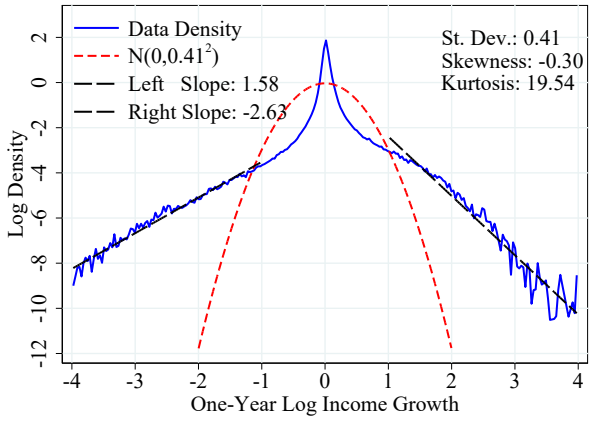
(d) Top 5%: Women



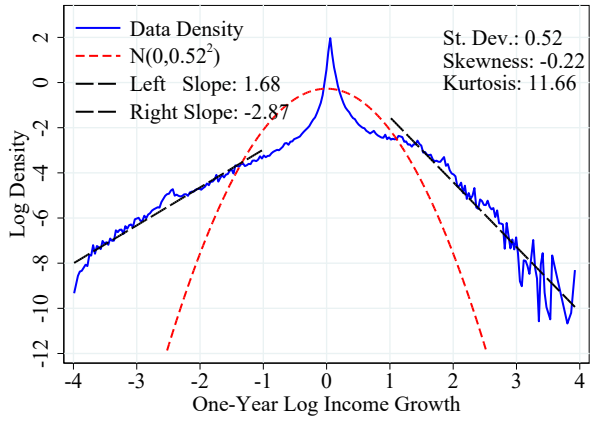
Notes: This figure shows the log of the inverse empirical CDF of log total income and a fitted linear regression line for observations with income in the top 1% and top 5% in the combined IAB-TPP data (CS sample). The absolute value of the slope of the regression line is the Pareto parameter above the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

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

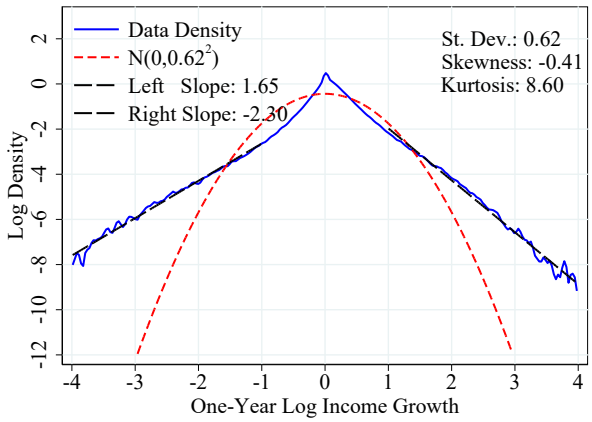
(a) Workers: Men



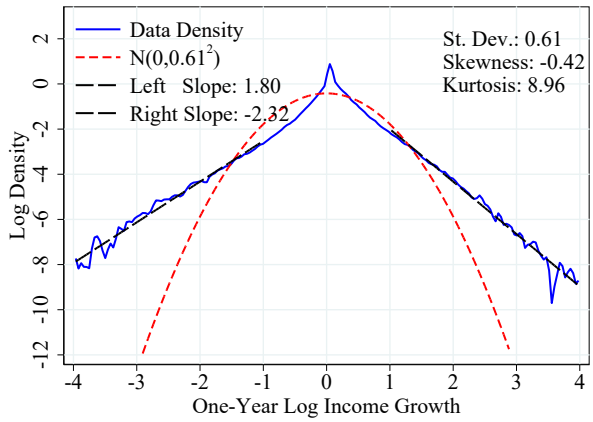
(b) Workers: Women



(c) Entrepreneurs: Men



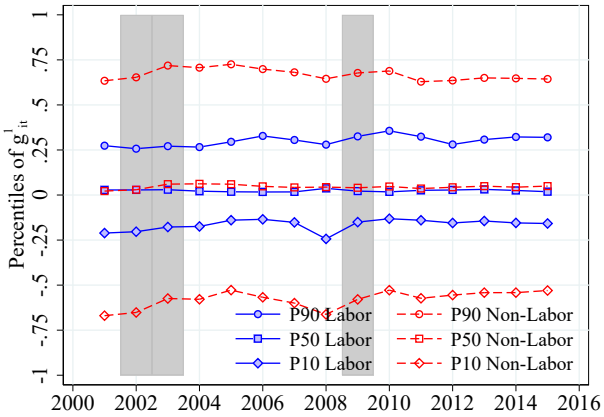
(d) Entrepreneurs: Women



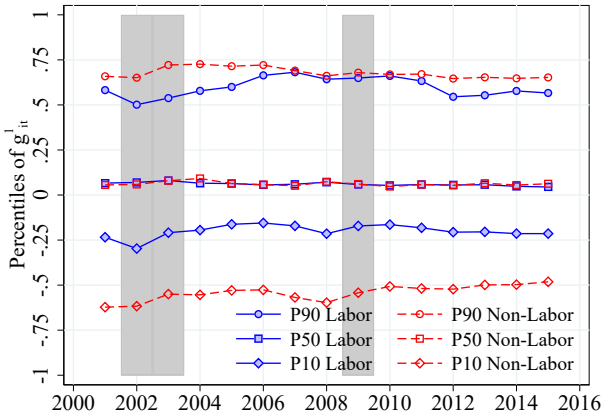
Notes: This figure shows the log density of 1-year changes of residualized log total income separately for workers (labor income as main income source) and entrepreneurs (non-labor income as main income source) and for men and women in the year 2005. LS sample of the combined IAB-TPP data. The dashed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

(a) Men



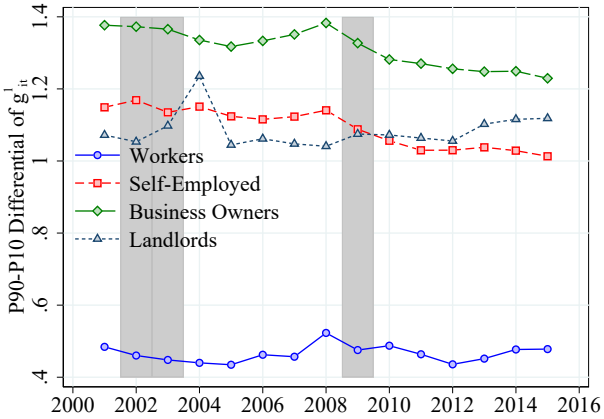
(b) Women



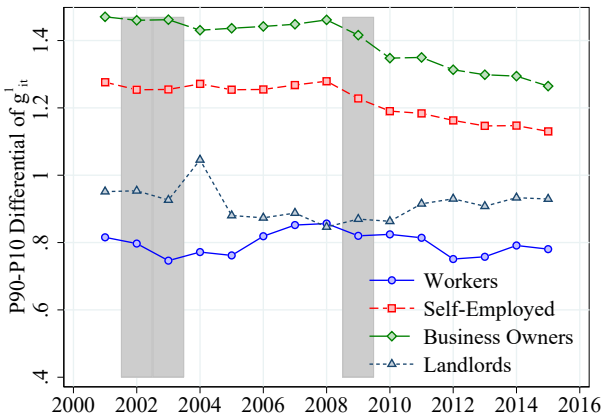
Notes: This figure the 90th, 50th and 10th percentiles of the distribution of 1-year changes in residualized log income (from $t-1$ to t) by main income source (workers vs. entrepreneurs) using the combined IAB-TPP data (LS sample).

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

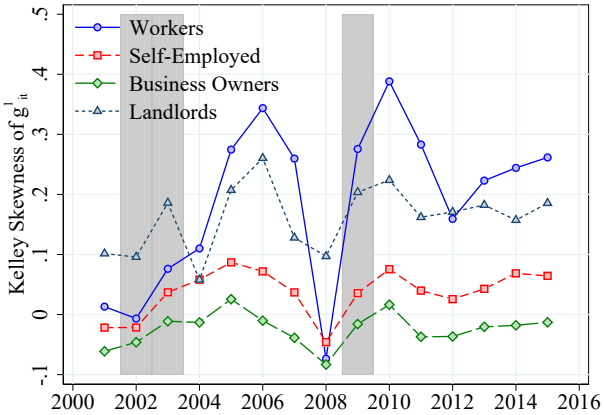
(a) P90–P10: Men



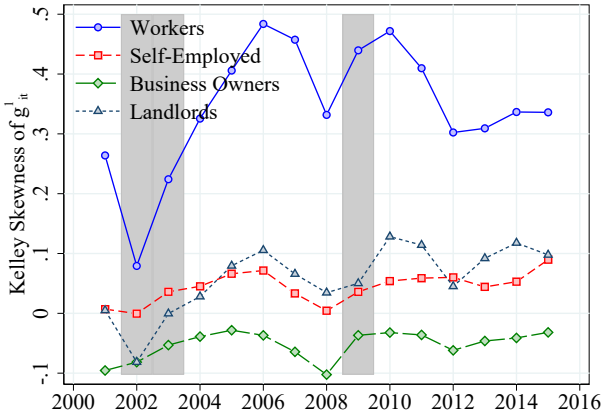
(b) P90–P10: Women



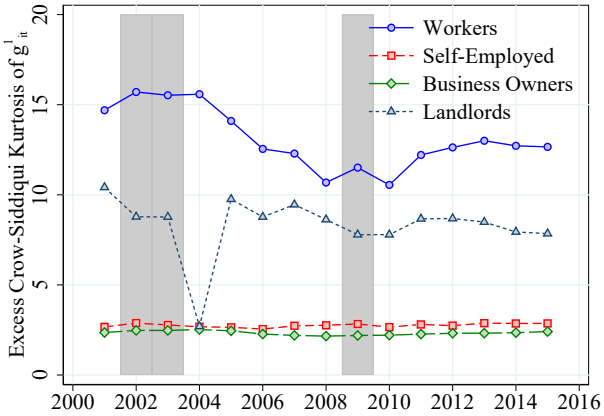
(c) Kelley Skewness: Men



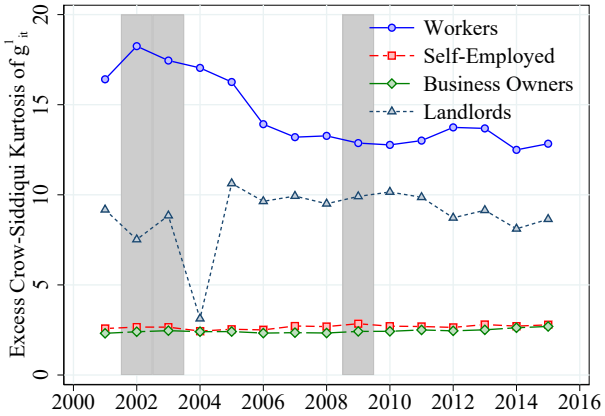
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



(f) Excess Crow-Siddiqui Kurtosis: Women

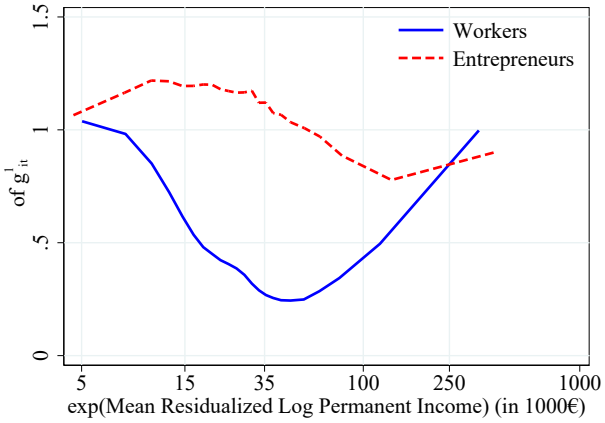


Notes: This figure shows the evolution of the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real annual total income (from $t-1$ to t) in the combined IAB-TPP data (LS sample) separately for men and women by main income source (workers, 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

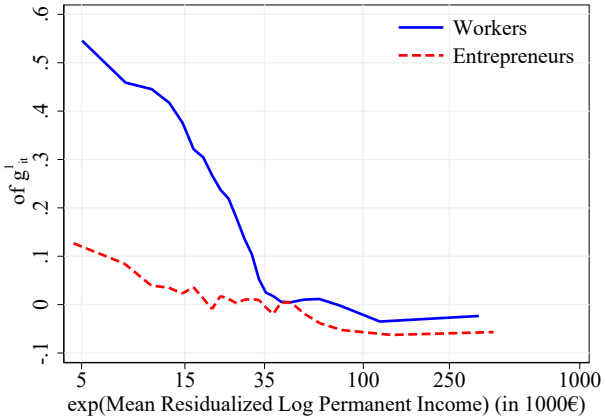
(a) P90-P10: Men

(b) P90-P10: Women

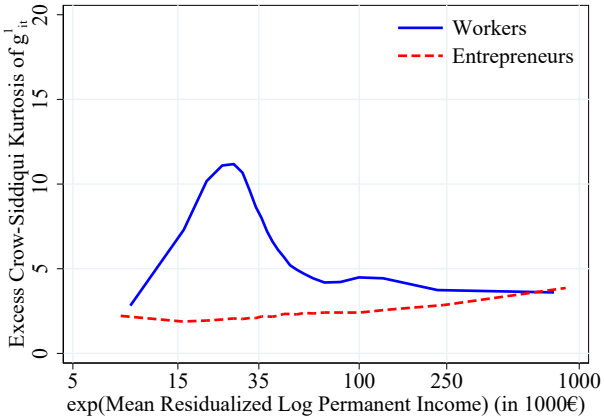


(c) Kelley Skewness: Men

(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men

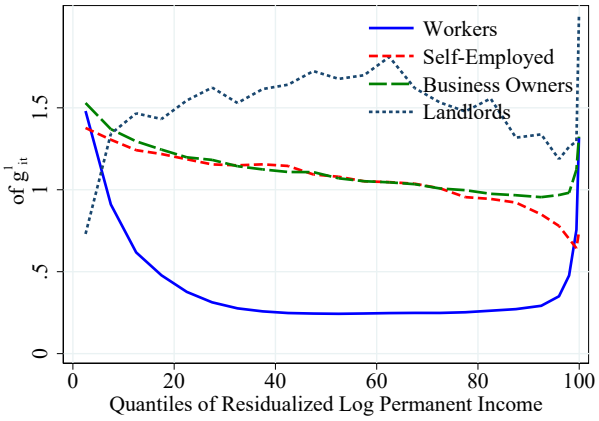


(f) Excess Crow-Siddiqui Kurtosis: Women

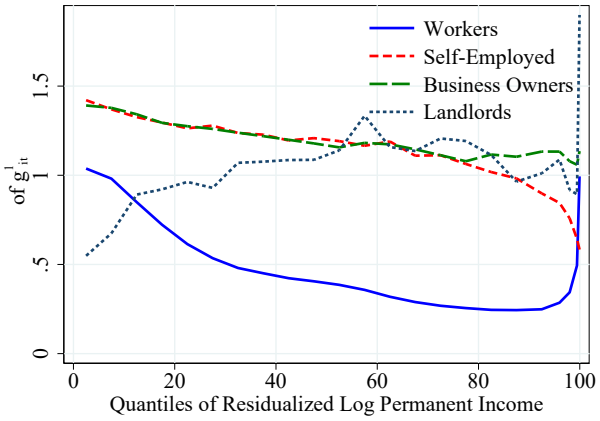
Notes: This figure shows the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real total income by permanent total income (from $t-1$ to t) in the combined IAB-TPP data (H Sample) as averages from 2004 to 2011 and separately for men and women 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

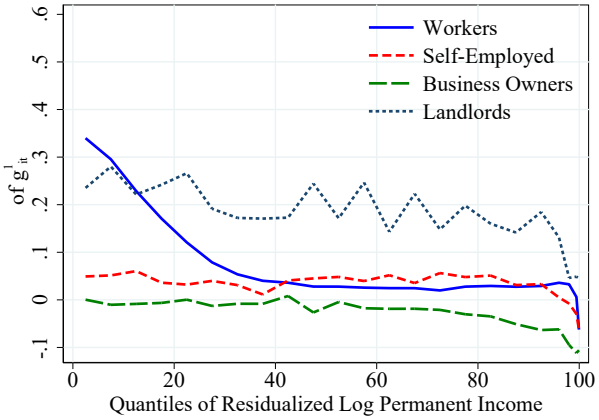
(a) P90-P10: Men



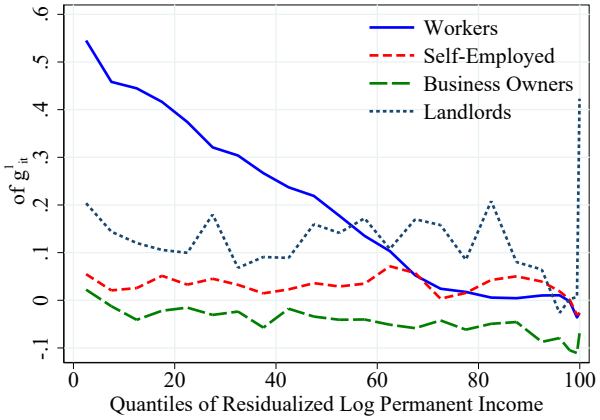
(b) P90-P10: Women



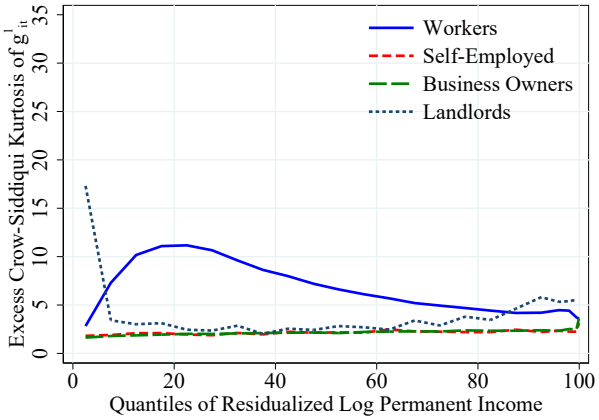
(c) Kelley Skewness: Men



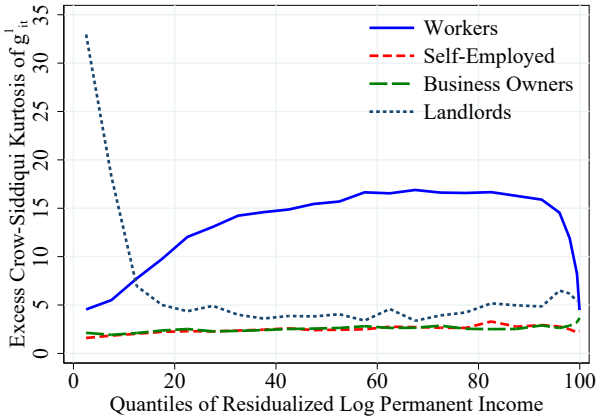
(d) Kelley Skewness: Women



(e) Excess Crow-Siddiqui Kurtosis: Men



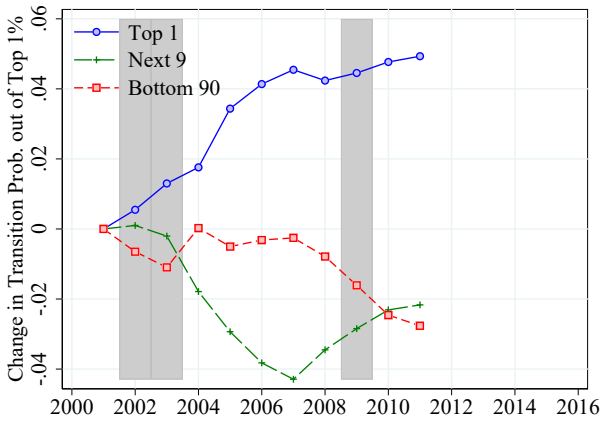
(f) Excess Crow-Siddiqui Kurtosis: Women



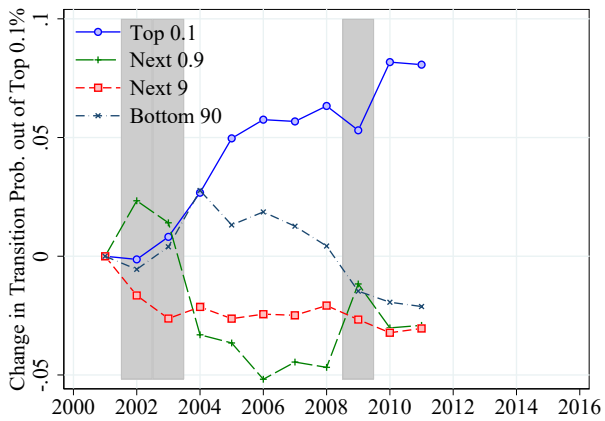
Notes: This figure shows the P90-P10 differential, Kelley skewness and excess Crow-Siddiqui kurtosis of 1-year changes in residualized log real total income by quantiles of the distribution of permanent total income (from $t-1$ to t) in the combined IAB-TPP data (H Sample) as averages from 2004 to 2011 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

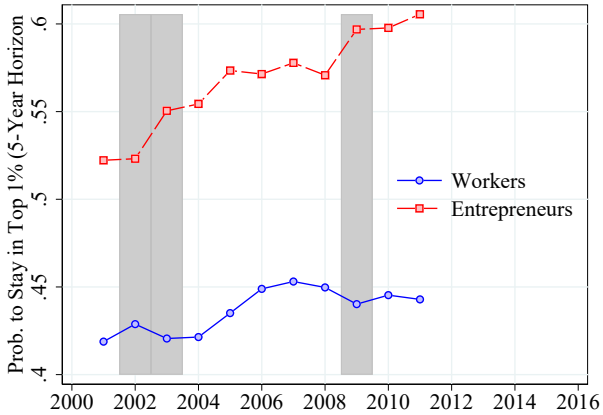
(a) Transitions out of the Top 1%



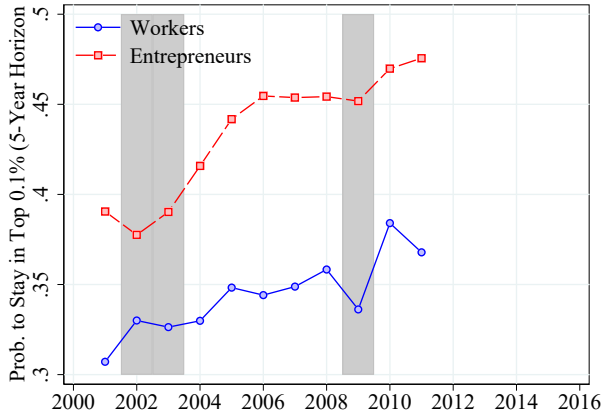
(b) Transitions out of the Top 0.1%



(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.