Unequal Fault Lines: Essays on Gender, Class and Regional Inequality

Inauguraldissertation

zur Erlangung des akademischen Grades einer Doktorin der Wirtschaftswissenschaft des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin

vorgelegt von
Theresa Neef, M.Sc.
aus Dresden

15. November 2023

Erstgutachter:	Prof. Dr. Dr. Giacomo Corneo Fachbereich Wirtschaftswissenschaft Freie Universität Berlin
Zweitgutachter:	Prof. Dr. Alexander Kriwoluzky Fachbereich Wirtschaftswissenschaft Freie Universität Berlin
Tag der Disputation:	01. Februar 2024

Zusammenarbeit mit Koautor*innen und Vorveröffentlichungen:

Kapitel 1 Distributional National Accounts (DINA) for Germany, 1992-2016

*Koautor*innen:* Stefan Bach (DIW Berlin) und Charlotte Bartels (DIW Berlin)

Frühere Version veröffentlicht als S. Bach et al. (2022). "Distributional National Accounts: A Macro-Micro Approach to Inequality in Germany." In: *Measuring Distribution and Mobility of Income and Wealth. Vol. 80.* Ed. by R. Chetty et al. University of Chicago Press, pp. 625–640.

Diese Version wurde veröffentlicht als S. Bach et al. (2023). *Distributional National Accounts for Germany, 1992-2016.* SSRN Working Paper. http://dx.doi.org/10.2139/ssrn.4487665.

Kapitel 2 When Capitalism Takes over Socialism: The Lasting Economic Divide between East and West Germany

*Koautor*innen:* Charlotte Bartels (DIW Berlin)

Keine Vorveröffentlichung.

Kapitel 3 The Long Way to Gender Equality: Gender Pay Differences in Germany, 1871-2021

Diese Version wurde veröffentlicht als T. Neef (2024). *The Long Way to Gender Equality: Gender Pay Differences in Germany, 1871-2021.* World Inequality Lab Working Paper No. 2024/02. Link.

Kapitel 4 Revenue Effects of the Global Minimum Tax under Pillar Two

*Koautor*innen:* Mona Baraké (EU Tax Observatory), Paul-Emmanuel Chouc (EU Tax Observatory) und Gabriel Zucman (UC Berkeley & EU Tax Observatory)

Frühere Versionen veröffentlicht als

> M. Baraké et al. (2021a). *Collecting the Tax Deficit of Multinational Companies: Simulations for the European Union.* EU Tax Observatory Report No. 1.
>
> M. Baraké et al. (2021b). *Minimizing the Minimum Tax? The Critical Effect of Substance Carve-Outs.* EU Tax Observatory Note No. 1.
>
> M. Baraké et al. (2021c). *Revenue Effects of the Global Minimum Tax: Country-by-Country Estimates.* EU Tax Observatory Note No. 2.

Diese Version wurde veröffentlicht als M. Baraké et al. (2022). "Revenue Effects of the Global Minimum Tax Under Pillar Two." In: *Intertax* 50.10, pp. 689–710. doi: https://doi.org/10.54648/taxi2022074.

Theresa Neef
Unequal Fault Lines: Essays on Gender, Class and Regional Inequality
Inauguraldissertation
Reviewers: Prof. Dr. Dr. Giacomo Corneo and Prof. Dr. Alexander Kriwoluzky
Freie Universität Berlin

School of Business & Economics
Department of Economics
Garystr. 21
14195 Berlin

Summary

Chapter 1 This paper presents Distributional National Accounts (DINA) for Germany from 1992 to 2016. Following the DINA method established by Piketty et al. (2018), my co-authors and I combine personal income tax microdata with SOEP survey data to build a representative microdataset for the German adult population. By aligning micro level incomes with their national accounts equivalents, we can estimate the distribution of national income from the bottom to the very top. We find that economic growth in Germany was pro-rich from 1992 to 2007 and pro-poor from 2007 to 2016. But despite rising labor incomes for the bottom 50% since 2007, the income gap between the bottom 50% and the top 10% has widened. This polarization is driven by continuously increasing labor incomes of the lower 9% of the top decile (P90-99). While Germany exhibits a significantly lower top 10% income share than the United States, Germany's top 0.1% income earners receive about 6% of national income, which is comparable to the United States and twice as high as in France. Germany's economic elite primarily holds firms as partnerships owned by two to four shareholders. In contrast, top income earners in France and the United States own shares in corporations with many shareholders. Due to the country-specific structure of German businesses, highly concentrated business profits can translate more directly into personal business incomes, shaping the high concentration at the very top.

Chapter 2 In this paper, we investigate the evolution of economic differences between East and West German residents along the regional income and wealth distributions from 1992 to 2016, employing the Distributional National Accounts (DINA) method that aligns microdata with internationally standardized national accounts. We find that East German residents still earn and own a fraction of their West German counterparts. This gap widens towards the top of the distribution, predominantly due to lower partnership and corporate incomes among East Germans. Exploring the causes of these persistent gaps, we find that East German residents have caught up in terms of the ownership rate and are now equally likely to be business owners as their West German peers, but differences in the level of business capital remain. Using the synthetic control method established by Abadie and Gardeazabal (2003), we show that West German top income earners benefited disproportionately from reunification. We link this finding to the privatization process of the 1990s, in which predominantly West German investors acquired formerly state-owned East German capital. We then examine differences in the characteristics of businesses owned by either East or West Germans, employing the reweighting approach by DiNardo et al. (1996). We conclude that persistent differences in the characteristics of businesses owned by East and West Germans explain an increasing share of sales and productivity gaps. This may impede the convergence of personal business incomes between East and West Germans.

Chapter 3 This paper provides the first long-run time series of the gender earnings ratio for the full-time employed workforce from 1871 to 2021, discusses possible drivers of the observed dynamics and compares the German path with the U.S. and Swedish cases. The industrialization period between 1871 and 1913 yielded slow advances in the gender earnings ratio. In contrast, the first half of the 20th century exhibited giant leaps towards earnings equality. In Germany,

the gender earnings ratio increased from about 46% in 1913 to 58% in 1937. Similar increases were observable in Sweden and the United States. In all three countries, the interplay between increased women's education and increased returns to education due to the expanding white-collar sector fueled increases in the gender earnings ratio. However, Germany's focus on on-the-job apprenticeships may have raised entry barriers for young women into expanding secondary schooling and slowed educational convergence compared to the United States. Germany's increase in the earnings ratio is better explained by women's migration from low-paid agricultural work to higher-paid white-collar jobs. The postwar period brought slower growth in the gender earnings ratio, but a growing participation rate due to the reentry of married women into the labor market. The trajectories of the United States and the European countries diverged significantly. While Sweden and Germany increased their earnings ratios between the late 1950s and 1980, the United States experienced stagnation from 1950 to 1980 and a rapid catch-up in the 1980s.

Chapter 4 In October 2021, 137 countries and jurisdictions agreed to implement a major reform of the international corporate tax system, a global minimum tax of 15% on the profits of large multinational enterprises (MNEs). My co-authors and I simulate the revenue effects of this tax, considering two possible scenarios regarding who collects the minimum tax: The country in which the headquarters are located based on the Income Inclusion Rule (IIR) or the host country of foreign affiliates as laid out under the Qualified Domestic Minimum Top-up Tax (QDMTT). We use the OECD's tabulated country-by-country report (CbCR) statistics, detailing the profits MNEs book and the taxes they pay in the headquarters and subsidiary jurisdictions. We complement these data with information on profits booked by MNEs in tax havens from Tørsløv et al. (2020). Our analysis highlights how the distribution of revenues depends on which country has the priority to collect the top-up tax. Under the IIR, headquarters countries could collect EUR 179 billion in annual revenues worldwide, of which EU Member States would receive EUR 67 billion. However, carve-outs, i.e., provisions that decrease the tax base for real economic activity, reduce potential revenues. The EU Member States' revenue would decrease to EUR 55 billion per year. Under the QDMTT, low-tax jurisdictions that have attracted affiliates of MNEs could be among the main beneficiaries of the reform.

Kurzfassung

Kapitel 1 In diesem Artikel schätzen wir *Distributional National Accounts* (DINA) für Deutschland von 1992 bis 2016. In Anlehnung an die von Piketty et al. (2018) entwickelte DINA-Methode verbinden meine Koautor*innen und ich Mikrodaten der Lohn- und Einkommensteuerstatistik mit Befragungsdaten des Sozio-oekonomischen Panels (SOEP), um einen repräsentativen Mikrodatensatz für die deutsche Bevölkerung ab 20 Jahren zu erstellen. Durch die Harmonisierung der mikrodatenbasierten Einkommen mit ihren Äquivalenten in den Volkswirtschaftlichen Gesamtrechnungen (VGR), können wir die Verteilung des Nationaleinkommens schätzen. Unsere Ergebnisse zeigen, dass von 1992 bis 2007 vor allem die Einkommensreichen vom Wirtschaftswachstum profitierten, während es von 2007 bis 2016 zugunsten der Einkommensschwächeren ausfiel. Aber trotz steigender Arbeitseinkommen für die einkommensschwächsten 50% seit 2007, hat sich die Einkommensschere zwischen den unteren 50% und den obersten 10% vergrößert. Die kontinuierlich steigenden Arbeitseinkommen der unteren 9% des obersten Dezils (P90-99) haben maßgeblich zu dieser Polarisierung beigetragen. Der Einkommensanteil der obersten 10% in Deutschland ist deutlich geringer als in den USA. Jedoch beziehen die obersten 0,1% der Einkommensbezieher*innen in Deutschland rund 6% des Nationaleinkommens. Dieser Anteil ist vergleichbar mit den USA und doppelt so hoch wie in Frankreich. Deutschlands Wirtschaftselite hält ihre Unternehmen überwiegend in Form von Personengesellschaften im Besitz von zwei bis vier Gesellschafter*innen. Im Gegensatz dazu halten die Spitzenverdiener*innen in Frankreich und den USA Anteile an Kapitalgesellschaften mit vielen Gesellschafter*innen. Aufgrund der landesspezifischen Unternehmensstruktur in Deutschland können sich konzentrierte Unternehmensgewinne direkter in persönliche Unternehmenseinkommen umsetzen. Dies prägt die starke Konzentration an der Spitze der Einkommensverteilung.

Kapitel 2 In diesem Beitrag untersuchen wir die Entwicklung der wirtschaftlichen Unterschiede zwischen Ost- und Westdeutschen entlang der regionalen Einkommens- und Vermögensverteilungen von 1992 bis 2016. Dabei verwenden wir die *Distributional-National-Accounts*-Methode (DINA), die Mikrodaten mit international standardisierten Volkswirtschaftlichen Gesamtrechnungen (VGR) verbindet. Wir zeigen, dass Ostdeutsche im Vergleich zu Westdeutschen immer noch einen Bruchteil des Einkommens verdienen und ein wesentlich geringeres Vermögen besitzen. Diese Differenz vergrößert sich am oberen Ende der Einkommens- und Vermögensverteilungen, vor allem aufgrund niedrigerer Unternehmenseinkommen aus Personengesellschaften und Körperschaften von Ostdeutschen. Wenn wir den Ursachen dieser persistenten Unterschiede nachgehen, finden wir, dass Ostdeutsche genauso häufig Unternehmer*innen sind wie Westdeutsche, aber die Unterschiede bei der Höhe des Unternehmenskapitals bestehen bleiben. Mit Hilfe der von Abadie und Gardeazabal (2003) entwickelten *Synthetic-Control*-Methode zeigen wir, dass westdeutsche Spitzenverdiener*innen überproportional von der Wiedervereinigung profitiert haben. Wir bringen dieses Ergebnis mit dem Privatisierungsprozess der 1990er Jahre in Verbindung, bei dem überwiegend westdeutsche Investor*innen ehemals staatliches ostdeutsches Kapital erwarben. Anschließend untersuchen wir mit Hilfe der *Reweighting*-Methode von DiNardo et al. (1996) Unterschiede in den Charakteristika von Betrieben in ost- und westdeutschem Besitz. Wir

kommen zu dem Schluss, dass anhaltende Unterschiede in Betriebsmerkmalen, wie Betriebsgröße und Rechtsform, einen zunehmenden Anteil der Umsatz- und Produktivitätsunterschiede zwischen ost- und westdeutschgehaltenen Betrieben erklären. Dies könnte die Konvergenz der persönlichen Unternehmenseinkommen verlangsamen.

Kapitel 3 Dieses Papier präsentiert die erste Zeitreihe des genderspezifischen Verdienstverhältnisses für Vollzeitbeschäftigte von 1871 bis 2021, diskutiert mögliche Einflussfaktoren auf deren Verlauf und vergleicht den deutschen Pfad mit dem der USA und Schwedens. Die Industrialisierungsperiode zwischen 1871 und 1913 brachte nur langsame Fortschritte im Verdienstverhältnis. In der ersten Hälfte des 20. Jahrhunderts kam es dagegen zu sprunghaften Fortschritten in Richtung Verdienstgleichheit. In Deutschland stieg das genderspezifische Verdienstverhältnis von ca. 46% im Jahr 1913 auf 58% im Jahr 1937. Ähnliche Steigerungen sind in Schweden und den USA zu beobachten. In allen drei Ländern hat das Zusammenspiel zwischen der zunehmenden Bildung von Frauen und der höheren Bildungsrendite aufgrund des expandierenden Angestelltensektors den Anstieg des genderbezogenen Verdienstverhältnisses befördert. Das in Deutschland lange Zeit dominierende Ausbildungssystem in Form einer Lehre könnte jedoch die Bildungskonvergenz zwischen Männern und Frauen im Vergleich zu den USA verlangsamt haben, da es für Frauen höhere Eintrittsbarrieren in die Sekundarbildung mit sich brachte. Der Anstieg des genderspezifischen Verdienstverhältnisses in Deutschland lässt sich besser durch die Abwanderung von Frauen aus gering bezahlter landwirtschaftlicher Arbeit in besser bezahlte Angestelltenberufe erklären. Die Nachkriegszeit war geprägt durch ein langsameres Wachstum des genderspezifischen Verdienstverhältnisses bei einer steigenden Frauenerwerbsquote aufgrund des Wiedereintritts verheirateter Frauen in den Arbeitsmarkt. Die Entwicklungen in den USA und Europa verliefen dabei unterschiedlich. Während Schweden und Deutschland ihre genderspezifischen Verdienstquoten zwischen den späten 1950er Jahren und 1980 steigerten, stagnierte das Verdienstverhältnis in den USA von 1950 bis 1980 und holte in den 1980ern Jahren rasch auf.

Kapitel 4 Im Oktober 2021 einigten sich 137 Staaten und Gebiete auf die Umsetzung einer Reform des internationalen Körperschaftsteuersystems: eine globale effektive Mindeststeuer von 15% auf die Gewinne großer multinationaler Unternehmen. Meine Koautor*innen und ich simulieren die Aufkommenseffekte dieser Steuer, wobei wir zwei mögliche Szenarien in Betracht ziehen, wer die Steuer erhebt: Das Land, in dem sich der Firmenhauptsitz befindet, auf Grundlage der *Income Inclusion Rule* (IIR), oder das Land, in dem sich ausländische Tochtergesellschaften ansiedeln, unter Anwendung der *Qualified Domestic Minimum Top-up Tax* (QDMTT). Wir verwenden die tabellarischen Country-by-Country-Berichte (CbCR) der OECD, in denen die von multinationalen Unternehmen deklarierten Gewinne und ihre tatsächlich gezahlten Steuern in den Ländern des Hauptsitzes und der Tochtergesellschaften aufgeführt sind. Wir ergänzen diese Daten mit Informationen über die von multinationalen Unternehmen in Steueroasen verbuchten Gewinne aus Tørsløv et al. (2020). Unsere Analyse zeigt, wie die Verteilung der Einnahmen davon abhängt, welches Land bei der Steuererhebung Priorität hat. Im Rahmen der IIR könnten die Hauptsitzländer weltweit jährliche Einnahmen in Höhe von 179 Mrd. Euro erzielen, wovon 76 Mrd. Euro auf die EU-Mitgliedsstaaten entfielen. Sogenannte *Carve-outs*, Ausnahmeregelungen auf Basis realer Wirtschaftstätigkeit, reduzieren jedoch die Bemessungsgrundlage und damit die potenziellen Einnahmen. Die Einnahmen der EU-Mitgliedstaaten würden auf 55 Mrd. Euro pro Jahr sinken. Im Rahmen der QDMTT könnten Niedrigsteuerländer, die Tochtergesellschaften von multinationalen Unternehmen angezogen haben, zu den Hauptbegünstigten der Reform gehören.

Declaration

Ich erkläre hiermit gemäß § 10 Abs. (3) der Promotionsordnung zum Dr. rer. pol. des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin vom 13. Februar 2013, dass ich mich noch keinem Promotionsverfahren unterzogen habe oder um Zulassung zu einem solchen beworben habe. Darüber hinaus habe ich die Dissertation mit dem Titel "Unequal Fault Lines: Essays on Gender, Class and Regional Inequality" in der vorliegenden oder überarbeiteten Form keiner Prüfungskommission, keinem Fachbereichsvertreter und keiner anderen Fakultät einer anderen Universität zur Begutachtung vorgelegt. Ich habe bei der Erstellung dieser Dissertation verschiedene Hilfsmittel und Ressourcen verwendet, die ich im Folgenden aufführe:

1. Stata 15-17
2. TEXLive 2023 in Overleaf und TEXStudio
3. Microsoft Office 2016
4. Spotify Premium

Berlin (Germany), 15 November 2023

Theresa Neef

Contents

Acknowledgements

I am immensely grateful for numerous persons who shared part of the way with me and supported me in both my academic and personal growth in the past years.

I would like to thank Giacomo Corneo, my first supervisor, for his support and feedback. Further, I am thankful for Alexander Kriwoluzky's support on the last stretch. I am grateful for Miriam Beblo's feedback and for inviting me to her PhD seminars. This thesis would not have been possible without Charlotte Bartels' steady support. Charlotte offered me the first opportunity to pursue my interest in inequality topics as her research assistant in 2016 and encouraged me to apply for the PhD scholarship in 2017. She was a constant throughout these PhD years as an engaged co-author and mentor. I am immensely thankful for the knowledge, motivating words and advice she shared with me. I want to express my gratitude to Thomas Piketty for giving me the opportunity to become part of the World Inequality Lab and Paris School of Economics communities and for his support and feedback. I am equally indebted to Gabriel Zucman who gave me the chance to widen my research horizon to international taxation by letting me become part of the EU Tax Observatory.

I would like to thank my co-authors, Stefan Bach, Mona Baraké, Charlotte Bartels, Paul-Emmanuel Chouc, Anne-Sophie Robilliard and Gabriel Zucman for fruitful cooperations and sharing their skills and insights with me. In particular, I have learned so much about the German tax-transfer system and the personal income tax data in long sessions with my co-author, Stefan Bach. I am grateful for his patience to share his knowledge with me. Two papers would not have been possible without the restricted access to the German personal income tax data. I would like to thank the teams of the Federal Statistical Office and the Research Data Centre of the Statistical Offices of the Federal States (FDZ) in Berlin. I gratefully acknowledge generous financial support from the Hans Böckler Foundation whose PhD scholarship allowed me to embark on this journey.

My work and living contexts have changed twice during the past six years and I am happy to have always found kind and supportive communities along the way. Thank you, Max, for sharing the office in the Boltzmann basement with me for the first year. I consider myself immensely fortunate to have become part of the amazing community of passionate inequality researchers of the World Inequality Lab. I have learned so much about inequality research and beyond from all of you. Thank you for all your knowledgeable, motivating and kind words and for going through the Covid years together, dear Rowaida, Oscar, Thomas, Tom, Álvaro, Morten, Anne-Sophie, Clara, Nacho, Nitin, Alice, Olivia, Diane, Amory, Marc and Yajna. It was exciting to embark on a new topic with a completely new institution, the EU Tax Observatory. I am thankful for the atmosphere of research passion and support I found there. Thank you, Bluebery, Gabriel, Giulia, Lauren, Leo, Manon, Mona, Panos, Paul-Emmanuel, Sarah and Sofía. Lastly, I am happy to have returned to Berlin, my favorite city, and to have found an encouraging team at the macroeconomics department of DIW. Thank you, Timm, Pia and Jan for welcoming me.

Last but not least, I would like to thank my parents and my friends who kept me content despite many long work days: Thank you, Judith, Lulu and Stef! I am grateful for our long friendship. Jule, thank you for your constant positive energy and support! Rowaida, Oscar, Vivi, Lennart, Thiago and Marcela, thank you for making Paris feel like home! Thank you, Lukas and Carina Lota, for finding the right words and sometimes the right questions. Iris, Luisa, Nikita, Paula, Pia, Seb and Magda, thank you for welcoming me back in Berlin and sharing my passion for the cultural scene of this city.

Introduction

Rising economic inequality in numerous countries across the globe has put the topic of inequality back at the center of political and societal debates. The distribution of wealth and income touches upon a crucial question: In which society do we want to live? At the same time the inequality debate is not only important for equity considerations but also under the aspect of efficiency. A central narrative of 20th century economics postulated the fundamental trade-off between equity and efficiency (Okun, 1975). It was supported by prominent studies highlighting the positive effects of inequality by increasing work incentives (Lazear and Rosen, 1981; Mirrlees, 1971) and aggregate savings and investment resulting in higher incomes for all society members (Bourguignon, 1981). However, in the past decades, more and more evidence challenges this trade-off by showing that inequality might also yield a negative externality at different levels of the economy. At the micro level, highly unequal income and wealth distributions might inhibit investments in human capital and limit individuals' occupational choice due to credit constraints (Banerjee and Newman, 1993; Galor and Zeira, 1993). At the firm level, inequality and perceived unfairness in pay might decrease workers' morale and potential productivity (Breza et al., 2017; Card et al., 2012). At the macro level, economic and political systems might be hampered by polarized income and wealth distributions due to political unrest, increased uncertainty discouraging investments and rising household and government debt (Alesina and Perotti, 1996; Bénabou, 1996; Benhabib and Rustichini, 1996; Mian et al., 2020, 2021). All these studies show that economic inequality can slow economic growth and development.

Economic inequality is first and foremost a political and societal decision – as prominent scholars such as Atkinson (2015) and Piketty (2020) emphasize. It can be shaped by well-informed and adequate policies. In this context, learning from historical inequality research about which environments, institutions and policies propel or restrain economic inequality is crucial for constructive political debates and effective policies.

The initial question of every study of inequality is: Inequality of what among whom? The empirical studies assembled in this dissertation investigate pretax income inequality along three different fault lines in Germany: inequality between different groups along the German income distribution, regional inequality between East and West German residents and gender earnings inequality.

The greater availability of country-specific inequality studies has made salient general movements of income inequality observed in many countries, namely the U-shaped development of top income shares over the 20th century. From the analysis of cross-country similarities, we have learned about the global importance of the rise of the welfare state and progressive income and inheritance taxation to keep inequality in check (Piketty, 2020, 2022). Country-specific case studies and cross-country comparisons can also help us investigate unique institutional differences and political settings, informing us about the contexts in which inequalities take shape. The first and third chapters of this dissertation use cross-country comparisons, namely with the well-researched U.S. benchmark, to shed light on how country-specific institutions have shaped the level of income inequalities. The second chapter adds a case study examining the impact of Germany's unique privatization process on long-run inequality dynamics.

In *Chapter 1* "*Distributional National Accounts (DINA) for Germany, 1992-2016*", my co-authors, Stefan Bach and Charlotte Bartels, and I investigate inequality dynamics between different income groups – from the bottom to the very top – of the German income distribution from 1992 to 2016. This paper contributes to the understanding of country-specific structures shaping inequality patterns by closely investigating differences in the income composition between U.S. and German top incomes. Our analysis underlines the importance of firm structures for the level of top income concentration. We build upon recent methodological advances in inequality research using the Distributional National Accounts method established by Piketty et al. (2018). By combining personal income tax microdata with SOEP survey data, we build a representative microdataset for the German adult population. In a last crucial step, we align incomes at the micro level with their national accounts equivalents. Due to the harmonization with internationally standardized national accounts, the DINA method allows to compare inequality estimates across countries and to explore the question of who benefits from national income growth. We find that the gap between the top 10% and the bottom 50% has widened. From 1992 to 2007, income growth was pro-rich. Incomes of the bottom 50% declined in real terms and their income share declined from 22% to 17%. Although growth has been pro-poor since 2007, the bottom 50% did not reach their income share of the early 1990s in 2016. This polarization can be attributed to the continuous increase in labor incomes among the lower 9% of the top decile (P90-99) since 1992. In cross-country comparison, we find that Germany's economic elite receives an extraordinarily high income share. The top 0.1% receives approx. 6% of national income, which is comparable to the United States and twice as high as in France. Germany's economic elite primarily holds firms as partnerships owned by two to four shareholders, while top income earners in France and the United States hold shares in corporations with numerous shareholders. Due to the country-specific structure of German businesses, highly concentrated business profits can translate more directly into personal business incomes. This contributes to the observed strong concentration of personal business incomes at the very top.

Chapter 2 "*When Capitalism Takes over Socialism: The Lasting Economic Divide between East and West Germany*" turns to regional inequality. The "equivalence of living conditions" across the entire federal territory is anchored in the German constitution (Art. 72(2) GG). However, income differences between East and West German residents persist three decades after reunification. My co-author, Charlotte Bartels, and I estimate these differences along the regional income and wealth distributions from 1992 to 2016 and explore possible causes of the lasting divide. Our analysis contributes a rigorous case study on how Germany's unique privatization process has impacted inequality. It emphasizes the importance of the design of privatization policies for long-run inequality dynamics. We find that East German residents still earn and own a fraction of their West German counterparts. This gap widens towards the top of the distribution, predominantly due to lower partnership and corporate incomes among East Germans. We explore the causes of these persisting differences from two perspectives. First, we find that East German residents have caught up in terms of the ownership rate and are now equally likely to be business owners as their West German peers. Therefore, we turn to firm characteristics influencing the level of the owned business capital. Previous studies have shown that the overriding majority of formerly state-owned East German firms was sold to West German investors and companies, often operating in the same or similar industries (Dornbusch et al., 1992; Mergele et al., 2020; Windolf, 1996). Substantial tax reliefs aimed at West German top income earners spurred investment in East German firms, housing and land. Using the synthetic control method established by Abadie and Gardeazabal (2003), we quantify the impact of reunification on West German income

concentration. We find that West German top income earners indeed benefited disproportionately from reunification. The top 1% income share in West Germany soared to reach about 20% in the early 2000s, a development not visible in the counterfactual. Against the historical backdrop, we interpret this abrupt increase as indicative evidence for West German top earners investing in East Germany which increased their medium- and long-run capital incomes. We then investigate differences in characteristics of business establishments *owned* by either East or West Germans. To do so, we employ the reweighting approach developed by DiNardo et al. (1996) using the IAB Establishment Panel. We find that an important fraction of the sales and productivity gaps between East and West German owned establishments arises from differences in establishment characteristics, such as size, legal form and firm structure. While the observed productivity and sales gaps declined between 2000 and 2019, the share of these gaps explained by the above-mentioned characteristics gained more importance. This highlights that rather permanent features of East German owned establishments inhibit the convergence of business incomes between East and West Germans.

Chapter 3 "*The Long Way to Gender Equality: Gender Pay Differences in Germany, 1871-2021*" highlights another persistent fault line in the German society: Gender earnings inequality. I provide the first long-run time series of the gender earnings ratio for the full-time employed workforce from 1871 to 2021, discuss possible drivers of the observed dynamics and compare the German path with the U.S. and Swedish cases. The field of gender economics has grown immensely in the past years, rigorously investigating the current causes of gender disparities in the labor market and proposing effective policies. Claudia Goldin (1990)'s work, awarded with this year's Nobel Prize in Economic Sciences, has shown that the long-run perspective is a crucial basis for understanding current gender disparities. It sheds light on fundamental labor demand and supply shifts and pivotal institutions, such as laws, norms and educational systems, promoting or inhibiting gender equality. By comparing the German path towards gender pay equality since the 19th century with the U.S. time series by Goldin (1990), this chapter underscores the importance of country-specific institutions, particularly the educational system's structure, in promoting or restraining gender pay equality. I find that the industrialization period between 1871 and 1913 yielded slow advances in the gender earnings ratio due to women's delayed inclusion in the industrial workforce in Germany. In contrast, the first half of the 20th century exhibited giant leaps towards earnings equality. In Germany, the gender earnings ratio increased from about 46% in 1913 to 58% in 1937. Similar increases were visible in Sweden and the U.S. In all three countries, the interplay between increased women's education and increased returns to education due to women's rapid migration into the expanding white-collar sector fueled the increases in the gender earnings ratio. However, Germany's focus on on-the-job vocational training may have raised entry barriers for young women into expanding secondary schooling and slowed educational convergence compared to the United States. Germany's nonetheless similar increase in the earnings ratio is better explained by women's migration from low-paid agriculture to higher-paid white-collar jobs, which substantially increased female relative earnings. The postwar period brought slower growth in the gender earnings ratio, but a growing participation rate due to the reentry of married women into the labor market. The trajectories of the United States and the European countries diverged significantly. While Sweden and Germany increased their earnings ratios between the late 1950s and 1980, the U.S. experienced stagnation from 1950 to 1980 and a rapid catch-up in the 1980s. Economic conditions and policy action shaped the country-specific developments: While the U.S. benchmark narrative discusses the strong increase in married women's labor supply as the most remarkable development of the early postwar period,

Germany's postwar recession delayed a similar development. In the 1970s, Sweden took the lead in terms of women's labor supply and gender earnings equality due to a set of policies propelling married women's labor supply, absent in Germany and the U.S.

Chapter 4 "*Revenue Effects of the Global Minimum Tax under Pillar Two*" turns to the question of adequate policy action. Increased possibilities to shift profits to low-tax jurisdictions have raised fairness concerns of national tax systems and decreased national tax revenues. My co-authors, Mona Baraké, Paul-Emmanuel Chouc and Gabriel Zucman, and I explore the revenue potential of a global minimum tax of 15% on corporate profits. This policy was agreed upon among 137 countries and jurisdictions under the umbrella of the OECD in October 2021. We simulate the revenue effects of the global minimum tax, considering two possible scenarios regarding who collects the tax: The country in which the headquarters are located based on the Income Inclusion Rule (IIR) or the host country of foreign affiliates as laid out under the Qualified Domestic Minimum Top-up Tax (QDMTT). We use tabulated country-by-country report (CbCR) statistics, detailing the amount of profits multinational enterprises book and the taxes they pay in the headquarters and subsidiary jurisdictions, provided by the OECD. We complement these data with information on profits booked by multinational companies in tax havens from Tørsløv et al. (2020). Our analysis accentuates the significance of policy design: We find that the distribution of revenues depends on which country has the priority to collect the top-up tax. Under the IIR, headquarters countries could collect EUR 179 billion in annual revenues worldwide, of which EU Member States would receive EUR 67 billion. However, carve-outs, i.e., provisions that decrease the tax base for real economic activity, reduce potential revenues. The EU Member States' revenue would decrease to EUR 55 billion per year. Under the QDMTT, low-tax jurisdictions that have attracted affiliates of multinational companies could be among the main beneficiaries of the reform.

1 Distributional National Accounts (DINA) for Germany, 1992-2016*

**Coauthored with Stefan Bach (DIW Berlin)
and Charlotte Bartels (DIW Berlin)**

Abstract

This paper presents Distributional National Accounts (DINA) for Germany from 1992 to 2016. We combine personal income tax files with SOEP survey data to estimate the distribution of national income following the DINA method established by Piketty et al. (2018). We find that economic growth in Germany has been pro-rich from 1992 to 2007 and pro-poor from 2007 to 2016. But despite rising labor incomes for the bottom 50% since 2007, the income gap between the bottom 50% and the top 10% has widened. The polarization is driven by continuously increasing labor income of the lower 9% of the top decile (P90-99). The top 0.1% receives 6% of national income, which is similar to the United States and twice as high as in France.

JEL Classification: D3, E01, H2 H5, J3

1.1. Introduction

Who benefits from economic growth? This question is difficult to answer as national accounts, measuring the economic performance of a country's economy, are not linked to micro data reflecting the distribution of income across the country's population. The Distributional National Accounts (DINA) methodology first developed for the United States by Piketty et al. (2018) has established a new framework to measure the distribution of national income as defined by the System of National Accounts (SNA).[1]

This paper presents a high-quality DINA series for Germany from 1992 to 2016. We combine personal income tax files with SOEP survey data to create an individual income database for the

* The authors would like to thank Thomas Blanchet, Giacomo Corneo, Stefan Hauf, Wojciech Kopczuk, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman for helpful comments. Furthermore, we thank conference and seminar participants at the ECINEQ 2019, NBER's Conference on Research in Income and Wealth 2020, IARIW 2021, the World Inequality conference 2021, the workshop of the social policy committee of the Verein für Socialpolitik 2022, and the seminars of the German council of economic experts, at Paris School of Economics and at the Stockholm Institute of Transition Economics. Lotte Maaßen provided excellent research support. We gratefully acknowledge financial support from the Hans-Böckler-Foundation, Düsseldorf.

[1] There is also an effort led by the OECD and the European Commission, which is called EG DNA. See Coli et al. (2022) for EG DNA results for Germany and other countries for the year 2015. DINA and EG DNA differ in two ways: First, DINA series combine household surveys and income tax data to estimate percentile distributions with even smaller fractiles at the top, while EG DNA use only survey data to estimate quintile shares. Second, DINA distributes incomes of all sectors to private households, while EG DNA is limited to the household sector. See Bartels and Waldenström (2020) for more details on the differences between DINA and EG DNA.

entire adult population in Germany. Subsequently, we meticulously align each microdata income component with its national account equivalent to estimate the distribution of national income.

Our DINA series reveals two significant findings regarding the evolution of Germany's national income distribution from 1992 to 2016. First, the gap between the top 10% and the bottom 50% has widened. From 1992 to 2007, income growth was pro-rich. Incomes of the bottom 50% declined in real terms and their income share declined from 22% to 17%. Although growth has been pro-poor since 2007, the bottom 50% did not reach their income share of the early 1990s in 2016. The polarization between the bottom 50% and the top 10% can be attributed to the continuous increase in labor income among the lower 9% of the top decile (P90-99) throughout the entire period from 1992 to 2016. Second, Germany's economic elite receives an extraordinarily high income share. The top 0.1% receives approx. 6% of national income, which is comparable to the United States and twice as high as in France. Germany's economic elite primarily holds firms as partnerships owned by two to four shareholders, while top income earners in France and the United States hold shares in corporations with many shareholders. This contributes to a strong concentration of business incomes at the very top. For example, owning a partnership in Germany with up to three shareholders which earns more than EUR 2.4 million, immediately puts the owners into the top 0.1%, which started at about EUR 800,000 in 2016.

This paper makes a twofold contribution. First, we provide a high-quality DINA series for Germany that allows us to compare the distribution of national income in an internationally standardized way. Currently, only a handful of country teams are able, similar to our research, to combine personal income tax files with other administrative or survey data creating a comprehensive micro income database for the entire population that can be aligned to national accounts by component. The teams that have achieved this are Piketty et al. (2018) for the United States, Garbinti et al. (2018) and Bozio et al. (2023) for France and Bruil et al. (2022) for the Netherlands for 2016. The majority of DINA series employ a simplified approach of top-correcting household surveys using income tax tabulations and then uprating the top-corrected survey income distribution in broad categories to match national accounts. This simplified approach is adopted when personal income tax files are not available, which still applies to most countries.[2] Second, we adjust the DINA standard to the German data and economic structure and discuss the methodological implications of these modifications and research caveats.

The paper is organized as follows. Section 1.2 introduces our data sources, income concepts and methodology. A particular focus is on the imputation of national income components that are missing from one or both of our microdata sources, such as retained earnings. Another focus is on the mismatch and subsequent alignment between microdata and national accounts. Section 1.3 introduces our DINA series for Germany from 1992 to 2016. We present and discuss income share series in comparison with France and the United States, for which DINA series based on detailed microdata exist. Section 1.4 is on the robustness of our series towards alternative assumptions, particularly on the distribution of retained earnings and the alignment of microdata and national accounts by component. Section 1.5 compares our new DINA series to previous national income share series for Germany. Section 1.6 concludes. The paper is accompanied by an exhaustive Data Appendix that documents the details of the German national accounts, income tax returns and survey data, our methodological procedures and assumptions.

[2] Examples are, among others, Jestl and List (2020) for Austria, Guzzardi et al. (2022) for Italy, Novokmet et al. (2018) for Russia, Piketty et al. (2019b) for China, Chancel and Piketty (2019) for India and, last but not least, Blanchet et al. (2019) for 28 European countries. Researchers are also extending the DINA method to Latin American (De Rosa et al., 2022) and African countries (Chancel et al., 2023).

1.2. DINA concepts, methods and data sources

We base our analysis on the DINA method established by Piketty et al. (2018) and Blanchet et al. (2021). Yet, country-specific income data landscapes and economic structures require country-specific adjustments and modifications. For example, Germany's heavy reliance on non-corporate family firms requires us to modify the standards developed for the United States and France, where stock-listed firms are widely spread. The ultimate goal of our modifications to the DINA standard is to have a consistent series over time, ensuring a harmonized comparison with other countries. This section is a summary of our procedures; the Data Appendix to this paper explains data sources and methodological decisions in detail.

Section 1.2.1 briefly introduces the four income concepts of the DINA methodology. In Section 1.2.2, we describe our procedure to build an individual income database representative of the entire adult population in Germany, combining personal income tax returns and household survey data. Personal income tax (PIT) returns record incomes of the tax-filing population, while incomes of the non-filing population are obtained from the Socio-Economic Panel (SOEP).[3] Bringing together income tax and survey data requires data adjustments such as the common definition of the observation unit and income concepts. In Section 1.2.3, we discuss the most challenging aspect of the DINA estimation procedure: aligning microdata with national accounts' aggregates.

We face two challenges here. First, incomes like imputed rents or retained earnings are not visible in one (PIT) or both of our microdata sources. Subsections 1.2.3.1 and 1.2.3.2 discuss our strategies to simulate, impute or distribute these incomes. For example, imputed rents are recorded in the SOEP, but not in PIT data. Other examples are retained earnings or indirect taxes that have no microdata counterpart so that we have to decide on a rule how to distribute the aggregate. Assumptions how national accounts components without microdata counterpart are distributed crucially affect the level of income inequality and also its evolution over time.[4] Therefore, we undertake several robustness checks to study the sensitivity of our main series towards alternative distributional assumptions.

Second, labor, rental, capital and business incomes are visible in microdata and national accounts, but their aggregates do not necessarily align. Section 1.2.3.3 elaborates on how we align each income component at the micro and macro level. We collected additional subcomponents beyond the published sector accounts so that we can align incomes on narrowly defined income concepts. For example, for the component of net mixed income (B3n), we requested additional information to further split this component into rental income and sole proprietorship income.[5] As a robustness check, we vary the aggregation of the subcomponents when uprating microdata to national accounts' aggregates (see Section 1.4).

[3] Household survey data like the SOEP used in this paper are known to offer a good picture of the broad population, but suffer from undercoverage at the top of the income distribution, particularly of the top 1% (Bartels and Metzing, 2019). PIT data capture these top income earners, but exclude the poorer, non-filing part of the population.

[4] For example, Riedel and Stichnoth (2022) estimate that distributing government expenditure lump-sum instead of proportional reduces the gap in posttax income shares between the top 10% and bottom 50% in the United States from about 20 to 10 percentage points in recent years.

[5] Against the standards of United Nations et al. (2009), German national accounts pool sole proprietor and rental income in B3n and only include imputed rents in B2n (S14).

1.2.1. DINA income concepts

Following Piketty et al. (2018) and Blanchet et al. (2021), the DINA methodology comprises four income concepts: Pretax factor income (1), pretax national income (2) and posttax national income (3). All three concepts sum up to net national income which equals gross domestic product (GDP) after removing capital depreciation and adding net foreign income. Posttax disposable income (4) excludes in-kind transfers, e.g. for health, education and defense, and is thus smaller in aggregate than net national income. This version of the paper focuses on pretax factor and pretax national income.

Pretax factor income is gross market income from labor and capital including wages and salaries, self-employment and business incomes, dividends and interest as well as incomes from renting and leasing including owner-occupied housing rents (= imputed rent). Social insurance contributions are added to labor incomes. Pretax factor income measures market income before any operation of the social insurance or tax-transfer system. Pensioners – about 20% of the German population – receive very small amounts or even zero pretax factor income. Consequently, we display the distribution of pretax factor income for the population aged 20 to 64 years.

Pretax national income is the benchmark income concept. It takes the operation of the social insurance system into account. We add replacement incomes of the German social insurance system such as old-age pensions, unemployment benefits (*Arbeitslosengeld I*) and sickness benefits (*Krankengeld*) and subtract paid social insurance contributions.

Posttax disposable income is pretax national income minus all taxes plus monetary means-tested transfers. This concept excludes in-kind transfers, e.g., public expenditures for health, education and internal security. Thus, the aggregate of posttax disposable income is smaller than net national income (and the other three income concepts).

Posttax national income results from adding in-kind transfers and collective expenditures to posttax disposable income.

The four income concepts differ in how their components are allocated across our microdata observations. For example, we expect that pretax national income is more concentrated at the top of the distribution than posttax national income, which takes the operation of the tax-transfer-system into account.

1.2.2. Combining PIT and household survey data

Germany is one of the few countries which offers access to both the universe of PIT files and high-quality household survey data. Our combined income distribution covers both the tax-filing and non-filing population (e.g., low-income households and transfer recipients). We restrict our sample to the adult population of 20 and more years, following the DINA guidelines by Blanchet et al. (2021). Being able to draw on a variety of income components at the individual level in both survey and PIT data, we can align incomes much closer with United Nations et al. (2009) definitions than country studies based on grouped PIT statistics. Further, we can decompose pretax and posttax incomes into components to better understand the dominant inequality dynamics.

Microfiles of PIT returns are provided by the federal statistical office for the years 1992, 1995, 1998, 2001, 2004, 2007, 2010 and 2016.[6] For this version of the paper, we use a stratified 10% sample which provides complete coverage of the top of the income distribution. Germany has a joint income tax system so tax units are either married couples or singles.[7] Between 1992 and 2001, PIT microfiles include all tax units that filed a PIT return. Between 2001 and 2007, the statistical office subsequently added payroll taxpayers that did not file a PIT return.[8] As a result, the fraction of all potential tax units (= married couples + singles of 20^+ years) that is covered by PIT microfiles increased from 61% to 76% (see Appendix Figure 1.14).

We add the non-filing population from the SOEP, which is a representative annual survey of German households since 1984 recording information on socio-demographics, education, labor market status and detailed individual and household income sources (Goebel et al., 2019; Schröder et al., 2020). Here, we build on Bach et al. (2009) who combined individual income tax returns with SOEP household survey data to obtain gross and net income distribution series for Germany 1992-2005. We construct tax units in the SOEP data and, then, identify non-filers (& pure payroll cases) using a microsimulation model that simulates total income (*Gesamtbetrag der Einkünfte*), taxable income and its components for every SOEP tax unit. We identify likely filers and non-filing payroll cases based on the total income, income types, marital status and various deduction possibilities, e.g. for dependent children or high commuting costs. We add SOEP non-filers (& payroll cases) to PIT data until we match the observed number of households in official population statistics along the following groups: single/married x East German vs. West German residence x 5-year-age-groups of the household head from 20 to 70+ years.[9] Our microdata (PIT+SOEP) cover a population of about 67 million individuals and 60-63% of national income (see Appendix Figure 1.14).

In our benchmark series, we equally split incomes between married couples (equal-split series) which assumes equal sharing of incomes between spouses.

1.2.3. Aligning income tax, survey data and national accounts

The next step is to align income components in our combined microdata (PIT + SOEP) with national accounts aggregates. Subsections 1.2.3.1 and 1.2.3.2 explain how we impute labor and capital income components that cannot be observed in one or both of our micro datasets. Table 1.1 summarizes our allocation strategies.

1.2.3.1. Labor income

Employer's social insurance contributions are not part of taxable earnings. We simulate employer's contributions applying the prevailing contribution rates.

[6] RDC of the federal statistical office and statistical offices of the federal states, DOI: 10.21242/73111.(1992-2016).00.00.2.1.0

[7] For couples, incomes and other information are provided separately for each partner.

[8] The payroll tax (*Lohnsteuer*) on employment income is withheld at source. Employees only have to file a PIT return if they want to claim particular tax allowances, declare deductions and/or if they have other incomes than employment income.

[9] Before 2007, we additionally add pure payroll cases from SOEP as these are excluded from PIT files 1992- 2001 and subsequently added between 2001 and 2007.

Pensions are taxable to different degrees depending on the year of retirement. In PIT returns, we observe the taxable share of social insurance, occupational and private pensions. We recover the full amount by applying the prevailing taxable share.[10]

CEO compensation is visible as employment income in PIT returns and part of distributed income of corporations (D42, S14) in national accounts. We move CEO compensation from D42 (S14) to employment income D1 (S1). This modification slightly increases the labor share of net national income. In 2001 the German fiscal court set a ceiling for CEO compensation by sector to mitigate hidden profit distribution. To estimate the total of CEO compensations (which is not provided by the statistical office), we use these ceilings and multiply the sector-weighted average of CEO compensation with the number of limited liability companies (*GmbHs*) in each year. We estimate that CEO compensations amount to about EUR 43 billion in 1992 and EUR 105 billion in 2013.

1.2.3.2. Capital income

Imputed rent , i.e., the implicit rental income from owner-occupied housing, is not recorded in PIT, but in SOEP data. We estimate the distribution of imputed rents by income fractile, marital status, 5-year age group of the tax unit head and region (West/East Germany) in SOEP and assign these mean-values to likely homeowners among PIT observations by group.

Dividend and interest income is subject to different tax regimes during our observation period which affects their coverage in PIT files: 1) Before 2001, dividends were fully taxable; 2) From 2001 to 2008, only 50% of dividends,[11] but 100% of interest were taxable under the personal income tax rate so that we have to single out dividends and double them (*Halbeinkünfteverfahren*); 3) Since 2009, dividends and interest income are taxed at source with a withholding tax of 25% (*Abgeltungssteuer*) and not systematically reported in income tax returns anymore.[12] Since 2009, we therefore extrapolate the distribution of dividends and interest incomes using the average share of dividends and interest in total dividend and interest income between 2001 and 2007 by income fractile, region, marital status and gender. This means we assume that the incidence of dividends and interest income across those groups remained stable since 2007, while the aggregate evolves in line with national accounts. Note that dividends and interest income are comparably small in Germany, even for top income earners. We will see in Subsection 1.3.1 that German top income earners dominantly derive their income from non-corporate, closely-held firms.

Dividend and interest income below the saver's allowance is usually not declared. The saver's allowance (*Sparerfreibetrag*) was raised from 600 DM (EUR 307) in 1992 to 6,000 DM from 1993 and reduced to EUR 801 in 2009. We mean-value impute dividend and interest income under the saver's allowance from SOEP data for PIT filers who indicate zero dividend and interest income in their tax return.

[10] The taxable share of pensions varies with the year of retirement and has increased annually since 2005. For example, the taxable share was 50% in 2005 and increased by 2%-points each year until 2020 (80%). Since 2021, it increases by 1%-point yearly until 100% of pensions will be taxable from 2040.

[11] The so-called Halbeinkünfteverfahren (half-income method) was applied to income from foreign equity investments since 2001 and to domestic equity investments in corporations since 2002.

[12] For some tax filers, it is still beneficial to declare dividends, for example, if dividends are counter-balanced by large losses or expenses that decrease taxable dividends, or if the personal tax rate is lower than the withholding tax rate.

Retained earnings are profits withheld within the firm, i.e., not distributed to the owners. In DINA for France and the United States, where large firms are usually held as a corporation, retained earnings are allocated proportionally to distributed corporate profits, i.e., dividends. In Germany, not only corporations such as limited liability companies (*Gesellschaft mit beschränkter Haftung - GmbH*) and public companies (*Aktiengesellschaft - AG*) but also partnerships play an important role, even for large firms. Partnerships are pass-through entities, i.e., profits are distributed and taxed at the personal tax rate of the shareholder. However, a special form of a limited partnership, the *GmbH & Co KG*, can retain profits since the general partner (with unlimited liability) is a limited liability corporation (*GmbH*).[13]

In our baseline scenario, we modify the DINA guidelines rule and distribute the national accounts aggregate of retained earnings (B5n, S11 + S12) proportional to received dividends + business income from partnerships, which are classified as quasi-corporations.[14] We also compute alternative scenarios inspired by Alstadsæter et al. (2023) in Figure 1.11: (1) We display the top 1% income share without retained earnings, removing retained earnings also from total income (denominator). (2) We distribute retained earnings lump-sum to all individuals receiving either dividends or business income from partnerships. (3) We distribute retained earnings lump-sum to all individuals receiving either dividends or business income from partnerships in the top 1%. We discuss these alternative scenarios together with the results in Section 1.4.

Private insurance income (D441 + D442, S14 + S15) is the profit margin of private health, life and pension insurance which is neither declared in PIT nor SOEP data. Main holders of private health and pension insurance are civil servants, entrepreneurs and top-earning employees, while most employees are covered by the social insurance system. Therefore, we identify likely private insurance policy holders based on income and status (e.g., civil servants). Despite recent fiscal incentives to invest in private pension insurances, this type of pension is still of minor importance in Germany.

Paid interest on mortgages is part of the overall interest paid by households (D41, S14 + S15). To account for the tax deductibility of interest payments for landlords and business owners, we directly subtract interest paid by landlords and business owners from rental and sole proprietor income (B3n, S14) in national accounts data. Interest paid on mortgages for owner-occupied housing is recorded in SOEP data. We mean-value impute this information to PIT cases.

Capital gains from pure asset price changes are not part of national income. Realized capital gains from selling a partnership or sole proprietorship are taxable in Germany so that we have to deduct these gains from PIT business incomes to match NA income definitions.

[13] Appendix Figure 1.13 illustrates the particular construction of the *GmbH & Co KG*: It is a partnership where the general partner is a previously established limited liability corporation (*GmbH*), which is usually held by the same owners. As a partnership, i.e., a pass-through entity like the limited partnership (*Kommanditgesellschaft - KG*), profits are considered as non-corporate business income which is subject to personal income tax. However, if the *GmbH* manages the *GmbH & Co KG*, then only 60% of the dividends of the *GmbH* are taxable (*Teileinkünfteverfahren*). It should be added that setting up and maintaining a *GmbH & Co KG* is costly due to additional formal requirements such as two separate balance sheets. Appendix Figure 1.16 illustrates the importance of partnerships owned by less than 4 shareholders at the top of the German partnership income distribution.

[14] The corporate sector (S11 + S12) in German national accounts also includes partnerships like *KG* and *OHG*, which national accounts classify as *quasi-corporations*. We cannot differentiate between these partnership forms in PIT data so that, in our benchmark scenario, retained earnings are distributed to all partnership owners.

Table 1.1.: Aligning income tax, survey data and national accounts

Labor income components	
Non-filer labor income	Estimated taxable income below income tax allowance in SOEP (= identified non-filers) (DA A.4.4)
Employer contributions	Simulated using individual's earnings in PIT & SOEP (DA A.3.2.1)
Social insurance, occupational and private pensions	Taxable share of pensions upscaled to full amount
CEO compensation	Estimated aggregate of CEO compensation moved from D42 to D1 in NA data
Capital income components	
Imputed rent	Mean-value imputation based on imputed rent from SOEP to PIT-filers
Dividends + interest < savings allowance	Mean-value imputation based on SOEP capital incomes to PIT-filers who did not declare capital income (DA A.3.6.2)
Retained earnings	Corporate sector income (personal component) distributed proportionally to the sum of positive dividend and shareholder income from partnerships.
Private insurance income	D441(private health + life insurance) distributed proportionally to income of civil servants, self-employed and business owners D442 (private pensions) distributed proportionally to income of self-employed and top earner employees (above SI ceiling)
Paid interest on mortgages	Mean-value imputation based on SOEP information to PIT filers

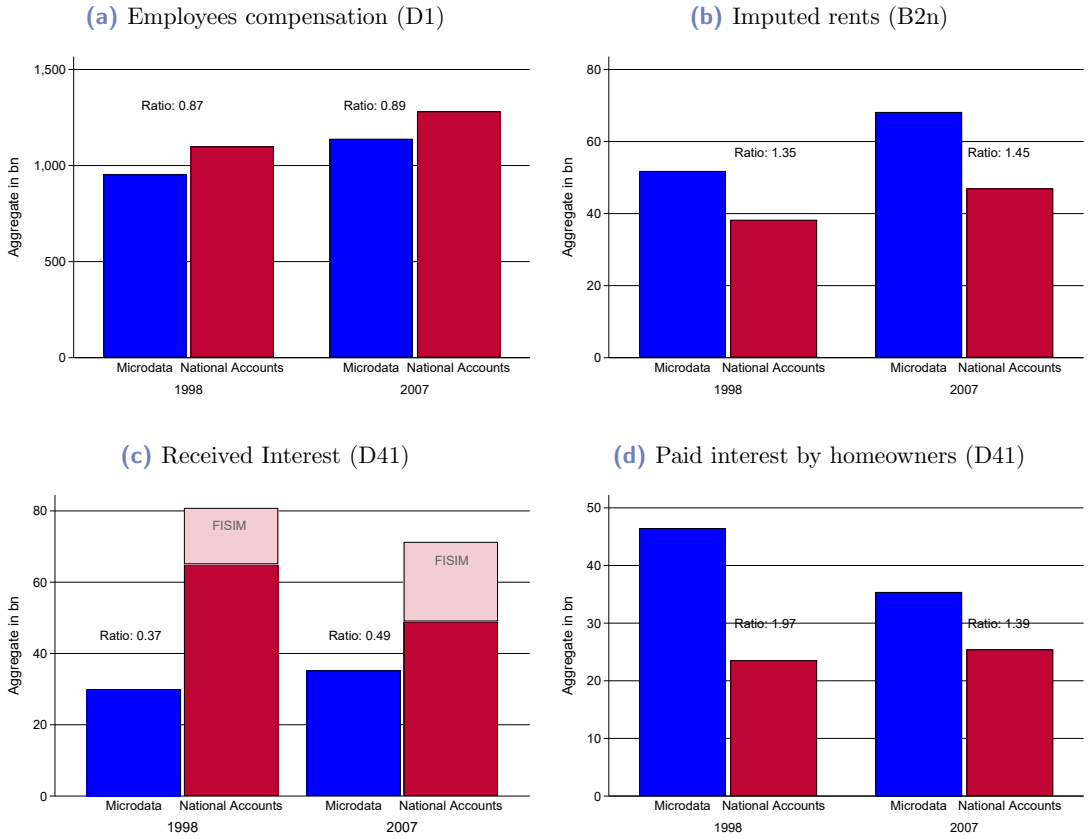
1.2.3.3. Uprating microdata to national accounts

Income components visible in microdata and NA are employment income (D1), interest income (D41), dividend and partnership income (D42), imputed rent (B2n) and mixed income (B3n), which includes sole proprietor income and rental income. Our microdata (PIT+SOEP) list all of these items separately so that we can compute the ratio between the micro and macro aggregate for each component. However, even after aligning income concepts as closely as possible, a micro-macro-mismatch remains.

While our micro aggregate of employment income matches the national accounts aggregate (D1) well (see Figure 1.2a), there is a sizable gap for other income components. This raises the question which components to combine when uprating microdata to national accounts' aggregates.

Figure 1.1 shows the aggregates and the micro-to-macro ratio for employment income (D1), interest income (D41) and imputed rent (B2n). Our microdata cover approx. 90% of employment income. Imputed rents are higher in our microdata than in national accounts: microdata suggest 40% more than national accounts indicate. A sizable mismatch also appears for received and paid interest.

Figure 1.1.: Micro-macro-mismatch for D1, B2n and D41, 1998 and 2007



Note: Financial intermediation services indirectly measured (FISIM) accounts for service fees of financial institutions. FISIM is supposed to capture the difference between interest received and paid by households and the interest that households would have received or paid under the riskfree interbank rate. Thus, FISIM effectively increases interest received and decreases interest payments by the household sector. FISIM is booked in the household sector and thus allocates income to households that effectively flow to financial institutions (and their employees and shareholders). Interest paid by company owners and landlords (D41 paid, S14 + S15) is deducted from B3n in NA to align with microdata income concepts, so that D41, as shown here, only includes interest paid by homeowners.

Table 1.2.: Firms' categorization into the corporate or the household sector by legal form.

	Corporations		Partnerships				Sole
	AG	GmbH	KG	GmbH&Co KG	OHG	GbR	proprietor
National accounts	co	co	co	co	co	n-co	n-co
Income tax law	co	co	n-co	n-co	n-co	n-co	n-co

Note: This table displays the categorizations of the most widely used legal forms of firms in Germany according to national accounts and to income tax law. In national accounts, partnerships like the limited partnership (*Kommanditgesellschaft - KG*), *GmbH & Co KG* and the general partnership (*Offene Handelsgesellschaft - OHG*) are defined as *quasi-corporations* and categorized as part of the corporate sector. Partnerships under civil law (*Gesellschaft bürgerlichen Rechts -GbR*) and sole proprietorships are categorized as part of the household sector. German income tax law distinguishes between corporate income, i.e., dividends from limited liability companies (*Gesellschaft mit beschränkter Haftung - GmbH*) and public companies (*Aktiengesellschaft - AG*) and business income.

The line between dividend and partnership income (D42, S14+S15[15]) and mixed income (B3n, S14+S15), which includes sole proprietor income and rental income, is drawn differently by income tax legislation and national accounts. Table 1.2 displays how national accounts and income tax law differently categorize legal forms. German national accounts distinguish between the corporate sector (S11+12) and the household sector (S14), but include those forms of partnerships commonly used by German family firms as *quasi-corporations* in the corporate sector. This means that income from these partnerships is part of D42 or remains as retained earnings in the corporate sector in German national accounts.

We find notable disparities between micro and macro data for dividend and partnership income (D42) and sole proprietor and rental income (B3n).[16]

Reports of the statistical office acknowledge substantial uncertainty about the amount and composition of business and property income in German national accounts. All net operating surplus – including business incomes from sole proprietors and partnerships, interest and dividends – is computed as a residual in German national accounts by deducting employment income (D1), intermediate consumption, depreciation and net production taxes from output (P1) in the production account (Statistisches Bundesamt, 2016).[17] The estimated business income is then distributed between the corporate sector (including quasi-corporations) and the household sector according to the turnover share of corporations vs. non-corporate firms. In our view, using turnover shares of corporate vs. non-corporate firms likely allocates to much profit to the corporate sector that generates higher turnover but not necessarily higher profits. For example, VAT statistics show that sole proprietors and partnerships contribute about 26% of total turnover, while business tax statistics show that sole proprietors and partnerships contribute about 46% of total business profits (*Gewerbeertrag*). In line with this hypothesis, we find that the microdata aggregate of sole proprietor income is close to or even exceeds the national accounts aggregate, while the dividend and partnership microdata income only reaches 50 to 60% of the macro aggregate. In our benchmark estimation, we resolve this issue by pooling dividends, partnership, sole proprietor and rental incomes before aligning with national accounts. Figure 1.2a shows that

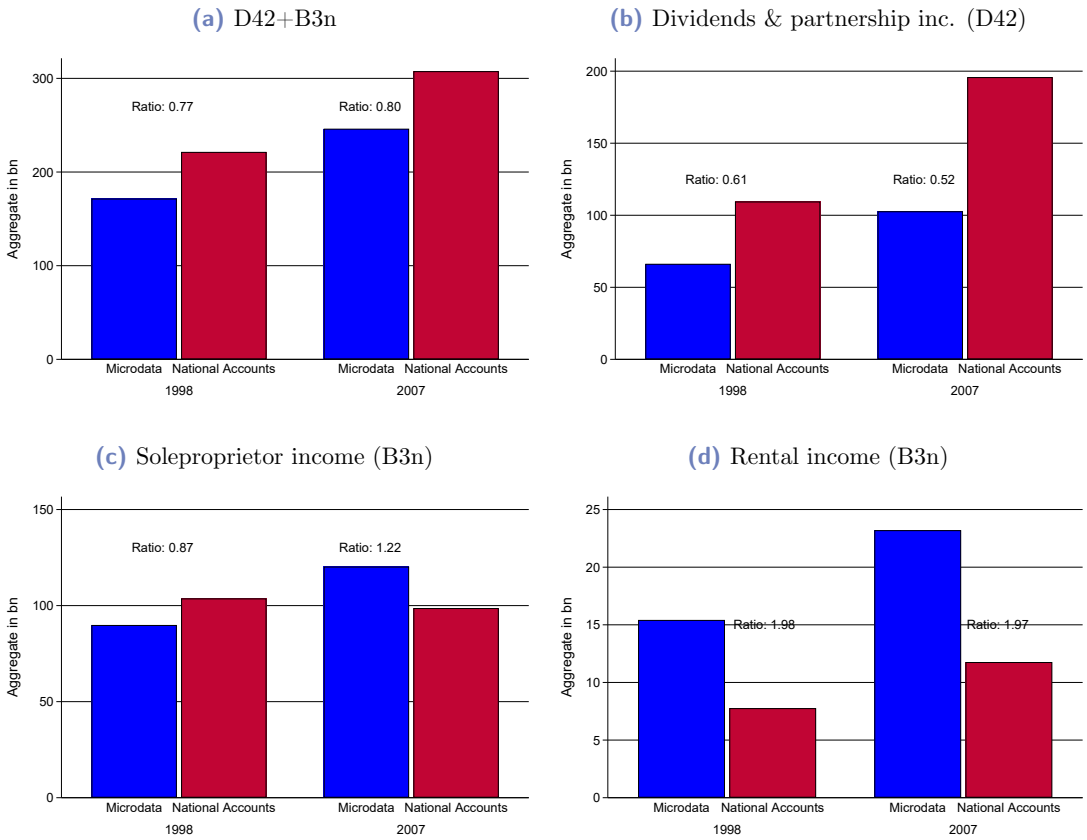
[15] German sector accounts pool the household (S14) and non-profit sector (S15).

[16] In PIT data, we can distinguish between dividend income (D42), business income from partnerships (D42) and sole proprietors (B3n) and rental income (B3n). In national accounts, we have separate information for rental and sole proprietor income (B3n), but only pooled information for dividends and partnership income (D42).

[17] Bach et al. (2009) also find significant discrepancies between their microdata-based aggregate and sector accounts as well and come to the same explanation. Bach (2013) highlights that conceptual differences in the income definitions and determination rules result in higher business income in national accounts compared to tax data.

we capture approx. 80% of the resulting D42+B3n aggregate in our microdata. We show and discuss results from uprating these items separately in Section 1.4.

Figure 1.2.: Micro-macro-mismatch for B3n and D42, 1998 and 2007



Note: The national accounts aggregates of sole proprietor and rental income (B3n) are reduced by the interest paid by company owners and landlords (D41 paid, S14 + S15) to harmonize with micro data income concepts.

1.3. Distributional National Accounts for Germany, 1992-2016

Table 1.3 presents the income threshold, average income and income shares of five groups in the income distribution in 2016. Average income of individuals aged 20+ (67 million individuals) is approx. EUR 40,000. The bottom 50% earned about EUR 15,000, which is less than a third of average income, and 19% of national income. Moving to the top reveals a very skewed income distribution.

Within the top 1%, incomes are almost equally split between the lower 0.9% (P99-99.9) and the top 0.1%, whose income shares are 7.7% and 6.3%, respectively. Within the top 0.1%, the same pattern emerges between the lower 0.09% (P99.9-99.99) and the top 0.01%, whose income shares are 3.3% and 3.0%, respectively. The top 0.01% comprises a group of around 6,700 individuals whose income is at least EUR 3.8 million and who earned, on average nearly EUR 12 million in 2016.

Figure 1.3 presents the pretax national income shares between 1992 and 2016 for four groups: bottom 50%, middle 40% (P50-99), the next 9% (P90-99) and the top 1%, 0.1% and 0.01%. To immediately place the German levels and trends into international comparison, we also display

Table 1.3.: Pretax national income by income group, 2016

Income Group	Number of adults	Income Threshold	Average Income	Income Share
Full population	66,837,765		39,705€	100%
Bottom 50%	33,418,829		14,729€	18.5%
Middle 40%	26,735,161	28,999€	44,375€	44.7%
P90-99	6,015,396	70,200€	100,262€	22.7%
P99-99.9	601,541	209,462€	341,868€	7.7%
P99.9-99.99	60,153	811,589€	1,440,982€	3.3%
Top 0.01%	6,685	3,832,683€	11,933,152€	3.0%

Note: Population aged 20+. Income in current prices.

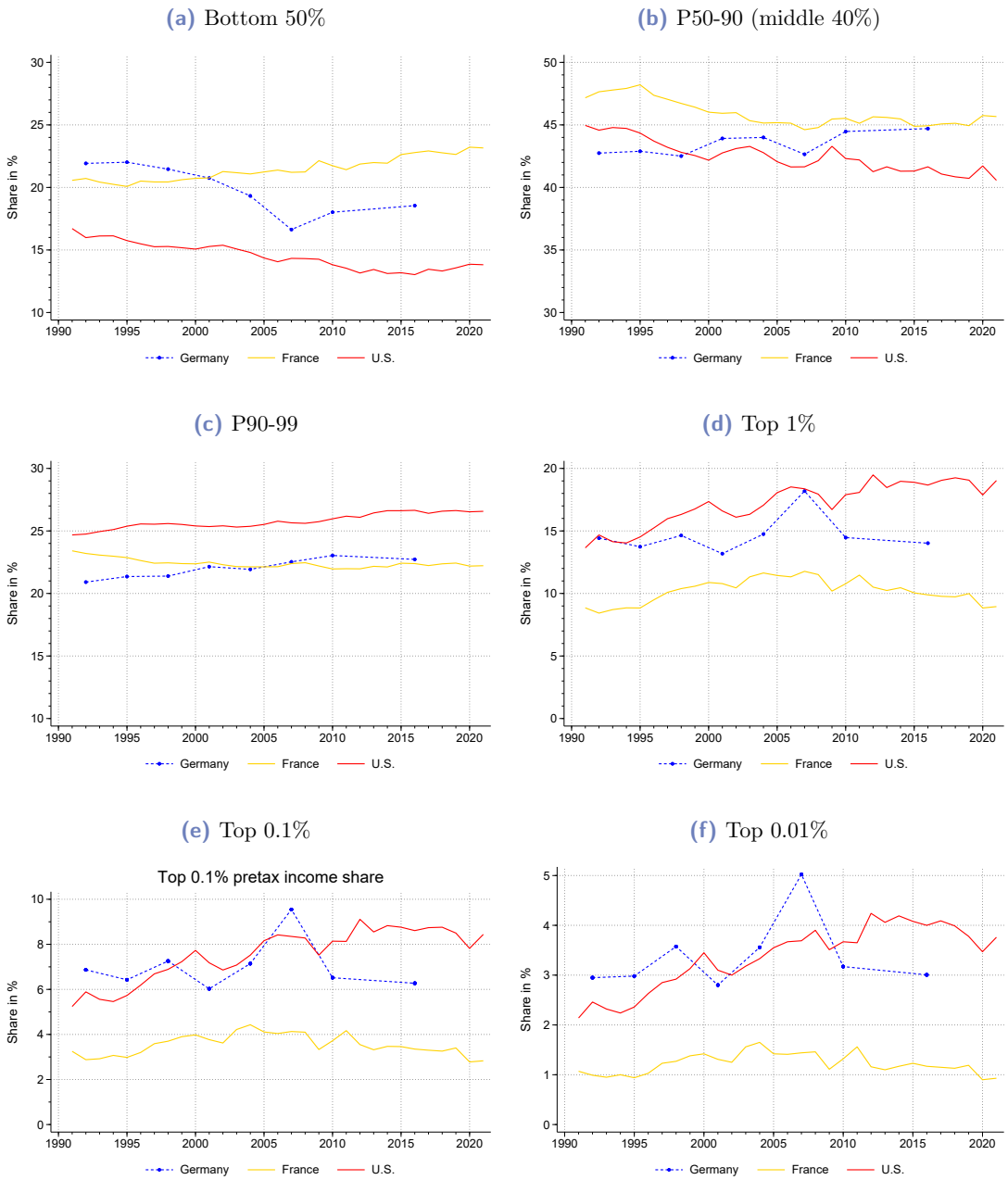
the respective income shares of France and the United States. For these countries, DINA series are based on a combination of microdata, most importantly PIT microfiles, which arguably provide the best possible coverage of top income earners in a country.

The bottom 50% income share in Germany declined from 22% in 1992 to 17% in 2007. This trend reversed thereafter so that the share reached almost 19% in 2016 (Figure 1.3a). This downward trend is also documented by Blanchet et al. (2022a) adopting a simplified DINA approach, yet they do not find a trend reversal in 2007. The evolution of total national income in Germany presented in Figure 1.6 reveals that gross wages and salaries virtually stagnated between 1992 and 2007 and subsequently started to increase. We discuss this evolution in more detail in Section 1.3.1. In France, the bottom 50% received between 20% and 26% of national income during this period.

The middle 40% received a comparably stable share of pretax national income of similar magnitude (approx. 45%) in Germany and France. In constrast, this group clearly lost out in the United States where their income share has declined from 45% to 41% between 1991 and 2021 (Figure 1.3b).

The top decile's income share followed quite different paths across countries: It has increased in Germany and the United States, whereas it has remained rather stable in France. While the top 1% drive this trend in the United States (Figures 1.3d), the P90-99 gained more in relative terms in Germany (Figures 1.3c). All in all, Germany stands between the United States and France. From a European perspective, Germany's income concentration at the top is high and the income share of the bottom 50% is low. Both in the United States and in Germany, the top 0.01% receives about 3 to 4% of national income. We discuss this further in Section 1.3.1.

Figure 1.3.: Pretax national income shares in international comparison

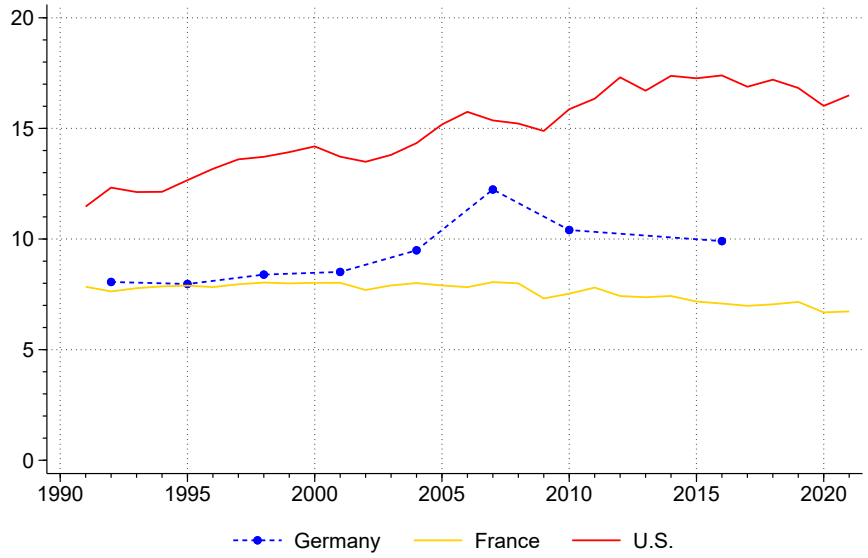


Note: Series for France by Bozio et al. (2023) and Garbinti et al. (2018) and for the United States by Piketty et al. (2018).

Our analysis shows that the bottom 50% lost out in relative terms in Germany and the United States, while the top 10% have gained. How does this picture evolve if we directly compare average incomes of the top 10% and the bottom 50% in these countries? Figure 1.4 displays this ratio and reveals a clear increase in this ratio for Germany and the United States, while the trend is flat for France. Germany started in the 1990s at par with France, when the top 10% earned eight times the average income of the bottom 50%. Since the early 2000s the German ratio has increased and reached a value of about twelve in 2007 and then declined to stabilize at

ten. For the United States, this ratio has almost continuously increased from a value of about twelve in the 1990s to between 16 and 17 since 2011.

Figure 1.4.: Top 10% average income to bottom 50% average income

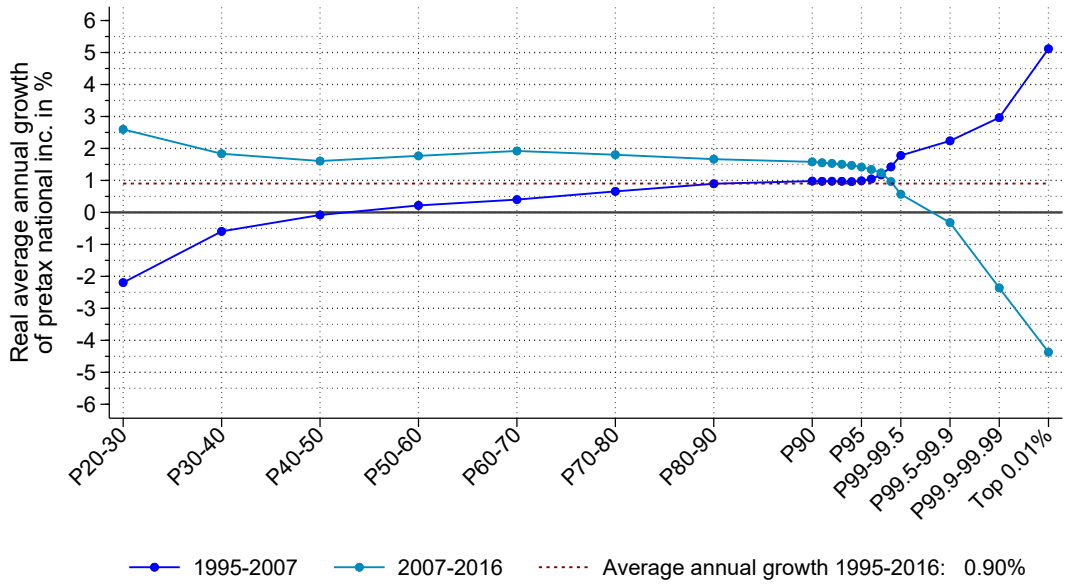


Note: Top 10% average income divided by bottom 50% average income. Series for France by Bozio et al. (2023) and Garbinti et al. (2018) and for the United States by Piketty et al. (2018).

Who has benefited from economic growth in Germany? The results presented up to this point reveal a turning point in 2007. Growth incidence curves of consecutive data years (see Appendix Figure 1.22) confirm such a pattern. We, therefore, display growth incidence curves for two periods, 1995-2007 and 2007-2016, in Figure 1.5. Average real annual growth was less than 1% during this period.[18] Growth has been pro-rich between 1995 and 2007 and pro-poor between 2007 and 2016.

[18] Note that DINA measures net national income growth, and not gross domestic product growth which indicates higher growth rates.

Figure 1.5.: Growth incidence curve, pretax national income, 1995 - 2016



Note: Figure 1.5 shows the growth incidence along the pretax national income distribution for 1995 to 2007 and from 2007 to 2016. Both time periods show diametrical growth patterns. While growth was dominantly pro-rich from the 1990s to 2007, it is pro-poor growth since 2007. For separate graphs for consecutive years, see Figure 1.22.

1.3.1. Breakdown by income source

To better understand the differing growth dynamics across the income distribution, we decompose pretax factor income by components for different groups along the distribution. In this section, we use pretax factor income, i.e. the sum of all market incomes before any operation of the social insurance or tax-transfer system, to uncover heterogeneous market income sources across distributional groups that might have contributed differently to the observed inequality dynamics.

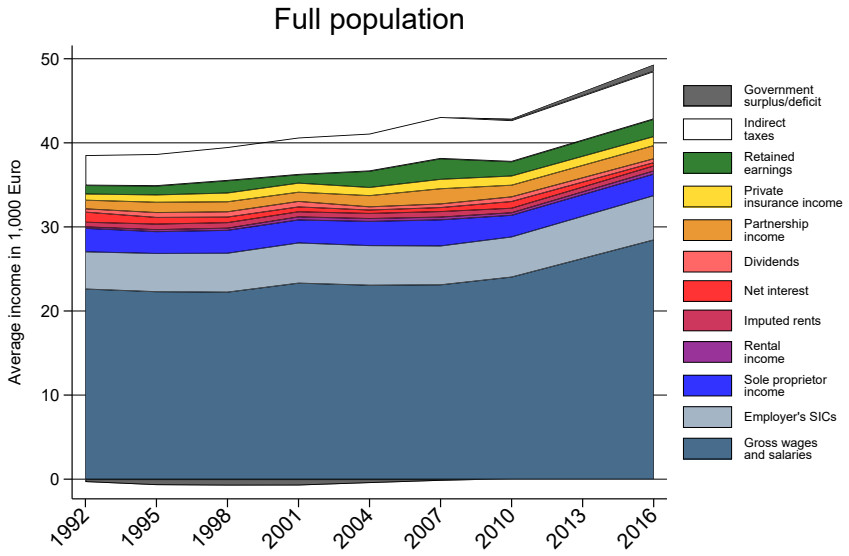
Figure 1.6 presents the evolution of pretax factor income per capita from 1992 to 2016. Pretax factor income per capita increased from almost EUR 40,000 in 1992 to almost EUR 50,000 in 2016 (in prices of 2015). About 70% of pretax factor income is gross wages and salaries plus employer's social security contributions. Business incomes from partnerships contribute with 3 to 4% and from sole proprietors with 5 to 7% to pretax factor income. Compared to the importance of partnership and sole proprietorship income, dividends and interest income play a minor role in Germany. Dividends represent less than 2% of pretax factor income throughout the period.

Between 1992 and 2016, firms increasingly retained their earnings, particularly since the early 2000s. Retained earnings increased from 2.6% of pretax factor income in 1992 to almost 6% in 2007 and stabilized around 4% in the 2010s.[19] Between 1992 and 2007, gross wages and salaries almost stagnated in real terms at about EUR 23,000. Since 2007, gross wages and salaries started to grow, reaching approx. EUR 28,500 in 2016.

While Figure 1.6 shows a slow but steady increase in real pretax income, this averaged visualization disguises radically different patterns for different income groups. Figures 1.7 and

[19] This share only includes retained earnings that we attribute to the household sector. We split the aggregate of retained earnings (before tax, B5n, S11+S12) into a share received by households and another received by the government (included in the government surplus/deficit) based on the aggregate ratio of equity and investment fund shares held by each sector (see Data Appendix A).

Figure 1.6.: Composition of pretax factor income per capita



Note: Pretax factor income per capita in EUR 1,000 (2015 prices). Pretax factor income comprises all market incomes before any operation of the social insurance and tax-transfer system. Excluding the population aged 65+ that largely has no market income.

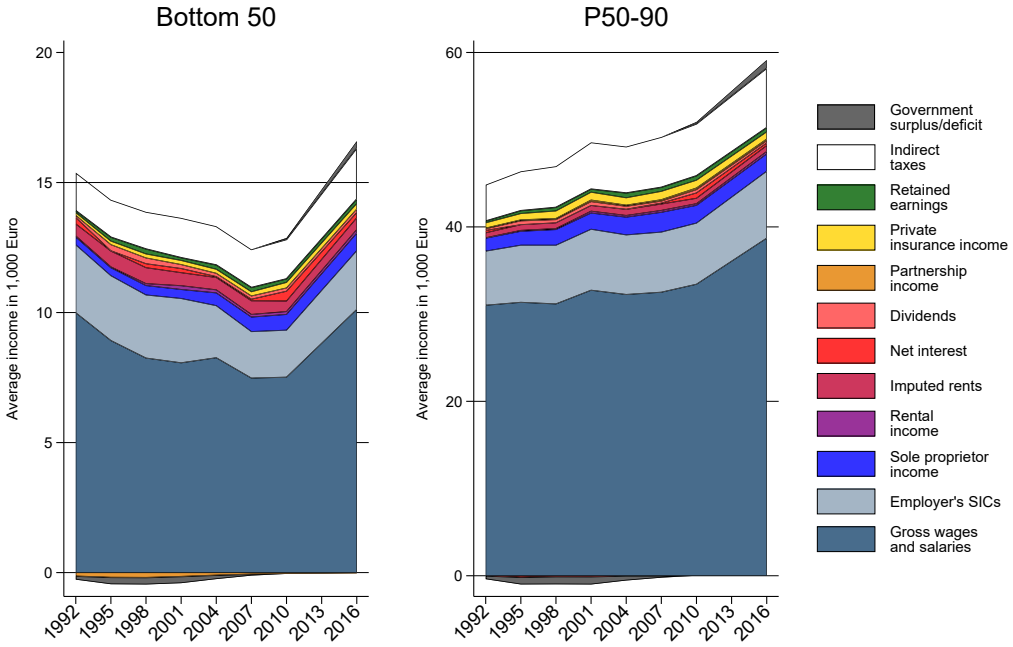
1.8 display the composition of factor incomes from bottom to the very top of the German income distribution. We start with the bottom 90% of the income distribution in Figure 1.7, which we further split into the bottom 50% (left-hand graph) and the middle 40% (P50-90, right-hand graph). For the bottom 90%, incomes largely consist of gross wages and salaries (between 73 and 83% for the bottom 50% and between 78% and 84% for the middle 40%). Between 1992 and 2007, the income dynamics were, however, remarkably different for the two groups: Real average pretax factor income of the bottom 50% dropped from approx. EUR 15,000 to EUR 12,500 (in 2015 prices), while the average income of the middle 40% slightly increased from EUR 44,500 to EUR 50,000 (in 2015 prices). The year 2007 marked a turning point after which growing real hourly wages and monthly earnings raised incomes throughout the bottom 90%. Grabka (2022) names two reasons: First, nominal wage hikes translated into real wage growth in an environment of economic growth, increasingly tight labor markets and low inflation. Second, Germany's low-wage sector stagnated since 2007.

Figure 1.8 displays the top 10% average income composition, which we further split into the lower 9% of the top decile (P90-99, left-hand graph) and the top 1% (right-hand graph). The income composition of the lower 9% looks very similar to the bottom 90%, with the majority of income stemming from labor. The main difference comes from the larger importance of sole proprietor income, and, to a smaller extent, of partnership income, dividends and interest income. Between 1992 and 2016, real average income of the lower 9% of the top decile have continuously expanded from approx. EUR 88,500 to EUR 126,500 (in 2015 prices).

The top percentile's income composition is significantly different from the bottom 99%, analysed so far. Wages and salaries represent less than 25% of this group's average income. Dividends add 2.5 to 5%. The share of sole proprietorship income has decreased from about 20% in the early

1990s to 10% in 2016. A similar decline is visible in net interest income from 12% in 1992 to 1.5% of this group's income in 2016. Throughout the time period, the bulk of top incomes stems from closely held partnerships and their retained earnings, reflecting the particular structure of the German economy. Even among Germany's largest firms with respect to their turnover, almost 40% are partnerships, like *KG* and *GmbH & Co KG*, and only about 15% are public companies, i.e., *Aktiengesellschaften* (Appendix Figure 1.20).[20] At the very top of the German income distribution (top 0.01%), more than 70% of income is derived from partnerships and their retained earnings (see Appendix Figure 1.19).

Figure 1.7.: Pretax factor income decomposition, Bottom 90%



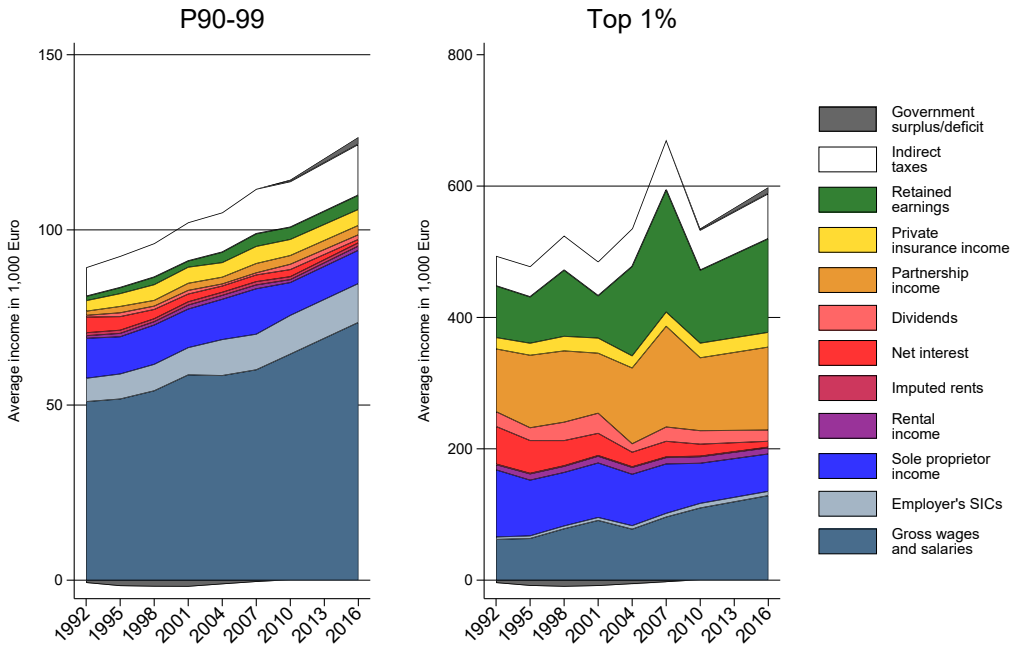
Note: Pretax factor income in EUR 1,000 (2015 prices). Pretax factor income comprises all market incomes before any operation of the tax-transfer system. Excluding the population aged 65+ that largely has no market income.

The fact that German top income earners hold their firms as partnerships instead of corporations is a candidate explanation for the persistence of German top incomes. Persistence rates of persons in top income fractiles in Germany are higher than in Canada, France and the United States (Jenderny, 2015).[21] In contrast to (stock-listed) corporations, ownership of partnerships is less flexible and more concentrated. Among those partnerships in Germany earning more than EUR 2.5 million in 2007 (around 6,500 partnerships), about two-thirds were owned by two to four shareholders (left-hand graph of Appendix Figure 1.16). Around 30% of the income of these top-earning partnerships was distributed to three or four shareholders and approx. 30% to only two shareholders (right-hand graph of Appendix Figure 1.16). This means that, for example, owning a partnership with up to three shareholders which earns more than EUR 2.4 million,

[20] This structure is very stable over time (see Appendix Figure 1.21).

[21] A picture of persistence also emerges when looking at the value of family firms by their founding year. The most valuable German firms today were founded during the period of rapid industrialization between 1850-1914 (Appendix Figure 1.15).

Figure 1.8.: Pretax factor income decomposition, Top 10%



Note: Pretax factor income in EUR 1,000 (2015 prices). Pretax factor income comprises all market incomes before any operation of the tax-transfer system. Excluding the population aged 65+ that largely has no market income.

immediately puts the owners into the top 0.1% of the personal income distribution, which started at approx. EUR 800,000 in 2016.

Several features of the German tax system incentivize the creation of partnerships over corporations:[22] (1) Partnership profits are not subject to the corporate tax of 15% (*Körperschaftsteuer*).[23] (2) The local business tax (*Gewerbesteuer*), amounting to 15 to 20%, includes an allowance and is (partly) credited against the personal income tax, which is not possible for corporate owners. Taking (1) and (2) together, creates a lower income tax burden for partnership income compared to corporate income. (3) Further, partnerships, including *GmbH & Co KG*, can be exempt from inheritance and gift taxation if the recipient's share exceeds 1%, while a tax-exempt transfer of corporate shares is restricted to heirs receiving shares exceeding 25%.

The particularity of Germany's reliance on closely held family firms becomes apparent when comparing the income source distribution with the United States. Figure 1.9 shows income sources by pretax factor income decile in Germany and the United States, zooming into smaller groups at the top of the distribution. In both countries, the lower 95% of the income distribution derive their income from labor.[24] While the importance of capital income in the form of interest

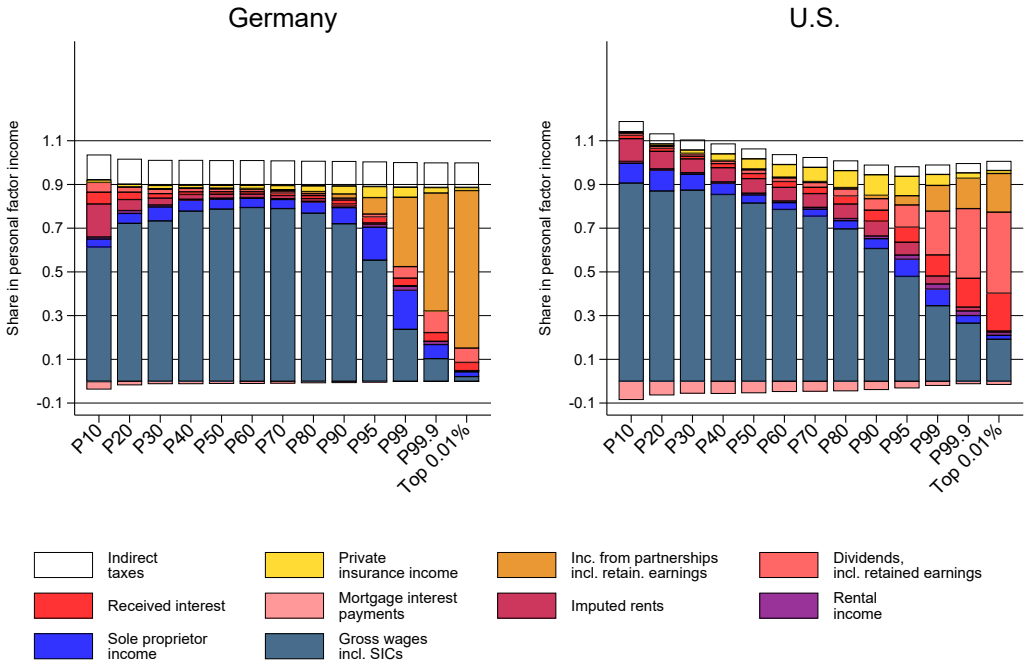
[22] This tradition goes back at least until the Nazi regime that viewed the *anonymous capital* of corporations as problematic and incentivized the conversion of corporations into partnerships through corporate tax reforms starting in 1934 (Banken, 2018, p. 406). As a consequence, the number of corporations fell by two-thirds between 1933 and 1939 (see Data Appendix Section 6 of Albers et al. (2022) for further elaborations).

[23] The *GmbH* of the *GmbH & Co KG* is usually constructed as profit-neutral (*gewinnneutral*) so that corporate taxation does not apply.

[24] The higher share of home ownership in the United States results in both higher debt service for mortgages (negative net interest income) and higher shares of imputed rent.

income and dividends continuously increases when moving towards the top of the U.S. income distribution, these income types are of minor importance in Germany, even for top income earners. Dividends represent almost 40% of top 0.01% incomes in the United States, but only approx. 7% in Germany. Partnership income represents 72% of top 0.01% incomes in Germany, but only about 18% in the United States.

Figure 1.9.: Pretax factor income composition in Germany and the United States

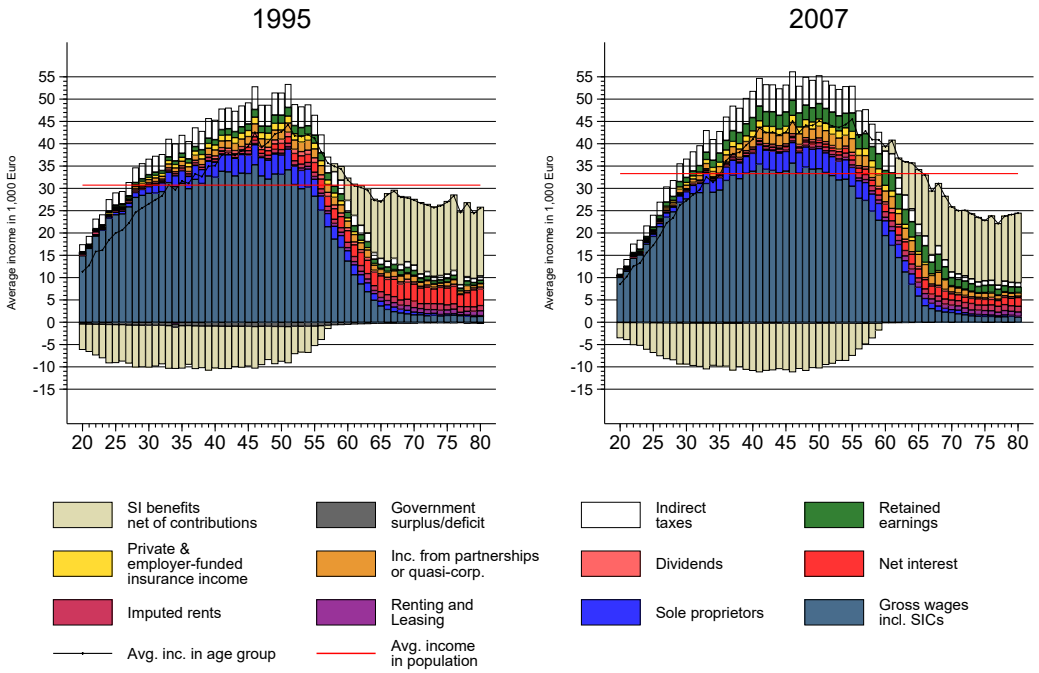


Note: Pretax factor income distribution in 2007. Sources: U.S. data based by Piketty et al. (2019a). For Germany, own calculations based on RDC of the Federal Statistical Office and Statistical Offices of the Federal States, DOI: 10.21242/73111.(2007).00.00.2.1.0 & SOEP. German partnerships that are constructed as *GmbH & Co KG* can retain earnings in the *GmbH* so that retained earnings are also distributed to partnership owners in Germany (see Section 1.2.3 and Appendix Figure 1.13).

1.3.2. Breakdown by age

Figure 1.10 displays the distribution of pretax national income by age for two exemplary years, 1995 and 2007. Several findings are worth noting. First, we see a typical inverse u-shape of lifetime income that peaks between ages 40 and 50 and then declines after age 55. Second, sole proprietorship and partnership incomes gain importance in the late 30s. While average wages decline quickly after age 55, sole proprietorship and partnership incomes decline much slower and remain important even after the official retirement age. Third, from age 20 to 60, the average German is a net contributor to the social insurance system and, then, becomes a net recipient. Fourth, retiree income has moved further away from the average income (red line). Net interest income contributed a sizable portion of average retirement income in 1995 but is hardly visible in 2007.

Figure 1.10.: Pretax national income by age, 1995 and 2007



Note: Pretax national income in EUR 1,000 (2015 prices). Social insurance benefits include old-age pensions, unemployment benefits and sickness benefits.

1.4. Robustness towards alternative assumptions

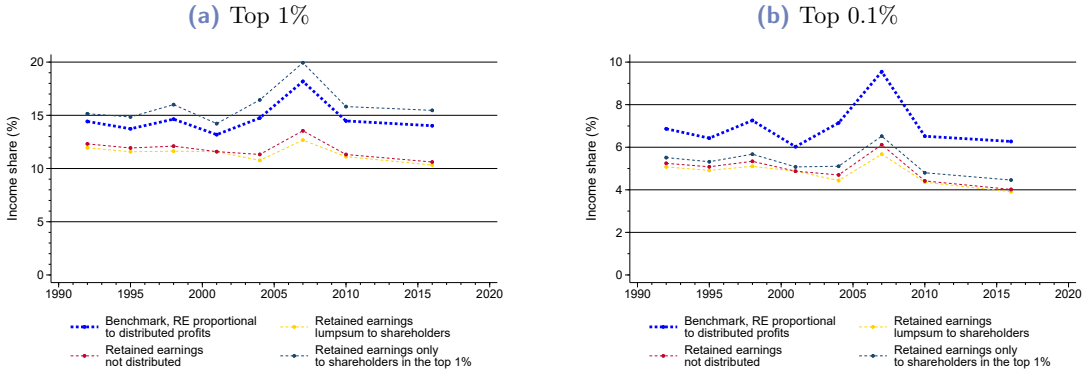
Constructing DINA involves a number of assumptions with potentially large effects on levels and trends of income inequality. This subsection reviews the robustness of our series towards alternative assumptions regarding the distribution of retained earnings and the uprating basis for dividend, business and rental income.

Retained earnings are distributed proportionally to distributed profits in the form of received dividends and business income from partnerships in our benchmark scenario. This strategy is in line with the standard DINA methodology (Blanchet et al., 2021; Piketty et al., 2018). However, this assumption has been contested: Alstadsæter et al. (2023) link individual and firm administrative data from Norway and find that distributed profits perform well as a proxy for retained earnings in boom years, but not in recessions. We compute three alternative scenarios based on Alstadsæter et al. (2023)'s findings[25]: (1) We discard retained earnings, i.e., we compute the top 1% income share excluding retained earnings from both the numerator and the denominator. Contrasting this scenario with the benchmark scenario reveals the cyclicality of retaining profits within the firm. (2) We distribute retained earnings lump-sum to all individuals receiving either dividends or business income from partnerships. (3) We distribute retained earnings exclusively to the top 1% (similar to the blue line in figure 11 in Alstadsæter et al. (2023)). More precisely, we assign a lump-sum amount to all individuals in the top 1% if they receive either dividends or business income from partnerships. Figure 1.11 compares the three alternative scenarios with our benchmark scenario.

[25] See Alstadsæter et al. (2023, figure 11).

We find that our benchmark scenario and the lumpsum distribution to the top 1% produce relatively similar results and higher income concentration than the scenario without retained earnings.[26] However, moving to the top of the income distribution increases the importance of capital income. For the top 0.01%, our benchmark scenario – distributing proportionally to distributed profits – produces the highest level of income concentration.

Figure 1.11.: Pretax national income shares varying the distributional assumptions for retained earnings



Note: Pretax national income shares based on RDC of the Federal Statistical Office and Statistical Offices of the Federal States, DOI: 10.21242/73111.(2007).00.00.2.1.0 & SOEP. For other distributional groups, see figures 1.17.

Uprating dividends (D42), business and rental income (B3n) to national accounts aggregates is not straightforward in Germany because national accounts and income taxation differently distinguish between corporate and non-corporate business income. Recall that, in contrast to labor income, there is a substantial micro-macro gap for D42 and B3n (see Figure 1.2). We already discussed this challenge of uprating these income components in subsection 1.2.3.3. In our benchmark scenario, we pool dividends, business income and rental income in our microdata base and uprate these incomes jointly to match the national accounts aggregate of D42+B3n (S14 + S15). In our robustness checks in Figure 1.18, we compare joint and separate uprating and find that the joint uprating tends to produce top 1% shares that are up to 2 percentage points lower than when uprating all three income components (dividends, business income, rental income) separately.

1.5. Comparison with previous series

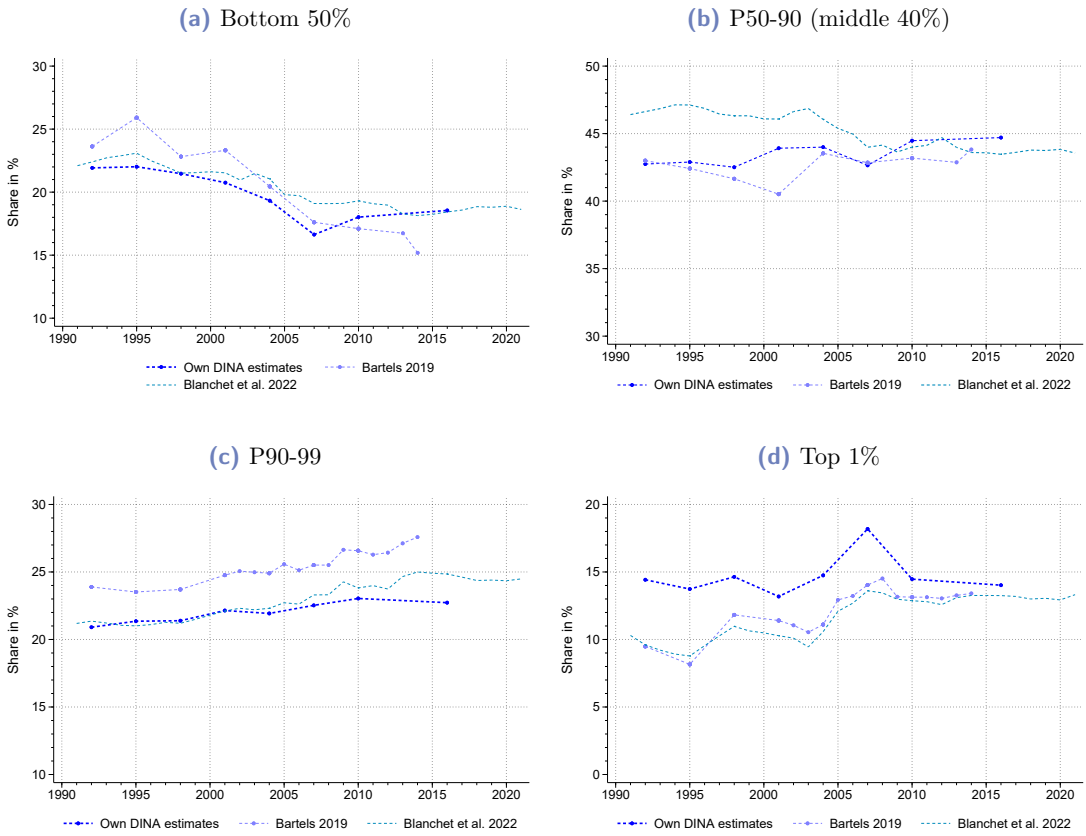
Previous long-run income share series for Germany have been estimated using income tax tabulations and/or survey data applying a harmonized methodology.[27] For Germany, Bartels (2019) estimated fiscal income shares using tabulated income tax statistics and national accounts. Blanchet et al. (2022b) estimated pretax national income shares using EU-SILC survey data applying a top-correction based on Bartels (2019). We will refer to this series as simplified DINA shares. Figure 1.12 compares our new DINA series with these previous series.

[26] The lumpsum distribution to the top 1% produces higher top 1% shares than the benchmark because it increases the share of the P99-P99.9 at the expense of the top 0.1%.

[27] These series are collectively available at the World Inequality Database.

All three series have in common that they show an increasing gap between the bottom 50% and the top 10%. Still, our newly estimated series shows two visible deviations from previous series. First, all studies find the decline of the bottom 50% income share, but Bartels (2019) and Blanchet et al. (2022b) document a continuous decline of the bottom 50% income share, while the new DINA series suggests a trend reversal after 2007. Second, the DINA top 1% income share is higher than the fiscal top 1% income share by Bartels (2019) and the simplified DINA share by Blanchet et al. (2022b), but this gap closes over time because the DINA series exhibits an overall stable pattern while Bartels (2019) and Blanchet et al. (2022b) suggest an increase in income concentration from a lower level.

Figure 1.12.: Pretax national income shares in comparison with previous series



Note: Series from Bartels (2019) based on fiscal income per tax unit from official PIT tabulations. Series from Blanchet et al. (2022b) based on EU-SILC survey data top-corrected with estimates from Bartels (2019) and uprated to pretax national income.

Several potential reasons for the gap between these series exist:

Observation unit is the individual (equal-split) in our DINA series, but it is the tax unit in Bartels (2019). The share of married couples in total potential tax units has declined from 41% in 1992 to 36% in 2016. Does the increase in fiscal top income shares arise from a tendency that top income earners are still more likely to be married than bottom income earners?

Income concepts differ between the series. For example, the fiscal series is based on distributed income, thereby excluding retained earnings. As retained earnings have gained importance over time, the fiscal series' total income became relatively smaller than net national income. The

DINA income total equals by definition the net national income aggregate. The fiscal income total amounts to 69% in 1992 and 64% in 2016 in net national income. For a given sum of top 1% income, a larger income total generates relatively lower top 1% fiscal shares in the 1990s and relatively higher top 1% fiscal shares since 2004, after which retained earnings have gained importance.

1.6. Conclusion and outlook

In this paper, we have combined individual income tax returns, SOEP household survey data and national accounts to estimate Distributional National Accounts (DINA) for Germany from 1992 to 2016, following the methodology established by Piketty et al. (2018) for the United States. Because country-specific income data landscapes and economic structures require country-specific adjustments and modifications, we discuss, adjust and modify the DINA standards to the fine-grained micro and macro data of Germany.

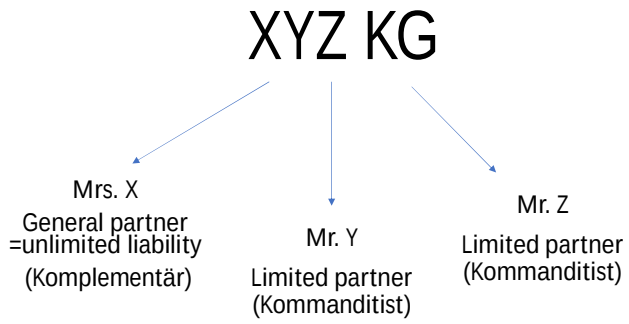
Our DINA series show that economic growth has been pro-rich from 1992 to 2007 and pro-poor from 2007 to 2016. But although incomes of the bottom 50% have resumed to grow since 2007, the income gap between the bottom 50% and the top 10% has widened between 1992 and 2016. The ratio of top 10% to bottom 50% average incomes has increased from eight to ten. This polarization between the bottom and the top is driven by continuously increasing labor income of the lower 9% of the top decile (P90-99).

In Germany, top incomes flow to a highly concentrated elite. Germany's top 0.1% and 0.01% income shares are similar to the United States and far above France. Germany's top business income recipients primarily hold firms as partnerships predominantly owned by two to four shareholders, while top business income earners in the United States and France hold shares in corporations. This country-specific business structure might contribute to more concentrated business incomes at the very top. Nonetheless, the top 0.1% and 0.01% income earners have seen relatively small income gains in relative terms (although large in absolute terms). For example, real average income of Germany's top 0.1% increased from EUR 2.2 million in 1992 to EUR 2.5 million in 2016.

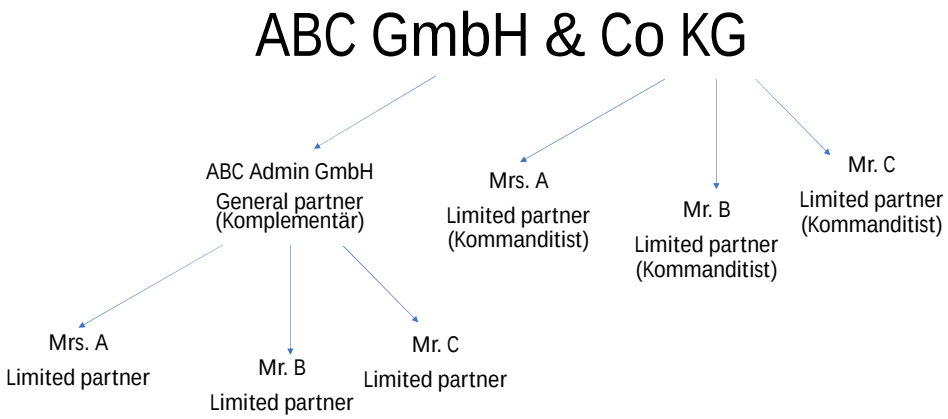
1.7. Appendix figures & tables

Figure 1.13.: Ownership and liability structure of Gmbh & Co KG

(a) Limited partnership: *Kommanditgesellschaft - KG*



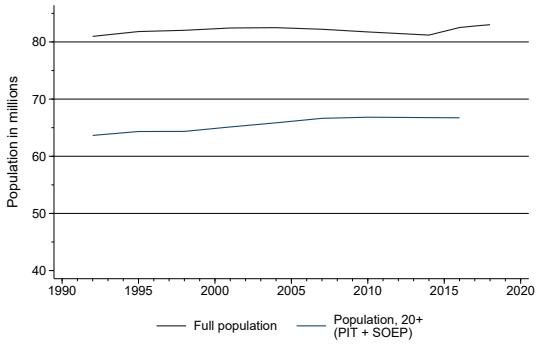
(b) Limited personal liability of all partners: *GmbH & Co KG*



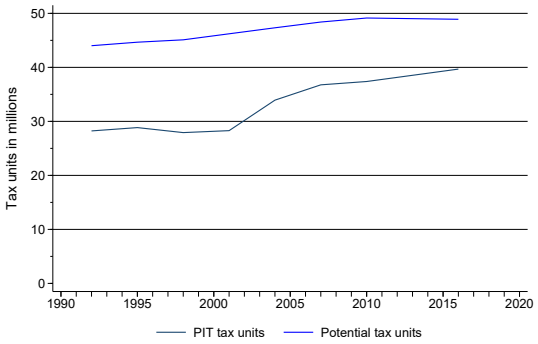
Note: The limited partnership (*Kommanditgesellschaft - KG*) is a partnership of partners with limited personal liability and at least one general partner with unlimited personal liability. The *GmbH & Co KG* is a partnership where the general partner is a previously established limited liability corporation (*Gesellschaft mit beschränkter Haftung - GmbH*) (in our example: ABC Admin GmbH) with a minimum capital requirement of EUR 25,000 and which is usually held by the same owners (in our example: Mrs. A, Mr. B and Mr. C). The *GmbH & Co KG* is a partnership, i.e., a pass-through entity, like the *KG* so that profits are considered as non-corporate business income which is subject to personal income tax. However, if the ABC Admin GmbH operates the ABC GmbH & Co KG, then only 60% of the dividends of the ABC Admin GmbH are taxable (*Teileinkünfteverfahren*).

Figure 1.14.: Population and income coverage in microdata (PIT+SOEP)

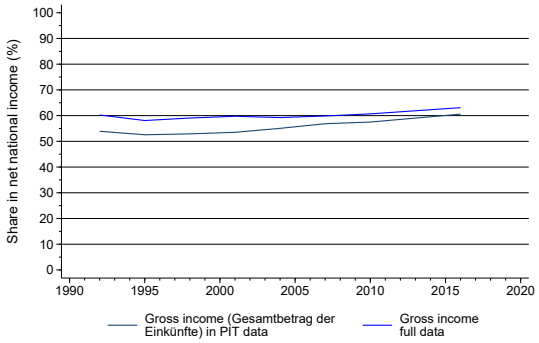
(a) Population



(b) Tax units

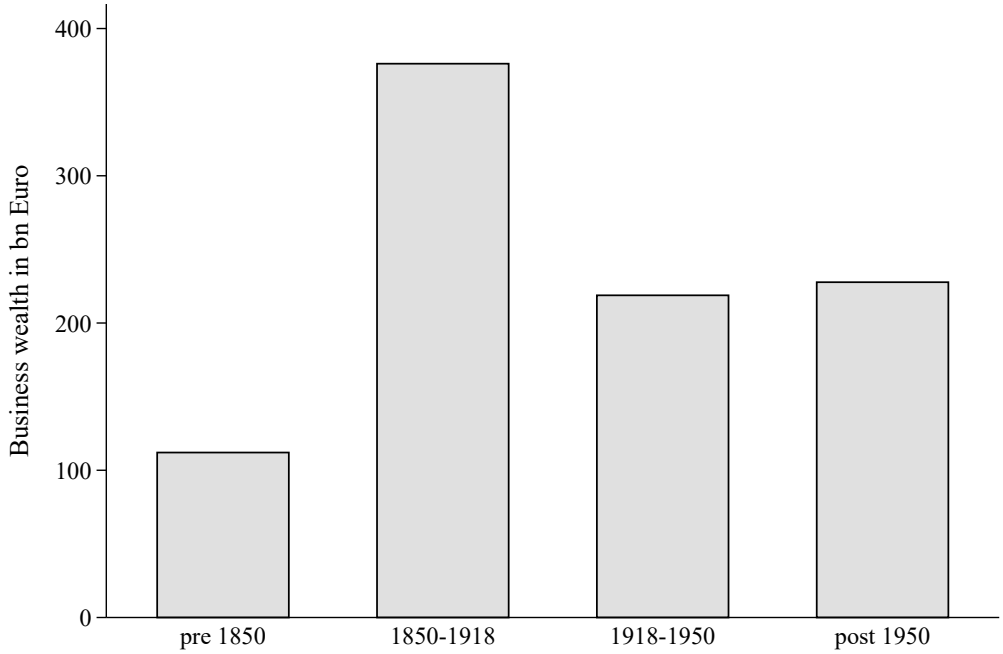


(c) Gross income



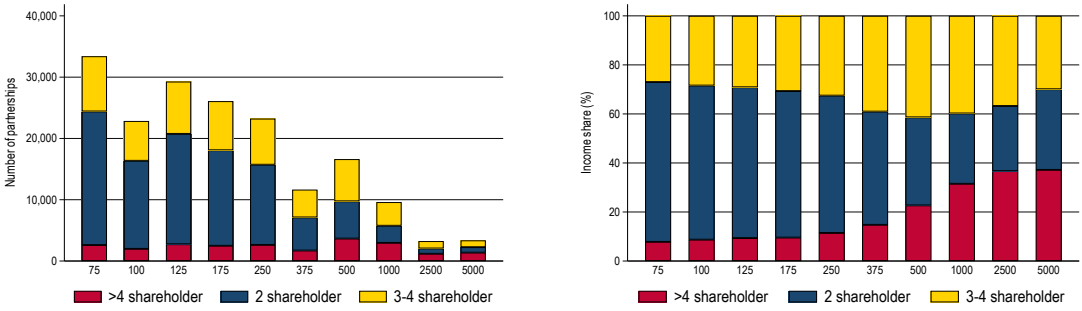
Note: Graph (a) shows the population 20+ and the total population in Germany. Graph (b) contrasts potential tax units (singles+married couples) with tax units recorded in PIT data. Between 2001 and 2007, payroll taxpayers (who do not have to file a tax return) are continuously added to PIT data so that the number of PIT tax units increases. Graph (c) depicts gross income in PIT and in our final microdata (PIT+SOEP) as share of net national income. Different to gross market income, net national income also comprises income components such as retained earnings and indirect taxes.

Figure 1.15.: Business wealth by founding year of the family business



Source: MM-list 2018, additional information on founding years of the family businesses is kindly provided by Andreas Bornefeld (co-author of the MM-List 2017 and 2018).

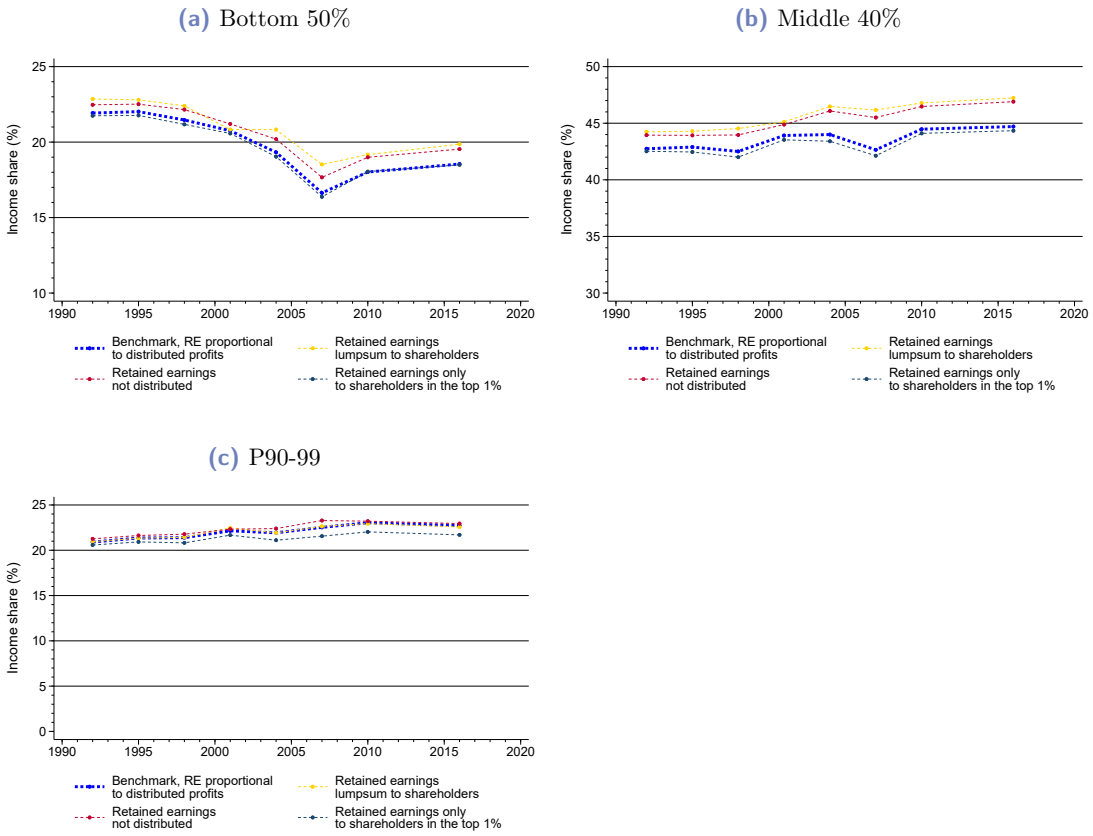
Figure 1.16.: Number of shareholders and income share by partnership income group



(a) Number of shareholders (b) Income share

Note: The figure displays the number of shareholders and the income share by income group of the respective partnership type in 2007. The income group is marked by the lower income thresholds in EUR 1,000. For example, panel (a) shows that 2,793 partnerships earning between EUR 1 and 2.5 mio. are held by only two shareholders. Panel (b) shows that the partnerships earning between EUR 1 and 2.5 million held by only two shareholders contribute almost 30% of total profits of this income group.
Source: Statistisches Bundesamt (2012) Lohn- und Einkommensteuer – Statistik über die Personengesellschaften/Gemeinschaften 2007.

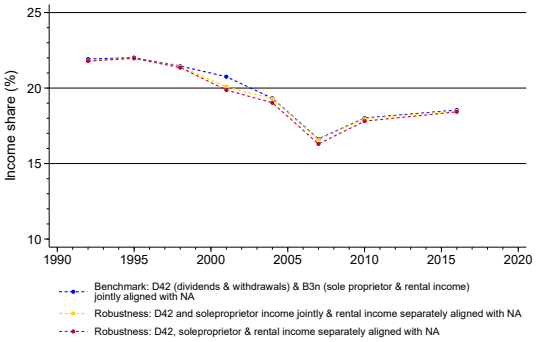
Figure 1.17.: Pretax national income shares varying the distributional assumptions for retained earnings



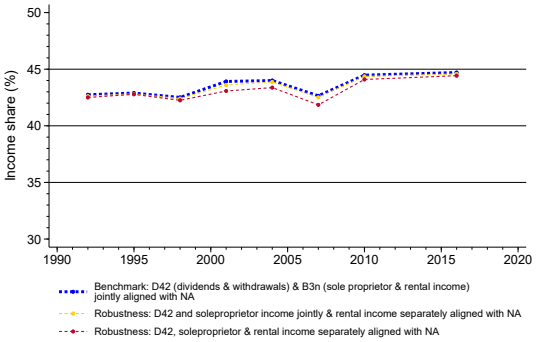
Note: Pretax national income shares based on RDC of the Federal Statistical Office and Statistical Offices of the Federal States, DOI: 10.21242/73111.(2007).00.00.2.1.0 & SOEP.

Figure 1.18.: Pretax national income shares, uprating robustness checks

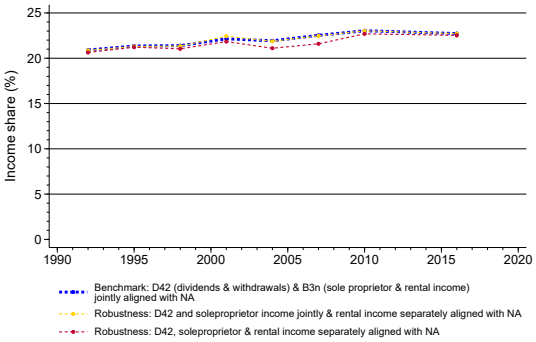
(a) Bottom 50%



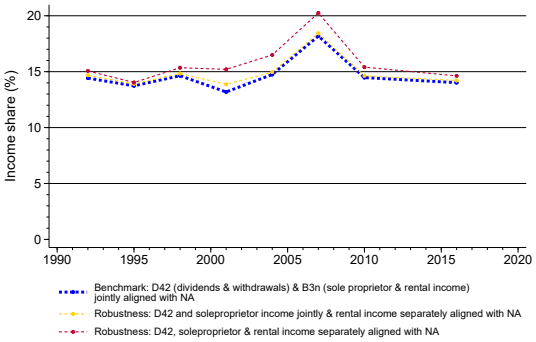
(b) P50-90 (middle 40%)



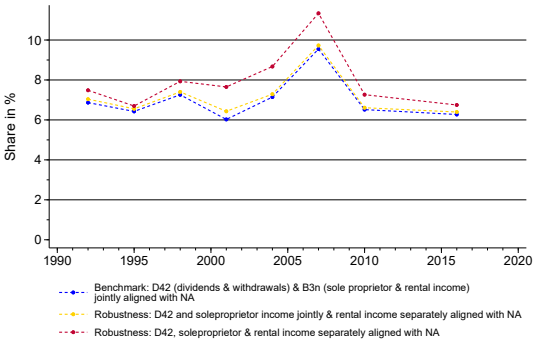
(c) P90-99



(d) Top 1%



(e) Top 0.1%



(f) Top 0.01%

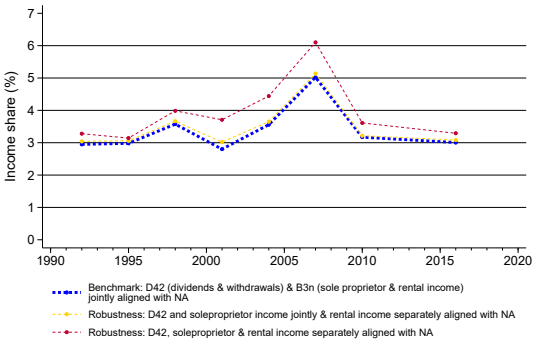
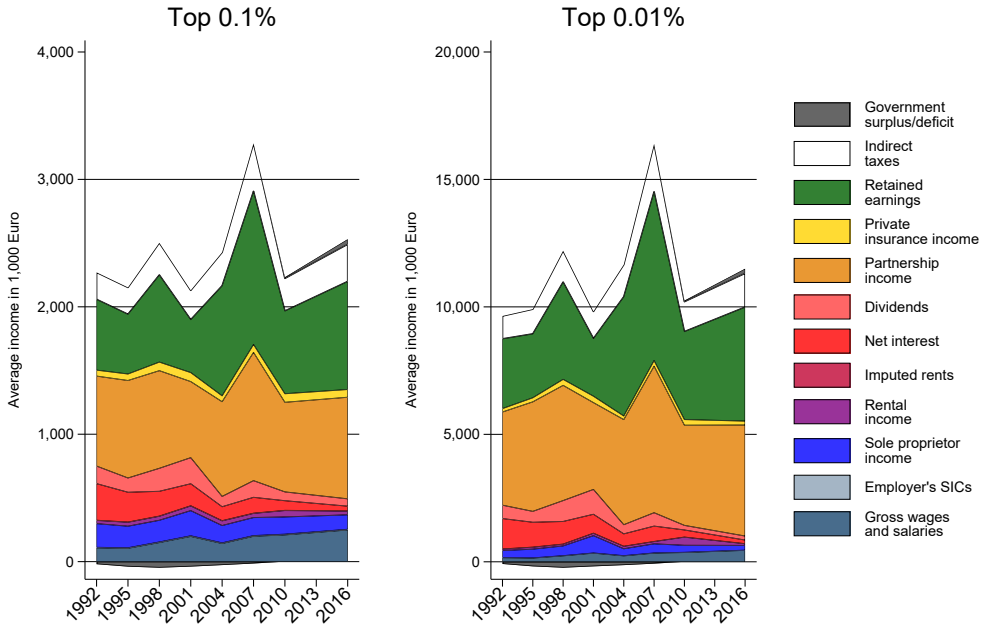
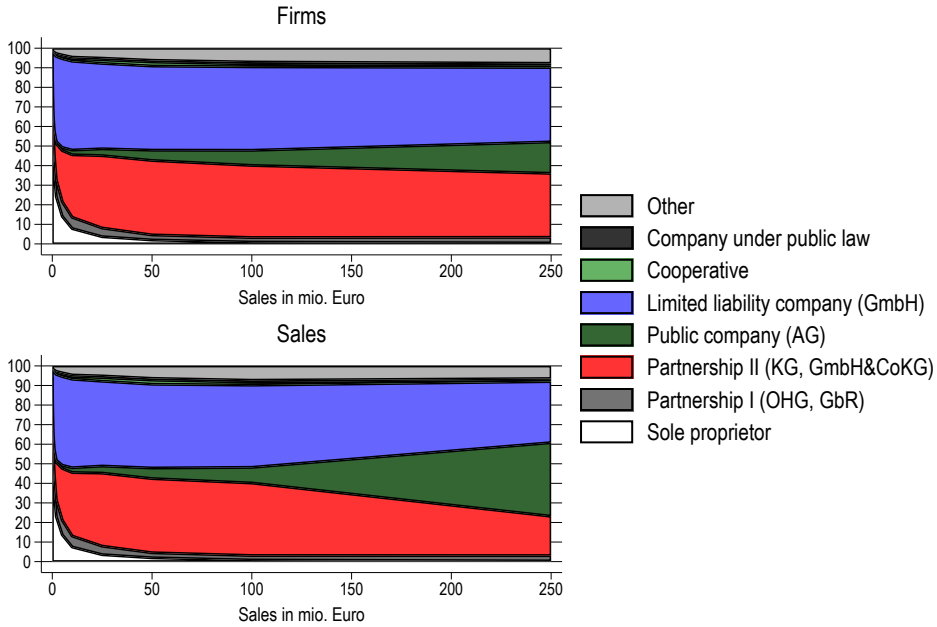


Figure 1.19.: Pretax factor income decomposition, Top 0.1% and 0.01%



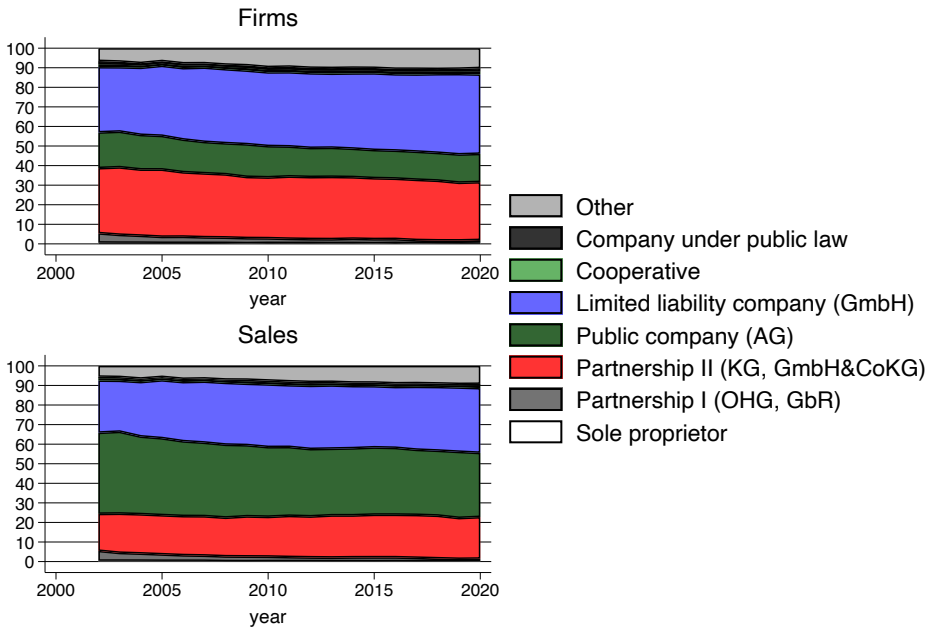
Note: Pretax factor income in EUR 1,000 (2015 prices). Pretax factor income comprises all market incomes before any operation of the tax-transfer system. Excluding the population aged 65+ that largely has no market income.

Figure 1.20.: Distribution of legal form



Note: Firms with sales exceeding EUR 17,500 per year (approx. 3 million firms in Germany) are subject to sales tax (*Umsatzsteuer*). The upper graph shows the share of firms by firm type across the sales distribution. The lower graph shows the share in total sales across the sales distribution.
Source: Statistisches Bundesamt (2009)

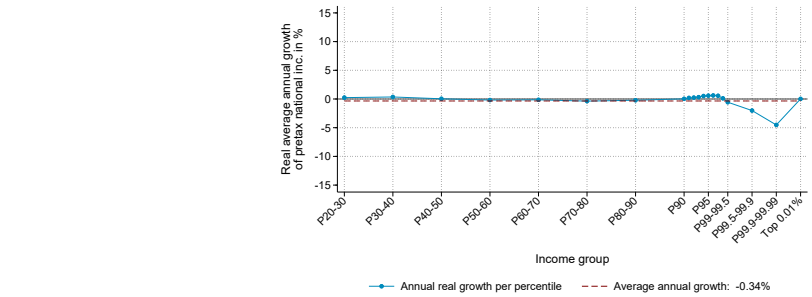
Figure 1.21.: Legal form of top firms, 2002-2020



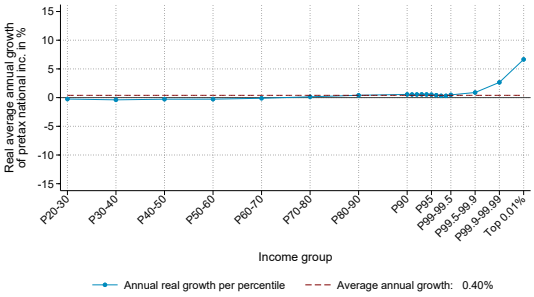
Note: Firms with sales exceeding EUR 17,500 per year (approx. 3 million firms in Germany) are subject to sales tax (*Umsatzsteuer*). The figure depicts the legal form of firms with sales exceeding EUR 250 mio.. The upmost graph shows the number of firm type by sales bracket and the lower graph shows the share in total sales.
Source: Statistisches Bundesamt (2009).

Figure 1.22.: Growth incidence 1992-2016

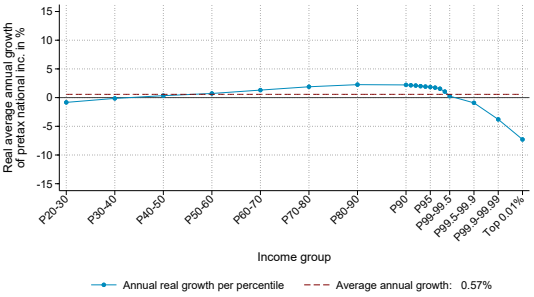
(a) 1992-1995



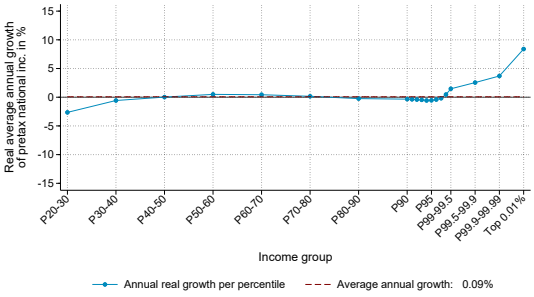
(b) 1995-1998



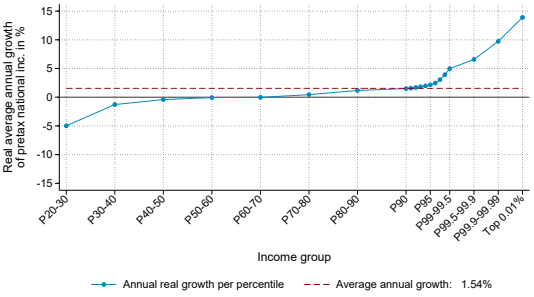
(c) 1998-2001



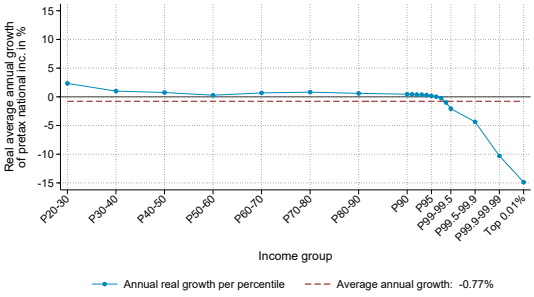
(d) 2001-2004



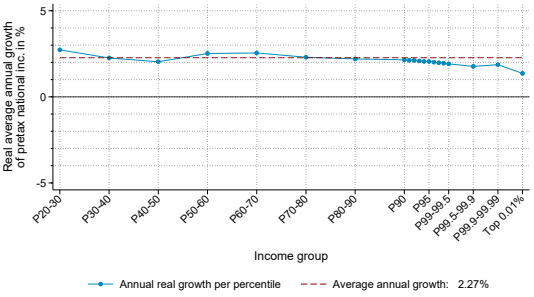
(e) 2004-2007



(f) 2007-2010



(g) 2010-2016



2 When Capitalism Takes over Socialism: The Lasting Economic Divide between East and West Germany*

Coauthored with Charlotte Bartels (DIW Berlin)

Abstract

In this paper, we investigate the economic divide between East and West German residents along the regional income and wealth distributions employing the Distributional National Accounts (DINA) method which aligns microdata with internationally standardized national accounts. We find that East German residents still earn and own a fraction of their West German counterparts. The gap expands towards the top of the distribution due to East Germans' lower business income and wealth. Investigating the causes, we find that reunification boosted West German top incomes. We link this to the privatization process of the 1990s in which predominantly West German investors acquired formerly state-owned East German capital. Further, persistent characteristics of businesses owned by either East or West Germans such as the number of employees, the complexity of the firm structure and legal form explain a rising share of the productivity gap which might impede the convergence in personal business incomes between East and West Germans.

JEL Classification: D3, E01, H2 H5, J3

2.1. Introduction

Thirty years after German reunification, a sizable economic divide between East and West German residents persists, contrary to the promises of swift convergence of living standards made in the early 1990s. East German wages are still about 25% lower than West German wages (Brüll and Gathmann, 2020). Average wealth of East German households is still less than 50% of the West German average (Albers et al., 2022).[1]

* The authors would like to thank Thomas Blanchet, Giacomo Corneo, Stefan Hauf, Wojciech Kopczuk, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman for helpful comments. Furthermore, we thank conference and seminar participants at the ECINEQ 2023, NBER's Conference on Research in Income and Wealth 2020, IARIW 2021, the World Inequality conference 2021, the workshop of the social policy committee of the Verein für Socialpolitik 2022, and the seminars of the German council of economic experts, at Paris School of Economics and at the Stockholm Institute of Transition Economics. Lotte Maaßen provided excellent research support. We gratefully acknowledge financial support from the Hans-Böckler-Foundation, Düsseldorf.

[1] Other persistent differences between East and West Germany have been documented, for example, for financial literacy (Bucher-Koenen and Lamla-Dietrich, 2018), preferences for redistribution (Alesina and Fuchs-Schündeln, 2007), egalitarian sex-role attitudes (Bauernschuster and Rainer, 2012), solidarity behavior (Brosig-Koch et al., 2011), social trust (Heineck and Süssmuth, 2013), and inflation expectations (Goldfayn-Frank and Wohlfart, 2019).

Countries of the former Eastern Bloc adopted different privatization strategies ranging from voucher privatizations, manager-buyouts or auctions to foreign investors (Sutela, 1998; Ther, 2016; World Bank, 1996). The East German privatization process was unique in its centralized organization via the "Treuhandanstalt", which was headed by West German managers and politicians, and its swift privatization of the majority of the East German capital within only three years (Priewe, 1993). Substantial tax reliefs on real estate and business investments – mostly directed at West German top income earners and firms – fostered investment flows going to East Germany in the 1990s.[2] In effect, a large share of the former GDR's state capital was acquired by West-German investors, i.e., the citizens from the former neighbouring capitalist state that simultaneously transferred its institutions to the former GDR.[3] Only in recent years a critical debate about the unachieved convergence and the adequacy of economic policies accompanying the German reunification has evolved.

This paper, for the first time, estimates the evolution of the entire income and wealth distribution in East and West Germany since reunification. In a second step, we study the causes behind the lasting economic divide and investigate how the process of German reunification contributed to cementing economic differences between East and West Germany. While several papers have investigated the *labor* income gap between East and West Germany (Biewen, 2001; Brüll and Gathmann, 2020; Fuchs-Schündeln et al., 2010; Hunt, 2001, 2002; Orlowski and Riphahn, 2009), we extend the analysis to *capital* income and ownership, which has received little attention so far. The German reunification provides a unique case study on the long-run distributional consequences of large-scale privatization of formerly state-owned land, housing and firms.

We first study the evolution of economic differences between East and West German residents along the regional income and wealth distributions from 1992 to 2016 and 2018, respectively. To achieve this, we employ the Distributional National Accounts (DINA) method which aligns micro data with internationally standardized national accounts. Income distribution results are our own calculations by region extending our project on Distributional National Accounts for Germany (Bach et al., 2023) following the methodology established by Piketty et al. (2018). Wealth distribution results are our own calculations by region based on Albers et al. (2022).

We find that East German residents still earn and own a fraction of their West German counterparts with this gap expanding towards the top of the income and wealth distribution. For example, West German top 1% income earners earned, on average, about EUR 650,000 in 2016, while the East German top 1% income earners earned approx. EUR 320,000, which is mostly due to East Germans receiving much lower incomes from partnerships and corporations.

To analyse the causes of the lasting economic divide, we proceed in two steps. First, employing the synthetic control method by Abadie and Gardeazabal (2003), we estimate if and by how much reunification has disproportionately increased top 1% incomes in West Germany. We find that reunification substantially increased income concentration in West Germany. This suggests that West German investors increased their capital income by acquiring formerly state-owned East German capital.

Second, we study differences in characteristics of business establishments *owned* by either East or West Germans employing the reweighting approach developed by DiNardo, Fortin, and Lemieux (1996, DFL) using the IAB Establishment Panel. Although the productivity gap between

[2] The economic literature on German reunification has highlighted these large investment flows and fiscal transfers from West to East Germany. See, e.g,. Dornbusch et al., 1992; Von Hagen et al., 2002; Burda and Hunt, 2001; Snower and Merkl, 2006.

[3] Restitution to former owners and sales to non-German investors were marginal (Mergele et al., 2020; Priewe, 1993).

businesses *located* in East vs. West Germany has been discussed and studied since reunification, the literature is far from conclusive. Three explanations for the East-West productivity gap appear worth mentioning: smaller establishments, lower concentration of managers (related to fewer headquarter) and cheaper products depress productivity in East Germany relative to West Germany (Burda and Severgnini, 2018; Mertens and Mueller, 2022). However, the previous analyses have focused on establishments or firms *located* in East Germany, while location and ownership can substantially diverge due to investments across regions, particularly given the nature of the privatization process of formerly state-owned East German capital in the 1990s. We find that – on top of establishment size – legal form and firm structure explain an increasing share of 25% in 2001 and almost 60% in 2019 of the productivity gap. East German owned establishments are more likely to be sole proprietorships and single entities instead of partnerships or corporations and part of a multi-establishment firm structure. This paper is organized as follows. Section 2.2 gives an overview of the privatization process following German reunification in 1990. Section 2.3 presents our results on the economic divide between East and West German residents along the regional income and wealth distributions. Section 2.4 investigates possible drivers behind the lasting economic divide, namely the process of reunification and business establishment characteristics. Section 2.5 concludes.

2.2. The economic reunification of Germany

On May 18 1990, the treaty of the monetary, economic and social union (MESU) of the Federal Republic of Germany (FRG) and the German Democratic Republic (GDR) was signed. East German fiscal and monetary sovereignty was transferred to West Germany and the economic order of the FRG was transplanted to the GDR. As Collier and Siebert (1991) note, reunification meant "merging a large open economy, relatively well-endowed with capital and technology, with a smaller, semi-autarkic economy, relatively well-endowed in labor and land."

For the former GDR, reunification kicked-off a "dramatic process of de-industrialization" (Von Hagen et al., 2002, p. 13). Industrial production fell by two-thirds as the capital stock was largely judged obsolete and production techniques outmoded (Burda and Hunt, 2001; Priewe, 1993). Employment declined by one-third between 1989 and 1992 (Burda and Hunt, 2001, p. 1). Those who kept their jobs benefited from an unprecedented wage hike achieved through negotiation by West German labor unions that aimed at reaching parity between East and West German wage levels by 1994 (Burda and Hunt, 2001, p. 4).

Reunification also meant large-scale privatization of formerly state-owned firms, housing and land. Privatization procedures varied greatly across the former Eastern Bloc countries, from voucher privatizations to manager-buyouts to the auctioning of big companies to national and international investors (Sutela, 1998; Ther, 2016). While, for example, Poland and Hungary pursued quick and early privatization and reforms with a high share of sales to foreign investors, the Czech Republic and Slovenia – the countries with the lowest income inequality in Eastern Europe today (Morgan and Neef, 2020) – postponed radical reforms, privatized slowly and under strict government control, liberalized foreign trade in several stages so firms could adjust, and regulated the housing market heavily (Brown et al., 2006; Ther, 2016; World Bank, 1996).

The East-German privatization process was unique, first, in its centralization via a state-owned trust agency, second, in its rapidity, and, third, in its low share of capital ownership transferred to the citizens of the former GDR. The last GDR government created a state-owned trust agency (*Treuhandanstalt*) to preserve and manage the national capital stock. This agency was tasked

with the privatization, restructuring and closure of formerly state-owned businesses. Already in March 1990, even before the economic and monetary union of East and West Germany, the *Treuhandanstalt*, that from late 1990 was mainly led by West German personnel, "became the owner of 126 former centrally-managed combines and of 95 regionally-managed combines, including [more than] 8,000 firms with about 45,000 plants [4] [and] of an estate of 62,000 km^2" (Priewe, 1993, p. 337) or about 57% of the total GDR territory. By late 1992, 83% of these firms were already privatized or closed (Priewe, 1993).

The overriding majority of firms was sold to West German investors and companies (Dornbusch et al., 1992; Mergele et al., 2020; Windolf, 1996), while the share of restitution to former owners or the share of sales to foreign investors remained low (Priewe, 1993). Investments in East German real estate and businesses were fostered by substantial tax reliefs – mostly directed at West German top income earners. With hindsight, "firms with higher baseline productivity [were] more likely to be privatized, [...], more often acquired by West German investors, and more likely to remain in business even [after] 20 years" (Mergele et al., 2020). The largest West German investments went to manufacturing, construction and the service sector in East Germany. "[M]ost firms were sold to enterprises that operate in the same or similar industries." (Dornbusch et al., 1992). Dornbusch et al. (1992, p. 244) highlight "the immediate and strong infusion of market skills and state-of-the-art technology at the level of the firm" from the sale to outside investors. At the same time, transfer of ownership and control to Western enterprises further increased the concentration of means of production ownership in Germany. Sinn and Sinn (1994) summarize the economic reunification process as follows: "Property rights worth mentioning have not been assigned to East Germans, but unrealistically high wages have been promised – a combination well designed to prevent investment and to maximize unemployment."

But still, the switch to a market economy induced a start-up boom in East Germany. During the 1990s, the self-employment rate in East Germany grew rapidly and reached the West German level in 2004. Yet, these new East German firms were on average smaller (IWH, 2010) and less successful when compared to their West German counterparts (Brixy and Grotz, 2004; Fritsch, 2004). A relatively high share of the newly emerging businesses in East Germany was in industries such as retailing, hospitality and catering, which are characterized by low entry barriers in terms of financial resources and required qualifications (Fritsch et al., 2014).

In sum, the economic reunification process in Germany in the 1990s opened up unique large-scale investment opportunities for West German top earners to acquire land, housing and firms in East Germany. Much of the formerly state-owned firms and real estate were acquired by West German investors. On the other hand, these investments also infused market skills and technology that might have boosted the growth of East German owned firms.

2.3. The economic divide between East and West Germany

2.3.1. Data, concepts and methods

In this section, we describe the data, concepts and methods to measure the income and wealth divide between East and West German residents. For both income and wealth distributions, we build on series that are consistent with macroeconomic aggregates. The unit of observation is the individual aged 20 and above.

[4] As a result, about 41% of the total GDR work force (about 4.1 million employees), were working in *Treuhandanstalt* firms in mid-1990 (Priewe, 1993, p. 337).

For the income distribution, we estimate pre-tax income distributions for reunified Germany and separately for East and West Germany excluding Berlin. We build on the procedures developed for the Distributional National Accounts (DINA) project for Germany (Bach et al., 2023).[5] In that project, we combined personal income tax files (PIT) since 1992[6], SOEP survey data and the sectoral accounts published by the national statistical office to construct an income distribution that is consistent with and sums up to net national income.

We use two concepts of pretax income: 1) pretax factor income which comprises all gross market incomes, namely wages and salaries gross of employers' and employees' social insurance contributions, self-employment and business incomes including retained profits, dividends and interest, rental incomes as well as imputed rents of homeowners. Pensioners often exhibit zero or low market incomes. When using pretax factor income concept we restrict our analysis to the population of 20 to 64 years of age. 2) Pretax national income adds insurance-based replacement incomes such as old-age pensions, unemployment and sickness benefits (*Arbeitslosengeld I*, *Krankengeld*) to pretax factor income and subtracts social security contributions. We equally split income between married couples so that our series are individual incomes.

For the wealth distribution, we build on Albers et al. (2022) who constructed the wealth distribution series for Germany, 1895-2018. They produced a household wealth series based on a combination of household surveys, macroeconomic aggregates and rich lists.[7]

To analyze the income and wealth distribution in East and West Germany, we first rank all residents in region r (East or West Germany), N^r, according to their income or wealth, respectively, to categorize them into four groups: bottom 50%, middle 40% (P50-90), P90-99 and top 1%. Then, we analyze average income and wealth by group, region and asset type. For example, the average income from asset j of residents belonging to fractile p of the income distribution of region r can be written as

$$y_j^{p,r} = \frac{1}{N^{p,r}} \sum_{i=1}^{N^{p,r}} \alpha_{j,i} \cdot K_j \cdot r_j \tag{1}$$

where K_j denotes the total stock of asset j owned by German residents and $\alpha_{j,i}$ is the share owned by resident i. Note that the sum of shares across regions and fractiles is equal to one, i.e., $\sum_{i=1}^{N} \alpha_{j,i} = 1$. Note that the DINA method distributes net national income so that K_j encompasses the national capital stock owned by households, the corporate, non-profit and government sector alike.[8] r_j is the rate of return on asset j. We follow Saez and Zucman (2016) assuming that the rate of return only differs across asset types, but neither by wealth level nor by region.[9]

[5] The Distributional National Accounts methodology has been introduced by Piketty et al. (2018) and Blanchet et al. (2021) and has been applied to several countries by now. Bach et al. (2023) construct Distributional National Accounts (DINA) for Germany, 1992-2016, and provide a detailed discussion of the method, country-specific adjustments, specific micro-macro alignment steps, and underlying assumptions in the main paper and in the accompanying Data Appendix.

[6] Current results are based on the 10% sample with full coverage of top incomes.

[7] The estimation of the wealth distribution by capitalizing the DINA incomes following the capitalization method established by Saez and Zucman (2016) is in the process, but official results are probably only publishable by the end of this year.

[8] Government surplus/deficit is a separate DINA income category so that we view this income type as stemming from a separate asset type j.

[9] Fagereng et al. (2020) provide evidence that returns positively correlate with wealth using Norwegian tax records. It is debated if and how to estimate heterogeneous returns. Saez and Zucman (2016) argue in favor of homogeneous returns. Smith et al. (2022) introduce return heterogeneity to the estimation of capitalized wealth based on income tax records in the United States and find that accounting for this

Rearranging Eq. 1 gives the average wealth in asset j of residents i of region r (East or West Germany) belonging to fractile p of the wealth distribution in region r:

$$w_j^{p,r} = \frac{y_j^{p,r}}{r_j} = \frac{1}{N^{p,r}} \sum_{i=1}^{N^{p,r}} \alpha_{j,i} \cdot K_j \tag{2}$$

The corresponding ownership rate s_j^r of a given asset j in region r can be formulated as

$$s_j^r = \frac{1}{N^r} \sum_{i=1}^{N^r} \mathbb{1}[\alpha_{j,i} > 0] \tag{3}$$

To understand the economic divide between East and West German residents, we are interested in two factors: 1) the ownership rate s_j^r (extensive margin) and the level of capital $w_j^{p,r}$ (intensive margin) owned by the respective populations of region r. In other words, East German residents might receive less capital income, on average, because they are less likely to own capital or because they own lower levels of the capital.

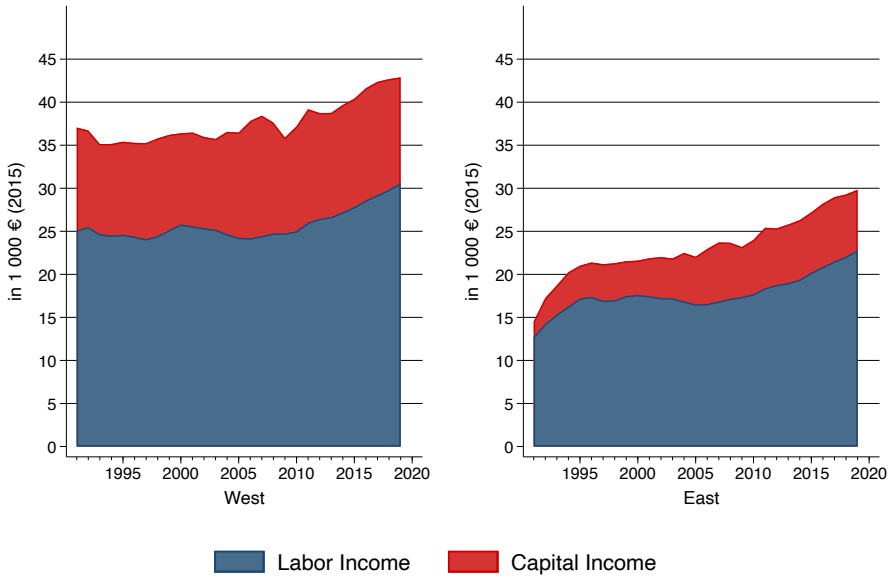
2.3.2. The income divide

We start by comparing the average labor and capital income of East and West German residents in Figure 2.1. Thirty years after reunification, net national income per capita in East Germany stands at about 70% of the West German level. Comparing these levels across federal states uncovers even larger gaps: National income per capita exceeds EUR 30,000 in the two southern states (Baden-Wuerttemberg and Bavaria) and in the independent city of Hamburg in the north. In all East German states, national income is below EUR 25.000 and in two East German states even below EUR 20.000 (Mecklenburg-Western Pomerania and Saxony-Anhalt) (see Appendix Figure 2.13).

Both lower capital income and lower labor income contribute to the persistent income gap between East and West German incomes. While the share of East German labor income has reached 75% of West German labor income in 2018, East German capital income is even lower in comparison reaching less than 60% of the West German level.

heterogeneity does not change the fundamental story for top wealth shares and their growth. We will explore return heterogeneity in a future version of this paper.

Figure 2.1.: Net national income per capita: capital and labor income



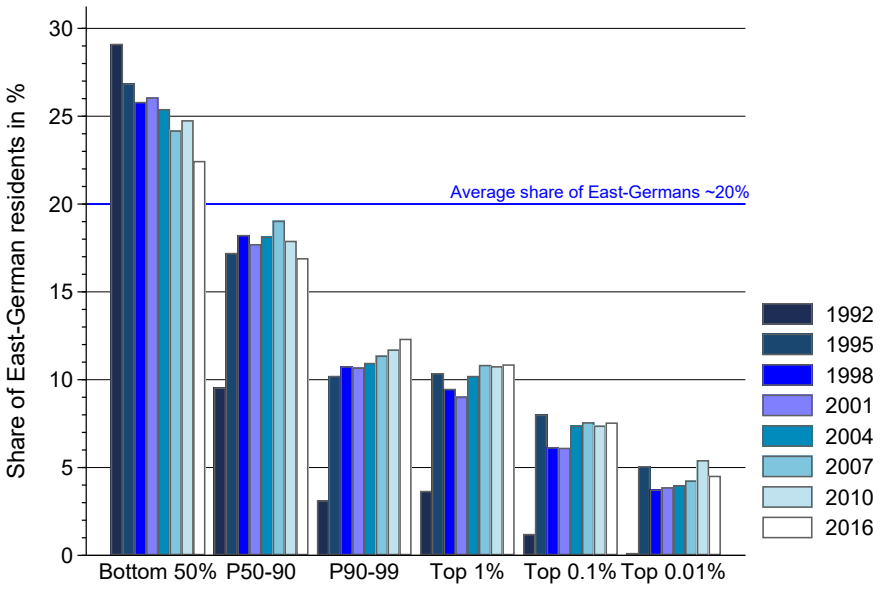
Note: Own calculations based on the national accounts of federal states (*VGR der Länder*), Statistische Ämter der Länder, excluding Berlin. In line with the DINA methodology, the average income is computed as net national income divided by the number of residents of at least 20 years of age.

To uncover the reasons for the capital income gap between East and West German residents, as a first step, we now disentangle DINA for Germany (Bach et al., 2023) by income group and region of residence. Figure 2.2 shows how East German residents sort into the reunified German pretax national income distribution. This picture gives a first impression in which fractiles of the national income distribution East Germans are over- or underrepresented. We see that East German residents are still slightly overrepresented in the bottom 50% of the national income distribution and slightly underrepresented in the middle 40% (P50-90). Their share is about 20% of the German adult population (highlighted by the blue horizontal line),[10] while their share is about 23% in the bottom 50% and 17% in the middle 40%. Moving further to the top of the distribution, East Germans are increasingly underrepresented. Their share declines from about 12% in the upper middle class (P90-99) to less than 5% in the top 0.01%. Strikingly, we see a continuous convergence pattern for the bottom 50%. However, this might be partially due to the migration of young, single and low-income East Germans, as described by Fuchs-Schündeln and Schündeln (2009), to West Germany who, to some degree, enter the bottom 50% in West Germany.[11] In contrast, the representation of East Germans among the top 50% is dominated by a one-time increase during the first five years after reunification and has stagnated since then.

[10] Note that our data restricted to the population of 20+ which results in an East German share of 19% in 1992 (12m of 63.8m), 21% in 1995 (13.8m of 64.4m), 21% in 2010 (13.9m of 66.9m) and 19% in 2016 (12.8m of 66.8m). Official population statistics show a lower share of East German residents because these statistics also include the population below the age of 20.

[11] In our integrated database of PIT and survey data uprated to national accounts, we find a drain of about 1.8m East German residents from the bottom 50% between 1992 and 2016. Using the SOEP, we find a net migration of East Germans to West Germany (net of West Germans moving East) of the same order of magnitude for the same period.

Figure 2.2.: East Germans in the German pretax national income distribution, 1992 - 2016

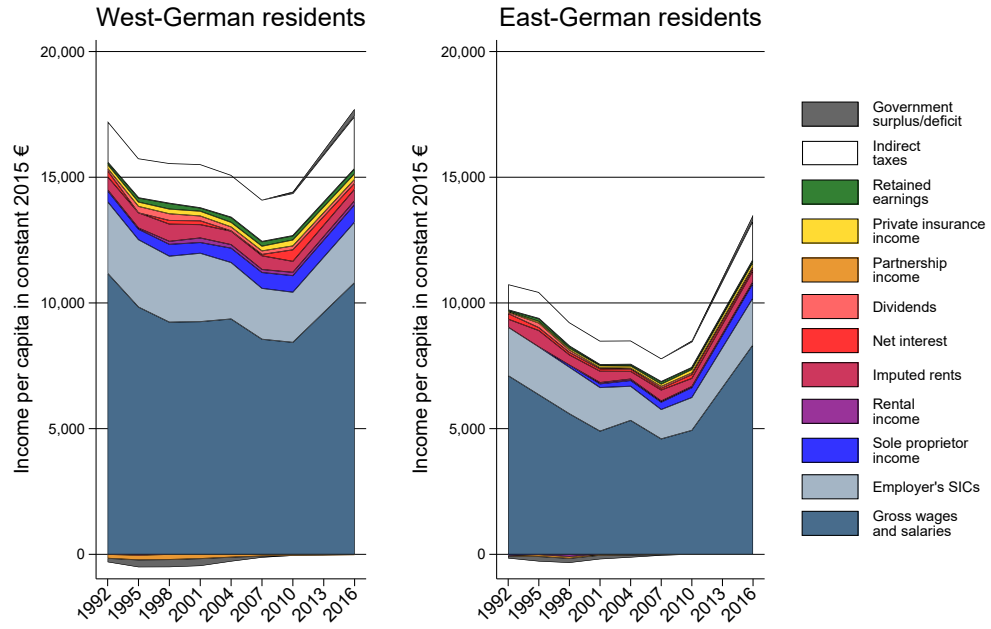


Note: Own calculations based on our integrated database of PIT and survey data uprated to national accounts.

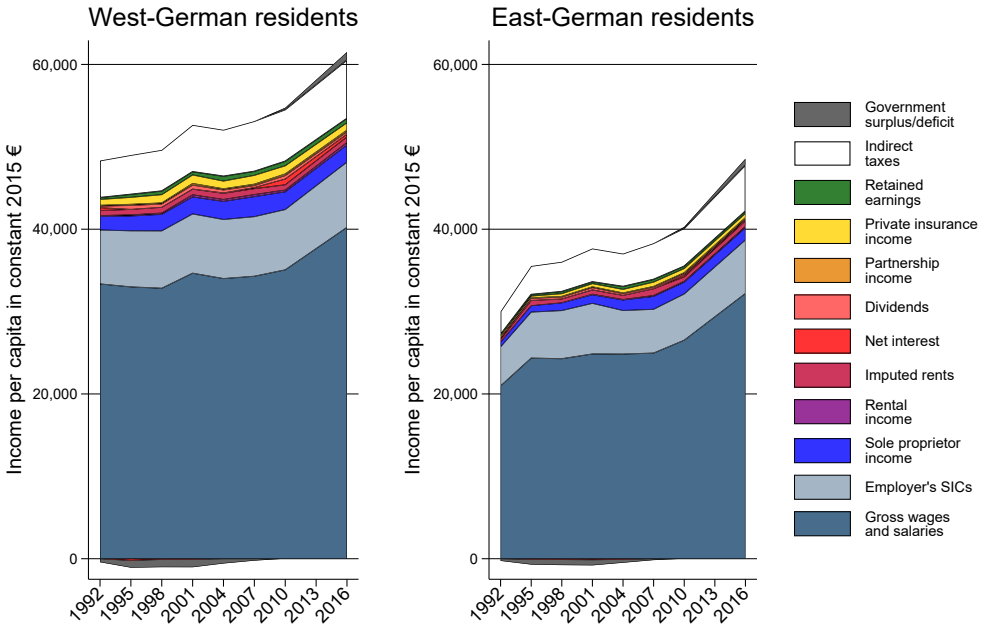
We now analyse separate regional distributions: Figures 2.3 and 2.4 show the composition of pretax factor income per capita in East and West Germany for four groups along the East and West German distributions of pretax factor national income. Income differences between East and West Germans in the bottom 90% are largely explained by labor income differences. The bottom 50% in West and East German have seen their real labor incomes decreased from the early 1990s until the mid-2000s and have seen a pronounced increase since. Moving further to the top of the income distribution, labor income differences become relatively less important and capital income differences emerge as the central source of the gap.

Figure 2.3.: Pretax factor income composition in West and East Germany by income group, Bottom 90%

(a) Bottom 50%



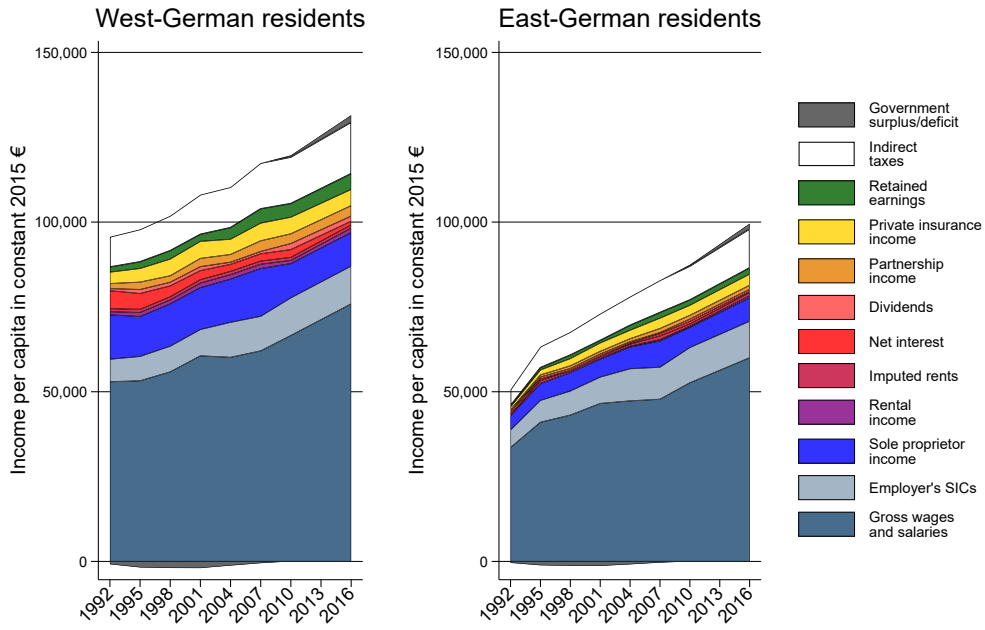
(b) P50-90



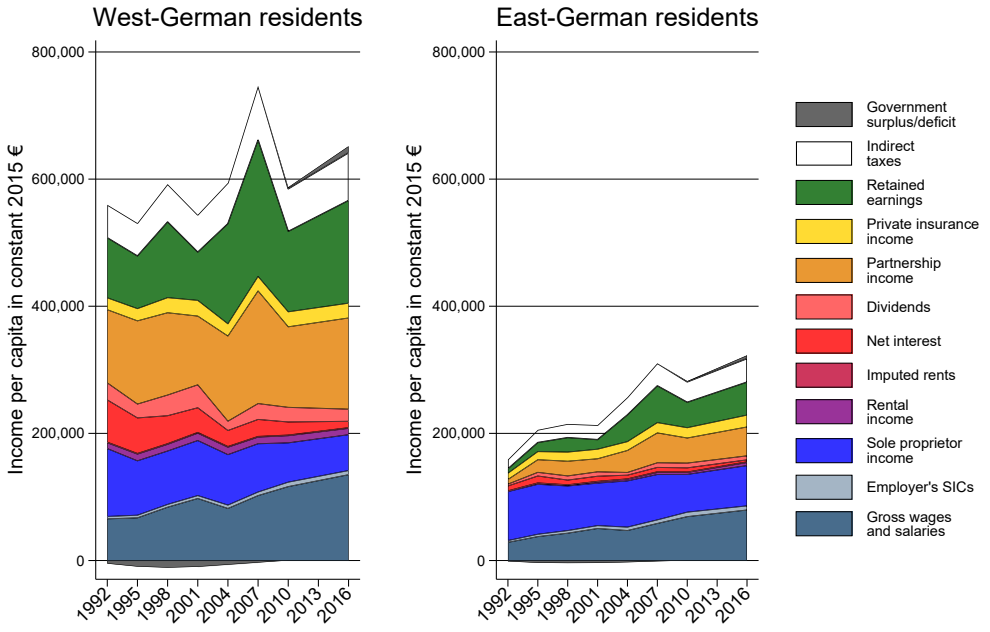
Note: Own calculations based on our integrated database of PIT and survey data uprated to national accounts. Sample population restricted to 20 to 64 years of age. Incomes in constant 2015 EUR.

Figure 2.4.: Pretax tax factor income composition in West and East Germany by income group, Top 10%

(a) P90-99



(b) Top 1%



Note: Own calculations based on our integrated database of PIT and survey data uprated to national accounts. Sample population restricted to 20 to 64 years of age. Incomes in constant 2015 EUR.

Figure 2.4 shows the income composition of the top decile broken down into the bottom 9% of the top decile and the top 1%. Average income of the top 1% was about EUR 650,000 in West Germany and EUR 320,000 in East Germany in 2016. The top percentile's average income from wages and self-employment is of almost comparable magnitude in East and West Germany, about EUR 200,000 for the West German and about EUR 150,000 for the East German top 1%. Capital income from corporate, quasi-corporate and non-corporate firms as well as interest income flowing to West German residents greatly exceeds East German levels. Partnership income including retained earnings amounts to about EUR 305,000 or 47% of the average pretax income of the West German top 1%, while it sums to approx. EUR 97,000 or about 30% of the pretax income of East German top 1% earners. Thus, capital income explains the difference in top incomes between East and West Germany.

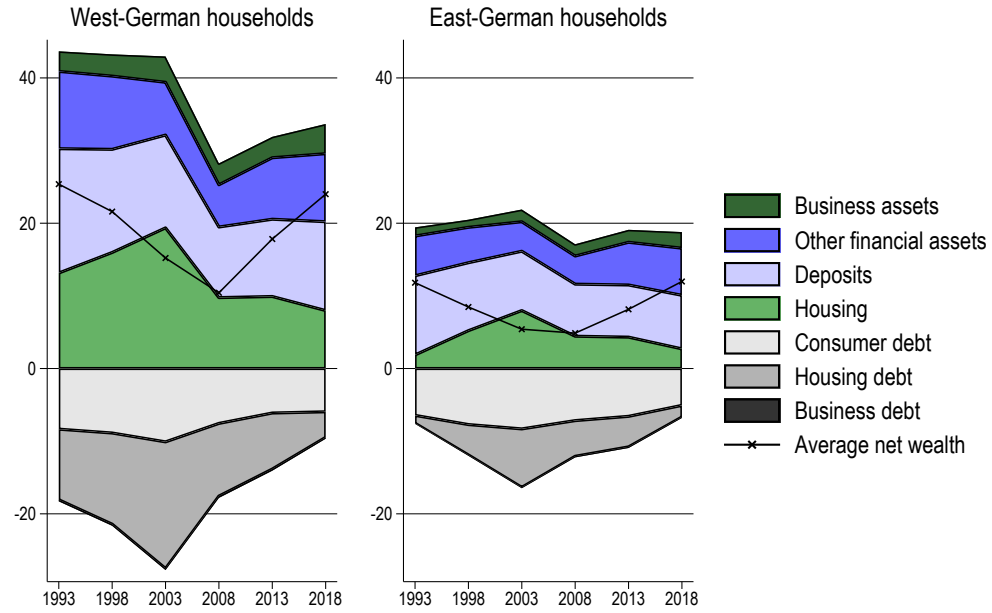
2.3.3. The wealth divide

After having identified the capital income gap as major factor for the persistent income differences between East and West German top earners, we turn to the source of those incomes: wealth. The estimation of the wealth distribution by capitalizing the DINA incomes following the capitalization method established by Saez and Zucman (2016) is in the process, but official results are probably only publishable by the end of this year. For the time being, we present and discuss the wealth results by region from Albers et al. (2022) in this subsection. Note that the unit of observation differs from the rest of this paper: Albers et al. (2022) analyze household wealth (not equal-split individual wealth). While equal-split does not generate any re-ranking of individuals compared to the household wealth distribution, average household wealth levels are higher than average equal-split levels.

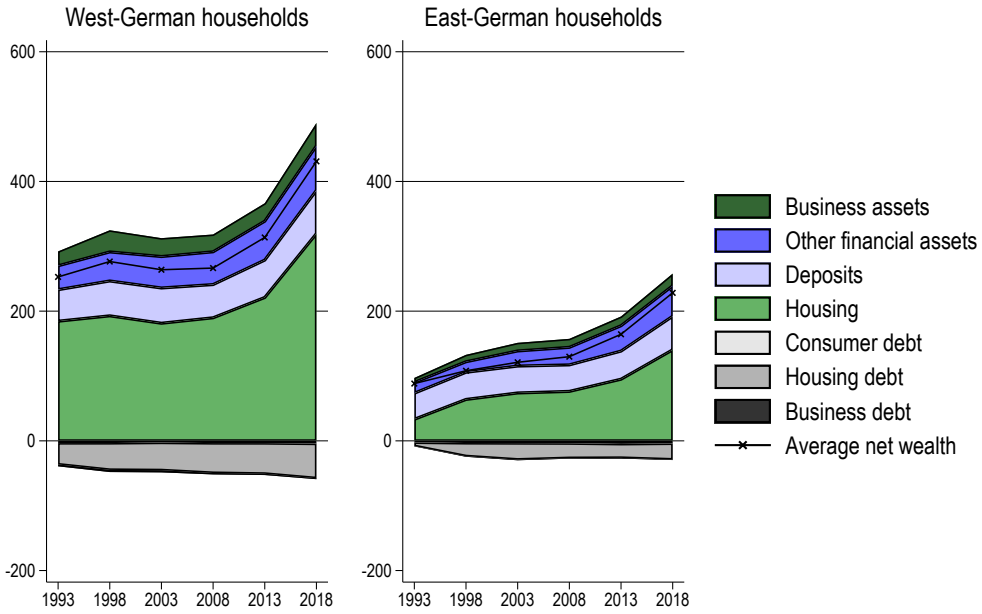
Figure 2.5 shows the average portfolio of West and East German households belonging to the bottom 90% of the wealth distribution in their respective region. Figure 2.6 shows the respective averages for the top 10%. East German households belonging to the poorer 90% are about half as rich as their West German counterparts. The wealth gap is particularly high at the very top of the distribution: In 2018, West German households belonging to the top 1% owned an average net wealth of more than EUR 12 million, while East German households of the same group own approx. EUR 3 million. While the West German affluent are richer across all asset types, they own higher levels of business and housing wealth in particular.

Figure 2.5.: Wealth composition in West and East Germany by wealth group, Bottom 90%

(a) Bottom 50%



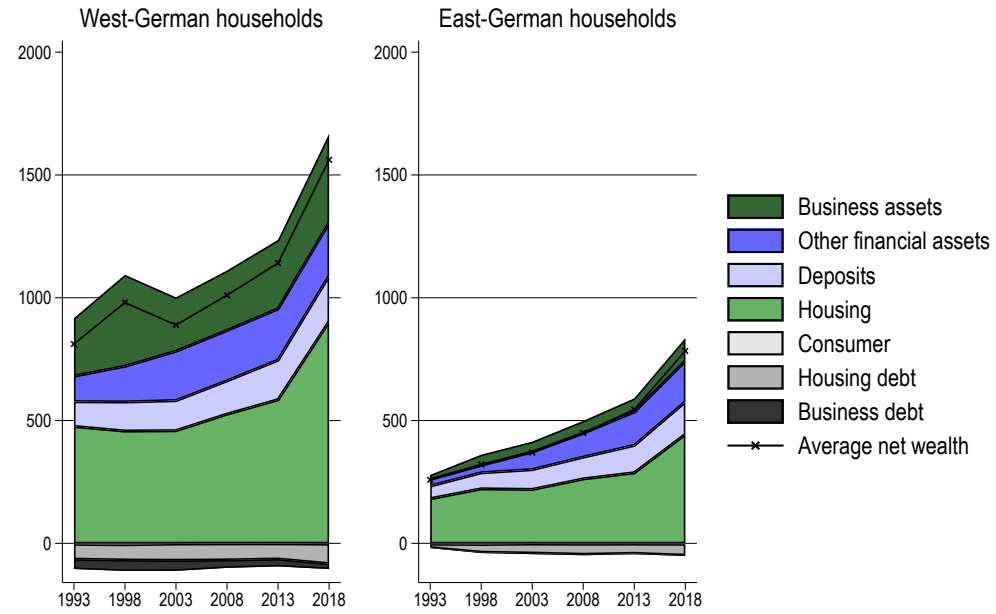
(b) P50-90



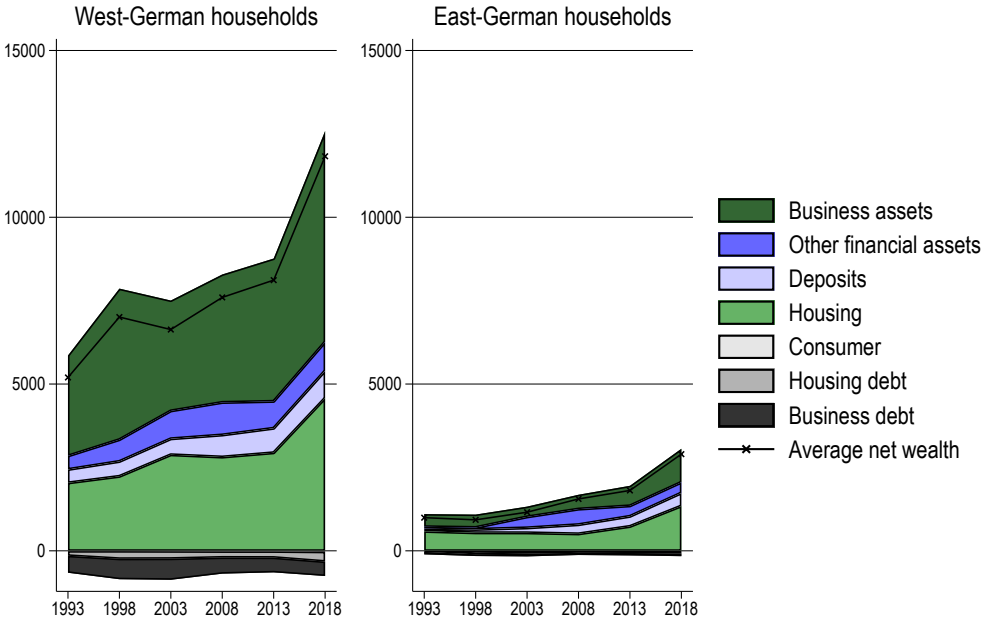
Note: Own illustration based on calculations by Albers et al. (2022) in EUR 1,000 and in prices of 2015.

Figure 2.6.: Wealth composition in West and East Germany by wealth group, Top 10%

(a) P90-99



(b) Top 1%



Note: Own illustration based on calculations by Albers et al. (2022) in EUR 1,000 and in prices of 2015.

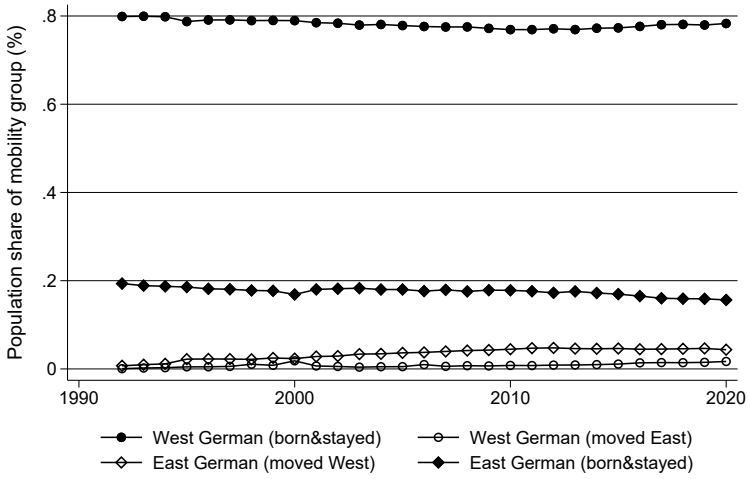
2.4. Explaining the lasting economic divide

2.4.1. Capital ownership

The two previous sections have shown that East German residents hold less wealth and receive less capital income than West German residents, on average. The gap widens when moving to the top of the German income or wealth distribution and is most visible for real estate and business capital. This section investigates to what extent differences in ownership ($s_{j,r}$ in Eq. 3) for these two asset types contribute to the economic divide. We base our analysis on SOEP data which allows us to distinguish four population groups: West German (born & stayed), moved to East Germany, moved to West Germany, East German (born & stayed). In accordance with the DINA series in Section 2.3.2, income and wealth is equally split between household members 20+.

Figure 2.7 shows that a negligible part of the West German born population, which makes up over 80% of the entire German population, moved to East Germany. On the other hand, about one-fifth (or about 2.7m) of the East German born population has migrated to West Germany since reunification. Young, single and college-educated individuals were more likely to migrate West as well as those from low-income counties (Fuchs-Schündeln and Schündeln, 2009).

Figure 2.7.: German population by residence and birth region



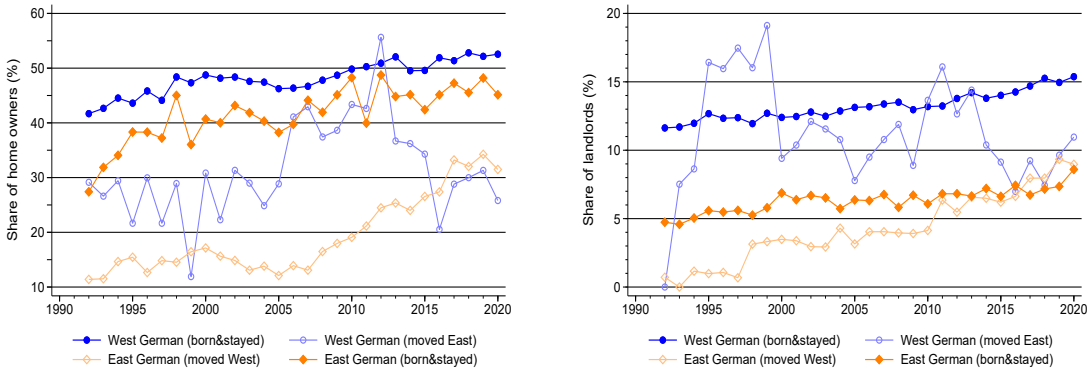
Note: SOEP v37. Group indicator based on the comparison between current region of residence and birth region.

Figure 2.8 presents the share of home owners (receiving imputed rent) (Figure 2.8a) and landlords (receiving rental income, Figure 2.8b) among the four population groups. The probability to own real estate is highest for West Germans who stayed in West Germany. In recent years, the home ownership rate exceeds 50%. However, East Germans have almost caught up. East Germans who stayed in East Germany raised their ownership rate from less than 30% in 1992 to almost 50% and East Germans who moved West increased their ownership rates from around 10% in the early 1990s to slightly above 30% in recent years.[12] Note that the group of West Germans moving East is small, particularly after excluding Berlin (to be consistent with the DINA analysis in Section 2.3.2), so that time trends are jumpy. The share of landlords (renting

[12] Different to residents of several other countries of the former East Bloc where it was common practice that residents acquired the apartment they were living in during transition as their property, this was not the case in East Germany.

out real estate) shows lasting differences between East and West Germans. West Germans who stayed in West Germany show the highest rates of around 15% in 2020, while slightly less than 10% of East Germans (both movers and stayers) are landlords. An increasing trend for these three groups, but no clear convergence is visible. But as for home ownership, a sizable catch-up is visible, particularly for East German movers, who increased their incidence of being landlords from nearly zero to almost 10% during the past three decades.

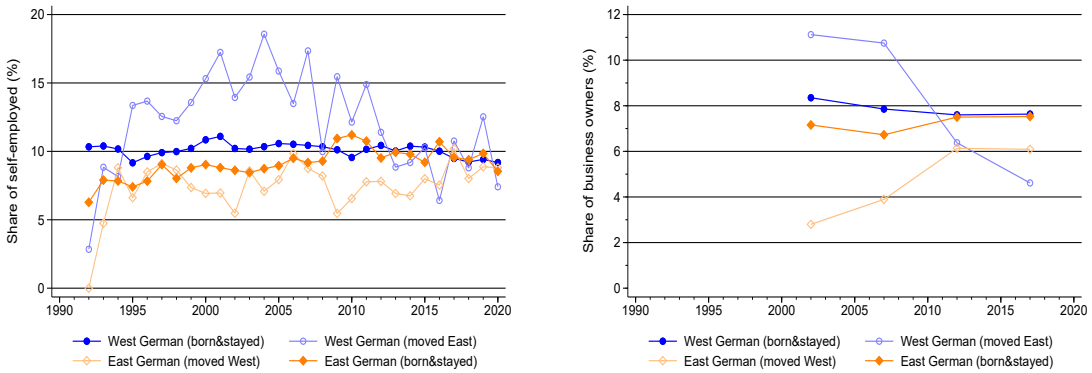
Figure 2.8.: Real estate ownership by residence and birth region



(a) Home ownership (imputed rent) (b) Rental income

Note: SOEP v37. Home ownership is defined as receiving imputed rent. Landlords are identified by receiving positive rental income.

Figure 2.9.: Business ownership by residence and birth region



(a) Self-employed income (b) Business owner

Note: SOEP v37. Self-employed is equal to one if earning income from self-employment. Business owner is equal to one if owning positive business wealth according to the wealth questionnaire asked in 2002, 2007, 2012 and 2017.

Figure 2.9 displays the share of self-employed (earning self-employment income) and business owners (owning positive business wealth). Both indicators show a convergence of East Germans (born & stayed) with West Germans (born & stayed) in the 1990s and early 2000s so that in recent years business ownership rates are similar. The probability of earning self-employment income is approx. 10% for both groups and stands at 7% for owing business assets. Ownership rates are slightly lower for East German movers, for whom rates have increased and reached 8%

and 6%, respectively. We do not comment on the share of West German movers, because the shares are jumpy given the small number of observations.

In sum, differing real estate ownership rates can explain some of the economic East-West divide. Lower land prices and rents in East Germany probably further add to the economic divide. In contrast, self-employment and business ownership rates are quite similar for East and West Germans, particularly if they stayed in the birth region. Thus, the ownership rate s_j^r does not help to explain the persistent business wealth and business income differences that we documented in Subsections 2.3.3 and 2.3.2. The vast differences in business wealth and incomes must then be due to differences in the firms' characteristics which we investigate in the following two subsections.

2.4.2. West German investors going East

As described in Section 2.2, the overriding majority of formerly state-owned East German firms was sold to West German investors and companies, often operating in the same or similar industries. Substantial tax reliefs aimed at West German top income earners spurred investment in East German firms, housing and land. This section aims to quantify the impact of reunification on West German top incomes.

Given the policy background, we would expect that tax reliefs, granting substantial write-offs for the acquisition of East German capital, reduced West German top incomes in the first years after reunification.[13] If those investments were on average profitable in the medium and long run, we would expect that the (fiscal) income concentration would increase in later years, compared to the counterfactual that Germany had not been reunified.

Employing the synthetic control method by Abadie and Gardeazabal (2003), we estimate if and by how much reunification has affected income concentration in West Germany which would not have taken place without the external shock of reunification. We construct a synthetic West Germany and its evolution of the top 1% fiscal income share, our outcome variable, absent of the reunification shock. This counterfactual West Germany is built by the weighted combination of unaffected, or so-called, donor countries. The respective weights of the donor countries are computed so that they minimize the difference between West Germany and its synthetic counterfactual in the pre-intervention period of 1980-1989 for the outcome and predictor variables. This ensures that the synthetic control can be used as a reliable predictor in the post-event period. Following Abadie et al. (2015) who analyzed the impact of reunification on West German GDP, we create a donor pool consisting of 15 OECD countries: Australia, Austria, Belgium, Denmark, France, Greece, Italy, Japan, the Netherlands, New Zealand, Portugal, Spain, Switzerland, the United Kingdom, and the United States.[14] Sources to all predictor variables are given in table 2.11.[15] We restrict our prediction period to 1990 to 2004.

Table 2.1 shows the non-zero weights assigned to each donor country. Austria, Japan and the

[13] Privatization of firms via the *Treuhandanstalt* started in mid-1990. The law with generous deduction possibilities (*"Gesetz über Sonderabschreibungen und Abzugsbeträge im Fördergebiet"*) came into effect in mid-1991.

[14] We select the same countries as Abadie et al. (2015), leaving out a number of countries because of their disruptive structural developments that might interfere with our analysis. In contrast to Abadie et al. (2015), we exclude Norway because a tax reform in 1992 creates a break in the top income series which might lead to "interpretational difficulties" as Aaberge and Atkinson (2010) point out.

[15] Since our outcome variable, the top 1% fiscal income share for Germany, is based on triennial data, we restrict our analysis to triennial data points since 1980.

Table 2.1.: Synthetic control weights for West Germany

Country	Synthetic control weight
Austria	.335
Belgium	.012
Japan	.298
Switzerland	.063
United States	.292

Note: This table shows only donors with non-zero weights.

United States exhibit by far the highest weights. Our synthetic West Germany will be the weighted average of these five countries' top share values.

Table 2.2 compares the pre-intervention averages of all predictor variables and of the top 1% income share between 1980 and 2004 for actual West Germany and its synthetic control. The synthetic control shows reasonably similar values to West Germany for the top 1% income share, GDP per capita and the industry share in value added. Consumer price index and investment rate and the trade openness measure show some divergence. As Abadie et al. (2015) point out, the Federal Republic had the lowest inflation rate among OECD countries in the 1980s, so this indicator cannot be perfectly produced with the donor pool.

Figure 2.10 shows the evolution of West Germany's top 1% income share compared to its synthetic control from 1980 to 2004. The counterfactual reproduces very closely the outcome variable in the pre-intervention years. The root mean square prediction error is 0.69 percentage points.

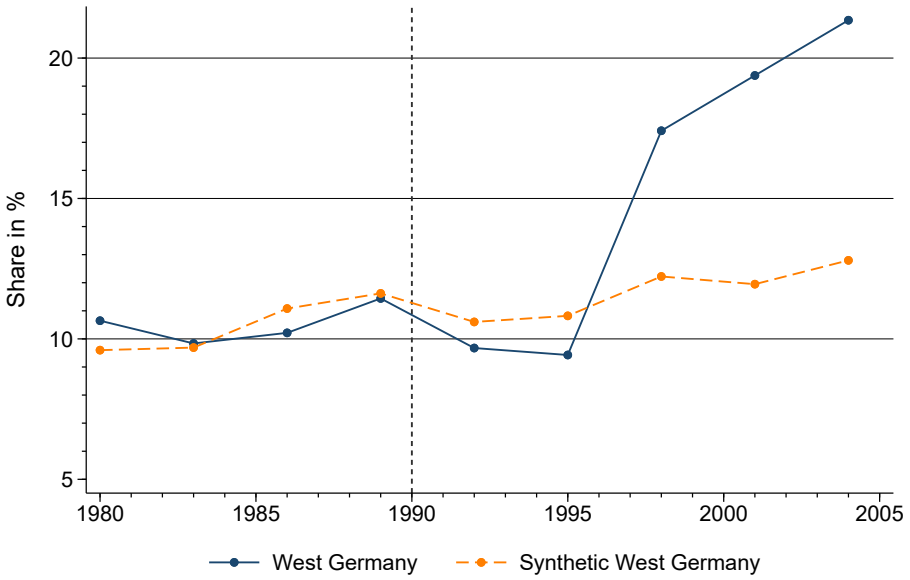
Table 2.2.: Top 1% income share predictor means before reunification

	Treated	Synthetic
Top 1% share	10.54	10.5
GDP per capita	15735.0	15367.5
Consumer Price Index	76.7	59.1
Trade openness	55.4	43.3
Investment rate	21.4	27.7
Life expectancy males	71.6	72.0
Industry share in value-added	34.9	35.3
Labor share in GDP	.668	.621

Note: Because top 1% income based in income tax data are only available in 1980, 1983, 1986, and 1989, all other variables are from these years as well. Sources are listed in table 2.11.

Between 1989 and 1995, the observed top 1% fiscal income share decreases and stands about two percentage points below the counterfactual level until 1995. This is in line with our previous thoughts on the effect of tax reliefs on taxable income. After 1995, the actual West German top income share doubles until 2004, while the synthetic control shows only a slight increase. This suggests that reunification and subsequent privatization policies increased income concentration in West Germany.

Figure 2.10.: Top 1% Share: West Germany vs synthetic West Germany



Note: Own calculations, sources in appendix, table 2.11.

To test for statistical validity, we perform four robustness checks: 1) an in-time placebo test, 2) an in-space placebo test, 3) we compute the ratio between post-and pre-treatment root mean square prediction error (RMSPE) and 4) a leave-one-out analysis.

For the in-time placebo test (1), we reassign the reunification shock to 1984, i.e., six years before the actual event.[16] Appendix Figure 2.16 shows that the placebo synthetic control follows reasonably closely the actual evolution of the West German top share not only until 1984, but until 1989, the last year before the treatment. The placebo synthetic control is slightly higher than our actual control. Nonetheless, similar to the benchmark counterfactual, the evolution of the top shares diverges substantially between the placebo synthetic control and the actual West German top share after 1990. This suggests that the divergence of top shares reflects the effect of reunification and not the lack of predictive power of our counterfactual.

The in-space placebo test (2) reassigns the reunification shock to countries in the donor pool which should be unaffected by the treatment. The treatment effect is deemed significant if the estimated effect for Germany is exceptionally large compared to the estimated effects of the donor pool countries. Panel (b) of Appendix Figure 2.17 shows that the difference between the actual and synthetic top 1% income share is exceptional for West Germany relative to other countries. The only other country with a similarly large deviation, but a very distinct and since the mid-1980s arising upward trend is the United States.

The ratio between the post-treatment RMSPE and the pre-treatment RMSPE (3) allows to evaluate the exceptionality of the treatment effect. The intuition behind this measure is that a large treatment effect is less indicative if the difference between actual and synthetic top 1% share is already large in the pre-treatment period (Uliczka, 2023). West Germany stands out with the highest post-RMSPE to pre-RMSPE ratio suggesting that the income concentration

[16] 1984 is the earliest possible year for the in-time placebo test because top income share series for several countries only start in 1980.

effect in West Germany is indeed exceptional and not plainly driven by a low prediction power of our estimation (see Appendix Figure 2.18).

The leave-one-out analysis (4) tests if our results are driven by one specific donor country by subsequently excluding countries with a positive weight from the donor pool (Abadie et al., 2015). Appendix Figure 2.19 shows that all restricted synthetic controls are reasonably similar to our benchmark counterfactual. When Austria, Switzerland or Japan are left out, the synthetic control increases by about one to two percentage points above the benchmark, but still shows a similar trend to our benchmark control. This suggests that no single country drives the synthetic control trend.

Overall, our synthetic control estimation suggests that West German top income earners benefited disproportionately from reunification. Against the backdrop of the reunification and privatization process described in Section 2.2, we tentatively interpret this abrupt increase in income concentration as indicative evidence for West German top income earners investing in formerly state-owned capital in East Germany which contributed to their increasing capital income in the medium and long run. To neatly pin down this mechanism, more microdata-based research is necessary.

2.4.3. Characteristics of East vs West German owned establishments

The previous analyses have shown that business ownership rates have converged in the 2000s (Section 2.4.1), yet average business income and business wealth of East German residents remain persistently lower than for West German residents (Sections 2.3.2 and 2.3.3). What explains the persistent business income and wealth gap?

We explore sales and productivity differences between establishments *owned* by either East or West Germans employing the reweighting approach developed by DiNardo, Fortin, and Lemieux (1996, DFL). The IAB Establishment Panel provides information on sales, wage sum, ownership (East German, West German or other majority ownership), legal form (sole proprietorship, partnership, corporation), number of employees, industry (manufacturing/service) and firm structure (independent entity or part of a multi-establishment firm). While sales, wages and workers are variables measured at the establishment level, legal form, ownership and industry are variables measured at the firm-level.

The goal of our analysis is to assess the extent to which business income differences between East and West German residents can be explained by differences in the distribution of establishment characteristics. In the following, we explain how we adapt the DFL method to our purposes. Let each establishment be characterized by a vector (y, z, c) comprising a continuous variable y (sales or productivity), a vector of attributes z (e.g., establishment size, single entity or part of a multi-establishment firm structure, legal form), and an ownership identifier c indicating either East or West German. The joint distribution of sales (productivity) and attributes of establishments owned by the population of c then is $F(y, z, c)$, while $F(y, z|c)$ denotes the joint distribution of sales and attributes conditional on ownership c. Following DiNardo et al. (1996), the density of sales (productivity) of establishments owned by the population of c, $f_c(y)$, can be written as

$$f_c(y) \equiv f(y; c_y = c, c_z = c). \tag{4}$$

The notation allows us to express the density of sales (productivity) y of establishments owned by one population subgroup conditional on the distribution of attributes z of establishments owned

by the other population subgroup. For example, while $f(y; c_y = \mathrm{W}, c_z = \mathrm{W})$ denotes the actual density of sales (productivity) of West German owned establishments (W), $f(y; c_y = \mathrm{W}, c_z = \mathrm{E})$ is the counterfactual density of sales (productivity) of West German owned establishments (W), applying the distribution of attributes of East German owned establishments (E). The aim of the DFL reweighting method is to estimate the counterfactual density, which (taking the example for E and W) is defined as

$$f(y; c_y = \mathrm{W}, c_z = \mathrm{E}) = \int f(y|z, c_y = \mathrm{W}) dF(z|c_z = \mathrm{E}) \tag{5}$$

$$= \int f(y|z, c_y = \mathrm{W}) \phi_z(z) dF(z|c_z = \mathrm{W}) \tag{6}$$

where $\phi_z(z)$ denotes the reweighting function.

$$\phi_z(z) = \frac{dF(z|c_z = \mathrm{E})}{dF(z|c_z = \mathrm{W})} = \frac{Pr(c_z = \mathrm{E}|z)}{Pr(c_z = \mathrm{W}|z)} \cdot \frac{Pr(c_z = \mathrm{W})}{Pr(c_z = \mathrm{E})}. \tag{7}$$

The probability of observing an establishment owned by subpopulation c, given establishment attributes z, can be estimated with a probit model:

$$Pr(c_z = c|z) = Pr(\epsilon > -\beta' H(z)) = 1 - \Phi(-\beta' H(z)). \tag{8}$$

where $\Phi(\cdot)$ is the cumulative normal distribution and $H(z)$ is a vector of covariates.

We use the cross-sectional sample of the dataset comprising between 13,900 (2000) and 15,400 observations (2019). Data exists from 1993 to 2019 for West Germany and since 1996 for East Germany. We restrict our analysis to the period from 2000 to 2019 due to data irregularities in earlier years. We exclude Berlin and only keep establishments with either East- or West-German ownership, excluding establishments of public, foreign or unclear ownership which restricts the sample to between 8,600 (2000) and 9,900 observations (2019). Our selected sample reveals a rather stable evolution of sales compared to the unrestricted sample in the early 2000s (see Appendix Figure 2.15).

Sales is the most suitable proxy for business income in the IAB Establishment Panel.[17] We only include observations that indicate sales.[18] For the reweighting by different variables, we keep individual companies (*Einzelunternehmen*), partnerships (like *KG*, *OHG*, *GmbH & Co. KG*), and corporations (like *GmbH* and *AG*) and exclude other marginal legal forms (see Appendix Table 2.6 for variable documentation).

Size has been emphasized in previous studies as a crucial explanatory factor for lower productivity of firms or establishments *located* in East Germany. As a first step, we check if we find a similar important role of size when considering ownership rather than location. We distinguish between four establishment size groups based on the number of employees.[19] Table 2.3 shows that East German owned establishments are smaller than their West German owned counterparts. Both West and East German owned establishments have increased in size between 2000 and 2019.

We first reweight West German owned establishments to match the size distribution of East German owned establishments. Figure 2.11 shows a substantial and persisting sales gap between

[17] Ideally we would like to analyze profit per shareholder so that we could link our findings to the previous analysis at the personal income level (Figure 2.4).

[18] The dataset provides a non-harmonized variable of business volume that predominantly indicates sales, but in several cases also the total budget or the aggregate premium.

[19] Sample sizes are given in Appendix Table 2.7.

Table 2.3.: Distribution of establishment sizes by ownership

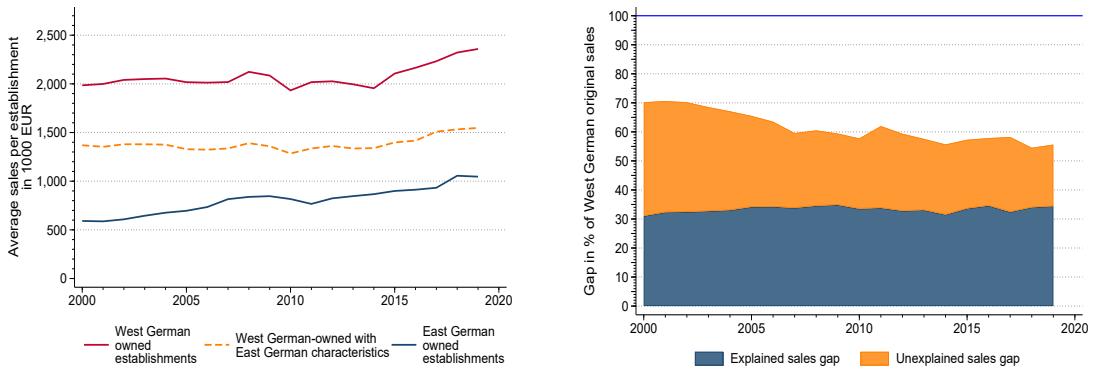
Size	2000 West-owned	2000 East-owned	2019 West-owned	2019 East-owned
1-4 employees	63.4%	68.2%	61.5%	65.9%
5-9 employees	18.6%	18.3%	18.2%	18.4%
10-99 employees	16.5%	13.2%	18.7%	15.0%
100+ employees	1.4%	0.4%	1.6%	0.7%
No. of establ.	1,351,632	303,070	1,340,015	252,342

Note: Own calculations based on IAB Establishment Panel.

East and West German owned establishments. Figure 2.11a) shows the absolute gap and Figure 2.11b) shows the relative gap. In 2000, average sales of West German owned establishments were more than three times those of East German owned establishments (EUR 2m vs. EUR 600k). Until 2019, this gap has declined to a factor of about two. The dashed line shows that average sales of reweighted West German owned establishments would have been EUR 1.5m in 2019 if they had been as small as East German owned establishments.

Figure 2.11b shows that the relative sales gap declined from 70% to about 55% of the West German owned sales level. However, the gap that is explained by the fact that East German residents own smaller establishments (blue area) remained stable at about 30%. The closing of the observed sales gap is thus due to characteristics other than size (orange area).

Figure 2.11.: Sales gap between East and West German owned establishments



(a) Absolute gap (b) Relative gap

Note: Own calculations based on the IAB Establishment Panel. DFL reweighted by establishment size.

To further understand the gap that cannot be explained by size of the establishments, we now ask how differences in legal form and firm structure affect the productivity gap.[20] Figure 2.4 has shown that East German residents earn, on average, similar sole proprietor incomes compared to West German residents but lack incomes from partnerships and corporations. The establishment panel data confirms that East German owned establishments are more often held

[20] Similar to other studies in the literature (Mertens and Mueller, 2022), we interpret sales per employee as a proxy for productivity.

Table 2.4.: Distribution of legal form by ownership

	West German owned		East German owned	
Year	Sole proprietorships	Partnerships & corporations	Sole proprietorships	Partnerships & corporations
2000	59.3%	40.7%	70.4%	29.6%
2005	59.3%	40.7%	70.8%	29.2%
2010	58.6%	41.4%	67.7%	32.3%
2015	57.2%	42.8%	68.2%	31.8%
2019	53.9%	46.1%	64.3%	35.7%

Note: Own calculations based on IAB Establishment Panel. Absolute values in table 2.8.

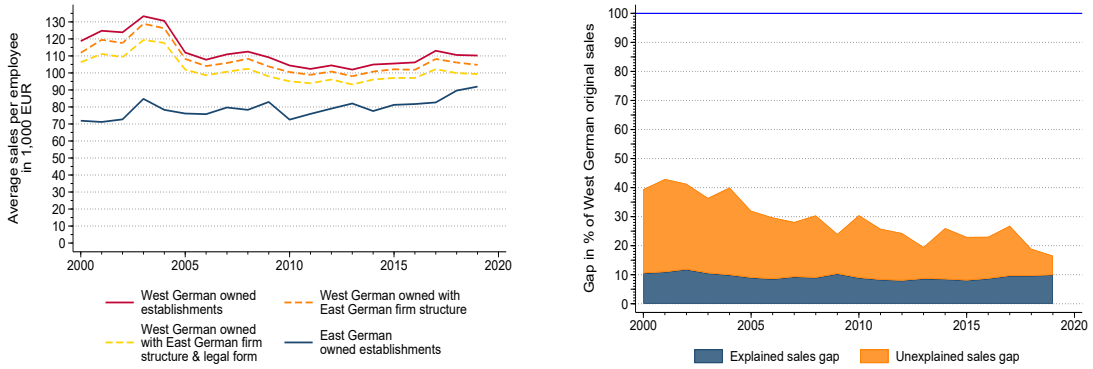
Table 2.5.: Distribution of firm structure by ownership

	West German owned		East German owned	
Year	Single entity	Multi-establishment structure	Single entity	Multi-establishment structure
2000	88.5%	11.5%	97.4%	2.6%
2005	90.4%	9.6%	96.2%	3.8%
2010	89.4%	10.6%	94.3%	5.7%
2015	89.3%	10.7%	95.0%	5.0%
2019	88.5%	11.5%	92.6%	7.4%

Note: Own calculations based on IAB Establishment Panel. Absolute values in table 2.9.

in sole proprietorship than West German owned ones (Table 2.4) and are more often independent units and not part of a multi-establishment firm structure (Table 2.5).

Figure 2.12.: Productivity gap between East and West German owned establishments



(a) Absolute gap (b) Relative gap

Note: Own calculations based on the IAB Establishment Panel. Reweighted by legal form (sole proprietor vs partnership & corporations) and firm structure (single- vs. multi-establishment structure). We define productivity as sales per employee.

Figure 2.12a shows productivity (sales per employee) of East and West German owned establishments and productivity of West-German establishments reweighted to match East German owned establishment characteristics. East German owned establishments' productivity amounted to about 60% of the West German level (about EUR 72k vs EUR 120k) in the early 2000s but has converged to above 80% in 2019.[21] The orange dashed line denotes productivity of West German owned establishments reweighted to resemble the firm structure distribution

[21] Sample sizes for this reweighting exercise are given in Table 2.10.

of East German owned establishments. The yellow dashed line additionally takes legal form differences into account. Reweighting reduces productivity by about EUR 10,000 across the years. Figure 2.12b brings these findings together: Although the relative gap declined from 40% to about 17% of the productivity level of West German owned establishments, the explained gap remained stable at about 10%. This suggests that rather sticky factors like firm structure and legal form play an increasing role for the productivity gap of East German owned relative to West German owned establishments. In contrast to partnerships or corporations, sole proprietorships bring no risk-sharing with other investors, no possibility to bring in new investors and the owner is fully personally liable, which might contribute to liquidity constraints. We undertake several robustness checks. Figure 2.14 shows that a more detailed split of legal forms does not increase the magnitude of the explained gap.[22]

Overall, an important fraction of the sales and productivity gap between East and West German owned establishments is due to differences in establishment characteristics like size, legal form and firm structure. While the observed gap has declined between 2000 and 2019, the share of the gap explained by the above characteristics has gained more importance. This implies that rather permanent features of East-German owned establishments inhibit the convergence of business incomes between East and West Germans.

2.5. Conclusion

In this paper, we documented economic differences between East and West German residents along the regional income and wealth distributions. For this, we employed the Distributional National Accounts (DINA) method which aligns microdata with internationally standardized national accounts. We found that East German residents still earn and own a fraction of their West German counterparts with this gap expanding towards the top of the distribution. At the top of the income distribution, East-West German income differences are predominantly due to East Germans residents' lack of capital incomes, particularly partnership and corporate incomes.

We then showed that East German residents have caught up with West German residents in their business ownership rate. However, their business (and real estate capital) is persistently less valuable. To explore the causes of these persistent gaps, we first conduct a synthetic control analysis to estimate how reunification has affected West German top incomes. For this, we contrasted the observed evolution of the West German top 1% income share with the evolution of the top 1% share of a synthetic West Germany absent the reunification shock. We concluded that reunification significantly boosted West German top incomes and their concentration and linked this to privatization policies in the 1990s in which West German investors prominently increased their medium-run capital income with the acquisition of formerly state-owned East German capital.

Lastly, we estimated how establishment characteristics such as size, legal form and firm structure, influence sales and productivity levels of East and West German owned establishments using the reweighting approach by DiNardo et al. (1996). East German owned establishments

[22] We are currently working on further robustness checks. These comprise 1) using a measure of profit per establishment instead of sales. This is not trivial due to a lack of an adequate variable provided in the dataset. 2) A robustness check contrasting the benchmark of reweighting East and West German owned establishments with East and West German located establishments. Most of the previous literature uses the distinction by location rather than ownership. 3) Another reweighting including the presence of managers in an establishment, a factor identified as influential for total factor productivity in previous literature (Burda and Severgnini, 2018).

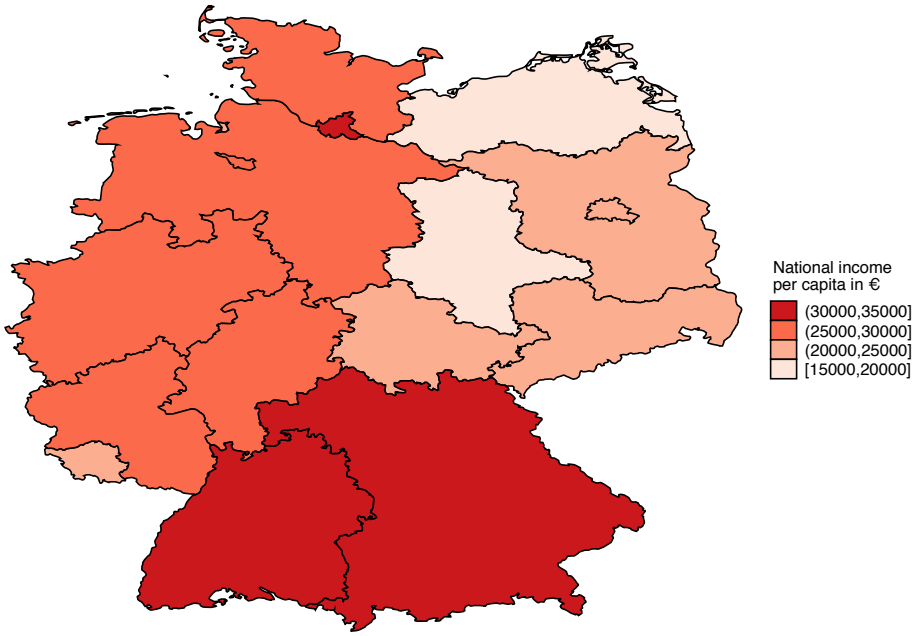
have on average a lower number of employees, are more often held as sole proprietorships and are more often single entities and not part of a multi-establishment firm structure compared to West German owned establishments. These structural differences in characteristics of establishments *owned* by East and West Germans explain an increasing share of the sales and productivity gaps, which might inhibit the convergence of business incomes between East and West Germans. Understanding the interactions of legal form and firm structure with productivity, innovation and credit constraints remain open questions for future research.

Finally, our study highlights the importance of the design of privatization policies. The German reunification did not just bring together a resident population that could accumulate wealth for several decades with a much poorer population, but additionally fostered wealth accumulation among West German top earners. The design of investment and privatization policies facilitated the acquisition of business and real estate assets for high income West German residents. A window of opportunity was missed to have East Germans participate in the income and capital gains generated by their formerly state-owned land, housing and firms from the start. Prices were much lower in the early 1990s so that any catch-up in ownership is much costlier for current and future generations with a GDR background. This speaks to the importance of institutions and policy design for inequality, particularly for countries in transition.

2.6. Appendix figures & tables

2.6.1. Population and income

Figure 2.13.: Net national income per capita, 2018



Note: Own calculations based on the national accounts of federal states (*VGR der Länder*), Statistische Ämter der Länder.

2.6.2. Reweighting: East-West business incomes

Table 2.6.: Variables, reweighting, IAB Establishment Panel

Variable	Description
Sales per establishment	Original variable: *bvole_c*, Business volume (DM/EUR); Cleaning: Variable comprises different types of business volume: 1) turnover/sales (excl. VAT) (kept), 2) balance sheet total (dropped), 3) aggregate premium (dropped), 4) total budget (dropped);
Profit per establishment	Sales minus gross wage/total salary (*bvole_c* - *annual wage sum*), $annual\ wage\ sum = 12 \times 1.2 \times wsum\_c$ (annualization and addition of employer's share of social insurance contributions)
Profit/sales per employee	Division of sales/profit per establishment by *empt_c*: no. of employees
Establishment size	Original variable: *sizegr_d*, 9 categories of establishment size by the number of employees; Cleaning: reduction to 4 categories: 1) 1-4 employees, 2) 5-49 emp., 3) 50-199 emp., 4) 200+ employees
Legal form	original variable: *legfo_d*,from 2017 on *legfo17_d* comprising 1) individual comp, 2) partnership, 3) GmbH & Co. KG, 4) corporation, 5) corporation of public law, 6) other; Cleaning: reduction to 3 categories: 1) Individual firms, 2) partnerships, GmbH & Co. KG and corporations 3) corp. of public law and other (dropped for reweighting)
Headquarter	Original variable: *estpar_d* comprising 4 categories: 1) autonomous, 2) head office, 3) subsidiary, 4) intermediary; Cleaning: reduction to 2 categories: 1) autonomous/single entity (1), 2) multi-establishment firm structure (2-4))

Table 2.7.: Reweighting by establishment size: Sample sizes per reweighting category, IAB Establishment Panel

Unweighted			
Type	Group	West German owned	East German owned
1	1-4 employees	44,722	24,725
2	5-9 employees	17,594	9,423
3	10-99 employees	53,695	20,502
4	100+ employees	25,462	3,202
Weighted			
1	1-4 employees	17,248,979	3,602,830
2	5-9 employees	4,784,636	906,801
3	10-99 employees	4,579,598	708,742
4	100+ employees	395,851	27,267

Table 2.8.: Distribution of legal form by ownership, absolute

	West German owned		East German owned	
Year	sole proprietorships	partnerships & corporations	sole proprietorships	partnerships & corporations
2000	818,219	561,697	216,512	90,916
2005	785,574	538,500	186,608	77,143
2010	786,294	554,571	175,799	83,996
2015	790,066	590,437	169,393	79,015
2019	724,440	619,154	162,981	90,449

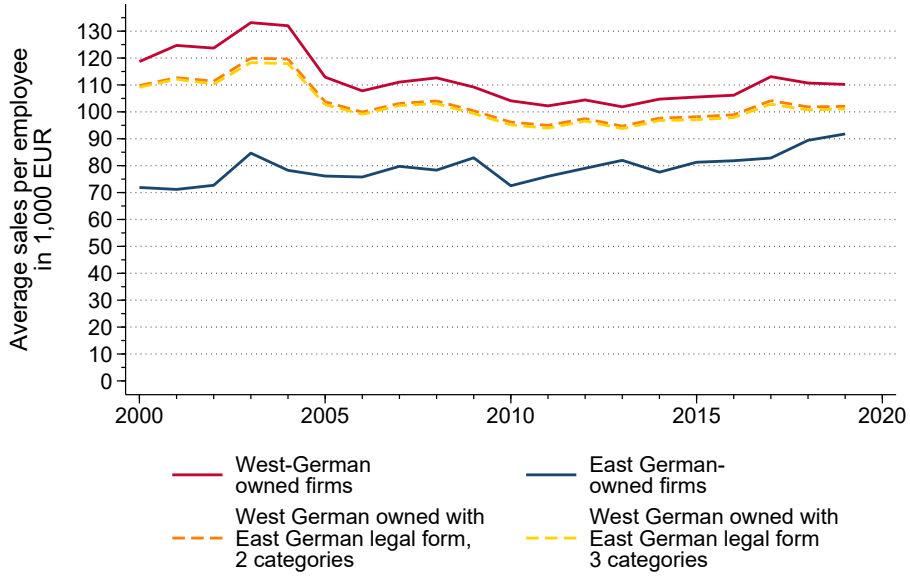
Table 2.9.: Distribution of firm structure by ownership, absolute

	West German owned		East German owned	
Year	Single entity	Multi-establishment structure	Single entity	Multi-establishment structure
2000	1,220,976	158,940	299,443	7,985
2005	1,196,703	127,371	253,831	9,920
2010	1,198,406	142,459	245,085	14,710
2015	1,232,323	148,180	235,927	12,481
2019	1,188,421	155,173	234,639	18,791

Table 2.10.: Reweighting by firm structure and legal form: Sample sizes per reweighting category, IAB Establishment Panel

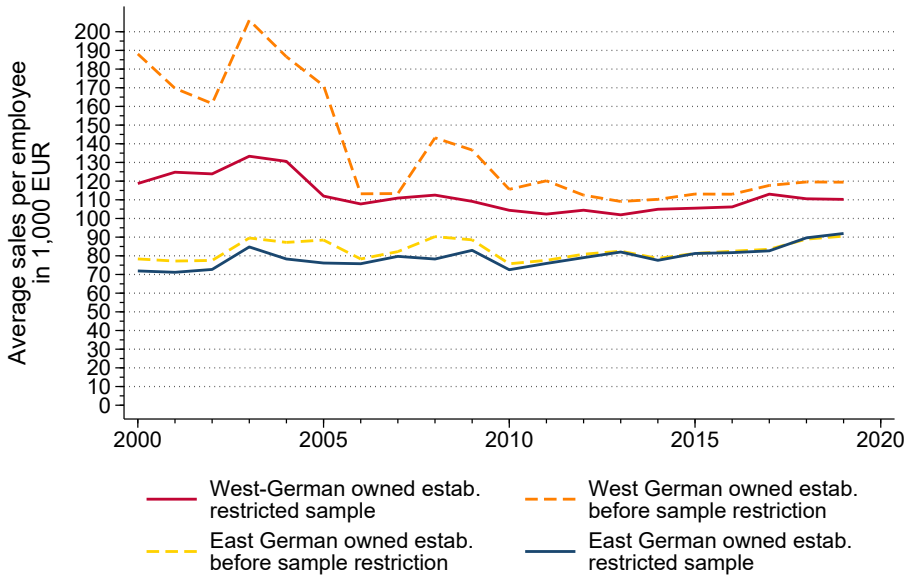
Unweighted			
type	Group	West German owned	East German owned
1	sole proprietorship & single entity	43,568	27,871
2	sole proprietorship & part of multi-estab. firm structure	1,824	661
3	partnership or corporation & single entity	64,420	26,189
4	partnership or corporation & part of multi-estab. firm structure	31,055	2,910
Weighted			
1	sole proprietorship & single entity	15,170,330	3,519,766
2	sole proprietorship & part of multi-estab. firm structure	475,569	70,969
3	partnership or corporation & single entity	8,912,376	1,478,122
4	partnership or corporations & part of multi-estab. firm structure	2,301,862	156,760

Figure 2.14.: Robustness: Reweighting by legal form



Note: 2-category legal form (benchmark): 1) sole proprietorship & 2) partnerships and corporations; 3-category legal form (robustness): 1) sole proprietorship & 2) partnerships, and 3) corporations. Corporations of public law were excluded as well as other marginal legal forms.

Figure 2.15.: Sales per employee before and after sample restriction



2.6.3. Synthetic control: West German top incomes

Table 2.11.: Predictor variables and sources

Variable	Source
Top 1% share (fiscal income)	World Inequality Database 2023, we add fiscal income share for Austria by Altzinger et al. (2012)
GDP per capita	OECD: https://stats.oecd.org/
Trade openness	Abadie et al. (2015)
Consumer Price Index	OECD: https://stats.oecd.org/
Gross capital formation	OECD: https://stats.oecd.org/ & Federal Statistical Office: Lange Reihen
Life expectancy males	World Bank & Sensch (2014): HISTAT online data compilation
Industry share in value-added	Abadie et al. (2015)

Figure 2.16.: Robustness of the synthetic control: In-time placebo test

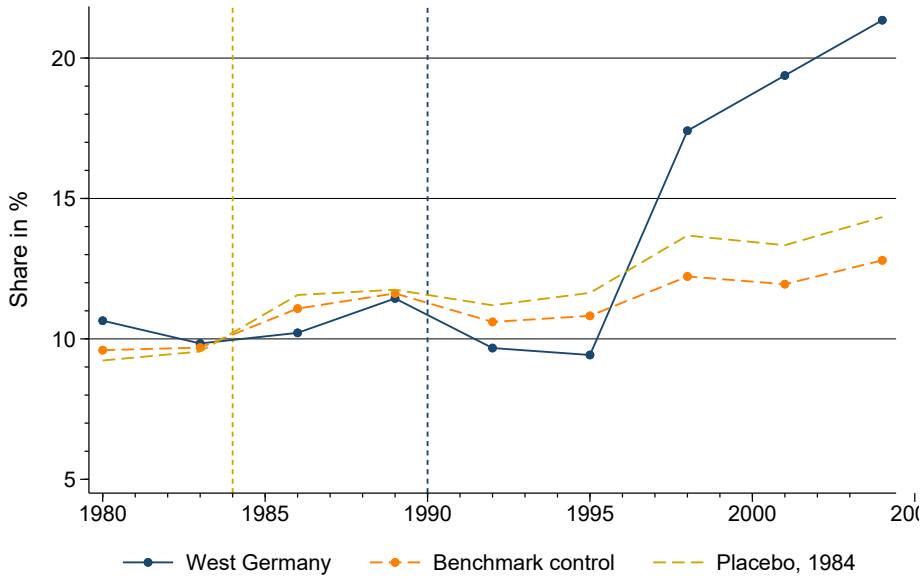
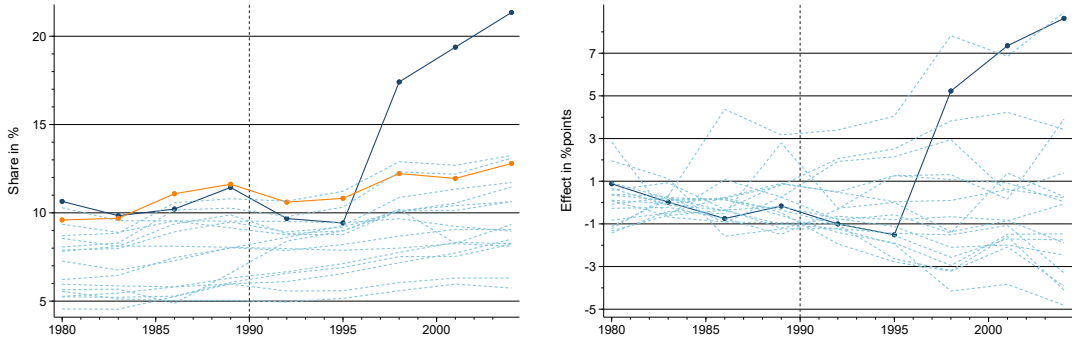


Figure 2.17.: Robustness: In-space placebo tests

(a) Top 1% share

(b) Difference actual and synthetic



Note: Panel (a) shows the development of the top 1% share for West Germany (dark blue), the synthetic control of West Germany (orange) and the synthetic control of each country of the sample (light blue). Panel (b) shows the difference between the actual and synthetic top 1% income share for each country in the sample. The dark blue line shows the German case. The donor country with a similarly large difference between the actual and synthetic share is the United States. However, it is visible that the deviation for the United States emerged in the early 1980s and thus, shows a very different pattern from West Germany.

Figure 2.18.: Robustness of the synthetic control: Post-/Pre-RMSPE ratio

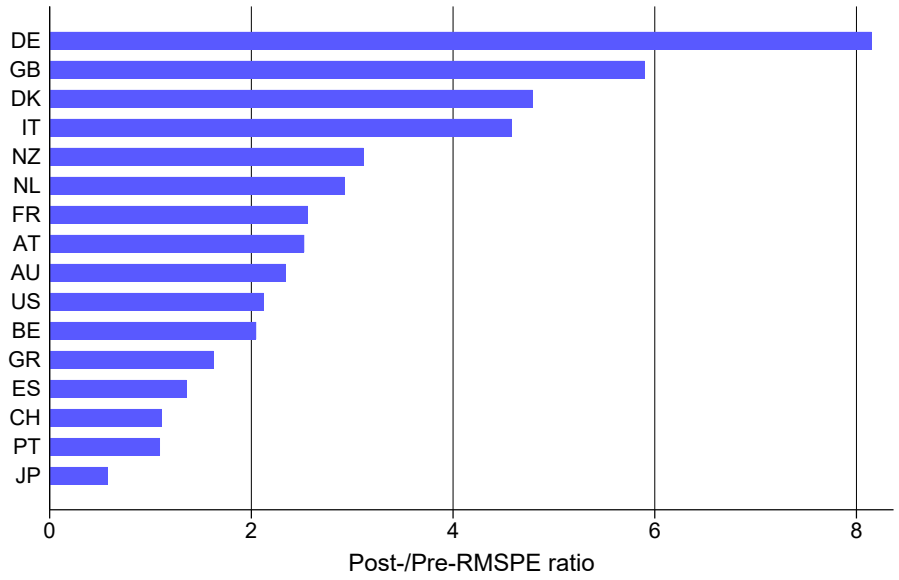
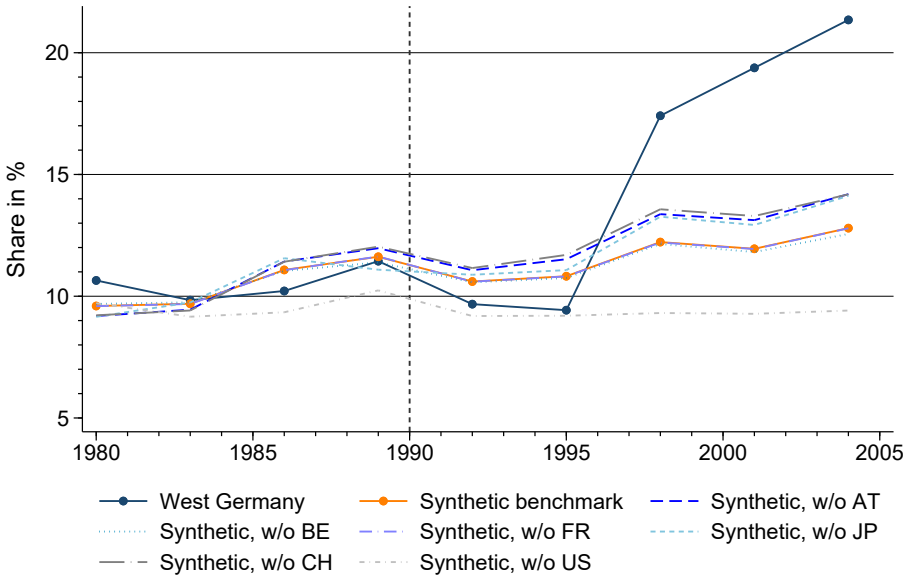


Figure 2.19.: Robustness of the synthetic control: Leave-one-out analysis



3 The Long Way to Gender Equality: Gender Pay Differences in Germany, 1871-2021*

Abstract

This paper provides the first harmonized time series of the gender earnings ratio for the full-time employed workforce in Germany since the 1870s and compares Germany's path towards gender earnings equality with the Swedish and U.S. cases. The industrialization period until 1913 yielded slow advances in economic gender relations due to women's delayed inclusion in the industrial workforce. In contrast, the first half of the 20th century exhibited giant leaps towards pay equality. In Germany, the gender earnings ratio increased from 46% in 1913 to 58% in 1937. Similar increases are visible in Sweden and the United States. In all three countries, the interplay between increased women's education and increased returns to education due to the expanding white-collar sector fueled pay equalization. Yet, women's educational convergence has been slowed compared to the United States, due to the dominance of on-the-job vocational training in Germany. Germany's increase in the earnings ratio is better explained by women's migration from low-paid agricultural work to higher-paid white-collar jobs. The postwar period brought slower growth of the gender pay ratio and diverging developments between Germany, Sweden and the United States attributable to differences in economic conditions and policy action.

JEL Classification: J16, J31, I24, N33, N34

3.1. Introduction

Germany still exhibits one of the highest gender earnings gaps in Europe. While the EU average stands at 13%, gross hourly earnings in Germany show an average unadjusted gender gap of 18% (EUROSTAT, 2023, year 2021). While gender pay[1] differences are currently the focus of intense academic and policy discussion, only few studies take a long-run perspective despite its potential. Long-run time series and their international comparison can help us answer up-to-date pressing questions in a broader historical framework: Which factors and contexts propel gender pay equalization and which do hinder the path towards equality? In which institutional and economic conditions do increasing women's education and labor market participation translate into more pay equality? Which pathways have already increased pay equality and where is still space to act?

Despite research efforts in several countries, Claudia Goldin's (1990) work on the long-run development of the gender earnings ratio in the United States is still the benchmark for understanding gender earnings differences since the industrialization to the present. Svensson (2003)

* I thank Klaus Harney for providing the data on vocational schooling before 1945 and Lotte Maaßen and Teo Ernst for their excellent research assistance. Further, I am thankful for comments from the seminar participants at Paris School of Economics and Roma Tre as well as at the DIW gender economics workshop 2018, the EHES conference 2019, the EHS annual conference 2020 and ECINEQ 2021.

[1] I will use the words *earnings* and *pay* as equivalents throughout this article, while I use *wage* for official statistics providing a pay rate.

and Stanfors and Goldscheider (2017) have added accounts on the development of the gender earnings ratio and the female labor force participation for Sweden.[2]

This paper presents the first comprehensive long-run time series of the gender earnings ratio for Germany since the 1870s. I relate the observed developments with three key influencing factors: 1) educational convergence between men and women, 2) women's occupational migration from lower-paid to better-paid jobs, and 3) changes in female labor force participation. I compare the German evolution of the gender earnings ratio and its influencing factors with the Swedish and U.S. cases. By adding another country-specific account and discussing the effect of prominent influencing factors in comparison to Sweden and the United States, I can reevaluate the U.S. benchmark conclusions. The comparison between Germany, Sweden and the United States is exceptionally fruitful as these countries diverge substantially in structural factors such as different speeds of industrialization, cyclical events such as booms and busts and institutions such as differences in the education system as well as in the policy action and norms towards working wives and mothers.

For Germany, only scattered evidence on the long-run evolution of the gender earnings ratio exists. Relevant studies either focus on specific occupational or sector groups (Bajohr, 1979; Baldauf, 1932; Gärtner, 2014; Maier, 2007; Suhr, 1930; Ziegler, 2010) or explore gender earnings differences for the entire workforce using microdata, which is only available since the 1970s (e.g. Fitzenberger and Wunderlich 2002; Kunze 2005; Gärtner 2014). Gómez León and de Jong (2019), providing a time series for the gender earnings ratio for the employed workforce from 1900 to 1950, is the closest contribution to this paper. My study goes beyond Gómez León and de Jong's study since I have assembled a substantially more granular earnings database and I contextualize the ratio's development by relating it to crucial influencing factors.

The gender earnings ratio is defined as the share of women's average gross hourly earnings in comparison to men's. I have assembled data on gendered gross earnings from a rich variety of sources due to the absence of continuous and consistent earnings data over the researched time period. For the period before World War I, I draw on surveys commissioned by professional associations and contemporary research works. For the interwar period, I use data by professional white-collar and agricultural organisations and the first official wage statistics for blue-collar industrial workers. Only the postwar period holds more comprehensive official earnings statistics for a large share of the employed workforce in agriculture, industry and services. This paper is supplemented by an extensive data appendix, detailing all used data sources as well as the assumptions and harmonization strategies applied.

My final dataset comprises men's and women's average gross earnings for three groups: 1) agricultural workers, 2) industrial blue-collar wage receivers and 2) white-collar workers (salaried personnel) in industry, commerce, banking and insurance. To arrive at a representative estimate of the gender earnings ratio for the entire employed workforce, I use the social tables method. The main idea is to weight information on the male and female average earnings of these three occupational groups by their relative size in the male and female employed workforce.

I trace gender earnings equality throughout five distinct periods which provided very different environments for working women and their work's valuation in Germany: (I) the era of rapid

[2] Further time series of gender earnings differences exist for other periods, among others for unskilled workers in England between 1260 and 1850 (Humphries and Weisdorf, 2015), and during the industrial revolution from 1720 to 1850 in Britain (Burnette, 2008). Boter (2017) provides a comparable series to Humphries and Weisdorf (2015) for the Netherlands. Pleijt and Zanden (2021) provide evidence on agricultural earnings differences in Europe from 1300 to 1800. Recently, De la Rosa Ramos (2021) built a time series for teachers, unskilled workers and industrial workers in the 20th century in Mexico.

industrialization from 1871 to 1913; (II) World War I (1914-1918), a time of tumultuous change in the industrial workforce, and the Weimar Republic (1918-1933) with its progressive institutions such as compulsory secondary schooling and strong labor unions; (III) the Nazi reign between 1933 and 1945 in which married women were first urged to leave the labor market but soon demanded in the industrial workforce in the context of the rearmament; (IV) the postwar period in the Federal Republic of Germany between 1949 and 1990 characterized by single-earner households during the intensive child-rearing phase and increasing numbers of married women reentering the labor market thereafter; and (V) the post-reunification period since October 1990 which brought together two countries and populations with different wage levels, gender earnings ratios and skill sets.

The industrialization period since the 1870s yielded slow advances in the gender earnings ratio due to women's delayed inclusion in the industrial workforce in Germany. In contrast, the first half of the 20th century exhibited giant leaps towards earnings equality. In Germany, the gender earnings ratio increased from about 46% in 1913 to 58% in 1937. Similar increases were visible in Sweden and the United States. In all three countries, the interplay between increased women's education and increased returns to education due to women's rapid migration into the expanding white-collar sector fueled the increases in the gender earnings ratio. In Germany, women's educational convergence was less successful compared to the United States. Germany's focus on on-the-job apprenticeships might have posed higher entry barriers for young women into secondary schooling compared to the school education centered system of the United States. While educational convergence predominantly pushed the gender earnings ratio upwards in the United States (Goldin, 1990), Germany's similar increase in the earnings ratio is better explained women's migration from low-paid agricultural work to higher-paid white-collar jobs in the context of Germany's late industrialization. The postwar period brought about slower growth in the gender earnings ratio, but a growing participation rate due to the reentry of married women into the labor market. The paths of the United States and the European countries diverged substantially. While Sweden and Germany increased their earnings ratios between the late 1950s and 1980, the United States experienced stagnation from 1950 to 1980 and a rapid catch-up in the 1980s. In the United States, the strong reentry of married women with low work experience depressed the female earnings potential (Goldin, 1990), while Germany's postwar recession delayed married women's reentry into the labor market and slow increases in the gender earnings ratio were observable. Sweden took the lead in terms of gender equality in the 1970s due to a set of policies propelling married women's labor supply, absent in Germany and the United States.

To better interpret the results of this study, it is crucial to understand its blind spots and limitations. Specific to German history is the changing extent of the nation. The territory of Germany from the Reich to reunified Germany underwent several changes from ceding territories after World War I, the annexation of territories by the Nazis and the reshaping of the postwar borders as well as the separation and the reunification of the two Germanys. I estimate the earnings ratio for the respective historical territory of Germany. Further, my sample population comprises only full-time employees in the private sector and outside of the house. Civil servants, self-employed, assisting family members and domestic servants are not included. Overall, constructing a harmonized time series means combining heterogeneous data sources. Despite carefully harmonizing earnings data and taking care that representativeness is given along geographical and occupational dimensions, the stated estimates remain approximations. Still, they can shed light on crucial historical dynamics.

The article is organized as follows: In section 3.2, the data, its representativeness, the earnings concept and empirical strategy are laid out. Section 3.3 discusses the main development of the gender earnings ratio since the 1870s in Germany and in comparison to the Swedish and U.S. time series. The subsections of chapter 3.4 take a detailed look at influencing factors of the gender earnings ratio's evolution in international comparison, namely increased education of young women (section 3.4.1), occupational migration into better-paid jobs (section 3.4.2) and female labor force participation of married women (section 3.4.3). Section 3.5 concludes.

3.2. Data, earnings concept and representativeness

I construct a long-run time series of the gender earnings ratio which is equal to women's gross hourly earnings W_f across all sectors and occupations divided by men's gross hourly earnings W_m:

$$Gender\ earnings\ ratio\ = \frac{W_f}{W_m} \tag{1}$$

For most of the 150 years since the foundation of the German Reich, wage and earnings statistics in Germany were published separately for three broad occupational groups across different sectors of the German employed workforce: 1) (blue-collar) agricultural workers, 2) blue-collar industrial workers, and 3) white-collar workers (salaried personnel) in industry, commerce, banking and insurance.[3] To construct the male and female gross hourly earnings across all occupations in the sample population, W_f and W_m, I apply the social tables methodology. This method was previously used to explore consumption and income inequality of the active population in different countries based on budget surveys, tax records or earnings (Gómez León and de Jong, 2019; Lindert and Williamson, 2016; Milanovic et al., 2010). But also, Goldin (1990) and Svensson (2003) have applied the key steps of this method to construct male and female earnings and the gender pay ratio of the employed workforce.[4] The main idea of this method is to compile information on the average earnings of different social or economic groups of the population, in this case, the three occupational groups, and weight these group-specific average earnings by the relative size of each group in the sample population (Gómez León and de Jong, 2019):

$$W_g = \sum_{o=1}^{O} \phi_{o,g} \cdot w_{o,g} \tag{2}$$

$w_{o,g}$ denotes the gross hourly earnings by gender g of each of the three occupational groups o. $\phi_{o,g}$ indicates the relative share of each occupational group in the female and male sample population, with $\phi_{o,g} = \frac{n_{o,g}}{N_g}$ and $\sum_{o=1}^{O} \phi_{o,g} = 1$.

To construct the respective weights $\phi_{o,g}$, I use information from the occupational censuses conducted in 1882, 1895, 1907, 1925, 1933, 1939, 1950 and 1961. For 1882, 1895 and 1907, I use corrected census data on the composition of the employed workforce outside of the home provided

[3] Data are available for agricultural workers, *(Industrie-)Arbeiter*, which are blue-collar wage receivers, and *Angestellte* or salaried personnel, which consist mainly of white-collar workers. Throughout the article, I will use white-collar workers and salaried personnel as equivalents. These three main groups had separate insurance schemes and advocacy or labor unions which were the main source of earnings information until the emergence of comprehensive official wage statistics.

[4] See Goldin (1990, p. 64) whose notation I follow.

by Müller et al. (1983).[5] Since the 1960s, I use comparable information based on the micro census (Statistisches Bundesamt 1964 - 1976 & Statistisches Bundesamt 1977 - 1984). These sources provide gendered information on the labor force by occupation, sector and industry in the main job. I interpolate linearly between census years. From 1984 onwards, I use the Socio-Economic Panel's yearly microdata. This dataset provides individualized earnings information which can be weighted using individual cross-sectional weights (Wagner et al., 2007). The geographical coverage changes with the borders of Germany: The censuses of 1882 to 1907 cover the resident population of the German Reich, the 1925 and 1933 censuses the population of the Weimar Republic excluding Posen, West Prussia, Saarland and Alsace and Lorraine; the census of 1939 comprises the territories of the Saarland again as well as Austria and Sudetenland. The 1950 census comprises the Federal Republic of Germany without Saarland and Berlin, the 1961 one holds the same population but includes the Saarland. The microcensus data since the 1960s represents the resident population of the entire Federal Republic. For more information, see table B.4.4 in the data appendix.

To construct the male and female gross hourly earnings of each occupational group $w_{o,g}$, I compile nominal gross earnings data from a rich variety of sources which range from surveys commissioned by professional associations and for research purposes, contemporaneous academic studies to official wage statistics of the Statistical Office. For the period before World War I, I mainly draw on surveys commissioned by professional associations and for research purposes. For agricultural workers, I use the detailed data of a survey commissioned by the association of farmers on agricultural workers' earnings across the German Reich in 1873 (von der Goltz, 1875). Later, I can draw on the academic work by Asmis (1919) who assembled agricultural workers' earnings for Prussia in 1849, 1873, 1892 and around 1910 from different previous works. For blue-collar workers' earnings, I rely on the data assembled by (Kuczynski, 1961, 1962, 1963).[6] In the interwar period, labor unions gained strength or, in the case of agricultural work, were founded. These and occupation-specific associations published detailed wage statistics for agricultural workers (Vorstand des Deutschen Landarbeiter-Verbandes, 1926a; 1929) and white-collar workers (Gewerkschaftsbund der Angestellten (GdA), 1931; Glaß and Kische, 1930; Suhr, 1930). Also, the first official wage statistics for blue-collar workers were regularly published during this period (Statistisches Reichsamt, 1929, 1930, 1931b, 1932). Only in the post-World-War-II period I can rely on broadly standardized publications by the statistical office of the Federal Republic of Germany (FRG) - foremost the structure of earnings surveys (*Verdienststrukturerhebung*) for blue- and white-collar workers published in the statistical yearbook (Statistisches Bundesamt, 1957-1990) and the salary and wage structure survey in agriculture (*Gehalts- und Lohnstrukturerhebung*) (Statistisches Bundesamt, 1958a, 1959, 1960, 1961a, 1963a,b, 1965). The structure of earnings survey has not incorporated the newly emerging occupations in the service sector and thus loses in coverage. Therefore, I use SOEP data as soon as it is available. Details on the data sources, coverage and harmonization strategies can be found in the data appendix B.

I use actual gross hourly earnings to compute the gender earnings ratio. Gross earnings are defined as all wages and salaries, including the employee's share of social insurance contributions. I exclude supplements for overtime work in industrial and white-collar work as these are detailed

[5] The authors harmonized categories over time and corrected the unsystematic representation of assisting family members in these early years.

[6] Hohls (1991) affirms that "In fact, all [earnings] series for the first years of the German Reich are based on Kuczynski's data", i.e. Desai (1968), Bry (1960) and Hoffmann (1965). Kuczynski assembled his data mainly based on labor unions' data on collectively agreed wages, reports of chambers of commerce and firm data on actual earnings (Hohls, 1991).

from the 1920s onwards. The gross earnings concept comprises social supplements such as marriage and child allowances mainly paid by employers to married men because these are usually not separately detailed in publications. I include in-kind payments (mainly board and lodging) to agricultural workers to make their earnings comparable to other occupational groups' earnings. These are often stated in the respective publications and are predominantly based on (stock) market prices (for details, see data appendix table B.1.1).

Before applying the social tables method (equ. 2), several harmonization steps are necessary: 1) a conversion to hourly earnings, 2) in several years, the construction of weighted average female and male earnings across all agricultural or blue-collar industrial workers requires weighting region-specific or industry-specific earnings, and 3) an inflation or deflation when data years for the three occupational groups do not overlap. In the following, I will describe briefly the main steps.

Official wage statistics recorded blue-collar industrial workers' earnings on an hourly basis since 1913.[7] However, white-collar workers' earnings were presented on a monthly basis and agricultural workers' earnings were paid on a daily basis (free day laborers) or annually based on their yearly contract (agricultural servants). I draw on information in the earnings data publications on working hours per day and working days per year, i.e., for agricultural workers on von der Goltz (1875), Statistisches Bundesamt (1958a), Baldauf (1932) and for white-collar workers on Kaiserliches Statistisches Amte, Abteilung für Arbeiterstatistik (1912), Gewerkschaftsbund der Angestellten (GdA) (1931), Statistisches Reichsamt (1935) and Statistisches Bundesamt (1966-2010). Gender differences in working hours are marginal since my sample only comprises full-time workers.

In the early years of the time series, I have to first construct the weighted average earnings $w_{o,g}$ for each of the three occupational groups because the original data source lists earnings for different geographical regions or different industries separately. For agricultural workers, I construct average gross earnings by weighting earnings information of different provinces and districts (*Regierungsbezirke*) using occupational census data for the strength of the agricultural workforce across districts for 1873, 1913 and the 1920s. Further, agricultural workers comprise different groups. Until 1930 the main groups were day laborers and servants.[8] I weight their earnings against each other using their share in the agricultural workforce based on data by Baldauf (1932). For blue-collar workers, until the late 1930s, gross hourly earnings are given for different industries respectively. I use the same logic as in equation 2: I construct weighted mean earnings of women (men) in the industrial sector by weighting the industry-specific average female (male) earnings by the share of the female (male) workforce in each industry relative to the entire female (male) industrial blue-collar workforce.

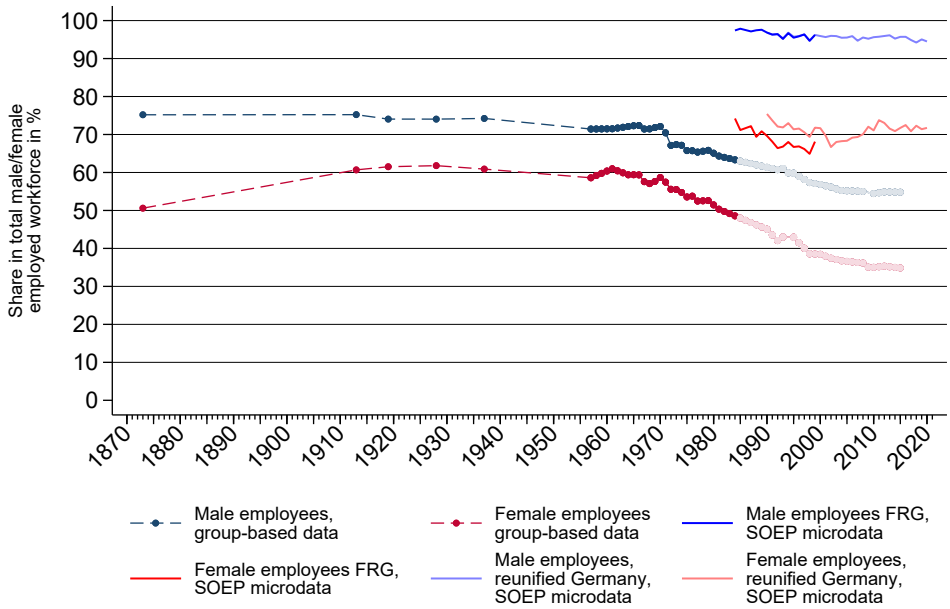
Further, for the early years of the time series, earnings data years for each of the three groups do not necessarily overlap. This is the case for the 1870s where I can draw on agricultural workers' earnings from 1873 but blue-collar workers' earnings of different industries from 1870 to 1880. To construct the gendered earnings for the entire sample, I need to bring together earnings for the

[7] My benchmark concept is average hourly gross earnings. I deviate from this only for the data point of 1873 where I use weekly earnings. Many previous studies discuss collectively agreed wages (Bajohr, 1979; Braunwarth, 1955, for example). Collectively agreed minimum wages can give a good overview of institutionalized differences in pay. However, employers had substantial liberty to go beyond these benchmarks such that the gender pay ratio of actual earnings can draw a different and more precise picture.

[8] I omit contracted day laborers as these had to provide labor power of two to three persons of the family and earned a family wage.

same year. Thus, I de- and inflate earnings data to 1873 using the index of living costs by Bry (1960, 325, table A-1). For the years after 1881, I can use the consumer price index provided by Sensch (2008).

Figure 3.1.: Coverage of male and female employees in assembled earnings dataset.



Note: This graph shows the coverage of the assembled earnings dataset in the employed male/female workforce. It is based on occupational censuses, which are linearly interpolated, and the comparison between the entire employed male/female workforce and the workers groups represented in earnings data, i.e., excluding blue-collar workers in commerce, transport and services, domestic services and agricultural workers after 1970 for published (group-based) data sources due to lack of gendered earnings data. The 1873 data point is estimated using the 1882 census. For the earnings dataset group-based average earnings published in the statistical yearbooks and other publications are used until 1983. SOEP microdata is used thereafter due to its more contemporary scope and comprehensiveness, including a broader range of service occupations.

My sample population N_g is restricted to the employed workforce outside the home and in the private sector. I do not include assisting family members, domestic servants outside of agriculture, nor civil servants. I also omit the marginal groups of blue-collar workers in agriculture, commerce and transport due to a lack of earnings information. I only consider full-time employed persons. For the postwar period until reunification, the analysis is restricted to the Federal Republic of Germany. Considering the occupational composition of the labor force, my dataset covers about 60% to 80% of the male employed workforce and 50% to 60% of the female employed workforce from 1873 to 1983. From 1984, I draw on SOEP survey data and reach a coverage of about 95% for employed men and above 70% for employed women (figure 3.1). Before 1900 domestic servants are the blind spot of my sample. Since domestic servants belong to the lowest earnings groups and women were overrepresented in this occupation, my estimate in the late 19th century, when domestic service was an important occupation field for young women, can be seen as an upper bound of the gender pay ratio.[9] Coverage drops in the 1970s, when the structure of earnings survey did not adapt to the newly emerging occupations in the service sector. Therefore, I use SOEP from 1984 onwards.

[9] Gómez León and de Jong (2019) assume pay in domestic services to be about 30% of average earnings in services.

To ensure robustness, I compare the benchmark time series with alternatives whenever different data sources are available. For the post-WW-II period, I can compare my benchmark series based on the yearly *Verdienststrukturerhebung* with results from the more comprehensive but irregularly conducted *Gehalts- und Lohnstrukturerhebung* for 1951 to 1990. From 1984 to the present day, I compare my benchmark series based on the Socio-Economic Panel with the continuing time series based on the structure of earnings survey. The results are presented in appendix section 3.6.1.

3.3. The gender pay ratio since the 19th century

This section provides the main results. The first part (subsec. 3.3.1) presents the German time series of the gender earnings ratio since the 1870s and discusses its evolution in historical context and in light of the gender earnings ratios in the three broad occupational groups. The second part (subsec. 3.3.2) compares the German series to the Swedish and U.S. cases.

3.3.1. The German path

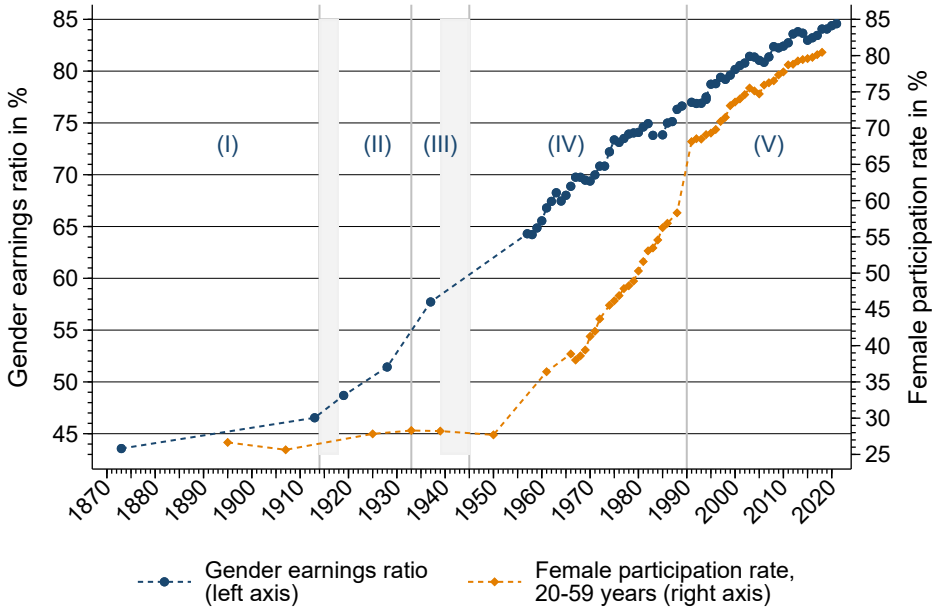
Figure 3.2 shows the gender earnings ratio for full-time employees from the foundation of the German Reich to the present day. I contrast the development of the earnings ratio with the female labor force participation rate as it indicates the selectivity and extent of women actively participating in the labor market.

Developments in five periods can be observed: (I) the period of high industrialization, from the 1870s until the onset of World War I, shows stagnation in the female labor force participation and slow increases in the gender earnings ratio. The relative female earnings stood at about 44% in the 1870s and at 46% in 1913. (II) The first half of the 20th century exhibited the most pronounced advances towards earnings equality while female participation in the labor market remained low. The year 1919, in the aftermath of World War I and the November revolution, exhibits an earnings ratio of about 49%, two percentage points higher than in 1913. The Weimar Republic brought about further increases to about 51% in 1928.[10] (III) The Nazi period could not stop the upward development of the gender earnings ratio, which reached almost 58% in 1937. (IV) Slower growth characterizes the postwar period in the Federal Republic of Germany. The gender earnings ratio stood at about 64% in 1957, the first year of comprehensive official earnings statistics in agriculture, blue- and white-collar work after the war, and stood at about 76% in 1990. Two more dynamic phases are visible: Between 1958 and 1966[11], and again between 1970 and 1975 the gender earnings ratio increased by approximately five percentage points, respectively. This dynamic slowed down in the 1980s. At the same time, female participation rate soared in the postwar period. (V) Despite the large economic changes brought about by the political and economic unification of the Federal Republic and the former GDR, the gender earnings ratio kept its longer-term moderate upward trend in the former Federal Republic (West Germany) as well as in reunified Germany from the mid-1990s. Only the participation rate shows a pronounced upward leap after the expansion of the resident population.

The development of the overall gender pay ratio results from developments of gender differences within sectors and the changing valuation of these sectors due to male and female employees' migration across sectors and occupations. Figure 3.3 shows gender pay ratios within the three

[10] And possibly more until 1932 which cannot be traced here.
[11] This is a period in which Germany just emerged from a severe recession with high unemployment (see figure 3.22)

Figure 3.2.: Gender earnings ratio and female labor force participation rate



Source: Own calculations based on sources listed in the data appendix. Female participation rate without assisting family members. Gender earnings ratio of reunified Germany since 1994. Participation rate of reunified Germany since 1991.

broad occupational groups. While the development of the overall pay ratio shows a relatively persistent upward trend, the developments within the three broad occupational groups were more volatile. Agricultural workers show the highest gender pay ratio. However, this group also exhibited the lowest average pay, often only about two-thirds of industrial workers' earnings. White-collar workers in services stood at the top of the average earnings hierarchy (see section 3.6.3 in the appendix).[12]

(I) In its **high industrialization period** since the 1870s, the German economy changed its mode of production from small-scale family handcraft businesses to industrial work in many parts of the country. In the first decades, mainly men were drawn to industrial work outside of low-paid agriculture or home-based craftsmanship. Women followed this trend only with a delay around the turn of the century and went from low-paid agriculture into consumption goods industries (Müller et al., 1983). The typical female workers outside of the home were young, single women who withdrew from the labor market and employment outside of the home upon marriage (Willms, 1983a).

(II) Despite the pivotal events of the **first World War** and the November revolution of 1918, the gender pay ratio only increased by about 2 percentage points between 1913 and 1919. World War I brought about temporary increases in the pay ratio in blue-collar industrial work which, however, ceased right after the demobilization (figure 3.3). While women were overrepresented in low-paid consumption industries (textile, clothing and food) before the war, during wartime the drafting of skilled men to the front decreased the average skill level of male workers, offered opportunities for women to take part in training programs (*Anlernprogramme*) to become skilled

[12] I discuss the influence of employees migration between the three occupational groups in detail in subsection 3.4.2)

workers replacing the drafted men and offered more opportunities for women to transfer to higher-paid, formerly male-dominated heavy industries. Figure 3.3 shows how the gender pay ratio of blue-collar workers in industry rose from September 1914 to September 1916 and then stabilized. However, as soon as men returned from the front, male workers' skill level rose again and many industrial jobs were reclaimed by them (Bajohr, 1979). The industrial gender pay ratio fell sharply from September 1918 to March 1919 to marginally above the 1914 level. The establishment of the **Weimar Republic** introduced a strong progressive institutional framework. Labor unions strengthened and established minimum collective wages in agriculture, industrial work and white-collar work. Some scholars highlighted the strong equalization tendencies sparked by the revolution (Karbe, 1928 and Kuczynski, 1963 as described by Bajohr, 1979). However, the collectively agreed minimum wages were still gendered, so that women's minimum collective wages amounted to 60 to 80 % of men's of the same qualification level in industrial work in the 1920s (Bajohr, 1979, 46ff).[13] Also, employers usually went beyond the minimum wages which decreased the pay ratio in effective earnings compared to collective minimum wages. Thus, the November revolution did not bring about instantaneous equalization tendencies. Still, figure 3.3 shows substantial advances in the gender pay ratio of blue-collar industrial workers and white-collar workers in the 1920s. Overall, the Weimar Republic brought about a higher level of the gender pay ratio within blue- and white-collar work and in the cross-sector gender pay ratio compared to the pre-war period.[14]

(III) The **fascist takeover** in the 1930s did not bring about substantial changes in the established earnings hierarchies nor did it stop the increase in the overall gender pay ratio. The introduced wage cuts in 1933 applied proportionally to all worker groups and the strict wage controls froze the established pay differences within industries and qualification levels (Bajohr, 1979, p. 56). "Collective bargaining" minimum wages show a stagnating gender pay ratio from 1933 to 1943 (Bajohr, 1979). Thus, it is likely that the small increase in blue-collar work between 1928 and 1935 that figure 3.3 suggests happened before 1933. In sections 3.4.1 and 3.4.2 I show that an interplay between women's increased training and migration of young women to higher-paid white-collar jobs are the main drivers of the increase in the gender earnings ratio across occupational groups in the interwar period. This dynamic continued into the Nazi period.

(IV) In the Federal Republic of Germany, the **postwar period** brought about slower increases in the gender pay ratio than before. However, the female labor force participation rate soared. Thus, more and more women participated in the labor market. In this period, Germany introduced several acts that might have propelled gender pay equality, such as the "equal rights act" in the constitution, the "equal pay for equal work" act of 1956 and the inclusion of women's right to receive equal pay and equal treatment in the workplace into civil law in 1980. However, I share Gärtner's (2014) conclusion that none of these regulations show a remarkable impact on the development of the gender pay ratio. Also, despite strong labor unions, these did not promote gender pay equality. Gärtner (2014) argues that some legislation was even undermined by employers and labor unions assigning disproportionate importance to physical strength in the construction of collective wage groups. In 1955 so-called *Leichtlohngruppen* were introduced in the industrial sector. These remuneration groups for work that did not require heavy bodywork

[13] Gendered pay rates were not only paid under the time wage regime but women also earned less than men under piece rate regimes, thus, for the same work output. Also, Bajohr (1979, 52f.) underlines that qualified women's wage rates (*gelernte und angelernte*) did not reach the wage level of unskilled men's in any of the 12 monitored industries in the late 1920s and early 1930s.

[14] This was already pointed out by contemporary scholars based on restricted blue-collar wage data (Karbe, 1928 and Braunwarth, 1955 as cited by Bajohr, 1979, p. 46)

were at the bottom of the pay scale. More than 80% of female blue-collar workers were assigned to them (Gärtner, 2014).

(V) While *reunification* brought about a marked jump in female labor force participation, the gender pay ratio does not show a remarkable disruption between the former FRG and reunified Germany as of 1994. This seems counterintuitive given the higher gender pay ratio within East Germany. However, different earnings developments and pay ratios between genders as well as East and West Germans overlap particularly in the early 1990s. Due to the inclusion of lower-paid East Germans in the early 1990s, the reunified earnings distribution widened. Since women are usually overrepresented at the lower end of the earnings distribution, a wider earnings structure decreases the pay ratio (Blau and Kahn, 1996). Hunt (2002) shows how the increase of the gender pay ratio by 10 percentage points in East Germany between 1990 and 1994 is in large part driven by the involuntary exit of low-skilled female workers during the transition that entailed an increasing positive selection into employment for women. This phenomenon is visible in the East German and reunified German time series in the early 1990s (see figure 3.14). From the mid-1990s, we see a stagnating gender pay ratio in East Germany, and only marginally growing pay ratios in the former FRG and reunified Germany. Gallego-Granados and Wrohlich (2020) find a substantial and increasing positive selection bias into full-time employment for women in West Germany between 1990 and 2014[15], which might explain the marginal increases in the pay ratio since the mid-1990s.

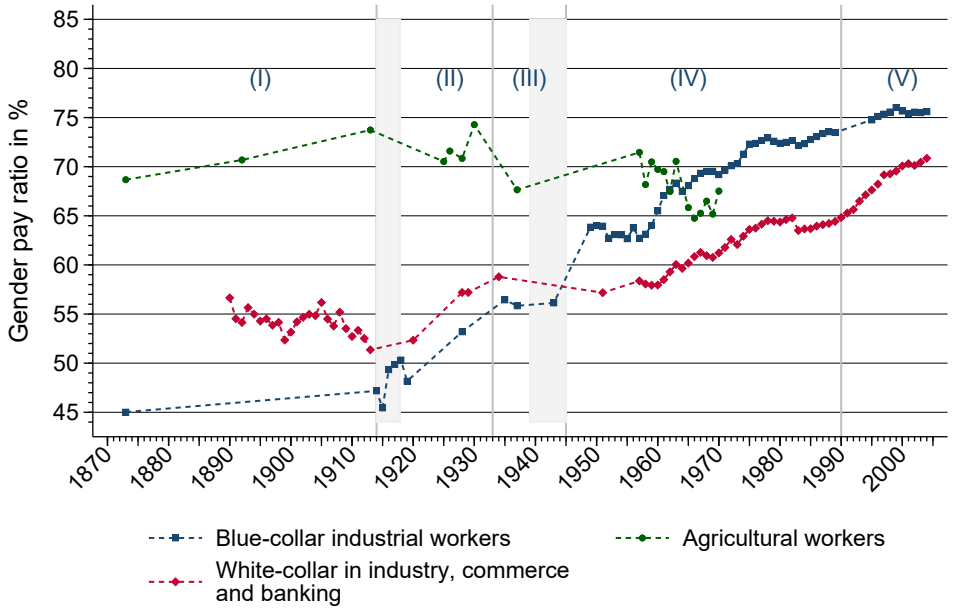
All in all, the most dynamic increase in the overall gender pay ratio across occupations and the pay ratios in blue- and white-collar work are seen in the first half of the 20th century. However, only a small number of women participated in the labor market back then. The postwar period brought slower growth of the gender pay ratio, but more and more women (re)entered the labor market and earned a wage or salary.

My results are also in line with gender differences in hourly earnings described by Bry (1960, p. 12), who finds a gender pay ratio of 58% in 1938 and of 64% in 1958. For the period of 1900 to 1945, Gómez León and de Jong (2019) construct a gender ratio of annual earnings using the social tables method. They find a deterioration of the gender pay ratio from about 59% in 1907 to 53% in 1920. This stands in contrast to my estimate of an increase from about 46% in 1913 to 49% in 1919. For the 1920s and 1930s they estimated a slow but steady increase of the gender pay ratio to 58% in 1937, very similar to my estimate in this year. Given the strong migration of women from low-paid agriculture to higher-paid white-collar work and women's increasing skill levels (see section 3.4.1 and 3.4.2), a deterioration of the gender gap seems unlikely for the 1910s. Differences could arise due to the use of yearly vs hourly gross earnings. My estimates on gender earnings differentials are based on more granular data for the different occupational groups, while Gómez León and de Jong include more occupational groups such as domestic servants.[16]

[15] For men, the positive selection bias was smaller but also increased over time.

[16] They base their analysis on non-gendered annual earnings from Hohls (1991) and Hoffmann (1965), weekly earnings from Statistisches Reichsamt (1913-1950). Gender differentials in hourly and weekly earnings are taken from Bry (1960). The authors use extensive imputations of average earnings levels but also gender differentials. They, for example, impute the gender differentials in agriculture from industry, while I find a much higher gender pay ratio in agriculture than in industry and a decrease in the pay ratio in this sector only with the rising mechanization and specialization in the postwar period. Also, they impute earnings of domestic servants as 30% of earnings in services.

Figure 3.3.: Gender pay ratios within three broad occupational groups 1873-2004



Note: Own calculations based on data listed in the data appendix. 1950-2004 based on the structure of earnings survey (*Verdienststrukturerhebung*). Reunified Germany from 1995 for blue-collar wage receivers and from 1998 for white-collar salaried personnel. No gendered data available for agricultural workers since 1971 when this group comprised less than 2% of the workforce. Details in the data appendix.

3.3.2. International comparison

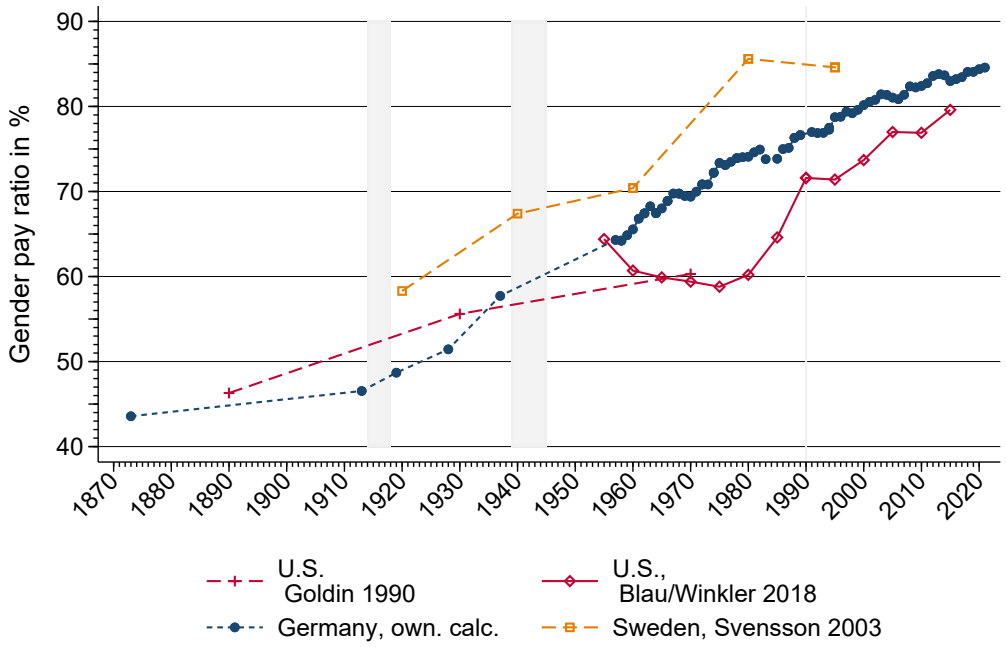
Comparing the German case to the U.S. (Blau and Winkler, 2018; Goldin, 1990) and the Swedish (Svensson, 2003) experiences uncovers similarities in developments and influencing factors before 1945 and quite substantial divergences in the postwar period.

In all three countries, the gender pay ratio rose substantially between 1890 and 1940. Despite different timings, the benchmark narrative of Goldin (1990) for the United States holds for all countries: the leap towards pay equality can be generally linked to a substantial entry of young women into better-paid white-collar and service jobs propelled by a broader access of young women to secondary education. However, in the German case, vocational on-the-job training instead of general secondary schooling was the main educational channel (see section 3.4.1).

The postwar period shows significant dissimilarities between the United States and the European experiences. In the United States, the gender pay ratio stagnated from the 1950s to 1980 and fell substantially behind the European developments. However, in the 1980s women overtook men in college completion rates, entered the labor market as high-skilled employees, chose the career path and increased the gender pay ratio by over 10 percentage points in just one decade (Goldin et al., 2006).

In Germany and Sweden, the upward trend of the gender pay ratio started in the early 1960s and stagnated after 1980. However, also between the European countries, the gender pay ratios diverged in the postwar period. In 1960 Germany's gender pay ratio stood at almost 66%. Sweden's pay ratio was only marginally higher at 70% in 1960 but quickly took off thereafter so that Sweden became the world leader in gender equality in the labor market by 1980. The narratives about how the Swedish equality miracle happened diverge: The main narrative for

Figure 3.4.: Germany's gender pay ratio in international comparison



Note: Germany and Sweden (Svensson, 1997) are based on average hourly earnings. Gender earnings ratio of reunified Germany since 1994. Goldin (1990) 1890-1970 constructed the weighted average of annual earnings of six occupational groups. Blau and Winkler (2018) show the median gender pay ratio of annual earnings. Gender pay ratio of full-time employed. More details in data appendix table B.5.6.

Sweden's swift equalization proposes institutions, particularly the strong power of the labor unions, wage policies, and ideological change as the main levers. The 1960s and 1970s were the heyday of the labor unions' solidarity wage policies. Unions and employers agreed upon the abolishment of gender-separated collective wage groups and instituted the "equal pay for equal work" norm in the 1960s. Additionally, Sweden changed to a system of separate taxation for spouses in 1971. Lastly, it massively expanded public affordable childcare in the 1970s and ran an ideological campaign to boost acceptance for working mothers and public childcare (Gärtner, 2014; Svensson, 2003). Gustafsson and Löfström (1991) conclude that about 50% of the observed increase in the gender pay ratio in industrial work can be attributed to policy action. Gärtner (2014) names diverging gender norms concerning the social acceptance towards working mothers as one of the main factors for the observed differences between Sweden and Germany. Svensson (2003) challenges these narratives by pointing out that the upward trend emerged before the introduction of the discussed policies. He claims that the increase in the gender pay ratio already took place between 1960 and the mid-1970s and was fueled by foreign competition and excess demand for unskilled female labor in the Swedish industry which seized after the structural crisis of the 1970s.[17] He concludes that market forces and not institutions have propelled the equalization process of the postwar period.

[17] Svensson (2003) assumes fixed relative earnings throughout this period.

3.4. Potential drivers

The labor and gender economics literature has highlighted several "traditional" factors influencing the gender earnings ratio over the 20th century, such as differences in education and work experience, female labor force participation, occupational, industry and firm segregation, the gendered division of labor within families and discrimination (Blau and Kahn, 2017).[18] This section discusses three key influencing factors of the gender earnings ratio in the past 150 years in comparison between Germany, Sweden and the United States: 1) educational expansion, 2) occupational and sectoral segregation and how migration between sectors and occupations influences the relative female earnings, and 3) labor force participation, particularly of married women.

3.4.1. Education

Educational expansion and women's catch-up with men's educational attainment has been discussed as one of the main drivers for the equalization of men's and women's earnings from a historical perspective (Blau and Winkler, 2018; Goldin, 1990; Goldin and Polachek, 1987).

In the U.S. case, women have benefited from two waves of educational expansion since the late 19th century: First, due to the expansion of high school in which women swiftly overtook men in the early 20th century (Goldin, 1990) and second, when women overtook men in college graduation in the 1980s (Goldin et al., 2006). While the expansion in women's secondary schooling in the early 20th century and the parallel increase in the gender earnings ratio can be observed in all three countries, the effect of the tertiary education expansion on the pay ratio is only visible in the United States.

During the first decades of the 20th century, women's increasing secondary schooling was met with the expansion of the white-collar sector in all three countries. This broadened employment opportunities for skilled women and increased returns to education. While Goldin (1990) emphasizes that the rising gender skill and earnings ratios within the white-collar sector were the predominant factors increasing the overall earnings ratio across occupational groups, for Germany educational convergence was slower and the migration between the three broad occupational groups was more important for the increasing gender earnings ratio (as I show in subsection 3.4.2).

While the secondary schooling expansion in the United States and Sweden took the form of school education, in Germany this expansion took place in vocational on-the-job training. Despite laws obliging youths of all genders to attend a secondary school (incl. vocational schools), the educational convergence was much less pronounced in Germany than in the United States. This might be due to the nature of vocational training which is traditionally centered around apprenticeships. I argue that the observed difference in educational gender convergence between both countries can be made comprehensible based on the human capital model and the theory of statistical discrimination.

The secondary schooling expansion happened quite parallelly in Germany, Sweden and the United States in the first decades of the 20th century. In Germany, changes in trade regulations (*Reichsgewerbeordnung*) before World War I extended compulsory vocational training to female

[18] In recent years, factors such as gender norms, differences in psychological attributes (e.g., risk aversion) and noncognitive skills are getting more attention to understand the still remaining unexplained part of the gender earnings gap better (Blau and Kahn, 2017).

industrial workers.[19] The Weimar Constitution of 1919 postulated compulsory attendance of a secondary school, including vocational schools, until the age of 18 years. These policy changes had the potential to increase women's education. However, their effectiveness is debated due to a loss of parliamentary majorities in 1920 and the lack of standardization across federal states (Flora, 1975; Herrmann, 2006, p. 111). Still, in the decades following the reforms, women's numbers and their share in the male-dominated part-time vocational schools and the newly emerging full-time vocational schools increased (figure 3.5). The United States were the front-runner in the expansion of generally educating secondary schools in the early 20th century. The so-called high school movement increased the supply of secondary schools throughout the country. Between 1910 and 1940, the secondary schooling rate of 18-year-olds increased from 9% to 50% (Goldin and Katz, 2009). In 1927 Sweden passed an educational reform opening public secondary schools for young women and thus effectively broadened their educational attainment.[20]

This increase in educational attainment was met in all three countries with a swift expansion of service and white-collar occupations which increased returns to education for these young, trained women,[21] while returns to education beyond literacy remained marginal in blue-collar work (Goldin, 1990, p. 106). In Germany, the share of women in white-collar office and sales jobs tripled from 1895 to 1939 (see section 3.4.2, figure 3.9). In the United States, white-collar office work was the single most important occupational group among white women after 1930.[22] In Sweden, new work opportunities for skilled female workers opened not only in the emerging commercial sector but also in the public sector, which only opened for female employees after an administrative reform in 1923 (Svensson, 2003).

Thus, in all three countries, increases in the gender earnings ratio emerged from the interplay between the rise in secondary education (incl. school education and vocational training) and the expansion of white-collar occupations which increased the return to secondary education. Goldin (1990, p. 70) argues that for the United States the increase in the gender skill ratio and the gender earnings ratio *within* the expanding white-collar occupations was the main driver pushing up the gender earnings ratio in the interwar period. In contrast, for Germany, advances in education lagged behind the U.S. development. For the still observed increase in the gender earnings ratio in the first decades of the 20th century, the migration from low-paid agriculture to higher-paid white-collar jobs was more important, which I will discuss more in section 3.4.2.

While the secondary schooling expansion in the United States and Sweden took the form of school education, in Germany vocational on-the-job training expanded. The most common way of becoming a skilled worker for most of the 20th century was by obtaining an apprenticeship in a firm, often accompanied by attendance of a part-time vocational school. However, in the early 20th century only very few young women could access this track (Mayer, 1999, 39f.). This observation can be interpreted along the lines of a model of statistical discrimination (Barron et al.,

[19] Novelle der Reichsgewerbeordnung vom 27.12.1911, §111.

[20] Female graduates increased their share "from 35% to 45% in lower secondary school and from 25% to 32% in upper secondary school" between 1913 and 1940 (Svensson, 2003).

[21] Goldin and Polachek (1987) argue that these increased returns to education were crucial for the increase in the gender earnings ratio between 1890 and 1970. Goldin (1984) also shows for the United States that both developments did not solely run parallel but about one-third of the occupational shift in the female labor force from manufacturing jobs to office work was due to increased high school education which provided the necessary skills to enter these jobs.

[22] The most affected cohort of U.S. women born between 1900 and 1920, did not only have higher skill levels but entered office jobs in greater numbers often directly after high school and stayed longer on their jobs than previous cohorts, even after marriage, increasing their work experience. These women were also the first cohorts that came back to the booming labor market of the 1950s and 1960s in large numbers and shaped the pattern of increasing labor force participation over the life cycle (discussed in section 3.4.3).

1993) based on the human capital model (Becker, 1964): In school education, which was heavily expanded during the U.S. high school movement, future employees (and the schooling institution) bear the costs of education. Germany's apprenticeship model is centered around on-the-job training, in which employees and employers jointly bear the costs. In such an arrangement high turnover is particularly costly for employers (Becker, 1964). Working women in the first half of the 20th century commonly quit their employment upon marriage, i.e., on average at the age of 25 years (figure 3.21). In a model of statistical discrimination, employers who perceive women on average as less stable employees with a higher quit rate than men would rather invest in young men with longer employment horizons to increase the returns to their offered on-the-job training and sort women into jobs that require less training (Barron et al., 1993). This would explain why women had little access to this path in the early 20th century.

The emergence of new school types and policy changes in the early 20th century had only limited potential to change this. First, the Weimar Constitution of 1919 postulated compulsory attendance of a secondary school for all youths irrespective of their gender but did not change incentives for employers to offer more apprenticeships to women. Still, women's share in the formerly male-dominated part-time vocational schools (which were the complement to apprenticeships) tripled from 10% in 1915 to 31 % in 1938 (figure 3.5, panel a). Second, as a substitute for the male-dominated path, full-time vocational schools for women, mainly focusing on commercial education or housekeeping tasks, emerged in the first decades of the 20th century (Herrmann, 2006). The rise of these full-time schools might have decreased women's entry barriers resulting from statistical discrimination for on-the-job training by providing school education for which the costs were entirely borne by students. Young women constituted the majority in this track of full-time vocational training. Nonetheless, these schools were until 1938 still marginal compared to part-time vocational schools (Herrmann, 2006 and figure 3.5, panel (b)). General secondary schooling (*Abitur*) levels remained low in the interwar period (fig 3.7).

This rather sluggish increase in women's secondary education stands in stark contrast to the United States where women outperformed men in high school graduation already in the first decades of the 20th century (National Center for Education Statistics (NCES), 2023)[23] The dominance of firm-centered vocational training might have inhibited women's catch-up with men in secondary schooling in Germany.

Thus, the specific design of Germany's education system might have inhibited a secondary schooling convergence for women comparable to the United States and Sweden. Also, it might have had long-run implications beyond the interwar period:

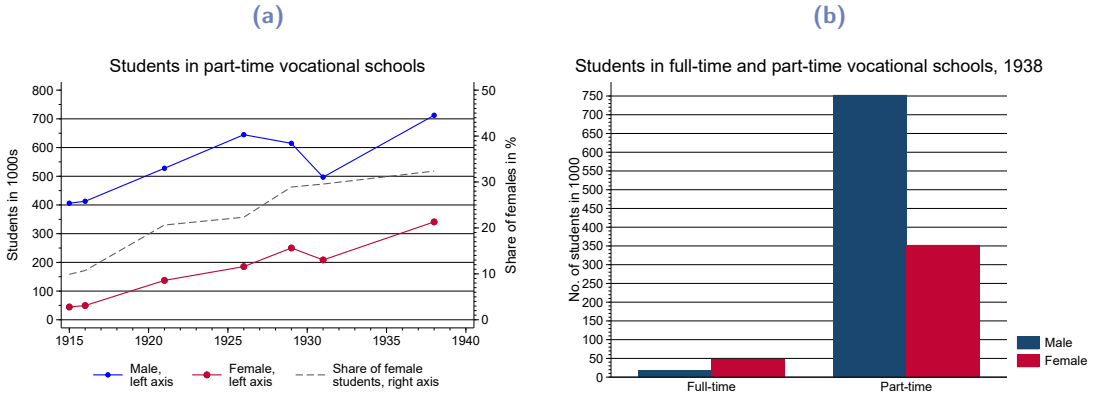
A strong specialization concerning the schooling content remained in Germany, while high school education in the United States was more general, even though not completely gender-neutral. While women overtook young men in the attendance of commercial schools transmitting training for sales and office personnel or attended vocational household tracks, solely installed for females to become kindergarten teachers, professional housekeepers or housewives, men stayed throughout in the majority in industrial schools training for specific skilled occupations in

[23] Already in the 19th century, women outperformed men in high school graduation because young men usually entered the labor force earlier. However, women's advantage only became relevant with the massive expansion of high school education, when more than a marginal fraction of the U.S. population could access high school education. (Goldin and Katz, 2009; National Center for Education Statistics (NCES), 2023).

blue-collar work (figures 3.6 and 3.15). This might have reinforced the long-term strong gender segregation observed in the German labor market (Müller et al., 1983).[24]

Vocational schools dominated the German schooling system long into the postwar period. General secondary schooling (*Abitur*), allowing for the attendance of universities, only increased late in the postwar decades. In Germany, women overtook men in high school graduation in the early 1980s (see figure 3.7) and in university enrollment only in the 2010s (figure 3.16 and Destatis, 2023). In the United States, the gender pay ratio saw an unprecedented increase of more than 10 percentage points in the 1980s linked to a second wave of women's educational expansion. Young U.S. women overtook men in college enrollment rates in the 1980s, chose more career-oriented majors and significantly increased their presence in managerial jobs (Blau and Kahn, 2017; Goldin et al., 2006). A similar second wave of educational expansion and its effect on the gender earnings ratio comparable to the United States is not visible, neither in Germany nor in Sweden in the postwar period. It remains open for further research if this is due to the later expansion of general secondary and tertiary schooling in Germany or different complementary factors.[25]

Figure 3.5.: Female and male students in part-time and full-time vocational schools (Berufsschulen)

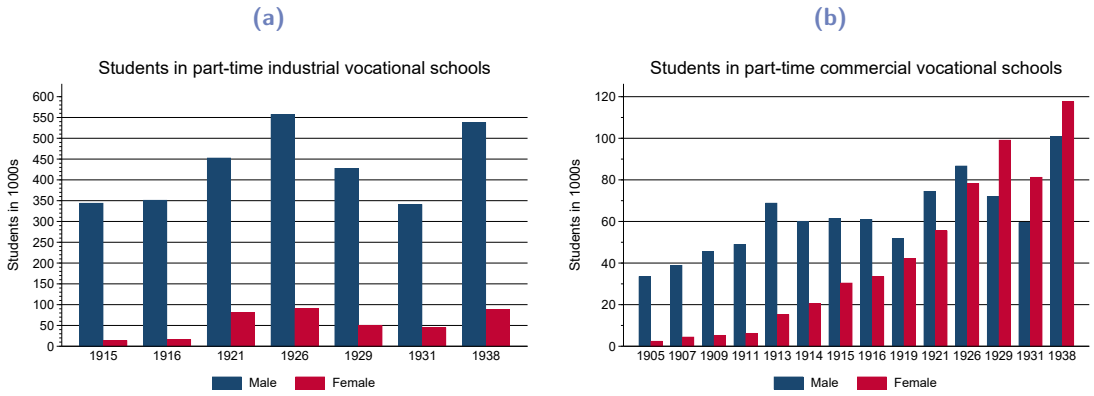


Source: Own calculations based on Herrmann et al. (2021) and Herrmann (2006).

[24] Willms (1983b) argues that women have never entered typically male-dominated occupations and industries in larger numbers, but have always entered old or newly emerging "womens' jobs".

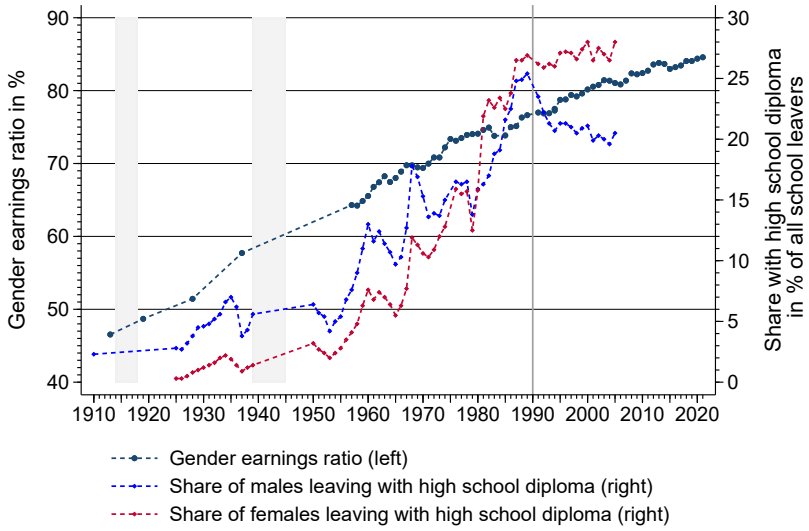
[25] Since the German vocational training system heavily relied on apprenticeships in firms, labor market conditions could influence the training level of the (female) population significantly. In an economic downturn with high unemployment, the availability of apprenticeships decreases. Müller et al. (1983, p. 97ff.) illustrates that by pointing out the declining training level of cohorts entering the market for apprenticeships in the 1940s and early 1950s. These cohorts wanted to enter vocational training in the economic recession of the late 1940s and early 1950s. Due to a lack of apprenticeships, particularly women of these cohorts show a lower average training level.

Figure 3.6.: Female and male students in part-time commercial and industrial vocational schools (Berufsschulen)



Source: Own calculations based on Herrmann et al. (2021) and Herrmann (2006).

Figure 3.7.: Female and male school graduates with high school diploma



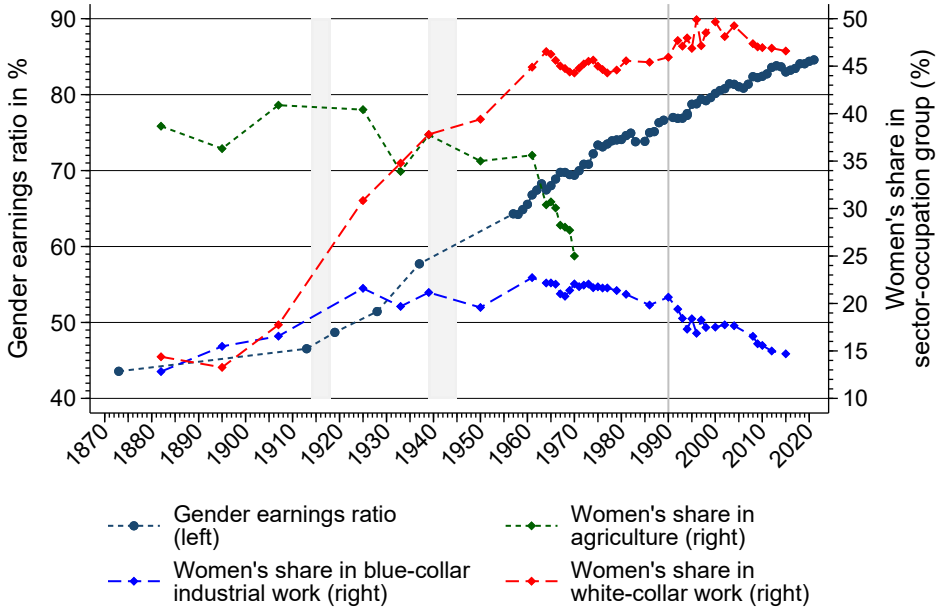
Source: Gender earnings ratio based on own calculations; reunified Germany since 1994. Education series are own visualizations based on data by Rahlf (2015), reunified Germany since 1991.

3.4.2. Sectoral and occupational migration

While pay ratios within the three broad occupational groups developed very differently (figure 3.3), the trend of the gender ratio across occupational groups has been increasing throughout the 20th century. One potent lever for this development was the migration of women workers from low-paid occupations, especially in agriculture, to higher-paid jobs in blue- and white-collar work. Figure 3.8 contrasts the share of women among all workers within the agricultural sector, blue-collar industrial work and white-collar work in industry, commerce, banking and insurance respectively (right axis) with the development of the gender pay ratio (left axis). Women moved out of agriculture at an increasing pace since the beginning of the 20th century. Women's share

in blue-collar work already peaked in the interwar period. Women never exceeded to be more than 23% of the blue-collar industrial workforce. The most dominant movement is the rapidly increasing share of women into white-collar work from 1895 to the 1960s.

Figure 3.8.: Women as share of all agricultural, industrial blue-collar and white-collar workers



Gender earnings ratio based on own calculations, reunified Germany since 1994. Sector-occupation data based on own calculations using Müller et al. (1983), Statistisches Reichsamt (1926-1943), Statistisches Bundesamt (1960-19), and Statistisches Bundesamt (1976-1997), reunified Germany since 1992.

As already touched upon in section 3.4.1, scholars agree that the strong increases in the gender earnings ratio in the first half of the 20th century were produced by the interplay between increased training and education in white-collar related tasks and the expansion of the white-collar occupations. Many many young women were drawn to these new jobs so that occupational patterns of the female employed workforce changed substantially. The main difference between the United States and Germany is the contribution of the factor education and the factor of migration from low-paid to higher-paid occupations for the observed increases in the gender earnings ratio. For Germany, due to a delayed industrialization, women's migration from low-paid agriculture to higher-paid white-collar work dominated the leap of the gender pay ratio from 1913 to 1928. Only thereafter increasing skill and earnings ratios *within* the three broad occupational groups become more important. For the United States, Goldin (1990, tab 3.2.) describes that the increase in the skill ratio and thus the gender pay ratio within white-collar work predominantly drives the increase in the overall gender pay ratio between 1890 and 1930. The difference arises not only because of the stronger educational convergence in the United States but also due to different migration patterns.

Figure 3.9 compares the composition of the female employed workforce across occupational groups between Germany and the United States. Due to very different databases, categories are not entirely congruent but several differences are clearly visible. The United States industrialized rapidly in the last decades of the 19th century which profoundly restructured its labor force composition. Already in 1890 only about 10% of employed women worked in agriculture, while

the largest share of women, above 30%, were employed in industrial blue-collar work, followed by personal service work (incl. domestic service). Between 1890 and 1930, U.S. women migrated from personal service and blue-collar work into office work and managerial tasks, which made up the majority of the female workforce since 1930.

Germany industrialized slower than many European countries and the United States (Ogilvie, 1996). Particularly the female labor force remained longer in the traditional and low-paid sectors while men moved swiftly to industrial work (Müller et al., 1983). In 1895 still about 35% of women worked in agriculture and about 40% in domestic service, the lowest-paid occupations of the economy, while blue-collar work made up about 15% of the female employed workforce. Between 1895 and 1933 the share of the female employed workforce in low-paid agriculture and domestic work halved, while women took on more and more jobs in blue-collar industrial and white-collar office work, which made up more than 50% of the female workforce in 1933.

These migration patterns influence the overall gender pay ratio because they decrease the relative weight $\phi_{\substack{\text{agri-}\\\text{culture},f}}$ attached to the lower average female earnings in agriculture $w_{\substack{\text{agri-}\\\text{culture},f}}$ and increase the relative weight $\phi_{\substack{\text{white}\\\text{collar},f}}$ of the higher average female earnings in white-collar work $w_{\substack{\text{white}\\\text{collar},f}}$. Further, average earnings in the agricultural sector were the lowest while blue-collar industrial workers usually stood the middle ground and white-collar service workers receive the highest average earnings.[26] Thus, German women bridged a larger earnings differential due to their migration from agricultural work to white-collar work than U.S. women mainly leaving blue-collar industrial work for the new white-collar jobs.

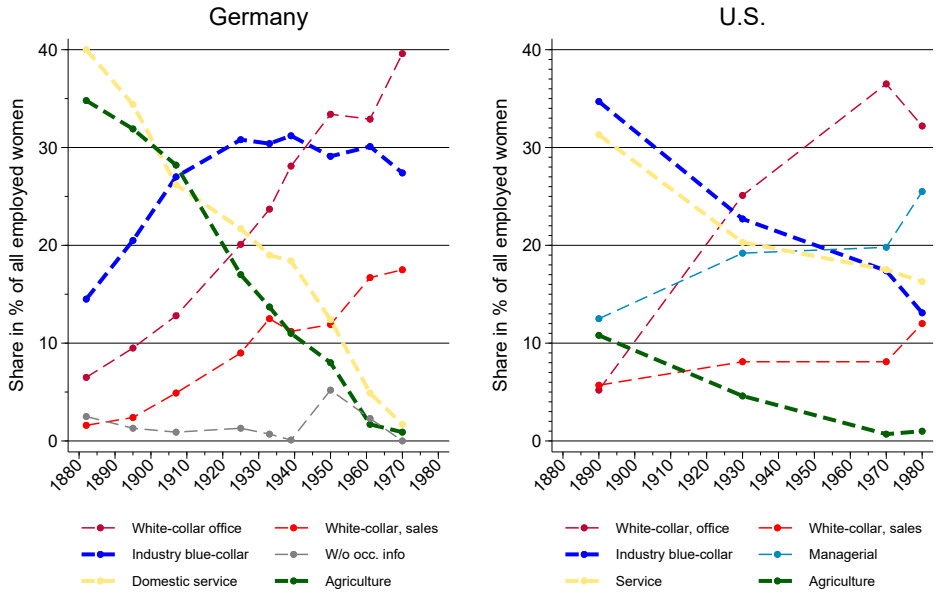
This is also confirmed when contrasting results for the United States by Goldin (1990, tab 3.2) with my tentative results for Germany (appendix section 3.6.3). Goldin's counterfactual analysis affirms that gender earnings equalization within occupational groups drove overall gender earnings equalization between 1890 and 1930 in the United States. For Germany, I find that the migration between the three occupational groups had a stronger impact on the overall gender earnings ratio between 1913 and 1928 (table 3.4).[27]

All in all, the strong upward trend in the gender earnings ratio in the United States and Germany in the first half of the 20th century can be linked to first, increases in education and second, occupational migration. However, their importance for the gender earnings ratio of the employed workforce differs between the two countries. In Germany, due to its late industrialization, the increase in female average earnings due to migration from the lowest-paid sector, agriculture, to more modern and higher-paid jobs in white-collar work increased women's relative earnings predominantly until 1928. Only thereafter, gender skill and pay ratios within the three occupational groups have become more important for the development of the gender pay ratio across occupational groups. This stands in contrast to the United States where the increasing gender pay ratio was predominantly propelled by women's increased education and an increasing gender earnings ratio within white-collar work.

[26] See Goldin (1990, tab 3.2., panel A) for the United States and table 3.1 for my analysis for Germany. It holds that $w_{agriculture,g} < w_{blue-collar,g} < w_{white-collar,g}$. Also, Gómez León and de Jong (2019, Appendix, figure S1.4 - 1.7) show that agriculture exhibited the lowest average earnings followed by industry wage earners (blue-collar) between 1900-1950 in Germany and Britain. White-collar service workers receive the highest average earnings. Further, they find that the earnings differences between these groups are more pronounced in Germany than in Britain. Which suggests a stronger effect of migration on the overall pay ratio.

[27] Between 1913 and 1937, increases in the overall gender earnings ratio benefit relatively equally from increasing within-occupation gender pay ratios and migration across occupational groups. This analysis should be seen as tentative because the reweighting between three very broad categories, comprising a great variety of occupations, cannot deliver more precise results.

Figure 3.9.: Share of female employment by occupational group



Source: Own visualisations based on Goldin (1990) & Müller et al. (1983). "White-collar office" comprises salaried personnel in industry, as well as in services like education, transport, banking, administration and hospitality for Germany. Sales employees were not split into blue- and white-collar in German data, but can be assumed to be generally white-collar.

3.4.3. Labor force participation

In this section, I discuss the different evolutions of female labor force participation in Germany, Sweden and the United States. Changes in labor force participation influence the gender earnings ratio by changing the average skill and experience level of the female workforce.

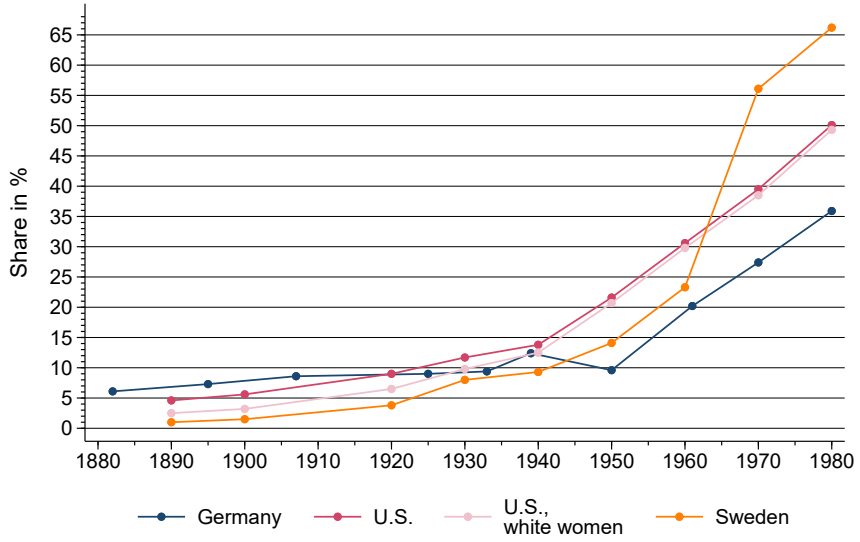
Figure 3.2 contrasts the development of the gender pay ratio with the labor force participation of working-age women (20 to 59 years). The female labor force participation was relatively stable from 1895 to 1939. This suggests a quite restricted pool of women in the labor market. In the early 20th century, the main female workers outside of the home were young, unmarried women who for the largest part dropped out of the labor market upon marriage in Germany (Willms, 1983a) as well as the United States (Goldin, 1990). The most striking evolution is the strong increase in female labor force participation in the postwar period despite a decrease in marriage age from the early 1950s until the mid-1970s (see figure 3.21). The described historical pattern holds true for Germany, the United States and Sweden (figure 3.19). But while all three countries exhibit similarly low female labor force participation rates until about 1950, the patterns diverged substantially thereafter.

The most defining event for women's labor force participation was marriage. More than 50% of single women were employed quite stably in Germany between 1882 and 1970 (Willms, 1983a, p. 33).[28] The majority of women in the late 19th and far into the 20th century exited the labor force upon marriage.[29] However, married women always constituted the majority of the working-age

[28] Only with the education expansion in the 1970s, this share dropped due to longer education trajectories.
[29] Goldin (1990, p. 16) emphasizes "marital status, more than any other characteristic save race [...] determined a woman's economic role." It "provides a clear dividing line for female labor force participation. The majority of women exited the labor force at or just after marriage."

population. Thus, patterns and changes in their labor market participation had large effects on the overall female labor force participation and the workforce's composition.

Figure 3.10.: Female labor force participation of married women, Germany, Sweden and the United States



Source: Own calculations based on Müller et al. (1983) and Goldin (1990) and Stanfors and Goldscheider (2017). Without assisting family members. The concept for Germany includes all market based activities including employed blue- and white-collar workers, civil servants, self-employed and service personnel (domestic services & maids).

Figure 3.10 shows that the labor force participation of married women was quite stably below 20% from the late 19th century until about 1940 in all three countries. While many married women worked as assisting family members, work outside of the home was a marginal phenomenon. However, these 10% of married women working outside of the home were likely to have had stable labor market attachments and thus, substantial labor market experience, an important factor for increasing earnings (Willms, 1983a). In the postwar period, similar to the patterns of the overall labor force, married women's labor supply increased strongly, but with distinct country patterns. Thus, the gender earnings ratio's evolution of the postwar period might have been significantly influenced by married women's labor force decisions.

The postwar period first shows a divergence in labor force participation rates from 1940 to 1950, introducing a wedge of about 10 percentage points between the United States and Germany. The upward trends between the two countries evolve quite in parallel thereafter with increases in the participation rate of around 10 percentage points per decades since 1940 (in the United States) and since 1950 in Germany. Sweden then took the lead in married women's labor force participation in the 1970s. Here, as for the gender pay ratio, discussed in section 3.3.2, stands the question if this increase was mainly influenced by excess demand for female labor in the industrial sector in the 1960s (Svensson, 2003) or due to policy action (Gustafsson and Löfström, 1991; Svensson, 2003).[30]

[30] Policies include the solidaristic labor union policies of the 1960s, the transition to individual taxation of married couples in 1971 and strongly increased publicly provided and subsidized childcare for preschool children in the 1970s.

What was the crucial difference that created a divide between the labor force participation rates of married women in the United States and Germany and produced very different labor force attachments of parallel cohorts in the two countries?

Several explanations for the strong increase in married women's labor force participation are outlined in the literature, mainly focused on the United States: 1) Goldin (1990) shows that a shift in labor demand in occupations where women were well represented was crucial for the initial shift between 1940 to 1960 and remained important thereafter. 2) Blau and Kahn (2007) show that in later postwar years, when women's average earnings were rising while men's were stagnating, the substitution effect due to increases in female wages outweighed the negative income effect due to increasing husband's earnings on married women's labor supply. 3) Further, labor supply effects due to the expansion of substitutes for home production and better household technology might have made married women's work outside of the home possible. 4) Blau and Kahn (2017) point out that preferences such as a stronger labor market attachment of women and norms might explain the residual part.

For Germany, several scholars argue that gender norms towards mothers reinforced the male breadwinner model in the postwar Federal Republic of Germany (Gärtner, 2014). However, both Willms (1983a) for Germany and Goldin (1990) for the United States find a pattern of a loosening stigma towards working wives already in the interwar period. For Germany, Müller (1983, 74f.) argues that married women increased their labor market participation due to the mobilizations of the Nazis. The marriage cohort of 1940 was the first to break the norm of immediate resignation from work upon marriage. For the United States, Goldin (1990) finds prolonged employment beyond the date of marriage emerging with the rise of office work in the 1920s and 1930s. Both authors argue that these cohorts of women, experiencing double-earner households, were the first shaping the newly emerging pattern of married women reentering the labor market after an intense childcare period visible in the postwar period. Thus, the basis for increased acceptance of working wives was set in both countries before 1945. Still, gender norms might have differed between the two countries.

The wedge between German and U.S. married women's labor force participation might be well explored based on Goldin's argument of increased labor demand. For the United States, Goldin (1990, p. 136ff.)[31] finds an outward shift in labor demand for occupations in which females were well represented between 1940 and 1960. This happened in the context of an economic boom and a labor squeeze which forced employers to offer more jobs for married women. The increased labor demand was met with an elastic labor supply function of married women. Goldin (1990) concludes that this interplay can explain the largest part of the observed increase in married women's labor market activity.

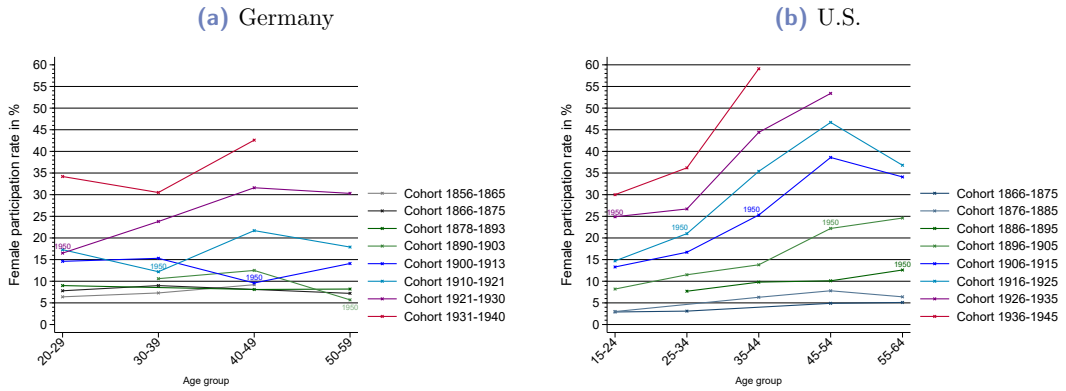
However, the economic conditions diverged fundamentally in Germany in the immediate postwar period. Women of the same birth cohorts in Germany could not reenter the labor market in the early 1950s due to a severe economic crisis with high unemployment of more than 10% (see figure 3.22).[32] Thus, the labor demand shift, crucial in Goldin's argument, was substantially delayed in Germany. Only around 1960, the FRG reached full employment and married women reentered more and more the labor market.

To support this argument, figure 3.11 compares married women's labor force participation

[31] Goldin builds on Mincer (1962).

[32] Willms (1983a) argues that many women who were looking for work were so discouraged by the severe labor market conditions that they declared themselves as housewives despite their ambitions to work outside the home.

Figure 3.11.: Labor force participation rate of married women across the life cycle, Germany and United States



Source: Germany, own calculations based on Willms (1983a). United States based on Goldin (1990, table 2.2.). Cohort profiles are constructed from information about certain age groups in different census years. Due to the irregularity of censuses in Germany, cohorts are sometimes broader and slightly overlap.

over their life cycle between Germany and the United States. A clear difference between the two country patterns centers around the year 1950. While U.S. married women's labor force participation is higher in 1950 than in their life decade previous to 1950, the opposite is true for all cohorts of German women, whose labor force participation decreases in the life decade around 1950. Thus, the early 1950s have increased labor force participation for U.S. women in all birth cohorts between 1900 to 1935 and may have aided the formation of the distinct pattern of increasing labor force participation over the life cycle in the United States. This means more married women worked in their 30s and 40s than in their 20s - a reversal of the historical stereotype of young working women of the prewar period. Contrary, for Germany, the early 1950s depressed labor force participation of all birth cohorts between 1890 and 1930 in different stages of the life cycle. This suggests that the decreased labor force participation was not entirely due to free choice and norms. It still remains up for further research why this gap between the U.S. and German labor force participation did not close in subsequent decades.

Overall, labor force participation rates of married women increased substantially in Germany, Sweden and the United States since the 1950s. However, the pace and patterns have differed. These different patterns might have shaped the average skill and experience level of the female workforce differently and in consequence the evolution of the gender earnings ratio.

In the U.S. benchmark narrative, Goldin (1990, p. 24) argues that due to the strong increase in the labor force participation of married women, who came back to work after a long absence and with low work experience, the average earnings potential of the female workforce was depressed. This translated into a stable gender pay ratio from 1950 to about 1980. Germany similarly shows a strong increase in married women's reentry into the labor market, however, with a delayed onset in the 1960s and a persistent gap compared to the U.S. participation rate of about 10 percentage points since then. If this delayed movement of less experienced women into the labor market might have exerted less downward pressure on the average experience level of the female workforce and made moderate pay ratio increases possible, is still up for exploration. Sweden's active policy to support married women's labor force participation from the 1970s onwards likely

shortened the period of skill depreciation because mothers could reenter the labor force even with young children. Similar policy action was not visible in Germany or the United States.

3.5. Conclusion

This article provides the first long-run time series of the gender earnings ratio for the full-time employed workforce from 1871 to 2021, discusses possible drivers of the observed dynamics and compares the German path to the U.S. and Swedish cases. The gender earnings ratio has increased substantially from about 44% in the 1870s to about 85% today. The most prominent leap is observable between 1913 and 1937, while the postwar period was characterized by slower growth.

In international comparison, Germany, Sweden and the United States show similar developments in the first half of the 20th century. In all three countries, strong increases in the gender earnings ratio can be linked to women's educational expansion and higher employment possibilities and returns to education in the expanding office and sales occupations. In the United States, increasing gender skill and earnings ratios in white-collar work were pushing the gender earnings ratio of the entire workforce upwards. German women's convergence in secondary schooling was substantially delayed compared to the United States, possibly due to a focus on vocational on-the-job training. While general public schooling as in the United States posed little entry barriers for women, entering an apprenticeship presented an obstacle for young women often assumed to have a high turnover rate. Thus, education was a less potent lever for the gender earnings ratio in Germany compared to the United States. Due to a delayed emergence of a modern industrial labor market in Germany, young women left the traditional low-paid agricultural sector later than their U.S. peers and directly migrated into higher-paid white-collar work. This was the main lever propelling the gender earnings ratio in the first half of the 20th century in Germany.

The evolution of the gender earnings ratio and labor force participation patterns diverged substantially between Germany, Sweden and the United States in the postwar period. Married women reentered the labor market in greater numbers in all three countries. Due to a severe crisis in the early postwar years in Germany, married women's labor force participation started rising a decade after their U.S. peers. The reentry of married women with little work experience and outdated training has most likely slowed growth of the gender earnings ratio in Germany and brought it to a standstill between 1950 and 1980 in the United States.

It remains open why the United States, despite being the leader in terms of a modern labor market due to a more rapid industrialization and an earlier general secondary schooling expansion, show a similar or lower level of the gender pay ratio than the European countries. One reason might be the higher labor earnings inequality in the United States which mechanically decreases the gender pay ratio (Blau and Kahn, 1996).[33] Another explanation may be the overlapping dynamics with the race earnings gap.

Overall, the German path towards gender earnings equality in the last 150 years is characterized by a delayed emergence of a modern labor market offering higher-paid jobs in blue- and white-collar work for women and a delayed educational convergence between women and men compared to the United States. In comparison to Sweden, the absence of effective policy action to support women's labor supply stands out in Germany.

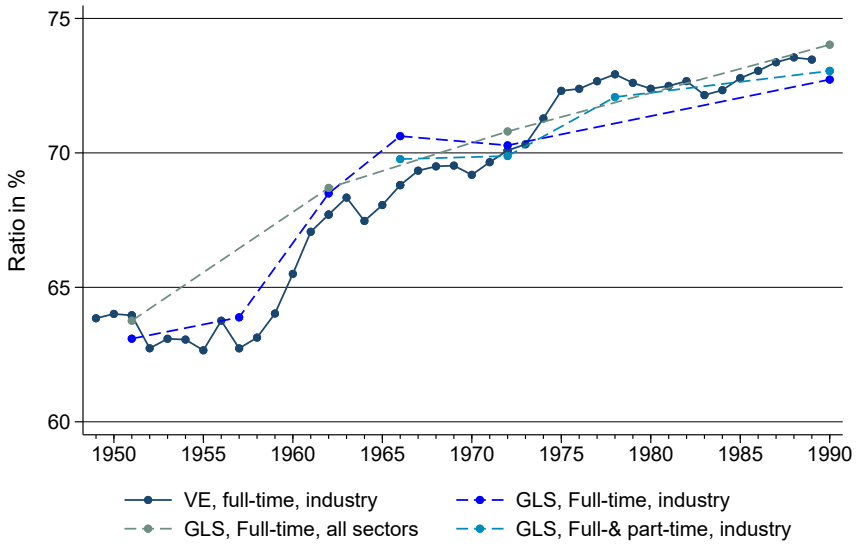
[33] The authors argue that the dispersion of the earnings distribution is one crucial factor for international differences in gender pay differences.

3.6. Appendix figures & tables

3.6.1. Robustness

Comparison of the structure of earnings survey (Verdienststrukturerhebung) and the salary and wage structure survey (Gehalts- und Lohnstrukturerhebung, GLS): To construct the benchmark series for the post-World-War-II period, I draw on the yearly structure of earnings survey (*Verdienststrukturerhebung, VE*). The *VE* is a yearly standardized survey since 1957 for blue- and white-collar workers.[34] It is based on earnings totals of specific worker groups. Blue-collar workers' earnings were surveyed in mining, quarrying and industry. White-collar workers were additionally surveyed in commerce, banking and insurance. The *VE* only comprises full-time workers. In comparison, the *GLS* is an individualized survey comprising blue- and white-collar workers in mining, quarrying, industry, commerce, transport, banking and insurance as well as selected services. From 1966 onwards, it comprises part-time workers in several survey waves, which became a prominent group among women from the 1950s onwards. Problematic about the GLS are 1) its sporadic availability every six to ten years and 2) its changing composition of industries. Different industries and sectors were surveyed in different waves. Part-time workers were only irregularly included (see data appendix). Figure 3.12 shows that the development and level of the gender pay ratio for blue-collar industrial workers is comparable for both data sources.

Figure 3.12.: Gender pay ratio of blue-collar industrial workers based on the structure of earnings survey (benchmark) and the salary and wage structure survey (*Gehalts- und Lohnstrukturerhebung, GLS*)

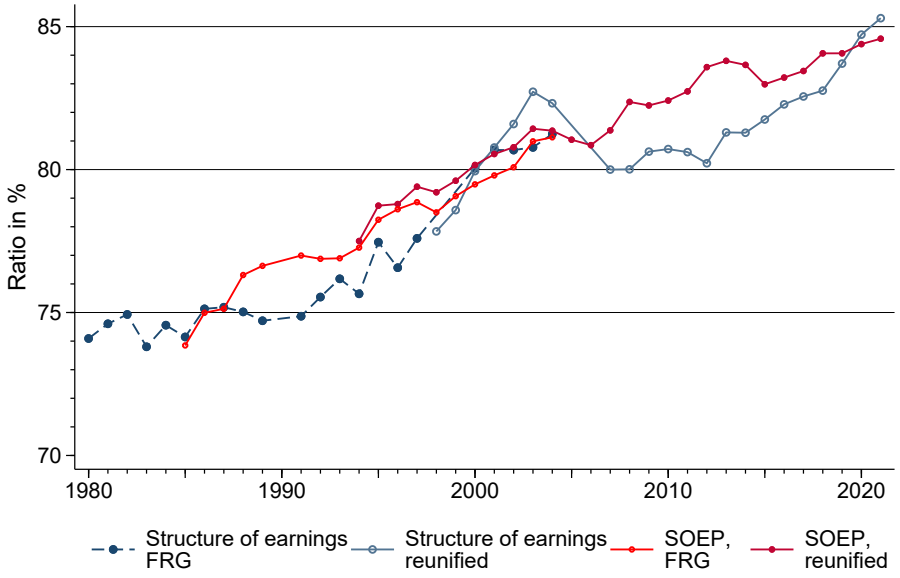


Comparison with the Socio-Economic Panel (SOEP): From 1984 to 2021, I use the Socio-Economic Panel because of its comprehensiveness. To ensure that this change in the database does not introduce a break in the time series, I contrast this new benchmark with the time series based on the structure of earnings survey (VE). Figure 3.13 compares the VE and SOEP series for the Federal Republic of Germany (FRG, West Germany) and reunified Germany.

[34] The statistical office of the FRG took up the methodology already used since Statistisches Reichsamt (1935) in 1946, but first only for blue-collar workers in a small range of industries. (see data appendix for more documentation).

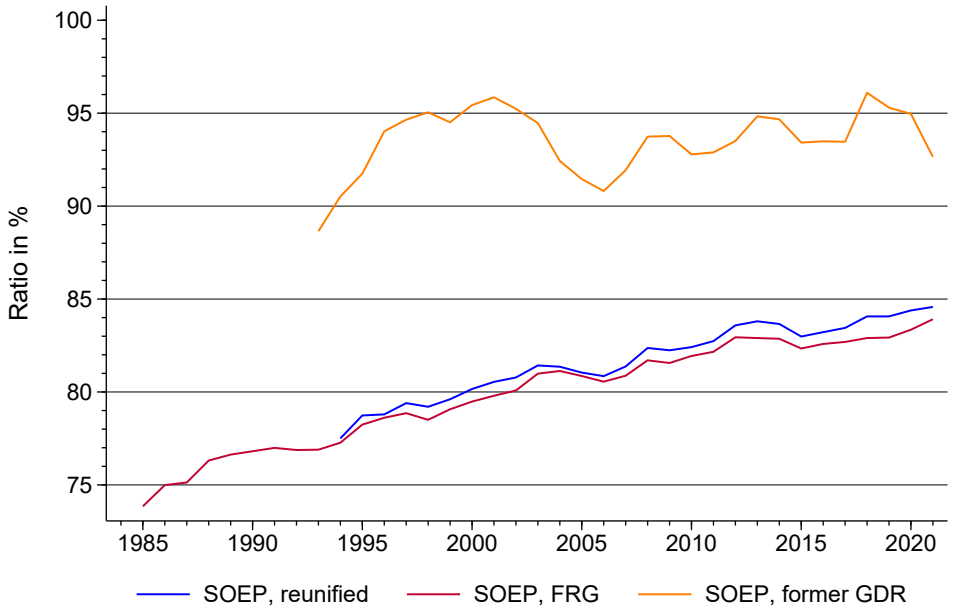
The SOEP series connects very well to the VE series in the mid-1980s with differences of less than one percentage point. The biggest deviation between the two series is visible from 1989 until 1992 when the gender earnings ratio based on the more comprehensive SOEP (including modern service occupations) stands two percentage points higher than based on the VE series. Thereafter both series deviate by 1 to 2 percentage points from each other but show the same trend. For reunified Germany, the VE only provides integrated data since 1998 while SOEP has integrated East German data from 1990 onwards. Due to data issues, I use SOEP data for reunified Germany from 1994 onwards. From 1998 until 2006, differences are negligible. From 2007 onwards, the SOEP-based gender pay ratio shows a significantly higher level of about two percentage points. I attribute those differences to the higher coverage of the dataset.

Figure 3.13.: Gender pay ratio based on SOEP (benchmark from 1985) and the structure of earnings survey.



Note: Restricted to full-time employed, SOEP treated with 3-year-MA process.

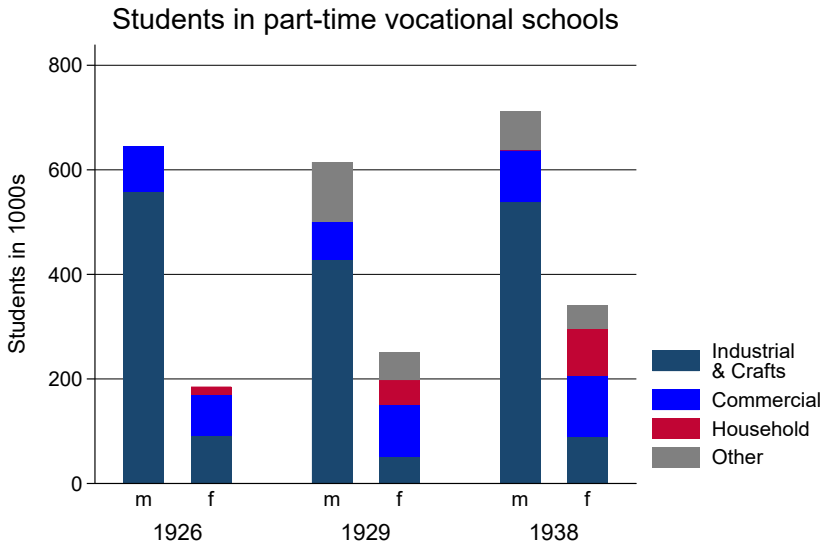
Figure 3.14.: Gender pay ratio based on SOEP (benchmark from 1985) for former FRG, GDR and reunified Germany



Note: Time series smoothed via a 3-year moving average.

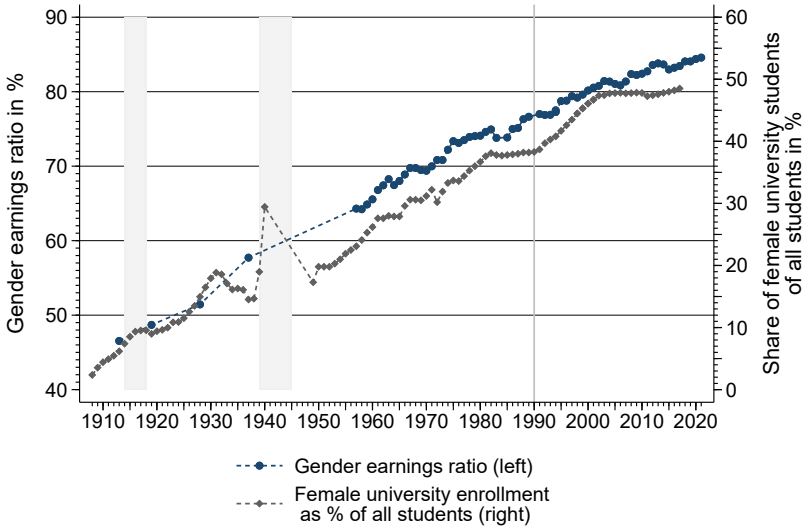
3.6.2. Education

Figure 3.15.: Composition of schooling content for female (f) and male (m) students in part-time vocational schools



Source: Own calculations based on Herrmann et al. (2021) and Herrmann (2006).

Figure 3.16.: Women's share among university students.



Source: Gender earnings ratio based on own calculations; reunified Germany since 1994. Education series based on data by Rahlf (2015), reunified Germany since 1991.

3.6.3. Occupational segregation

Table 3.1.: Gross hourly earnings in current Reichsmark

Occupational group	1913		1919		1928		1937	
	Male	Female	Male	Female	Male	Female	Male	Female
Agriculture	0,20	0,15	1,03	0,72	0,31	0,22	0,40	0,27
White-collar	0,69	0,35	1,79	0,94	1,28	0,73	1,26	0,71
Blue-collar	0,49	0,24	1,69	0,75	1,08	0,56	0,86	0,48

Table 3.2.: Distribution of employees across the three broad occupational groups

Occupational group	1913		1919		1928		1937	
	Male	Female	Male	Female	Male	Female	Male	Female
Agriculture	18,1%	39,6%	15,6%	32,2%	13,6%	23,7%	12,0%	19,3%
White-collar	13,5%	12,9%	16,2%	18,1%	18,3%	25,0%	17,4%	28,6%
Blue-collar	68,4%	47,4%	68,2%	49,7%	68,2%	51,3%	70,6%	52,0%

Table 3.3.: Gender pay ratio within three broad occupational groups

Occupational group	1913	1919	1928	1937
Agriculture	73,7%	70,5%	70,8%	67,6%
Blue-collar	44,0%	44,4%	51,5%	55,8%
White-collar	51,4%	52,3%	57,2%	56,2%

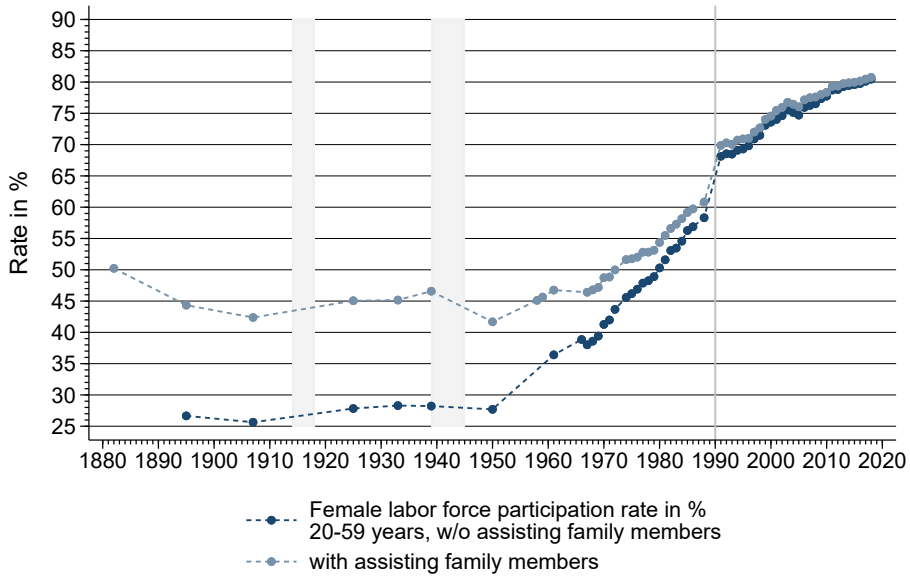
Table 3.4.: Gender pay ratio across all three occupational groups and counterfactuals

	Occupation 1913	Occupation 1919	Occupation 1928	Occupation 1937
Earnings 1913	**46,5**	48,0	50,3	51,6
Earnings 1919	48,3	**48,4**	48,9	49,1
Earnings 1928	46,1	48,4	**51,4**	53,0
Earnings 1937	51,3	53,2	55,9	**57,7**

Note: This table shows the observed and counterfactual gender earnings ratio across all occupational groups of the sample population. The observed ratio is highlighted in gray. Note that based on equation 2, the female (male) earnings are the weighted average of occupation-specific female (male) earnings. The observed gender earnings ratios are based on the earnings and occupational-group weights of the same year, while for the counterfactuals the year of the occupational distribution and the occupation-specific earnings diverge. Columns keep the occupation and sector distribution constant for the indicated year, while rows keep the gendered average earnings of the occupational groups constant for the indicated year. Holding the occupation-specific average earnings constant at the 1913 level and shifting the occupational distribution to 1928 shows a larger increase than keeping the occupational distribution stable at 1913 and varying the earnings to the 1928 level. Thus in this period, changes in the occupational distribution seemed to have played a more dominant role.

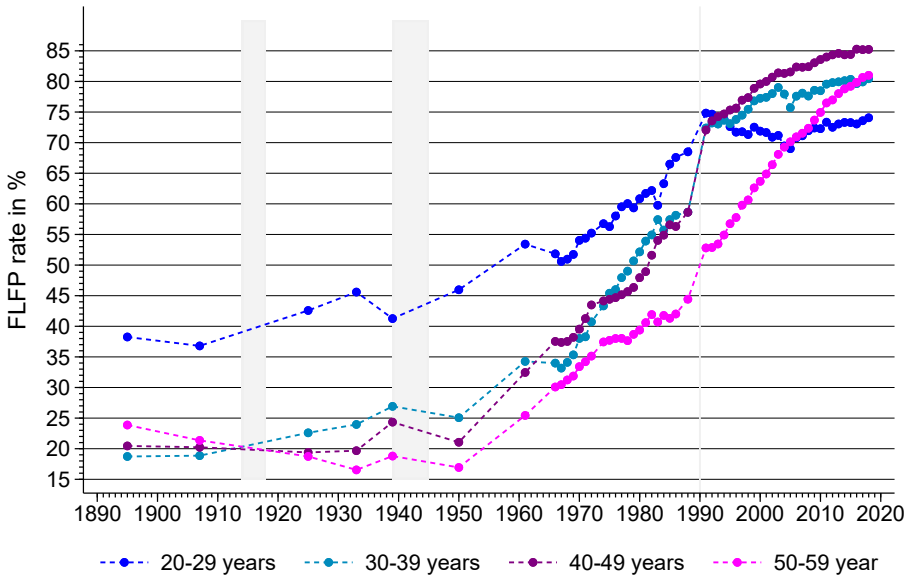
3.6.4. Labor force participation

Figure 3.17.: Female participation rate (with and without assisting family members)



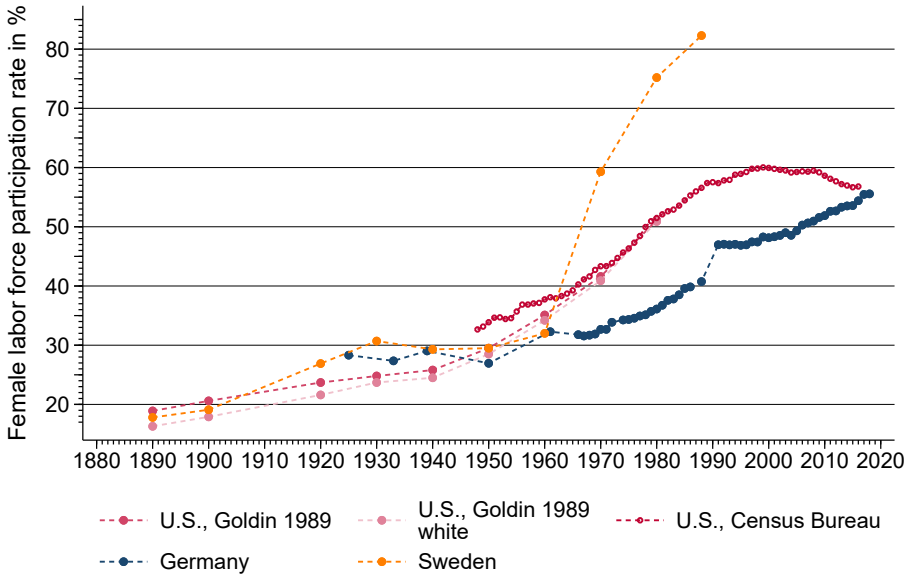
Note: The German occupational censuses included assisting family members. This graph shows the effect of this inclusion. Source: Occupational census (Berufszählung) 1882, 1895, 1907, 1925, 1933, 1939, 1950, 1961, 1970 and thereafter, micro census. Reunified Germany since 1991. For more details, see data appendix. Before 1925 assisting family members were not fully accounted for in the occupational censuses. I use data by Müller et al., 1983 who correct for this.

Figure 3.18.: Female participation rate by age group (without assisting family members).



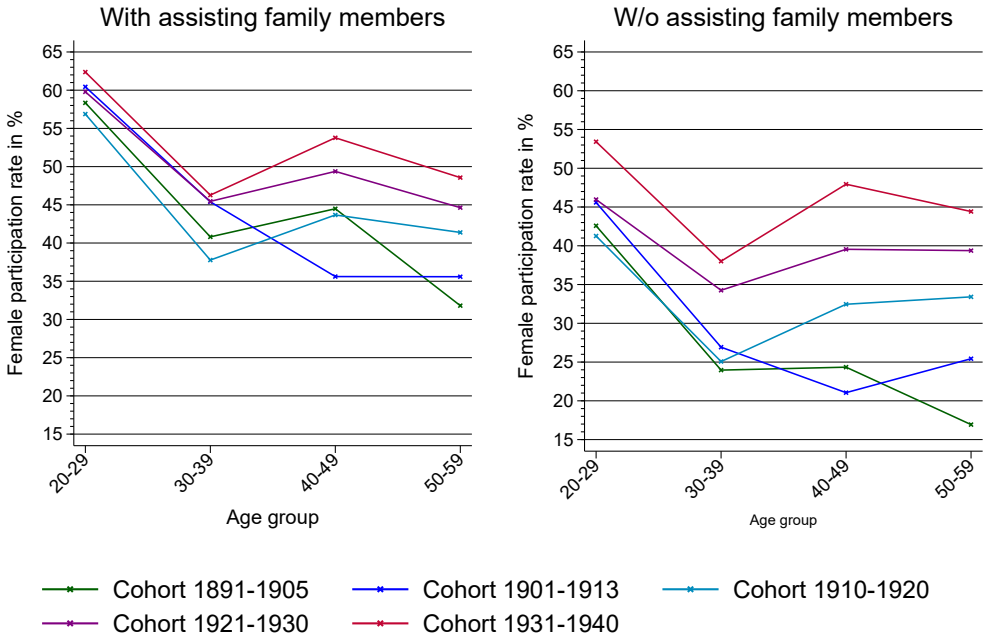
Note: Source: Occupational Census (Berufszählung) 1895, 1907, 1925, 1933, 1939, 1950, 1961, 1970 and thereafter, micro census. For more details, see data appendix. Reunified Germany since 1991.

Figure 3.19.: Female participation rate in the population of 15+ years (without assisting family members), Germany, Sweden and the United States



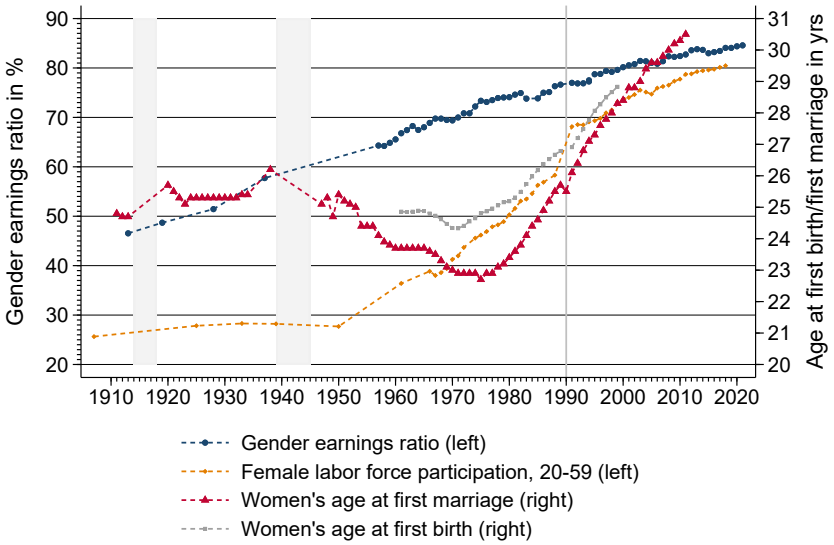
Note: United States taken from Goldin (1990) & Census Bureau. Germany own calculations based on occupational censuses. Share of women in the labor force of 15 years and above. For Germany including employed and unemployed, 1882-1939. Reunified Germany since 1991.

Figure 3.20.: Female participation rate (with and w/o assisting family members) over the life cycle.



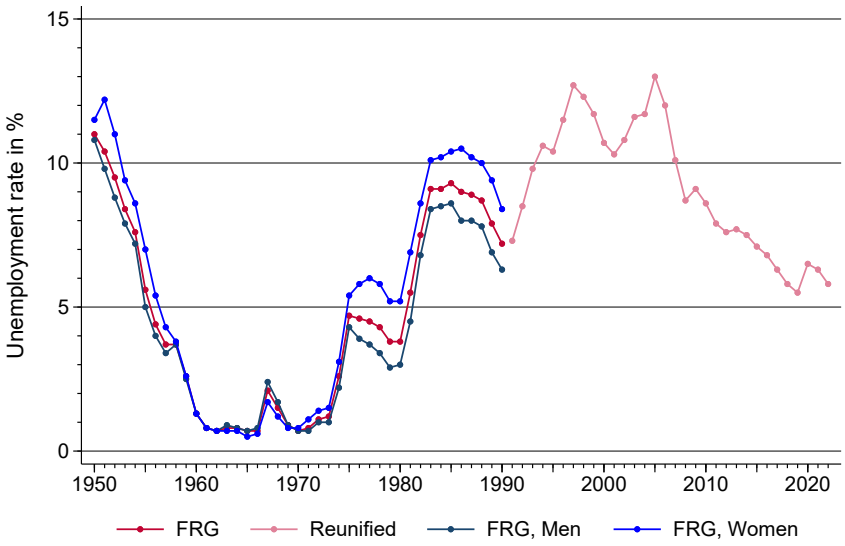
Note: Own calculations based on occupational censuses.

Figure 3.21.: Gender pay ratio, female labor force participation and marriage age.



Note: Gender earnings ratio and participation rate based on own calculations. For data documentation, see data appendix. Participation rate without assisting family members. Marriage age and age at first birth are visualizations based on (Rahlf, 2015). Reunified Germany since 1991 for participation rate, marriage age and age at first birth; since 1994 for gender earnings ratio.

Figure 3.22.: Unemployment rate.



Source: Own visualisation based on Bundesagentur für Arbeit (2023).

4 Revenue Effects of the Global Minimum Tax under Pillar Two*

coauthored with Mona Baraké (EU Tax Observatory), Paul-Emmanuel Chouc (EU Tax Observatory), and Gabriel Zucman (UC Berkeley & EU Tax Observatory)

Abstract

In October 2021, 137 countries and jurisdictions agreed to implement a major reform of the international corporate tax system, i.e., a global minimum tax of 15% on the profits of large multinational companies. This article presents simulations of the revenue effects of the global minimum tax. Two possible scenarios are considered regarding who collects the minimum tax: The country in which the headquarters are located based on the Income Inclusion Rule (IIR) or the host country of foreign affiliates as laid out under the Qualified Domestic Minimum Top-up Tax (QDMTT). The Organization for Economic Cooperation and Development's (OECD's) tabulated country-by-country report (CbCR) statistics are complemented with data by Tørsløv, Wier, and Zucman (2020). Based on a sample of eighty-three parent countries, it is estimated that headquarters countries could collect a total revenue of EUR 179 billion globally. The EU Member States could receive EUR 67 billion from a 15% minimum top-up tax. Carve-outs, provisions that decrease the tax base for real economic activity, reduce the potential tax revenues by approximately 14% to 22% over the entire sample. Under the current agreement, the European Union can expect a total tax revenue of EUR 55 billion yearly. The analysis accentuates how the distribution of revenues varies depending on which country has the priority to collect. Under the IIR in which the headquarters country collects the top-up tax, a country receives more revenues when it hosts more headquartered multinationals. With Qualified Domestic Top-up Taxes that give the host country of the foreign affiliate the priority to collect the top-up tax, low-tax jurisdictions that have attracted affiliates of many multinationals could be among the main beneficiaries of the reform. Static estimates, that take the distribution of profits and taxes paid as given, are presented. Thereafter, possible behavioural effects that may affect the estimates are discussed.

Published as:
M. Baraké et al. (2022). "Revenue Effects of the Global Minimum Tax Under Pillar Two." In: *Intertax* 50.10, pp. 689–710.
doi: https://doi.org/10.54648/taxi2022074.

Due to copyright restrictions, this chapter is not included in the online version of the dissertation.

* The authors would like to thank the participants in the World Inequality Lab Conference (2021) and the Hybrid Intertax & Cideeff Seminar on Pillar Two (2022) for their helpful comments and discussions. They also gratefully acknowledge Gaspard Richard for excellent research assistance.

A | Data Appendix: Distributional National Accounts (DINA) for Germany, 1992-2016*

**Coauthored with Stefan Bach (DIW Berlin)
and Charlotte Bartels (DIW Berlin)**

Abstract

This Data Appendix supplements our paper "Distributional National Accounts (DINA) for Germany, 1992-2016". It provides complete details on the methodology, data and programs.

A.1. Data sources

Distributional National Accounts for Germany are based on three main data sources – national accounts, personal income tax data, and SOEP survey data – which we describe in this section.

A.1.1. National accounts

German national accounts data (national income and sector accounts) are produced by the German federal statistical office (*Statistisches Bundesamt*, www.destatis.de). We supplement these data with additional macroeconomic indicators from Eurostat, the OECD and other publications by the federal statistical office. All macroeconomic series used in this project are included in the Excel file Macro_Germany.xlsx in the data input sheets Input1 to Input17.

Input 0: Edited DINA income concepts and subcomponents.

Input1 and Input2: net national income and its subcomponents, mainly based on domestic product accounts and sector accounts (Statistisches Bundesamt, 2022c).

Input3: tax aggregates by income source published in Statistisches Bundesamt (2022b): Fachserie 14, Reihe 7.1. "Finanzen und Steuern".

Input4: the number of firms liable under the corporate tax and the amount of profits (*Gesamtbetrag der Einkünfte*) and the liable corporate tax foll. Statistisches Bundesamt (2022a).

Input5: financial balance sheets of the German government and household sector provided by the OECD (2022a).

* This is the data appendix for chapters 1 and 2.

Input6: subcomponents of the sector accounts position D44 (Investment income attributable to insurance policy holders) provided by EUROSTAT (2022). These subcomponents are important for the construction of pretax national income.

Input7: GDP accounts (Statistisches Bundesamt, 2022e, Fachserie 18, Reihe 1.4) focusing on household's disposable income and expenditure for monetary social transfers to households.

Input8: family and children's allowances and benefits (*Familienleistungsausgleich*) (Bundesministerium für Finanzen, 2022).

Input 9: GDP accounts (Statistisches Bundesamt, 2022e, Fachserie 18, Reihe 1.4) focusing on public consumption expenditure and its subcomponents (COFOG).

Input 10: effective interest rates granted to private households, particularly home owners asking for loans, provided by Deutsche Bundesbank (2022).

Input 11: information directly provided to us by the statistical office on 1) net operating surplus and net mixed income (incl. the split between the subcomponents rental income and self-employment income) of households and 2) interest received and paid by households incl. information on included FISIM.

Input 12: Data on the property tax (*Grundsteuer A & B*) from Statistisches Bundesamt (1993-2022).

Input 13: Information on firm size (no. of employees) by WZ08 sector classification from the company register (Statistisches Bundesamt, 2006-2020) and yearly information on the number of GmbHs (private limited liability corporations) (Statistisches Bundesamt, 1995-2021)

Input 14: the number of employed in "Minijobs" according to SOEP survey data and the Federal Employment Agency (Bundesagentur für Arbeit, 2021).

Input 15: consumer price index provided by Statistisches Bundesamt (2022d).

Input 16: gross operating surplus and gross mixed income of the household sector from EUROSTAT (2022).

Input 17: government's aggregate expenditure for social assistance and its subcomponents (Statistisches Bundesamt, 2023b) and war victims welfare (Statistisches Bundesamt, 2023a).

A.1.2. Personal income tax data

We can access the universe of personal income tax (PIT) returns via remote data processing and the 10% stratified sample with full coverage of top incomes on-site for the tax years 1992, 1995, 1998, 2001, 2004, 2007, 2010 and 2016 (until September 2023 the last year available). The DINA concept hinges on the national or resident concept (vs. the domestic concept). Consequently, we exclude all PIT units with restricted tax liability (*beschränkt steuerpflichtig*) who are most likely foreigners without a permanent residence in Germany. Our sample consists of tax units

with unrestricted tax liability (keep ef17 == 1-3) and permanent residence in Germany (drop if ef1==10). In line with DINA guidelines by Blanchet et al. (2021), we only keep tax units of 20 years and above. After these sample restrictions, the datasets provide information on items stemming from tax forms for approximately 37 mio. tax units (tax year 2007).
German personal income tax law distinguishes seven income sources:

1. Income from agriculture and forestry
2. Business income from unincorporated businesses like sole proprietorships or partnerships (*GbR, OHG, GmbH & Co KG* etc.)
3. Self-employment income
4. Employment income
5. Capital income from interest income or dividends from corporations (*GmbH, AG*)
6. Income from renting and leasing
7. Other incomes (e.g., private pensions)

Since the introduction of a withholding tax on dividend and interest income in 2009, capital income is not systematically documented in personal income tax returns. In section A.3.7, we describe our strategy to impute capital income after 2009.

In Germany, married spouses can file jointly. Nonetheless, items are recorded individually for each spouse. This makes a thorough gender analysis possible. The bandwidth of items included in the tax data increases in more recent years such that analyses can be conducted more precisely for more recent years due to more detailed information on, e.g., pensions, child benefits or the local business tax (*Gewerbesteuer*).

A.1.3. Survey data

We use the German Socio-Economic Panel (SOEP v35) for two purposes: First, we identify non-filers in order to merge this part of the population with the tax-filing population included in PIT data. Second, we conduct mean-value imputations on items missing in PIT data (imputed rents, paid property incomes such as mortgages, capital income below the savings allowance, social assistance benefits for low-income earners). The SOEP is available on an annual basis since 1984 and for East German household since 1990. The dataset includes a wide variety of income items such as primary incomes, taxes, insurance-based and means-tested benefits as well as imputed incomes such as imputed rents from owner-occupied housing. While most income items are individualized, the following are only available at the household level: income and losses from renting and leasing as well operating costs for rented apartments, interest and dividend incomes and losses from capital investment as well as most social and means-tested benefits, such as child benefit, childcare subsidy and additional child benefits, housing benefit, long-term care allowances, social assistance benefits and unemployment benefits II (Grabka, 2019).

We split households into tax units to align the unit of observation with income tax data. To weight tax units, we use the enumerated individual weight (*$$phrfe*) of the main earner for the tax unit. Enumerated weights have less missing values than individual weights in the SOEP. When summing up the newly built weights, we arrive at a reasonable sum of about 49 million potential tax units (incl. both filers and non-filers) of 20 years and above in Germany in 2007, which is in line with the top fiscal income series by Bartels (2019).

A.2. Creating DINA in national accounts

In this section, we describe how we arrive at the four main income concepts – pretax factor income, pretax national income, posttax national income and posttax disposable income. In general, we closely follow the seminal paper by Piketty et al. (2018) and the guidelines by Blanchet et al. (2021) in order to construct internationally comparable DINA series for Germany. We point out German particularities in each subsection. Our main challenge in constructing DINA for Germany is to conceptually harmonize and align the smallest possible subcomponents of national income across national accounts, income tax data and survey data. For example, we split net mixed income of the household sector (B3n, S14 +S15) into sole proprietor income and income from renting and leasing of households and align microdata with NA respectively. In some cases, we must reallocate incomes from one component to another in the NA to harmonize micro- and NA data. The ultimate goal is to reach the national accounts definition of the respective main income concept (pretax factor, pretax national, posttax national or posttax disposable income). This means, reallocating between the subcomponents leaves the main income concept unaffected.

In the following, we first give the official German definition of the available subcomponents and point out where German statistics deviate from the international standard outlined in United Nations et al. (2009). Second, we identify where income tax data and national accounts differently allocate income subcomponents. For example, German national accounts classify CEO compensation as property income (D42, S14) and do not list this item separately, while income tax statistics classify CEO compensation as employment income. We then explain how we correct those deviations, for example, by reallocating subcomponents using additional information. Third, we also explain how subcomponents of national accounts for which we do not have microdata counterparts, are distributed across the income distribution.

Note that German national accounts do not differentiate between the household sector (S14) and non-profit institutions serving households (NPISHs, S15). This means that we cannot add NPISH income to government incomes (as suggested by Blanchet et al. (2021)). It remains included in household incomes in Germany.

A.2.1. Pretax factor income

A.2.1.1. Pretax personal factor labor income

Pretax personal factor labor income is computed as follows:

	Concept	SNA 2008	Sector
	Compensation of employees received by residents	D1	S1
+	Labor share (70%) of net mixed income	B3n	S14+15
+	Labor comp. of imputed taxes on production net of subsidies	D2-D3	S13
=	Pretax factor labor income		

D1 includes compensation of employees received by residents. To align NA with tax and survey microdata, we reallocate an estimated aggregate of employees compensation of CEOs, which is originally booked in distributed income of corporations (D42, S14 +S15), to employees compensation (D1, S1).

D2-D3 is imputed taxes on production (D2) less subsidies (D3). Imputed taxes on production comprise two main components: (1) taxes on goods and services including the VAT as its largest component (D21), and (2) other taxes on production, mainly property taxes (*Grundsteuer A + B*). We assume that the labor component of imputed taxes minus subsidies is equal to the share of labor income in net national income based on factor prices (*Volkseinkommen*). This share fluctuates between 75% and 80% between 1992 and 2018.

B3n, S14 + S15 net mixed income in Germany consists of 1) self-employment income and 2) income from renting and leasing. German definitions differ significantly from United Nations et al. (2009, SNA 2008) and other countries' implementation in two ways: First, self-employment income is defined as incomes of unincorporated entities without a legal identity such as sole proprietors, sole traders, members of the free professions as well as civil law partnerships (*GbR*), group practices of lawyers, doctors and architects. Thus, the legal form is the sorting criterion into either household or corporate sector. In United Nations et al. (2009, SNA 2008) and other countries' implementation, the size of the firm is crucial for the allocation of the firms' income to either household sector (B3n, S14) or corporate sector (B5n, S11+S12). Second, German net mixed income comprises income from renting and leasing (Statistisches Bundesamt, 2016, p. 306). In United Nations et al. (2009, SNA 2008) and other countries' implementation, renting and leasing incomes are part of the net operating surplus of households (B2n, S14).

To align B3n in macro- and microdata, we adjust the national accounts subcomponents to align with the PIT concepts. First, we deduct interest paid by landlords and self-employed because these income sources are net of interest paid in PIT data. Accordingly, we neutralize this specific paid interest in D4 in NA data.

A.2.1.2. Pretax personal factor capital income

Pretax personal factor capital income is outlined as follows in Blanchet et al. (2021):

	Concept	SNA 2008	Sector
	Capital share (30%) of net mixed income	B3n	S14+15
+	Net operating surplus of the household sector	B2n	S14+15
+	Property income received by households	D4, resources	S14+15
−	Property income paid by households	D4, uses	S14+15
+	Household component of primary income of the corp. sector	B5n	S11+12
+	Capital comp. of imputed taxes on production net of subsidies	D2-D3	S13
=	Pretax factor capital income		

B2n, S14+15 Net operating surplus of the household sector comprises imputed rents for owner-occupied housing.

B5n, S11+12 We follow the DINA framework (Blanchet et al., 2021), and distribute the primary income of the corporate sector (B5n, S11+12) to individuals and the government because eventually all corporations are owned by individuals or the government. We split the corporate

income aggregate (B5n, S11 +S12), into a share received by households and another received by the government. We estimate the household share as:

$$\text{Household share in corporate sector income} = \frac{\text{Equity \& investment fund shares held by households}}{\text{Equity \& invest. fund shares held by hhs \& the gov.}} \times [B5n, S11 + 12] \tag{1}$$

Data was retrieved from financial balance sheets provided by the OECD (Input5 in Macro_Germany.xlsx). This share varies between 81% and 63% over the 1992-2018 period. The household share is added to pretax personal factor capital income, while the remaining government share is added to pretax government factor capital income (see subsection A.2.1.3).

D4, S14+15, received Received property income in national accounts consists of five subcomponents (D41 to D45). Eurostat offers a more detailed break down of D44 into D441 to D443, which we use as complementary source because this differentiation enables us to match the PIT microdata more precisely to macro aggregates.

- **D41, S14+15** Interest income
- **D42, S14+15** Distributed income of corporations, i.e., dividends from capital stock and from companies (shareholding in companies) and withdrawals from income of quasi-corporations
- **D43, S14+15** Reinvested earnings on foreign direct investment (note that this is zero for households)
- **D44, S14+15** Investment income disbursements:
  - **D441, S14+15** Investment income of insurance policy holders such as private health and life insurance payments
  - **D442, S14+15** Investment income of insurance policy holders such as private pensions
  - **D443, S14+15** Investment income attributable to collective investment fund share holders. Following Blanchet et al. (2021), we add this to D42, S14+S15.
- **D45** comprises rents from agricultural land, water bodies, mining and oil extraction land as well as other types of land (recreational areas). Data on this item is fragmentary at the macro level and non-existent at the microdata level (Statistisches Bundesamt, 2001, p. 41).

D4, S14+15, paid Property incomes paid by households are decomposed into two subcomponents:

- D41, S14+15: Interest paid includes 1) mortgage payments of landlords renting out their property, 2) mortgage payments of home-owners and 3) interest payments of sole proprietors. To mirror our correction to B3n, S14+ S15, we subtract interest paid by landlords and sole proprietors from this component, leaving only mortgage payments of home-owners, which we distribute according to the distribution in the SOEP.
- D45, S14-15: Land leasing paid.

Note that we do not correct for Financial Intermediation Services Indirectly Measured (FISIM). FISIM increases interest received by households in order to account for the financial intermediation

services. This is problematic from a distributional point of view, as FISIM allocates income earned by financial institutions (and their employees and shareholders) to households. Vice versa, FISIM reduces the interest paid by indebted households.

A.2.1.3. From pretax personal factor income to pretax factor income

Pretax personal factor capital income and pretax personal factor labor income add up to pretax personal factor income. To reach pretax factor income (equal to net national income), we add pretax government income:

	Concept	SNA 2008	Sector
	Property income received	D4, res.	S13
–	Property income paid	D4, uses	S13
+	Government comp. of primary income of the corporate sector	B5n	S11+12
=	Pretax government factor capital income		

Here, we differ from the DINA guidelines by Blanchet et al. (2021). Because we cannot distinguish incomes of households and NPISHs, we cannot add NPISH income to government incomes (as suggested by Blanchet et al. (2021)). It remains included in personal factor incomes in Germany. The definition of the government component of primary income of the corporate sector is laid down in Equation 1.

A.2.2. Pretax national income

Pretax national income is computed as pretax factor income after the operation of the social insurance system but before the operation of the tax-transfer system. Note that – while leaving the aggregate income unaffected – including the social insurance system in the income concept of pretax national income, will have effects on the *distribution* of income. Conceptually, pretax national income is the sum of pretax labor income and pretax capital income which will be explained in the following subsections.

A.2.2.1. Pretax labor income

	Concept	SNA 2008	Sector
	Pretax factor labor income		
–	Pension and other social contributions	D61, uses	S14+15
+	Pension and other social insurance benefits (labor share)	D621 +D622 - D442	S14+15
+	Labor component of pension and other social insurance surplus (100%)		
=	Pretax labor income		

D61 paid, S14+15 The sector accounts provide the subcomponents of D61:

1. employer's actual social contributions (D611)
2. employer's imputed social contributions (D612, incl. assumed civil servants pension contributions and company pensions)

3. households' actual social contributions (D613)
4. households' social contribution supplements (D614) and
5. social insurance scheme service charges (D61SC, substracted)

We use this information to uprate employer's contributions (D611 + D612) separately from employee's contributions (D613 + D614 - D61SC).

D621 + D622, S14+15 : Blanchet et al. (2021) advise to either use a narrow (only old-age pensions) or broad definition of social insurance. As Germany has a broad social insurance system covering old-age pension, unemployment insurance, sick leave etc., we use the broad concept. It includes all replacement incomes distributed via the social insurance system, such as public old-age and disability pension, insurance-based unemployment benefits (*Arbeitslosengeld I*), benefits for sickness leave and long-term care allowance. We split social insurance benefits into a labor and a capital share assuming that the amount of investment income payable to pension entitlements (D442) captures the capital share of this item.

Pension and other social insurance surplus is the difference between contributions and benefits. We attribute 100% of this surplus to labor income because capital-based private pensions are marginal in Germany.

	Concept	SNA 2008	Sector
	Pension and other social contributions	D61, uses	S14+15
+	Investment income payable to pension entitlements	D442	S14+15
–	Pension and other social insurance benefits	D621 +D622	S14+15
=	Pension and other social insurance surplus		

A.2.2.2. Pretax capital income

	Concept	SNA 2008	Sector
	Pretax personal factor capital income		
–	Investment income payable to pension entitlements	D442	S14+15
+	Capital share of pension and other social insurance benefits	D442	S14+15
+	Capital component of pension and other social insurance surplus (0%)		
–	Pretax government factor income		
=	Pretax capital income		

A.2.3. Posttax national income

Posttax national income is the sum of three components which we describe in the following: posttax disposable income (A.2.3.1), in-kind transfers and collective expenditures (A.2.3.2) as well as government surplus (or deficit) (A.2.3.3). Note that — while leaving the aggregate income unaffected – including taxes and means-tested benefits in cash, will have effects on the *distribution* of income.

	Concept
	Posttax disposable income (see A.2.3.1)
+	Public spending and in-kind transfers (see A.2.3.2)
+	Government surplus (primary) (see A.2.3.3)
=	Posttax national income

A.2.3.1. Posttax disposable income

Posttax disposable income results from pretax national income after the operation of taxes and means-tested benefits in cash. Note that social insurance-based monetary transfers are already distributed in the pretax national income concept (see section A.2.2). Starting from pretax national income, means-tested benefits in cash (= social assistance benefits in cash) are added. Imputed taxes on production net of subsidies as well as current taxes on income and wealth are deducted.

	Concept	SNA 2008	Sector
	Pretax national income		
+	Social assistance benefits in cash	D623	S14+15
-	Imputed taxes on production net of subsidies	D2-D3	S13
-	Current taxes on income and wealth received by the government	D5	S13
=	Posttax disposable income		

D623, S14+15 Social assistance benefits in cash include include the following categories in German NA:

1. Social assistance and war victims welfare (*Sozialhilfe & Kriegsfürsorge*) comprising the following components in order of their quantitative importance:
    a) integration assistance for disabled people (*Eingliederungshilfe für Behinderte*) comprising subsidies for services in recognized workshops, services for the participation in life in the community, and a broad range of subsidies for disabled persons from medication to education-related expenses;
    b) basic subsistence income for the elderly (*Grundsicherung im Alter*);
    c) care assistance (*Hilfe zur Pflege*) for people that cannot afford to pay for care for themselves and their relatives and for which the care insurance does not cover the full amount required;
    d) regular social assistance for working age individuals, but classified as unable to live from own earnings (*Hilfe zum Lebensunterhalt*), which differs from the group of individuals receiving unemployment assistance, who are classified as able to work (see below);
    e) subsidies for healthcare of persons not integrated in to the social insurance system like homeless persons (*Hilfen zur Gesundheit*)
    f) social assistance for special circumstances (*Sozialhilfe in anderen Lebenslagen*)
    g) benefits for asylum seekers (*AsylbLG*)
2. War victim pensions (*Kriegsopferversorgung*)

3. Unemployment assistance (*Arbeitslosenhilfe*) which includes needs-based unemployment benefits (*Arbeitslosenhilfe until 2004, ALG II thereafter*) for individuals classified as able to work (as opposed to social assistance, see above). Since 2011, this item also includes the education package for low-income families (*Bildungs- und Teilhabepaket*) for children in transfer-receiving households

4. the NA aggregate of child benefit (*Kindergeld*) includes child benefits (*Kindergeld*) and the child surcharge (*Kinderzuschlag*). In Germany, child benefits are granted in two ways: A child benefit (*Kindergeld*) that is paid monthly to every household with children and a child tax allowance (*Kinderfreibetrag*) that reduces taxable income and is granted when filing the PIT return. Before 1995, both the benefit and the tax allowance were granted to all tax units with children (dual system). Since the child benefit reform of 1996, the tax authorities test and apply the most favorable option (*Günstigerprüfung*). If in the yearly tax return, the child tax allowance is more favorable, the difference between the tax allowance and the already received benefit is credited to the taxpayer. Most tax units only receive the child benefit. Only high-income earners benefit additionally from the tax allowance. The German national accounts and statistical office divide the child benefits and allowance in two parts which do not reflect the effective transfer as just described: 1) Exemption of the subsistence income of the child (*Freistellung des Existenzminimums*) which is the theoretical tax allowance for all tax units irrespective if the benefit or allowance is more favorable; 2) the *Förderanteil*, i.e., the difference between the paid child benefit and the modelled subsistence income of the child. Only the *Förderanteil* is registered as child benefit under D623, S14+S15. We combine both parts to pool the full amount of paid child benefit in D623 using information on the aggregates of both concepts from Bundesministerium für Finanzen (2022, p. 53). More precisely, we add the aggregate of the *Freistellung des Existenzminimums*, decreased by the additional tax relief of the tax allowance (*Freibetrag*), to the child benefit (D623, S14+S15). This assumes that everyone has received only the monthly child benefit, abstracting from the additional tax relief for high earners via the tax allowance. To balance out this reallocation, we add this item back to taxes on income and wealth (D5, S13, resources) because it had reduced taxes. Since D623, S14 is added to individuals' incomes and D5, S13 is subtracted in the microdata, the aggregate posttax income is not affected.

5. Housing benefit (*Wohngeld*)

6. Other benefits which include parental leave benefits (*Erziehungsgeld; Elterngeld*), students' stipends and alimony advance payments (*Unterhaltsvorschuss*)

To benchmark our microdata aggregates, we use the aggregates provided by (Statistisches Bundesamt, 2022e, table 3.4.4.4, Fachserie 18, Reihe 1.4.) and aggregates based on expenses for social assistance benefits based on the *Sozialhilfestatistik* available at Destatis Genesis. Note that the NA aggregates are based on the *Finanzstatistik* and only include in-cash (but not in-kind) transfers. The *Sozialhilfestatistik* includes expenses for both in-cash and in-kind transfers.

A.2.3.2. In-kind transfers and collective expenditures

Consumption expenditure is defined as follows:

	Concept	SNA 2008	Sector
	Individual consumption expenditure	P31	S13
+	Collective consumption expenditure	P32	S13
=	Final consumption expenditure	P3	S13

P3, S13: Final consumption expenditure. The item can be broken down in two ways. First, into *individual consumption expenditure* **(P31, S13)** which includes in-kind transfers to individuals, especially health services and education, against *collective consumption expenditure* **(P32, S13)** which includes spending of non-individualized societal infrastructure such as defense and police spending as well as housing expenditures.

Alternatively, P3, S13 can be broken down using the Eurostat Classification of the Functions of Government (COFOG) into (EUROSTAT, 2014):

1. General and public services
2. Defense
3. Public order and safety
4. Economic affairs
5. Environmental protection
6. Housing and community amenities;
7. Health;
8. Recreation, culture and religion;
9. Education and
10. Social protection.

To distinguish between individual and collective consumption expenditure within those social purposes, we use COFOG two-digit data by Statistisches Bundesamt (2022e). Following United Nations et al. (2009, SNA 2008, p. 191): "All of classes 1 to 6 are collective services, as are section 7.5 and 7.6 of health, sections 8.3 to 8.6 of recreation, culture and religion, sections 9.7 and 9.8 of education, and sections 10.8 and 10.9 of social protection." The remaining divisions and groups are treated as individual consumption (P.31, S14+S15).

A.2.3.3. Government surplus

Government surplus (or deficit) is computed as follows:

	Concept	SNA 2008	Sector
	Imputed taxes on production net of subsidies received by the government	D2-D3	S13
+	Current taxes on income and wealth rec. by the government	D5	S13
–	Social assistance benefits in cash rec. by households	D623	S14+15
–	Individual expenditures	P31	S13
–	Collective expenditures	P32	S13
=	Government surplus (primary)		

A.3. Creating DINA concepts from PIT data

In this section, we explain three methodological procedures. (1) We explain how we build DINA income concepts in PIT data ensuring conceptual harmonization with German NA income concepts. (2) We detail how we impute national accounts items via mean-value imputation which are not recorded in PIT data but in SOEP, such as imputed rent. (3) We describe how we distribute items for which we have no microdata basis such as public spending. Note that uprating of the combined distribution of tax filers (PIT, this section) and non-filers (SOEP, section A.4) to national account aggregates is described in section A.5. German PIT law distinguishes seven income sources:

1. Income from agriculture and forestry
2. Business income from unincorporated businesses like sole proprietorships or partnerships (*GbR, OHG, GmbH & Co KG* etc.)
3. Self-employment income
4. Employment income
5. Capital income from interest income or dividends from corporations (*GmbH, AG*)
6. Income from renting and leasing
7. Other incomes (e.g., private pensions)

While [4.] is labor income and [5.] is capital income, [1.], [2.], [3.], [6.] and [7.] contain both a labor and a capital income component. More specifically, [5.] includes dividends from corporations (*Kapitalgesellschaften: AG, GmbH*) and interest income and categories [1.]-[3.] include income from sole proprietorhips (*Einzelunternehmer: Einzelkauffrau/-mann, Freiberufler*) and partnerships (*Personengesellschaften: GbR, OHG, KG, GmbH & Co KG, GmbH & Co OHG*). We follow national account definitions and classify civil law partnerships (*GbR*) as part of sole proprietors and *OHG, KG, GmbH&Co KG, GmbH & Co OHG* as quasi-corporations.

For agricultural incomes, business income, and self-employment income, we distinguish between the income from sole proprietorship which adds to net mixed income *B3n, S14+15* and income from quasi-corporations which adds to property income *D42, S14+15*.

A.3.1. Pretax factor income

A.3.1.1. Pretax personal factor labor income

	Concept	SNA 2008	strategy
	Compensation of employees	D1, S1	micro & uprate
+	Labor share (70%) of net mixed income	B3n, S14+15	micro & uprate
+	Labor comp. of imputed taxes on production net of subsidies	D2-D3, S13	distribute
=	Pretax factor labor income		

D1, S1: Compensation of employees received by residents (variable *ce**):

- Income from employment (income category [4.]) reduced by civil servant pensions and company pensions, which is also included in this tax laws' income category (*lohnbr = c65163 + c65164 - c66206 - c66207*)
- Tax-exempt income according to double-taxation treaties (*c66202 + c66203*)
- Employer's social insurance contributions which we simulate using the legislative contribution rates and assessment ceilings (Do-file *sozbeitrsaetze.do*). We separately simulate social insurance contributions for civil servants for which we multiply legislative contribution rates by 2.53 for two reasons: First, the absence of social security contributions for civil servants implies that we have to simulate both employee's and employer's contributions at the same time, which roughly doubles contributions. Second, their pension benefits are, on average, higher than for employees due to *Vollversorgung*. 2.53 is the ratio that is used in official national accounts calculations to impute social security contributions for civil servants.

B3n, S14+15: Labor share (70%) of net mixed income (variable *nmi*=soleprop + vuvbr*)

- Income from sole proprietorship (*Einzelunternehmer*, *soleprop*) from forestry and agriculture [1.], business (*Gewerbe*) [2.], and earnings from liberal professions (*Freiberufler*) and other self-employed activities (e.g. for membership in supervisory boards) [3.]. We add tax-free incomes from abroad[1] and deduct tax reliefs (see section A.3.5 for list of tax reliefs considered)[2].
  1. Income from sole proprietorship in agriculture and forestry [1.] (*c65103 + cc65104*)
  2. Income from sole proprietorship in business incomes [2.] (c65123 + c65124) + share of tax-free incomes from abroad (c66204/5, main share attributed to quasi-corporations)
  3. Income from free professions and other self-employed activities (board membership) [3.] (*c65143 + c65144 + c65147 + c65148*) + share of tax-free incomes from abroad (c66204/5, main share attributed to quasi-corporations)
- Income from renting and leasing (*vuvbr*) is taxable income from renting and leasing (*c65241 + c65242*). We account for the generous tax optimization schemes for real estate investments by cutting all losses above 5,000 €.

[1] We do not have information on the source of foreign incomes, except that foreign labor incomes are filed in a different form (Anlage N-AUS, variable c66202/3) and dividends and interest from abroad are directly declared in forms KAP for capital incomes. To make a reasonable assumption on the source of foreign tax-free incomes, we first, fully attribute to agricultural income, business (Gewerbe), self-employment income or income from renting and leasing based on the source of the highest absolute income (loss) a tax unit earns. Second, within the incomes from the dominant source, we primarily attribute incomes from abroad to quasi-corporations because we assume that German tax residents are less likely to be present as sole proprietors abroad. Only if a tax unit solely has business income as sole proprietor (and zero inc. from partnerships) in Germany, we attribute these foreign incomes to sole proprietorship.

[2] Tax reliefs are fully attributed to agriculture/business/self-employment income based on the source of the highest absolute income (loss) a tax unit earns. Second, we attribute tax reliefs proportional to either income from sole proprietorships or quasi-corporation. In cases with negative income from one and positive income from the other legal form, we prioritize the negative income.

D2–D3, S13: Labor share of net taxes on production : For pretax factor income, we distribute net taxes on production proportional to pretax personal factor income[3], as recommended in Blanchet et al. (2021, 58f.). In pretax factor income, no consumption choice was yet made, thus product taxes should only shift the level of incomes (from factor to market prices), but not yet affect the distribution.

A.3.1.2. Pretax factor capital income

	Concept	SNA 2008	Strategy
	Capital share (30%) of net mixed income	B3n, S14+15	micro & uprate
+	Net operating surplus of the household sector	B2n, S14+15	micro & uprate
+	Consolidated property income of households	D4, ress. - uses, S14+15	micro & uprate, mean-value imp. of mortgages
+	Personal component of primary income of the corporate sector	B5n, S11+12	distribute
+	Capital comp. of imputed taxes on production net of subsidies	D2-D3, S13	distribute
=	Pretax factor capital income		

B3n, S14+15: Capital share (30%) of net mixed income (variable *nmi* = *soleprop* + *vuvbr*)

- see "labor share of net mixed income" in Section A.3.1.1

B2n, S14+15: Net operating surplus of the household sector (variable *nos*) = Imputed rents of owner-occupied housing

- not included in tax data - mean-value imputed from SOEP data see Section A.3.6.1

D4, S14+15: Property income :

1. **D41, S14+15: Interest income, received** (var. *D41intr*) is not explicitly provided in the PIT microdata. We compute interest income as the residual of gross capital income (*kapbrovg*) minus dividends. We use gross capital income (Einnahmen) and not taxable income (Einkünfte) because we assume that dividends and interest income do not produced high operating costs, so that declared expenses (Werbungskosten) are mainly tax-related expenses such as consulting fees.
    - 1992-1995: gross capital income ($kapbrovg = (c65223 + c65224)$, see below) is not given in those years, but taxable capital income (c65221 + c65222) is. We mean-value

[3] Pretax personal income equals compensation for employees (ce*) + net mixed income (nmi*) + net operating surplus (nos*) + net property income (= D41int + D42div + d43fdi + d44ins + d45) + personal component of the corporate sector (picshh*).

impute the difference between both concepts, i.e. expenses (Werbungskosten) by quantiles of gross income (GdE), marital status and region (East/West) using the distribution of 1998; interest is then computed as gross capital income minus dividends (see below)

- 1998: $kapbrovg = (c65223 + c65224) - (c65653/0.3)$ where *c65223/4* is taxable capital revenue (= gross capital income above the saver's allowance)
- 2001: $kapbrovg = [(c65223 - (c54014 + 0.5 * c54016 + c54046 + 0.5 * c54048)] + [c65224 - (c54015 + 0.5 * c54017 + c54047 + 0.5 * c54049)]$ where *c65223/4* is capital income *c54014/5, c54046/7* are fully taxable dividends and *c54017/8, c54048/9* dividends of which only 50% are included in taxable capital income
- 2004-2007: $kapbrovg = [(c65223 - (c54014 + 0.5 * c54016 + c54046 + 0.5 * c54048 + 0.5 * c54054)] + [c65224 - (c54015 + 0.5 * c54017 + c54047 + 0.5 * c54049 + 0.5 * c54055)]$ where *c65223/4* is capital income *c54014/5, c54046/7* are fully taxable dividends and *c54017/8, c54048/9, c54054/5* dividends of which only 50% are included in taxable capital income
- after 2009: After the dual income tax reform of 2009, capital incomes in PIT data are not systematically documented anymore. We thus set all PIT-reported capital incomes to zero and fully impute dividends and interest (see section A.3.7).

2. **D42, S14+15: Dividends from corporations and income from quasi-corporations** (var. *D42div*). Our dividend concept is gross dividends, i.e. dividends before business taxes (*Gewerbe- und Körperschaftsteuer*). Table A.3.1 summarizes taxation regimes for dividends from corporations directly received by the personal income taxpayer (upper part) and dividends received through quasi-corporations (lower part). We add back business taxes to pretax incomes because these taxes levied at the firm level decrease the absolute amount of profits possibly distributed to shareholders. We also add tax-free incomes from abroad and tax reliefs as described for B3n, S14+15.

   - Dividends from corporations were subject to two tax reforms (2001/2002 and 2009), so that we must harmonize three different dividend concepts recorded in the PIT data over time. Our consistent definition of dividends is gross dividends, which is dividends before local business and corporate taxes. Thus, we have to add back business taxes[4] in the years, where only cash dividends are recorded in the PIT microdata, to obtain gross dividends.
   - Incomes from quasi-corporations (*c65125 + cc65126*) or tax groups (*Organschaften, c65133 + c65134*) from business incomes [2.], agriculture and forestry [1.], (*c65105 + c65106*), and self-employment income [3.] (*c65145 + c65146*) + share of tax-free incomes from abroad (*c66204/5*) and tax reliefs (as described for B3n, S14+S15)
   - Dividends received by the quasi-corporations, which is recorded as a separate item in business [2.] and self-employment [3.]
     - 1992-1998: First, we identify dividends using the tax credit granted for corporate income tax paid at the corporate level, which is recorded in PIT data (*c65653*,

[4] The corporate tax (*Körperschaftsteuer*) is levied on profits of corporations. Local business tax (*Gewerbesteuer*) is levied on corporate and unincorporated businesses excluding the free professions, small-scale self-employed and agricultural businesses. In Section A.3.4, we explain how we compute the local business tax.

see Table A.3.1). This item includes dividends received either directly by the individual taxpayer or indirectly through quasi-corporations. Second, we assign the dividend according to the major income source (corporate dividends or quasi-corporate income). The resulting dividend is already gross of the corporate tax.

  - 2001-2007: We see the taxable 50% of net dividends in variables *c44024 + c44025*; We also know that 50% of net dividends are already included in business incomes (variables *c65125/6 or c65133/4*) - which we also include in D4 (see above). Thus, we only add the variable *c44024/5* once to D4 (i.e., 50% of dividends) and add 100% of the business tax liability on these dividends to take into account gross dividends. We proceed analogously with dividends flowing to quasi-corporations of self-employed with variables *c22062 + c22063*.
  - 2010: Dividends received through quasi-corporations are still fully visible in the tax data, which is in contrast to dividends from corporations received by individual taxpayers that are subject to a withholding tax since 2009. We observe the tax-free 40% of net dividends in variables *c44024 + c44025*; We also know that the taxable 60% of net dividends are already included in taxable business incomes (variables *c65125/6 or c65133/4*) - which we also include in D4 (see above). Thus, we only add the variable *c44024/5* once to D4 (i.e., 40% of dividends) and add 100% of the business tax liability on these dividends to take into account gross dividends. We proceed analogously with dividends flowing to quasi-corporations of self-employed with variables *c22062 + c22063*.

3. **D43, S14+15 Reinvested earnings on foreign direct investment** (Var. *D43fdi*): The aggregate is zero.

4. **D44, S14+15 = D441 + D442 + D443: Investment income disbursements** var. *D44ins* - aggregate fully distributed

   - **D441** = Investment income attributable to insurance policyholders: private health + life insurance: We distribute the macro aggregate proportionally to the income of likely policy holders. These are all individuals not covered by public health insurance scheme, e.g., civil servants, entrepreneurs and capital rentiers.
   - **D442** = Investment income payable on pension entitlements: private pensions. Piketty et al. (2019a, p. 31) distribute this item proportional to private pension wealth. As we lack this information, we distribute this item proportionally to income (employment income, capital incomes, sole proprietor income and income from renting and leasing) of likely insurance policy holders. These are entrepreneurs, high-earning employees who want to increase their pensions above the maximum public amount and other groups like capital rentiers.
   - **D443**: Investment income attributable to collective investment fund share holders. We added this item to the national accounts aggregate of D42, S14 + S15.

5. **D45, S14+15**: Rents from agricultural land, water bodies, mining and oil extraction land and other types of land. The macro aggregate is fully distributed proportional to dividends and income from quasi-corporations in business and agriculture. We assume that dividends and income from quasi-corporations as well as land rents are primarily shared between tax units at the upper end of the distribution and thus, have similar distributional patterns.

We mean-value impute income from interest and dividends (D41, received + D42, S14) below the saver's tax allowance and mortgages paid by homeowners (D41, paid, S14) using distributional information in the SOEP, because those items are not recorded in income tax data (see sections A.3.6.1 & A.3.6.2).

Table A.3.1.: Taxation of dividends

	pre-2001	2001/02–2008	since 2009
	Dividends received directly by personal income taxpayers		
Tax base	100%	$(1-t_{lb})\cdot(1-t_{corp}*(1+t_{soli}))$	$(1-t_{lb})\cdot(1-t_{corp}*(1+t_{soli}))$
Deductions	100%	50%	40%
Tax rate	PIT	PIT	$min(w, PIT)$
LBT deductable from CIT	*yes*	*yes*	*no*
Corp. tax credit in PIT	*yes*	*no*	*no*
Income source	capital inc. (5.)	capital inc. (5.)	capital inc. (5.)
Code	(1)	(2)	(3)
	Dividends received through quasi-corporations		
tax base	100%	$(1-t_{lb})\cdot(1-t_{corp}*(1+t_{soli}))$	$(1-t_{lb})\cdot(1-t_{corp}*(1+t_{soli}))$
deductions	100%	50%	40%
tax rate	PIT	PIT	PIT
income source	business inc. (2.)	business inc. (2.)	business inc. (2.)
t_{corp}(%)	30%	25%	15%
t_{soli}(%)	5.5% of CIT	5.5% of CIT	5.5% of CIT
t_{lb} (%, nat. average)	16.6%	16.6%	14%

Note: Business taxes decrease the absolute amount of profits available for distribution among shareholders. Thus, we add these taxes back to obtain gross dividends. The tax rate of the withholding tax w on dividends introduced in 2009 is 25%. t_{corp} indicates the corporate tax rate (Körperschaftsteuer), t_{soli} the solidarity surcharge, t_{lb} the local business tax rate (*Gewerbesteuer*).
Code (1) $brutdiv = (c65653/(t_{corp} * (1 + t_{soli})/(1 - t_{lb}))$;
Code (2) $brutdiv = \frac{cashdiv/(1-t_{corp}*(1+soli))}{1-t_{lb}}$
2001, 2004: $cashdiv = c54014 + c54015 + c54016 + c54017 + c54046 + c54047 + c54048 + c54049$;
2007: $cashdiv = c54014 + c54015 + c54016 + c54017 + c54046 + c54047 + c54048 + c54049 + c54054 + c54055$
Code (3) $brutdiv = \frac{cashdiv}{(1-t_{lb}-t_{corp}*(1+t_{soli}))}$; $cashdiv = 0.5 \cdot [c65223 - (c54211 + c54231 + c54241 + c54251 + c54261 + c54271 + c54274) + 65224 - (c54411 + c54431 + c54441 + c54451 + c54461 + c54471 + c54474)]$ Ass.: Dividend/Interest in capital inc. = 50/50.

B5n, S11/S12 : Personal component of primary income of corporate sector: This component is distributed from the macro aggregate proportional to individuals' positive dividends and income from quasi-corporations.

D2–D3, S13, capital share : See section A.3.1.1.

A.3.1.3. Pretax government capital income

The macroeconomic aggregate of pretax government factor capital income is distributed proportionally to pretax personal factor income which is the sum of pretax personal factor labor income and pretax personal factor capital income after uprating these concepts proportionally to meet NA.

	Concept	SNA 2008	Sector
	Property income received	D4, res.	S13
–	Property income paid	D4, uses	S13
+	Government component of primary income of the corporate sector	B5n	S11+12
=	Pretax government factor capital income		

A.3.2. Pretax national income

A.3.2.1. Pretax national labor income

We use the broad definition of replacement incomes from the DINA guidelines (Blanchet et al., 2021) including insurance-based pensions, insurance-based unemployment benefits and sick leave benefits.

	Concept	SNA 2008	Strategy
	Pretax personal factor labor income		
–	Employer's social insurance contributions	D611 +D612, S14 + S15	microdata uprate
–	Employee's social insurance contributions	D613 + D614 - D61SC, S14 + S15	microdata uprate
+	Pension and other social insurance benefits	D621 +D622, S14+15	microdata uprate
+	Social insurance surplus		macro agg. distribution
=	Pretax national labor income		

D61, S14+15, paid : Social insurance contributions: We simulate employee's (D613 +D614 - D61SC) and employer's contributions (D611 + D612) at the micro level and uprate them separately to NA.

- We simulate employee's (variable *ansozbtr*) and employer's contributions (variable *agsozbtr*) for pension, unemployment and compulsory health and long-term care insurance for self-employed, employees, and civil servants applying the prevailing legislative contribution rates and assessment ceilings (*sozbeitrsaetze.do*) both for SOEP (file *7_ stsm.do*) and tax data (file *8_ sozbeitr.do*). In the statutory insurance schemes, employers and employees share the contributions for pension and unemployment insurance. In the case of the health insurance, employees shoulder a larger share. Long-term care insurance contribution rates depend on the federal state, the age and the number of children of the employee.

- Civil servants are not contributing to the public pension insurance scheme. Their pension benefits are shouldered by their employer, the (federal) state. Nonetheless, their "contributions" are included in NA under the item "Employers' imputed social contributions" (D612, S13 paid). To include civil servants contributions, we assume that pension contributions for civil servants amount to 2.35 times the statutory employees' social insurance contributions and are 100% shouldered by their employer. This number amounts to the employee's and the employer's share of contributions and adds 0.35 for the, on average, substantially higher

pension payments compared to employees and additional health benefits for active and retired civil servants and their dependants (*Beihilfe*). Civil servants usually contribute to a private health and long-term care insurance scheme. Since contribution rates depend on health status of a person and we do not have information on health, we assume similar contribution rates as for employees under the statutory schemes.

- For business owners and self-employed, there are no strict rules which we could use to simulate social insurance contributions. Thus, we use the variable *c65416* (insurance contributions). We subtract the simulated insurance contributions for employee's (based on employment income and pensions) from all declared insurance contributions: $untsozbtr = c65416 - ansozbtr$. Thereby, we can distinguish within the tax unit between social insurance contributions from employment and pension income (*ansozbtr*) and private social insurance contributions (*untsozbtr*). For most self-employed and entrepreneurs with no employment income nor pensions, social insurance contributions will amount to: $untsozbtr = c65416$.[5]

D442, S14+15: Investment income payable to pension entitlements

- see section A.3.1.2.

D621 + D622, S14+15 : Social insurance benefits

- We only observe the taxable share of pensions (*Ertragsanteil*), which until 2004 amounted to 30% and thereafter gradually increased to 76% in 2018 and depends on the year of retirement. Thus, we simulate the full amount of pensions (*3_ pensions.do*) using the person's age to estimate the year of retirement and the taxable share of pensions. We also take into account that pension increases were 100% taxable. The taxable share of pensions is given by the following variables in PIT data:
  - 1992 -1995: v856/6
  - 1998-2004: *c65263/4,c65265/6,c65267/8,c65269/70*;
  - 2007-: *c65263/4, c65269/70*
- Civil servants pensions (*Versorgungsbezüge*, var. *vgbzbr*) include the taxable share of civil servants pensions before subtracting the personal tax allowance (*c66206/7*) + the professional allowance (*c65172/3*, Werbungskostenpauschale); Variable *c65172/3* includes professional allowances for employees and pensioners.
- Replacement incomes (*Lohn- und Einkommensersatzleistungen*, var. *replace*) are recorded in tax files because they contribute to the progression proviso (*Progressionsvorbehalt*). Replacement incomes include insurance-based unemployment benefits (*ALG I*), sickness pay (*Krankengeld*), short-time work allowance (*Kurzarbeitergeld*), early retirement transition benefit (*Altersübergangsgeld*), transitional allowance (*Übergangsgeld*), as well as parental leave and maternity benefits (*Elterngeld, Mutterschaftsgeld*). We subtract parental leave and maternity benefits from replacement incomes because national accounts classify these benefits as *social assistance benefits in cash* (D623, S14+15), which are added at the stage of posttax disposable income.
  - 1992-1995: (v890 + v891) + (v915 + v916 + v917 + v918) if female taxpayer's age between 20 and 40 (v54) & receiving child benefit (v42)

[5] We do not use the variable *c65416* (insurance contributions) for employee's and civil servants' contributions, because *c65416* also covers further insurances unrelated to pension, unemployment and sickness. Hence, simulations applying legislative contribution rates provide a more accurate estimate.

- 1998-2001: (c18020 + c18021 + c66200 + c66201) - (c18020 + c18021 + c66200 + c66201) if birthyear of child $\geqslant$ (tax year - 1)
- 2004-: (c18120 + c18121 + c66200 + c66201) - (c18020 + c18021 + c66200 + c66201) if birthyear of child $\geqslant$ (tax year - 1)

Social insurance surplus : We distribute the macro aggregate proportional to employer's and employee's social insurance contributions and social insurance benefits (pensions and replacement incomes): $ansozbtr + agsozbtr + pensbr + replace$. See subsection A.2.2.1 for the NA definition.

A.3.2.2. Pretax national capital income

	Concept	SNA 2008	Strategy
	Pretax personal factor capital income		
–	Investment income payable to pension entitlements	D442, S14+15	distribute, see sec.A.3.2.1
+	Capital share of pension and other social insurance benefits	D442, S14+15	
+	Capital component of pension and other social insurance surplus	0%	
–	Pretax government factor income		distribute, see sec. A.3.1.3
=	Pretax capital income		

A.3.3. Posttax national income

Posttax national income results from deducting direct taxes and adding the value of cash and in-kind transfers (cash non-insurance benefits, in-kind transfers, collective expenditures). Posttax disposable income deducts direct taxes and only adds cash transfers (*excluding* all in-kind transfers and collective expenditures).

A.3.3.1. Posttax disposable income

Posttax disposable income deducts direct taxes and only adds cash transfers (*excluding* all in-kind transfers and collective expenditures).

	Concept	SNA 2008	Strategy
	Pretax national income		
+	Imputed taxes on production net of subsidies	D2-D3, S13	distribution
–	Current taxes on income and wealth received by the government	D5, S13	microdata uprate & distribution
+	Social assistance benefits in cash	D623, S14+15	microdata uprate & distribution
=	Posttax disposable income		

D2-D3, S13, rec.: Different to pretax income, in posttax income we distribute this component proportional to consumption. Net taxes on production include consumption taxes with its main

component VAT, but also property taxes. We distribute both components differently. First, we distribute consumption taxes (mainly VAT) proportionally to consumption using the consumption patterns by decile from *Einkommens- und Verbrauchsstichprobe (EVS) 2018*. Second, we distribute property taxes (*Grundsteuer A+B*) proportionally to to the sum of imputed rents (*nos**) and income from renting and leasing (*vuvbr*).

D5, S13, rec.: Current taxes on income, wealth, etc. received by the government (S13) are paid by 1) the corporate sector (S11 + S12), 2) resident households and non-profit organizations (S14 + S15) and 3) non-residents net of residents paying taxes to another country (S2). In the microdata we can identify the following tax items:

- Personal income tax including the solidarity surcharge (*Solidaritätszuschlag*); variable *estsoli = c65613 + (c66975/100)*[6].
- Local business tax (*Gewerbesteuer*) (LBT) on business incomes (variable *gewstgew*; for more details see section A.3.4)
  - 1992 - 1998: *gewst* is simulated (see section A.3.4).
  - 2001-2007: The dataset provides information on the product of basic rate × taxable profits (= Gewerbesteuermessbetrag *c44085/6*) and the LBT credit in accordance with §35 EStG (*c65630*). We add the average local scaling factor (*Hebesatz*) for each federal state to the database and estimate the tax to be paid as follows: LBT = basic rate × taxable profits (= Gewerbesteuermessbetrag *c44085/6*) × avg. local scaling factor.
  - 2010-: The dataset provides information on the product of basic rate × taxable profits (=Gewerbesteuermessbetrag, *c44064/5 + c44068/9 + c44085/6*), the LBT credit in accordance with §35 EStG (*c65630*) and the LBT liability (*zu zahlende Gewerbesteuer, c44066/7 + c44070/1 + c44081/2*). We use the latter directly as our LBT measure. We identify outlier as observations with an implicit local scaling factor of less than 2 (statutory minimum) or more than 5 (2010), 7.5 (2016) or 9 (in Rheinland-Pfalz) and replace their LBT with the simulated amount as described for 2001-2007.
- Business taxes (*Gewerbesteuer*, *Körperschaftsteuer*) on dividends from corporations (see Table A.3.1 for procedure to compute taxes from observed net dividends); resulting variables *gewstdiv, gewstdivgew2, gewstdivsa2, kstsolidiv1 and kstsolidivgew2, kstsolidivsa2* for taxes on dividends on capital and business respectively
- Business taxes (*Gewerbesteuer*, *Körperschaftsteuer*) on retained earnings which we cannot observe in microdata.

We adopt different strategies to harmonize the subcomponents of D5, S13 with NA aggregates across the PIT filer income distribution.

PIT After summing up the observed income tax payments, a small gap to the NA aggregate remains. This gap likely arises from the PIT payments of non-residents (D5, S2), which we don't observe in the PIT microdata. We proportionally uprate the microdata aggregate of the personal income tax including the solidarity surcharge to the income tax paid by the household and NPISH sector and non-residents (D5 paid, S14 + S15 + S2). We are aware that non-residents are not

[6] Solidarity surcharge in cents in dataset.

part of our microdata. However, we have to attribute this aggregate to a group to reach the full amount of taxes received by the government. This is one of the small conceptual inconsistencies in the DINA method.

Business taxes: We do not have information on how much business taxes fall on distributed vs undistributed profits. Therefore, we do not uprate the simulated business taxes based on observed business incomes and dividends directly. Instead, we start with the NA aggregate *direct taxes paid by the corporate sector* (D5, paid, S11+S12). First, we deduct the simulated businesses taxes based on microdata business incomes and dividends (*gewstgew + gewstdiv + gewstdivgew2 + gewstdivsa2 + kstsolidiv1 + kstsolidivgew2 + kstsolidivsa2*). Second, we separate the remaining aggregate into a) the household component of business taxes and b) the government component of business taxes. The assumption here is that households as well as the government hold shares in the corporate sector. To separate both components, we use the same share already used to divide the primary income of the corporate sector into the household and government component based on the equity & investment fund shares held by the households vs the government (see section A.2.1.2, equation 1). Last, we distribute the household component of business taxes proportional to the personal component of the primary income of the corporate sector (B5n, S11 + S12), that is proportional to positive dividend and quasi-corporate income. Implicitly, the distribution of remaining business taxes follows our assumptions on the distribution of retained earnings. The government component of business taxes is distributed proportional to government income, i.e., proportional to pretax personal factor income.

D623, S14+15: Social assistance benefits in cash

- D623 consists of means-tested and other social benefits. The PIT data hardly provides information on means-tested benefits since these are mainly directed towards non-filer tax units which we observe in SOEP data (see section A.4). Only, two non-means-tested benefits, that are included in D623, can be identified in PIT data:
  - Parental leave benefits (*Elterngeld*, *Betreuungsgeld*) are already identified at the stage of pretax national income as subcomponent of replacement incomes (*Einkommens-/Lohnersatzleistungen*). We assume that replacement incomes (*c18020,c18021,c66200, c66201*) are entirely due to parental leave for those tax units who had a child in the tax year or the year preceding it: [$c18020+c18021+c66200+c66201$] if birthyear of child $\geqslant$ (tax year $-$ 1); for 1992 and 1995, we do not observe the age of the children. We therefore assume that mothers (identified as child benefits recipients (*ef42*)) between the age of 20 and 40 receiving replacement entirely receive parental leave benefits ($par = 1$ if $v54 >= 2$ & $v54 <= 3$ & $v42 > 0$ & $v42! = .$).
  - Child benefit (*Kindergeld*)
    - 1992 - 1995: Before the child benefit reform in 1996, both child tax allowance and child benefit were granted to all families. We observe the number of children eligible for the child tax allowance (*kz850, kz851, kz624-kz627*) and the amount of child tax allowance (*kz455, kz628*), but not the resulting tax relief. Based on the number of eligible children and the official child benefit amounts (1st child 70 DM, 2nd child 130 DM, 3rd child 220 DM, 4th child 240DM), we simulated the resulting child benefit. Only the child benefit for the first child was independant of income, for all other children we simulate reduction amounts which increase with net income above an allowance. Additionally we add the

child benefit supplement (*Kindergeldzuschlag*) targeted at low-income households. We simplify the legislation by attributing the supplement to tax units with a statutory (tarifliche) income tax of zero.[7] Note that we only simulate the child benefit (Kindergeld) but not the tax relief resulting from child tax allowance (Kinderfreibetrag) as we cannot clearly deduct the income dependant tax relief from the general tax allowance.

* 1998-2007: We observe the granted amount of child benefits (*c65880*) as well as the birth year of each child in the household (*c36*16*). For filers, we additionally observe the possibly granted child allowance (Kinderfreibetrag, *c66998*). For most cases, we use information from the variable *c65880* (granted child benefits) directly. However, when the granted child benefit *c65880* is unreasonably low given the number and age of children in the household, we correct this variable using own estimations based on the number of children and the nominal monthly child benefit. The challenge is that not all children in the household are fully eligible for child benefits due to 50% eligibility for separated parents with shared custody. For the correction, we, first, deduce the number of children eligible for child benefit (*anzukind*) using information on the possibly granted child allowance (Kinderfreibetrag, *c66998*). For payroll tax cases, for which we do not observe *c66998*, we use the number of children living in the tax unit's household. Second, we simulate the theoretical amount of child benefits per tax unit from the amount officially granted for each child (external information) and the estimated number of eligible children (*anzukind*): *childbent* = official amount per child × anzukind. This correction is only applied if the amount of child benefit (*c65880*) is below the subsistence level. We do not take into account the additional tax relief for high income earners due to the more favorable effect of the *Kinderfreibetrag* as we cannot clearly deduce the tax relief from the tax allowance amount. This means that this additional tax relief still reduces taxes paid and is not added back to the child benefit. This is analogous to our treatment in the national accounts.

* 2010-2016: We observe the granted amount of child benefit (*c65880*) as well as the birth year of each child in the household (*c36*16*) and the granted amount of child benefits for each child separately (*c36*15*). For most cases, we use information from the variable *c65880* (granted child benefits of tax unit) directly. We correct for cases where this variable is unreasonably low given the information on children and their eligibility. For the correction, we, first, deduce the number of eligible children from the granted amount of child benefits for each child (*c65*15*). Second, we estimate the amount of child benefits per tax unit from the amount officially granted for each child (external information) and the estimated number of eligible children (*anzukind*): *childbent* = official amount per child × anzukind. This correction is only applied if the amount of child benefit (*c65880*) is below the subsistence level. We do not take into account

[7] Full legislation: https://www.bgbl.de/xaver/bgbl/start.xav?start=%2F%2F*%5B%40attr_id%3D%27bgbl190s0149.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl190s0149.pdf%27%5D__1678817325882 and here https://www.bgbl.de/xaver/bgbl/start.xav?start=//*%5B@attr_id=%27bgbl194s0168.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl194s0168.pdf%27%5D__1678817327788.

the additional tax relief for high income earners due to the more favorable effect of the *Kinderfreibetrag* as we cannot clearly deduce the tax relief from the tax allowance amount. This means that this additional tax relief still reduces taxes paid and is not added back to the child benefit. This is analogous to our treatment in the national accounts.

– We mean-value impute social assistance and ALG II benefits for filers whose income does not exceed the living wage and who thus can supplement their income via social assistance and ALG II (so-called *Aufstocker*)

A.3.3.2. Public Spending and in-kind transfers

	Concept	SNA 2008	Sector
	Individual consumption expenditure	P31	S13
+	Collective consumption expenditure	P32	S13
=	Public Spendings and in-kind transfers		

We use information on the subcomponents of P3, S13 (Final consumption expenditure by the government) given as the Eurostat Classification of the Functions of Government (COFOG) in the national accounts (see section A.2.3.2) to provide the following distribution scenarios:

1. We distribute the NA aggregate P3, S13 proportional to posttax disposable income (*variable psikt1*). This only increases the income level, but keeps the disposible income distribution as is.
2. We distribute all final consumption expenditure lumpsum to individuals (*variable psikt2*).
3. We distribute education expenditure proportional to posttax disposable income given the high correlation between income (of parents) and educational achievements. Education expenditures are given as separate COFOG division 9. All other expenditure is distributed lumpsum (*variable psikt3*).
4. Following the benchmark series of Piketty et al. (2018) and Garbinti et al. (2018), we distribute health expenditure (COFOG No. 7) lump-sum and all other expenditures proportional to posttax disposable income (*variable psikt4*).

A.3.3.3. Primary government surplus

The macro aggregate of government surplus or deficit is distributed following Piketty et al. (2018, p. 21). We add 50% to taxes paid on income and wealth (D5, S13, *variable govd5up*) and 50% to means-tested benefits received (D623, S14+15, *variable govd623up*).

A.3.4. From net to gross business incomes

We have to estimate business taxes for two reasons: First, business incomes in PIT are observed net of business taxes. If we want to obtain gross business incomes, we have to add back business taxes. Second, to go from pretax national income to posttax national income, we have to deduct business taxes.

There are two types of business taxes in Germany, which we simulate: Local business tax (LBT, *Gewerbesteuer*) and corporate tax (*Körperschaftsteuer*).

Local business tax is levied on corporate and unincorporated firms with exceptions for agricultural businesses, the free professions and the public sector. PIT returns record business incomes (*Einkünfte aus Gewerbe, Einkünfte aus selbständiger Tätigkeit*) and dividends net of LBT. In the following, we describe our procedure to add back LBT to business income from unincorporated firms (Section A.3.4.1) and to dividends from corporations (Section A.3.4.2).

The LBT law (*GewStG*) determines the LBT liability in three steps:

1. Business profits (*Gewerbeertrag*) at the level of the firm is defined as the sum of profits (*Gewinn gem. EStG bzw. KStG*) and additions (*Hinzurechnungen*, §8 GewStG) reduced by deductions (*Kürzungen*, §9 GewStG) and losses of preceding years (*Gewerbeverlust aus Vorjahren*) and annual trade tax exemption (*Freibetrag*)

2. The resulting business income is multiplied by the basic rate (*Steuermesszahl*) set at the federal level to arrive at the *Steuermessbetrag*:

$$\text{Steuermessbetrag} = \text{business profits} \times \text{basic rate} \tag{2}$$

   Until 2007, the basic rate was 5% for corporations. For business partnerships and sole proprietors, the basic rate was increasing with business income

   - 1% for business incomes up to 12,000€
   - 2% for business incomes up to 24,000€
   - 3% for business incomes up to 36,000€
   - 4% for business incomes up to 48,000€
   - 5% for business incomes exceeding 48,000€

   In 2008, the basic rate was set uniformly to 3.5% of business income.

3. The *Steuermessbetrag* is then multiplied by the local scaling factor (*Hebesatz*) set at the municipal level. These multipliers greatly vary across municipalities with a legislative mininum of 200% (§16 GewStG):

$$\text{LBT} = \text{Steuermessbetrag} \times \text{local scaling factor} \tag{3}$$

A.3.4.1. LBT on business incomes

LBT liability is recorded in PIT data since 2010. For the preceding years, we have to simulate LBT.

1990s: We only have business income of the PIT filer. We do not have information on business profits (*Gewerbeertrag*) at the level of the entire firm nor on the amount of *Gewerbesteuermessbetrag* nor do we know the number of shareholders of the business that produced the individual's business income. Thus, we simulate the LBT payment of each business income recipient using the observable taxable business incomes (*c65123-c65133*) in three steps :

1. Since the local business tax is assessed at the firm level, we roughly impute the average number of shareholders of partnerships and business income groups using the official statistics of income tax for partnerships (*Statistik über die Personengesellschaften und Gemeinschaften*). These statistics provide the total income of partnerships by the number of shareholders.

2. The *Steuermessbetrag* is computed from the simulated business income at the firm-level as explained in Equation 2 (*gewsttarif2.do*). Then, we allocate the individual's share in the firm's *Steuermessbetrag* to the business income taxpayers.

3. The LBT is computed as explained in Equation 3. We use federal state averages as proxy for the local scaling factor (*gewsthebesatz.do*).

2001, 2004 and 2007: The *Steuermessbetrag* is given in the PIT data (*c44085/6*).The LBT is computed as explained in Eq. 3. We use federal state averages as proxy for the local scaling factor (*gewsthebesatz.do*) because we observe the federal state of the taxpayer (and not the municipality of the firm). In Germany's family-firm structured economy, it is likely that the owner (=taxpayer) resides close to the firm, i.e., in the same federal state.

2010-: LBT is given in PIT data (*c44066/7*, *c44070/1* & *c44081/2*).

A.3.4.2. LBT on dividends

We add back the local business tax (LBT) to dividends from corporations received by the personal taxpayer or at the level of the partnership. For dividends, we apply the national average of the local scaling factor (approx. 400%). As opposed to business income, it is unlikely that the recipient of dividends from corporations resides in the same federal state as the corporation. Hence, we use annual national averages instead of federal state averages. The basic rate on corporations is 5% until 2007 and 3.5% since 2008. The average LBT rate then can be estimated by: $basic\ rate \times average\ scaling\ factor$. Table A.3.2 lists average scaling factors and the resulting average LBT rate by year.[8] We use the cash dividend, i.e., net of corporate taxes to obtain the gross dividend:

$$\text{Gross dividend} = \frac{\text{Cash dividend}}{1 - \text{average LBT rate}} \tag{4}$$

Table A.3.2.: National averages of basic rate (*Steuermesszahl*), local scaling factor (*Hebesatz*) and local business tax (*Gewerbesteuer*) rates. Source for local scaling factor: Statistisches Bundesamt (1993-2022).

Year	Basic rate	Average scaling factor	Average LBT rate
1992	5% for corp.	x 370%	= 16.6%
1995	5% for corp.	x 376%	= 16.6%
1998	5% for corp.	x 390%	= 16.6%
2001	5% for corp.	x 388%	= 16.6%
2004	5% for corp.	x 388%	= 16.6%
2007	5% for corp.	x 389%	= 16.6%
2010	3.5% for all	x 390%	= 14%

A.3.4.3. Corporate tax on dividends

We add back the corporate tax including the solidarity surcharge to dividends received directly from corporations (*variable kstsolidiv1* and indirectly via quasi-corporations and sole proprietorships (*variable kstsolidivgew2 & kstsolidivsa2*) applying the tax rates displayed in Table A.3.1.

[8] Before 2008, the LBT was deductible from business income of the corporations, thus the LBT rate is computed as: $LBT = basic\ rate(m) \times average\ scaling\ factor(h)\ \ (business\ income(Y) - LBT) = Y \times \frac{mh}{1+mh} = Y \times \frac{5\% \times 400\%}{1+5\% \times 400\%} = 16.6\% \times Y$

A.3.5. Treatment of high losses in the light of tax reliefs

In the German tax law, numerous tax allowances and reliefs reduce taxable income. This applies, in particular, to business and self-employment incomes as well as incomes from renting and leasing. Tax reliefs for business and real estate investments represent an important tool in German tax law to foster investments in small and medium-sized firms. These tax reliefs work through special - often unrealistically high - depreciation rates on business assets or real estate. These include, for example, a special write-off for small and medium-sized enterprises (SMEs) (*§7g EstG Sonderabschreibung zur Förderung kleiner und mittlerer Betriebe*) or the a tax-free depreciation reserve for SME (*§7g, p. 3 EstG Ansparabschreibung für KMU*). These tax reliefs can create high negative taxable incomes despite positive incomes from business or from renting and leasing. Our aim is to add back these tax reliefs to pretax incomes. Table A.3.3 lists the tax reliefs, for which we have information in the tax data and the years they were legally effective.

Table A.3.3.: Tax reliefs

	'92	'95	'98	'01	'04	'07	'10	'16
Tax reliefs targeted at businesses								
§7g EstG Sonderabschreibung zur Förderung kleiner und mittlerer Betriebe (*c35065*)	✓	✓	✓	✓	✓	✓	✓	
§7g, par. 3 EstG Ansparabschreibung für kleinere und mittlere Betriebe (*c35026*)	✓	✓	✓	✓	✓	✓		
§2 -§4 FoerdG Sonderabschreibungen für betriebliche Investitionen (*c35031*)	✓	✓	✓	✓	✓	✓		
§5 FoerdG Gewinnabzug bei Land- und Forstwirten (*c35032*)	✓	✓	✓	✓	✓	✓		
§82f EStDV Bewertungsfreiheit für Seeschiffe und Luftfahrzeuge (*c35043*)	✓	✓	✓	✓	✓			
§80 EStDV Bewertungsabschlag für Güter ausländischer Herkunft (*c35035*)	✓	✓	✓	✓	✓	✓	✓	
§7g Abs. 7 EStG Ruecklage Existenzgründer (*c35027*)				✓	✓	✓		
Tax reliefs targeted at real estate								
§3 & 4, p. 1-2 FörderGG Special depreciation (Acquisition and construction) (*c35029*)	✓	✓	✓	✓	✓	✓		
§3& 4, p. 3 FörderGG Special depreciation (subsequent construction costs) (*c35030*, 2010: *c25*32*)	✓	✓	✓	✓	✓	✓	✓	

Note: We top-code the variable *c35065* at EUR 3 mio. and *c35026* at EUR 800,000.

Importantly, such tax reliefs were used to fiscally support investment in East German federal states after reunification in 1990. Many of these special tax reliefs were initiated in the 1990s and granted until the early 2000s. For example, the *Fördergebietsgesetz* (*FörderGG*) laid down special depreciation rates on business investments and real estate which could reach up to 50% of the initial acquisition value. As a result, many personal taxpayers were able to deduct high write-offs from comparably small incomes from renting and leasing, thereby creating high losses from renting and leasing, which in turn greatly reduced their overall taxable income. This scheme was widely used among West German top income earners to reduce taxable income in the 1990s. As a result, we observe very small, often negative incomes from renting and leasing in PIT data.

The aggregate of taxable income from renting and leasing in Germany was negative throughout the 1990s (Statistisches Bundesamt, 2022b).
As these negative incomes do not stem from *real* losses, but rather from the possibility to deduct a substantial proportion of the acquisition value, we want to correct for this effect. We take all information on these schemes available in PIT data into account and add back the full amount of tax reliefs that allow for unrealistically high depreciation rates to taxable income. We top-code the first two items listed in Table A.3.3. The special write-off for small and medium-sized enterprises (*§7g EstG Sonderabschreibung zur Förderung kleiner und mittlerer Betriebe*) is granted to firms with business assets below 235,000€. We top-code this variable (*c65065*) at 3 million €. Legally, the write-off can reach up to 40% of acquisition costs for an investment declared by the firm. Our maximum of 3 million € would then result from an investment of 7.5 million. We argue that larger investments are quite unlikely for small and medium-sized firms. For the tax-free depreciation reserve for SMEs (*§7g, par. 3 EstG Ansparabschreibung für KMU*), we restrict the maximum value considered realistically to 800,000€. Conditional on having received this type of income, the first group of tax reliefs listed in Table A.3.3 is added back to business income and the second group of tax reliefs is added back to income from renting and leasing.

A.3.6. Imputation of tax-exempt capital income

A.3.6.1. Imputation of imputed rents (B2n, S14) and interest paid on mortgage payments of home owners (D4, S14+15 paid)

In PIT data, we do not observe imputed rents (B2n, S14) and interest paid on mortgages (D4, S14+15 paid) only to a very limited extent. We impute both items using mean-value imputation based on information of the SOEP across the following groups: quintiles of gross income (*Gesamtbetrag der Einkünfte*) x marital status x 5-year age groups (20-70+ years) of the main earner and region (East/West Germany). We use the distribution of imputed rent and interest payment on mortgages by homeowners observed in SOEP data to impute this item to tax units with similar characteristics observed in PIT microdata. Before we can mean-value impute the interest paid on mortgages, we first have to deduce it from the information on mortgage payments in the SOEP: To extract the interest share of mortgage payments, we use a simple model based on information on mean effective interest rates for housing mortgages to private households (Deutsche Bundesbank, 2022, Tab. III. Geld- und Kapitalmarkt, 3. Einlagen- und Kreditzinssätze deutscher Banken) and assuming a credit period of 25 years (see *Input 10* in Table *Macro_ Germany_ 08_ 2019.xlsx*). We multiply the estimated interest share with mortgage payments by homeowners recorded in the SOEP (see section A.4.1.2). We then only transfer the information on the interest share of mortgages to PIT data.

Interest on mortgage payments contribute to the item of property income paid by households (D4, S14+15 paid). The NA aggregate includes mortgages and interest payments by home-owners for owner-occupied housing, landlords that rent out their property and firms paying interest on loans. We solely impute mortgage payments for home-owners, but not for landlords or firms because mortgage payments for tenant-occupied housing or firm investments can be deducted in tax returns and are, thus, already consolidated with profits in either income from renting and leasing (*c65240*) or business income (*c65120, c65100, c65140)*).

A.3.6.2. Imputation of capital incomes below the savings allowance

Dividend and interest income must be declared only if exceeding the annual savings allowance. As a consequence, dividend and interest income below the savings allowance is not reported and not included in PIT data. The year-specific savings allowance is listed in Table A.3.4.

Table A.3.4.: Savings allowance and deductible professional expenses

Year	Single filer	Splitting couple	Professional expenses
1992	600 DM (306.78€)	1200 DM (613.55€)	100 DM (51.13€)
1993 - 1999	6000 DM (3067.75€)	12,000 DM (6135.50€)	100 DM (51.13€)
2000-2001	3000 DM (1533.88€)	6000 DM (3067.75€)	100 DM (51.13€)
2002-2003	1550 €	3100€	51 €
2004-2006	1370 €	2740€	51€
2007-2008	750€	1500 €	51 €
since 2009	801 €	1602 €	51 €

Note: Conversion rate of 1.95583 DM per EUR.

We use a mean-value imputation based on the SOEP taking into acount all tax units declaring annual capital incomes between EUR 50 and the year-specific savings allowance.[9] In a next step, we split this subsample into groups according to 5-year-age groups of the main earner (20-70+ years), 7 gross income groups (*GdE*, excluding capital incomes, P0,20,40,60,70,80,90), marital status and region (East/West Germany). For each group, we compute the number of tax units with positive capital income (below the savings allowance) and their mean annual capital income. We transfer this information to the PIT microdata. Last, we randomly assign the mean capital income to tax units with formerly zero capital income in the PIT data so that the number of tax units as well as the sum and mean of capital income by group matches the SOEP data information.

Unfortunately, the capital income variable (*divdy$$*) in SOEP data is at the household level and does not distinguish between dividends and interest income. For splitting couples, we equally distribute individual interest and dividends between spouses. Imputed capital incomes are randomly labeled as either 100% interest income (66% of tax units) or 100% dividends (34% of tax units).

A.3.7. Imputation of dividends and interest income after the introduction of dual income taxation in 2009

Since the introduction of dual income taxation in 2009, dividend and interest income is subject to a flat withholding tax of 25% and not systematically documented anymore in personal income tax data.[10]

[9] The lower threshold is to exclude minimal interest payments which might be prone to measurement error in a survey.

[10] The dualization of the income tax schedules does not necessarily lead to a lack of data on top incomes: in Scandinavian countries, information on capital income and other income can be linked using the individual taxpayer-ID. In Germany, data linkage by taxpayer is not available. For some tax units it is beneficial to declare capital income in their income tax declaration despite the reform, e.g., if the flat rate exceeds their personal income tax rate. However, the size of capital income still documented in PIT microdata is negligible.

We transfer the averaged distribution of dividends and interest of 2001-2007 to income years after 2008. More precisely, we build the share of dividends (interest) that each group receives of the total of dividends (interest):

$$share_{i,dividends} = \frac{dividends_i}{\sum_i^I dividends} \tag{5}$$

And the share of recipients of dividends (interest) in the population:

$$share_{dividend\ recipients} = \frac{dividend\ recipients}{population} \tag{6}$$

And the share of recipients of dividends (interest) in each group compared to the full population of recipients:

$$share_{i,dividendrecipients} = \frac{dividend\ recipients_i}{\sum_i^I dividend\ recipients} \tag{7}$$

i denotes one of 108 groups defined by fractiles of gross incomes[11] (P0,P10,...,P80, P90, P95,P96, P97, P98, P99, P99.5, P99.9, P99.99), the gender of the tax units (for single units), the marital status and the region (east/west).

For income years after 2008, we set the unsystematically reported dividends and interest in the tax data to zero and overwrite the whole distribution of dividends and interest above the savings allowance. We keep capital incomes under the savings allowance in the dataset. We then directly distribute the national accounts aggregates of dividends (share of D42, S14 + S15[12]) and received interest (D41, S14+S15, received)according to the estimated distributions of 2001-07.

The imputation method relies on two assumptions: First, we assume that the distribution of dividend and interest income over the groups i does not substantially change over time. Yet, the tax reform might have induced income shifting between sources, probably even to different degrees across the distribution. Then, the assumption would be challenged: if richer tax units shift more income towards dividend income, i.e. incorporate former unincorporated firms, our imputation would understate top income shares and our results would mark a lower bound. Second, we assume that the aggregate development of dividends and interest in national accounts reflects the development of taxpayers' dividend and interest income in the PIT definition even though income concepts in PIT and NA differ slightly.

[11] We define gross income here as: gross inc = Gross business incomes + Gross wages (without employer's SSC) + Income from renting and leasing + imputed rents + Pensions (incl. civil servants pensions.

[12] We take the share of dividends relative to dividends and income from partnerships 2001-2007 to compute the macro share.

A.4. Creating DINA concepts with SOEP

We use the SOEP data for two purposes: First, we elicit non-filing households to complement the tax-filing population covered by the PIT microdata. Second, we extract information on the distribution of national account subcomponents which are not recorded in PIT microdata to impute those incomes for tax filers. While we described our procedures for the second purpose in section A.3.6, this section describes how we elicit non-filing households from SOEP and align their incomes to meet national account definitions.

We mainly use the SOEP-equivalent file (pequiv) if not indicated differently. Our variable names and SOEP variable names are in italics. To match the PIT data structure, we create tax units in SOEP data and create all DINA concepts analogously to PIT data. However, SOEP data in some cases is not as detailed as PIT data, e.g. we cannot distinguish between dividends (*D42, S14+15*) and interest (*D41, S14+15*) or between business income of sole proprietors (*B3n, S14+15*) and business income from partnerships (*D42, S14+15*) in SOEP data.

A.4.1. Pretax factor income

A.4.1.1. Pretax factor labor income

	Concept	SNA 2008	strategy
	Compensation of employees	D1, S1	micro & uprate
+	Labor share (70%) of net mixed income	B3n, S14+15	micro & uprate
+	Labor comp. of imputed taxes on production net of subsidies	D2-D3, S13	distribute, see sec. A.3.1.1
=	Pretax factor labor income		

D1, S1: Compensation of employees received by the household sector (variable *ce**)

- Gross wage (variable *lohnbr*) includes :
  - Wages or salary from the main job of the previous year (*ijob1$$*)
  - Income from secondary employment of previous year (*ijob2$$*)
  - 13th and 14th monthly salary of previous year (*i13ly$$, i14ly$$*)
  - Christmas bonus of previous year (*ixmas$$*)
  - Vacation bonus of previous year (*iholy$$*)
  - Profit-sharing of previous year (*igray$$*)
  - Other bonuses in previous year (*iothy$$*)
  - Indemnity payments of previous year (*idemy$$*)
  - Commuting expenses, travel grants of previous year (*itray$$*)
- Employer's social insurance contributions are simulated from yearly information of contribution rates and the contribution assessment ceiling using the same external information (*sozbeitrsaetze.do*), assumptions and method as for PIT data (see sections A.3.1.1). Earnings

from "mini-jobs"[13] are included in wage income items of *lohnbr*. Employer's contributions rates for mini-jobs are different from regular social security contribution rates. We identify "Minijobber" using the monthly income ceiling (7_stsm_mod.do) and separately simulate employer's social contributions for mini-jobs (*33_zusammenedit.do*)

B3n, S14+15 Labor share (70%) of net mixed income (variable *nmi* = *soleprop* + *vuvbr*)

- Income from sole proprietorship (*Einzelunternehmer, var. soleprop*): We use variable *iself$$*, i.e., self-employment incomes. As only low-income observations from SOEP enter our full distribution of both SOEP and PIT microdata, we assume that these low-incomes stem from sole proprietors and not from partnerships.
- Income from renting and leasing (*vuvbr*) is gross income from renting and leasing net of maintenance costs and mortgage payments: *renty$$ - opery$$ - mortgagerent* ; We subtract mortgage payments on rented apartments to harmonize the variable with tax data where mortgage payments on rented apartments are deductible; *mortagerent* is from SOEP dataset {wave}h: the variable name depends on the income year, income year = survey year - 1; income year 1992: *jh4202, 1995: mh4202, 1998: ph4202, 2001: sh4002, 2004: vh3902, 2007: yh4002, 2010: bbh4002*

A.4.1.2. Pretax factor capital income

	Concept	SNA 2008	Strategy
	Capital share (30%) of net mixed income	B3n, S14+15	micro & uprate
+	Net operating surplus of the household sector	B2n, S14+15	micro & uprate
+	Property income consolidated of households	D4, res. - uses, S14+15	micro & uprate, mean-value imp. of mortgage interest
+	Personal component of primary income of the corporate sector	B5n, S11+12	distribute, see sec. A.3.1.1
+	Capital comp. of imputed taxes on production net of subsidies	D2-D3, S13	distribute, see sec. A.3.1.1
=	Pretax factor capital income		

B3n, S14+15: Capital share (30%) of net mixed income (variable *nmi** = *soleprop* + *vuvbr*)

- see "labor share of net mixed income" in section A.4.1.1

[13] Mini-jobs are limited by a monthly earnings ceiling of EUR 450 (2020) or temporary employment of up to 70 days per year (2020). Minijobber are not obliged to pay social insurance contributions but their employer's have to pay contributions to social security.

B2n, S14+15: Net operating surplus (variable *nos*)

- Imputed rent (*i11105$$*) - is kept for non-filer cases & used for mean-value imputation of filer cases to tax data (see section A.3.6.1)

D4, S14+15, consolidated (received - paid): (variable *npi=pir-pip*)

- Interest & dividend income at the household level (*divdy$$*): We assume that capital incomes of all non-filers (incl. those paying only payroll taxes) receive only interest income (D41, S14+15) and zero dividends. Since we only use low-income tax units from the SOEP for our final dataset, this is a reasonable assumption. Only 10% of German households own shares in corporations and these households are located at the top of the income distribution. Shareholding is in general more dominant for middle- and high-income households, in Germany in particular.
- Interest and mortgage payment for owner-occupied housing from dataset *h* is used to estimate the interest share of mortgage payments (part of D41 paid). This interest share is mean-value imputed to PIT data (see section A.3.6.1) and substracted from interest received. Variable name depends on the income year, income year = survey year - 1; income year *1992: jh3301, 1995: mh3301, 1998: ph33, 2001: sh32, 2004: vh29, 2007: yh30, 2010: bbh30*.

A.4.2. Pretax national income

A.4.2.1. Pretax national labor income

	Concept	SNA 2008	Strategy
	Pension and other social contributions	D61, uses, S14+15	simulated from microdata
+	Investment income payable to pension entitlements	D442, S14+15	distribute, see sec. A.3.1.1
–	Pension and other social insurance benefits	D621 +D622, S14+15	microdata
=	Pension and other social insurance surplus		

D61, S14+15 paid : Social insurance contributions paid

- We simulate employee's (variable *ansozbtr*) and employer's contributions (variable *agsozbtr*) for pension, unemployment and compulsory health and long-term care insurance based on income liable for insurance contributions, assessment ceilings and socio-economic status (self-employed, employee, civil servant) in a consistent way for SOEP and tax data. Our simulation results are comparable to the SOEP-variable *i11112$$* (household social security taxes).
- Business owners' and self-employed contributions: *untsozbtr= 0* due to missing information. Due to only using non-filer cases with income below the personal tax allowance from SOEP, this item is negligible.

D621 + D622, S14+15: Social security benefits

- Pensions (var. *pensbr*) include old-age & disability pensions (*ioldy$$*), widows and orphans pensions (*iwidy$$*) and company pensions (*icomp$$*). We subtract civil servants pensions (Versorgungsbezüge, variable *iciv1$$ + iciv2$$*) from this aggregate and include it as a separate item to make the variable comparable to the PIT variable.
- Civil servants pensions (Versorgungsbezüge): $iciv1\$\$ + iciv2\$\$$ (var. *vgbzbr*)
- Private pensions: *iprvp$$* (var. *privpensbr*)
- Replacement incomes (*replace*) include insurance-based unemployment benefits (*iunby$$*), old-age transition benefit (*ieret$$*), insurance-based subsistence allowance (*Übergangsgeld*, part of the insurance-based unemployment scheme; variable *isuby$$*) and long-term care allowance (*Pflegegeld, nursh$$*)

A.4.3. Posttax national income

A.4.3.1. Posttax disposible income

	Concept	SNA 2008	Strategy
	Pretax national income		
+	Imputed taxes on production net of subsidies	D2-D3, S13	distribute, see sec. A.3.1.1
–	Current taxes on income and wealth received by the government	D5, S13	microdata & uprate
+	Social assistance benefits in cash	D623, S14+15	microdata & uprate
=	Posttax disposable income		

D5, S13: Taxes on income and wealth

- Income tax including solidarity surcharge (*Solidaritätszuschlag*) : original = *i11111$$*, constructed var.: *estsoli**,)
- Local business tax (*Gewerbesteuer*) on business incomes (variable *gewstgew*) - not included in SOEP, assumption gewstgew = 0 for low-income, non-filer observations picked from the SOEP
- Business taxes (*Gewerbesteuer*, *Körperschaftsteuer*) on dividends from corporations and received at the level of the partnership = 0 because we assume that non-filers' capital income is 100% interest income and 0% dividend income.

D623, S14+15: Social assistance benefits in cash D623 consists of means-tested and other social benefits. While in PIT data, we can only observe the non-means-tested child benefits and parental leave benefits, the SOEP provides more details on means-tested benefits relevant for non-filers. Below, we list all means-tested and non-means-tested benefits observed in SOEP data as categorized by NA into six subgroups (see sec. A.2.3.1):

1. Social assistance and war victims' welfare (*Sozialhilfe & Kriegsfürsorge*): We can only directly observe part of the subcomponents of this concept in SOEP data, namely
   a) basic subsistence income for the elderly (Grundsicherung im Alter): *ssold*$$,

b) regular social assistance (Hilfe zum Lebensunterhalt): *subst*$$,

c) social assistance for special circumstances (Sozialhilfe in anderen Lebenslagen): *sphlp*$$,

d) benefits for asylum seekers (foll. AsylbLG): *asyl*$$.

Overall the coverage in SOEP data comparable to the NA aggregates. We compare the micro aggregate of those items with national accounts, but do not harmonize them directly with the NA aggregate. The reason is that the declaration of benefits as social assistance vs unemployment benefit II is not always adequate at the municipality level. So-called "Optionsgemeinden" distribute benefits autonomously without assistance by the Federal Employment Agency. Those municipalities might not make a precise distinction between both benefit types. We thus harmonize social and unemployment assistance jointly with NA.

For other items, we impute benefits based on the NA aggregate and the number of (possible) recipients. Those include:

a) Eingliederungshilfe für behinderte Menschen: We use the information on employment status which includes an indication of working at a workshop for the disabled of the {wave}p dataset and distribute the NA aggregate to these tax units.

b) Hilfe zur Pflege indicate means-tested transfers for households that do not receive sufficient care allowances by the statutory care insurance to ensure sufficient long-term care in the house or in institutions. We distribute the macroeconomic aggregate to households indicating a person needing care in the household in the {wave}h dataset (wave year 2008, variable yh56). We only include households in the bottom 18% of the pretax national income distribution to align our number of recipients with the recipients indicated in the *Sozialhilfestatistik* of Destatis Genesis (Statistisches Bundesamt, 2023b).

c) Hilfen zur Gesundheit are provided to persons that are not a member of the statutory health insurance. The main group are homeless persons. We simulate this transfer as a latent transfers, ie distributing the macro aggregate to all homeless people as they potentially have access to this benefit. Based on data by the *BAG Wohnungslosenhilfe*, we infer about 900k homeless in 1992, about 250k in 2007 and 2010 and about 400k in 2016 (not including another 400k homeless refugees). We distribute the macro aggregate to the number of potential recipients who are single and with the lowest incomes in the SOEP dataset.

The NA aggregate also include in-cash benefits for child and youth welfare (*Jugendhilfe*) which we cannot identify in the microdata.

2. War victim pensions (*Kriegsopferversorgung*) includes war veteran pensions: $war1\$\$ + iwar2\$\$$

3. Unemployment assistance (*Arbeitslosenhilfe*) includes means-tested unemployment benefits (until 2004: *Arbeitslosenhilfe*, 2005-: *Arbeitslosengeld II*) and benefits of the education package (*Bildungs- und Teilhabepaket*): $iunay\$\$ + alg2\$\$ + educpac\$\$$

4. Child benefit (*Kindergeld*) includes child benefit and additional child benefit for low-income families: $chspt\$\$ + adchb\$\$$. Since we only use non-filer observations from SOEP, an additional tax relief from the child tax allowance (*Kinderfreibetrag*) is not applicable to this group.

5. Housing benefit (*Wohngeld*): *house*$$. We correct housing benefits when persons report unemployment benefits II and housing benefits as it is not possible to claim both. Recipients of unemployment benefits II receive assistance for renting a flat under the umbrella of the unemployment benefits. We thus add stated housing benefits to unemployment benefits II if the tax unit also receives unemployment benefits II.

6. Other benefits include maternity and parental leave benefits (*Mutterschaftsgeld*, *Elterngeld*), child care subsidy in income year 2013-15 (*Betreuungsgeld*), students' grants and advance child maintenance payment (*Unterhaltsvorschuss*): imaty$$ + chsub$$ + istuy$$ + iachm$$

Note that SOEP survey data suffers from an underrepresentation of persons in community accomodations, like student housing and homes for the elderly.

A.4.4. Imputation of non-filers

Our PIT microdata (*Lohn- und Einkommensteuerstatistik*) comprises approximately 36 million tax units (single filers and splitting couples, 2007) that either file a tax return or are included as employees liable solely to payroll tax (*reine Lohnsteuerfälle*). Employees liable solely to payroll tax do not have to file a tax return as their payroll tax is automatically withheld by the employer. Until 2004, tax units not filing income tax returns were asked to submit their information to the tax office. Because the submission was not enforced, we miss about 10% of the payroll tax paid in PIT data until 2001. Since 2007, payroll taxpayers are automatically included through the information by their employers.

To arrive at a distribution covering the entire population of 20+ years, we add non-filer observations from the SOEP to PIT data to construct a representative database. For 1992 to 2004, we additionally add payroll tax cases (*reine Lohnsteuerfälle*) from the SOEP.

The number of observations to be identified from SOEP is predetermined by the gap between population statistics and PIT data observations. We have to make sure that our group of non-filers matches population characteristics. For example, the sum of single women in our integrated SOEP+PIT database in a particular federal state should not exceed their total number in official population statistics. Based on population statistics, we take into account the composition of the population by age, gender, marital status, and federal state.

We use enumerated individual weights ($phrfe$) to weight SOEP observations. For married-couple tax units we assign the enumerated weight of the main earner to the entire tax unit. These weights match best the number of tax units in the population with a total of 49.5 million tax units of 20 years and above (see Bartels (2019)'s data appendix for a discussion on the number of potential tax units). For the imputation, we take into account the post-stratification procedure of the SOEP, i.e., the generation of cross-sectional weights. SOEP cross-sectional weights are adjusted to fit the marginal distributions of several variables based on the micro census (Kroh et al., 2018). For the imputation of non-filers, we add observations according to federal state[14], which is post-stratified in the SOEP at the household level, gender and 5-year-age groups (20+, 25+, 30+,..., 70+), which are stratified at a personal level in the SOEP (Kroh et al., 2018, 66ff.).

Non-filers are identified as follows:

1. We align the survey data to the format of the tax data, i.e., we create tax units from households according to individuals' information on marital status (*d11104$$, partz$$*

[14] To fit the post-stratification procedure, we pool observations for Berlin and Brandenburg, for Hamburg and Schleswig-Holstein, for Bremen and Lower Saxony and for Saarland and Rhineland-Palatinate.

from dataset pgen) and intra-household relations (*{wave}stell of dataset pbrutto*) (*4_taxunit_soep.do*).

2. We identify non-filers, filers and pure payroll cases (*reine Lohnsteuerfälle*) in the SOEP using information on business, employment and capital income, total earnings (*Gesamtbetrag der Einkünfte*), the personal tax allowance (*Grundfreibetrag*) and deduction possibilities (commuter's allowance, child allowance). Payroll cases are those tax units that earn above the personal tax allowance (*Grundfreibetrag*), but solely receive employment income. Thus, they do not have to file a tax return as taxes are deducted at the employer level.
3. Starting in 2007, those payroll cases are fully included in PIT data due to automatic information transfer from employers. Thus, we only consider non-filers from SOEP. Before 2007, non-filing payroll cases were underrepresented in PIT data and we also consider non-filers as well as pure payroll cases from SOEP.
4. We fill the gap between PIT data observations and official population statistics based on the following categories:
   - For single filers: By gender and 5-year-age group and federal state.
   - For splitting couples: By 5-year-age group of the male spouse and federal state.

If too many non-filer cases are observed in the SOEP for a specific category, we draw observations randomly. If too little non-filers are left in the SOEP, we adjust the weights proportionally (up to a factor of 2).

A.5. Uprating incomes to national accounts

In a final step of our analysis, we have to align the aggregates of our micro distribution created from the combination of PIT and SOEP data with NA aggregates. When we aggregate PIT+SOEP microdata and compare the aggregate to national accounts subcomponents, substantial gaps are revealed which run in both directions. In some cases, the national accounts' aggregate is higher than the PIT+SOEP aggregate, but also vice versa. The reasons for the gap between micro-data fiscal income and macro-data national income can be grouped into three components: tax-exempt labor income, tax-exempt capital income and production taxes. Yet, even after estimating as many of the tax-exempt components as possible, income tax data aggregates do not exactly match the national accounts aggregates. Hence, we have to uprate income tax data aggregates to the national account aggregate. For this, we uprate the raw income concepts described in sections A.4 and A.3 to match the NA aggregate. The sources of NA aggregates are described in subsection A.1.1.

A.5.1. Investigating and resolving mismatches between national accounts and microdata aggregates

We find significant gaps between national accounts and microdata (tax & survey data) aggregates. In particular, we find that net mixed income (B3n, S14 + S15) is substantially lower in national accounts than our microdata aggregate. On the other hand, distributed incomes of partnerships and (quasi-)corporations (D42, S14 & S15) are much higher in national accounts than in our comprehensive microdata base. Based on details how German national accounts are computed, which the German statistical office shared with us, we undertook two modifications to the national accounts aggregates.

Wages of low-income workers In German national accounts, low-income workers (*Minijobbers*) are assumed to be employed exclusively by the household sector (S14+S15), i.e., sole proprietors, self-employed and private households. Yet, the Federal Statistical Office's *Verdiensterhebung*, the Federal Employment Agency's *Betriebe nach Betriebsgrößenklassen* and our own calculations using SOEP data indicate that low-income workers are also employed by larger firms, i.e., partnerships and corporations. This implies that B3n is too small as too many Minijob wages are deducted. Instead, a fraction of Minijob wages should be deducted from the profits of the corporate sector. We calculate overall wage costs by multiplying the total number of low-income workers published by the Federal Employment Agency in their report *Beschäftigte nach ausgewählten Merkmalen* with the average wage published by the Federal Statistical Office in their *Verdiensterhebung*. We find that overall wage costs for low-income workers sum to approx. EUR 2.5 bn in 2017. We extrapolate the total number of low-income workers backwards from 2004 to 1992 using trends documented by the SOEP, since the Federal Employment Agency does not offer data for this period. We extrapolate the average wage of low-income workers backwards using the share of the average low-income wage observed in 2019, which is 75% of the low-wage threshold. The *Verdiensterhebung* indicates that about 64% of *Minijobbers* are employed in firms with 10 employees and above. Hence, we deduct 64% of the low-income wages, i.e., EUR 1.6 bn in 2017, from corporate income (B5n, S11 & S12) and add it back to B3n, S14 & S15.

CEO compensation In national accounts, CEO compensation of limited liability companies (*GmbHs*), which is not subject to social security contributions, is treated as distributed corporate income (D42). In contrast, these incomes are declared as income from employment in the personal income tax. This implies that corporate income in national accounts is higher than our corporate income aggregate in microdata. Hence, we want to reattribute these incomes from "distributed incomes from corporations" (D42, S14) to "employees compensation" (D1, S1) in the national accounts. While in other countries, e.g. in the Italian NA, distributed corporate profits (D42, S14) are further detailed in three sub-categories: 1. D421 dividends, 2. D422 income withdrawn from quasi-corporation, 3. D423 other distributed income, which comprises mainly CEO compensations, in the German NA, D42 (S14+S15) is not further decomposed. Thus, we first have to approximate the amount of CEO compensation comprised in D42 (14+S15).

We compute CEO compensation as follows. The ruling of the fiscal court of Baden-Württemberg in 2001 introduced a cap for CEO compensation by industry and number of employees allowing an annual growth of 3%. For example, a CEO of a firm with up to ten employees in manufacturing was allowed to earn a maximum of EUR 131,000 in 2001. We first collect the number of firms by industry and number of employees assessed by the Federal Statistical Office on the basis of the firm registry (*Unternehmensregister*[15]) in 2020, which records a total number of firms – approx. 3.4m. The VAT statistics (*Umsatzsteuerstatistik*) record the number of limited liability companies (*GmbHs*). Assuming that the distribution of limited liability companies over industries and size is similar to the overall distribution recorded by the firm registry, we calculate the number of limited liability companies by industry and size and multiply this number with the compensation maximum by industry and size introduced by the fiscal court. According to our computations, CEO compensation amounts to 105bn in 2013.

[15] Table 52111-0002, "Rechtliche Einheiten (Unternehmensregister-System): Deutschland, Jahre, Wirtschaftszweige (Abschnitte), Beschäftigtengrößenklassen" drawn from the commercial register.

B | Data Appendix: The Long Way to Gender Equality: Gender Pay Differences in Germany, 1871-2021*

Abstract

This Data Appendix supplements my paper "The Long Way to Gender Equality: Gender Pay Differences in Germany, 1871-2021". It provides comprehensive details on the data sources, and harmonization steps to construct the weighted mean of gross hourly earnings across the three broad occupational groups for which earnings data exists since the 1870s: 1) agricultural blue-collar workers, 2) blue-collar industrial workers and 3) salaried personnel (white-collar workers) in industry, commerce, banking and insurance.

B.1. Agricultural workers

Until the broader mechanization and specialization of agricultural work, the three main groups of agricultural worker before the 1930s were: 1) contracted day laborers (*kontraktliche Tagelöhner*, also called *Dienstleute, Instleute or Gärtner*), 2) free day laborers (*freie Tagelöhner* also called *Einlieger, Heuerlinge or Häusler* if they had their own house) and 3) servants (*Gesinde*, i.e. *Knechte* and *Mägde*) . Each group had a specific remuneration schemes based on their living conditions and work arrangements. Thus, it is good to know some key facts about these different groups. Contracted day laborers have to provide their own and their families labor power, usually the husband and a youth (*Scharwerker* or *Hofgänger*) and occasionally also their wife's labor, throughout the year to one employer and often earn substantial in-kind deputats, e.g. farmland for self-management, fodder, wood, grain and other food, for their family on top of the cash wage. This type of workers emerged from serfs (*Leibeigene*). In the Eastern provinces, where big agricultural businesses were abundant contracted day laborers were dominant. Free day laborers usually lived in villages or on the estate of a farmer for rent or in their own house, often with a little plot of farmland (*Häusler*). While in the north-east of the Reich hardly any owned their own plot of land, this was more common in the south and west of the German Reich (von der Goltz, 1872). This type of workers could work for different employers as they are not contractually bound and often only earned a small share or no in-kind payments. Free day laborers were distributed unequally across the national territory. This group hardly existed in the North-Western, but dominated the Western and Southern states like Hessen or Baden-Württemberg (Asmis, 1919; von der Goltz, 1872). Servants lived in their landlords estate and earned a substantial share, often more than the cash wage, in-kind (board and lodging, sometimes clothing). (von der Goltz, 1872). Due to the close living arrangement, this group usually had the longest work hours (Baldauf, 1932). In this study, I exclude contracted day laborers from the calculation

* This is the data appendix for chapter 3.

because work contracts were focused on the family, i.e. while the husband and father earned a substantial cash wage and receives deputats (food, farmland, animals) for the family, all other family members (wife and children) earn minimal cash earnings. In this setting, the calculation of a gender wage gap is not sensible as already Asmis (1919) pointed out.

Work hours as well as remuneration throughout the year were quite volatile, especially for free day laborers. Before 1919, no law or collective wage agreement agreement set specific work hours, nor remuneration. This gap was only closed with the Provisional Agricultural Labor Act (*Vorläufige Landarbeiterverordnung, LAO*) (Wunderlich, 1961). At the end of the 19th century, work hours were documented as 10 to 12 hours in summer and between dawn and dusk in winter (7 to 8 hours). earnings were substantially higher during harvest season than in winter - especially for free day laborers (Asmis 1919 foll. von der Goltz 1875 & Verein für Socialpolitik 1892). Women as servants or free-day laborers usually worked similar amounts of daily hours as men (Baldauf, 1932, p. 29). Women's time devoted to employed work was determined by their marital status. Unmarried women often worked full-time as servants or year-round free day laborers. Agricultural worker's families often had own farm land at their disposal. Women stayed at home many days to manage their domestic agriculture. However, this effect might have affected yearly earnings of women, but less so daily earnings which are used for free day laborers.

Table B.1.1.: Agricultural workers: Data sources & harmonization.

Year	Documentation
1873	**Source & Coverage**: von der Goltz (1875): based on 1352 questionnaires mainly answered by agricultural associations, in some cases also farmers or civil servants like mayors in agricultural communities.[1] **Geography**: Earnings information for about 80 districts (*Regierungsbezirke*) in all provinces of the Reich **Categories:** servants and free day laborers (contracted day laborers information not used) Free day laborers: restricted to year-round working (many married women only work sporadically throughout the year), *original* earnings concept: average daily summer and winter earnings (with and w/o in-kind payments) in *Silbergroschen*, *used* benchmark: daily wage of those w/o in-kind payments; von der Goltz (1875, p. 446) suggests that information for this group is more accurate, in-kind payments yearly in *Thaler* Servants: *original information*: yearly cash earnings and monetized in-kind earnings in *Thaler* of different types of servants (*Aufseher, Knechte/Mägde, Jungen*) on smaller and bigger estates

[1] In 1872, the congress of German farmers (*Congress deutscher Landwirthe*) agreed to conduct a survey to inquire about the economic situation of German agricultural workers. A commission under the lead of Freiherr von der Goltz (university professor in Königsberg) created suitable questionnaires and send 11,000 pieces to all provinces of the German Reich. The study published in 1875 lists detailed summer and winter daily earnings for free day laborers with and without food & board, yearly cash and in-kind remuneration of servants (*Gesinde*) for small and big agricultural businesses and an estimate of the yearly income of a family of contractual day laborers.

Table B.1.1.: Agricultural workers contd.

Year	Documentation
	Work time harmonization: Equal weighting of winter and summer earnings (ass. winter season from Oct. to March), harmonization to weekly earnings assuming 300 workdays and 50 work weeks per year (von der Goltz, 1875); daily work hours for free day laborers in summer (around 12-14 hours) and winter (about 8-9 hours) from von der Goltz (1875, Part B) **Currency:** 30 Silbergroschen = 1 Taler; Taler = 3 Reichsmark (Gesetz über die Münzverfassung in den preußischen Staaten 1821) **Harmonization** across districts and worker categories: 1) Harmonization across different servant types: simple arithmetic mean wage across represented categories (*Aufseher, Knechte/Mägde, Jungen*) 2) Harmonization across servants on smaller and bigger estates: weighting by the share of small (up to 50ha in Prussia, up to 20ha in the rest of the Reich) and bigger estates foll. Galloway (2007, 1882 census) for Prussia and Hohls and Kaelble (1989) based on 1895 census for all other provinces and districts 3) Geographic harmonization: regional earnings information weighted by size of agricultural workforce by district (*Regierungsbezirk*) using data by Hohls and Kaelble (1989) based on occupational censuses 1882 (Kaiserlich Statistisches Amt, 1884) 4) Harmonization across agricultural workers types: Weighting of servants' vs free day laborer's earnings: information on relative strength of groups from Baldauf (1932, 3ff.) originally taken from Kaiserlich Statistisches Amt (1884)
1913 (1892, 1873, 1849)	**Source**: Asmis (1919) **Geography**: Prussia, this is more restricted than in 1873 based on von der Goltz, 1875. Figure B.1.1 shows that estimated average earnings are very close. **Workers groups**: free day laborers, unmarried servants, (contractual day laborers were excluded for the analysis) **Original earnings concept**: minimum and maximum daily earnings of free day laborers which do not receive board (same as for 1873); min. and max. yearly cash and in-kind earnings of servants **Worktime**: harmonization to hourly earnings assuming 300 workdays, 12 hours per day for servants, 10 hours per day for free day laborers (Baldauf, 1932) **Currency**: Reichsmark, Silbergroschen, Pfennige, conversion as for 1873; original information for years 1907-1912 converted to 1913 RM using CPI by Sensch (2008) **Harmonization:** 1) Geographic harmonization: weighting by geographic location using no. of agricultural workers at province level from census 1907 (1895, 1882) 2) Harmonization across agricultural workers types: Weighting of servants' vs free day laborer's earnings: information on relative strength of groups from Baldauf (1932, 3ff.) for 1907 (1882, 1895)

Table B.1.1.: Agricultural workers contd.

Year	Documentation
1919*	for this year, no original gendered data exists but non-gendered hourly earnings, to construct the average gross earnings across agriculture, blue-collar work in industry and white-collar work in industry, commerce, transport, banking and insurance gendered earnings information by Baldauf (1932) for 1924/5 is used and deflated to the 1919 using the hourly wage level in agriculture provided in Vorstand des Deutschen Landarbeiter-Verbandes (1926a, p. 17), the following describes the harmonization of the non-gendered data of 1919 **Geography**: 15 Prussian provinces (covering about 80% of agriculture workers acc. to occupational census 1925 **Workers groups**: all agricultural workers excluding forestry workers, no sub-categories given **Original earnings concept**: effective hourly cash and in-kind earnings for 1919 and 1923, not gendered **Currency**: Mark **Harmonization:** 1) Geographic harmonization: weighting by no. of agriculture workers per province based on using occupational census 1925 2) Deflation: The 1925 average hourly gendered earnings (male earnings = 28.76 Pf) based on Baldauf (1932) (see below) are deflated to the 1919 hourly (non-gendered) earnings level (deflated male earnings 102.67 Pf, defl. female earnings 72.4 Pf) alternative deflation using CPI by Sensch (2008) yields similar results (1925 male earnings deflated to 1919: 99 Pf; female deflated earnings: 69.9 f; the overall gender pay ratio across all sectors is 48.4% in benchmark and 48.1% for alternative deflation procedure
1924-1930	**Source**: Baldauf (1932, p. 121) based on Vorstand des Deutschen Landarbeiter-Verbandes (1926b), Vorstand des Deutschen Landarbeiter-Verbandes (1929) **Geography**: 10 Prussian provinces & 6 free states (one district per province);[2] **Workers groups**: four categories: 1) free day laborers, 2) servants of 18 years 3) young laborers of 18 yrs (young *Freiarbeiter*innen* in south and west, *Hofgänger*innen* in east and north) (excluded) 4) contracted day laborers and their wifes (excluded) **Original earnings concept**: hourly cash and in-kind earnings for free day laborers (young laborers and contracted day laborers); monthly cash earnings for servants; in-kind benefits monetized using stock market prices by Vorstand des Deutschen Landarbeiter-Verbandes (1926b) **Worktime**: hourly (free day laborers) and monthly (servants) earnings converted to hourly ass. 26 work days per month and 12 hours per day (Baldauf, 1932) **Currency**: Mark and Pfennig

[2] Measured in different collective bargaining districts: one bargaining district exhibiting the average earnings level was chosen in each province. In this district, usually a community with higher earnings level was surveyed (Baldauf, 1932).

Table B.1.1.: Agricultural workers contd.

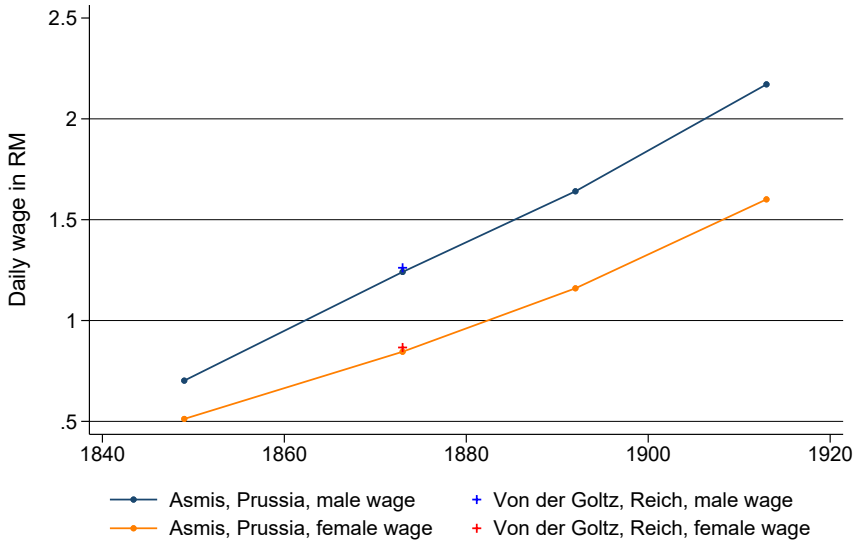
Year	Documentation
	Harmonization: 1) Imputation of deputats/in-kind pay (board and lodge) of servants: assumed as 140% of the cash wage following an example by Baldauf (1932, p. 78); 2) Geographic harmonization: weighting by no. of male and female agricultural workers in each province taken from Hohls and Kaelble (1989).
1937	**Source**: Statistisches Reichsamt (1941) based on individual survey (small sample: 90k), **Geography**: Reich as of 1937 **Workers groups**: 12 groups: 1) Married male agricultural workers with deputat (lodging, farming plot), 2) Married agricultural workers without deputat 3) unmarried male & female laborers, 4) female married laborers 5) female youth laborers 6) male agricultural foremen (Landwirtschaftsmeister), 7) male milking foremen, 8) male shepherd foremen, 9) male pig foremen, 10) male milkers, male cattle feeders, pig feeders, and shepherd assistants, 11) machine operator, tractor operator, driver, 12) craftsmen on farms **Original earnings concept**: effective yearly gross cash and in-kind earnings, in-kind payments provided in publication, assessed using consumer prices by Statistisches Reichsamt (1941); lodging costs for families were attributed to married men, to correct for this I redistribute 150RM of the married male deputat to married women assuming about 300 RM of yearly lodging costs for a family **Worktime**: harmonization to hourly earnings ass. 300 work days per year and 10 hours per day (sample population restricted to workers with 300 or more yearly working days) **Harmonization:** 1) Across workers types: weighting based on the relative sample strengths of different workers categories given in the publication; *benchmark*: including all workers categories esp. male high wage groups that women could not access:[3] resulting hourly male earnings 0.396 RM; female hourly earnings: 0.2677; gender earnings ratio within agriculture: 68% *robustness*: restricted sample similar to the historical categories (married and unmarried laborers and servants): resulting male earnings: 0.357 RM, gender earnings ratio within agriculture 75% For the overall gender pay ratio the restriction has a marginal effect: benchmark ratio: 57.7%, restricted agricultural workers sample: 58.0%
1953, 1957-63	**Source**: Statistisches Bundesamt (1958a, 6, tab. 2), Statistisches Bundesamt (1959), Statistisches Bundesamt (1960), Statistisches Bundesamt (1961a), Statistisches Bundesamt (1963a), Statistisches Bundesamt (1963b), Statistisches Bundesamt (1965); **Geography**: Federal Republic of Germany, w/o Bremen, Hamburg, Berlin (West), Saarland, farms with less than 20 ha are excluded

[3] I include those because it takes into account the mechanization of agriculture in the early 20th century. Usually only men had access to these better paid positions.

Table B.1.1.: Agricultural workers contd.

Year	Documentation
	Workers groups: agricultural workers in monthly or hourly pay schemes and divided into normal and specialized workers (*Facharbeiter*) on small (20-50ha) and bigger farms (50+ ha); original sample population restricted to men of 21+ years and women of 18+ years **Original earnings concept**: effective hourly and monthy gross cash earnings depending on remuneration scheme, in-kind payments given for 1953; *imputation of in-kind pay*: based on the 1953 ratio of in-kind to cash earnings separately for workers in hourly pay scheme (with marginal average in-kind payment) and workers in monthly pay scheme (usually receiving board and lodging) (as suggested in Statistisches Bundesamt (1958a, p. 5) **Worktime**: monthly work hours of male workers based on Statistisches Bundesamt, 1958a, 7, Tab 4, I assume women's work time to be 95% of men's work time **Harmonization:** 1) Harmonization across worker types: Weighting of workers in hourly and monthly remuneration scheme against each other using labor force strength given in publication (reasonable dur to the representative nature of the survey)
1964-1970	**Source**: Statistisches Bundesamt (1966a, 1967, 1968, 1969, 1970a, 1971, 1972) **Geography**: Federal Republic excluding Hamburg, Bremen, Saarland Berlin **Workers groups**: normal vs. medium-skilled (*angelernte*) workers in monthly remuneration scheme on small (20-50ha) and big (50+ ha) farms and normal, medium-skilled (*angelernte*) and specialized male workers (*Facharbeiter*) in hourly remuneration scheme on big farms (reform of collective wage groups and thus different workers groups from before) **Original earnings concept**: effective hourly and monthly gross earnings including in-kind payments of firms that signed a collective agreement; original sample population restricted to highest collective wage age level, i.e. males from 20 years, female from 18 years **Worktime**: earnings of workers in monthly remuneration scheme converted ass. 225-230 monthly work hours for men of which about 10 hours are overtime Statistisches Bundesamt (1966a, 1967, 1968, 1969, 1970a, 1971, 1972), I assume that women do not work overtime and thus work about 95% of men **Harmonization:** Weighting of workers in hourly and monthly pay scheme against each other using labor force strength in publication (representative nature of the survey)
1971-	After another reform of the salary and wage structure survey, since 1971 the official statistics only present earnings information in a non-gendered way or for male workers; however, agricultural workers made up 2% of male workers and 1.5% of female workers of the sample in 1970; the effect on the gender earnings ratio of the employed workforce is marginal

Figure B.1.1.: Robustness Check: Agricultural daily earnings of free day laborers and servants based on von der Goltz (1875) for the entire Reich vs. Asmis (1919) for Prussia.



B.2. Blue-collar industrial workers

Table B.2.2.: Blue-collar industrial workers time series documentation

Year	Documentation
1873	**Women's earnings** **Source**: Reichskanzler-Amt (1877): representative survey (226k female workers of 16+ years), survey answered by mainly by factory owners and supervisors and in some cases by the workers themselves **Geographical coverage**: German Reich w/o Alsace, Lorraine, Mecklenburg-Strelitz, Reuß jüngere Linie, Lübeck, and Bremen **Workers groups**: All female industrial workers within a district (*Regierungsbezirk*), 73 districts surveyed **Covered industries**: 16 industries: mining; metallurgy; brickworks, pottery and glass; ignition goods (Zündwaren); haberdashery, toys, steel springs and needles; yarns and twine; silk and velvet products, bleaching and dying and print of fabrics and yarn; cotton wool products; tulle, lace, embroidery and knit ware; paper and carton goods; straw goods; tobacco, cigar and cigarette production; chocolate and chicory goods; and beet sugar factories. From this list, it becomes clear that metal work, machinery and chemical industry do not seem well covered. However, the typically female-dominated industries are represented. **Original earnings concept**: average weekly earnings of female factory workers in summer and winter respectively; I take the simple mean of both data. **Work time**: benchmark for this year: weekly earnings; however, information on average weekly work hours are given in publication

Table B.2.2.: Blue-collar workers contd.

Year	Documentation
	Harmonization: 1) Geographical harmonization: weighting by relative workforce strength of female industrial workers in district (Regierungsbezirk) based on the census 1882 (benchmark) (Kaiserlich Statistisches Amt, 1884); Robustness: weighting by industrial census (1875, Gewerbezählung); 2) Benchmark: sample restriction to districts for which also male earnings data is available. This increases the average earnings by about 5%, probably due to better coverage of male earnings in high-wage, more urban locations. 3) Deflation from 1874 to 1873 using the index of living costs by Bry (1960, 325, Table A-1) **Men's earnings** **Source**: Kuczynski (1962, pp. 419–430) and Kuczynski (1963)[4] **Geographical coverage**: different locations of the German Reich, mainly urban areas **Workers groups**: 1) building trade workers' weekly earnings in different districts in Prussia, Saxony, Hamburg and Bavaria, 2) carpenters' daily earnings in Berlin and Nuremberg, Hamburg, Rostock, Quedlinburg, Kiel, Chemnitz; 3) unskilled construction workers' daily earnings in Bochum, Gelsenkirchen, Hamburg, Hoerde, Nuremberg, Rostock; 4) metal workers' daily earnings in Essen, Stettin, Nuremberg, their yearly earnings in Stettin, Bochum, Munich, Augsburg; their weekly earnings in Chemnitz, Hannover; 5) textile workers' yearly earnings in Hof, Allgaeu, Augsburg, their weekly earnings in Rhineland; 6) wood workers' daily earnings in Wuerttemberg, Hamburg, Nuremberg, Berlin, their weekly earnings for Chemnitz, Breslau, Stolp, Halle, Zeitz; 7) printers' weekly earnings in Berlin, Breslau, Frankfurt a. M., Hamburg, Hannover, Leipzig, Stuttgart; 8) chemical industry workers' yearly and weekly in Berlin, Stettin and Wuerttemberg; 9) transport (drivers') weekly earnings in Berlin; 10) Mining: shift earnings for Saar, Aachen, Wuerttemberg (Right Rhine), Alsace (Left Rhine), Mansfeld (copper); 11) Food, brewery and bakery workers' weekly earnings in Eschwege, their daily earnings in Stettin; 12) cement industry workers' daily earnings in the Rhineland and Stettin **Covered industries**: eight industries + construction + mining **Original earnings concept**: effective daily, weekly, yearly earnings

[4] Hohls (1991) affirms that "In fact, all [earnings] series for the first years of the Empire are based on Kuczynski's data", i.e. Desai (1968), Bry (1960) and Hoffmann (1965). Kuczynski assembled his data mainly based on labor unions' data on collectively agreed earnings, reports of chambers of commerce and firm data on actual earnings paid (Hohls, 1991).

Table B.2.2.: Blue-collar workers contd.

Year	Documentation
	Work time: conversion to weekly earnings assuming 300 workdays or 50 weeks per year, 6 work days per week **Harmonization**: 1) Average earnings within industry: Simple arithmetic mean across districts (*Regierungsbezirk*) within industry (e.g. across all construction workers, or printers); 2) Harmonization across industries: Weighting by strength of male workforce in each industry based on the census 1882 (benchmark) (Kaiserlich Statistisches Amt, 1884); robustness: based on industrial census (1875, Gewerbezählung); 3) In-/Deflation: Earnings of the years 1870-80 in-/deflated to 1873 using the index of living costs by Bry (1960, 325, Table A-1) **Alternative data:** yearly earnings based on Kuczynski (1963), Desai (1968) and Hoffmann (1965) provided in Hohls (1991, p. 88). All of these series show a much lower yearly earnings level as the original data by Kuczynski (1963, pp. 419–430) suggests. Thus, I use the original data mainly based on weekly effective earnings directly. See figure B.2.2 for comparison.
1914-1918	**Source**: Statistisches Reichsamt (1921, p. 18), Statistisches Reichsamt, Abteilung für Arbeiterstatistik und Reichsamt für Arbeitsvermittlung (1920, 64f.), 63-65, volontary survey conducted by statistical office **Geographical coverage**: German Reich (but small sample of only 350-375 firms) **Workers groups**: adult male and female workers in 12 industries **Covered industries**: 12 industries (stone and clay, metal, chemical, textile, leather, wood & carving, food, clothing, duplication, paper, machinery, electronics) **Original earnings concept**: effective daily earnings **Work time**: conversion to hourly earnings assuming 9.5 hours/day female workers (Reichsamt des Innern, 1905) and 10.5 hours for male workers (Meinert, 1958) **Harmonization**: 1) Harmonization across industries: weighting by the strength of female/male workforce in each industry based on 1907 occupational census; 2) Update of industrial composition of the workforce between 1914-1918 using the employment index (based on workload in days (*Arbeitertagewerke*) provided in Statistisches Reichsamt, Abteilung für Arbeiterstatistik und Reichsamt für Arbeitsvermittlung (1920, 63f.) (see figure B.2.4); *Robustness:* this benchmark is compared to average average earnings across all 12 industries computed by the Statistisches Reichsamt, Abteilung für Arbeiterstatistik und Reichsamt für Arbeitsvermittlung (1919, p. 621), see figure B.2.3 3) Deflation to 1913: I use the earnings of March 1914 without deflating earnings
1927-1930	**Source**: Statistisches Reichsamt (1930, pp. 289–296), Statistisches Reichsamt (1931b, pp. 273–4), Statistisches Reichsamt (1932, pp. 271–2), Statistisches Reichsamt (1929): official and representative wage survey based on the *Verordung zur Ausführung des Gesetzes betreffend Lohnstatistik vom 14. Juli 1927 (RGBl. I, 185)*, survey based on individual worker's information (*Individualerhebung*) **Geographical coverage**: territory of the Weimar Republic. ie. excl. Posen, West Prussia, Saarland, Alsace and Lorraine)

Table B.2.2.: Blue-collar workers contd.

Year	Documentation
	Workers groups: Skilled and unskilled workers of different occupations (e.g. mason, carpenter, cement worker, painter within construction) in 12 industries separately, sample population: workers of 16+ years, excluding home workers, part-time workers and apprentices (Statistisches Reichsamt, 1929) **Covered industries**: 12 industries (textile, clothing, chemical , metalworking (*metallverarbeitend*), steel (*Eisen- und Stahlerzeugung*), wood, lithography, printing, shoe, brewery, food, and paper) + construction (only male workers) + mining; imputation of clothing (= wage level of textile) and stone (= wage level of mining) **Original earnings concept**: hourly earnings (under time and piece rate scheme) including & excluding allowances for overtime and extra hours worked Used benchmark: hourly wage *excluding* overtime supplements **Work time**: original data already hourly **Harmonization**: 1) harmonization within industries: weighting of different occupations and by time vs piece rate pay schemes *within* industry and skill groups by strength of different groups in the same publication (reasonable due to representative survey) 2) harmonization within industry across skill groups: weighting of female/male skilled and unskilled workers' earnings according to female/male workforce by skill level and industry based on Willms-Herget (1985, p. 163) 3) harmonization across industries: weighting by female/male workforce strength in each industry based on by census 1925 using data by Willms (1982) 4) de-/inflation: to 1928 using CPI by Sensch (2008), the 1930 data points are still comparable to wage of previous years because nominal hourly earnings only declined slowly after the depression and were most effectively from 1931/2 onwards when collective wage rates were decreased, the most direct decreases of remuneration after 1929 are seen in weekly earnings due to lower work opportunities (Statistisches Reichsamt, 1938)
1935-1943	**Source**: Statistisches Reichsamt (1938, p. 160) 1937 gross earnings in RM & earnings development indeces by gender and skill level 1935-1943 based on Statistisches Reichsamt (1939, 1940, 1943 1943), survey based on wage sums for specific workers groups (men, women, un/skilled, industries) (*Lohnsummenverfahren*) **Geographical coverage**: German Reich (as of 1935) **Workers groups**: skilled and unskilled male workers and female workers *across* industries **Covered industries**: 16 industries (mining, iron and steel production, metalworking, chemical, construction, sawing industry, carpentry and furniture making, paper making, paper processing, printing, lithography, textile, clothing, shoe, confectionery & bakery, brewing) **Original earnings concept**: effective gross hourly (benchmark) & weekly effective earnings **Work time**: original data hourly

Table B.2.2.: Blue-collar workers contd.

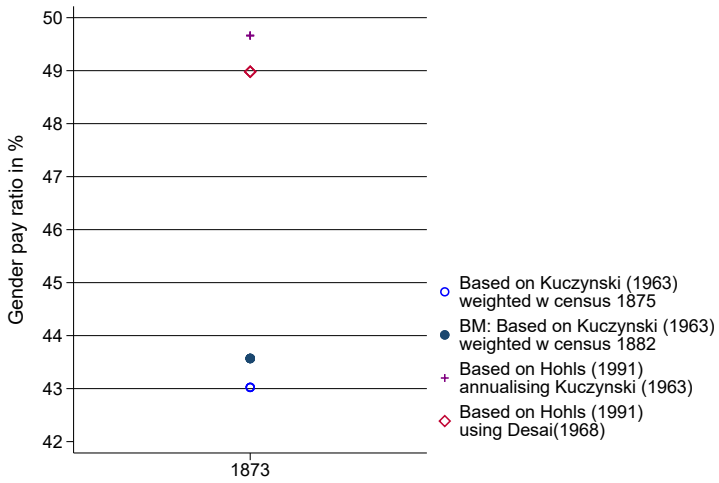
Year	Documentation
	Harmonization: 1) Harmonization across skill levels: Weightingbased on approximate share of female/male skilled and unskilled workers in industrial sector by Bry (1960) based on Statistisches Reichsamt (1931a, pp. 99, 101)
1957-1990	**Source**: Statistisches Bundesamt (1957-1990): yearly Structure of earnings survey (*Laufende Verdienststrukturerhebung in Industrie (und Handel)*) **Geographical coverage**: Federal Republic of Germany incl. Berlin (West), Saarland from 1958 **Workers groups**: cross-industry earnings given for full-time employed, excluding part-time employees and executive employees as well as master craftsmen and apprentices **Covered industries**: 24 industries - mining, energy industry, production goods industries (stone and earth, iron and steel industry, chemical industry, rubber and asbestos processing, sawmill and wood processing, paper production), capital goods industries (steel and mechanical engineering, shipbuilding, road construction, electrical engineering, optics and precision mechanics), consumer goods industries (processing of artificial materials, glass industry, wood processing, paper processing, printing, leather production, shoe industry, textiles, clothing, production of musical instruments and toys), food and beverage industries (brewery, food, tobacco); part of the construction industry (building, civil and structural engineering). **Original earnings concept**: effective gross hourly earnings, ie. all amounts received by employees on a regular basis, incl. other allowances and supplements & turnover commission if applicable, but excluding one-off payments, such as 13th month's salary, expenses, severance pay **Benchmark vs alternative data sources**: *Structure of earnings survey (benchmark):* based on wage and salary sums of predefined workers groups (Lohnsummenverfahren); broad industry coverage and yearly frequency but reduced level of detail, methodological information: Statistisches Bundesamt (1957).[5]

[5] The survey was already picked up again in 1946 based on the methodology applied in the 1930s. Until 1956 only a small group of industries were surveyed. With the reform of 1956/7, white-collar workers were included and surveys industries for blue-collar workers was substantially increased.

Table B.2.2.: Blue-collar workers contd.

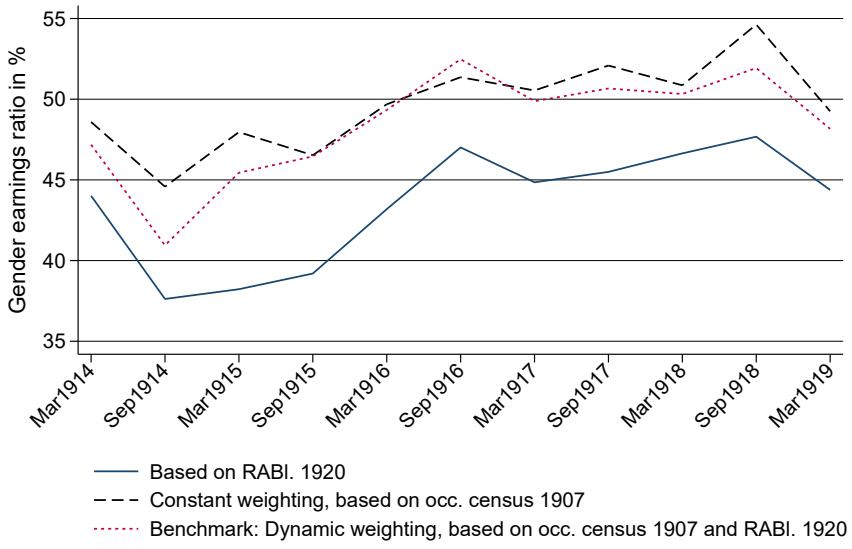
Year	Documentation
	Alternative data sources: Salary and wage structure survey: 1951, 1957, 1962, 1966, 1972, 1978, 1990 (*Gehalts- und Lohnstrukturerhebung*) • Individualized earnings surveys, thus higher level of detail but lower frequency • Includes part-time employees since 1966 • In several years higher coverage of sectors beyond manufacturing • Drawback: changing composition of industries: 1951 - industry, commerce, banking/insurance, transport and selected services; 1957 - industry and construction ; 1966 - industry, construction, banking and insurance, selected services; 1972 - industry, construction, banking and insurance, and selected services, women only in industry, part-time workers only in industry; 1978 - industry and construction ; 1990 - all sectors and full- and part-time SIAB: microdata since 1975, only daily earnings, no work hours variable (dummy for full- vs part-time)
1984-2020	**Source**: Socio-economic Panel (SOEP) (Wagner et al., 2007): microdata **Geographical coverage**: Federal Republic (West) 1984-; reunified Germany: data is available since 1990, due to the inclusion of East German households stretched over the first few years, I show estimates for reunified Germany from 1994 on **Workers groups**: blue- and white-collar workers in all sectors, agricultural workers are not sufficiently surveyed **Original wage concept**: individual monthly gross earnings of the last month (*pglabgro*) **Harmonization**: 1) Sample restriction: to full-time employees with 30+ contractual work hours per week and non-negative earnings; 2) Conversion to gross hourly earnings: monthly earnings divided by 4.3 work weeks and individual actual (contractual) hours worked; 3) Weighting of individual gross hourly earnings by cross-sectional weights ($w11101$ in *pequiv* file) 3) Time series smoothed using a 3-year moving average **Benchmark vs alternative data sources**: SOEP has a higher coverage of modern service occupations compared to Structure of Earnings Survey **Alternatives**: Structure of Earnings Survey (Statistisches Bundesamt, 1991) • Coverage: until 2004: FRG (Statistisches Bundesamt, 1991-2005) • 1996-2004 reunified Germany (Statistisches Bundesamt, 1991-2005) • 2007- Statistisches Bundesamt (2007-2015, Table 1.2. 1 Durchschnittliche Verdienste und Arbeitszeiten nach Wirtschaftsabschnitten) SIAB: microdata since 1975, only daily earnings, no work hours variable (dummy for full- vs part-time)

Figure B.2.2.: Robustness Check: Blue-collar workers 1873.



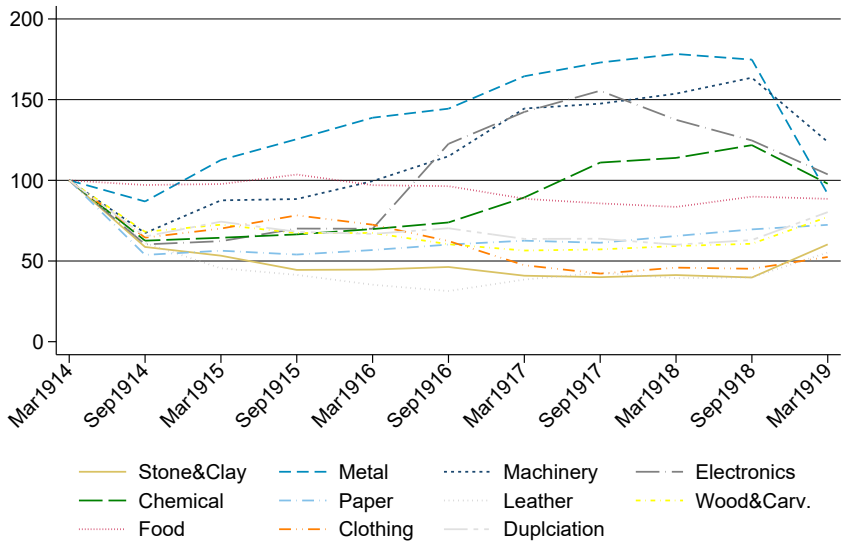
Note: For this year, different data sources and weights were tried. However, as Hohls (1991) pointed out "all [earnings] series for the first years of the Empire are based on Kuczynski's data". My benchmark estimates are based on mainly weekly and daily (sometimes yearly) earnings of different industry and occupation groups gathered by Kuczynski (1962, pp. 419–430). Weighting by the 1875 census or 1882 census shows marginal differences. However, the annualized earnings of male workers estimated by Hohls (1991) based on Kuczynski (1962) and Desai (1968) seem quite low in comparison to the weekly earnings. Thus, the higher earnings ratio.

Figure B.2.3.: Robustness Check: blue-collar workers' earnings in World War I.



Note: For 1914-1919 blue-collar workers daily gross earnings were given for 12 industries. This graph contrasts the benchmark (red), weighting the industries against each other using the 1907 census updated using the employment index 1914-1919, with the weighting by the census 1907 (black) and the gender ratio based on average male/female earnings across industries given in the original publication (blue).

Figure B.2.4.: Robustness Check: Blue-collar workers employment shifts in World War I.



Note: This graphs shows the employment shifts across the 12 surveyed industries between 1914 and 1919 based on the employment index used for harmonization across industries.

B.3. White-collar workers/salaried personnel

Table B.3.3.: White-collar workers /Salaried personnel

Year	Documentation
1873*	**Source**: Hohls (1991, 60, Tab. 2.4) for 1890 Due to a lack of reliable data on white-collar workers earnings in the 1873, I impute these using information of Hohls, 1991 for 1890 see below). 1) *benchmark:* imputing white-collar workers earnings as twice the blue-collar wage. This ratio was taken from the comparison of annual earnings of these groups in the 1890s from Hohls (1991, 60, Tab. 2.4). This ratio seems reasonable for the late 19th century because the marginal number of white-collar workers were more likely to be secretaries of firm owners conducting a wide range of skilled tasks and with an overview of the business. Only around the turn of the century, due to the automation and mechanization of many tasks due to the introduction of the typewriter etc., a wider class of white-collar workers with more specialized tasks emerged. This development let white-collar workers earnings converge to blue-collar workers earnings Goldin (1990, p. 106) 2) *Robustness:* Deflating white-collar workers earnings from 1890 to 1873 using the CPI of Sensch, 2004 and the cost of living index by Bry (1960), 3) *Robustness:* omitting white-collar workers which were then a rather marginal group. **Figure B.3.5** contrasts the different options and shows that they only produce marginal differences in the gender earnings ratio across the entire employed workforce.
1890-1913	**Source**: Hohls (1991, 60, Tab. 2.4) correcting the time series by Pierenkemper (1987) based on social insurance data (*Reichsversicherungsanstalt f. Angestellte*) **Geographical coverage**: German Reich **Employees groups**: average earnings across full sample population; sample population : white-collar employees in manufacturing, trade, transport and private services + part of female staff in public services (teachers + nursing staff), excl. domestic servants, incl. civil servants of state railroads, employees of private railroad companies and salaried employees in Prussian mining industry[6] **Original earnings concept**: yearly effective earnings **Work time**: conversion to hourly earnings (1913) assuming 310 workdays per year, 9 work hours per day for both genders; Kaiserliches Statistisches Amt, Abteilung Arbeiterstatistik (1912) shows that employees in the legal sector worked on average 8 hours per day; however, sales personnel was known to work substantially more, thus I added one hour, no gender differences in work time visible in data; conversion to weekly earnings (1873) via division by 50 work weeks

[6] Hohls (1991, pp. 59–60) revises data by Pierenkemper (1987) for white-collar employees in manufacturing, trade, transport and private services, part of female staff in public services (teachers + nursing staff), excl. domestic servants (Hausangestellte, who after 1913 are part of the white-collar employee insurance scheme (Angestelltenversicherung) and adds information on civil servants of state railroads, employees of private railroad companies and salaried employees in Prussian mining industry; as a benchmark, I use Hohls, 1991 because his estimate is nearer to actually measured salaries by the insurance scheme (*Reichsversicherungsanstalt für Angestellte*) despite the broader sample

Table B.3.3.: White-collar workers /Salaried personnel contd.

Year	Documentation
	Harmonization: no further in-/deflation or weighting needed
1920	**Source**: Statistisches Reichsamt (1921, pp. 40, 42, 46): obligatory survey (employers could be fined for not responding), 11,600 surveyed firms of all sizes, 1.6 million blue-collar workers and 226.5k white-collar/salaried personnel, **Geographical coverage**: German Reich, incl. occupied territories **Employees groups**: Earnings across sample population (benchmark); sample population: only full-time workers that worked the full survey month of February, detailed earnings information for commercial, office and technical employees respectively available in publication **Original earnings concept**: effective monthly salary **Work time**: conversion to hourly assuming 26 workdays per month and 46 weekly work hours (Gewerkschaftsbund der Angestellten (GdA), 1931; Statistisches Reichsamt, 1935) **Harmonization**: 1) deflation from 1920 to 1919 using CPI by Sensch (2008) - only necessary for the construction of the overall gender earnings ratio
1929-30	**Source**: three large white-collar organizations: 1) Gewerkschaftsbund der Angestellten (GdA) (1931): *Gewerkschaftsbund der Anstellten*, survey year: 1929, 150k individual survey answers ; largest white-collar organization accepting men and women ; 2) Suhr (1930): survey year: early 1930, *Zentralverband der Angestellten*, one of the biggest socialist white-collar organizations, only female earnings; 3) Glaß and Kische (1930): survey year: early 1929, 24.6k individual answers, *Verband der weiblichen Handels- und Büroangestellten*, one of the biggest christian-national white-collar associations, only female earnings (information on different association based on Yu (1994, 211f.)) **Geographical coverage**: territory of the Weimar Republic, ie. excl. Posen, West Prussia, Saarland, Alsace and Lorraine) **Employees groups**: Suhr (1930): only female, mainly commercial personnel; Glaß and Kische (1930): only female, mainly commercial and office personnel, Gewerkschaftsbund der Angestellten (GdA), 1931: male and female commercial, technical, office and foremen (Yu, 1994, p. 212) **Original earnings concept**: effective average monthly salaries (incl. family supplements usually received by married men) **Work time**: from monthly to hourly earnings assuming 4.3 weeks per month and 46 weekly hours as stated in Gewerkschaftsbund der Angestellten (GdA) (1931, p. 202) (no significant gender differences in hours) **Harmonization**: 1) harmonization between groups surveyed in the three studies: female average earnings as weighted average of Gewerkschaftsbund der Angestellten (GdA) (1931) (50%), Suhr (1930) (25%), Glaß and Kische, 1930 (25%);[7] male wage based on Gewerkschaftsbund der Angestellten (GdA) (1931) 2) in-/deflation: since all surveys were conducted between early 1929 and early 1930, I do not in-/deflate between data points

[7] I give higher weight to the first source as this was one of the biggest organisations and because the male wage is defined by the GDA's scope as the other sources do not include male wage information.

Table B.3.3.: White-collar workers /Salaried personnel contd.

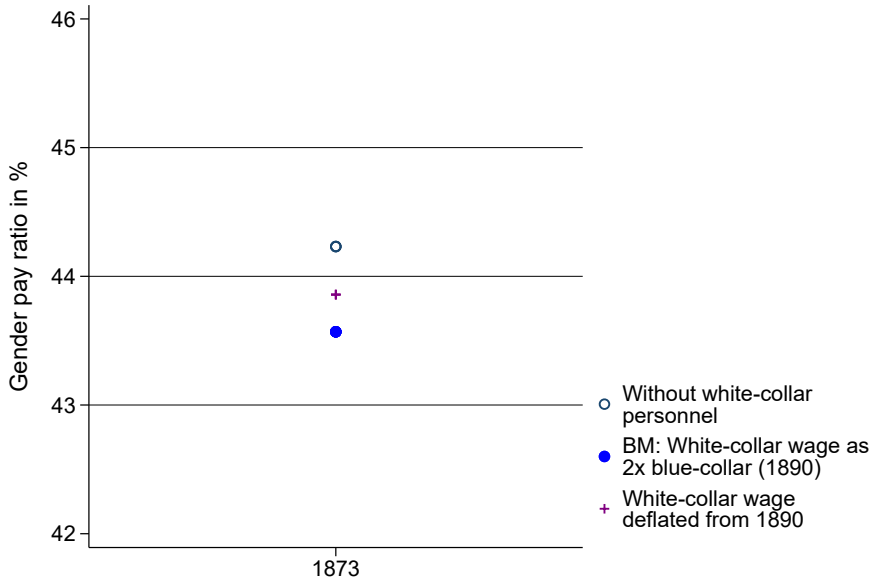
Year	Documentation
	3) preparation for 1928 gender earnings ratio across entire employed workforce: deflation from 1929 to 1928 using CPI by Sensch (2008)
1934-1936	**Source**: Statistisches Reichsamt (1937, 1939a): payroll tax data **Geographical coverage**: Reich, incl. Saarland from 1936 onwards **Employees groups**: all employees registered in salaried personnel/white collar insurance scheme (*Angestellte*), above exemption threshold[8] **Original earnings concept**: annual income from employment (incl. supplements for overtime, hazard pay, allowances and compensation) after deduction of income-related expenses (*Pauschbetrag*) **Work time**: Conversion to hourly earnings assuming 4.3 weeks per month/ 51.6 weeks per year, 46 weekly work hours based on Statistisches Reichsamt, 1935 (w/o substantial gender differences) **Harmonization**: 1) Conversion for workforce gender earnings ratio: Inflation from 1936 to 1937 using CPI by Sensch, 2008 **Alternative data sources**: Statistisches Reichsamt (1935, 136ff.): salaried personnel in banking 1934, representative survey for the Reich, all commercial personnel in banking up to salary of 600RM per month, individual survey (Individualerhebung) in 935 locations, 2,195 bank establishments, 49k employees, effective monthly salaries (incl. and excl. supplements (given only for employees under collective agreement)) – not used because banking personnel was a high-income group among salaried personnel and would overestimate the average salary of this group and artificially inflate gender pay ratio
1957-1990	**Source**: Statistisches Bundesamt (1957-1990): yearly Structure of earnings survey (*Laufende Verdienststrukturerhebung in Industrie und Handel*) **Geographical coverage**: Federal Republic of Germany incl. Berlin (West), Saarland from 1958 **Workers groups**: cross-industry-sector earnings given for full-time employed, excluding part-time employees and executive employees **Covered industries**: 24 industries (see tab. B.2.2 for details) + commerce, banking and insurance **Original earnings concept**: effective monthly salaries; salaries comprise all amounts received by employees on a regular basis, incl. other allowances and supplements & turnover commission if applicable, but excluding one-off payments, such as 13th month's salary, expenses, severance pay **Work time**: conversion to hourly earnings assuming 4.3 work weeks per month and btw. 46.6 weekly hours (1957) and 39.4 hours (1983) for men; full-time female employees assumed to work 97% of men (Statistisches Bundesamt, 1966-2010) **Benchmark vs alternative data sources**: see tab. B.2.2

[8] The income exemption limit was lowered substantially in 1936, so that tax data got more representative esp. for women being overrepresented in low-income brackets. Tax data for salaried employees comprise about one-third women, incl. married women under special tax regime for co-earning wives.

Table B.3.3.: White-collar workers /Salaried personnel contd.

Year	Documentation
	Alternative data sources: Salary and wage structure survey: 1951, 1957, 1962, 1966, 1972, 1978, 1990 (*Gehalts- und Lohnstrukturerhebung*) Statistisches Bundesamt (1954, 1961b, 1966b, 1970b, 1981, 1993) SIAB: microdata since 1975, only daily earnings, no work hours variable (dummy for full- vs part-time)
1984-2020	see table B.2.2

Figure B.3.5.: Robustness check: Gender earnings ratio across three occupational groups (agriculture, blue-collar industry and white-collar) under different imputation procedure for white-collar workers earnings, 1873.



Note: Due to a lack of adequate data, I try different imputation strategies for average earnings of white-collar workers. The graphs shows that the overall gender earnings ratio across all three occupational groups is only marginally affected by different imputation strategies.

B.4. Occupational census data

Table B.4.4.: Sources for employed population structure used to construct weights $\phi_{o,g}$ used in social table method.

Year	Documentation
1882, 1895, 1907	**Source**: Occupational censuses data drawn from Müller et al. (1983, Tab A2, p. 176/77). These authors have already restricted the data to employed persons outside the home. **Geographical coverage**: German Reich **Labor force concept**: Labor force (*Erwerbspersonen*) comprising employed and unemployed
1925, 1933, 1939	**Source**: Länderrat des Amerikanischen Besatzungsgebiets (1949, p. 32f, tab. 13. Die Erwerbspersonen nach Wirtschaftsabteilungen und Stellung im Beruf 1939, 1933 und 1925.) **Geographical coverage**: 1925, 1933: Weimar Republic, excl. Posen, West Prussia, Saarland, and Alsace and Lorraine; 1939: German Reich incl. Saarland and territories of Austria and Sudetenland **Labor force concept**: Labor force (*Erwerbspersonen*) comprising employed and unemployed
1950	**Source**: Occupational census 1950 drawn from Statistisches Bundesamt (1958b, 117ff.) **Labor force concept**: Labor force (*Erwerbspersonen*) comprising employed and unemployed **Geographical coverage**: Federal Republic of Germany without Saarland and Berlin
1961	**Source**: Occupational census drawn from Statistisches Bundesamt (1964b, 142, Tab 6) **Labor force concept**: Gainfully employed (*Erwerbstätige*) **Geographical coverage**: Federal Republic of Germany without Berlin
1963-1975	**Source**: Statistisches Bundesamt (1964a) to Statistisches Bundesamt (1976): yearly information **Labor force concept**: Gainfully employed (*Erwerbstätige*) **Geographical coverage**: Federal Republic of Germany (West)
1976 - 1983	**Source**: Statistisches Bundesamt (1977), Statistisches Bundesamt (1984): yearly information, data available for the FRG until 1997 (Statistisches Bundesamt, 1997) **Labor force concept**: Gainfully employed (*Erwerbstätige*) **Geographical coverage**: Federal Republic of Germany
1984 - 2020	**Source**: SOEP, version 37, variable: w11101 (individuals cross-sectional weight) of pequiv file (Wagner et al., 2007) **Labor force concept**: Gainfully employed (Erwerbstätige with non-zero wage or salary **Geographical coverage**: Federal Republic of Germany and reunified Germany from 1990

Table B.4.5.: Employed population structure and weighting contd.

Year	Documentation
	An equivalent definition of the three historical occupational groups is not possible in SOEP data. Individual weights (irrespective of occ. group) are used to construct average male/female earnings for the entire workforce

B.5. Earnings data U.S. and Sweden

Table B.5.6.: International Comparison, time series documentation for the US and Sweden

Country	Year	Documentation
U.S.	1890, 1930, 1970	**Source**: Goldin (1990, p.68ff. and tab. 3.2. on p. 64) **Workers groups/sample population**: full-time employees in six occupational groups (agriculture, manufacturing, clerical, sales, professional, service) **Original earnings concept:** Annual earnings **Harmonization**: 1) Across occupational groups: gender- & occupational-group-specific earnings weighted by employment shares in each gender-occupation group (similar to method laid out in section 2 of the main paper)
U.S.	1955-2015	Blau and Winkler (2018, Tab 7-5, p. 172) **Earnings concept:** Weekly and annual median earnings **Workers groups**: full-time employees, earnings across all occupational, sector and industry groups
Sweden	1920, 1940, 1960, 1980, 1995	**Source**: Svensson (2003) based on Swedish Social Welfare Board & Swedish Board of Commerce for blue- and white-collar workers in industry, service sector from a variety of sources or imputation from similar groups of labor **Earnings concept:** Gross hourly earnings **Workers groups/sample population**: employees of seven sectors (agriculture, industry and crafts, commerce and hotel industry, transport, banking and insurance, public sector and liberal professions, domestic services), blue-collar and white-collar workers, unclear if part-time workers included **Harmonization**: 1) Across occupational groups: gender- & occupational-group-specific earnings weighted by employment shares in each gender-occupation group (similar to method laid out in section 2 of the main paper)

Bibliography

Aaberge, R. and A. Atkinson (2010). "Top Incomes in Norway." In: *Top Incomes in Global Perspective*. Ed. by A. B. Atkinson and T. Piketty. Oxford University Press, pp. 448–481.

Abadie, A., A. Diamond, and J. Hainmueller (2015). "Comparative Politics and the Synthetic Control Method." In: *American Journal of Political Science* 59.2, pp. 495–510.

Abadie, A. and J. Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country." In: *American Economic Review* 93.1, pp. 113–132.

Albers, T., C. Bartels, and M. Schularick (2022). *Wealth and its Distribution in Germany, 1895-2018*. CEPR Discussion Paper 17269.

Alesina, A. and R. Perotti (1996). "Income Distribution, Political Instability, and Investment." In: *European Economic Review* 40, pp. 1203–1228.

Alesina, A. and N. Fuchs-Schündeln (2007). "Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences." In: *American Economic Review* 97.4, pp. 1507–1528.

Almeida, R. (2007). "The Labor Market Effects of Foreign Owned Firms." In: *Journal of International Economics* 72.1, pp. 75–96.

Alstadsæter, A., M. Jacob, W. Kopczuk, and K. Telle (2023). "Accounting for Business Income in Measuring Top Income Shares: Integrated Accrual Approach Using Individual and Firm Data from Norway." Unpublished manuscript.

Altzinger, W., C. Berka, S. Humer, and M. Moser (2012). "Die langfristige Entwicklung der Einkommenskonzentration in Österreich, 1957-2009. Teil II: Methodik und Ergebnisse." In: *Wirtschaft und Gesellschaft* 38.1, pp. 77–101.

Asmis, W. (1919). *Zur Entwicklung der Landarbeiterlöhne in Preußen*. Verlag für Landwirtschaft, Gartenbau und Forstwirtschaft, Berlin.

Atkinson, A. B. (2015). *Inequality. What Can Be Done?* Harvard University Press, Cambridge, US & London.

Bach, S. (2013). *Has German Business Income Taxation Raised too Little Revenue Over the Last Decades?* DIW Discussion Paper 1303. DIW Berlin.

Bach, S., C. Bartels, and T. Neef (2022). "Distributional National Accounts: A Macro-Micro Approach to Inequality in Germany." In: *Measuring Distribution and Mobility of Income and Wealth. Vol. 80*. Ed. by R. Chetty, J. N. Friedman, J. Gornick, B. Johnson, and A. Kennickell. University of Chicago Press, pp. 625–640.

– (2023). *Distributional National Accounts for Germany, 1992-2016*. SSRN Working Paper. http://dx.doi.org/10.2139/ssrn.4487665.

Bach, S., G. Corneo, and V. Steiner (2009). "From Bottom to Top: The Entire Income Distribution in Germany, 1992-2003." In: *Review of Income and Wealth* 55.2, pp. 303–330.

Bajohr, S. (1979). *Die Hälfte der Fabrik. Geschichte der Frauenarbeit in Deutschland 1914-1945*. Verlag Arbeiterbewegung und Gesellschaftswissenschaft, Marburg.

Baldauf, E. (1932). "Die Frauenarbeit in der Landwirtschaft." PhD thesis. Christian-Albrechts-Universität zu Kiel.

Banerjee, A. V. and A. F. Newman (1993). "Occupational Choice and the Process of Development." In: *Journal of Political Economy* 101.2, pp. 274–298.

Banken, R. (2018). *Hitlers Steuerstaat*. De Gruyter, Oldenbourg.

Baraké, M., P.-E. Chouc, T. Neef, and G. Zucman (2022). "Revenue Effects of the Global Minimum Tax Under Pillar Two." In: *Intertax* 50.10, pp. 689–710.

Baraké, M., T. Neef, P.-E. Chouc, and G. Zucman (2021a). *Collecting the Tax Deficit of Multinational Companies: Simulations for the European Union*. EU Tax Observatory Report No. 1.

– (2021b). *Minimizing the Minimum Tax? The Critical Effect of Substance Carve-Outs*. EU Tax Observatory Note No. 1.

– (2021c). *Revenue Effects of the Global Minimum Tax: Country-by-Country Estimates*. EU Tax Observatory Note No. 2.

Barron, J. M., D. A. Black, and M. A. Loewenstein (1993). "Gender Differences in Training, Capital, and Wages." In: *The Journal of Human Resources* 28.2, pp. 343–364.

Bartels, C. (2019). "Top Incomes in Germany, 1871–2014." In: *The Journal of Economic History* 79.3, pp. 669–707.

Bartels, C. and M. Metzing (2019). "An Integrated Approach for a Top-corrected Income Distribution." In: *Journal of Economic Inequality* 17.2, pp. 125–143.

Bartels, C. and D. Waldenström (2020). "Inequality and Top Incomes." In: *Handbook of Labor, Human Resources and Population Economics*. Ed. by K. F. Zimmermann. Springer International Publishing, Cham, pp. 1–28.

Bauernschuster, S. and H. Rainer (2012). "Political Regimes and the Family: How Sex-role Attitudes Continue to Differ in Reunified Germany." In: *Journal of Population Economics* 25.1, pp. 5–27.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Columbia University Press, New York.

Becker, J. and J. Englisch (2021). *GloBE Minimum Taxation: Calculating the Local ETR with Carve-outs*. Kluwer International Tax Blog. https://kluwertaxblog.com/2021/12/17/globe-minimum-taxation-calculating-the-local-etr-with-carve-outs/.

Beer, S., R. De Mooij, and L. Liu (2020). "International Corporate Tax Avoidance: A Review of the Channels, Magnitudes, and Blind Spots." In: *Journal of Economic Surveys* 34.5, pp. 660–688.

Bénabou, R. (1996). "Inequality and Growth." In: *NBER Macroeconomics Annual* 11, pp. 11–74.

Benhabib, J. and A. Rustichini (1996). "Social Conflict and Growth." In: *Journal of Economic Growth* 1, pp. 125–142.

Biewen, M. (2001). "Measuring the Effects of Socio-Economic Variables on the Income Distribution: An Application to the East German Transition Process." In: *The Review of Economics and Statistics* 83.1, pp. 185–190.

Blanchet, T., L. Chancel, I. Flores, and M. Morgan (2021). *Distributional National Accounts Guidelines. Methods and Concepts Used in the World Inequality Database*. Updated and extended version of the DINA Guidelines first published as WID Working Paper 2016/2.

Blanchet, T., L. Chancel, and A. Gethin (2019). *How Unequal Is Europe? Evidence from Distributional National Accounts, 1980-2017*. WID.world Working Paper 2019/06.

– (Oct. 2022a). "Why Is Europe More Equal than the United States?" In: *American Economic Journal: Applied Economics* 14.4, pp. 480–518.

Blanchet, T., E. Saez, and G. Zucman (2022b). *Real-Time Inequality*. NBER Working Paper 30229.

Blau, F. D. and L. M. Kahn (1996). "Wage Structure and Gender Earnings Differentials: An International Comparison." In: *Economica* 63.250, S29–S62.

– (2007). "Changes in the Labor Supply Behavior of Married Women: 1980–2000." In: *Journal of Labor Economics* 25.3, pp. 393–438.

– (2017). "The Gender Wage Gap: Extent, Trends and Explanations." In: *Journal of Economic Literature* 55.3, pp. 789–865.

Blau, F. D. and A. E. Winkler (2018). *The Economics of Women, Men, and Work (Eighth edition)*. Oxford University Press, New York & Oxford.

Blouin, J. and L. Robinson (2020). *Double Counting Accounting: How Much Profit of Multinational Enterprises is Really in Tax Havens?* SSRN Working Paper. http://dx.doi.org/10.2139/ssrn.3491451.

Boter, C. (2017). "Dutch Divergence? Women's Work, Structural Change, and Household Living standards in the Netherlands, 1830-1914." PhD thesis. Wageningen University.

Bourguignon, F. (1981). "Pareto Superiority of Unegalitarian Equilibria in Stiglitz' Model of Wealth Distribution with Convex Saving Function." In: *Econometrica* 49.6, pp. 1469–1475.

Bozio, A., B. Garbinti, J. Goupille-Lebret, M. M Guillot, and T. Piketty (2023). "Predistribution vs. Redistribution: Evidence from France and the U.S." In: *American Economic Journal: Applied Economics* forthcoming.

Bratta, B., V. Santomartino, and P. Acciari (2021). *Assessing Profit Shifting Using Country-by-Country Reports: A Non-linear Response to Tax Rate Differentials*. DF Working papers. 11/2021.

Braunwarth, H. (1955). *Die Spanne zwischen Männer- und Frauenlöhnen: tatsächliche Entwicklung und kritische Erörterung ihrer Berechtigung*. Bund-Verlag, Köln-Deutz.

Breza, E., S. Kaur, and Y. Shamdasani (2017). "The Morale Effects of Pay Inequality." In: *The Quarterly Journal of Economics* 133.2, pp. 611–663.

Brixy, U. and R. Grotz (2004). "Differences of the Economic Performance of Newly Founded Firms in West- and East Germany." In: *Advances in Interdisciplinary European Entrepreneurship Research*. Ed. by M. Dowling, J. Schmude, and D. zu Knyphausen-Aufsess. LIT Verlag, Münster, pp. 143–152.

Brosig-Koch, J., C. Helbach, A. Ockenfels, and J. Weimann (2011). "Still Different After All these Years: Solidarity Behavior in East and West Germany." In: *Journal of Public Economics* 95.11-12, pp. 1373–1376.

Brown, J. D., J. S. Earle, and Á. Telegdy (2006). "The Productivity Effects of Privatization: Longitudinal Estimates from Hungary, Romania, Russia, and Ukraine." In: *Journal of Political Economy* 114.1, pp. 61–99.

Bruil, A. et al. (2022). *Inequality and Redistribution in the Netherlands*. Netherlands Bureau for Economic Policy Analysis CPB Discussion Paper March 2022.

Brüll, E. and C. Gathmann (2020). *Evolution of the East German Wage Structure*. ZEW Discussion Paper No. 20-081.

Bry, G. (1960). *Wages in Germany 1871-1945*. Princeton University Press, Princeton.

Bucher-Koenen, T. and B. Lamla-Dietrich (2018). "The Long Shadow of Socialism: Puzzling Evidence on East-West German Differences in Financial Literacy." In: *Economic Notes: Review of Banking, Finance and Monetary Economics* 47.2-3, pp. 413–438.

Bundesagentur für Arbeit (2021). *Beschäftige nach ausgewählten Merkmalen (Zeitreihe Quartalszahlen)*. Table 8. Geringfügig entlohnte Beschäftigte nach ausgewählten Merkmalen. https://statistik.arbeitsagentur.de/SiteGlobals/Forms/Suche/Einzelheftsuche_Formular.html?topic_f=beschaeftigung-sozbe-zr-ausgewmerkmale-altersgr.

– (2023). *Arbeitsmarkt in Zahlen*. Arbeitslosigkeit im Zeitverlauf: Entwicklung der Arbeitslosenquote. https://statistik.arbeitsagentur.de/Statistikdaten/Detail/Aktuell/iiia4/alo-zeitreihe-dwo/alo-zeitreihe-dwo-b-0-xlsx.xlsx.

Bundesministerium für Finanzen, ed. (2022). *Datensammlung zur Steuerpolitik 2022*. https://www.bundesfinanzministerium.de/Datenportal/Daten/offene-daten/steuern-zoelle/Datensammlung-zur-Steuerpolitik-2010-2022/datensaetze/2022.pdf?__blob=publicationFile&v=2.

Burda, M. C. and J. Hunt (2001). "From Reunification to Economic Integration: Productivity and the Labor Market in Eastern Germany." In: *Brookings Papers on Economic Activity* 2001.2, pp. 1–92.

Burda, M. C. and B. Severgnini (2018). "Total Factor Productivity Convergence in German states since Reunification: Evidence and Explanations." In: *Journal of Comparative Economics* 46.1, pp. 192–211.

Burnette, J. (2008). *Gender, Work and Wages in Industrial Revolution Britain*. Cambridge University Press, Cambridge.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." In: *American Economic Review* 102.6, pp. 2981–3003.

Chancel, L., D. Cogneau, A. Gethin, A. Myczkowski, and A.-S. Robilliard (2023). "Income Inequality in Africa, 1990–2019: Measurement, Patterns, Determinants." In: *World Development* 163, p. 106162.

Chancel, L. and T. Piketty (2019). "Indian Income Inequality, 1922-2015: from British Raj to Billionaire Raj?" In: *Review of Income and Wealth* 65.1, pp. 33–62.

Clausing, K. (2020). "Profit Shifting Before and After the Tax Cuts and Jobs Act." In: *National Tax Journal* 73.4, pp. 1233–1266.

Clausing, K., E. Saez, and G. Zucman (2021). *Ending Corporate Tax Avoidance and Tax Competition: A Plan to Collect the Tax Deficit of Multinationals*. UCLA School of Law, Law-Econ Research Paper (20-12).

Coli, A., R. Istatkov, H. Jayyousi, F. Oehler, and O. Tsigkas (2022). *Distributional National Account Estimates for Household Income and Consumption: Methodological Issues and Experimental Results*. European Union & OECD statistical working papers. https://ec.europa.eu/eurostat/documents/3888793/14615020/KS-TC-21-010-EN-N.pdf/c26a1e27-7178-3f2e-c488-ff708bc90442?t=1654781916208.

Collier, I. L. and H. Siebert (1991). "The Economic Integration of Post-wall Germany." In: *American Economic Review* 81.2, pp. 196–201.

Conseil federal (2022). *Rapport Explicatif Relatif à l'Arrét Fédéral sur une Imposition Particuliére des Grands Groupes d'Entreprises (Mise en Œuvre du Projet Conjoint de l'OCDE et du G20 sur l'Imposition de l'Économie Numérique)*. https://www.newsd.admin.ch/newsd/message/attachments/70603.pdf.

De la Rosa Ramos, E. (2021). "Understanding the Gender Wage Gap in Mexico, 1880-1980." Unpublished manuscript.

De Rosa, A., I. Flores, and M. Morgan (2022). *More Unequal or Not as Rich? Revisiting the Latin American Exception*. Stone Center on Socio-Economic Inequality Working Paper No. 53.

Desai, A. (1968). *Real Wages in Germany 1871-1913*. Cambridge University Press, Cambridge.

Destatis (2023). *Studierende insgesamt und studierende Deutsche nach Geschlecht. Lange Reihen mit Jahresergebnissen ab 1975. Tabellen zu Bildung und Forschung mit Originalwerten und Veränderungsraten.* URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/lrbil01.html#242472.

Deutsche Bundesbank (2022). *Lange Zeitreihen. Zur Wirtschaftsentwicklung in Deutschland. Aktualisierte Ausgabe*. Table III. Geld- und Kapitalmarkt 3. Einlagen- und Kreditzinssätze deutscher Banken.

Devereux, M. P., J. Vella, and H. Wardell-Burrus (2022). *Pillar 2: Rule order, Incentives, and Tax Competition*. Oxford University Centre for Business Taxation Policy Brief. SSRN Working Paper http://dx.doi.org/10.2139/ssrn.4009002.

Devereux, M. P., J. Vella, M. Simmler, and H. Wardell-Burrus (2021). *What Is the Substance-Based Carve-Out under Pillar 2? And How Will It Affect Tax Competition?* EconPol Policy Brief 39.

Devereux, M. P. et al. (2020). *The OECD Global Anti-Base Erosion Proposal*. Oxford University Centre for Business Taxation Report.

Dharmapala, D. (2014). "What Do We Know About Base Erosion and Profit Shifting? A Review of the Empirical Literature." In: *Fiscal Studies* 35.4, pp. 421–448.

DiNardo, J., N. Fortin, and T. Lemieux (1996). "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." In: *Econometrica* 64, pp. 1001–1044.

Dornbusch, R., H. Wolf, and L. Alexander (1992). "Economic Transition in Eastern Germany." In: *Brookings Papers on Economic Activity* 1992.1, pp. 235–272.

European Central Bank (ECB) (2021). *Euro Foreign Exchange Reference Rates. ECB/Eurosystem Policy and Exchange Rates*. https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.html.

EUROSTAT (2014). *Europäisches System Volkswirtschaftlicher Gesamtrechnungen. ESVG 2010*. http://ec.europa.eu/eurostat/documents/3859598/5925749/KS-02-13-269-DE.PDF/0f8f50e6-173c-49ec-b58a-e2ca93ffd056.

– (2022). *Non-financial Transactions [nasa_10_nf_tr]*. https://ec.europa.eu/eurostat/data/database?node_code=nasa_10_nf_tr.

– (2023). *Gender pay gap statistics. The unadjusted gender pay gap, 2021.* URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Gender_pay_gap_statistics.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020). "Heterogeneity and Persistence in Returns to Wealth." In: *Econometrica* 88.1, pp. 115–170.

Ferreira Liotti, B., J. W. Ndubai, R. Wamuyu, I. Lazarov, and J. Owens (2022). "The Treatment of Tax Incentives under Pillar Two." In: *Transnational Corporations Journal* 2.2, pp. 25–46.

Fitzenberger, B. and G. Wunderlich (2002). "Gender Wage Differences in West Germany: A Cohort Analysis." In: *German Economic Review* 3.4, pp. 379–414.

Flora, P. (1975). *Indikatoren der Modernisierung. Ein historisches Datenhandbuch*. Studien zur Sozialwissenschaft No. 27. Springer Fachmedien, Wiesbaden.

Fritsch, M. (2004). "Entrepreneurship, Entry and Performance of New Business Compared in Two Growth Regimes: East and West Germany." In: *Journal of Evolutionary Economics* 14.5, pp. 525–542.

Fritsch, M., E. Bublitz, A. Sorgner, and M. Wyrwich (2014). "How much of a socialist legacy? The re-emergence of entrepreneurship in the East German transformation to a market economy." In: *Small Business Economics* 43.2, pp. 427–446.

Fuchs-Schündeln, N., D. Krueger, and M. Sommer (2010). "Inequality Trends for Germany in the Last Two Decades: A Tale of Two Countries." In: *Review of Economic Dynamics* 13.1, pp. 103–132.

Fuchs-Schündeln, N. and M. Schündeln (2009). "Who Stays, Who Goes, Who Returns? East-West Migration Within Germany since Reunification." In: *Economics of Transition* 17.4, pp. 703–738.

Fuest, C., F. Hugger, and F. Neumeier (2022). "Corporate Profit Shifting and the Role of Tax Havens: Evidence from German Country-by-Country Reporting Data." In: *Journal of Economic Behavior & Organization* 194, pp. 454–477.

Gallego-Granados, P. and K. Wrohlich (2020). *The Gender Wage Gap across the Distribution and Over Time*. SOEPpapers on Multidisciplinary Panel Data Research 1070-2020. German Socio-Economic Panel at DIW Berlin.

Galloway, P. R. (2007). *Galloway Prussia Database 1861 to 1914*. URL: www.patrickgalloway.com.

Galor, O. and J. Zeira (1993). "Income Distribution and Macroeconomics." In: *The Review of Economic Studies* 60.1, pp. 35–52.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2018). "Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA)." In: *Journal of Public Economics* 162, pp. 63–77.

Garcia-Bernardo, J., P. Janský, and G. Zucman (2022). *Did the Tax Cuts and Jobs Act Reduce Profit Shifting by US Multinational Companies?* NBER Working Paper 30086.

Gärtner, S. (2014). "German Stagnation versus Swedish Progression: Gender Wage Gaps in Comparison, 1960–2006." In: *Scandinavian Economic History Review* 62.2, pp. 137–162.

Gewerkschaftsbund der Angestellten (GdA) (1931). *Die wirtschaftliche und soziale Lage der Angestellten. Ergebnisse und Erkenntnisse aus der großen Erhebung des Gewerkschaftsbundes der Angestellten*. Sieben-Stäbe-Verlag, Berlin.

Glaß, F. and D. Kische (1930). *Die wirtschaftlichen und sozialen Verhältnisse der berufstätigen Frauen. Erhebung 1928/29 durchgeführt von der Arbeitsgemeinschaft Deutscher Frauenberufsverbände*. Carl Heymann, Berlin.

Goebel, J. et al. (2019). "The German Socio-Economic Panel (SOEP)." In: *Jahrbücher für Nationalökonomie und Statistik* 239.2, pp. 345–360.

Goldfayn-Frank, O. and J. Wohlfart (2019). "Expectation Formation in a New Environment: Evidence from the German Reunification." In: *Journal of Monetary Economics* 115, pp. 301–320.

Goldin, C. (1990). *Understanding the Gender Gap. An Economic History of American Women*. Oxford University Press, Oxford & New York.

Goldin, C. and S. Polachek (1987). "Residual Differences by Sex: Perspectives on the Gender Gap in Earnings." In: *The American Economic Review: Papers & Proceedings* 77.2, pp. 143–151.

Goldin, C. (1984). "The Historical Evolution of Female Earnings Functions and Occupations." In: *Explorations in Economic History* 21.1, pp. 1–27.

Goldin, C. and L. F. Katz (2009). "Why the United States Led in Education: Lessons from Secondary School Expansion, 1910 to 1940." In: *Human Capital and Institutions*.

Ed. by C. Goldin, L. Katz, D. Eltis, F. Lewis, and K. Sokoloff. Cambridge University Press, Cambridge, pp. 143–178.

Goldin, C., L. F. Katz, and I. Kuziemko (2006). "The Homecoming of American College Women: The Reversal of the College Gender Gap." In: *The Journal of Economic Perspectives* 20.4, pp. 133–156.

Gómez León, M. and H. J. de Jong (2019). "Inequality in Turbulent Times: Income Distribution in Germany and Britain, 1900–50." In: *The Economic History Review* 72.3, pp. 1073–1098.

Grabka, M. M. (2019). *SOEP 2017 – Codebook for the PEQUIV File 1984-2017: CNEF Variables with Extended Income Information for the SOEP*. SOEP survey paper series D - variable description and coding. German Socio-Economic Panel (SOEP).

– (2022). *Löhne, Renten und Haushaltseinkommen sind in den vergangenen 25 Jahren real gestiegen*. DIW-Wochenbericht No. 23.

Gustafsson, S. and A. Löfström (1991). "Policy Changes and Women's Wages in Sweden." In: *Stability and Change in Six Industrialized Countries*. Ed. by S. L. Willborn. International Review of Comparative Public Policy. JAI Press, Greenwich, USA.

Guzzardi, D., E. Palagi, A. Roventini, and A. Santoro (2022). *Reconstructing Income Inequality in Italy: New Evidence and Tax Policy Implications from Distributional National Accounts*. World Inequality Lab Working Paper 2022/02.

Heckemeyer, J. H. and M. Overesch (2017). "Multinationals' Profit Response to Tax Differentials: Effect Size and Shifting Channels." In: *Canadian Journal of Economics* 50.4, pp. 965–994.

Heineck, G. and B. Süssmuth (2013). "A Different Look at Lenin's Legacy: Social Capital and Risk Taking in the Two Germanies." In: *Journal of Comparative Economics* 41.3, pp. 789–803.

Herrmann, U. G., J. Friedrich, and K. Harney (2021). "Datenhandbuch zur deutschen Bildungsgeschichte: Band IV: Das berufsbildende Schulsystem in Deutschland 1815-1945." Unpublished manuscript.

Herrmann, U. G. (2006). "Schulische Berufsbildung für die weibliche Jugend 1900-1938." In: *Zeitschrift für Erziehungswissenschaft* 9.Beiheft 7/2006, pp. 105–125.

Heyman, F., F. Sjöholm, and P. Tingvall (2007). "Is There Really a Foreign Ownership Wage Premium? Evidence from Matched Employer-Employee Data." In: *Journal of International Economics* 73.2, pp. 355–376.

HM Treasury (2022). *OECD Pillar 2 - Consultation on Implementation*. https://www.gov.uk/government/consultations/oecd-pillar-2-consultation-on-implementation.

Hoffmann, W. G. (1965). *Das Wachstum der deutschen Wirtschaft seit der Mitte des 19. Jahrhunderts*. Springer, Berlin & Heidelberg.

Hohls, R. (1991). "Arbeit und Verdienst. Entwicklung und Struktur der Arbeitseinkommen im Deutschen Reich und in der Bundesrepublik (1885-1985)." PhD thesis. Freie Universität Berlin.

Hohls, R. and H. Kaelble (1989). *Die regionale Erwerbsstruktur im Deutschen Reich und in der Bundesrepublik 1895-1970*. Scripta Mercaturae Verlag, St. Katharinen.

Horst, T. and A. Curatolo (2020). "Assessing the Double count of Pretax Profit in the IRS Summary of CbC Data for Fiscal 2017." In: *Tax Notes International* 98.4.

Humphries, J. and J. Weisdorf (2015). "The Wages of Women in England, 1260–1850." In: *The Journal of Economic History* 75.2, pp. 405–447.

Hunt, J. (2001). "Post-Unification Wage Growth in East Germany." In: *The Review of Economics and Statistics* 83.1, pp. 190–195.

– (2002). "The Transition in East Germany: When Is a Ten-Point Fall in the Gender Wage Gap Bad News?" In: *Journal of Labor Economics* 20.1, pp. 148–169.

International Labor Organization (ILO) (2021). *Mean Nominal Monthly Earnings of Employees by Sex and Economic Activity (Local Currency) – Monthly*. ILO Statistics, https://ilostat.ilo.org/data/national-sources-catalogue/.

International Monetary Fund (IMF) (2021). *World Economic Outlook (WEO) Database*. https://www.imf.org/en/Publications/WEO/weo-database/2021/April.

IWH (2010). *Ostdeutschlands Transformation seit 1990 im Spiegel wirtschaftlicher und sozialer Indikatoren. 2. aktualisierte und verbesserte Auflage*. IWH-Sonderheft 1/2009. Institut für Wirtschaftsforschung Halle.

Janeba, E. and G. Schjelderup (2023). *The Global Minimum Tax Raises More Revenues Than You Think, or Much Less*. CESifo Working Paper No. 10318.

Jenderny, K. (Feb. 2015). "Mobility of Top Incomes in Germany." In: *Review of Income and Wealth* 62.2, pp. 245–265.

Jestl, S. and E. List (2020). *Distributional National Accounts (DINA) for Austria, 2004-2016*. World Inequality Lab - Working Paper 2020/23.

Johannesen, N. (2022). "The Global Minimum Tax." In: *Journal of Public Economics* 212.

Johansson, Å., Ø. B. Skeie, S. Sorbe, and C. Menon (2017). *Tax Planning by Multinational Firms: Firm-level Evidence from a Cross-country Database*. Tech. rep. 1355. OECD Economics Department Working Papers.

Kaiserlich Statistisches Amt (1884). *Berufsstatistik nach der allgemeinen Berufszählung vom 5. Juni 1882. 1. Teil. Berufsstatistik des Reichs und der kleineren Verwaltungsbezirke*. Statistik des Deutschen Reichs No. 2. Puttkammer & Mühlbrecht, Berlin.

Kaiserliches Statistisches Amt, Abteilung Arbeiterstatistik (1912). *Erhebung über die Arbeitsverhältnisse der Angestellten der Rechtsanwälte. Veranstaltet im Februar 1911.* Drucksachen des Kaiserlichen Statistischen Amtes, Abteilung Arbeiterstatistik, Erhebungen No. 9. Carl Heymann, Berlin.

Kaiserliches Statistisches Amte, Abteilung für Arbeiterstatistik (1912). *Erhebung über die Arbeitsverhältnisse der Angestellten der Rechtsanwälte*. Carl Heymanns Verlag, Berlin.

Karbe, A. (1928). *Die Frauenlohnfrage und ihre Entwicklung in der Kriegs- und Nachkriegszeit*. Hamburger Wirtschafts- und sozialwissenschafltiche Schriften No. 6. Carl Hinstorffs Verlag, Rostock.

KPMG (2021). *Corporate Tax Rates Table*. https://kpmg.com/sg/en/home/services/tax/tax-tools-and-resources/tax-rates-online/corporate-tax-rates-table.html.

Kroh, M., S. Kühne, R. Siegers, and V. Belcheva (2018). *SOEP-Core - Documentation of Sample Sizes and Panel Attrition (1984 until 2016)*. SOEP Survey Papers 480.

Kuczynski, J. (1961). *Darstellung der Lage der Arbeiter in Deutschland von 1789 bis 1849*. Die Geschichte der Lage der Arbeiter unter dem Kapitalismus. Teil 1. Die Geschichte der Lage der Arbeiter in Deutschland von 1789 bis zur Gegenwart, Band 1. Akademie-Verlag, Berlin.

– (1962). *Darstellung der Lage der Arbeiter in Deutschland von 1871 bis 1900*. Die Geschichte der Lage der Arbeiter unter dem Kapitalismus. Teil 1. Die Geschichte der Lage der Arbeiter in Deutschland von 1789 bis zur Gegenwart, Band 3. Akademie-Verlag, Berlin.

– (1963). *Studien zur Geschichte der Lage der Arbeiterin in Deutschland von 1700 bis zur Gegenwart*. Die Geschichte der Lage der Arbeiter unter dem Kapitalismus. Teil 1. Die Geschichte der Lage der Arbeiter in Deutschland von 1789 bis zur Gegenwart, Band 18. Akademie-Verlag, Berlin.

Kunze, A. (2005). "The Evolution of the Gender Wage Gap." In: *Labour Economics* 12.1, pp. 73–97.

Laffitte, S., J. Martin, M. Parenti, B. Souillard, and F. Toubal (2021). *Taxation of Multinationals: Design and Quantification*. Conseil d'analyse économique Focus 064bis-2021.

Länderrat des Amerikanischen Besatzungsgebiets (1949). *Statistisches Handbuch von Deutschland 1928-1944*. Ehrenwirth, München.

Lazear, E. P. and S. Rosen (1981). "Rank-Order Tournaments as Optimum Labor Contracts." In: *Journal of Political Economy* 89.5, pp. 841–864.

Lindert, P. H. and J. G. Williamson (2016). *Unequal Gains: American Growth and Inequality since 1700*. The Princeton Economic History of the Western World. Princeton University Press, Princeton.

Maier, F. (2007). *The Persistence of the Gender Wage Gap in Germany*. Discussion Paper 01, 12/2007. Harriet Taylor Mill-Institut für Ökonomie und Geschlechterforschung.

Mayer, C. (1999). "Entstehung und Stellung des Berufs im Berufsbildungssystem." In: *Beruf und Berufsbildung. Situation, Reformperspektiven, Gestaltungsmöglichkeiten*. Ed. by K. Harney and H.-E. Tenorth. Zeitschrift für Pädagogik No. 40. Beltz Verlag, Weinheim & Basel, pp. 35–60.

Meinert, R. (1958). "Die Entwicklung der Arbeitszeit in der deutschen Industrie 1820 - 1956." PhD thesis. Westfälische Wilhelms-Universität zu Münster.

Mergele, L., M. Hennicke, and M. Lubczyk (2020). *The Big Sell: Privatizing East Germany's Economy*. CESifo Working Paper 8566.

Mertens, M. and S. Mueller (2022). "The East-West German Gap in Revenue Productivity: Just a Tale of Output Prices?" In: *Journal of Comparative Economics* 50.3, pp. 815–831.

Mian, A., L. Straub, and A. Sufi (2020). *The Saving Glut of the Rich*. NBER Working Paper 26941.

– (2021). "Indebted Demand." In: *The Quarterly Journal of Economics* 136.4, pp. 2243–2307.

Milanovic, B., P. H. Lindert, and J. G. Williamson (2010). "Pre-Industrial Inequality." In: *The Economic Journal* 121.551, pp. 255–272.

Mincer, J. (1962). *Labor Force Participation of Married Women: A Study of Labor Supply*. Princeton University Press, Princeton.

Mirrlees, J. A. (1971). "An Exploration in the Theory of Optimum Income Taxation." In: *The Review of Economic Studies* 38.2, p. 175.

Morgan, M. and T. Neef (2020). *What's New About Income Inequality in Europe (1980-2019)?* World Inequality Lab - Issue Brief 2020/04.

Müller, W. (1983). "Frauenerwerbstätigkeit im Lebenslauf." In: *Strukturwandel der Frauenarbeit 1880 bis 1980*. Ed. by W. Müller, A. Willms-Herget, and J. Handl. Campus, Frankfurt am Main, pp. 55–105.

Müller, W., J. Handl, and A. Willms, eds. (1983). *Strukturwandel der Frauenarbeit 1880-1980*. Campus, Frankfurt am Main.

National Center for Education Statistics (NCES) (2023). *Table 111. High School Graduates, by Sex and Control of School: Selected Years, 1869-70 through 2020-21*. URL: https://nces.ed.gov/programs/digest/d11/tables/dt11_111.asp.

Neef, T. (2024). *The Long Way to Gender Equality: Gender Pay Differences in Germany, 1871-2021*. World Inequality Lab Working Paper No. 2024/02.

Noked, N. (2022). *The Case for Domestic Minimum Taxes on Multinationals*. SSRN Working Paper https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4053407.

Novokmet, F., T. Piketty, and G. Zucman (2018). "From Soviets to Oligarchs: Inequality and Property in Russia 1905-2016." In: *Journal of Economic Inequality* 16, pp. 189–223.

OECD (2020). *Tax Challenges Arising from Digitalisation - Economic Impact Assessment: Inclusive Framework on BEPS*. Paris: OECD Publishing, Paris).

– (2022a). *710. Financial Balance Sheets - Consolidated - SNA 2008.* Published at https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE610. Table 710.

– (2022b). *Corporate Tax Statistics. Table 1 - Aggregate Totals by Jurisdiction.* https://doi.org/10.1787/71d6bf8b-en.

– (2022c). *Important Disclaimer Regarding the Limitations of the Country-by-Country Report Statistics*. https://www.oecd.org/tax/tax-policy/anonymised-and-aggregated-cbcr-statistics-disclaimer.pdf.

Ogilvie, S. (1996). "The Beginnings of Industrialization." In: *Germany: A New Social and Economic History. Vol. II: 1630-1800*. Edward Arnold, London, pp. 263–308.

Okun, A. M. (1975). *Equality and Efficiency: The Big Tradeoff*. Brookings Institution Press, Washingtin.

Orlowski, R. and R. T. Riphahn (2009). "The East German Wage Structure after Transition." In: *Economics of Transition* 17.4, pp. 629–659.

Pierenkemper, T. (1987). *Arbeitsmarkt und Angestellte im Deutschen Kaiserreich 1880-1913. Interessen und Strategien als Elemente der Integration eines segmentierten Arbeitsmarktes*. F. Steiner Verlag, Wiesbaden.

Piketty, T., E. Saez, and G. Zucman (2018). "Distributional National Accounts: Methods and Estimates for the United States." In: *Quarterly Journal of Economics* 133.2, pp. 553–609.

– (2019a). *Distributional National Accounts: Methods and Estimates for the United States. Data Supplments. Appendix*. http://gabriel-zucman.eu/usdina.

Piketty, T. (2020). *Capital and Ideology*. Belknap Press of Harvard University Press, Cambridge, US & London.

– (2022). *A Brief History of Equality*. Belknap Press of Harvard University Press, Cambridge, US & London.

Piketty, T., L. Yang, and G. Zucman (2019b). "Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015." In: *American Economic Review* 109.7, pp. 2469–2496.

Pleijt, A. and J. L. Zanden (2021). "Two Worlds of Female Labour: Gender Wage Inequality in Western Europe, 1300–1800." In: *The Economic History Review* 74.3, pp. 611–638.

Priewe, J. (1993). "Privatisation of the Industrial Sector: The Function and Activities of the "Treuhandanstalt"." In: *Cambridge Journal of Economics* 17.3, pp. 333–348.

Rahlf, T., ed. (2015). *Deutschland in Daten. Zeitreihen zur Historischen Statistik*. Bundeszentrale für Politische Bildung, Bonn.

Reichsamt des Innern (1905). *Die Arbeitszeit der Fabrikarbeiterinnen. Nach Berichten der Gewerbeaufsichtsbeamten bearbeitet im Reichsamt des Innern*. R. v. Deckers Verlag, Berlin.

Reichskanzler-Amt (1877). *Ergebnisse der über die Frauen- und Kinder-Arbeit in den Fabriken auf Beschluß des Bundesraths angestellten Erhebungen*. Carl Hegemann Verlag, Berlin.

Riedel, L. and H. Stichnoth (2022). *Government Expenditure in the DINA Framework: Allocation Methods and Consequences for Post-Tax Income Inequality*. ZEW Discussion Paper No. 22-004.

Saez, E. and G. Zucman (2016). "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data." In: *The Quarterly Journal of Economics* 131.2, pp. 519–578.

– (2020). *The Triumph of Injustice - How the Rich Dodge Taxes and How to Make Them Pay*. WW Norton, New York.

Schröder, C. et al. (2020). "The Economic Research Potentials of the German Socio-Economic Panel Study." In: *German Economic Review* 21.3, pp. 335–371.

Sensch, J. (2004). *Histat-Datenkompilation online: Erwerbstätigkeitsstatistik in der Bundesrepublik Deutschland. Ausgewählte Daten von 1950 bis 2000.* https://histat.safe-frankfurt.de/.
– (2008). *Histat-Datenkompilation online: Preisindizes für die Lebenshaltung in Deutschland 1924 bis 2001. Verbraucherpreise seit 1881. Table C1 Entwicklung der Verbraucherpreise seit 1881 (1881-2001).* https://histat.safe-frankfurt.de/.
Sinn, G. and H.-W. Sinn (1994). *Jumpstart: The Economic Unification of Germany.* MIT Press, Cambridge, Massachusetts.
Smith, M., O. Zidar, and E. Zwick (2022). "Top Wealth in America: New Estimates Under Heterogeneous Returns." In: *The Quarterly Journal of Economics* 138.1, pp. 515–573.
Snower, D. and C. Merkl (2006). "The Caring Hand that Cripples: The East German Labor Market after Reunification." In: *American Economic Review Papers and Procedings* 96.2, pp. 375–381.
Stanfors, M. and F. Goldscheider (2017). "The Forest and the Trees: Industrialization, Demographic Change, and the Ongoing Gender Revolution in Sweden and the United States, 1870-2010." In: *Demographic Research* 36, pp. 173–226.
Statistisches Bundesamt (1954). *Die Verdienste der Angestellten in der gewerblichen Wirtschaft im November 1951 (Ergebnisse der Gehalts- und Lohnstrukturerhebung 1951/52).* Statistik der Bundesrepublik Deutschland No. 91. Statistisches Bundesamt, Wiesbaden.
– (1957). "Die methodischen Grundlagen der laufenden Verdiensterhebungen in Industrie und Handel." In: *Wirtschaft und Statistik, Volume 9, No. 6*, pp. 522–527.
– (1958a). *Die Arbeiterverdienste in der Landwirtschaft. März und September 1957, März 1958.* Fachserie M. Reihe 14 No. 1. W. Kohlhammer, Mainz & Stuttgart.
– (1958b). *Statistisches Jahrbuch für die Bundesrepublik Deutschland 1957.* W. Kohlhammer, Mainz & Stuttgart.
– (1959). *Die Arbeiterverdienste in der Landwirtschaft. September 1959.* Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1960). *Die Arbeiterverdienste in der Landwirtschaft. September 1960.* Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1961a). *Die Arbeiterverdienste in der Landwirtschaft. September 1961.* Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1961b). *Gehalts- und Lohnstrukturerhebung 1957. Heft 2. Verdienste der Angestellten in der gewerblichen Wirtschaft im Oktober 1957.* Statistik der Bundesrepublik Deutschland No. 246. W. Kohlhammer, Mainz & Stuttgart.
– (1963a). *Die Arbeiterverdienste in der Landwirtschaft. September 1962.* Fachserie M. Reihe 14 No. 1. W. Kohlhammer, Mainz & Stuttgart.
– (1963b). *Die Arbeiterverdienste in der Landwirtschaft. September 1963.* Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1964a). *I. Entwicklung der Erwerbstätigkeit (Ergebnisse des Mikrozensus) April 1963.* Fachserie A. Reihe 6. W. Kohlhammer, Mainz & Stuttgart.

– (1964b). *Statistisches Jahrbuch für die Bundesrepublik Deutschland 1963*. W. Kohlhammer, Mainz & Stuttgart.
– (1965). *Die Arbeiterverdienste in der Landwirtschaft. September 1964*. Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1966a). *Die Arbeiterverdienste in der Landwirtschaft. September 1965*. Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1966b). *I. Gewerbliche Wirtschaft und Dienstleistungsbereich. Angestelltenverdienste 1962*. Fachserie M. Reihe 17. W. Kohlhammer, Mainz & Stuttgart.
Statistisches Bundesamt (1967). *Die Arbeiterverdienste in der Landwirtschaft. September 1966*. Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1968). *Die Arbeiterverdienste in der Landwirtschaft. September 1967*. Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1969). *Die Arbeiterverdienste in der Landwirtschaft. September 1968*. Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1970a). *Die Arbeiterverdienste in der Landwirtschaft. September 1969*. Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1970b). *I. Gewerbliche Wirtschaft und Dienstleistungsbereich. Angestelltenverdienste 1966*. Fachserie M. Reihe 17. W. Kohlhammer, Mainz & Stuttgart.
– (1971). *Die Arbeiterverdienste in der Landwirtschaft. September 1970*. Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1972). *Die Arbeiterverdienste in der Landwirtschaft. September 1971*. Fachserie M. Reihe 14. W. Kohlhammer, Mainz & Stuttgart.
– (1976). *I. Entwicklung der Erwerbstätigkeit Mai 1975 (Ergebnisse des Mikrozensus aus der EG-Arbeitskräftestichprobe)*. Fachserie A. Reihe 6. W. Kohlhammer, Mainz & Stuttgart.
– (1977). *Stand und Entwicklung der Erwerbstätigkeit. Mai 1976 (Ergebnisse des Mikrozensus)*. Fachserie 1. Reihe 4.1. W. Kohlhammer, Mainz & Stuttgart.
– (1981). *Gehalts- und Lohnstrukturerhebung. Arbeiter- und Angestelltenverdienste im Produzierenden Gewerbe, im Groß- und Einzelhandel, bei Kreditinstituten und im Versicherungsgewerbe*. Fachserie 16. W. Kohlhammer, Mainz & Stuttgart.
– (1984). *Stand und Entwicklung der Erwerbstätigkeit Mai 1983 (Ergebnisse des Mikrozensus)*. Fachserie 1. Reihe 4.1. W. Kohlhammer, Mainz & Stuttgart.
– (1991). *Statistisches Jahrbuch für das vereinte Deutschland*. Metzler-Poeschel, Stuttgart.
– (1993). *Gehalts- und Lohnstrukturerhebung. Arbeiter- und Angestelltenverdienste im Produzierenden Gewerbe, im Groß- und Einzelhandel, Bank und im Versicherungsgewerbe 1990*. Fachserie 16. Heft 1. Statistisches Bundesamt, Wiesbaden.
– (1997). *Stand und Entwicklung der Erwerbstätigkeit 1997 (Ergebnisse des Mikrozensus)*. Fachserie 1. Reihe 4.1.1. W. Kohlhammer, Mainz & Stuttgart.
– (2001). *Methode und Verfahren. Volkswirtschaftliche Gesamtrechnungen. Einkommensrechnungen. Überblick über die Methoden und Grundlagen in der Bundesrepublik Deutschland*. Statistisches Bundesamt.

– (2009). *Umsatzsteuerstatistik (Voranmeldungen) 2007*. Fachserie 14, Reihe 8.1. Statistisches Bundesamt, Wiesbaden.
– (2012). *Lohn- und Einkommensteuerstatistik. Statistik über die Personengesellschaften und Gemeinschaften 2007*. Statistisches Bundesamt, Wiesbaden.
– (2016). *Volkswirtschaftliche Gesamtrechnungen. Inlandsprodukt und Nationaleinkommen nach ESVG 2010. Methoden und Grundlagen.* Fachserie 18, Reihe S.30. Statistisches Bundesamt, Wiesbaden.
– (2022a). *Körperschaftsteuerpflichtige, Gesamtbetrag der Einkünfte, Körperschaftsteuer: Deutschland, Jahre, Rechtsformen, Gewinnfälle und Verlustfälle. Table 73211-0001.* Destatis Genesis Database. https://www-genesis.destatis.de/genesis/online/.
– (2022b). *Lohn- und Einkommensteuer 1992-2018*. Fachserie 14, Reihe 7.1. Statistisches Bundesamt, Wiesbaden.
– (2022c). *National Accounts. Sectoral Accounts. Annual results 1991 onwards. 1991 to 2021*. Statistisches Bundesamt, Wiesbaden.
– (2022d). *Preise. Verbraucherpreisindizes für Deutschland. Lange Reihen ab 1948*. Statistisches Bundesamt, Wiesbaden.
– (2022e). *Volkswirtschaftliche Gesamtrechnungen. Inlandsproduktberechnung. Detaillierte Jahresergebnisse, 1991-2021.* Fachserie 18, Reihe 1.4. Statistisches Bundesamt, Wiesbaden.
– (2023a). *Ausgaben und Einnahmen der Kriegsopferfürsorge. Table 22731-0002.* Destatis Genesis Database. https://www-genesis.destatis.de/genesis/online.
– (2023b). *Bruttoausgaben, Einnahmen, Nettoausgaben der Sozialhilfe. Tabelle 22111-0001 - 22111-0004 & 22111-0023.* Destatis Genesis Database. https://www-genesis.destatis.de/genesis/online.
– (1993-2022). *Realsteuervergleich. Realsteuern, kommunale Einkommen und Umsatzsteuerbeteiligungen, 1992-2021*. Fachserie 14, Reihe 10.1. Statistisches Bundesamt, Wiesbaden.
– (1957-1990). *Statistisches Jahrbuch für die Bundesrepublik Deutschland*. W. Kohlhammer, Mainz & Stuttgart.
– (1991-2005). *Statistisches Jahrbuch für die Bundesrepublik Deutschland*. Metzel-Poeschel, Stuttgart.
– (2006-2020). *Statistisches Unternehmensregister. Unternehmen nach Wirtschaftsabschnitten und Größenklassen der sozialversicherungspflichtig Beschäftigten.* https://-www.destatis.de/DE/Themen/Branchen-Unternehmen/Unternehmen/Unternehmensregister/Tabellen/stat-unternehmen-beschaeftigten-groessenklassen-wz08.html.
– (1995-2021). *Umsatzsteuerstatistik (Voranmeldungen)*. Fachserie 14, Reihe 8.1. Statistisches Bundesamt, Wiesbaden.
– (2007-2015). *Verdienste und Arbeitskosten. Arbeitnehmerverdienste*. Fachserie 16. Reihe 2.3. Statistisches Bundesamt, Wiesbaden.

Statistisches Reichsamt (1921). *Lohn- und Gehaltserhebung vom Februar 1920*. Statistik des deutschen Reichs No. 293. Puttkammer & Mühlbrecht, Berlin.

Statistisches Reichsamt (1928). *Statistisches Jahrbuch für das Deutsche Reich No. 47*. Verlag von Reimar Hobbing, Berlin.
– (1929). "Hauptergebnisse der amtliche Lohnerhebung in der Schuhindustrie." In: *Wirtschaft und Statistik. Volume 9, No. 20*, pp. 838–849.
– (1930). *Statistisches Jahrbuch für das Deutsche Reich No. 49*. Verlag von Reimar Hobbing, Berlin.
– (1931a). "Die Nachprüfung und Neubearbeitung der amtlichen Tariflohnstatistik." In: *Vierteljahreshefte zur Statistik des Deutschen Reichs. No. 40*, pp. 94–113.
– (1931b). *Statistisches Jahrbuch für das Deutsche Reich No. 50*. Verlag von Reimar Hobbing, Berlin.
– (1932). *Statistisches Jahrbuch für das Deutsche Reich No. 51*. Verlag von Reimar Hobbing, Berlin.
– (1935). "Einkommen der Angestellten im Bankgewerbe. Gesamtergebnis der amtlichen Gehaltserhebung im Bankgewerbe für Februar 1934." In: *Wirtschaft und Statistik. Volume 15, No. 4*, pp. 136–140.
– (1937). *Der Steuerabzug vom Arbeitslohn in den Jahren 1932 und 1934*. Statistik des deutschen Reichs No. 492. Verlag für Sozialpolitik, Wirtschaft und Statistik, Paul Schmidt, Berlin.
– (1938). "Die Entwicklung der Arbeitsverdienste in den letzten 10 Jahren." In: *Wirtschaft und Statistik. Volume 18, No. 4*, pp. 157–161.
– (1939a). *Der Steuerabzug vom Arbeitslohn im Jahre 1936*. Statistik des deutsches Reichs No. 530. Verlag für Sozialpolitik, Wirtschaft und Statistik, Paul Schmidt, Berlin.
– (1939b). "Die Arbeitsverdienste im September 1938." In: *Wirtschaft und Statistik. Volume 19, No. 6*, p. 235.
– (1940). "Die Entwicklung der Arbeitsverdienste im Jahre 1939." In: *Wirtschaft und Statistik. Volume 20, No. 18*, pp. 433–434.
– (1941). "Die amtliche Lohnerhebung in der Landwirtschaft für das Jahr 1937." In: *Wirtschaft und Statistik. Volume 21, No. 21*, pp. 403–406.
– (1943a). "Die Arbeitsverdienste im März 1943." In: *Wirtschaft und Statistik. Volume 23, No. 11*, pp. 279–280.
– (1943b). "Die Arbeitsverdienste im September 1942." In: *Wirtschaft und Statistik. Volume 23, No. 3*, pp. 82–84.
Statistisches Reichsamt, Abteilung für Arbeiterstatistik und Reichsamt für Arbeitsvermittlung (1919). "Arbeitsbedingungen. Zusammensetzung und Löhne der Arbeiterschaft während des Krieges." In: *Reichsarbeitsblatt, Volume 17, No. 8*, pp. 618–623.
Statistisches Reichsamt, Abteilung für Arbeiterstatistik und Reichsamt für Arbeitsvermittlung (1920). "Kriegsverdienste der Arbeiterschaft bis Anfang 1919 nach den Jahresberichten des Gewerbeaufsichtsbeamten." In: *Reichsarbeitsblatt, Volume 18, No. 1*, pp. 56–65.

Suhr, S. (1930). *Die weiblichen Angestellten. Arbeits- und Lebensverhältnisse. Eine Umfrage des Zentralverbandes der Angestellten*. Verlag Zentralverband der Angestellten, Berlin.

Sutela, P. (1998). *Privatization in the Countries of Eastern and Central Europe and the Former Soviet Union*. UNU WIDER Working Papers 146.

Svensson, L. (1997). "En Historia om Löneutjämning - Kvinnors och Mäns Löner i Långsiktigt Perspektiv." In: *Kvinnors och Mäns Löner - Varför så Olika*. Ed. by I. Persson and E. Wadensjö. Fritzes, Stockholm, pp. 14–44.

Svensson, L. (2003). "Explaining Equalization. Political Institutions, Market Forces, and the Reduction of the Gender Wage Gap in Sweden, 1920-95." In: *Social Science History* 27.3, pp. 371–395.

Ther, P. (2016). *Europe since 1989. A History*. Princeton University Press, Princeton.

Tørsløv, T. R., L. S. Wier, and G. Zucman (2019). *The Missing Profits of Nations. Estimates updated to 2016 and 2017*. Tables available on https://missingprofits.world/.

– (2018). *The Missing Profits of Nations*. NBER Working Paper 24701.

– (2020). *The Missing Profits of Nations*. Unpublished update of NBER Working Paper Nr. 24701 (2018).

Uliczka, N. (2023). "When the Celtic Tiger Relaxed its Corporate Tax Bite: An Analysis of Effects on Top and Upper Middle Income Shares in Ireland." In: *Economics Letters* 222, p. 110933.

UN (2020). *World Economic Situation and Prospects 2020*. https://www.un.org/development/desa/dpad/publication/world-economic-situation-and-prospects-2020/.

United Nations, European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, and World Bank (2009). *System of National Accounts 2008*. http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf.

van der Straaten, K., N. Pisani, and A. Kolk (2020). "Unraveling the MNE Wage Premium." In: *Journal of International Business Studies* 51, pp. 1355–1390.

Verein für Socialpolitik (1892). *Die Verhältnisse der Landarbeiter in Nordwestdeutschland, Württemberg, Baden und in den Reichslanden*. Die Verhältnisse der Landarbeiter in Deutschland No. 1. Duncker & Humblot, Leipzig.

von der Goltz, T. (1872). *Die ländliche Arbeiterfrage und ihre Lösung*. Verlag von A.W. Kasemann, Danzig.

von der Goltz, T. (1875). *Die Lage der ländlichen Arbeiter im Deutschen Reich. Bericht an die vom Congress deutscher Landwirthe niedergesetzte Commission zur Ermittlung der Lage der ländlichen Arbeiter im Deutschen Reich*. Verlag von Wiegandt, Hempel & Parey, Berlin.

Von Hagen, J., R. R. Strauch, and G. B. Wolff (2002). *East Germany: Transition with Unification - Experiments and Experiences*. ZEI Working Paper B 19-2002.

Vorstand des Deutschen Landarbeiter-Verbandes (1926a). *Bericht des Verbandsvorstandes über die Verbandsarbeit in den Jahren 1920 bis 1922. Niederschrift von der 3. Generalversammlung des Deutschen Landarbeiter-Verbandes 17. und 18. September 1923*.

Schriften des Deutschen Landarbeiter-Verbandes No. 16. Vorstand des Deutschen Landarbeiter-Verbandes, Berlin.
– (1926b). *Bericht des Verbandsvorstandes über die Verbandsarbeit in den Jahren 1923 bis 1925*. Schriften des Deutschen Landarbeiter-Verbandes. Vorstand des Deutschen Landarbeiter-Verbandes, Berlin.
– (1929). *Bericht des Verbandsvorstandes über die Verbandsarbeit in den Jahren 1926 bis 1928*. Schriften des deutschen Landarbeiter-Verbandes. Vorstand des Deutschen Landarbeiter-Verbandes, Berlin.
Wagner, G. G., J. R. Frick, and J. Schupp (2007). "The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements." In: *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 127.1, pp. 139–169.
Willms, A. (1982). "Modernisierung durch Frauenarbeit." In: *Historische Arbeitsmarktforschung*. Ed. by T. Pierenkemper and R. Tilly. Vandenhoeck & Ruprecht, Göttingen, pp. 37–71.
– (1983a). "Grundzüge der Entwicklung der Frauenarbeit von 1880 bis 1980." In: *Strukturwandel der Frauenarbeit 1880-1980*. Ed. by W. Müller, J. Handl, and A. Willms. Campus Verlag, Frankfurt & New York, pp. 25–54.
– (1983b). "Segregation auf Dauer? Zur Entwicklung des Verhältnisses von Frauenarbeit und Männerarbeit in Deutschland, 1882-180." In: *Strukturwandel der Frauenarbeit 1880-1980*. Ed. by W. Müller, J. Handl, and A. Willms. Campus Verlag, Frankfurt & New York, pp. 107–182.
Willms-Herget, A. (1985). *Frauenarbeit: zur Integration der Frauen in den Arbeitsmarkt*. Campus, Frankfurt am Main.
Windolf, P. (1996). "Die Transformation der ostdeutschen Betriebe." In: *Berliner Journal für Soziologie* 4, pp. 467–488.
World Bank (1996). *From Plan to Market. World Development Report 1996*. Oxford University Press, Oxford.
Wunderlich, F. (1961). *Farm Labor in Germany 1810-1945. Its Historical Development Within the Framework of Agricultural and Social Policy*. Princeton University Press, Princeton.
Yu, C. (1994). "Search for a New Identity: Female White-Collar Employees in Germany, 1918-1933." PhD thesis. University of Pennsylvania.
Ziegler, P. (2010). "Der geschlechtsspezifische Lohnunterschied in den Vereinigten Staaten und der Europäischen (Wirtschafts-)Gemeinschaft/Union - ein Vergleich." PhD thesis. Universität Wien.

List of Figures

List of Tables

Colophon

This thesis was typeset with the typesetting system LaTeX 2$_\varepsilon$ using

- the editor TeXstudio,
- the engine LuaLaTeX,
- the documentclass `scrreprt` from KOMA-Script,
- the text fonts CM-Unicode and the math fonts Latin Modern Math and STIX2,
- the bibliography package BibLaTeX,
- the bibliography engine `Biber`.