Accepted: 26 April 2024

DOI: 10.1111/labr.12274

ORIGINAL ARTICLE

The gender gap in lifetime earnings: A microsimulation approach

Rick Glaubitz[1,2] | Astrid Harnack-Eber[1] | Miriam Wetter[1]

[1]School of Business & Economics, Freie Universität Berlin, Berlin, Germany

[2]Center for Evidence-Based Healthcare (ZEGV), University Hospital and Faculty of Medicine Carl Gustav Carus at TU Dresden, Dresden, Germany

Correspondence
Rick Glaubitz, Center for Evidence-Based Healthcare (ZEGV), University Hospital and Faculty of Medicine Carl Gustav Carus at TU Dresden, Fetscherstraße 74, 01307, Dresden, Germany.
Email: rick.glaubitz@ukdd.de

Abstract

To obtain a more complete understanding of the persisting gender earnings gap in Germany, this paper investigates both the cross-sectional and lifetime dimension of gender inequalities. Based on a dynamic microsimulation model, we analyse how gender differences accumulate over work lives to examine the lifetime dimension of the gender gap. We estimate an average gender gap in lifetime earnings of 51.5 per cent for birth cohorts 1964–72. We show that this unadjusted gender lifetime earnings gap increases strongly with the number of children, ranging from 17.3 per cent for childless women to 68.0 per cent for women with three or more children. Results from a counterfactual analysis approach show an adjusted gender gap in lifetime earnings of around 10 per cent, suggesting that the gender gap in lifetime earnings is rather driven by gender differences in observable characteristics than by differences in rewards.

JEL CLASSIFICATION

D15, D31, J16, J22, J31

1 | INTRODUCTION

As most research on the gender pay gap has focused on differences in monthly or annual earnings data, evidence on how gender inequalities add up over the life course is still limited. In contrast to cross-sectional analysis, which generally only focuses on a snapshot of an individual's employment career, a life cycle perspective acknowledges that earnings are transient across individuals' careers. For example, earnings can be temporarily low during educational training or even zero during times of unemployment or labour market inactivity. The analysis of accumulated earnings over the entire career ('lifetime earnings') offers more comprehensive insights into individuals' long-term position in the earnings distribution and is more closely linked to individuals' life chances (see, e.g., Corneo, 2015; Tamborini et al., 2015). Therefore, the concept of lifetime earnings is suitable to examine the extent to which gender inequalities accumulate over the life cycle. However, due to high data requirements, there is only scarce empirical evidence on gender lifetime earnings gaps (e.g., Boll et al., 2017; Guvenen et al., 2021, 2022). In addition, these studies are often limited by their use of administrative data and subsequent lack of family-related information (e.g., number of children, marital status). Because the average labour market participation of women is lower than that of men at both the intensive and extensive margin due to, e.g., different effects of parenthood (see, e.g., Goldin, 2014; Kleven et al., 2019), an analysis of the household context is necessary for a more comprehensive understanding of the underlying drivers of gender differentials in lifetime earnings.

This study uses the Socio-Economic Panel (SOEP) to shed light on the role of women's family backgrounds in gender differences, from both a cross-sectional and a lifetime perspective. Using an Oaxaca–Blinder decomposition, we find that the gaps can largely be explained by both the extensive and intensive margins of labour. On average, women have less work experience and work fewer hours than men, explaining a large part of women's lower earnings.

To further take advantage of the detailed socioeconomic and family background information in the SOEP survey compared with administrative data sources, we use a dynamic microsimulation model to obtain full employment biographies, and subsequently lifetime earnings data. In contrast to the existing studies using administrative data, this allows us to analyse the extent to which gender gaps in lifetime earnings vary by family background (number of children). Furthermore, this approach also leads to a more comprehensive sample than the ones of earlier studies for Germany (Boll et al., 2017; Bönke et al., 2015) as we are, for the first time, able to include self-employed individuals, civil servants and women with longer unemployment/inactivity spells. We find that women earn on average 51.5 per cent less than men over their work life. This unadjusted gender gap in lifetime earnings correlates largely with the number of children and ranges from 17.3 per cent for childless women to 68.0 per cent for women with three children or more.

To investigate which part of the observed gender gap in lifetime earnings can be explained by the observable differences in the distribution of characteristics (e.g., work experience, level of education) across gender and which part is due to differences in labour market returns to characteristics, we estimate women's counterfactual lifetime earnings. We find that around 80 per cent of the observed lifetime earnings gap can be explained by different characteristics across men and women, leading to an adjusted gender lifetime earnings gap of 10 per cent.

Our paper is related to three different strands of literature. First and most generally, it contributes to the extensive literature on the gender gap in pay and its drivers. Existing studies show that a large extent of the pay gap can be attributed to fewer hours worked and higher discontinuity of female employment biographies (e.g., Bertrand et al., 2010; Blau & Kahn, 2017).[1] The persistence of this gender earnings inequality is mainly due to different effects of parenthood on men's and women's labour market behaviour, and consequently their earnings (see, e.g., Angelov et al., 2016; Bütikofer et al., 2018; Kleven et al., 2021; Kleven & Landais, 2017; Waldfogel, 1998). In line with previous studies (e.g., Gallen et al., 2019; Goldin, 2014; Juhn &

McCue, 2017), we confirm that gender differences in annual earnings increase during the period of family formation, peak around age 40 and slowly decrease until retirement, leading to an inverse U shape of the gender annual earnings gap over the work life.

Studies for Germany show that the cross-sectional earnings gap between mothers and non-mothers are largely driven by domestic work and childcare duties (e.g., Beblo & Wolf, 2002; Ejrnæs & Kunze, 2013). Strikingly, child penalties on women's pay are high in Germany compared with other countries (see, e.g., Kleven et al., 2019). This is often attributed to longer maternal leave entitlement and a higher rate of part-time work for women in Germany (see, e.g., Gangl & Ziefle, 2009; Harkness & Waldfogel, 2003). However, more recent studies also stress the influence of relative conservative gender norms in Germany in this context (e.g., Kleven et al., 2019, 2020).

Second, our study adds to the scarce literature on lifetime earnings and specifically to what extent these differ by gender. Due to the high data requirements, the literature on the gender pay gap and its evolution has primarily focused on cross-sectional hourly wages, annual earnings or earnings over a short time period. Using administrative data for the United States, Guvenen et al. (2021) show that the fraction of women among lifetime top earners is significantly lower than that of men for birth cohorts 1956–58. On average, lifetime top earners in the United States tend to be individuals who experience high earnings growth over the first half of their life cycle—the period when the gender gap increases the most, likely due to family-related reasons. In a later study, Guvenen et al. (2022) provide evidence that the large gender lifetime earnings gap is narrowing over time, with women's median lifetime earnings increasing while men's median lifetime earnings decreases for younger birth cohorts.

Using administrative data from the German Pension Register (VSKT), Bönke et al. (2015) find evidence that intragenerational lifetime earnings inequality for West German men born between 1935 and 1969 has increased, largely due to losses in the bottom of the lifetime earnings distribution. They also supplement their work with additional results on West German women. However, due to data restrictions, their data only includes women with stable employment biographies. Therefore, the VSKT data is not representative for most women mainly due to the high rate of inactivity among women of older cohorts and should not be used for estimating the gender lifetime earnings gap in Germany. Closest to our paper is the study by Boll et al. (2017) analysing the gender lifetime earnings gap in Germany. Using the administrative Sample of Integrated Labour Market Biographies (SIAB), they estimate an unadjusted gender lifetime earnings gap of 46 per cent for West German birth cohorts 1950–64. They show that the gender gap widens significantly during the age of family formation and that gender differences in work experience and hours worked explains around two-thirds of this overall gender lifetime earnings gap. However, SIAB data does not offer any information about individuals' family background. Hence, to the best of our knowledge, our study is the first to extensively examine the influence of parenthood in the context of gender differentials in lifetime earnings.

Third, our study contributes from a methodological point of view to the literature on the implementation of dynamic microsimulation models for the simulation of missing information (e.g., Levell & Shaw, 2016; Li & O'Donoghue, 2013; Zucchelli et al., 2012). A dynamic microsimulation approach refers to a regression-based simulation which predicts the transition probabilities of different units (e.g. individuals or households) for moving from one state to another between two different points in time. Therefore, in contrast to studies using a splicing approach (e.g., Grabka & Goebel, 2017; Westermeier et al., 2012) where sequences of existing biographies are stitched together to construct full life-cycle data, the microsimulation approach typically 'ages' the data year by year (Li & O'Donoghue, 2013). We apply a dynamic microsimulation model to SOEP survey data to obtain complete earnings biographies, which facilitates lifetime earnings analyses. Combining simulation models with survey data is a well-established method to deal with missing observations and panel attrition, which often impede using survey data to conduct long-term analyses (see, e.g., Brown et al., 2009; Coronado et al., 2011). For Germany, e.g., there are existing studies simulating employment biographies using SOEP data (e.g., Bonin et al., 2015; Geyer & Steiner, 2014; Hänisch & Klos, 2016).

The next section introduces our dataset and starts by analysing cross-sectional gender differences in hourly wages and annual earnings over the work life by using an Oaxaca–Blinder decomposition. Section 3 describes our microsimulation approach to obtain full work biographies and presents our estimates for the unadjusted and adjusted gender lifetime earnings gap. Section 4 concludes.

2 | CROSS-SECTIONAL ANALYSIS

The cross-sectional analysis explores how gender gaps in hourly wages and annual earnings develop with increasing age and to investigate if short-term differences already follow certain patterns across gender. This first step is crucial to subsequently better understand how gender inequalities in labour market characteristics and earnings add up or equalize over the entire work life.

2.1 | Data and methodology

Our study is based on the German SOEP. The SOEP is a representative annual panel survey questioning about 30,000 individuals across 15,000 households since 1984. In contrast to administrative data, the SOEP includes a rich set of socioeconomic variables, detailed labour market information and household background including information on the partner and children.[2] Specifically, we use the 35th wave of the SOEP comprising data for the years 1984–2018.

We restrict our cross-sectional analysis to birth cohorts 1940–79. These are the same birth cohorts used for the underlying regressions of our microsimulation model in Section 3. We observe these cohorts at least once between the ages of 38 and 44 in the SOEP. This age restriction is crucial as it is the age frame when individuals' cross-sectional earnings show the highest correlation with lifetime earnings (Björklund, 1993; Bönke et al., 2015) and is therefore needed to successfully simulate life-cycle profiles in Section 3. Furthermore, we focus on West German individuals because those born in East Germany were only included in the SOEP after German reunification in 1990. The poor comparability of the Federal Republic of Germany and the German Democratic Republic with respect to labour market institutions and economic systems does not allow us to simulate missing information for East Germans before 1990. It is important to note that, by focusing only on employed individuals, we are solely analysing the observed distribution of earnings and not the counterfactual distribution that would be observed if everyone were in employment and had positive earnings.

Section 2 focuses on the evolution of cross-sectional hourly wages and annual earnings with increasing age over the work life. This approach sheds light on two main components of the gender gap in lifetime earnings; the gender gap in hourly wages shows the differences in the compensation between women and men for 1 h of their work, while the gap in annual earnings reveals dissimilarities driven by the variation in working hours.

We use an Oaxaca–Blinder decomposition (see Blinder, 1973; Oaxaca, 1973) to investigate how much of the difference in the observed gender gap is driven by different observable characteristics between men and women and how much can be attributed to different returns to characteristics within the labour market.[3] Using this decomposition approach the gender gap G in the labour market outcome variable L (here, logarithmic hourly wage and logarithmic annual earnings) is defined as:

$$G_x = E(L_{mx}) - E\left(L_{fx}\right). \tag{1}$$

Therefore, G is the gender differential between the means of outcome L for men (m) and women (f) at age x. We can then divide the gender gap into two parts. First, the endowment part, which is the component of the gender gap which is due to differences in the distribution of observable characteristics between men and women. And second, the coefficient part, which accounts for differences in returns to characteristics. Hence, the coefficient part shows the gender driven difference of the labour market's willingness to pay for the same characteristics obtained by either men or women. However, note that the coefficient part may also include gender differences that remain unexplained in our model due to data and model restrictions. We run the following regression model separately by sex (s) and age (x) for the labour outcome L[4]:

$$L_{s,i,x} = \alpha_{s,i,x} + \beta_{s,i,x} Z_{s,i,x} + \epsilon_{s,i,x}, \quad E(\varepsilon_{s,x}) = 0, \quad s \in \{F, M\}, x \in [20, 60], \tag{2}$$

where Z is a vector of control variables including work experience measured as number of working years, full-time or part-time work, work sector, highest education level, marital status and number of children. In addition, we control for cohort and time effects.[5]

2.2 | Hourly wage

Overall, employed men have significantly higher hourly wages than employed women (see Table A1). At the beginning of their work life at age 20, men earn on average €9.37 per hour while women's average wage is only €7.97 per hour. In line with results found by the Federal Statistical Office (Statistisches Bundesamt, 2017), the average hourly wages of men in our sample then almost triples over the work life to €26.13 per hour at age 60. In contrast, women's hourly wages only increase to €17.48, already showing significant gender differences in wage growth over the work life.

The solid line in Figure 1 shows the evolution of the gender gap in hourly wages in log points from age 20 to 60. Notably, the gender gap remains stable over the early years of work life. At age 25, men's hourly wages are only 0.059 log points higher than women's and the difference is still insignificant (see also Table 1). However, during the time of family formation and childcare, this gap drastically widens up to a highly significant difference of 0.378 log points at age 45.[6] Afterwards, the growth of the gender gap in hourly wages slows down and remains relatively stable with a peak at age 55. This finding is consistent for all cohorts (see Figure A2). In line with our findings, previous studies also documented a widening of the gender wage gap over the life cycle (e.g., Anderson et al., 2002; Angelov et al., 2016; Tyrowicz et al., 2018).

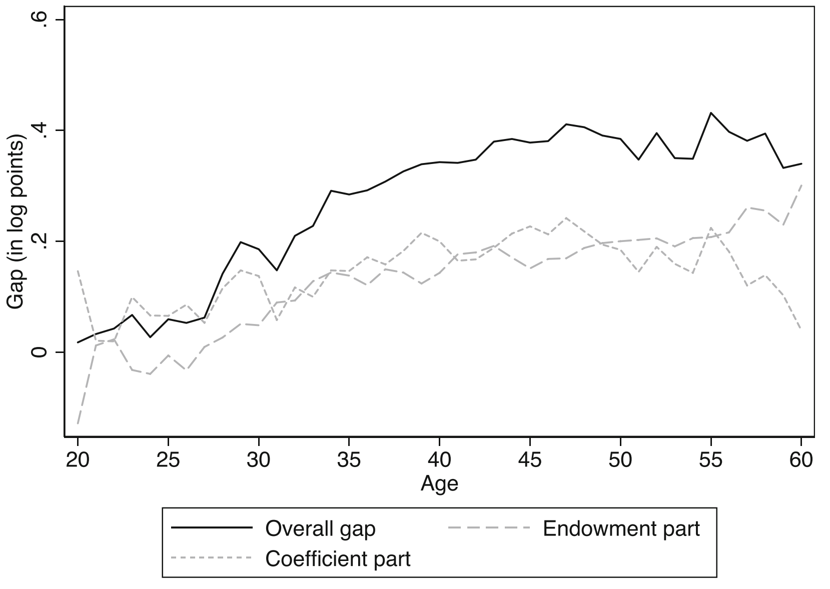

FIGURE 1 Gender gap in hourly wages. Only employed individuals are considered. Cohorts 1940–79, weighted sample. *Source:* Own calculations based on SOEP v35.

The results of the Oaxaca–Blinder decomposition are displayed by the grey lines in Figure 1 and also in Table 1. Visibly, the widening of the gender gap in hourly wages over the work life can be explained by the increase in the endowment part, while the coefficient part of the gender gap shapes its overall trend. At younger ages, the different distribution of characteristics does not play a role yet. Therefore, at the beginning of work life all wage differences between men and women are due to different returns to labour market characteristics. Main differences in characteristics such as work experience or family background widen only later in life; after age 25, the high and significant coefficients for work experience in Table 1 show that the increase of the endowment part can largely be explained by women's lower gain of work experience with increasing age. By the age of 60, men have accumulated on average 37.32 years of full-time and 1.09 years of part-time work experience, whereas women have accumulated on average only 19.65 years of full-time and 13.32 years of part-time work experience (see Table A1). Our results show that these large differences in work experience are crucial in explaining the gender gap in hourly wages. By the end of work life, differences in work experience account for 0.309 log points of the overall gender wage gap of 0.340 log points. Hence, around 90 per cent of the overall gender gap of 40.5 per cent in hourly wages can be explained by differences in work experience.

In contrast to the stable growth of the endowment part, the evolution of the coefficient part follows a slight inverse U-shape. At age 20, the gender gap cannot be explained through differences of characteristics across genders, but the coefficient part amounts to 0.126 log points. This means that even if we observed the same labour market characteristics in women and men, men's wages would be 0.126 log points (13.4 per cent) higher than women's wages at this age. The coefficient part of the gender gap peaks at 0.247 log points (28.0 per cent) at age 45 and then declines again to a difference of 0.042 log points (4.3 per cent) just before retirement.[7] In contrast to the endowment part, none of the variable groups have a constant significant influence on the overall gender gap, including the constant itself.[8] Therefore, not one individual

TABLE 1 Oaxaca–Blinder decomposition of hourly wage gender gap.

	(1) Age 20	(2) Age 25	(3) Age 30	(4) Age 35	(5) Age 40	(6) Age 45	(7) Age 50	(8) Age 55	(9) Age 60
Overall									
Men	1.963*** (0.040)	2.563*** (0.021)	2.771*** (0.015)	2.912*** (0.012)	2.980*** (0.012)	3.008*** (0.013)	3.019*** (0.016)	3.054*** (0.022)	3.003*** (0.026)
Women	1.945*** (0.033)	2.503*** (0.024)	2.586*** (0.019)	2.628*** (0.017)	2.637*** (0.016)	2.630*** (0.017)	2.634*** (0.017)	2.622*** (0.020)	2.663*** (0.031)
Difference	0.018 (0.052)	0.059 (0.031)	0.186*** (0.024)	0.284*** (0.021)	0.343*** (0.020)	0.378*** (0.021)	0.385*** (0.024)	0.432*** (0.030)	0.340*** (0.041)
Endowment	−0.108** (0.040)	−0.031 (0.018)	0.033* (0.015)	0.107*** (0.018)	0.107*** (0.020)	0.131*** (0.022)	0.196*** (0.026)	0.200*** (0.028)	0.297*** (0.043)
Coefficient	0.126* (0.050)	0.091** (0.033)	0.152*** (0.028)	0.177*** (0.026)	0.235*** (0.029)	0.247*** (0.028)	0.189*** (0.036)	0.231*** (0.040)	0.042 (0.060)
Endowment									
Children	0.003 (0.005)	−0.000 (0.003)	0.001 (0.001)	−0.002 (0.002)	−0.004 (0.003)	−0.009* (0.004)	−0.015* (0.007)	−0.014 (0.011)	−0.013 (0.013)
Married	0.001 (0.004)	−0.001 (0.004)	0.004 (0.002)	0.002 (0.002)	0.003 (0.002)	0.005* (0.002)	0.000 (0.002)	−0.002 (0.003)	0.010 (0.008)
Experience	−0.068* (0.030)	0.024* (0.011)	0.084*** (0.013)	0.167*** (0.019)	0.207*** (0.022)	0.224*** (0.028)	0.228*** (0.026)	0.264*** (0.032)	0.309*** (0.049)
Part time	0.002 (0.008)	−0.041** (0.013)	−0.027 (0.015)	−0.067*** (0.019)	−0.096*** (0.019)	−0.066*** (0.019)	−0.026 (0.019)	−0.056* (0.024)	−0.030 (0.024)
Education	−0.008 (0.007)	−0.019** (0.006)	−0.009 (0.006)	0.021** (0.007)	0.019** (0.006)	0.022*** (0.006)	0.031*** (0.007)	0.030*** (0.008)	0.045*** (0.011)
Cohort	−0.001 (0.005)	−0.002 (0.003)	−0.000 (0.002)	0.001 (0.002)	0.003 (0.002)	0.001 (0.001)	0.002 (0.001)	0.003 (0.002)	−0.001 (0.002)
Sector	−0.037 (0.029)	0.009 (0.011)	−0.019* (0.008)	−0.017** (0.006)	−0.025*** (0.007)	−0.045*** (0.007)	−0.024*** (0.007)	−0.024* (0.010)	−0.022* (0.011)
Coefficient									

(Continues)

TABLE 1 (Continued)

	(1) Age 20	(2) Age 25	(3) Age 30	(4) Age 35	(5) Age 40	(6) Age 45	(7) Age 50	(8) Age 55	(9) Age 60
Children	0.003 (0.005)	0.030 (0.020)	0.095*** (0.025)	0.063 (0.034)	0.002 (0.038)	0.007 (0.046)	−0.075 (0.044)	−0.048 (0.053)	0.080 (0.088)
Married	−0.006 (0.010)	0.021 (0.027)	0.041 (0.031)	0.052 (0.034)	0.090* (0.038)	−0.008 (0.041)	0.100* (0.045)	0.047 (0.056)	0.020 (0.074)
Experience	0.243*** (0.063)	0.207 (0.129)	0.067 (0.128)	−0.094 (0.127)	−0.058 (0.112)	0.005 (0.170)	−0.150 (0.226)	0.196 (0.280)	−0.973 (0.995)
Part time	0.013 (0.027)	0.031 (0.037)	0.008 (0.023)	−0.021 (0.019)	0.032 (0.021)	0.032 (0.021)	0.030 (0.025)	0.145*** (0.036)	0.036 (0.054)
Education	0.329 (0.195)	−0.144 (0.174)	−0.210 (0.141)	−0.105 (0.230)	−0.564* (0.256)	0.183 (0.237)	0.568 (0.295)	0.131 (0.840)	0.722 (0.648)
Cohort	0.083 (0.044)	0.023 (0.085)	0.002 (0.036)	−0.007 (0.125)	0.091 (0.049)	−0.019 (0.034)	0.030 (0.041)	−0.030 (0.038)	0.002 (0.040)
Sector	0.222 (0.193)	−0.445** (0.139)	−0.436** (0.160)	−0.053 (0.116)	−0.066 (0.147)	−0.139 (0.142)	−0.037 (0.124)	−0.085 (0.188)	−0.153 (0.192)
Constant	−0.762** (0.276)	0.368 (0.283)	0.587* (0.255)	0.343 (0.308)	0.708* (0.322)	0.186 (0.317)	−0.276 (0.383)	−0.125 (0.887)	0.308 (−1.090)
N	765	1782	3053	4323	5356	5592	4304	2866	1758

Note: Standard errors in parentheses; The stars refer to the following significance level: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. The different drivers are summarized as followed: 'Children': Number of children; 'Married': Dummy variable on marital status, 'Experience': Total years of working full time, part time or being inactive (also squared); 'Part time': Dummy variable indicating full time or part time work; 'Education': Dummy variables indicating highest level of educational attainment, 'Sector': Occupational sector; 'Cohort': Cohort dummies. Cohorts 1940–79, weighted sample. *Source*: Own calculations based on SOEP v35.

effect dominates the coefficient part of the overall gender gap, but the coefficient part is instead a combination of many individual influences including those not controlled for in this regression model.

In summary, the gender gap in hourly wages is determined by two factors: first, women have in sum less favourable labour market characteristics compared with men, and second, even if they have the same observable characteristics, the labour market rewards women worse than men. The influence of differences in characteristics grows significantly with age, mainly through increasing differences in accumulated work experience across gender. Of the observed gender gap of 40.5 per cent (0.340 log points) at age 60, different characteristics account for 87 per cent (0.297 log points). This leads to an adjusted gender gap in hourly wages of 5.3 per cent.

2.3 | Annual earnings

In addition to earning less per hour, women also work on average fewer hours than men do. Therefore, the gender gap in annual earnings might be even wider than the gap in hourly wages due to gender differences in the intensive margin of work.

Figure 2 shows the overall gender gap in annual earnings, the part of the gap due to different characteristic across gender (endowment part) and the part of the gender gap due to differences in coefficients (coefficient part). Visibly, the gender gap in annual earnings is significantly higher than the gender gap in hourly wages. At the peak of the gap at age 40 (0.829 log points corresponding to 129.1 per cent), men's average annual earnings are more than twice as high than women's. Similar to the gender gap in hourly wages, the gender gap in annual

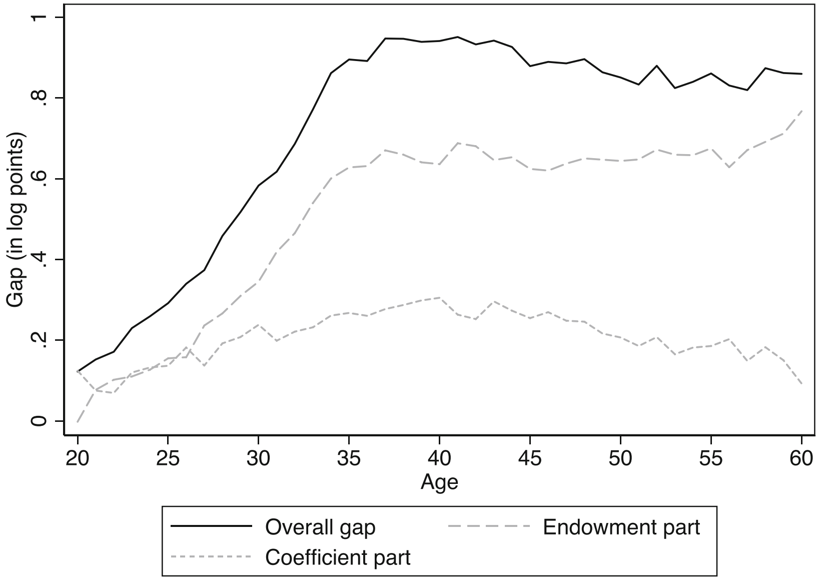

FIGURE 2 Gender gap in annual earnings. Only employed individuals are considered. Does not include values of zero annual earnings. Cohorts 1940–79, weighted sample. *Source*: Own calculations based on SOEP v35.

TABLE 2 Oaxaca–Blinder decomposition of the annual earnings gender gap.

	(1) Age 20	(2) Age 25	(3) Age 30	(4) Age 35	(5) Age 40	(6) Age 45	(7) Age 50	(8) Age 55	(9) Age 60
Overall									
Men	9.462*** (0.041)	10.155*** (0.024)	10.460*** (0.017)	10.623*** (0.012)	10.695*** (0.013)	10.713*** (0.015)	10.717*** (0.017)	10.693*** (0.022)	10.542*** (0.030)
Women	9.424*** (0.039)	9.950*** (0.026)	9.923*** (0.026)	9.854*** (0.024)	9.867*** (0.021)	9.915*** (0.021)	9.904*** (0.024)	9.868*** (0.028)	9.775*** (0.040)
Difference	0.038 (0.057)	0.205*** (0.036)	0.537*** (0.031)	0.769*** (0.027)	0.829*** (0.024)	0.797*** (0.025)	0.812*** (0.030)	0.825*** (0.036)	0.766*** (0.050)
Endowment	−0.081 (0.046)	0.102*** (0.026)	0.318*** (0.028)	0.538*** (0.025)	0.555*** (0.025)	0.529*** (0.028)	0.609*** (0.034)	0.657*** (0.036)	0.747*** (0.050)
Coefficient	0.119* (0.046)	0.103** (0.036)	0.219*** (0.036)	0.231*** (0.028)	0.274*** (0.031)	0.269*** (0.033)	0.203*** (0.042)	0.168*** (0.042)	0.019 (0.053)
Endowment									
Children	−0.000 (0.005)	0.001 (0.003)	0.001 (0.002)	−0.001 (0.002)	−0.003 (0.003)	−0.010* (0.005)	−0.018* (0.007)	−0.020 (0.011)	−0.012 (0.012)
Married	0.001 (0.004)	0.001 (0.003)	0.005 (0.003)	0.001 (0.001)	0.001 (0.002)	0.001 (0.003)	−0.002 (0.002)	−0.004 (0.003)	0.003 (0.008)
Experience	−0.058 (0.030)	0.033* (0.014)	0.129*** (0.015)	0.203*** (0.018)	0.244*** (0.022)	0.273*** (0.026)	0.306*** (0.028)	0.334*** (0.033)	0.351*** (0.040)
Hours worked	0.023 (0.018)	0.082*** (0.015)	0.214*** (0.024)	0.327*** (0.022)	0.313*** (0.021)	0.282*** (0.023)	0.310*** (0.025)	0.331*** (0.029)	0.371*** (0.034)
Education	−0.003 (0.006)	−0.021** (0.007)	−0.011 (0.007)	0.023** (0.007)	0.020** (0.006)	0.023*** (0.007)	0.032*** (0.007)	0.033*** (0.008)	0.048*** (0.012)
Cohort	−0.001 (0.005)	−0.002 (0.003)	0.000 (0.002)	0.001 (0.002)	0.002 (0.002)	0.000 (0.001)	0.002 (0.001)	0.003 (0.003)	−0.002 (0.002)
Sector	−0.043 (0.026)	0.008 (0.011)	−0.019* (0.009)	−0.016* (0.006)	−0.021** (0.007)	−0.040*** (0.008)	−0.021** (0.008)	−0.020 (0.010)	−0.012 (0.011)

TABLE 2 (Continued)

	(1) Age 20	(2) Age 25	(3) Age 30	(4) Age 35	(5) Age 40	(6) Age 45	(7) Age 50	(8) Age 55	(9) Age 60
Coefficient									
Children	0.000 (0.004)	0.037* (0.015)	0.155*** (0.028)	0.069 (0.037)	−0.005 (0.040)	0.025 (0.049)	−0.093 (0.048)	−0.074 (0.057)	0.069 (0.081)
Married	−0.004 (0.010)	0.033 (0.025)	0.040 (0.033)	0.049 (0.036)	0.092* (0.040)	−0.012 (0.048)	0.113* (0.048)	0.075 (0.055)	0.060 (0.074)
Experience	0.224*** (0.063)	0.146 (0.139)	0.045 (0.146)	−0.139 (0.156)	−0.159 (0.121)	−0.150 (0.177)	−0.169 (0.269)	0.378 (0.423)	−0.964 (0.966)
Hours worked	−0.348 (0.222)	−0.260 (0.184)	−0.625*** (0.160)	−0.811*** (0.105)	−0.801*** (0.100)	−0.783*** (0.143)	−0.891*** (0.144)	−1.152*** (0.145)	−0.621*** (0.151)
Education	0.356* (0.173)	−0.082 (0.182)	−0.240 (0.151)	−0.165 (0.245)	−0.595* (0.263)	0.229 (0.310)	0.720* (0.332)	0.932 (0.700)	0.809 (0.699)
Cohort	0.067 (0.044)	0.002 (0.051)	−0.038 (0.039)	0.028 (0.124)	0.134** (0.046)	−0.005 (0.036)	0.039 (0.044)	−0.030 (0.040)	0.003 (0.040)
Sector	0.166 (0.180)	−0.295* (0.144)	−0.396* (0.166)	−0.126 (0.130)	−0.055 (0.159)	−0.091 (0.150)	−0.008 (0.142)	−0.120 (0.213)	−0.098 (0.217)
Constant	−0.344 (0.330)	0.521 (0.304)	1.278*** (0.310)	1.326*** (0.358)	1.663*** (0.343)	1.056* (0.413)	0.492 (0.465)	0.159 (0.834)	0.759 (−1.112)
N	765	1782	3053	4323	5356	5592	4304	2866	1758

Note: Standard errors in parentheses; the stars refer to the following significance level: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. The different drivers are summarized as followed: ‘Children’: Number of children; ‘Married’: Dummy variable on marital status, ‘Experience’: Total years of working full time, part time or being inactive (also squared); ‘Hours worked’: Hours worked per year; ‘Education’: Dummy variables indicating highest level of educational attainment, ‘Sector’: Occupational sector; ‘Cohort’: Cohort dummies. Cohorts 1940–79, weighted sample. *Source*: Own calculations based on SOEP v35.

earnings increases rapidly until age 35 and remains on a constant high level during the years of child rearing. Afterward, it only declines slightly until retirement. This finding is in line with earlier studies for the United States providing evidence for a similar course of the cross-sectional gender gap in annual earnings over the work life (Goldin, 2014; Juhn & McCue, 2017).

When decomposing the overall gender gap in annual earnings, we find that the larger gap (in comparison to the gap in hourly wages) can be explained by the significantly higher endowment part. While the gender gap due to differences in coefficients is only slightly higher than in the model for hourly wages, the endowment part has more than tripled.[9] This result underlines the importance of differences in the intensive labour margin across gender.

Table 2 shows that the endowment part of the gender gap in annual earnings also includes the lesser work experience that women have accumulated over their life cycle. Moreover, the lower number of hours worked by women per year at all ages influences the gender gap to an even greater extent. These findings are in line with previous studies (e.g., Bertrand et al., 2010; Gallen et al., 2019).

At age 35, women's annual earnings are on average 0.327 log points lower than men's due to their lower number of work hours.[10] In addition, women's earnings are on average 0.203 log points lower than men's due to the lesser work experience they have accumulated up to this age. This means that at this point around half of the overall gap can be explained by the distribution of working hours and around a quarter can be explained by the different distribution of work experience. The influence of work experience steadily increases over the life cycle and peaks at age 60 with 0.351 log points. In contrast, differences in the level of education or family background play a smaller role.

The coefficient part of the gender gap in annual earnings is positive throughout the life cycle. This means that, besides worse characteristics, women also face less beneficial coefficients in their wage regression (see Tables 2, A4, and A5). This is especially pronounced between ages 30 and 45. Potential explanations may include employers' fear of women's higher risk of work absence due to pregnancy and child rearing (see, e.g., Correll et al., 2007) and more women opting for less financially rewarding positions in return for higher work flexibility after having children (see, e.g., Goldin, 2014). However, interestingly, for individuals aged 60 the coefficient part of the gender gap is very small in magnitude and no longer statistically significant, indicating that at this point the gender gap in annual earnings can almost be entirely explained by observable differences in endowments.

3 | MICROSIMULATION AND LIFETIME ANALYSIS

The previous section gave an analysis of the cross-sectional gender gaps in hourly wages and annual earnings, their development with increasing age and drivers. In this section, we investigate how gender earnings differentials might accumulate or balance out over the complete work life by looking at lifetime earnings. This allows us to shed light on the lifetime dimension of the gender earnings gap.

3.1 | Data and methodology

We continue to use the SOEP as it offers panel data containing not only detailed labour market but also family background information, which administrative data cannot offer. However, the SOEP suffers from panel mortality. Only around 10 per cent of the participants have been observed for at least 20 years or more, with an average participation period of 9.36 years (see Figure A3). To investigate lifetime earnings for a larger sample, we implement a dynamic microsimulation approach to fill in the missing data of non-observed years during an individual's work live. This approach yields complete earnings data for the observation period which we can combine with the rich set of socioeconomic characteristics and family information in the SOEP.

To implement our dynamic microsimulation model successfully, we need to add the following restrictions to our cross-sectional sample: First, our lifetime earnings investigation focuses on birth cohorts 1964–72 only. This approach gives us the opportunity to observe the cohorts starting at age 20 until at least age 45. This restriction is important as we know in the German context that only lifetime earnings up to age 45 and older are sufficient proxies for complete lifetime earnings up to age 60 (Bönke et al., 2015). Second, we exclude individuals who were only observed prior to turning 30 because labour market patterns of individuals in their twenties are very unstable and could yield a life-cycle bias (see, e.g., Bönke et al., 2015; Brenner, 2010; Haider & Solon, 2006). Furthermore, the probability of observing the highest educational attainment accurately increases significantly with age 30 and older (Autorengruppe Bildungsberichterstattung, 2018) and observing the true educational attainment is crucial as education levels and earnings patterns over the work life are highly correlated (see, e.g., Bhuller et al., 2011; Bönke et al., 2015; Brunello et al., 2017). Third, we also exclude individuals without at least two consecutive observation years in the SOEP. Otherwise, no panel information is available and a distinction between individual short- and long-term labour shocks would not be possible. After eliminating those observations, we are left with a sample of 3315 women and 3212 men across birth cohorts 1964–72 (see Table A6) for the dynamic microsimulation.

3.1.1 | Dynamic microsimulation model

We apply a dynamic microsimulation model to fill in missing information in non-observed years based on the individual's employment biography and socioeconomic characteristics. The general idea and structure of our microsimulation approach follows the approach proposed by Levell and Shaw (2016). To exploit our data to its fullest extent, we use both forward- and backward-looking simulations. The simulation starts either at an individual's first or last observed year in the data. As shown in Figure 3, we impute the missing variables in time $t+1$ or $t-1$ by

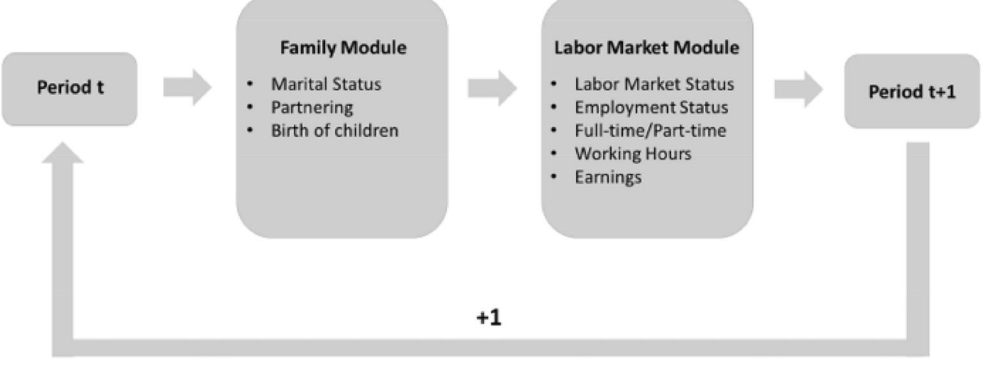

FIGURE 3 Dynamic microsimulation model. *Source*: Own diagram.

running the regressions for our dynamic microsimulation in two consecutive steps: First, missing observations of marital status, fertility (i.e. number of children) and partners are simulated in the Family Module (Module 1). Second, the obtained information from Module 1 is used in addition to other provided data to simulate individuals' labour market information in the Labour Market Module (Module 2). Completing both modules yields the successful imputation of all relevant information in time $t+1$ or $t-1$. Afterwards, the process moves forward to the simulation of the next years, i.e., $t+2$ or $t-2$, $t+3$ or $t-3$ and so on. The simulation ends after reaching 1984 in the backward looking and 2017 in the forward-looking process. We obtain a full dataset without any missing earnings or family information between 1984 and 2017.

In addition, investigating complete lifetime patterns for our sample requires us to extend our simulation for 15 additional years until 2032, when the youngest birth cohort 1972 turns 60. The prediction of employment biographies after 2017 is based on regression parameters of observed individuals from older cohorts, while we assume that general labour market characteristics (e.g., unemployment rate) remain stable after 2017. We also account for differences in trends using cohort and age fixed effects in our regressions. Nevertheless, this prediction comes naturally with a certain level of uncertainty due to the assumption that trends remain stable—an assumption that neglects, e.g., labour market effects related to the COVID-19 pandemic. The simulation ends when all missing information between 1984 and 2032 is simulated.

Within each module, the simulation of variables is based on estimating transition probabilities between 2 years, e.g., if marital status changes from year t to $t+1$. The estimation of a change of a variable j between two periods is then implemented by using a random process (see, e.g., Neufeld, 2000; Plümper & Troeger, 2007; Zucchelli et al., 2012). For each individual observation i we simulate the transition probability from time t to $t+1$ or $t-1$ and then draw a random number N_{it} from a uniform $[0,1]$ distribution. If the calculated transition probability P_{it} is larger than the drawn random number N_{it} $(P_{it} > N_{it})$, a transition occurs. In contrast, no transition takes place if $P_{it} \leq N_{it}$. Therefore, high transition likelihoods do not always induce actual transitions and even low transition probabilities may still lead to transitions. This approach helps to account for the uncertainty that comes with a simulation. In addition, we use a Monte Carlo simulation approach to test the robustness of our results (see Figures A5 and A6). The results of the Monte Carlo simulation confirm the high robustness of our simulation outcomes.

Next, we will give brief summaries about both simulation modules. Detailed information on all the regression models of every simulation step can be found in Table A7.

3.1.2 | Module 1: Family Module

Empirical evidence shows that family background strongly influences women's labour market behaviour (e.g., Kleven & Landais, 2017). Therefore, we need information on individual's family background before simulating earnings for non-observed years. Naturally, simulation of family background comes with a certain level of uncertainty. However, all individuals in our sample completed entry questionnaires including questions on marital status and, if applicable, birth years of children before entering the survey. This allows us to reconstruct full family histories which eliminates the necessity of the backward looking simulation component in this module. Consequently, missing data occurs exclusively after individuals left the survey and is therefore less of a problem for individuals observed at an older age. Around 80 per cent of the women in our sample are observed at age 40 or older, which means that the number of simulated child births is limited.

The Family Module then consists of two steps: predicting marital status, including a partnering module when necessary, and predicting births of children for individuals with missing information. First, we run logistic regressions separately by gender s (female or male) and marital status m (single or partnered) in year t to predict the individual transition probability p^{married} to change the marital status from year t to the missing year $t+1$:

$$p^{\text{married}}_{m,s,t+1} = \beta_0 + \beta_1 X_{m,s,t} + \epsilon_{m,s,t}, \quad E(\epsilon_{m,s,t}) = 0, m \in \{S,P\}, s \in \{F,M\}, t \in [1984, 2017]. \tag{3}$$

The regression consists of a set of explanatory variables X_t including socioeconomic characteristics (e.g., education, age, migration background) and labour market behaviour (e.g., employment status). In addition, we control for the number of years that an individual's marital status has remained unchanged until year t. Table A7 gives a detailed overview about all covariates included in each regression-based simulation step.

Recall that if $P_{it} \leq N_{it}$, the marital status stays the same and if $P_{it} > N_{it}$, the marital status changes. Therefore, this simulation step has four possible outcomes: First, a person who is single in year t can remain single in $t+1$. Second, married individuals can stay married. Here we assume that their partners remain the same. Third, married individuals in period t can get divorced and become single in $t+1$.[11] And fourth, singles in year t can get married in $t+1$. In this last case, we run a Partner Module to assign a partner.[12] This allows us to account for partners' characteristics when simulating family and labour market decisions. Using Mahalanobis distance matching (Mahalanobis, 1936) we identify five ‘best’ partners based on age, education and region for each observation. We then randomly assign one of the five potential partners to the individual. Our matching procedure is not unique, i.e., one individual can serve multiple times as a ‘donor’ for partner characteristics. In this way, we ensure a sufficient pool of potential partners.

Next, we simulate whether a woman will give birth to a child in the next non-observed period $t+1$ by marital status m:

$$p^{\text{birth}}_{m,t+1} = \beta_0 + \beta_1 X_{m,t} + \epsilon_{m,t}, \quad E(\epsilon_{m,t}) = 0, m \in \{S,P\}, t \in [1984, 2017]. \tag{4}$$

Again, X_t represents a set of explanatory variables including socioeconomic characteristics like information on existing children and labour market information. The simulation is similar to the approach described in the simulation of the marital status. Afterwards, the information on an individual's number of children is updated accordingly. In contrast to our marriage simulation, births are only simulated for women. Children are then attached to men depending on women's family background.

Because we estimate transition likelihoods for $t+1$ by using information available in period t, the likelihood of a change of the marital status or a childbirth in $t+1$ do not influence the transition probability of one another. Therefore, the order in which we implement fertility and marital transitions is irrelevant and does not alter our results.

Completing the Family Module for years 1984–2032 results in a sample with full information on family characteristics. Figure 4 shows that our simulated data (dashed line) replicates the initial distributions before the simulation (solid line) very accurately. In Panel A, the women's average number of children increases strongly until age 35. Then, the growth rate slows down and comes to a natural stop between ages 40 and 45 due to biological reasons.

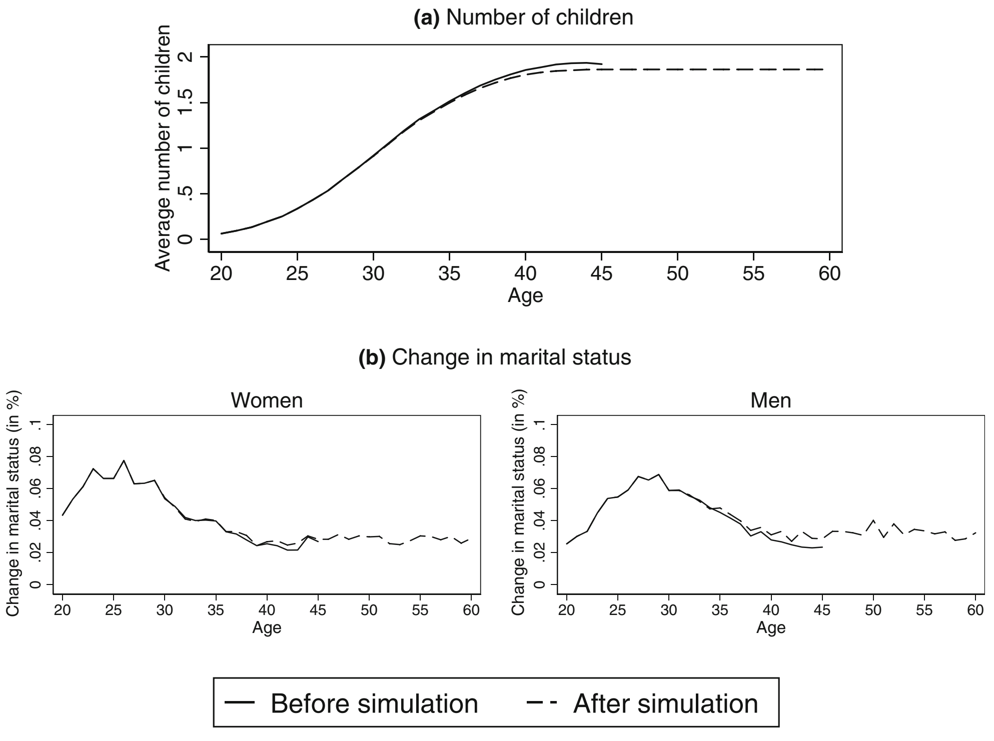

FIGURE 4 Family information before and after simulation. Panel A shows the average number of children of women by age before and after the simulation. Panel B demonstrates the share of individuals in our sample changing their marital status before and after the simulation. *Source*: Own calculations based on SOEP v35.

Panel B displays the percentage change in marital status by age. Obviously, both men and women follow the same trend over the life cycle. Most changes in marital status happen in the beginning of life.

3.1.3 | Module 2: Labour Market Module

The Labour Market Module generates complete information on an individual's employment biography through five stages: labour market participation, employment status, type of work arrangement (full-time or part-time), annual working hours and annual earnings. In this module, we use both forward and backward simulation as the introductory survey questionnaires do not allow us to construct sufficient work histories. Our model description will focus on the forward-looking simulation component. However, the backward-looking part of the simulation follows the same methodology.

In general, the logic and structure of this module is very similar to our approach in the Family Module. We start with the estimation of $p^{\text{lmp}}_{(m,t+1)}$, the probability for an individual of marital status m to change the labour market participation lmp from year t to year $t+1$. The labour market participation dummy variable is equal to 1 if individuals are unemployed or employed and equal to 0 if they are not attached to the labour market (e.g., due to parental or sick leave). We run the estimation separately by gender s and marital status m:

$$p^{\mathrm{lmp}}_{s,m,t+1} = \beta_0 + \beta_1 p^{\mathrm{lmp}}_{s,m,t} + \beta_2 p^{\mathrm{lmp}}_{s,m,t-1} + \beta_3 X_{s,m,t} + \epsilon_{s,m,t},$$
$$E(\epsilon_{s,m,t}) = 0,\ s \in \{F, M\},\ m \in \{S, P\},\ t \in [1984, 2017]. \tag{5}$$

$X_{(s,m,t)}$ is again a vector of control variables with socioeconomic characteristics like marital status, partner's earnings and their own labour market information. Furthermore, we include lagged dependent variables to account for path dependencies over the work life while still modelling a dynamic data generating process.[13] If individuals are recorded as not participating in year $t+1$, we directly record their earnings as zero for $t+1$ and do not include them in the subsequent steps. For individuals who are active in the labour market, we next run a regression to estimate the probability to change their employment status $p^{\mathrm{emp}}_{(s,m,e,t+1)}$ (employed/unemployed) from year t to year $t+1$. The following model is run separately by gender s, marital status m and employment status e:

$$p^{\mathrm{emp}}_{s,m,e,t+1} = \beta_0 + \beta_1 p^{\mathrm{emp}}_{s,m,e,t} + \beta_2 p^{\mathrm{emp}}_{s,m,e,t-1} + \beta_3 X_{s,m,e,t} + \epsilon_{s,m,e,t},$$
$$E(\epsilon_{s,m,e,t}) = 0,\ s \in \{F, M\},\ m \in \{S, P\},\ e \in \{0, 1\},\ t \in [1984, 2017]. \tag{6}$$

Once more, the regression contains a set of explanatory variables $X_{(s,m,e,t)}$ including information on family and the socioeconomic background. Also included in the control vector is the work history of individuals. To this end, we measure work experience by years of full-time work, part-time work and years without any work until year t to account for the different levels of labour market experience.

Individuals recorded as unemployed in year $t+1$ after this first regression step receive zero earnings in $t+1$ and are excluded from further estimations. For all employed individuals, the dynamic microsimulation moves forward with a logistic regression simulating if individuals worked full- or part-time in year $t+1$. In the next step, we estimate the probability of changing full-time or part-time arrangements from year t to year $t+1$:

$$p^{\mathrm{wt}}_{(s,m,t+1)} = \beta_0 + \beta_1 p^{\mathrm{wt}}_{s,m,t} + \beta_2 p^{\mathrm{wt}}_{s,m,t-1} + \beta_3 X_{s,m,t} + \epsilon_{s,m,t},$$
$$E(\epsilon_{s,m,t}) = 0,\ s \in \{F, M\},\ m \in \{S, P\},\ t \in [1984, 2017]. \tag{7}$$

Again, $X_{(s,m,t)}$ includes the usual control variables in addition to the labour market history. We can now move on to estimate the precise number of annual working hours in $t+1$ separately for part-time and full-time workers. We use an OLS regression model following the same logic as the earnings regression model as introduced in Equation (2.8).[14]

Finally, we use an earnings regression to estimate the annual earnings $y_{(s,m,t+1)}$ by gender s and marital status m:

$$y_{(s,m,t+1)} = \beta_0 + \beta_1 y_{s,m,t} + \beta_2 y_{s,m,t-1} + \beta_3 X_{s,m,t} + \epsilon_{s,m,t},$$
$$E(\epsilon_{s,m,t}) = 0,\ s \in \{F, M\},\ m \in \{S, P\},\ t \in [1984, 2017]. \tag{8}$$

$X_{(s,m,t)}$ now includes information about the work history in years of full-time work, part-time work or unemployment, working hours in t and, if applicable, partner and child information. All earnings are price-adjusted and presented in 2015 Euro. The simulation then moves to the next year, e.g., $t+2$ or $t-2$. After completing all five steps of the Labour Market Module between 1984 and 2017, all individuals have complete employment and earnings information for previously unobserved years. Afterwards, we continue the simulation until 2032 to obtain complete biographical data up to age 60.

Figure 5 shows that our simulated data (dashed line) replicates the original SOEP data (solid line) well, particularly for Panel D (Full-time work), Panel E (Working hours) and Panel F (Earnings). Panel A (Labour Market Participation), Panel B (Employment) and Panel C

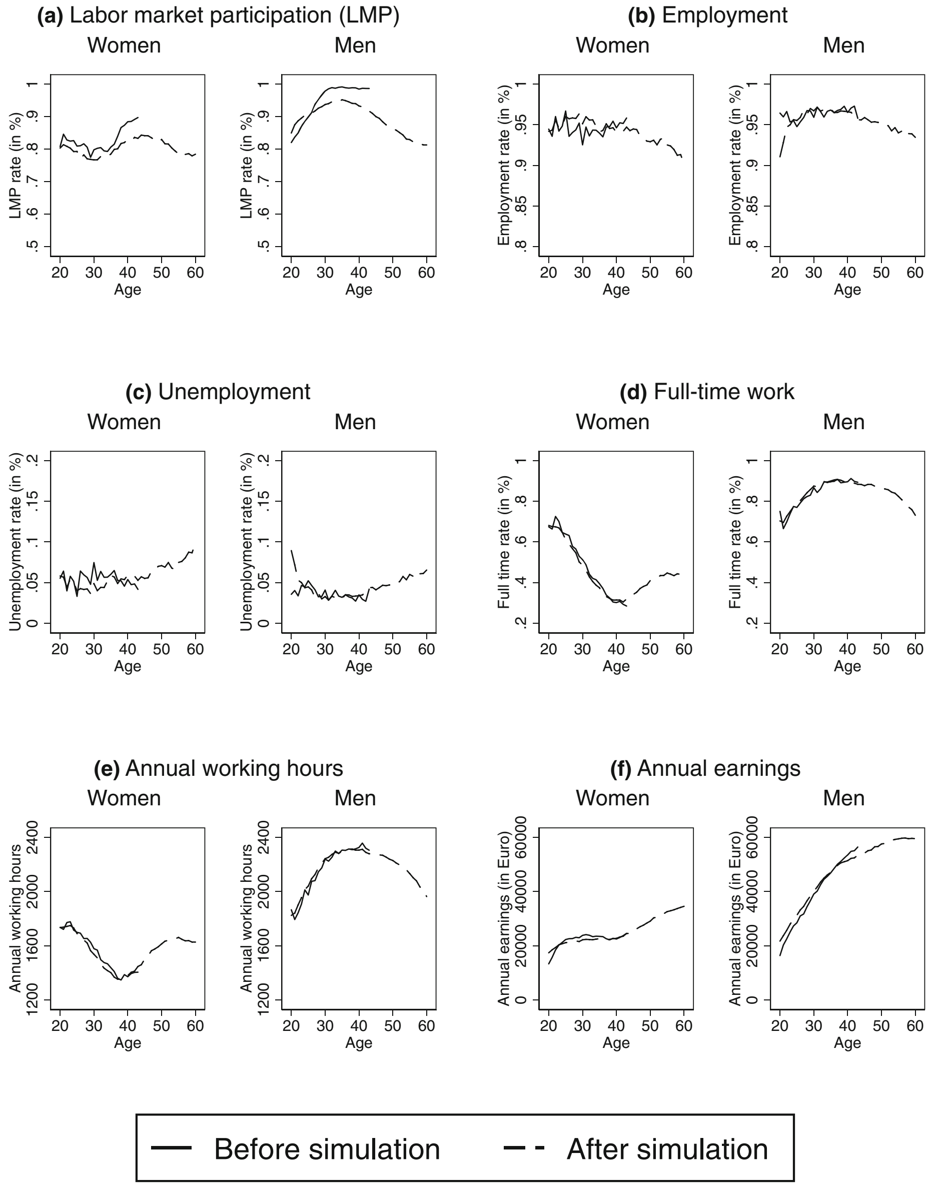

FIGURE 5 Labour market information before and after simulation. Only employed individuals are considered. Does not include values of zero annual earnings. Cohorts 1940–79, weighted sample. *Source*: Own calculations based on SOEP v35.

(Unemployment) show small deviations. Most of these differences occur in the beginning of the work life. These differences do not necessarily diminish the quality of our microsimulation for the following two reasons: First, our sample restriction to individuals observed at least once at age 30 or older leads to fewer observations in individuals' early twenties. As a result, our SOEP sample before the simulation is not very reliable for this age range due to a small sample size, and therefore comparisons may be misleading. Second, as depicted in Figure 5, earnings are on average relatively low at the beginning of an individual's work life and they increase over their careers. Consequently, earnings at young age only account for a small share of lifetime earnings. Generally, lifetime earnings estimates might be more reliable for individuals whom we observe in their mid-30s to mid-40s as existing studies provide evidence that during this period the (rank) correlation between cross-sectional and lifetime earnings is particularly high (see, e.g., Björklund, 1993; Bönke et al., 2015; Haider & Solon, 2006).[15]

After the completion of both modules of our dynamic microsimulation model, we obtain all relevant labour market and household information for birth cohorts 1964–72 from age 20 to 60 to proceed with our lifetime analysis.[16] Overall, the simulated data mirrors the data patterns before simulation and our simulation results are robust. Additional robustness checks based on a Monte Carlo simulation approach (Figure A5 and Figure A6) and the simulation of pseudo-missings (Figure A4) can be found in the Appendix.

3.2 | Lifetime analysis

Although we have already shown that women face lower hourly wages and annual earnings, and are less active on the labour market, the cross-sectional analysis only shows a snapshot of an individual's employment biography. A cross-sectional analysis does not reveal how these different factors add up over the life cycle. For a better understanding of when and how in life the gender gap develops, we investigate differences in accumulated earnings over the life cycle for birth cohorts 1964–72 using their complete biography data from age 20 to 60 obtained from our microsimulation. To analyse the accumulation of earnings over the work life we follow Bönke et al. (2015) and use the ‘up-to-age-X’ (UAX) concept. UAX earnings refer to accumulated price-adjusted (in 2015 Euro) gross annual earnings up to a certain age X. In line with the study by Bönke et al. (2015), we define lifetime earnings as UA60 earnings.

3.2.1 | Gender gap in lifetime earnings

To analyse the gender gap in lifetime earnings, we now focus on non-logarithmic incomes rather than logarithmic incomes as used in the Oaxaca–Blinder decomposition in Section 2.[17] Using logarithmic incomes would lead to the exclusion of zero earnings and, thus, periods of inactivity.[18] Because especially women accumulate periods of inactivity over life through motherhood and child rearing, those parts of their employment biographies without any earnings play a crucial role for the gender lifetime earnings gap and need to be included in this analysis.

The gender gap G in the labour market outcome variable L (here, hourly wages, annual earnings, UAX earnings) in percent for men m and women f at age x is now defined as:

$$G_x = \left[\left(\overline{L}_{m,x} - \overline{L}_{f,x} \right) / \overline{L}_{m,x} \right] \times 100. \tag{9}$$

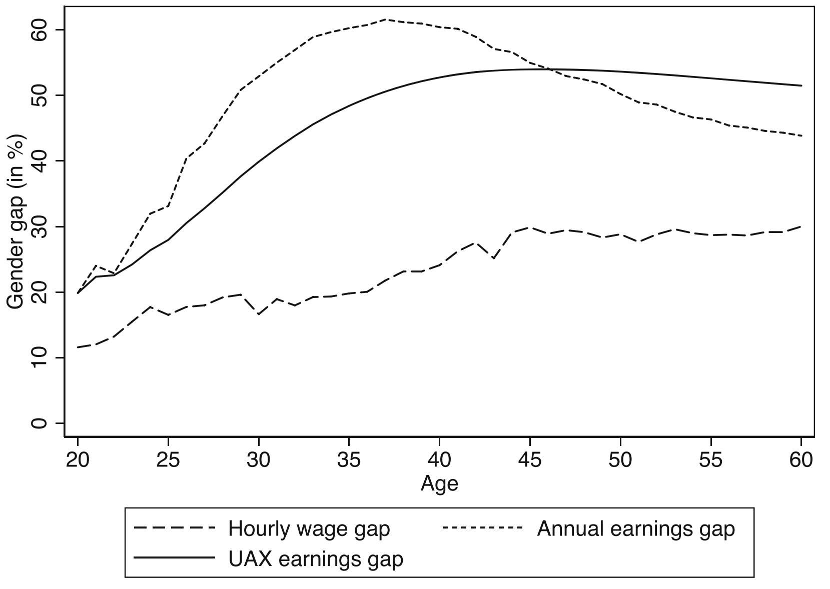

FIGURE 6 Gender gaps in wages, annual earnings and UAX earnings over the life cycle. Individuals with zero UAX earnings are included in the calculation. For annual earnings, employed and unemployed individuals are considered. For hourly wages, only employed individuals are considered. Cohorts 1964–72. *Source*: Own calculations based on SOEP v35.

Based on our new sample obtained from the microsimulation, Figure 6 shows the gender gaps in hourly wages, annual earnings and UAX earnings for ages 20–60 for birth cohorts 1964–72. As expected, despite the same trend, we see several differences when we compare the gender gaps in hourly wages and annual earnings using this microsimulation sample to our results based on the cross-sectional sample discussed in Section 2.

At early ages, the gender gap in hourly wages rather low but then increases steadily until retirement. However, we can observe differences in levels, which are driven by the more confined cohort restriction in our microsimulation sample and the varying definition of the gender gap (logarithmic vs. non-logarithmic income). Comparing the gender gaps in annual earnings reveals more pronounced differences between the cross-sectional and lifetime approach. First, the inversely U-shaped gender gap in annual earnings in Figure 6 is significantly larger than the gender gap shown in Figure 2. This difference is largely driven by the inclusion of inactive labour periods with zero earnings in this lifetime analysis, while we excluded those in our cross-sectional analysis in Section 2.[19] Including periods with zero earnings leads to a decline in women's average earnings, and thus to an increase in the gender gap. Naturally, this difference is especially pronounced in the years of family formation when women, on average, have longer spells of labour market inactivity due to child rearing. Second, in contrast to the gender gap we observe in our cross-sectional data, Figure 6 shows a pronounced decline of the gender gap in annual earnings between ages 40 and 60. Again, these different results are driven by the different composition of our two samples. While the cross-sectional sample includes all birth cohorts 1940–79, the lifetime sample is restricted to younger cohorts. Due to the higher labour market participation rates for women of younger cohorts, the gender gap in annual earnings declines again before retirement once we restrict our sample to younger cohorts, because more women reentered the labour market after times of inactivity during family formation.

Finally, the solid line in Figure 6 shows the gender gap in UAX earnings as the sum of the annual earnings up to age X. Ultimately, the UA60 earnings coincide with our definition of lifetime earnings. Hence, the higher the age X, the closer UAX earnings are to lifetime earnings. At the beginning of the work life, women earn on average 20 per cent less than men do. The difference in earnings accumulates over the life course and increases to a gender gap in UA40 earnings of 52.7 per cent. After that, the gap remains stable, which results in a gender gap in lifetime earnings of 51.5 per cent (UA60). At this point in life, women have earned on average around €732,000—slightly less than half of the average income that men were able to accumulate (€1,510,000).[20]

The evolution of the gender gap in UAX earnings is by construction following the shape of the annual earnings gender gap curve. UAX earnings are less volatile because the marginal effect of adding an additional year of annual earnings to the UAX earnings decreases with increasing age. Hence, the gender gaps in annual and UAX earnings both experience large growth until age 40, but when the gender gap in annual earnings declines again, the UAX gender gap remains at its high level.

The profound difference in lifetime earnings is largely the result of differences in the extensive and intensive margin of labour supply of women over their lives. One can discuss how labour supply is influenced by own decisions or forced by personal and social circumstances. Previous studies have shown a strong relationship between gender gaps in income and children (e.g., Adda et al., 2017; Angelov et al., 2016; Kleven & Landais, 2017). This can be partially explained by the close connection between women's labour market decisions and the number of children they have (Ejrnæs & Kunze, 2013; Kühhirt & Ludwig, 2012). In line with these studies, we also find that mothers face higher earning losses with every additional child, while fatherhood does not seem to affect men's earnings. Hence, observed earnings differences between childless women and men are smallest and grow wider with every additional child (see Figure A8). This observation also holds true when we analyse the evolution of UAX earnings by number of children (Figure A9).

Figure 7 shows the gender gap in hourly wages (Panel A), the gender gap in hours worked (Panel B), the gender gap in annual earnings (Panel C) and the gender gap in UAX earnings (Panel D) over the life cycle by number of children. In the beginning, the gender gap in hourly wages shows only small gender differences for men and women with and without children but widens over the life cycle. In Section 2, we have shown that this can be largely explained by the lesser work experience women with children gain over their life courses. The gender gap in annual earnings clearly differs by the number of children throughout the entire life cycle (see Figure 7, Panel C), exacerbating the gap in hourly wages associated with mother's lower intensive margin of work (see Figure 7, Panel B).

The gender gap in lifetime earnings also increases with the number of children. While childless men and women experience a gender gap of 17.3 per cent, the gap is significantly higher for men and women with three or more children (68.0 per cent at age 60). The significant widening of the gender gap between UA20 and UA35 earnings thereby coincides with the increase in the cross-sectional gender gaps in annual hours worked, and consequently annual earnings. These results are in line with existing studies finding evidence for motherhood penalties and fatherhood premiums (e.g., Budig & England, 2001; Killewald & Garca-Manglano, 2016; Killewald & Gough, 2013). Therefore, descriptive evidence clearly hints that motherhood might be a key driver of gender earnings inequality over the life cycle.

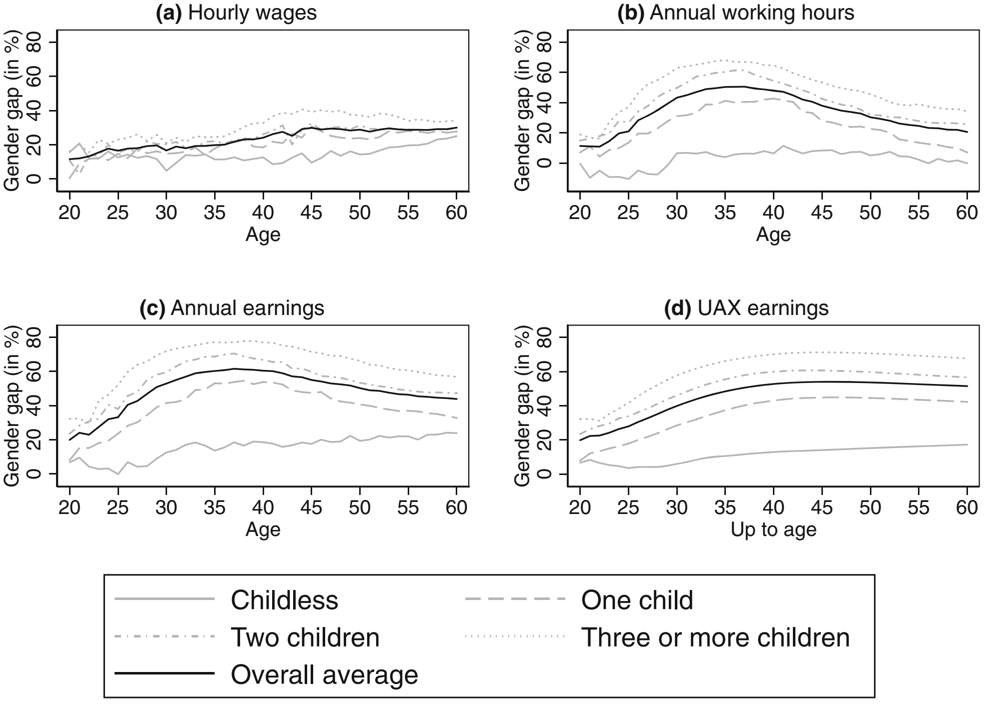

FIGURE 7 Gender gaps over the life cycle by children. Number of children refers to the total number at age 60. Gender gaps in accumulated earnings are earnings up to a given age. Individuals with zero annual and UAX earnings are included in the calculation. *Source*: Own calculations based on SOEP v35.

3.2.2 | Counterfactual analysis

In the last step, we want to estimate, which part of the observed gender gap in lifetime earnings can be associated with differences in the distribution of observable characteristics across gender and which part is associated with differences in returns to characteristics. To investigate this issue further, we will predict counterfactual lifetime earnings for women in the following two steps.[21]

First, we use a slightly modified version of the earnings regression results from our micro-simulation model, estimated for male M and female F individuals separately[22]:

$$\hat{y}_{s,t} = \hat{\beta}_{0,s} + +\hat{\beta}_{1,s} X_{s,t}, \quad s \in \{F, M\} \text{ and } t \in [1984, 2017]. \tag{10}$$

Second, we then set all annual earnings for men and women to missing and re-estimate women's counterfactual annual earnings $\hat{y}_f^C$ by using the coefficients obtained from the male regression model in the women's Mincer earnings regression:

$$\hat{y}_{f,t} = \hat{\beta}_{0,m} + \hat{\beta}_{1,m} X_{f,t}, \quad t \in [1984, 2017]. \tag{11}$$

For the estimation of male earnings, we use coefficients from the male regression model. Women's counterfactual annual earnings in year t then represent the salary women would have

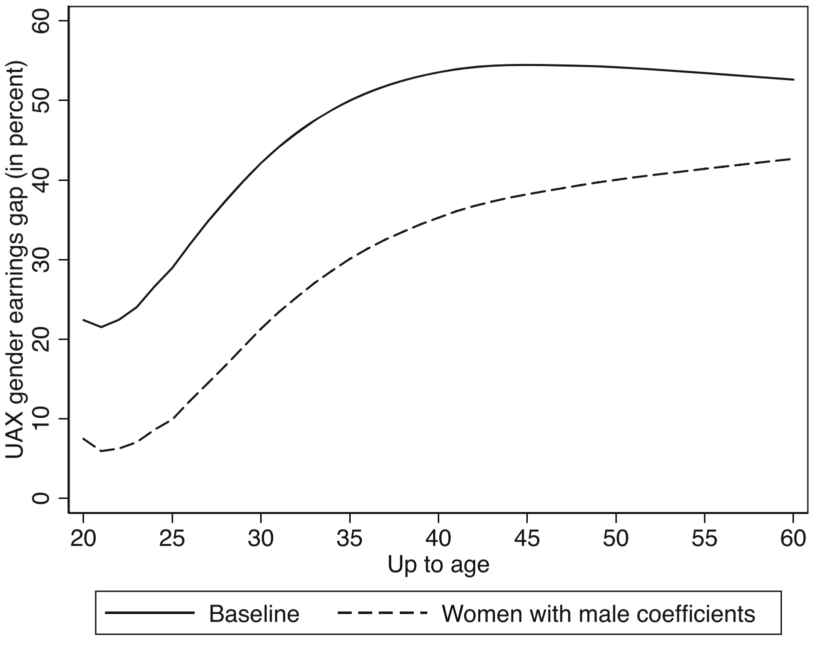

FIGURE 8 Counterfactual estimation of the lifetime earnings gap. Baseline and counterfactual gender gap in UAX earnings. Gender gaps in accumulated earnings are earnings up to a given age. Individuals with zero UAX earnings are included in the calculation. *Source*: Own calculations based on SOEP v35.

earned if their characteristics were rewarded the same as men's. Adding up the counterfactual annual earnings for each woman over the life course then yields women's counterfactual UAX earnings. Furthermore, we also calculate women's UAX earnings for a baseline scenario where we use the coefficients from the female earnings regression model (Equation 10). We then calculate the baseline UAX earnings gap for which men's and women's respective coefficients are used and the counterfactual UAX earnings gap where women's earnings are estimated using male coefficients.[23] The counterfactual gender UAX gap is therefore solely based on different characteristics for men and women and not by different returns to characteristics.

Figure 8 depicts both the baseline (solid line) and counterfactual (dashed line) gender gap in UAX earnings. The difference between those two concepts can be interpreted as the unexplained part of the gender gap in UAX earnings (adjusted gender gap). In the beginning of the work life, the difference between both gaps shown in Figure 8 is 14.9 pp. Therefore, in early years, approximately one-third of the gender gap in UAX earnings is due to a different allocation of characteristics and two-thirds is due to a different reward or payment of characteristics. The adjusted gender gap then increases to about 20.8 per cent for UA30 earnings and declines afterwards to 10.0 per cent for lifetime earnings (UA60). Thus, until the years of family formation, the unexplained difference between women's and men's pay grows, whereas it declines towards retirement. Overall, 80 per cent of the (re-estimated) baseline gender lifetime earnings gap at age 60 can be explained by a different distribution of labour market characteristics of men and women. Consequently, one-fifth of the (re-estimated) baseline gender lifetime earnings gap at age 60 can be explained by less favourable rewards for women's labour market characteristics, leading to an overall adjusted gender lifetime earnings gap of around 10 per cent.

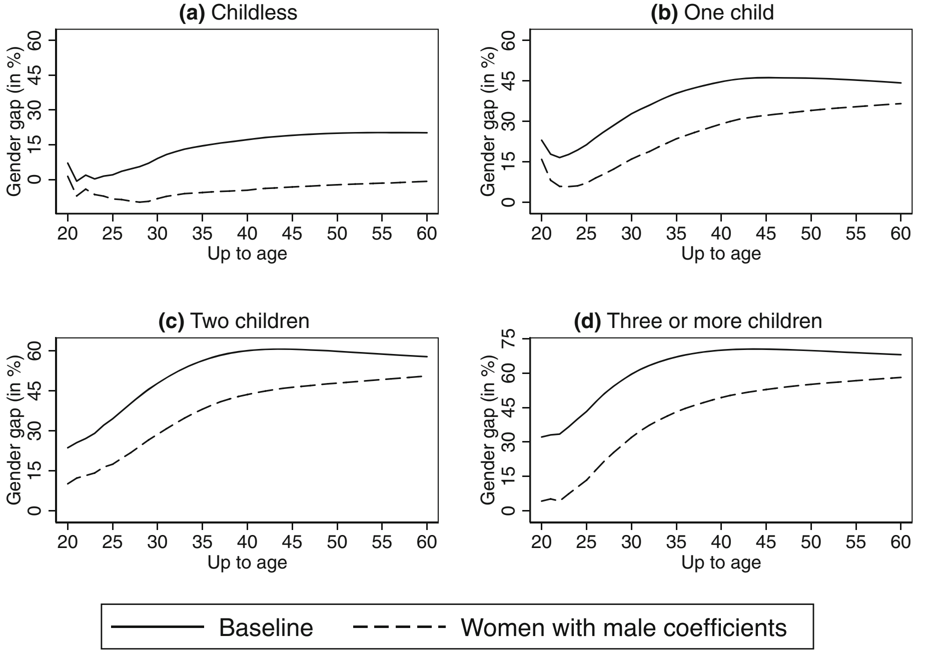

FIGURE 9 Counterfactual estimation of the lifetime earnings gap by number of children. Estimated and counterfactual gender gaps in UAX earnings. Gender gaps in accumulated earnings are earnings up to a given age. Individuals with zero UAX earnings are included in the calculation. *Source:* Own calculations based on SOEP v35.

The evolution of the adjusted gender gap indicates that rewards are least favourable for women in the first half of their work life.

Next, we want to investigate how parenthood influences the adjusted gender gap in lifetime earnings. Hence, Figure 9 compares the baseline and counterfactual gender gaps by the number of children. As already shown in Figure 7 (Panel D), the baseline gender gap in lifetime earnings is lowest for childless women and increases strongly with the number of children women have. But how much of the gender gap in lifetime earnings of women with and without children can be explained by a different distribution of characteristics, and what is the influence of the role of motherhood on the adjusted gender gap in lifetime earnings?

In stark contrast to the baseline UA60 gender gap, the adjusted UA60 gender gap only slightly differs between men and women with different numbers of children. The adjusted gender gap estimates are 7.7 per cent for one child and slightly higher for men and women with two children (7.3 per cent) and three or more children and women with three or more children (9.9 per cent). Hence, the large differences in the observed gender gaps of women with children are mainly driven by the different accumulation of characteristics rather than an additional unexplained penalty of motherhood. Our results in Section 2 and Figure 7 (Panel B) indicated that these differences might be mainly due to fewer working hours and less work experience, which women with children accumulate over their work life. However, our results suggest the opposite for childless individuals, for which the entire (re-estimated) gender gap in lifetime earnings (around 20 per cent) appears to be driven by differences in rewards.

4 | CONCLUSION

This paper underlines the importance of accounting for the lifetime dimension when analysing gender inequalities. First, our results show that cross-sectional gender differences are persistent over the work life. Comparing multiple dimensions of cross-sectional gender differences, we find that the gender gap in hourly wages is substantially smaller (less than half the size) than the gender gap in annual earnings. Using an Oaxaca–Blinder decomposition, we show that the gender gap in annual earnings can largely be explained by the extensive and intensive margin of labour, with women accumulating less work experience and working fewer hours.

We then applied a dynamic microsimulation model to obtain full lifetime earnings data including family background information. Using our simulated data, we observe a gender gap in lifetime earnings of 51.5 per cent. Furthermore, we show that the unadjusted gender gap in lifetime earnings increases with the number of children women have. While childless women face an average gender gap in lifetime earnings of 17.3 per cent, mothers with three or more children experience a gap of 68.0 per cent. Furthermore, we used the coefficients from the (modified) male earnings regression simulation model to estimate women's counterfactual earnings. As a result, all differences remaining were solely based on different characteristics of men and women and not by different returns to characteristics. The difference between the estimated baseline gender gap and the counterfactual gap then yielded the adjusted gender gap in lifetime earnings of 10 per cent. This means that women earn on average 10 per cent less than men over their lifetime due to different rewards for their observable characteristics in comparison to men. We find that in stark contrast to the observed gender gap in UAX earnings, the adjusted gender gap only differs slightly by the number of children mothers and fathers have. However, for childless men and women the adjusted gender gap is considerably higher and differences in rewards are the main driver of the gender gap in lifetime earnings.

The documented gender inequalities in lifetime earnings are high and therefore concerning for a variety of social and economic reasons. For example, fewer financial opportunities for women, and especially mothers, might create unhealthy dependency structures within households (see, e.g., Kalmuss & Straus, 1982). Furthermore, lower lifetime earnings result in significantly lower pensions and consequently a higher risk of poverty among elderly women (see, e.g., Fasang et al., 2013; Grabka et al., 2017). Against this background, it is of high importance to create the right conditions for women to have the opportunity and incentive to increase their labour market participation. One promising suggestion on how to increase work incentives for women in Germany is, e.g., a reform of *Ehegattensplitting*, the joint taxation of married couples or civil partners (see, e.g., Bach et al., 2017). Furthermore, the influential study by Olivetti and Petrongolo (2017) stresses the importance of the availability of childcare in this context. For Germany specifically, there is evidence that more extensive provision of adequate childcare would potentially positively influence mothers' labour market participation (e.g., Bauernschuster & Schlotter, 2015; Müller & Wrohlich, 2020). More broadly, fundamental changes in norms regarding the household division of labour are necessary as women still conduct the majority of housework and care-related tasks (see, e.g., Samtleben, 2019). In addition, employers should offer more flexible work arrangements in order to foster the compatibility of work and family. Indeed, recent studies indicate that such factors might have the potential to foster an increase in women's labour market participation as a considerable share of women who are currently working part-time have the (unrealized) desire to increase their working hours (e.g., Beckmannshagen & Schröder, 2022; Harnisch et al., 2018).

ACKNOWLEDGEMENTS

We thank Timm Bönke, Giacomo Corneo, Natalia Danzer, Richard Blundell, Carsten Schröder, Miriam Beblo, Daniel Waldenström and seminar participants at numerous venues for many valuable and helpful comments. Rick Glaubitz thanks the Hans-Böckler-Foundation for financial support. Open Access funding enabled and organized by Projekt DEAL.

ENDNOTES

[1] Past studies in this field focused on gender differences in human capital accumulation and discrimination as the main drivers of gender inequalities in labour markets. Altonji and Blank (1999) give an overview of the early literature in this field.

[2] See Goebel et al. (2019) for a detailed overview of the SOEP.

[3] A more detailed description of this methodological approach can be found in Section A.1 (Appendix).

[4] For comparability, we only control for variables that we can also use in our analysis of the lifetime gender gap in Section 3.

[5] Our pooled sample includes birth cohorts 1940–79. Therefore, we include cohort dummies into our estimation model. We do not find any consistent cohort effects in our analysis. Figure A1 also shows that gender gaps in labour market outcomes are generally stable over time in our sample of working women.

[6] A gender gap of 0.059 log points corresponds to a wage differential of $(e^{0.059}-1)\times 100 = 6.08$ per cent, while a gap of 0.378 log points corresponds to a wage differential of $(e^{0.378}-1)\times 100 = 45.94$ per cent.

[7] Tables A2 and A3 display the separate regression results for men and women which provide the basis for the difference in coefficients displayed in the Oaxaca–Blinder regression.

[8] The constant of the coefficient part also includes the effects of gender differences in unobserved predictors (Jann, 2008), e.g., different occupational choices or differences in employers.

[9] Please note that since this subsection focuses on the intensive margin of work, we now include the total hours worked per year for this model in contrast to the binary variable (part-time/full-time) used when we were analysing the gender gap in hourly wages. Consequently, this leads to an even more significant endowment part for the analysis of annual earnings as the total number of work hours is a key driver in the earnings difference across gender.

[10] It is crucial to note that our model does not control for endogenous choice. Hence, we do not differentiate whether women choose to work fewer hours or if they have trouble finding adequate employment. See, e.g., Harnisch et al. (2018) and Beckmannshagen and Schröder (2022) for studies on working hours mismatches in Germany.

[11] In this case we assume that the children stay with the mother. Empirical evidence by the Statistisches Bundesamt (2018) supports this assumption: The share of single fathers in the period since 1997 is only 10–13 per cent.

[12] For a few married individuals in our data, we cannot observe partner information since the partner did not participate in the survey, e.g., because they refused. In those cases, we also run the Partner Module as a preparation step before starting the Family Module.

[13] For this estimation strategy, we are only able to include individuals that have at least two observation years in the SOEP. Including additional lags would result in a reduced sample size since it would impose stricter sample restrictions (surveyed for at least 3 years in the SOEP).

[14] Again, see Table A7 for more detailed information.

[15] It is important to note that these studies only focus on men. For women, other patterns could emerge over the life cycle.

[16] Our sample after the microsimulation is significantly different from our original SOEP sample. Therefore, we cannot use the longitudinal weights initially provided by the SOEP. To maintain representativeness, we therefore use census data (Mikrozensus) to reweight our sample with regard to cohort, age, family and labour

market information. The Mikrozensus is considered highly representative for Germany, covering about 1 per cent of the entire German population through mandatory participation.

17 As stated in Section 2, the Oaxaca–Blinder decomposition is based on an OLS regression model using log hourly wage and log annual earnings.

18 The inverse hyperbolic sign (ihs) transformation represents an alternative concept. In contrast to the logarithmic transformation, it is also defined for negative and zero values (see, e.g., Burbidge et al., 1988; Pence, 2006). Due to these advantages, it is primarily used in the literature on wealth distributions (e.g., Grabka et al., 2015; Pence, 2006; Sierminska et al., 2018). However, we refrain from using this transformation as it is not easily interpretable and not a very commonly used concept in the literature on gender earnings gaps.

19 See Figure A7 for a direct comparison of the gender gap in annual earnings when including or excluding individuals with zero earnings.

20 Compare Figures A8 and A9 for the distribution of annual earnings and UAX earnings by men and women over the work life.

21 Here, we refrain from using an Oaxaca–Blinder approach, which we applied in our cross-sectional analysis in Section 2. The reason for this is that we want to avoid decomposing our complete life-cycle data (which itself is the result of regression-based dynamic microsimulation) with a regression-based decomposition approach.

22 In contrast to the earnings regression model used for our main analysis (Equation 8), here we refrain from including lagged earnings information since they include the lagged unexplained component. Including lagged earnings would, therefore, potentially compromise our analysis and lead to biased results of our counterfactual decomposition approach.

23 Note that due to the fact that we re-estimate all earnings for men and women (also the ones originally observed in the SOEP) and modify the earnings regression, our baseline results here vary slightly from our main results presented in Section 3.2.1. Here, the baseline UA60 earnings gap is 52.6 per cent while in our main analysis we find an UA60 earnings gap of 51.5 per cent.

REFERENCES

Adda, J., Dustmann, C. & Stevens, K. (2017) The career costs of children. *Journal of Political Economy*, 125(2), 293–337.

Altonji, J.G. & Blank, R.M. (1999) Race and gender in the labor market. In: Ashenfelter, O.C. & Card, D. (Eds.) *Handbook of labor economics*, Vol. 3. Amsterdam: Elsevier, pp. 3143–3259.

Anderson, D.J., Binder, M. & Krause, K. (2002) The motherhood wage penalty: which mothers pay it and why? *American Economic Review*, 92(2), 354–358.

Angelov, N., Johansson, P. & Lindahl, E. (2016) Parenthood and the gender gap in pay. *Journal of Labor Economics*, 34(3), 545–579.

Autorengruppe Bildungsberichterstattung. (2018) *Bildung in Deutschland 2018: ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung*. Bielefeld: wbv Publikation.

Bach, S., Fischer, B., Haan, P. & Wrohlich, K. (2017) Ehegattenbesteuerung: Individualbesteuerung mit übertragbarem Grundfreibetrag schafft fiskalische Spielräume. *DIW Weekly Report*, 84(13), 247–255.

Bauernschuster, S. & Schlotter, M. (2015) Public child care and mothers' labor supply—Evidence from two quasi-experiments. *Journal of Public Economics*, 123, 1–16.

Beblo, M. & Wolf, E. (2002) Die Folgekosten von Erwerbsunterbrechungen. *Vierteljahrshefte Zur Wirtschaftsforschung*, 71(1), 83–94.

Beckmannshagen, M. & Schröder, C. (2022) Earnings inequality and working hours mismatch. *Labour Economics*, 76, 102184.

Bertrand, M., Goldin, C. & Katz, L.F. (2010) Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3), 228–255.

Bhuller, M., Mogstad, M. & Salvanes, K.G. (2011) *Life-cycle bias and the returns to schooling in current and lifetime earnings* (Dept. of Economics Discussion Paper 4/2011). NHH Dept. of Economics.

Björklund, A. (1993) A comparison between actual distributions of annual and lifetime income: Sweden 1951–89. *Review of Income and Wealth*, 39(4), 377–386.

Blau, F.D. & Kahn, L.M. (2017) The gender wage gap: extent, trends, and explanations. *Journal of Economic Literature*, 55(3), 789–865.

Blinder, A. (1973) Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8, 436–455.

Boll, C., Jahn, M. & Lagemann, A. (2017) The gender lifetime earnings gap—Exploring gendered pay from the life course perspective. *Journal of Income Distribution*, 26(1), 1–53.

Bonin, H., Reuss, K. & Stichnoth, H. (2015) *Life-cycle incidence of family policy measures in Germany: evidence from a dynamic microsimulation model* (SOEPpapers 770).

Bönke, T., Corneo, G. & Lüthen, H. (2015) Lifetime earnings inequality in Germany. *Journal of Labor Economics*, 33(1), 171–208.

Brenner, J. (2010) Life-cycle variations in the association between current and lifetime earnings: evidence for German natives and guest workers. *Labour Economics*, 17(2), 392–406.

Brown, J.R., Coronado, J.L. & Fullerton, D. (2009) Is social security part of the social safety net? *Tax Policy and the Economy*, 23(1), 37–72.

Brunello, G., Weber, G. & Weiss, C.T. (2017) Books are forever: Early life conditions, education and lifetime earnings in Europe. *The Economic Journal*, 127(600), 271–296.

Budig, M.J. & England, P. (2001) The wage penalty for motherhood. *American Sociological Review*, 66(2), 204–225.

Burbidge, J.B., Magee, L. & Robb, A.L. (1988) Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83(401), 123–127.

Bütikofer, A., Jensen, S. & Salvanes, K.G. (2018) The role of parenthood on the gender gap among top earners. *European Economic Review*, 109, 103–123.

Corneo, G. (2015) Income inequality from a lifetime perspective. *Empirica*, 42, 225–239.

Coronado, J.L., Fullerton, D. & Glass, T. (2011) The progressivity of social security. *The B.E. Journal of Economic Analysis & Policy*, 11(1), 1–43.

Correll, S., Benard, S. & Paik, I. (2007) Getting a job: is there a motherhood penalty? *American Journal of Sociology*, 112(5), 1297–1339.

Ejrnæs, M. & Kunze, A. (2013) Work and wage dynamics around childbirth. *The Scandinavian Journal of Economics*, 115(3), 856–877.

Fasang, A.E., Aisenbrey, S. & Schömann, K. (2013) Women's retirement income in Germany and Britain. *European Sociological Review*, 29(5), 968–980.

Gallen, Y., Lesner, R.V. & Vejlin, R. (2019) The labor market gender gap in Denmark: sorting out the past 30 years. *Labour Economics*, 56, 58–67.

Gangl, M. & Ziefle, A. (2009) Motherhood, labor force behavior, and women's careers: an empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States. *Demography*, 46(2), 341–369.

Geyer, J. & Steiner, V. (2014) Future public pensions and changing employment patterns across birth cohorts. *Journal of Pension Economics & Finance*, 13(2), 172–209.

Goebel, J., Grabka, M.M., Liebig, S., Kroh, M., Richter, D., Schröder, C. et al. (2019) The German socio-economic panel (SOEP). *Jahrbücher für Nationalökonomie Und Statistik*, 239(2), 345–360.

Goldin, C. (2014) A grand gender convergence: its last chapter. *American Economic Review*, 104(4), 1091–1119.

Grabka, M.M. & Goebel, J. (2017) Realeinkommen sind von 1991 bis 2014 im Durchschnitt gestiegen: Erste Anzeichen für wieder zunehmende Einkommensungleichheit. *DIW Weekly Report*, 84(4), 71–82.

Grabka, M.M., Jotzo, B., Rasner, A. & Westermeier, C. (2017) Der gender pension gap verstärkt die Einkommensungleichheit von Männern und Frauen im Rentenalter. *DIW Weekly Report*, 84(5), 87–96.

Grabka, M.M., Marcus, J. & Sierminska, E. (2015) Wealth distribution within couples. *Review of Economics of the Household*, 13, 459–486.

Guvenen, F., Kaplan, G. & Song, J. (2021) The glass ceiling and the paper floor: changing gender composition of top earners since the 1980s. *NBER Macroeconomics Annual*, 35(1), 309–373.

Guvenen, F., Kaplan, G., Song, J. & Weidner, J. (2022) Lifetime earnings in the United States over six decades. *American Economic Journal: Applied Economics*, 14(4), 446–479.

Haider, S. & Solon, G. (2006) Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, 96(4), 1308–1320.

Hänisch, C. & Klos, J. (2016) *Long-run effects of career interruptions: a micro-simulation study* (Discussion Paper). Universität Freiburg.

Harkness, S. & Waldfogel, J. (2003) The family gap in pay: Evidence from seven industrialized countries. *Worker Well-Being and Public Policy*, 22, 369–413.

Harnisch, M., Müller, K.-U. & Neumann, M. (2018) Teilzeitbeschäftigte würden gerne mehr Stunden arbeiten, Vollzeitbeschäftigte Lieber Reduzieren. *DIW Weekly Report*, 85(38), 837–846.

Jann, B. (2008) The blinder–oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4), 453–479.

Juhn, C. & McCue, K. (2017) Specialization then and now: marriage, children, and the gender earnings gap across cohorts. *Journal of Economic Perspectives*, 31(1), 183–204.

Kalmuss, D.S. & Straus, M.A. (1982) Wife's marital dependency and wife abuse. *Journal of Marriage and the Family*, 44(2), 277–286.

Killewald, A. & Garca-Manglano, J. (2016) Tethered lives: a couple-based perspective on the consequences of parenthood for time use, occupation, and wages. *Social Science Research*, 60, 266–282.

Killewald, A. & Gough, M. (2013) Does specialization explain marriage penal—ties and premiums? *American Sociological Review*, 78(3), 477–502.

Kleven, H. & Landais, C. (2017) Gender inequality and economic development: Fertility, education and norms. *Economica*, 84(334), 180–209.

Kleven, H., Landais, C., Posch, J., Steinhauer, A. & Zweimüller, J. (2019) Child penalties across countries: evidence and explanations. *AEA Papers and Proceedings*, 109, 122–126.

Kleven, H., Landais, C. & Søgaard, J.E. (2021) Does biology drive child penalties? Evidence from biological and adoptive families. *American Economic Review: Insights*, 3(2), 183–198.

Kleven, H.J., Landais, C., Posch, J., Steinhauer, A. & Zweimüller, J. (2020) *Do family policies reduce gender inequality? Evidence from 60 years of policy experimentation* (NBER working paper 28082).

Kühhirt, M. & Ludwig, V. (2012) Domestic work and the wage penalty for motherhood in West Germany. *Journal of Marriage and Family*, 74(1), 186–200.

Levell, P. & Shaw, J. (2016) Constructing full adult life-cycles from short panels. *International Journal of Microsimulation*, 9(2), 5–40.

Li, J. & O'Donoghue, C. (2013) A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation*, 6(2), 3–55.

Mahalanobis, P. (1936) Mahalanobis distance. *Proceedings National Institute of Science of India*, 49, 234–256.

Müller, K.-U. & Wrohlich, K. (2020) Does subsidized care for toddlers increase maternal labor supply? Evidence from a large-scale expansion of early childcare. *Labour Economics*, 62, 101776.

Neufeld, C. (2000) Alignment and variance reduction in DYNACAN. *Contributions to Economic Analysis*, 247, 361–382.

Oaxaca, R. (1973) Male-female wage differentials in urban labor markets. *International Economic Review*, 14, 693–709.

Olivetti, C. & Petrongolo, B. (2017) The economic consequences of family policies: lessons from a century of legislation in high-income countries. *Journal of Economic Perspectives*, 31(1), 205–230.

Pence, K.M. (2006) The role of wealth transformations: an application to estimating the effect of tax incentives on saving. *The B.E. Journal of Economic Analysis & Policy*, 5(1), 1–26.

Plümper, T. & Troeger, V.E. (2007) Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis*, 15(2), 124–139.

Samtleben, C. (2019) Also on Sundays, women perform most of the housework and child care. *DIW Weekly Report*, 9(10), 87–92.

Sierminska, E., Piazzalunga, D. & Grabka, M.M. (2018) *Transitioning towards more equality? Wealth gender differences and the changing role of explanatory factors over time* (LISER Working Paper Series 2018-18).

Statistisches Bundesamt. (2017) *Verdienste im Überblick*. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2018) *Alleinerziehende in Deutschland*. Wiesbaden: Statistisches Bundesamt.

Tamborini, C.R., Kim, C. & Sakamoto, A. (2015) Education and lifetime earnings in the United States. *Demography*, 52(4), 1383–1407.

Tyrowicz, J., van der Velde, L. & van Staveren, I. (2018) Does age exacerbate the gender-wage gap? New method and evidence from Germany, 1984–2014. *Feminist Economics*, 24(4), 108–130.

Waldfogel, J. (1998) Understanding the “family gap” in pay for women with children. *Journal of Economic Perspectives*, 12(1), 137–156.

Westermeier, C., Rasner, A. & Grabka, M.M. (2012) *The prospects of the baby boomers: methodological challenges in projecting the lives of an aging cohort* (SOEPpapers 440).

Zucchelli, E., Jones, A.M. & Rice, N. (2012) The evaluation of health policies through dynamic microsimulation methods. *International Journal of Microsimulation*, 5(1), 2–20.

How to cite this article: Glaubitz, R., Harnack-Eber, A. & Wetter, M. (2024) The gender gap in lifetime earnings: A microsimulation approach. *LABOUR*, 38(4), 425–474. Available from: https://doi.org/10.1111/labr.12274

APPENDIX A

A.1 | Oaxaca Blinder decomposition

The Oaxaca Blinder decomposition was simultaneously introduced by Oaxaca (1973) and Blinder (1973) and divides the gender differential in labor market outcomes (here: hourly wage or annual earnings) into an *endowment part* and a *coefficient part*. The endowment part of the gender differential accounts for the part of the gap which can be attributed to differences in the allocation of characteristics (e.g., working hours, highest level of education) between men and women. In contrast, the coefficient part captures the gender differences in labor market returns to characteristics, and therefore in their coefficients. In other words, it states the gender differences of what the labor market is willing to pay for the same characteristics. This part is also called the raw or adjusted gender wage/earnings differential. This adjusted gap, however, also contains the effects of gender differences in unobserved predictors (Jann, 2008). The Oaxaca-Blinder decomposition approach enables us to analyze whether the gender gap in wages/earnings is mainly driven by the different distributions of productivity characteristics or by different rewards for these characteristics by gender.

The gender gap G_x is defined as the difference between the means of the labor market outcomes L at age x of men m and women f:

$$G_x = E(L_{mx}) - E\left(L_{fx}\right) \tag{12}$$

L_s for either sex (s) is based on the linear model

$$L_{sx} = Z'_{sx}\beta_{sx} + \epsilon_{sx}, \quad E(\epsilon_{sx}) = 0, \; S \in \{f, m\}, \tag{13}$$

where the vector Z includes all relevant characteristics, β is the estimation vector and ϵ is the error term. Inserting Equation (2.13) into Equation (2.12), the earnings differential can also be written as:

$$G_x = E(Z_{mx})'\beta_{mx} - E\left(Z_{fx}\right)'\beta_{fx}. \tag{14}$$

For the decomposition of the results, a non-discriminatory coefficient vector is needed, called β^*. Following Neumark (1988), the vector is determined as a pooled regression over both sexes. The gender gap can then be rewritten as:

$$G_x = \underbrace{\left[E(Z_{mx}) - E\left(Z_{fx}\right)\right]'\beta_x^*}_{\text{Endowment part}} + \underbrace{\left[E(Z_{mx})'(\beta_{mx} - \beta_x^*) + E\left(Z_{fx}\right)'\left(\beta_x^* - \beta_{fx}\right)\right]}_{\text{Coefficient part}} \tag{15}$$

where the first part of Equation (2.15) is the endowment part and the second part is the coefficient component of the gender gap in the labor market outcome.

A.2 | Robustness: Microsimulation

Pseudo missings To test the robustness of our simulation model further, we use the concept of pseudo missings. To that end, we set truly observed information for some part of the sample missing (pseudo missings) and predict their now missing observations again by using our dynamic microsimulation and the regression coefficients previously obtained. As we need a starting point of at least two observations for our models due to the lagged terms, we use the first two truly observed years for everyone before starting to create pseudo missings. Figure A4 shows the differences between the simulated pseudo missings (dashed line) and the truly observed information (solid line) for labor force status, employment status, annual working hours and annual earnings. In most graphs, the level of accuracy of the model is so high that it is hard to even tell the solid and dashed line apart. For labor market status, the model predicts 99.9% of all pseudo missings correctly. And even for employment status, where there appear to be bigger differences between pseudo missing and observations at a first glance, overall 97.7% of all cases are simulated correctly. These results further support the robustness of our simulation model.

Monte Carlo simulation Another way to validate the robustness of our dynamic microsimulation model is to make use of the underlying random process described in Subsection 3.1.1. We implement a Monte Carlo simulation approach by simulating each individual's employment biographies 100 times. By doing so, due to the underlying random process determining transitions in labor market outcome variables between $t-1$ and t, we simulate up to 100 different employment biographies for each individual. However, due to limited computational capacities we only simulate the employment variables (labor market status, employment status, full-time/part-time work, annual working hours and annual earnings) and keep the family information (number of children and marital status) constant for each of the 100 iterations. In the next step, we calculate lifetime earnings for each of the 100 simulated career paths per individual and compute the average lifetime earnings and the resulting UAX earnings gender gap in the population for each of the 100 runs. By deriving the 95% confidence intervals we can analyze whether average lifetime earnings vary significantly for different underlying random processes or whether they are robust. The results are presented in Figures A5 and A6. Figure A5 shows that lifetime earnings by cohorts are very robust. However, lifetime earnings of women vary more strongly than men's. Figure A6 provides evidence for a very narrow 95% confidence interval for the gender gap in UAX earnings. Consequently, the results of the Monte Carlo simulation confirm the high robustness of our simulation outcomes.

TABLE A1 Descriptive statistics—means by age

	Age								
Men	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
Annual earnings	15748.13	27727.89	37925.13	45217.80	51615.70	54204.14	54747.55	53969.63	51535.02
	(10972.17)	(13306.99)	(18571.57)	(24095.07)	(31182.61)	(38951.27)	(35380.01)	(33505.90)	(50496.35)
Hourly wage	9.37	15.13	18.12	20.72	23.06	23.95	24.24	25.83	26.13
	(7.72)	(18.11)	(20.76)	(16.32)	(16.25)	(17.91)	(14.47)	(31.96)	(28.33)
Hours worked per week	34.55	38.29	42.81	43.49	44.39	44.10	43.50	42.65	39.34
	(13.42)	(14.48)	(12.47)	(11.38)	(11.01)	(10.61)	(11.17)	(12.13)	(14.38)
Years in full-time work	1.20	4.75	8.54	12.97	17.71	22.58	27.43	32.69	37.32
	(1.28)	(2.60)	(3.77)	(4.37)	(4.85)	(5.31)	(5.71)	(5.81)	(5.74)
Years in part-time work	0.14	0.33	0.55	0.56	0.61	0.65	0.75	0.68	1.09
	(0.47)	(0.98)	(1.56)	(1.64)	(1.93)	(2.10)	(2.44)	(2.46)	(2.90)
Years in unemployment	0.13	0.31	0.39	0.43	0.45	0.47	0.51	0.49	0.45
	(0.38)	(0.70)	(0.98)	(1.21)	(1.37)	(1.64)	(1.85)	(1.78)	(1.65)
Years of education	8.97	10.61	11.84	12.44	12.62	12.65	12.67	12.57	12.73
	(3.86)	(3.03)	(3.25)	(3.17)	(3.03)	(2.96)	(2.92)	(2.84)	(2.92)
	Age								
Women	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
Annual earnings	12773.34	21115.69	22975.43	21925.18	22944.75	24975.61	26705.30	26475.69	24659.61
	(8683.31)	(12332.56)	(16720.75)	(19512.82)	(18626.65)	(20497.00)	(21713.56)	(25559.13)	(21236.77)
Hourly wage	7.97	12.87	15.19	15.63	16.23	16.23	16.82	16.54	17.48
	(6.58)	(9.59)	(12.19)	(13.18)	(12.02)	(10.88)	(12.49)	(13.18)	(14.99)
Hours worked per week	31.75	32.28	29.91	26.68	27.38	28.87	30.09	29.42	26.97
	(13.02)	(14.34)	(15.60)	(15.26)	(14.29)	(13.99)	(13.98)	(13.64)	(14.49)

(Continues)

TABLE A1 (Continued)

	Age								
Women	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
Years in full-time work	1.20	4.36	6.73	8.04	9.63	11.61	14.00	16.70	19.65
	(1.23)	(2.71)	(4.12)	(5.23)	(6.46)	(7.99)	(9.69)	(11.75)	(13.99)
Years in part-time work	0.21	0.82	1.91	3.81	5.69	7.60	9.45	11.66	13.32
	(0.55)	(1.61)	(2.60)	(3.76)	(4.87)	(6.16)	(7.75)	(9.77)	(11.79)
Years in unemployment	0.17	0.28	0.40	0.50	0.56	0.58	0.64	0.71	0.57
	(0.40)	(0.73)	(0.95)	(1.19)	(1.52)	(1.60)	(1.77)	(1.99)	(1.86)
Years of education	9.17	11.17	12.07	12.42	12.48	12.39	12.34	12.11	12.02
	(3.91)	(2.98)	(3.31)	(3.00)	(2.89)	(2.91)	(2.78)	(2.59)	(2.71)

Notes: Only employed individuals with hourly wages and annual earnings greater than zero were included. Cohorts 1940–1979, weighted sample. Annual earnings and hourly wages are price-adjusted and presented in 2015 Euro. Standard errors in parentheses.
Source: Own calculations based on SOEP v35.

TABLE A2 Regression results for hourly wages - women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
One child	−0.019	−0.061	−0.159***	−0.023	−0.003	0.075**	0.068*	0.104**	−0.058
	(0.158)	(0.055)	(0.046)	(0.036)	(0.038)	(0.034)	(0.040)	(0.048)	(0.070)
Two children	−0.461	−0.234**	−0.154***	−0.058	−0.000	0.073**	0.138***	0.088*	−0.014
	(0.501)	(0.107)	(0.058)	(0.041)	(0.039)	(0.036)	(0.040)	(0.048)	(0.071)
3 or more children		−0.171	−0.111	−0.167***	−0.030	0.036	0.129***	0.103*	−0.051
		(0.195)	(0.093)	(0.055)	(0.050)	(0.043)	(0.048)	(0.058)	(0.083)
Married	0.033	−0.038	0.054	0.008	0.004	0.068**	−0.052*	−0.031	0.055
	(0.100)	(0.036)	(0.034)	(0.030)	(0.029)	(0.027)	(0.029)	(0.035)	(0.048)
Years FT	0.445***	0.055**	0.059***	0.028***	0.035***	0.026***	0.027***	0.014**	0.030***
	(0.061)	(0.022)	(0.014)	(0.009)	(0.007)	(0.005)	(0.005)	(0.006)	(0.007)
Years FT (sq)	−0.054***	−0.003	−0.002***	0.000	−0.001*	−0.000	−0.000	−0.000	−0.000**
	(0.016)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years PT	−0.023	−0.039	−0.028	−0.019*	−0.013	−0.020***	−0.002	−0.001	−0.011
	(0.103)	(0.029)	(0.017)	(0.011)	(0.008)	(0.006)	(0.006)	(0.006)	(0.008)
Years PT (sq)	0.012	0.001	0.003*	0.002**	0.001**	0.001***	0.000*	0.000	0.000**
	(0.028)	(0.004)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years UE	−0.689***	−0.101**	−0.174***	0.013	−0.062***	−0.096***	−0.076***	−0.058***	−0.044*
	(0.209)	(0.043)	(0.039)	(0.021)	(0.018)	(0.018)	(0.018)	(0.017)	(0.024)
Years UE (sq)	0.236*	−0.000	0.034***	−0.005**	0.004**	0.007***	0.004**	0.001	0.001
	(0.140)	(0.006)	(0.009)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Part-time	0.244***	0.272***	0.110***	0.182***	0.161***	0.102***	0.023	−0.068*	0.069
	(0.081)	(0.049)	(0.042)	(0.032)	(0.031)	(0.027)	(0.031)	(0.039)	(0.054)

(Continues)

TABLE A2 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
Education	−0.068***	0.007	−0.033**	−0.063***	0.018	0.008	−0.023	0.012	−0.052
	(0.024)	(0.017)	(0.015)	(0.015)	(0.023)	(0.015)	(0.021)	(0.029)	(0.036)
Education (sq)	0.004*	0.001	0.003***	0.006***	0.002***	0.003***	0.004***	0.002**	0.005***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	2.241***	1.714***	1.233***	1.700***	1.825***	1.508***	1.033***	1.614***	1.898***
	(0.770)	(0.266)	(0.211)	(0.222)	(0.199)	(0.161)	(0.313)	(0.373)	(0.463)
Obs.	382	882	1307	1859	2493	2653	2043	1320	778
R-squared	0.323	0.127	0.187	0.240	0.192	0.219	0.205	0.213	0.248
Cohort-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Standard errors in parentheses; The stars refer to the following significance level:
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
Source: Own calculations based on SOEP v35.

TABLE A3 Regression results for hourly wages—men

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
One child	0.160	0.089	−0.017	−0.017	0.007	0.029	−0.043	0.026	0.053
	(0.349)	(0.055)	(0.031)	(0.024)	(0.022)	(0.024)	(0.029)	(0.040)	(0.053)
Two children	−0.952	0.134*	0.065*	0.069***	0.033	0.103***	−0.000	0.034	−0.010
	(0.751)	(0.079)	(0.035)	(0.024)	(0.022)	(0.023)	(0.027)	(0.037)	(0.049)
3 or more children	−0.006	0.139*	−0.024	0.015	0.050*	0.049*	−0.013	0.075	0.203***
	(0.173)	(0.075)	(0.045)	(0.033)	(0.028)	(0.029)	(0.035)	(0.049)	(0.067)
Married	−0.026	0.015	0.164***	0.085***	0.101***	0.069***	0.084***	0.020	0.105**
	(0.166)	(0.041)	(0.027)	(0.023)	(0.021)	(0.023)	(0.026)	(0.035)	(0.048)
Years FT	0.737***	0.164***	0.105***	0.058***	0.053***	0.045***	0.032***	0.040***	−0.030
	(0.061)	(0.023)	(0.013)	(0.008)	(0.007)	(0.008)	(0.009)	(0.012)	(0.036)
Years FT (sq)	−0.108***	−0.012***	−0.006***	−0.002***	−0.001***	−0.001***	−0.001***	−0.001**	0.001
	(0.015)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Years PT	0.225	−0.197***	−0.074***	−0.020*	−0.056***	−0.038***	−0.069***	−0.070***	−0.057***
	(0.181)	(0.036)	(0.018)	(0.012)	(0.010)	(0.011)	(0.010)	(0.014)	(0.017)
Years PT (sq)	−0.079	0.021***	0.006***	0.001	0.003***	0.003***	0.003***	0.003***	0.001*
	(0.071)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years UE	−0.279	−0.108**	−0.178***	−0.113***	−0.117***	−0.105***	−0.094***	−0.069***	−0.094***
	(0.195)	(0.050)	(0.029)	(0.015)	(0.013)	(0.012)	(0.014)	(0.019)	(0.036)
Years UE (sq)	0.099	−0.005	0.023***	0.004***	0.006***	0.006***	0.004***	0.002**	0.006
	(0.101)	(0.014)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Part−time	0.271***	0.422***	0.257***	0.173***	0.336***	0.251***	0.189***	0.389***	0.201***
	(0.075)	(0.054)	(0.043)	(0.034)	(0.031)	(0.031)	(0.039)	(0.045)	(0.054)

(Continues)

TABLE A3 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
Education	−0.031	−0.038***	−0.071***	−0.039***	−0.051***	0.040*	0.060***	0.002	0.020
	(0.030)	(0.013)	(0.010)	(0.010)	(0.013)	(0.022)	(0.023)	(0.055)	(0.067)
Education (sq)	0.003	0.003***	0.005***	0.003***	0.004***	0.001*	0.000	0.002	0.002
	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)
Constant	1.996***	1.591***	2.263***	2.331***	2.519***	1.703***	1.673***	1.875***	1.318
	(0.504)	(0.163)	(0.183)	(0.172)	(0.173)	(0.187)	(0.231)	(0.437)	(0.926)
Obs.	383	900	1746	2464	2863	2939	2261	1546	980
R-squared	0.449	0.231	0.185	0.229	0.277	0.283	0.252	0.184	0.208
Cohort-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Standard errors in parentheses; The stars refer to the following significance level:
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
Source: Own calculations based on SOEP v35.

TABLE A4 Regression results for annual earnings—women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
One child	0.113	−0.033	−0.130***	−0.008	−0.005	0.101***	0.059	−0.086	−0.044
	(0.150)	(0.051)	(0.045)	(0.036)	(0.037)	(0.034)	(0.039)	(0.047)	(0.071)
Two children	−0.485	−0.234**	−0.205*****	−0.044	0.007	0.092***	0.131***	0.066	−0.024
	(0.486)	(0.100)	(0.059)	(0.041)	(0.039)	(0.035)	(0.039)	(0.048)	(0.071)
3 or more children		−0.056	0.174*	−0.160***	−0.011	0.081*	0.132***	0.079	−0.034
		(0.183)	(0.093)	(0.056)	(0.049)	(0.043)	(0.047)	(0.058)	(0.083)
Married	0.040	−0.062*	0.038	0.044	−0.004	0.102***	−0.036***	−0.013	0.048**
	(0.096)	(0.034)	(0.034)	(0.031)	(0.029)	(0.026)	(0.028)	(0.035)	(0.048)
Years FT	0.466***	0.098***	0.065***	0.031***	0.036***	0.025***	0.023***	0.009*	0.031***
	(0.060)	(0.021)	(0.014)	(0.009)	(0.007)	(0.005)	(0.005)	(0.006)	(0.007)
Years FT (sq)	−0.059***	−0.006***	−0.003***	−0.000	−0.001**	−0.000	−0.000	0.000	−0.000**
	(0.015)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years PT	−0.099	−0.033	−0.005	−0.001	−0.002	−0.010*	0.002	−0.001	−0.011
	(0.103)	(0.027)	(0.016)	(0.011)	(0.008)	(0.006)	(0.006)	(0.006)	(0.008)
Years PT (sq)	0.034	0.002	0.001	0.001	0.001	0.001***	0.000	0.000	0.000**
	(0.028)	(0.004)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years UE	−0.524***	−0.076*	−0.174***	−0.010	−0.075***	−0.097***	−0.079***	−0.050***	−0.047*
	(0.197)	(0.040)	(0.039)	(0.021)	(0.018)	(0.018)	(0.018)	(0.017)	(0.024)
Years UE (sq)	0.141	−0.002	0.033***	−0.006**	0.005***	0.007***	0.004**	0.001	0.001
	(0.134)	(0.006)	(0.009)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Weekly hours	0.066***	0.054***	0.081***	0.093***	0.091***	0.087***	0.105***	0.113***	0.093***
	(0.009)	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)	(0.006)

(Continues)

TABLE A4 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
Weekly hours (sq)	−0.001***	−0.000***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	−0.046**	0.002	−0.034**	−0.071***	0.013	0.007	−0.024	0.012	−0.051
	(0.023)	(0.016)	(0.015)	(0.015)	(0.023)	(0.015)	(0.020)	(0.028)	(0.036)
Education (sq)	0.002	0.001	0.003***	0.006***	0.003***	0.003***	0.004***	0.002**	0.005***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	7.583***	8.273***	8.012***	7.637***	7.141***	7.141***	7.140***	6.986***	7.550***
	(0.197)	(0.137)	(0.130)	(0.126)	(0.169)	(0.119)	(0.153)	(0.213)	(0.277)
Obs.	382	882	1307	1859	2493	2653	2043	1320	778
R-squared	0.573	0.540	0.627	0.663	0.578	0.599	0.674	0.681	0.660
Cohort-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Standard errors in parentheses; The stars refer to the following significance level:
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
Source: Own calculations based on SOEP v35.

TABLE A5 Regression results for annual earnings—men

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
One child	0.189	0.061	−0.008	−0.014	0.014	0.028	−0.042	−0.003	0.073
	(0.332)	(0.052)	(0.028)	(0.023)	(0.021)	(0.024)	(0.028)	(0.038)	(0.051)
Two children	−1.006	0.139*	0.066**	0.067***	0.026	0.091***	−0.008	0.019	0.015
	(0.714)	(0.075)	(0.032)	(0.023)	(0.021)	(0.022)	(0.026)	(0.036)	(0.047)
3 or more children	−0.008	0.128*	−0.026	0.025	0.050*	0.051*	0.013	0.100**	0.157**
	(0.164)	(0.071)	(0.041)	(0.032)	(0.026)	(0.028)	(0.034)	(0.047)	(0.065)
Married	0.017	0.021	0.121***	0.078***	0.089***	0.054**	0.091***	0.028	0.099**
	(0.158)	(0.039)	(0.025)	(0.022)	(0.020)	(0.022)	(0.025)	(0.034)	(0.046)
Years FT	0.731***	0.179***	0.094***	0.058***	0.045***	0.043***	0.031***	0.052***	−0.033
	(0.058)	(0.022)	(0.012)	(0.008)	(0.006)	(0.008)	(0.009)	(0.012)	(0.035)
Years FT (sq)	−0.106***	−0.013***	−0.005***	−0.002***	−0.001***	−0.001***	−0.001**	−0.001***	0.001
	(0.014)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years PT	0.095	−0.191***	−0.044***	−0.020*	−0.060***	−0.042***	−0.057***	−0.079***	−0.056***
	(0.174)	(0.034)	(0.016)	(0.011)	(0.010)	(0.011)	(0.010)	(0.013)	(0.015)
Years PT (sq)	−0.051	0.024***	0.003	0.000	0.003***	0.003***	0.003***	0.003***	0.001*
	(0.067)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years UE	−0.204	−0.107**	−0.188***	−0.110***	−0.118***	−0.101***	−0.091***	−0.109***	−0.122***
	(0.186)	(0.047)	(0.026)	(0.014)	(0.012)	(0.012)	(0.014)	(0.018)	(0.035)
Years UE (sq)	0.040	−0.005	0.023***	0.004***	0.006***	0.005***	0.004***	0.004***	0.008**
	(0.097)	(0.014)	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Weekly hours	0.044***	0.049***	0.070***	0.052***	0.037***	0.054***	0.077***	0.053***	0.071***
	(0.009)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)

(Continues)

TABLE A5 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
Weekly hours (sq)	−0.000***	−0.000***	−0.001***	−0.000***	−0.000***	−0.000***	−0.001***	−0.000***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	−0.018	−0.040***	−0.076***	−0.041***	−0.057***	0.039*	0.053**	0.091*	0.019
	(0.028)	(0.012)	(0.009)	(0.009)	(0.012)	(0.021)	(0.023)	(0.053)	(0.065)
Education (sq)	0.003	0.004***	0.005***	0.004***	0.004***	0.002**	0.001	−0.001	0.002
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)
Constant	7.710***	8.367***	8.434***	8.715***	9.163***	7.965***	7.506***	7.527***	8.427***
	(0.195)	(0.113)	(0.099)	(0.103)	(0.115)	(0.174)	(0.201)	(0.394)	(0.730)
Obs.	383	900	1746	2464	2863	2939	2261	1546	980
R-squared	0.542	0.539	0.481	0.394	0.409	0.417	0.437	0.400	0.522
Cohort-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Standard errors in parentheses; The stars refer to the following significance level:
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
Source: Own calculations based on SOEP v35.

TABLE A6 Distribution of cohorts by gender

Birth Cohort	Number of Men	Number of Women	Total
1964	382	324	706
1965	373	383	756
1966	404	425	829
1967	385	401	786
1968	378	385	763
1969	388	387	775
1970	311	364	675
1971	303	342	645
1972	288	304	592
Total	3212	3315	6527

Source: Own calculations based on SOEP v35.

TABLE A7 Overview regression models of the dynamic microsimulation

Dependent Variables	Explanatory Variables
Child birth in $t + 1$ (Logit)	Number of children, age of youngest child, earnings; additionally, for married women: partner's age, highest level of education and earnings; run separately for married women and single women
Change in marital status (married/single) in $t + 1$ (Logit)	Marriage duration term interacted with age, number of children; additionally, for women: age of youngest child; additionally, for married individuals: Partner's age and highest level of education; run separately for men and women for each respective marital status
Change in labor force status in $t + 1$ (Logit)	Labor force status in t and $t - 1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men); additionally, for women: number of years since birth of last child; additionally, for married individuals: partner's labor force status and earnings in t; run separately for men and women for each respective marital status
Change in employment status (working/ unemployed) in $t + 1$ (Logit)	Employment status in $t - 1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men); additionally, for women: number of years since birth of last child; additionally, for married individuals: partner's employment status and earnings of the in t; run separately for men and women for each possible combination of marital and employment status in t
Transition in employment or unemployment in $t + 1$ after not participating in the labor market in t (Logit)	Employment status in $t - 1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men); additionally, for women: number of years since birth of last child; additionally, for married individuals: partner's employment status and earnings in t; run separately for men and women for each respective marital status (requirement: participating in the labor market in $t + 1$)
Transition full-time work/part-time work in $t + 1$ (Logit)	Labor force status in $t - 1$, dummy variable indicating full-time or part- time work in $t - 1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men); additionally, for women: number of years since birth of last child; additionally, for married individuals: partner's employment status and earnings of the partner in t; run separately for men and women for each possible combination of marital and full-time/ part-time status in t
Transition in full-time work/part-time work in $t + 1$ after not working in t (Logit)	Labor force status in $t - 1$, dummy variable indicating full-time or part-time work in $t - 1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men)-, additionally, for women: number of years since birth of last child; additionally, for married individuals: employment status and earnings of the partner in t; eun separately for men and women for each respective marital status (requirement: working in $t + 1$)

TABLE A7 (Continued)

Dependent Variables	Explanatory Variables
Number of working hours in t (OLS)	Annual hours worked in $t-1$ and $t-2$, annual earnings in $t-1$, dummy variable indicating full-time or part-time work in and labor market status $t-1$, number of children (not for unmarried men); additionally, for married individuals: earnings of the partner in $t-1$; eun separately for men and women for each respective marital and work (full-time/part-time) status
Annual earnings in t (OLS)	Annual earnings in $t-1$ and $t-2$, annual hours worked in t, $t-1$ and $t-2$, labor market history (years in full-time, part-time, unemployment), dummy indicating marital status; run separately for men and women

Notes: Explanatory variables which are included in every model: highest level of education and year of birth interacted with (quadratic) age, place of residency before 1989 (East or West Germany), immigration background (yes or no). This table depicts forward-looking simulations. Backward-looking simulations function analogously.
Source: Own calculations based on SOEP v35.

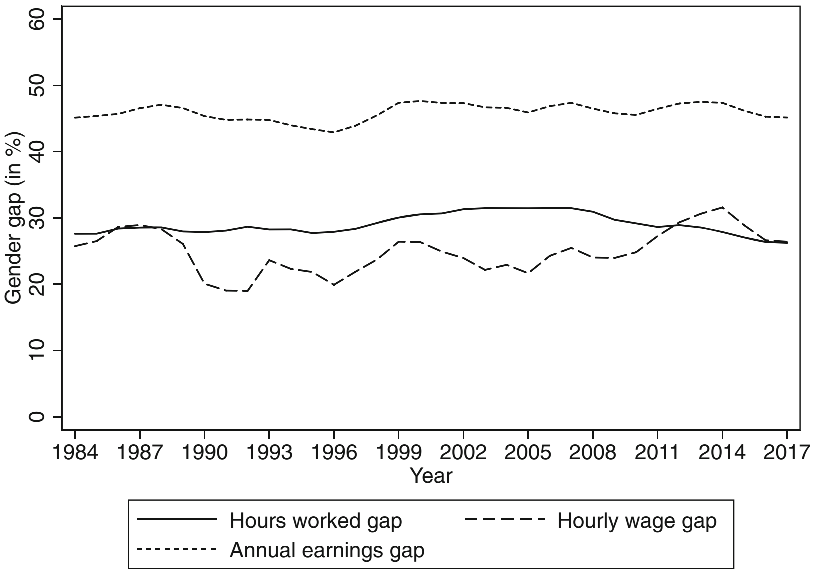

FIGURE A1 Gender gaps in labor market outcomes by survey year. Only employed individuals are considered. Does not include values of zero annual earnings. Cohorts 1940–1979, weighted sample. *Source:* Own calculations based on SOEP v35.

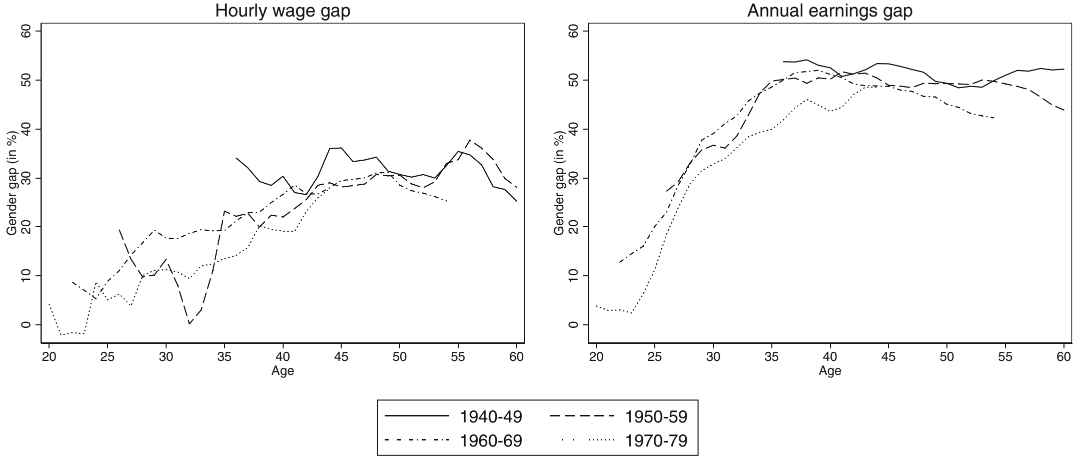

FIGURE A2 Gender gaps in hourly wages and annual earnings by cohort. Only employed individuals are considered. Does not include values of zero hourly wages or annual earnings. Cohorts 1940–1979, weighted sample. *Source:* Own calculations based on SOEP v35.

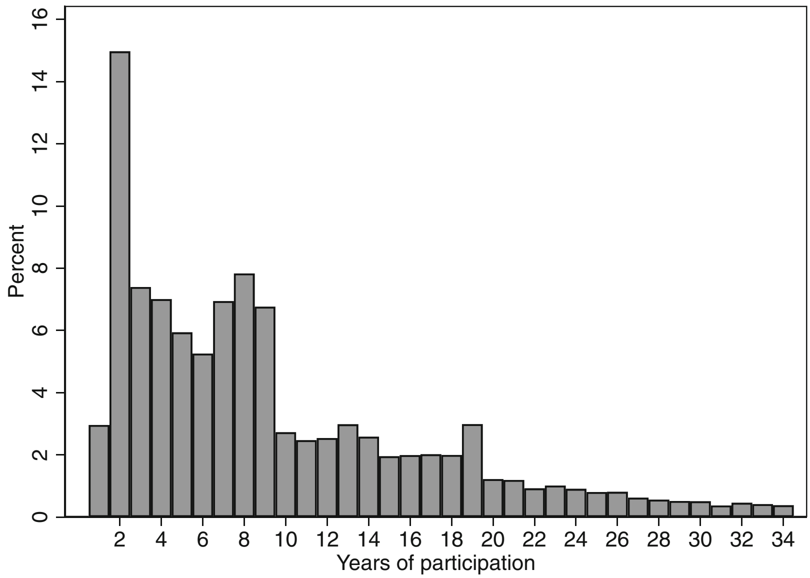

FIGURE A3 Distribution of participation years in the SOEP. *Notes*: Refers to participation years of the SOEP sample used in 2 of this paper. Restrictions for the microsimulation in Section 3 are not applied here. *Source:* Own calculations based on SOEP v35.

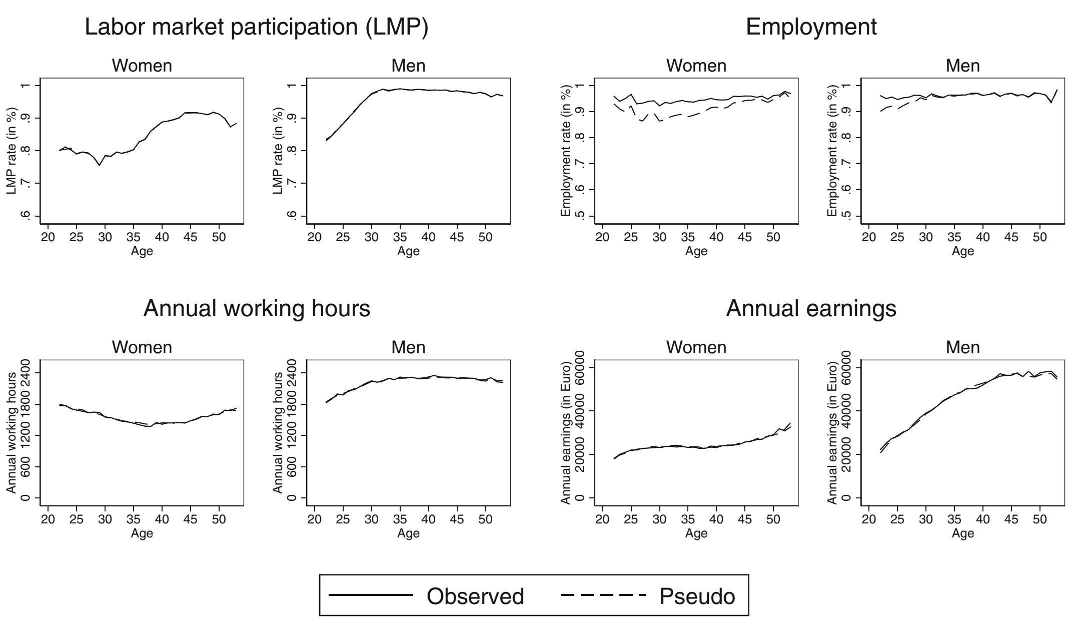

FIGURE A4 Pseudo missings for labor market outcomes. The graphs comparing truly observed and simulated pseudo information for annual working hours and annual earnings only focus on employed individuals. *Source:* Own calculations based on SOEP v35.

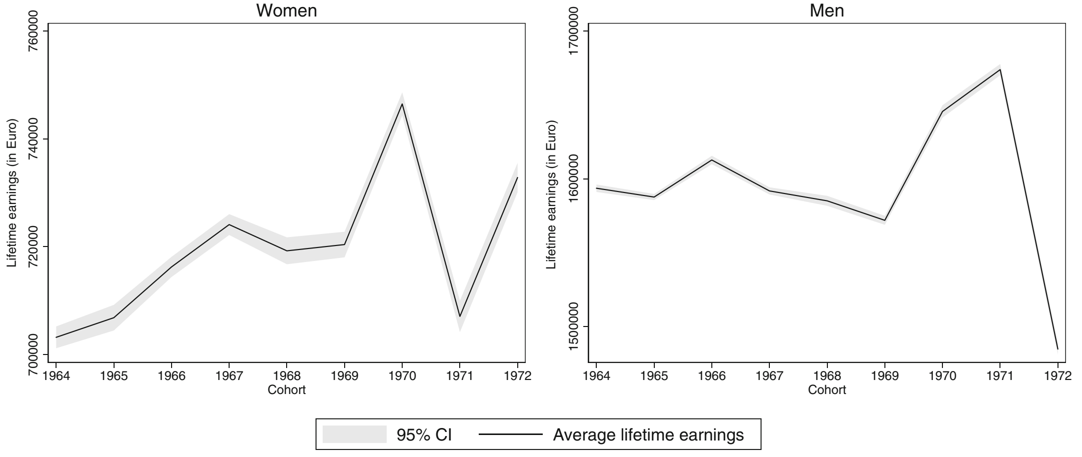

FIGURE A5 Monte Carlo simulation for earnings *Source*: Own *calculations* based on SOEP v35.

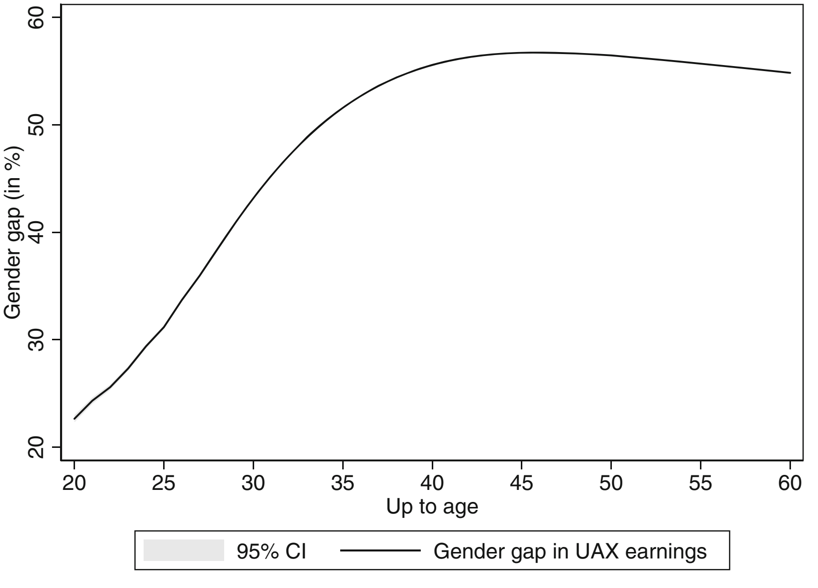

FIGURE A6 Monte Carlo simulation for the gender gap in UAX earnings *Source*: Own *calculations* based on SOEP v35.

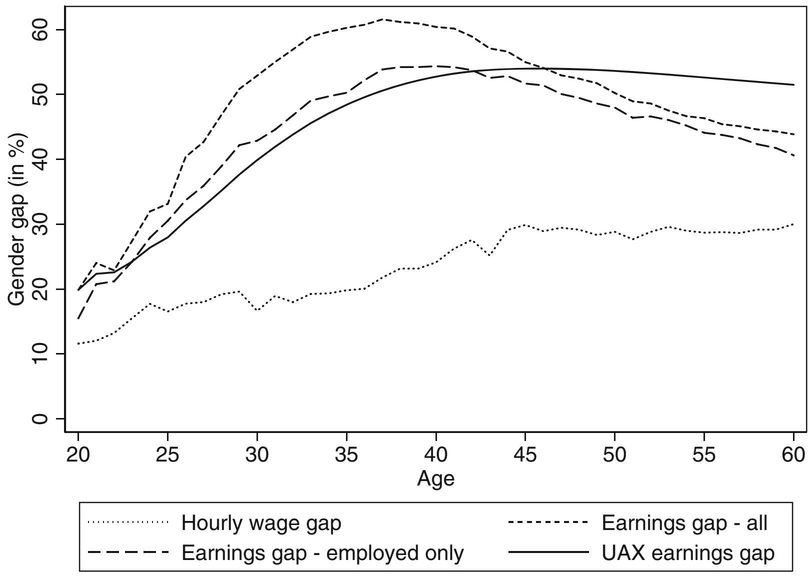

FIGURE A7 Gender gaps in earnings by different concepts. *Individuals* with zero UAX earnings are included in the calculation. For annual earnings gap, all employed and unemployed individuals are considered. Cohorts 1964–1972. *Source:* Own calculations based on SOEP v35.

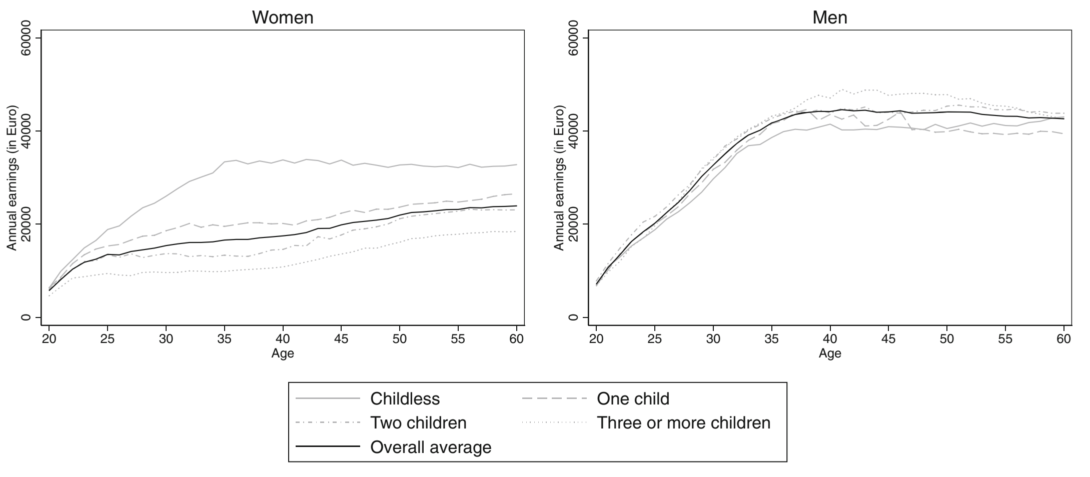

FIGURE A8 Annual earnings by gender and number of children. Employed and unemployed individuals are considered. Number of children refers to the total number at age 50. Cohorts 1964–1972. *Source:* Own calculations based on SOEP v35.

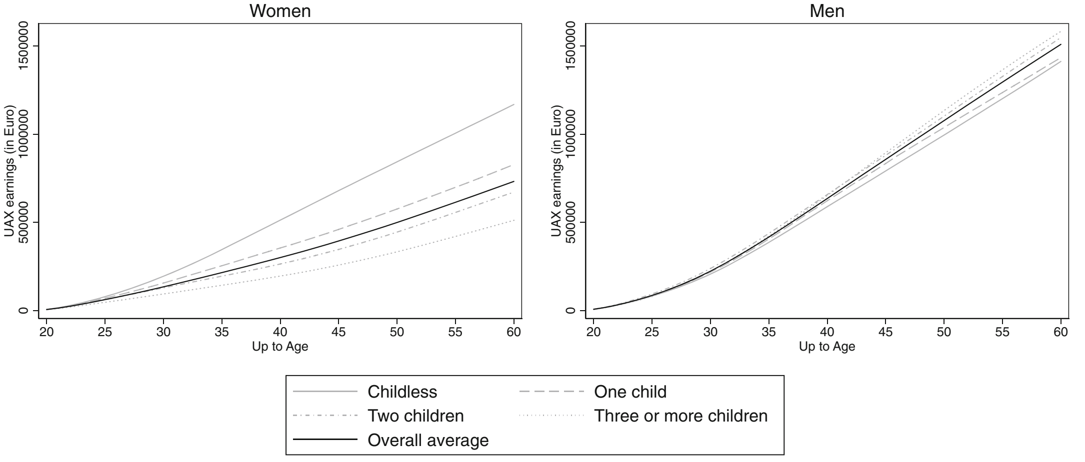

FIGURE A9 UAX earnings by gender and number of children. Employed and unemployed individuals are considered. Number of children refers to the total number at age 50. Cohorts 1964–1972. *Source:* Own calculations based on SOEP v35.