DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 52, ARTICLE 22, PAGES 689–740
PUBLISHED 9 APRIL 2025
https://www.demographic-research.org/Volumes/Vol52/22
DOI: 10.4054/DemRes.2025.52.22

Research Article

Uncovering what matters: Family life-course aspects and personal wealth in late working age

Nicole Kapelle

Carla Rowold

Contents

1	Introduction	690
2	Wealth accumulation and the life course	692
3	The relevance of family life courses to the accumulation of wealth	693
3.1	Linking marital status and transitions to wealth	693
3.2	Child-related wealth benefits and penalties	695
3.3	The interconnectedness of parenthood and marital status in the accumulation of wealth	696
4	The present study	697
5	Data and methods	698
5.1	Data and sample	698
5.2	Personal net wealth	699
5.3	Family life courses	699
5.4	Overview of methodological approach	701
5.5	Describing the process of trajectory's feature extraction and selection	703
6	Relevant family life-course predictors of wealth in late working age	704
7	Association of relevant family life-course features and wealth in late working age	706
8	Exploring potential variations by gender	709
9	Conclusion and discussion	712
10	Acknowledgements and funding	716
	References	717
	Appendix	725

Demographic Research: Volume 52, Article 22
Research Article

Uncovering what matters: Family life-course aspects and personal wealth in late working age

Nicole Kapelle[1]

Carla Rowold[2]

Abstract

BACKGROUND
Capturing the complexity of family life courses as predictors of later-life outcomes like wealth is challenging. Previous research has either (a) assessed a few selective but potentially irrelevant summary indicators, or (b) examined entire life-course clusters without identifying specific important aspects within and between them.

OBJECTIVE
Our aim is to investigate which family life-course variables that capture the order, duration, and timing of states and transitions are key personal wealth predictors for Western Germans aged 50 to 59, and to analyse the strength and direction of associations between the relevant variables and personal wealth, and whether these differ by gender.

METHODS
We used German Socio-Economic Panel (SOEP) data and combined feature selection, sequence analysis tools, and regression techniques.

RESULTS
We identified 23 family life-course variables as relevant predictors, with 2 – the time spent never-married, with and without children – deemed most relevant. Most family life-course variables were negatively associated with personal wealth and characterised by single parenthood, marital separation, or early marital transitions with or without fertility transitions. The prevalence and significance of some of the associations between these variables and personal wealth differed across genders. The results highlight the importance of previously concealed family life-course variables for wealth inequalities in late working age.

[1] Corresponding author. Department of Sociology, Trinity College Dublin, Ireland. Department of Social Sciences, Humboldt-Universität zu Berlin, Germany. Nuffield College, University of Oxford, UK. ORCID ID 0000-0001-5855-1153. Email: nicole.kapelle@tcd.ie.
[2] Nuffield College, University of Oxford, UK. Max Planck Institute for Demographic Research, Germany. Max Planck – University of Helsinki Centre for Social Inequalities in Population Health, Germany and Finland. ORCID ID 0000-0003-0292-6096. Email: rowold@demogr.mpg.de.

CONTRIBUTION

We extend previous research on the nexus between family demography and wealth stratification by using a novel, data-driven approach that more effectively explores family life-course complexities by considering the 'entire' universe of variables that describe such life courses and identifying those life-course variables that are relevant wealth predictors.

1. Introduction

Even in countries with formerly monolithic, generous public pension systems, the relevance of sufficient private wealth to supplement or even replace public pensions has drastically increased over recent decades as a response to the financial pressures on the public pension systems generated by an ageing population. However, wealth inequalities in late working age are soaring in most OECD countries. Thus, a substantial proportion of the population cannot meet the increasing personal responsibility to secure economic living standards throughout the life course, and particularly during older age. Understanding the drivers of wealth inequality is of utmost importance for policymakers and researchers alike. This understanding is crucial to reducing both economic disparities and reliance on welfare, thereby maintaining social cohesion and addressing the challenges associated with an ageing population.

Traditionally, debates and research on the drivers of wealth inequalities have focused on the role of social background and labour market position (Atkinson 1971; Bernardi, Boertien, and Geven 2018). More recently, a growing body of research has highlighted the relevance of the family as an important context for socioeconomic stratification and wealth inequality (e.g., Halpern-Manners et al. 2015; Kapelle and Vidal 2022; Ulker 2008; Wilmoth and Koso 2002; Zissimopoulos, Karney, and Rauer 2015). Overall, such studies have identified a distinct set of life-course aspects that are commonly associated with wealth advantages. These include, inter alia, continuous marriage that is entered at an average age and childbirth within marriage with two or – depending on the cultural and historical context – a maximum of three children. Such aspects reflect a culturally and contextually idealised family life course. On the contrary, family life-course patterns that break with this 'ideal' – for instance, through divorce or unmarried parenthood – have often been linked to less favourable wealth outcomes.

Family life courses exhibit a high degree of complexity due to their variation, which encompasses the types of family events and transitions encountered throughout an individual's life as well as the order, timing, and duration of these experiences. The analysis of the implications of this family complexity is becoming increasingly urgent, particularly when viewed through a gendered lens. Family complexity has increased, particularly for women (Van Winkle and Fasang 2021). Researchers have employed two strategies to analyse this complex association between family life courses and wealth. The first operationalises family life courses using a small range of summary indicators, such as age at first marriage, number of times married, duration of first marriage, and number of children, and these indicators are used to predict wealth levels (Halpern-Manners et al. 2015). Although research that follows this methodological tradition has provided critical initial analysis, the approach is highly selective and limited regarding the indicators and number of indicators that can be considered, due to multicollinearity issues (Rowold, Struffolino, and Fasang 2024). This means that the approach may not capture all relevant indicators of family life courses and provides an incomplete picture of the association of interest. The second strategy uses sequence analysis and cluster analysis to reduce the complex universe of family life-course trajectories into a set of meaningful and distinct groups of life courses. Identified clusters are subsequently used within a regression framework to predict wealth in late working age (Kapelle and Vidal 2022). While this approach offers a comprehensive, descriptive overview of the relationship between typical family life courses and wealth, it lacks specificity in identifying which particular life-course aspects, within or across these groups, are most relevant to the accumulation of wealth, and thus wealth levels in late working age.

Using a novel methodological approach, the present study addresses the shortcomings of previous research. Specifically, focusing on both marital and fertility life-course aspects, we (1) illustrate which family trajectory aspects out of a multitude of life-course summary measures are most relevant to personal wealth at ages 50 to 59 among cohorts of West Germans born between 1943 and 1967. We consider wealth at ages 50 to 59 because wealth penalties and advantages accumulate over the life course. As such, wealth inequalities should be particularly visible at pre-retirement age when wealth levels are expected to peak in anticipation of retirement (Alessie, Lusardi, and Aldershof 1997; Modigliani 1988). Additionally, we (2) investigate how the aspects identified as the most relevant relate to personal wealth in late working age. Specifically, we assess the extent to which and direction in which 'relevant' aspects predict personal wealth in late working age. Finally, we (3) examine whether the extent and direction of how family life-course aspects are associated with personal wealth in late working age differ by gender.

To address our research objectives, we use data from the German Socio-Economic Panel (SOEP v38, 1984–2021; Goebel et al. 2019) and apply a methodological approach that combines a feature selection algorithm and sequence analysis tools with regression techniques to identify wealth-relevant family life-course aspects (i.e., variables describing family life courses – referred to as 'features' in machine learning) and study the extent to which and direction in which they are associated with the personal wealth of women and men in late working age (Bolano and Studer 2020). Our analyses capture family life-course aspects experienced between the ages of 15 and 50. These aspects reflect the timing (i.e., age at which an event or transition takes place), order (i.e., the order in which events and transitions take place), and duration (i.e., how long individuals spend within a certain state) of family life-course events and transitions. Our data-driven, exploratory approach is linked to the life-course framework which highlights both the importance and challenges of capturing entire life courses, as opposed to focusing on snapshots of individuals' lives. Our methodological approach allows us to address these concerns within the life-course framework and uncover patterns and relationships between family life courses and wealth that are not pre-specified or constrained by existing theoretical frameworks and previous literature, thus contributing to the development of new theoretical insights.

We focus on West Germany as an intriguing setting for our study because its culture and institutions have continually supported traditional family structures involving stable marriage and male breadwinner ideologies, despite widespread changes in women's societal roles, endowments, and relationship and childbearing patterns (Trappe, Pollmann-Schult, and Schmitt 2015). East Germany, on the other hand, followed substantially different social policies regarding family life, gender equality, and wealth accumulation, resulting in very different family patterns and wealth accumulation profiles for the cohort under study (Kapelle and Weiland 2025). Therefore, it is not feasible to combine these two contexts in our study.

2. Wealth accumulation and the life course

Wealth accumulation is a dynamic process that occurs through three main pathways (e.g., Keister and Moller 2000; Killewald, Pfeffer, and Schachner 2017; Spilerman 2000). First, surplus income, originating from labour earnings, social welfare, or assets like rental income, interest, and dividends, may be saved or reinvested for exponential growth. Second, wealth may be obtained through financial transfers such as inter vivos transfers (i.e., transfers made during the grantor's life), inheritances, or other windfall profits. Finally, wealth may increase through capital appreciation, depending on the individual wealth portfolio and financial markets.

Wealth levels are expected to change over the life course as part of the dynamic process of wealth accumulation. Specifically, individuals commonly start with low or no wealth. Wealth then grows during working years before peaking before retirement. This may be considered the normative wealth accumulation pathway. Nevertheless, wealth levels in late working age are exceedingly heterogeneous and financial preparedness for retirement varies drastically (Halpern-Manners et al. 2015; Hurd 2002). In a life-course framework these wealth inequalities can be understood as the result of age differentiation processes (Bernardi, Huinink, and Settersten 2019; Dannefer 2003; O'Rand 1996): Initial comparative advantages or disadvantages at an early age restrict or enhance future wealth levels and accumulation potential. This path dependency means that individuals cumulate and compound disadvantages or advantages over the life course. Thus, wealth levels and accumulation potential progressively differentiate between individuals as they age.

In line with path dependencies, the experience of certain life-course events and transitions can generate enhancing or disruptive effects on an individual's wealth accumulation, with potentially lasting and flow-on effects for future opportunities. Very influential events or transitions are often denoted 'turning points' (Abbott 2001). However, it is not just whether enhancing or disruptive life-course events and transitions occur that determines an individual's ability to accumulate wealth over their life course, but also the order in which (e.g., being married before childbirth) and the time at which (e.g., getting married before the age of 20) they occur, and the duration an individual stays within a certain state (e.g., the time spent as an unmarried parent).

3. The relevance of family life courses to the accumulation of wealth

The present study focuses on family life-course features – the occurrence, timing, and order of events and transitions and the time spent in certain states – as structuring sources of opportunities or barriers to accumulate personal wealth until late working age. Parenthood and marital roles are known to affect labour income, consumption, investments, and wealth transfers which directly impact the process of wealth accumulation over the life course (e.g., Budig, Misra, and Boeckmann 2012; Kapelle and Lersch 2020; Leopold and Schneider 2011; Lersch, Jacob, and Hank 2017).

3.1 Linking marital status and transitions to wealth

Research consistently shows that married individuals accumulate more wealth throughout the life course than never-married, divorced, and re-married individuals (e.g., Addo and Lichter 2013; Halpern-Manners et al. 2015; Hao 1996; Wilmoth and Koso 2002). This 'marital wealth advantage' stems from several interconnected contextual and normative factors. First, depending on the institutional context, married spouses may benefit from favourable tax rates, joint insurances (e.g., health, car, or life insurance plans that cover both spouses and offer lower premiums), and pensions (Christl, De Poli, and Ivaškaitė-Tamošiūnė 2023). Germany's joint taxation system (*Ehegattensplitting*) exemplifies this (Schechtl and Kapelle 2024). Second, married couples are more likely to acquire assets such as housing property, facilitated by shared transaction costs and better borrowing terms (Thomas and Mulder 2016). Third, social norms surrounding marriage as a lifelong emotional and financial commitment encourage joint investments and resource pooling, creating economies of scale that make wealth accumulation more efficient (Knoll, Tamborini, and Whitman 2012; Vogler, Lyonette, and Wiggins 2008). Fourth, marriage increases the likelihood of receiving intergenerational transfers to support a couple's joint future (Leopold and Schneider 2011). Finally, marriage is associated with achieving normative financial milestones (Gibson-Davis, Gassman-Pines, and Lehrman 2018), such as stable income or asset ownership, which couples –but particularly men – may feel they need before marrying (Schneider 2011; Xie et al. 2003). Thus, couples have been shown to postpone marriage until they have the economic means.

Overall, marriage entry has been shown to reflect a relevant milestone in individuals' wealth accumulation process. However, it is not only the occurrence of marriage that seems to matter, but the timing. Marrying very early or late often reflects economic disadvantage due to aspects such as selection, economic insecurity, or limited educational and career opportunities, while moderate-age marriages align with better economic outcomes. Moreover, cumulative (dis)advantage theory suggests that the initial (e.g., intergenerational transfers) as well as the ongoing advantages of marriage (e.g., compounded interest effects, better access to mortgages, etc.) accumulate over time, leading to wealth advantages (Dannefer 2018; O'Rand 1996).

By contrast, marital dissolution through separation or divorce has negative short- and long-term effects on wealth accumulation (Addo and Lichter 2013; Kapelle 2022; Kapelle and Baxter 2021; Ozawa and Lee 2006). First, divorced individuals lose marriage-related financial advantages. Second, asset division, especially of major holdings like a home, is often difficult because spouses may lack other assets to offset the home's value or to qualify for a new mortgage individually. As a result, selling the home can incur high transaction costs, especially in unfavourable markets (Lersch and Vidal 2014). Third, ongoing financial obligations such as spousal maintenance can hinder wealth recovery, with men more likely to provide financial support due to child custody patterns. Finally, divorce is socially stratified, with economically stressed couples facing a higher risk of separation (Dew 2011; Dew, Britt, and Huston 2012; Eads and Tach 2016).

Thus, marital dissolution serves as a critical turning point in the wealth accumulation process. Its impact depends not only on whether divorce occurs but also on when and whether it is followed by remarriage, as both are crucial to economic recovery (Kapelle 2022; Ulker 2008; Wilson and Clarke 1992). Early divorces may involve fewer (joint) assets, making the separation less financially disruptive, and allow for longer recovery time. Early divorce is also linked to a higher likelihood of remarriage, which can facilitate financial recovery. Later divorces, however, are more financially taxing because assets are more interwoven, recovery time is shorter, and the likelihood of remarriage declines.

Individuals may also abstain from marriage. As previously highlighted, marriage tends to attract financially stable individuals. In turn, economically disadvantaged men and women are more likely to remain never-married, leading to lower wealth accumulation (Addo 2014; Carlson, McLanahan, and England 2004; Gibson-Davis, Edin, and McLanahan 2005; Smock, Manning, and Porter 2005). While staying unmarried results in a lack of all marital benefits, cohabiting individuals likely benefit from some, although not all, benefits, depending on the country context. However, long-term cohabitation was uncommon in the present study's cohorts of interest and country context, only playing a negligible role in the life courses of women and men (Le Goff 2002). It should be noted that cohabitation is not captured in the current study due to data limitations: instead, we focus on marital trajectories.

3.2 Child-related wealth benefits and penalties

The anticipation or presence of children is associated with a range of wealth-enhancing mechanisms, such as increased savings incentives, homeownership investments, and saving for children's education. It has also been shown to be associated with a higher likelihood of receiving intergenerational transfers (Leopold and Schneider 2011; Lusardi, Cossa, and Krupka 2001). However, despite these advantages and institutional support such as child allowances, parenthood entails significant direct and indirect costs.

Direct costs, borne disproportionately by mothers, include children's daily consumption expenses, childcare, and education costs (Bradbury 2011; Lanau 2023). The indirect financial costs of childbirth particularly emerge for women when they experience restrictions regarding employment opportunities, reduced working hours, and care-related career breaks (Budig and England 2001), all of which translate into mothers' but not fathers' lower income and reduced wealth accumulation (Lersch, Jacob, and Hank 2017). Although some of these intra-couple economic inequalities are likely compensated through pooling and sharing, previous research has highlighted that not all economic resources are pooled and shared – even within marriage – and substantial intra-couple gaps are visible for income and wealth (Hamplová, Le Bourdais, and Lapierre-Adamcyk 2014; Haupt and Strauß 2022; Kapelle and Lersch 2020). Additionally, employment offers a range of individual fringe benefits (e.g., access to private pension benefits or bonuses) that boost wealth accumulation (Chang 2010).

Early childbirth particularly increases indirect costs for women by limiting educational attainment and career prospects (Gough and Noonan 2013). Consequently, earlier births are more financially detrimental to mothers' wealth than later births, although men's wealth accumulation is also negatively affected by early fatherhood (Lersch, Jacob, and Hank 2017). In addition to the timing of childbirth, the number and spacing of births may matter, as child-related expenses increase with each additional child, and wider spacing may be more detrimental to women's careers due to longer consecutive periods out of the workforce.

3.3 The interconnectedness of parenthood and marital status in the accumulation of wealth

How childbearing is linked to the accumulation of wealth is closely intertwined with the transition in and out of marriage or the absence of such a transition. Parenthood within marriage makes financial transfers from husband to wife more likely (Eickmeyer, Manning, and Brown 2019) and increases the ability to save jointly for children, as married parents benefit from marital wealth premiums (Grinstein-Weiss et al. 2008). Particularly in the cohort and context studied here, parenthood within marriage was socially and institutionally encouraged, while childbirth outside of marriage was stigmatised (Le Goff 2002). As such, childbirth commonly followed marriage, or parents married shortly after childbirth if they were partnered.

While parents who marry after childbirth can fully benefit from all marital advantages once they enter marriage, lasting parenthood out of marriage (i.e., childbirth to unmarried individuals that stay unmarried after childbirth) is accompanied by a lack of marital wealth premiums as well as potentially reduced financial transfers. This is particularly the case between parents that do not live together, as previous research has illustrated that non- and under-payment of child support are common issues (Manning, Stewart, and Smock 2003). Unmarried parenthood can also be the result of marital dissolution (i.e., after separation and divorce), which is accompanied by a loss of marital premiums and partner's support in addition to any other divorce-related wealth penalties. Both pathways into unmarried parenthood – never-married parenthood and parenthood after marital separation and divorce – are highly selective of individuals with fewer economic resources (Upchurch, Lillard, and Panis 2002). As children commonly stay with the mother (Walper, Entleitner-Phleps, and Langmeyer 2021), the costs of unmarried parenthood are over-proportionally experienced by women, affecting women's wealth accumulation. The costs of unmarried parenthood may be reduced through (re)marriage. However, children in particular are a barrier and postponing factor for mothers' probability to enter a new marriage, but not for fathers (Di Nallo 2019).

Child-related costs and benefits regarding wealth accumulation are intertwined not only with marital status but also with the number of children. Generally, consumption costs increase with the number of children. This is accompanied by a simultaneous increase in the barriers to parents – particularly mothers – engaging in the labour market. Despite some institutional support for parents, the threshold at which child-related costs outweigh benefits also differs by marital status, with higher thresholds in marriage, where costs can be covered jointly, than outside of marriage (Zissimopoulos, Karney, and Rauer 2015). Thus, while moderate fertility is socially and institutionally supported, high fertility, with its associated child-related costs and the social perceptions surrounding it, may lead to penalties. Similarly, low fertility and childlessness are regarded as a violation of social norms and values, which can in turn evoke adverse wealth-relevant repercussions (e.g., fewer intergenerational transfers, discrimination, etc.). The wealth accumulation potential of both high fertility and low fertility and childlessness may be socially stratified.

4. The present study

Under the umbrella of the life-course framework, a large body of research has sought to connect earlier life-course aspects with later-life outcomes. Bernardi, Huinink, and Settersten (2019) emphasise that such research often conceptualises potentially relevant predictors as measures at a single point in time. This approach can be misleading as it overlooks the explanatory potential of other, potentially correlated predictors occurring earlier or later in life. Moreover, it often fails to fully capture all life-course dimensions, including not only the occurrence but also the timing, order, and duration of events or transitions. At the same time, Bernardi, Huinink, and Settersten (2019) acknowledge the increasing difficulty, both theoretically and methodologically, of capturing a large universe of interconnected predictors referring to an extended life-course timespan. As such, assessing which life-course variables are truly relevant becomes progressively more challenging. These challenges also impact the study of the association between family life courses and wealth in late working age. As illustrated in the previous sections, various family life-course aspects might influence wealth accumulation.

The present study navigates the challenges of capturing life-course complexity by using a new methodological approach. Considering the ‘entire’ universe of variables that describe family life courses, we use a data-driven approach to identify the variables that are the most relevant wealth predictors. However, the present study does not seek to highlight causal links between specific family life-course aspects and wealth. Instead, the aim is to provide a detailed description of the wealth-relevant features and explore the direction and magnitude of those features’ associations with respondents’ wealth in late working age. While we focus on family life courses, we acknowledge that they are closely interlinked with other life domains (e.g., work or educational trajectories). However, the interconnectedness of these domains is beyond the scope of the present study and should be considered an important avenue for future research.

5. Data and methods

5.1 Data and sample

The empirical analyses were based on longitudinal (prospective and retrospective) data from the German Socio-Economic Panel (SOEP v38, 1984–2021; Goebel et al. 2019). The SOEP is a large and nationally representative study that has tracked individuals living in eligible households annually since 1984. The dataset was suitable for our research purposes because it (1) collects information on a comprehensive set of wealth measures at the personal level in survey years 2002, 2007, 2012, and 2017, and (2) contains detailed information on marital and childbearing histories over the entire life course of the respondents.

We selected respondents who were aged 50 to 59 at any time between 2002 and 2017, who lived in West Germany in 1989, and who provided complete retrospective marital and fertility histories from age 15 to 50. Further, we restricted the sample to observations that had valid wealth information in at least one of the years in which wealth was measured (i.e., 2002, 2007, 2012, or 2017). For respondents who were captured more than once with a valid wealth wave between the ages of 50 and 59, we randomly selected one of the waves. This was necessary to reduce bias in the feature selection, as selected features would have been driven more strongly by the respondents with several waves. Note that respondents' wealth is measured at different ages (between ages 50 and 59). Additionally, family trajectories are only captured until age 50, meaning that there is potentially a gap between the last family life-course measurement and the measurement of respondents' wealth. This gap ranges from no gap to up to 9 years. While wealth should differ by age, it is also likely that family transitions, such as divorce or widowhood, occur between age 50 and when the respondents' wealth is measured. We account for these differences by including a range of covariates, which will be discussed in more detail later.

We further excluded 181 individuals (151 women, 30 men) who experienced widowhood before age 50, as the share of person-year-spells in widowhood was too small to be included. This led to 2.5% of the sample being excluded. As the SOEP has collected men's retrospective fertility data less frequently, the sample included fewer men than women. To prevent our results from being biased and driven by family states that are particularly important for women, we randomly selected the same number of women as men, establishing gender balance in our sample. This resulted in the deletion of 1,307 women. Robustness checks confirmed that women in the excluded group were similar to women in the sample group regarding family characteristics and other key measures. Our final sample consisted of 5,702 respondents with 2,851 women and the same number of men (see Figure A-2 for the number of individuals excluded at each sample selection step).

5.2 Personal net wealth

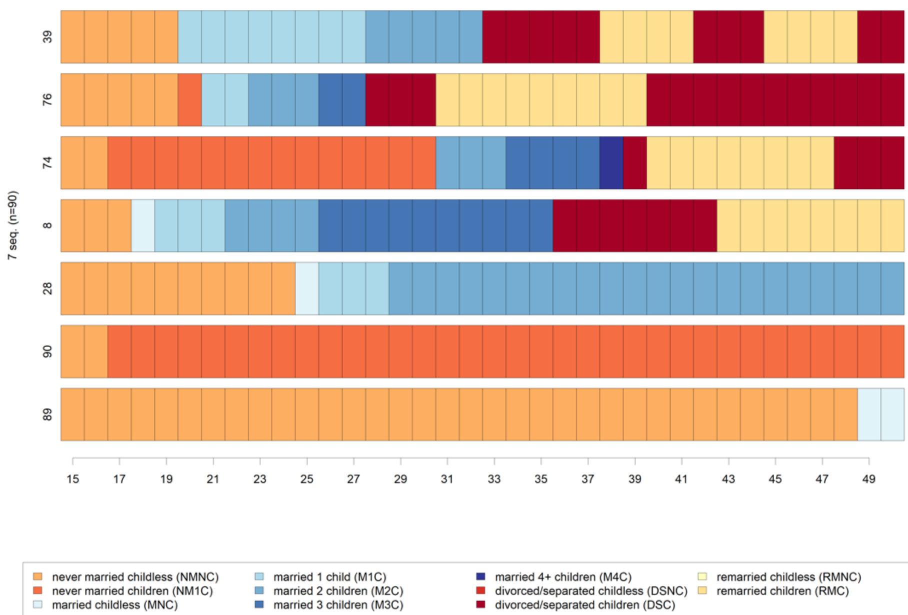
Our outcome variable of interest was a measure of personal net wealth, which was defined as the sum of all personally owned assets minus liabilities. Asset components in the SOEP include property assets, tangible and financial assets, private pensions, and business assets and collectables, while liabilities refer to consumer credits or mortgage debt. For each household member aged 17 and older, the SOEP collects personal wealth data in a three-step process: (1) a filter question is used to assess ownership of a certain wealth component, (2) the total market value of held wealth components is recorded, and (3) for jointly held wealth components, respondents are asked to provide the share they co-own. Our outcome measure thus explicitly includes the personal share of any assets and liabilities that are owned with other individuals. As liabilities were subtracted from assets, around 5% of respondents have a negative net worth, while around 11% have 0 net wealth. Figure A-3 in the Appendix gives an overview of the distribution of personal wealth by gender and Table A-1 provides more detail on the wealth distribution by gender, showing wealth levels at the mean, median, and 25th and 75th percentile.

We adjusted personal net wealth for inflation using the consumer price index of the German Statistical Office and top- and bottom-coded the extreme 0.1% of reported wealth measures to reduce the influence of outliers. These adjustments were done on the entire SOEP sample and thus before the sample restriction. Although wealth – similar to income – is often transformed for analyses to account for the skewness of the data (Killewald, Pfeffer, and Schachner 2017), we used the original, absolute wealth distribution in euros. This allowed us to better capture broader inequalities between respondents, which would be distorted through transformations. We used the imputed wealth variables provided by the SOEP team and took the mean value across the provided five imputed wealth sets (Grabka and Westermeier 2015).

5.3 Family life courses

We generated yearly indicators capturing the succession of family states over time from age 15 to 50 (see Figure 1 for seven example sequences extracted from our sample). To this end, we used biographical information on respondents' marital and fertility histories that were collected retrospectively and prospectively in the SOEP. The life-course states combined the two channels (partnership and fertility) into a total of 11 family states that could be experienced over the 46 years: (1) never-married and childless, (2) never-married with child(ren), (3) married and childless, (4) married with 1 child, (5) married with 2 children, (6) married with 3 children, (7) married with 4 or more children, (8) separated/divorced[3] and childless, (9) separated/divorced with child(ren), (10) remarried and childless, and (11) remarried with child(ren). Note that our analysis did not differentiate between children residing with the respondent at any specific time. Instead, the 'child(ren)' indicator was based on whether respondents reported having a child or children.

Figure 1: Example family sequences based on the analytical sample



Note: NMNC = never-married childless; NM1C = never-married with child(ren); MNC = married and childless, M1C = married with one child, M2C = married with two children, M3C = married with three children, M4C = married with four or more children, DSNC = divorced/separated and childless, DSNC = divorced/separated with child(ren), RMNC = remarried and childless, RMC = remarried with child(ren).
Source: Data are from the Socio-Economic Panel Survey v38 (2002, 2007, 2012, 2017).

[3] Separated refers to individuals who are legally married but live apart from their spouse, either without finalizing a divorce or before obtaining one. Since 1977, a one-year separation period has been a requirement for divorce in Germany.

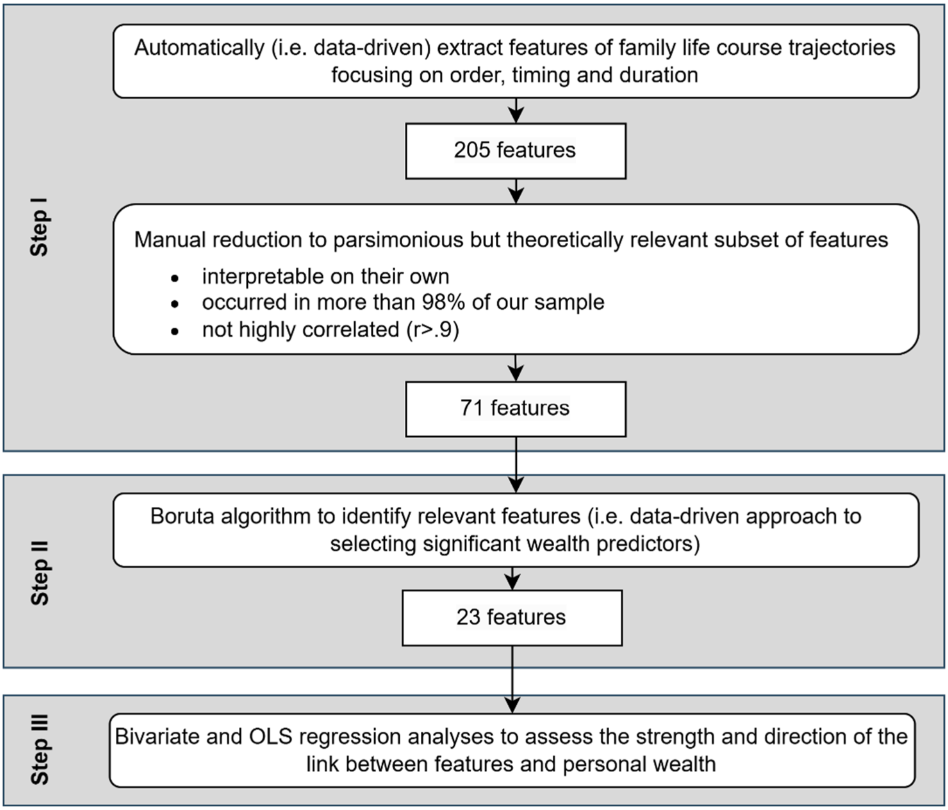
For methodological and theoretical reasons, the life-course states and thus complexity that could be captured had to be restricted. First, we only disaggregated the number of children within marriage but distinguished only between childless individuals and parents for the other three categories to ensure sufficient cell sizes. Second, our study focused on marital trajectories. Thus, cohabitation episodes were not captured. This limitation stems from the data collection approach of the SOEP, which emphasises retrospective partnership histories focusing on the formation and dissolution of marriages and does not collect detailed information on non-marital cohabitations. From a conceptual standpoint, we argue that the omission of cohabitation information is of minor concern for our study. For the cohort and context under study, individuals born between 1943 and 1967 in West Germany, cohabitation – especially long-term cohabitation and childbearing outside of marriage – was socially undesirable and discouraged (Le Goff 2002). Consequently, such states are likely to be rare in our data. Reinforcing this perspective, Rowold, Struffolino, and Fasang (2024) have demonstrated that in a cohort of slightly older West Germans and Italians, less than 1% of person-year spells between the ages of 18 and 65 were spent in non-married cohabitation.

5.4 Overview of methodological approach

In line with Bolano and Studer (2020), our empirical approach broadly followed three steps (see Figure 2). First, we automatically extracted a wide set of features of the family life-course trajectory.[4] Second, out of this large set we selected the features that were the most relevant wealth predictors. To this end, we used a data-mining feature selection algorithm, the Boruta algorithm (see next section). Third, the selected, smaller set of family life-course features (i.e., results from Boruta) were used as covariates in an Ordinary Least Square (OLS) regression model to estimate the direction and strength of each feature's association with wealth in late working age.

[4] Note that 'features' is the terminology used within machine learning to refer to variables. We use the terms 'features' and 'variables' interchangeably.

Figure 2: Overview of methodological process



For steps 2 and 3 of our approach, analyses were adjusted for a small set of key covariates. First, we added a continuous measure of respondents' age at the considered measurement of wealth ranging between age 50 and 59. Second, we added two dummy variables indicating whether respondents experienced a divorce or the death of their marital partner between the age of 50 and their considered wealth observation. Both were rare occurrences. In our sample of 5,702 respondents, 207 experienced a divorce (83 women, 124 men) and 70 the death of their partner (53 women, 17 men). Additionally, we accounted for respondents' socioeconomic background by including a continuous measure for their number of siblings, a dummy indicating whether they had a migration background (0 = no migration background, 1 = direct or direct migration background), and a categorical measure of their parents' educational achievement (0 = low, 1 = intermediate, 2 = high). Finally, we also added weights to our regression analyses to account for issues such as unequal probabilities of selection and over- or under-representation of certain groups.

The analytical steps 1 and 2 are discussed in more detail in the next section. The code for our analyses is available at https://osf.io/b9ya6/.

5.5 Describing the process of trajectory's feature extraction and selection

We used the *WeightedCluster* R package by Studer (2013) to automatically extract 205 family life-course variables (i.e., features) from our trajectories. These features covered the duration (1), ordering (2), and timing (3) of family-life states.

Duration was measured as the sum of years spent in the different life-course states. Using the sequences in Figure 1 as illustrative examples, individual 8, for instance, spent three years in the status 'married and having one child'. *Ordering* ('sub-sequences/ periods') captured whether and how often individuals experienced a period of the same state or sub-sequence of states. For example, individual 8 had one distinct period being 'divorced/separated and having children', and individuals 74 and 76 had two periods in this state. The variable *number of DSC–RMC sub-sequences* captures the sub-sequence of remarriage(s) following the state 'divorced/separated and having children'. For example, for individual 39 this sub-sequence occurred three times, as the first period of being divorced/separated is followed by two periods of being remarried and the second period of being divorced/separated is followed by one period of being remarried. We operationalised *timing* as the age range in which a state started in 5-year periods. For example, focusing on the start of 'being married and having no children', we can see that individual 28 experienced this between ages 25 and 29. Figure A-4 in the Appendix provides more examples.

The next aim was to reduce the rather long list of 205 automatically extracted features to a parsimonious but theoretically relevant subset of features. To this end, we followed suggestions by Bolano and Studer (2020) and removed, for instance, features that occurred in less than 2% of our sample and features that were highly correlated (see the Appendix for more detail). These restrictions resulted in a reduced set of 71 features.

To identify the family life-course features that were most relevant to personal wealth in late working age out of the list of 71 features, we applied the Boruta algorithm using the *Boruta* R package (Kursa and Rudnicki 2010). Following Bolano and Studer (2020), Boruta was deemed appropriate because it considers both various possible forms and potential interactions between features to assess their importance for the outcome and it has recently been shown to perform best, especially for low-dimensional data sets (Degenhardt, Seifert, and Szymczak 2019). The Boruta algorithm is a wrapping method based on a random forest (RF) approach. Essentially, an RF approach ranks all features according to their importance, but it usually does not determine a cut-off point distinguishing between important and unimportant features. Boruta provides such a cut-off point by generating shadow features which are randomised versions of the original features. The original features then compete against the shadow features and are only confirmed as being important for the outcome if they perform better in the random forest than the best-performing shadow feature. We provide a detailed description of the feature selection approach using the Boruta algorithm in the Appendix.

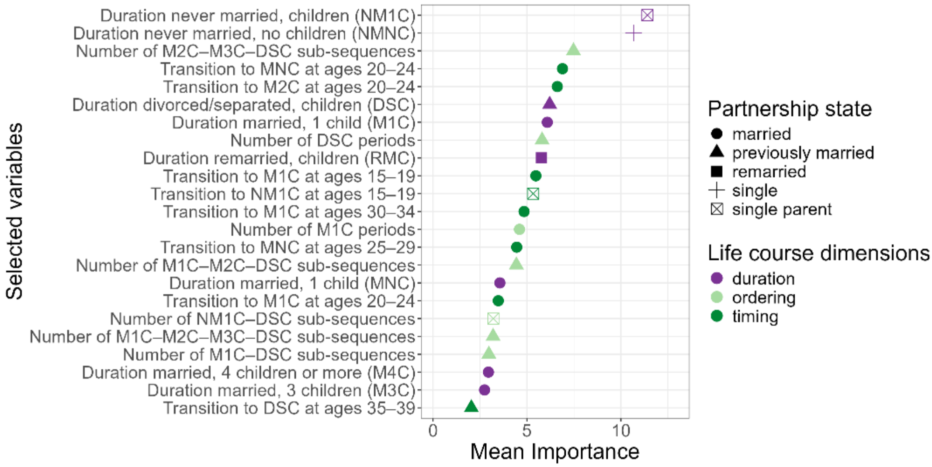
6. Relevant family life-course predictors of wealth in late working age

In total, the Boruta algorithm selected 23 family life-course features as relevant predictors of personal net wealth in late working age. These features are displayed in Figure 3, ranked by their mean 'importance' according to the Boruta algorithm.[5] Additional summary descriptives (i.e., median, minimum and maximum feature importance, means for continuous features, the prevalence of categorical features by gender) are provided in Table A-2 and Table A-3. As illustrated in Figure 3, features relating to the different life course dimensions – duration, order, timing – were selected equally often as relevant features (duration 8, order 7, timing 8). Additionally, features referring to being married (first time or higher order) or being unmarried (never married or divorced/separated) were also selected equally often, 12 and 11 times respectively.

Two features stand out as the most important wealth predictors according to the Boruta algorithm: duration spent in the statuses 'never-married without children' and 'never-married with child(ren)'. The high relevance of the duration spent in the status 'never-married with child(ren)' may highlight the persistent economic disadvantages associated with unmarried parenthood (Maldonado and Nieuwenhuis 2015; Sierminska 2018). The time spent in the status 'never-married without children' could be relevant for several reasons. It might be linked to a postponement of marriage and parenthood in favour of career progression and economic advantage (Amuedo-Dorantes and Kimmel 2005; Uecker and Stokes 2008). Conversely, time spent in this status could also reflect the selection of economically less successful individuals into singlehood (Gibson-Davis, Edin, and McLanahan 2005; Jalovaara and Fasang 2017).

[5] The mean importance in the Boruta algorithm is the average significance of a feature in making accurate predictions, as measured by its contribution to reducing uncertainty or error in the model. This measure does not directly capture the effect size or the prevalence of specific features but is influenced by both. It highlights features that consistently improve the model's accuracy across different random forest trees, thus emphasising their overall predictive contribution rather than the magnitude or commonness of the patterns they represent.

Figure 3: Features selected as relevant predictors of personal net wealth in late working age, ranked by their mean importance according to the Boruta algorithm



Note: Feature importance is measured as z-scores. NMNC = never-married childless; NM1C = never-married with child(ren); MNC = married and childless, M1C = married with one child, M2C = married with two children, M3C = married with three children, M4C = married with four or more children, DSNC = divorced/separated and childless, DSNC = divorced/separated with child(ren), RMNC = remarried and childless, RMC = remarried with child(ren).
Source: Data are from the Socio-Economic Panel Survey v38 (2002, 2007, 2012, 2017).

Three other results are noteworthy when considering the features selected as relevant. First, several life-course variables related to 'divorced/separated with child(ren)' are among the top 10 wealth predictors. Experiencing the sub-sequence of marital dissolution after having three children ('M2C–M3C–DSC') is the third most important family life-course proxy for predicting wealth. The duration and the number of times being divorced/separated and having children ('DSC') are also important. Second, and related to the first point, the features related to marital dissolution all pertained to the dissolution of marriages involving parents, while features of marital dissolution among childless respondents were not identified as relevant wealth predictors. This likely points to the high vulnerability of parents if their marriage breaks down. Third, concerning the timing of family events over the life course, the Boruta algorithm mostly identified transitions at early ages as important predictors of personal wealth. This may suggest that family transitions at early or very early ages (up to age 24) lay the groundwork for wealth accumulation throughout the life course. For example, marrying or already having two children and being married between ages 20 and 24 ('MNC' and 'M2C') are among the top 5 wealth predictors. By contrast, family transitions in mid-to-late adulthood appear to play a less substantial role, with only a few selected features referring to transitions at or after the age of 30.

Overall, while all life-course concepts are among the selected features, it is the duration within the statuses of 'never-married without children' and 'never-married with child(ren)' that were deemed the most important wealth predictors.

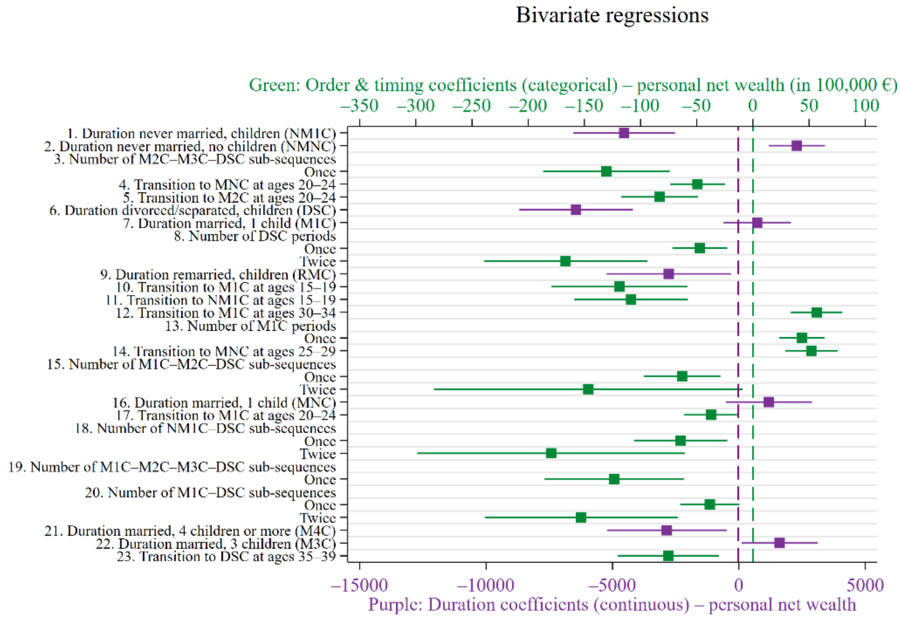
7. Association of relevant family life-course features and wealth in late working age

While the feature selection approach indicated which family life-course features are the most relevant predictors of personal wealth in late working age, it did not indicate the magnitude and direction of the association between relevant features and personal wealth. To further dive into this aspect and address our second research question, we moved to a regression approach. We started with bivariate regressions (23 regressions in total) and moved to stepwise multiple regressions that added up to 11 of the most relevant features into a single regression as explanatory variables for personal net wealth.

Figure 4 displays regression coefficients for the 23 bivariate associations. Note that we illustrate coefficients for duration features (in purple) on a different scale than coefficients for order and timing features (in green) due to the different variable types (i.e., continuous vs. categorical). As can be seen in Figure 4, most features are negatively associated with personal wealth in late working age. Negative associations are most often found for features relating to unmarried parenthood, marital separation, or comparatively early marital transitions with and without fertility transitions. Particularly strong negative associations can be found for features that refer to very early entry into parenthood (at ages 15–19, within or outside of marriage), the occurrence of serial marital separation with children, the duration spent as an unmarried parent after marital separation, and the duration spent as a never-married parent.[6] Figure 4 also highlights five features that are substantially positively associated with personal net wealth in late working age. Specifically, these features capture the duration of being never married without children, the duration as a married parent of three children, experiencing the state of being a married parent of one child, and relatively late transitions to first marriage without and with children.

[6] Note that 'never-married' may include a few cases of never-married cohabitation. However, as previously highlighted, never-married cohabitation was an uncommon occurrence and mostly short-lived experience for the cohort and context under study.

Figure 4: Bivariate regression results



Note: Whiskers indicate 95% confidence intervals. NMNC = never-married childless; NM1C = never-married with child(ren); MNC = married and childless, M1C = married with one child, M2C = married with two children, M3C = married with three children, M4C = married with four or more children, DSNC = divorced/separated and childless, DSNC = divorced/separated with child(ren), RMNC = remarried and childless, RMC = remarried with child(ren).
Source: Data are from the Socio-Economic Panel Survey v38 (2002, 2007, 2012, 2017). Weighted.

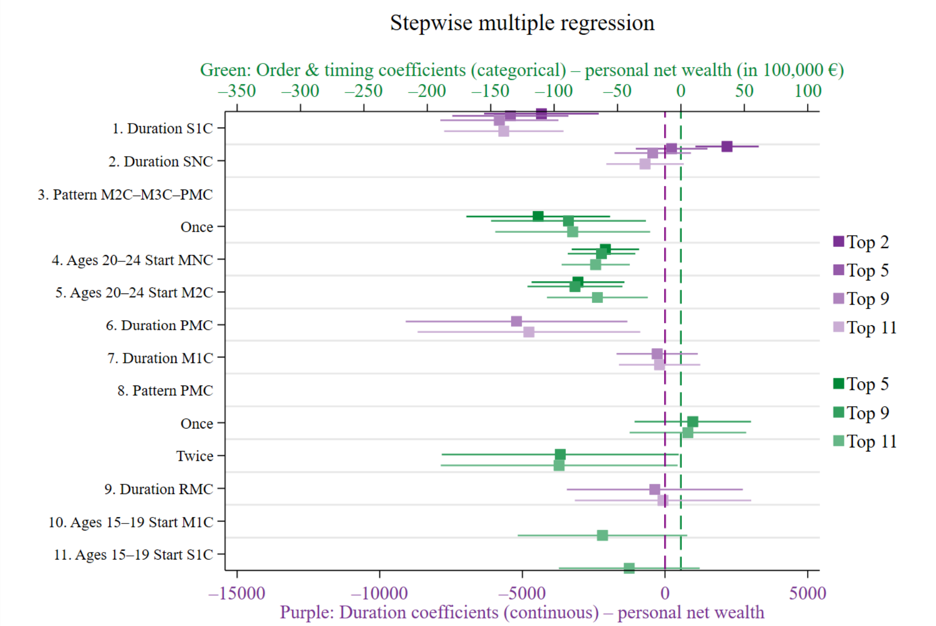
As a next step, we moved to multiple regressions that added features in a stepwise manner. For ease of comparison between the models, we illustrate the results in Figure 5 (see Table A-4 for the full regression results). We started with the two most relevant features – the duration spent as a never-married parent and the duration spent never-married without children – according to the mean importance of the feature selection analyses and moved on to the top 5, top 9, and finally top 11 features. We refrained from adding further features to the regression due to the decreasing importance of features and increasing multicollinearity issues.

Overall, results for these multiple regression models mirrored our bivariate results in Figure 4 and associations remained robust and in the expected direction. Some differences between the bivariate and the multiple regressions should be highlighted: the effect of the duration of never-married parenthood (‘NM1C’) is slightly more negative in the multiple than the bivariate result. Effects for experiencing the sub-sequence of marital dissolution after marriage with three children ('M2C–M3C–DSC') and the overall duration spent being separated/divorced ('DSC') are less negative in the multiple regression. And finally, the estimated effects for several features are smaller in magnitude overall and less precisely estimated, with confidence intervals that include zero (e.g., 'Number of DSC periods', 'Duration RMC', 'Transition to M1C at ages 15–19', 'Transition to NM1C at ages 15–19').

Focusing on the differences between the stepwise multiple models, we can see that some dependencies become visible as we start adding additional features. For instance, the association of the duration being never-married without children becomes less relevant and precisely estimated once we start adding the timing variables referring to whether the transition to first marriage – at that point of the transition, without children – takes place at ages 20 to 24 and whether the birth of a second child within marriage takes place within this age bracket. Thus, not adding the timing of marriage or any other features, the association between the years spent being never-married without children and wealth is ambiguous because a longer time being never-married without children likely also captures the positive association between late marriage and wealth – as can be seen in the bivariate associations (Figure 4).

Also, the effect of experiencing periods of being divorced/separated with children is likely absorbed by the duration of this pattern. To assess this, we re-ran the regression of the top 9 features dropping the duration indicator as a robustness check (results not shown but available upon request from the authors). Indeed, this leads to the expected results: periods of being divorced/separated becoming negatively correlated with wealth. Specifically, experiencing periods of being divorced/separated with children once or twice is associated with around €40,000 and €160,000 less personal wealth, respectively, once we no longer adjust for the duration in this status. This robustness check also reveals that the duration of remarried with children becomes more substantial with each year in this state and is associated with around €1,300 more wealth in later age. All other variables remain stable across the full 'top-9' model and our robustness check.

Figure 5: Stepwise multiple regression results



Note: Whiskers indicate 95% confidence intervals. NMNC = never-married childless; NM1C = never-married with child(ren); MNC = married and childless, M1C = married with one child, M2C = married with two children, M3C = married with three children, M4C = married with four or more children, DSNC = divorced/separated and childless, DSNC = divorced/separated with child(ren), RMNC = remarried and childless, RMC = remarried with child(ren). Weighted.
Source: Data are from the Socio-Economic Panel Survey v38 (2002, 2007, 2012, 2017).

8. Exploring potential variations by gender

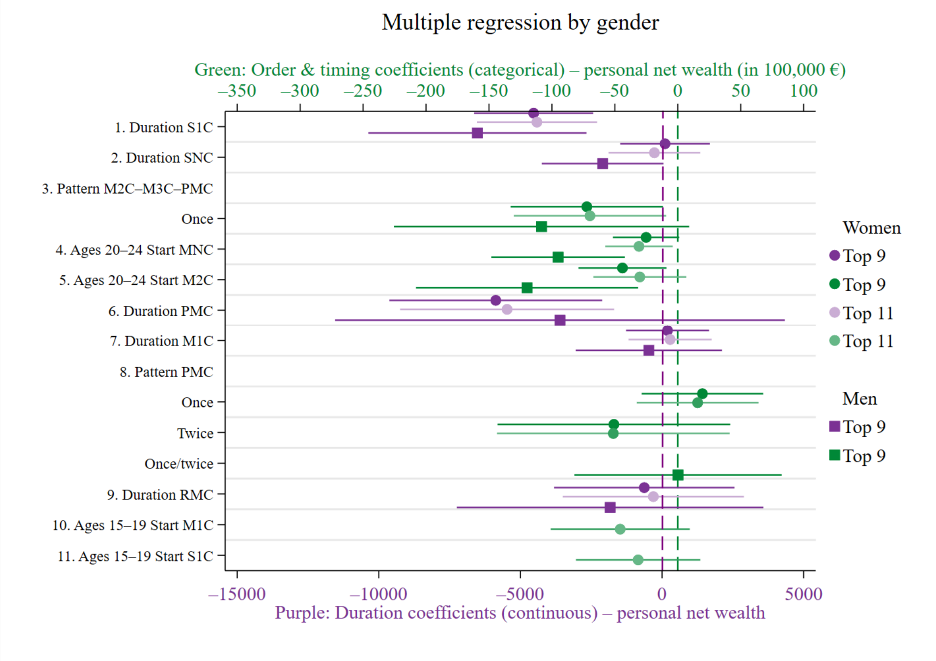
The average effects presented in the previous results section might differ radically for women and men. We thus move to our third research aim: exploring the extent to which the magnitude and direction of the association between specific features and personal wealth in late working age differ across genders. To this end, we ran gender-specific regression models. Note that cell sizes were not sufficient for all variables once disaggregated by gender (see Table A-5). For instance, transitions to never-married parenthood and having the first child within marriage at ages 15 to 19 were extremely uncommon for men but more common for women. Thus, we did not include these two features in the regressions for men. Similarly, only a few men reported being divorced/separated twice ('DSC'). We modified this variable for men to reflect experiencing this status at least once. Figure 6 shows the results of gender-specific regression models for the top 9 and 11 features for women and the top 9 features for men (see Table A-6 for the full regression results).

A range of gender differences are noteworthy (Figure 6). First, early marriage of childless individuals (age 20–24) and early transitions to a second child within marriage (age 20–24) were substantially more negatively associated with wealth for men, hinting at disadvantage or selectivity associated with comparatively early family transitions for men. Second, the time spent as never married without children was negatively associated with men's wealth while we found no substantial effect for women. This is in line with the idea that economically less well-off men with lower wealth accumulation potential are more likely to stay unmarried (Addo 2014; Carlson, McLanahan, and England 2004; Gibson-Davis, Edin, and McLanahan 2005). Third, the duration of being divorced/separated with children was slightly more negatively associated with wealth for women, highlighting the lasting disadvantage of marital dissolution for wealth accumulation for mothers especially (Kapelle 2022; Kapelle and Vidal 2022).

We found no substantial gender differences for the remaining family life-course features. This might seem surprising, at least for some features. For example, the duration in never-married parenthood was similarly negative for both men's and women's wealth, adjusting for other features and covariates. Overall, more women than men experienced this status and on average stayed longer in this status (Table A-3, Table A-5, and Table 8), in line with previous research (Walper, Entleitner-Phleps, and Langmeyer 2021). However, men who experience this state might be more highly selected than women, which may explain the similarly negative effects for women and men for this feature.

Underlying gender differences might have also biased our feature selection model because features that are oppositely associated with wealth by gender might cancel each other out. As a result, gender-specific features may be concealed as relevant wealth predictors in our main model. As a supplementary analysis, we addressed this issue by running our feature selection approach separately for women and men, resulting in analyses that focused on inequalities within each group rather than overall wealth inequality (see Table A-8 and Figure A-5 for the results).

Figure 6: Stepwise multiple regression results by gender



Note: Whiskers indicate 95% confidence intervals. NMNC = never-married childless; NM1C = never-married with child(ren); MNC = married and childless, M1C = married with one child, M2C = married with two children, M3C = married with three children, M4C = married with four or more children, DSNC = divorced/separated and childless, DSNC = divorced/separated with child(ren), RMNC = remarried and childless, RMC = remarried with child(ren). Weighted.
Source: Data are from the Socio-Economic Panel Survey v38 (2002, 2007, 2012, 2017).

In total, 13 and 14 features were deemed relevant in the separate models for women and men, respectively. Thus, a smaller set of features was selected as relevant in the gender-specific models than in our main model. However, the features that were selected were mostly also confirmed in the main model. Compared to our main model where features from all life-course dimensions were selected as relevant equally often, duration features were more often selected as relevant in both gender-specific models. The top five features in each gender group almost exclusively refer to the duration in certain states. Only among men was one order feature, periods spent in the status married with one child ('M1C'), also in the top 5. Hence, as for the main model, two duration features were deemed the most important in each gender model: for women this was the duration divorced/separated ('DSC') and the duration of never-married parenthood ('NM1C'). Overall features related to the life-course state of being a divorced/separated parent were more important wealth predictors for women than men. For men, the two most important features were duration in the state of being never-married without children ('NMNC') – ranked the second most important feature in the main model – followed by duration in remarriage with children ('RMC').

Thus, while some of the features that were selected in the main model appeared relevant to wealth inequalities in gender groups, the order of relevance differed between the main and gender-specific models. This also highlights that some features are more important for explaining within-group inequalities while being less important for explaining overall inequalities.

9. Conclusion and discussion

In this study, we explored which family life-course features (i.e., variables describing family life courses) are most relevant for personal net wealth at ages 50 to 59 among cohorts of West Germans born between 1943 and 1967. We delved into the complexity of family life courses, considering not only the occurrence of transitions or events but also their timing, order, and duration. Furthermore, we examined the strength and direction of the associations between relevant features and personal wealth. This investigation aimed to determine whether selected features predict wealth positively or negatively and the magnitude of these effects. Finally, we also explored the extent to which gender stratifies these processes.

Our theoretical and empirical approach was informed by notions embedded in the life-course framework. This framework has been used widely to predict individuals' later-life outcomes. Due to empirical limitations, life-course studies often focus on predictive factors at a single point in time and employ a limited set of predictors to represent entire life courses. However, the longer lives are studied, the more challenging it becomes to select a concise yet comprehensive set of predictors for later-life outcomes. To navigate these challenges, the current study adopted a novel, data-driven empirical approach. Using longitudinal (prospective and retrospective) data from the German Socio-Economic Panel, we first automatically extracted a broad set of family life-course features. Subsequently, we employed the Boruta algorithm to identify features that were statistically relevant predictors of personal wealth. Finally, we applied a regression framework to assess the direction and strength of each selected feature's association with wealth in late working age. This approach allowed us to incorporate theoretical perspectives on the complexity of family life courses as predictors of wealth accumulation, thereby reflecting these concepts within our empirical methodology and understanding wealth levels in late working age. Our main contribution to the literature lies in uncovering patterns and relationships between family life courses and later-life wealth that have not been pre-specified or constrained by existing theoretical frameworks or previous literature, thereby contributing to the development of novel insights into the link between family life courses and later-life wealth. Accordingly, our approach provides an empirical basis for identifying the most relevant components of family-life courses, which can be integrated into more structured, theory-driven studies. However, it is important to acknowledge that our approach is exploratory and does not establish causal links; therefore, the associations observed should be interpreted with caution, considering potential selection effects or reverse causality.

Overall, our methodological approach identified 23 features that were deemed relevant predictors of wealth and that were differently associated with wealth in the overall sample population, and to some degree by gender. A range of results are particularly noteworthy. First, the features that were deemed relevant wealth predictors were diverse. While previous research has predominately focused on whether a life-course transition occurred or not, our results highlight the importance of being more aware of the complexity of family life courses when considering the link between life courses and later-life outcomes. This diversity was also reflected in the fact that all three life-course dimensions – timing, order, and duration of events and transitions – were selected equally often as relevant.

Second, although all three dimensions appeared across the selected features, the duration spent in certain states – especially as a never-married parent or never-married without children – emerged as the most important aspect. This highlights that it is not only relevant whether and when adverse or beneficial transitions occur, but also how long individuals remain in these states. The particular significance of the duration in unmarried parenthood and unmarried singlehood underscores the cumulative wealth (dis)advantages faced over the life course.

Third, focusing on the timing of family transitions (i.e., age at transition), we showed that comparatively (very) early life-course events and transitions were deemed highly important wealth predictors for our cohort of interest. Regression results confirmed that these (very) early events and transitions were mostly negatively associated with wealth in late working age. This may hint at selection effects (i.e., economically less stable individuals transition earlier) but also at potential adverse effects for educational and career outcomes, with early transitions inhibiting prolonged education and career advancement (Amuedo-Dorantes and Kimmel 2005; Uecker and Stokes 2008). Thus, policies that aim at reducing wealth inequalities may want to focus particularly on the causes and consequences of early family transitions. This may even become more relevant for more recent cohorts where norms around the ‘appropriate’ age of family transitions have shifted and early family transitions may be deemed even more undesirable and selective.

Fourth, among the relevant 'sub-sequence' features, it was sequences that ended with 'being divorced/separated with children' that were most often selected. The duration spent in this state and whether the transition to this state took place at ages 35 to 39 were also deemed relevant. As illustrated in the regressions, features characterised by being separated or divorced negatively predicted wealth at a late working age. Thus, and in line with previous research, our results highlight the tremendous negative effects of marital dissolution on wealth outcomes (Kapelle 2022; Kapelle and Vidal 2022). However, our results may also hint at selection effects whereby financially stressed couples are more likely to experience divorce. Considering persistently high divorce rates, discussions and interventions need to focus on how the wealth penalties of divorce – particularly for parents and more so for mothers – can be reduced and economic self-reliance after divorce can be strengthened, particularly for individuals who have already experienced financial struggles during their marriage.

Finally, some gender differences prevailed. While early transitions were deemed particularly relevant, the definition of early differed between genders. For women it was transitions at ages 15 to 19, while for men it was transitions at ages 20 to 24. This is in line with observations that men tend to experience family transitions at slightly older ages than women (Ortega 2014). Additionally, a longer time spent as never-married without children was identified as particularly adverse for men, hinting at selectivity in the marriage market of economically more successful men into marriage (Xie et al. 2003). Lastly, life-course experiences related to the dissolution of parents' marriages were more important wealth predictors for mothers than fathers, and only negatively associated with their later wealth. This highlights the detrimental consequences mothers in particular face after divorce regarding their financial safety net.

Although direct comparison across different methodological approaches is not advisable due to their distinct aims, our study complements prior research by adding relevant detail. First, we highlight the importance of carefully considering a broader set of life-course indicators than previous studies, which often rely on a limited number of time-in-point measures, supplemented by only a few life-course-relevant indicators (e.g., overall duration of marriage). Second, we extend research using sequence and cluster analyses, as seen in Kapelle and Vidal (2022). Their approach, which is constrained by the number of clusters, results in some diverse clusters such as 3 characterized by divorce but differing in fertility and remarriage rates. These clusters cannot fully differentiate the specific order, duration, or timing of transitions. For example, our study finds that individuals experiencing union dissolution after high-fertility marital unions are particularly vulnerable economically. We also highlight the relevance of serial marital dissolutions, which Kapelle and Vidal (2022) do not capture. Similarly, while they group unmarried parenthood into a single cluster, our approach shows that both the overall duration in this state and the age at which the transition occurred are crucial. Overall, our approach should be viewed as a complementary tool in the methodological toolbox, rather than a replacement for previous methods.

Five notable limitations of the current study need to be highlighted. First, the SOEP respondents' cohabitation histories were not recorded retrospectively. Although we argue that the absence of cohabitation data in our sequences is not a significant concern for our study, given the social undesirability and discouragement of cohabitation within the studied cohort and context (Le Goff 2002), it should be noted that cohabitation may be a crucial factor in future research, particularly when examining more recent cohorts or more liberal contexts. Second, and connected to the previous point, although our study identifies wealth-relevant family life-course features for cohorts of West Germans born between 1943 and 1967, it remains unclear whether the identified features are similarly important wealth predictors for later or earlier cohorts or other contexts. This is an important avenue for future research. Third, we relied on retrospective fertility histories that capture whether a respondent had ever given birth to or fathered a child. However, these histories do not provide information on the residential status of those children. Nevertheless, it can be argued that, irrespective of a child's residential status, having mothered or fathered a child entails certain financial responsibilities that respondents without children would not incur (e.g., child maintenance, schooling costs, healthcare expenses, and childcare fees). Fourth, at least for the time being our methodological approach is limited in the number of life-course channels it can consider. Although our focus was on the family as a context for stratification, it is important to acknowledge that family life-courses are often closely intertwined with labour-market trajectories. We were unable to explicitly explore this interconnectedness in this study due to methodological restrictions, but as methods advance further future research may want to consider this interconnectedness more thoroughly. Finally, survey data on wealth, including those used in our study, are subject to several limitations commonly found in previous research. Issues of misreporting or nonresponse are frequent due to the sensitivity and complexity of such data (e.g., Grabka and Westermeier 2015; Riphahn and Serfling 2005). Additionally, respondents are required to report their share of potentially jointly held wealth for the collection of personal wealth data. This process can be susceptible to errors, as the clarity of property rights may not be apparent to each individual, and their perceived ownership may not correspond to legal ownership (Joseph and Rowlingson 2012). However, the SOEP data are uniquely valuable in providing full wealth information at the personal level, coupled with detailed family histories. This has enabled us to examine gender differences more appropriately than would have been possible with other survey data.

Overall, this study provides a thorough description of (1) which family life-course features are relevant for wealth levels in late working age and thus the accumulation of wealth over the life course for cohorts of West Germans born between 1943 and 1967,

(2) the magnitude and direction in which these features contribute to explaining wealth in late working ages, and finally (3) the gendered nature of the association between relevant life-course features and wealth. Understanding these aspects is crucial for the expanding body of literature on wealth inequality, family dynamics, and gender inequalities. While our study does not investigate causal relationships, our findings hint at potential causal connections and thus can contribute to the understanding of the potential causes and consequences of wealth inequalities. Our research results offer policy discussions valuable insights on ways to mitigate rising wealth inequalities and create an environment that enables individuals to build an economic safety net. Ensuring economic self-reliance throughout the life course independent of family ties, particularly in older age, is of tremendous importance for current and upcoming generations in the context of an ageing population, rising economic inequalities, and increasing welfare expenditures.

10. Acknowledgements and funding

The manuscript greatly benefited from discussions with participants of several conferences held in 2023, including the Population Association of America's (PAA) annual meeting (12th to 15th April 2023, New Orleans), the spring meeting of the Research Committee 28 on Social Stratification and Mobility (RC28) of the International Sociological Association (ISA) (24th to 26th May 2023, Paris), the annual conference of the European Consortium for Sociological Research (ECSR) (18th to 20th September 2023, Prague), and the annual conference of the Society for Longitudinal and Life course Studies (SLLS) (9th to 11th October 2023, Munich). Additionally, we would like to thank participants of the research colloquium at the State Institute for Family Research at the University of Bamberg (ifb) for comments and suggestions that greatly improved the study.

This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation program, Grant Agreement 681546. Carla Rowold acknowledges financial support from the Economic and Social Research Council grant ES/P0000649/1, from Nuffield College, and from grants to the Max Planck–University of Helsinki Center from the Max Planck Society, Jane and Aatos Erkko Foundation, Faculty of Social Sciences at the University of Helsinki, and Cities of Helsinki, Vantaa, and Espoo. The views expressed herein are those of the authors and are not necessarily those of the funding bodies.

References

Abbott, A. (2001). *Time matters: On theory and method*. Chicago, IL: University of Chicago Press.

Addo, F.R. (2014). Debt, cohabitation, and marriage in young adulthood. *Demography* 51(5): 1677–1701. doi:10.1007/s13524-014-0333-6.

Addo, F.R. and Lichter, D.T. (2013). Marriage, marital history, and black–white wealth differentials among older women. *Journal of Marriage and Family* 75(2): 342–362. doi:10.1111/jomf.12007.

Alessie, R., Lusardi, A., and Aldershof, T. (1997). Income and wealth over the life cycle: Evidence from panel data. *Review of Income and Wealth* 43(1): 1–32. doi:10.1111/j.1475-4991.1997.tb00198.x.

Amuedo-Dorantes, C. and Kimmel, J. (2005). The motherhood wage gap for women in the United States: The importance of college and fertility delay. *Review of Economics of the Household* 3(1): 17–48. doi:10.1007/s11150-004-0978-9.

Atkinson, A.B. (1971). The distribution of wealth and the individual life-cycle. *Oxford Economic Papers* 23(2): 239–254. doi:10.1093/oxfordjournals.oep.a041192.

Bernardi, F., Boertien, D., and Geven, K. (2018). Childhood family structure and the accumulation of wealth across the life course. *Journal of Marriage and Family* 81(1): 230–247. doi:10.1111/jomf.12523.

Bernardi, L., Huinink, J., and Settersten, R.A. (2019). The life course cube: A tool for studying lives. *Advances in Life Course Research* 41(100258). doi:10.1016/j.alcr.2018.11.004.

Bolano, D. and Studer, M. (2020). The link between previous life trajectories and a later life outcome: A feature selection approach. Lausanne: LIVES Swiss Centre of Expertise in Life Course Research. (LIVES Working paper 82). doi:10.12682/lives.2296-1658.2020.82.

Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J.M., and Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Information Sciences* 282: 111–135. doi:doi:10.1016/j.ins.2014.05.042.

Bradbury, B. (2011). Child costs. In: Ben-Arieh, A., Casas, F., Frønes, I., and Korbin, J.E. (eds.). *Handbook of child well-being*. Dordrecht: Springer: 1483–1507. doi:10.1007/978-90-481-9063-8_60.

Budig, M.J. and England, P. (2001). The wage penalty for motherhood. *American Sociological Review* 66(2): 204–225. doi:10.2307/2657415.

Budig, M.J., Misra, J., and Boeckmann, I. (2012). The motherhood penalty in cross-national perspective: The importance of work–family policies and cultural attitudes. *Social Politics: International Studies in Gender, State and Society* 19(2): 163–193. doi:10.1093/sp/jxs006.

Carlson, M., McLanahan, S., and England, P. (2004). Union formation in fragile families. *Demography* 41(2): 237–261. doi:10.1353/dem.2004.0012.

Chang, M.L. (2010). *Shortchanged: Why women have less wealth and what can be done about it*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195367690.001.0001.

Christl, M., De Poli, S., and Ivaškaitė-Tamošiūnė, V. (2023). Does it pay to say "I do"? Marriage bonuses and penalties across the EU. *Journal of European Social Policy* 33(3): 317–336. doi:10.1177/09589287231159492.

Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *The Journals of Gerontology: Series B* 58(6): S327–S337. doi:10.1093/geronb/58.6.S327.

Dannefer, D. (2018). Systemic and reflexive: Foundations of cumulative dis/advantage and life-course processes. *The Journals of Gerontology: Series B* 75(6): 1249–1263. doi:10.1093/geronb/gby118.

Degenhardt, F., Seifert, S., and Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics* 20(2), 492–503. doi:10.1093/bib/bbx124.

Dew, J. (2011). The association between consumer debt and the likelihood of divorce. *Journal of Family and Economic Issues* 32(4): 554–565. doi:10.1007/s10834-011-9274-z.

Dew, J., Britt, S., and Huston, S. J. (2012). Examining the relationship between financial issues and divorce. *Family Relations* 61(4): 615–628. doi:10.1111/j.1741-3729.2012.00715.x.

Di Nallo, A. (2019). Gender gap in repartnering: The role of parental status and custodial arrangements. *Journal of Marriage and Family* 81(1): 59–78. doi:10.1111/jomf.12527.

Eads, A. and Tach, L. (2016). Wealth and inequality in the stability of romantic relationships. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(6): 197–224. doi:10.7758/rsf.2016.2.6.10.

Eickmeyer, K.J., Manning, W.D., and Brown, S.L. (2019). What's mine is ours? Income pooling in American families. *Journal of Marriage and Family* 81(4): 968–978. doi:10.1111/jomf.12565.

Gibson-Davis, C.M., Edin, K., and McLanahan, S. (2005). High hopes but even higher expectations: The retreat from marriage among low-income couples. *Journal of Marriage and Family* 67(5): 1301–1312. doi:10.1111/j.1741-3737.2005.00218.x.

Gibson-Davis, C.M., Gassman-Pines, A., and Lehrman, R. (2018). 'His' and 'hers': Meeting the economic bar to marriage. *Demography* 55(6): 2321–2343. doi:10.1007/s13524-018-0726-z.

Goebel, J., Grabka, M.M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik* 239(2): 345–360. doi:10.1515/jbnst-2018-0022.

Gough, M. and Noonan, M. (2013). A review of the motherhood wage penalty in the United States. *Sociology Compass* 7(4): 328–342. doi:doi:10.1111/soc4.12031.

Grabka, M.M. and Westermeier, C. (2015). Editing and multiple imputation of item non-response in the wealth module of the German Socio-Economic Panel. (SOEP Survey Papers, Series C – Data Documentation 272). Berlin: DIW. http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0272.pdf.

Grinstein-Weiss, M., Hun Yeo, Y., Zhan, M., and Charles, P. (2008). Asset holding and net worth among households with children: Differences by household type. *Children and Youth Services Review* 30(1): 62–78. doi:10.1016/j.childyouth.2007.06.005.

Halpern-Manners, A., Warren, J.R., Raymo, J.M., and Nicholson, D.A. (2015). The impact of work and family life histories on economic well-being at older ages. *Social Forces* 93(4): 1369–1396. doi:10.1093/sf/sov005.

Hamplová, D., Le Bourdais, C., and Lapierre-Adamcyk, É. (2014). Is the cohabitation–marriage gap in money pooling universal? *Journal of Marriage and Family* 76(5): 983–997. doi:10.1111/jomf.12138.

Hao, L. (1996). Family structure, private transfers, and the economic well-being of families with children. *Social Forces* 75(1): 269–292. doi:10.2307/2580765.

Haupt, A. and Strauß, S. (2022). Long-term trends in the gender income gap within couples: West Germany, 1978–2011. *Social Politics: International Studies in Gender, State and Society* 29(3): 980–1008. doi:10.1093/sp/jxac019.

Hurd, M.D. (2002). Portfolio holdings of the elderly. In: Guiso, L., Haliassos, M., and Jappelli, T. (eds.). *Household portfolios*. Cambridge, MA: MIT Press: 431–472. doi:10.7551/mitpress/3568.003.0016.

Jalovaara, M., and Fasang, A. E. (2017). From never partnered to serial cohabitors: Union trajectories to childlessness. *Demographic Research* 36(55): 1703–1720. doi:10.4054/DemRes.2017.36.55.

Joseph, R. and Rowlingson, K. (2012). Her house, his pension? The division of assets among (ex-)couples and the role of policy. *Social Policy and Society* 11(1): 69–80. doi:10.1017/S147474641100042X.

Kapelle, N. (2022). Time cannot heal all wounds: Wealth trajectories of divorcees and the married. *Journal of Marriage and Family* 84(2): 592–611. doi:10.1111/jomf.12824.

Kapelle, N. and Baxter, J. (2021). Marital dissolution and personal wealth: Examining gendered trends across the dissolution process. *Journal of Marriage and Family* 83(1): 243–259. doi:10.1111/jomf.12707.

Kapelle, N. and Lersch, P.M. (2020). The accumulation of wealth in marriage: Over-time change and within-couple inequalities. *European Sociological Review* 36(4): 580–593. doi:10.1093/esr/jcaa006.

Kapelle, N. and Vidal, S. (2022). Heterogeneity in family life course patterns and intra-cohort wealth disparities in late working age. *European Journal of Population* 38(1): 59–92. doi:10.1007/s10680-021-09601-4.

Kapelle, N. and Weiland, A.P. (2025). Intra-couple gaps in retirees' financial resources: Their extent and predictors across Eastern and Western Germany. *European Societies* Online first. doi:10.1162/euso_a_00013.

Keister, L.A. and Moller, S. (2000). Wealth inequality in the United States. *Annual Review of Sociology* 26: 63–81. doi:10.1146/annurev.soc.26.1.63.

Killewald, A., Pfeffer, F.T., and Schachner, J.N. (2017). Wealth inequality and accumulation. *Annual Review of Sociology* 43(1): 379–404. doi:10.1146/annurev-soc-060116-053331.

Knoll, M.A.Z., Tamborini, C.R., and Whitman, K. (2012). I do … want to save: Marriage and retirement savings in young households. *Journal of Marriage and Family* 74(1): 86–100. doi:10.1111/j.1741-3737.2011.00877.x.

Kursa, M.B., Jankowski, A., and Rudnicki, W.R. (2010). Boruta – A system for feature selection. *Fundamenta Informaticae* 101: 271–285. doi:10.3233/FI-2010-288.

Kursa, M.B. and Rudnicki, W.R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software* 36(11): 1–13. doi:10.18637/jss.v036.i11.

Lanau, A. (2023). First the children, then the employed: Deprivation and intra-household inequality in Europe. *Journal of Poverty* 27(4): 331–349. doi:10.1080/10875549.2022.2065561.

Le Goff, J.-M. (2002). Cohabiting unions in France and West Germany: Transitions to first birth and first marriage. *Demographic Research* 7(18): 593–624. doi:10.4054/DemRes.2002.7.18.

Leopold, T. and Schneider, T. (2011). Family events and the timing of intergenerational transfers. *Social Forces* 90(2): 595–616. doi:10.1093/sf/sor030.

Lersch, P.M., Jacob, M., and Hank, K. (2017). Parenthood, gender, and personal wealth. *European Sociological Review* 33(3): 410–422. doi:10.1093/esr/jcx046.

Lersch, P.M. and Vidal, S. (2014). Falling out of love and down the housing ladder: A longitudinal analysis of marital separation and home ownership. *European Sociological Review* 30(4): 512–524. doi:10.1093/esr/jcu055.

Lusardi, A., Cossa, R., and Krupka, E.L. (2001). Savings of young parents. *The Journal of Human Resources* 36(4): 762–794. doi:10.2307/3069641.

Maldonado, L.C. and Nieuwenhuis, R. (2015). Family policies and single parent poverty in 18 OECD countries, 1978–2008. *Community, Work and Family* 18(4): 395–415. doi:10.1080/13668803.2015.1080661.

Manning, W.D., Stewart, S.D., and Smock, P.J. (2003). The complexity of fathers' parenting responsibilities and involvement with nonresident children. *Journal of Family Issues* 24(5): 645–667. doi:10.1177/0192513x03252573.

Modigliani, F. (1988). The role of intergenerational transfers and life cycle saving in the accumulation of wealth. *The Journal of Economic Perspectives* 2(2): 15–40. doi:10.1257/jep.2.2.15.

O'Rand, A.M. (1996). The precious and the precocious: Understanding cumulative disadvantage and cumulative advantage over the life course. *Gerontologist* 36(2): 230–238. doi:10.1093/geront/36.2.230.

Ortega, J.A. (2014). A characterization of world union patterns at the national and regional level. *Population Research and Policy Review* 33(2): 161–188. doi:10.1007/s11113-013-9301-x.

Ozawa, M.N. and Lee, Y. (2006). The net worth of female-headed households: A comparison to other types of households. *Family Relations* 55(1): 132–145. doi:10.1111/j.1741-3729.2006.00362.x.

Riphahn, R.T. and Serfling, O. (2005). Item non-response on income and wealth questions. *Empirical Economics* 30(2): 521–538. doi:10.1007/s00181-005-0247-7.

Rowold, C., Struffolino, E., and Fasang, A.E. (2024). Life-course-sensitive analysis of group inequalities: Combining sequence analysis with the Kitagawa–Oaxaca–Blinder decomposition. *Sociological Methods and Research* Online first. doi:10.1177/00491241231224226.

Saeys, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics* 23(19): 2507–2517. doi:10.1093/bioinformatics/btm344.

Schechtl, M. and Kapelle, N. (2024). The male marital earnings premium contextualized: Longitudinal evidence from the United States, Germany, and the United Kingdom. *Journal of Marriage and Family* 86(1): 176–198. doi:doi:10.1111/jomf.12937.

Schneider, D. (2011). Wealth and the marital divide. *American Journal of Sociology* 117(2): 627–667. doi:10.1086/661594.

Sierminska, E. (2018). The 'wealth-being' of single parents. In: Nieuwenhuis, R. and Maldonado, L.C. (eds.). *The triple bind of single-parent families: Resources, employment and policies to improve wellbeing*. Bristol: Bristol University Press: 51–80. doi:10.2307/j.ctt2204rvq.9.

Smock, P.J., Manning, W.D., and Porter, M. (2005). "Everything's there except money": How money shapes decisions to marry among cohabitors. *Journal of Marriage and Family* 67(3): 680–696. doi:10.1111/j.1741-3737.2005.00162.x.

Spilerman, S. (2000). Wealth and stratification processes. *Annual Review of Sociology* 26(1): 497–524. doi:10.1146/annurev.soc.26.1.497.

Studer, M. (2013). WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R. (LIVES Working Papers 24*)*. Lausanne: LIVES.

Thomas, M.J. and Mulder, C.H. (2016). Partnership patterns and homeownership: A cross-country comparison of Germany, the Netherlands and the United Kingdom. *Housing Studies* 31(8): 935–963. doi:10.1080/02673037.2016.1164832.

Trappe, H., Pollmann-Schult, M., and Schmitt, C. (2015). The rise and decline of the male breadwinner model: Institutional underpinnings and future expectations. *European Sociological Review* 31(2): 230–242. doi:10.1093/esr/jcv015.

Uecker, J.E. and Stokes, C.E. (2008). Early marriage in the United States. *Journal of Marriage and Family* 70(4): 835–846. doi:doi:10.1111/j.1741-3737.2008.00530.x.

Ulker, A. (2008). Wealth holdings and portfolio allocation of the elderly: The role of marital history. *Journal of Family and Economic Issues* 30(1): 90–108. doi:10.1007/s10834-008-9139-2.

Upchurch, D.M., Lillard, L.A., and Panis, C.W.A. (2002). Nonmarital childbearing: Influences of education, marriage, and fertility. *Demography* 39(2): 311–329. doi:10.1353/dem.2002.0020.

Van Winkle, Z. and Fasang, A.E. (2021). The complexity of employment and family life courses across 20th century Europe. More evidence for larger cross-national differences but little change across 1916–1966 birth cohorts. *Demographic Research* 44(32): 775–810. doi:10.4054/DemRes.2021.44.32.

Vogler, C., Lyonette, C., and Wiggins, R.D. (2008). Money, power and spending decisions in intimate relationships. *The Sociological Review* 56(1): 117–143. doi:10.1111/j.1467-954X.2008.00779.x.

Walper, S., Entleitner-Phleps, C., and Langmeyer, A.N. (2021). Shared physical custody after parental separation: Evidence from Germany. In: Bernardi, L. and Mortelmans, D. (eds.). *Shared physical custody: Interdisciplinary insights in child custody arrangements* (285–308). Cham: Springer International Publishing. doi:10.1007/978-3-030-68479-2_13.

Wilmoth, J. and Koso, G. (2002). Does marital history matter? Marital status and wealth outcomes among preretirement adults. *Journal of Marriage and Family* 64(1): 254–268. doi:10.1111/j.1741-3737.2002.00254.x.

Wilson, B.F. and Clarke, S.C. (1992). Remarriages: A demographic profile. *Journal of Family Issues* 13(2): 123–141. doi:10.1177/019251392013002001.

Xie, Y., Raymo, J.M., Goyette, K., and Thornton, A. (2003). Economic potential and entry into marriage and cohabitation. *Demography* 40(2): 351–367. doi:10.1353/dem.2003.0019.

Zissimopoulos, J.M., Karney, B.R., and Rauer, A.J. (2015). Marriage and economic well being at older ages. *Review of Economics of the Household* 13(1): 1–35. doi:10.1007/s11150-013-9205-x.

Appendix

Feature selection using the Boruta algorithm

Feature selection is a critical process in the field of machine learning and data analysis. It involves automatically selecting a subset of 'relevant' features (i.e., variables) from the input data for use in model construction, tailored to the specific research question one aims to answer. The primary goal of feature selection is to improve the model's performance by eliminating unnecessary noise from the input data, resulting in a simpler, faster, more effective, and potentially more accurate model (Saeys, Inza, and Larrañaga 2007).

There are two main approaches to feature selection. The 'minimum optimum feature selection approach' aims to find the minimal optimal subset of features that are sufficient for a model to predict the target variables. The 'all relevant feature selection approach' seeks to identify all relevant features contributing to the prediction of the target variable, offering insights into potential causal relationships for observed behaviours (Degenhardt, Seifert, and Szymczak 2019).

Three distinct types of feature selection methods can be distinguished: filter, wrapper, and embedded. Wrapper methods are particularly advantageous in their approach, as they directly analyse how subsets of variables perform within a specific predictive model, optimizing the selection based on actual model performance. This contrasts with filter methods, which independently assess features using statistical measures but do not account for model-specific interactions. While embedded methods like LASSO integrate feature selection into model training, wrapper methods offer a more focused evaluation, specifically optimising features in relation to the model's predictive accuracy, thus potentially yielding superior results in certain applications (Bolón-Canedo et al. 2014; Saeys, Inza, and Larrañaga 2007).

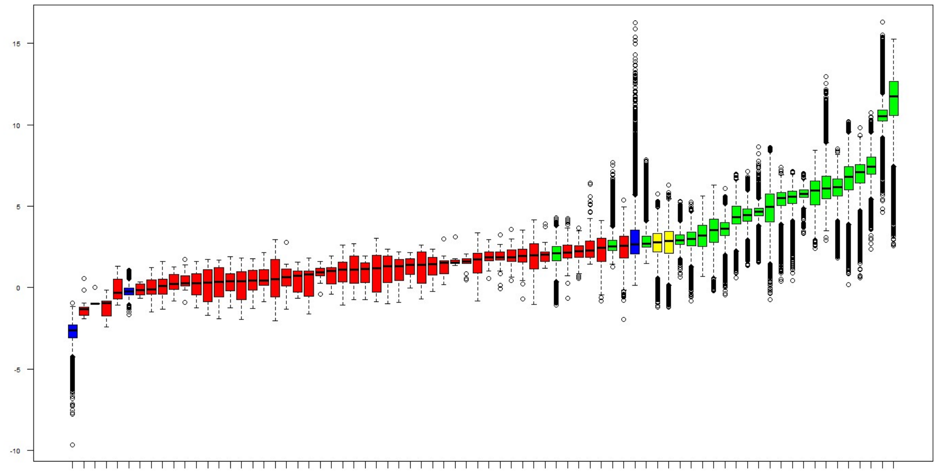
In our study we applied the Boruta algorithm, an 'all relevant feature' wrapper approach built around a random forest classification (Kursa, Jankowski, and Rudnicki 2010; Kursa and Rudnicki 2010). The Boruta algorithm aims to capture all relevant features in our dataset concerning our outcome variable – personal net wealth. This algorithm has a range of advantages – as already outlined in the main manuscript – that make it particularly suitable for our present study: its comprehensive approach ensures that no significant predictor is overlooked, its robustness against overfitting is crucial for handling our complex data, and its ability to deal with non-linear relationships and interactions is key, given the complex and intertwined ways life-course variables might be associated with wealth. Additionally, it has recently been shown to perform best among a set of selected algorithms, especially for low-dimensional data sets (Degenhardt, Seifert, and Szymczak 2019).

The Boruta algorithm is a multi-step process (Kursa, Jankowski, and Rudnicki 2010; Kursa and Rudnicki 2010). First, it expands the dataset by adding 'shuffled' duplicates of all features, effectively doubling the dataset's number of features. These added features, known as shadow features, are randomised versions of the original features and have no meaningful relationship with the response variable. Second, a random forest classifier is trained on this expanded dataset, and feature importance is evaluated. A random forest classifier is a versatile machine-learning model that constructs multiple decision trees and aggregates their results for improved accuracy and control of overfitting. Essentially, the results from multiple feature importance evaluations are combined to decide on the relevance of a feature, reducing the likelihood of error. Feature importance is commonly assessed using Mean Decrease Accuracy, expressed in standardized Z-scores (i.e., the mean of accuracy loss divided by the standard deviation of accuracy loss). The Maximum Z-Score Among the Shadow Attributes (MZSA) is calculated and used to determine whether each original feature's Z-score exceeds the MZSA. Original features with Z-scores significantly exceeding those of the MZSA are deemed important and retained, while the rest are pruned. Finally, the algorithm iterates the second step, removing 'unimportant' features across different iterations, and stops either when all features are deemed significant or insignificant, or when a specific number of iterations is reached.

Our Boruta algorithm results are depicted in Figure A-1. The Figure presents boxplots of feature importance. The blue boxplots represent the minimum, average, and maximum Z-scores of the shadow features. Meanwhile, the red, yellow, and green boxplots correspond to the Z-scores of rejected, tentative, and confirmed features, respectively.[7] A total of 23 features (depicted in green) are determined to have higher variable importance than the best-performing shadow feature, as indicated by Z-scores significantly exceeding those of the MZSA. These 23 features are considered in detail in the main manuscript.

[7] Note that the boxplots are ordered based on their median Z-score, although decisions are based on whether features scored significantly higher or lower than the MZSA in a run.

Figure A-1: Boruta feature selection results sorted by median variable importance



Note: Feature importance is measured as z-scores.
Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).

Description of steps of manual reduction to subset of features

Our analytical approach takes several steps, as outlined in Figure 2 in the manuscript. In step 1 of our approach, we first automatically extract all possible family life-course features relating to order, timing, and duration. Next, we conduct a manual selection to reduce the set of 205 automatically extracted features to a parsimonious but theoretically relevant subset of features that we can then use in the Boruta algorithm. To this end, we apply the following restrictions:

1. Remove constant indicators (standard deviation = 0; 1 variable).
2. Remove variables that occur for less than 2% of all sequences (40 variables).
3. Remove highly correlated variables based on theoretical consideration (93 variables).
   a. For variables referring to the order of family life-course events and transitions, features that capture a different number of children are correlated. Hence, we drop features with ‘gaps’. For example, because we have the sub-sequence MC1–MC2–MC3, we can drop MC1–MC3 because it is explicitly covered in the sub-sequence MC1–MC2–MC3.

b. We drop 'order' variables starting with being never married because these features are all highly correlated and not theoretically interesting since all family life courses start with 'being never-married'.
c. We also drop 'order' variables where the state of 'divorced/separated' is missing in the sequence between any state and 'remarried'. For example, we capture the sub-sequence MC1–DSC–RMC and as a result can drop MC1–RMC.
d. We drop other 'order' variables covering a sub-sequence of full sequence that are highly correlated with the full sequences.
e. And finally, we drop other 'order' variables that are still highly correlated >.9 but do not apply to the rules above

Additional tables

Table A-1: Summary statistics: Personal net wealth in 10,000 euros

		Mean	Median	p25	p75	Min	Max	N
Personal net wealth	Total	16.19	9.39	0.98	20.39	–113.11	425.81	5,702
	Men	18.47	10.49	1.55	22.06	–113.11	409.80	2,851
	Women	13.62	7.94	0.60	18.21	–92.62	425.81	2,851

Note: Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017). Weighted.

Table A-2: Boruta feature selection results

Features	Mean imp.	Min. imp.	Median imp.	Max. imp.	norm Hits
1. Duration never married, children (NM1C)	11.42	2.57	11.76	15.27	0.9942
2. Duration never married, no children (NMNC)	10.67	4.61	10.55	16.31	0.9958
3. Number of M2C–M3C–DSC sub-sequences	7.48	2.36	7.44	10.74	0.9805
4. Transition to MNC at ages 20–24	6.89	0.58	7.09	9.81	0.9622
5. Transition to M2C at ages 20–24	6.62	0.20	6.79	10.20	0.9525
6. Duration divorced/separated, children (DSC)	6.21	2.90	6.11	12.95	0.9563
7. Duration married, 1 child (M1C)	6.08	1.82	6.17	8.53	0.9506
8. Number of DSC periods	5.80	2.41	5.97	8.45	0.9418
9. Duration remarried, children (RMC)	5.77	3.31	5.77	7.01	0.9519
10. Transition to M1C at ages 15–19	5.47	0.42	5.59	7.14	0.9294
11. Transition to NM1C at ages 15–19	5.33	0.39	5.48	7.40	0.9177
12. Transition to M1C at ages 30–34	4.85	–0.73	4.95	8.60	0.8572
13. Number of M1C periods	4.60	1.58	4.66	8.65	0.8849
14. Transition to MNC at ages 25–29	4.45	1.35	4.47	7.13	0.8690
15. Number of M1C–M2C–DSC sub-sequences	4.44	0.62	4.34	6.99	0.8529
16. Duration married, no children (MNC)	3.56	–0.46	3.61	6.08	0.7264
17. Transition to M1C at ages 20–24	3.47	–0.18	3.53	6.32	0.6667
18. Number of NM1C–DSC sub-sequences	3.20	0.70	3.18	5.64	0.6234
19. Number of M1C–M2C–M3C–DSC sub-sequences	3.20	1.50	2.70	7.87	0.5773
20. Number of M1C–DSC sub-sequences	2.97	0.64	2.91	5.30	0.5887
21. Duration married, 4 children or more (M4C)	2.95	–0.84	3.01	5.23	0.5760
22. Duration married, 3 children (M3C)	2.74	1.26	2.52	7.70	0.4900
23. Transition to DSC at ages 35–39	2.04	–1.08	2.12	4.30	0.2820

Note: Mean, Min., Median, and Max. refer to the mean, minimum, median, and maximum importance of the estimated models. Hits refer to the share each feature had a higher importance than the MZSA. Feature importance is measured as z-scores.
Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).

Table A-3: Proportions for order and timing variables and averages for duration variables across all selected family life-course features, by gender

Features			Men	Women
Order (proportion)	Number of M2C–M3C–DSC sub-sequences	Never	98.11	96.77
		Once	1.89	3.23
	Number of DSC periods	Never	86.04	79.34
		Once	12.91	18.66
		At least twice	1.05	2.00
	Number of M1C periods	Never	28.17	25.96
		Once	71.83	74.04
	Number of M1C–M2C–DSC sub-sequences	Never	94.18	90.35
		Once	5.51	9.12
		Twice	0.32	0.53
	Number of NM1C–DSC sub-sequences	Never	95.62	93.62
		Once	3.86	5.79
		At least twice	0.53	0.60
	Number of M1C–M2C–M3C–DSC sub-sequences	Never	98.35	97.44
		At least once	1.65	2.56
	Number of M1C–DSC sub-sequences	Never	88.60	82.95
		Once	10.56	15.71
		At least twice	0.84	1.33
Timing (proportion)	Transition to MNC at ages 20–24	No	90.00	78.36
		Yes	10.00	21.64
	Transition to M2C at ages 20–24	No	96.81	88.64
		Yes	3.19	11.36
	Transition to M1C at ages 15–19	No	99.75	95.83
		Yes	0.25	4.17
	Transition to NM1C at ages 15–19	No	98.81	94.77
		Yes	1.19	5.23
	Transition to M1C at ages 30–34	No	77.31	85.48
		Yes	22.69	14.52
	Transition to MNC at ages 25–29	No	79.80	84.71
		Yes	20.20	15.29
	Transition to M1C at ages 20–24	No	89.62	76.53
		Yes	10.38	23.47
	Transition to DSC at ages 35–39	No	96.60	95.26
		Yes	3.40	4.74
Duration in years (mean)	Duration never married, children (NM1C)		1.19	1.56
	Duration never married, no children (NMNC)		15.63	11.58
	Duration divorced/separated, children (DSC)		0.96	1.85
	Duration married, 1 child (M1C)		4.30	5.06
	Duration remarried, children (RMC)		0.88	1.00
	Duration married, no children (MNC)		2.61	2.75
	Duration married, 4 children or more (M4C)		0.79	1.02
	Duration married, 3 children (M3C)		2.49	2.78

Notes: Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).

Table A-4: Multiple regression: Adding features as explanatory variables in a stepwise fashion to predict personal net wealth in late working age

	Top 2 B/(SE)/[CIs]	Top 5 B/(SE)/[CIs]	Top 9 B/(SE)/[CIs]	Top 11 B/(SE)/[CIs]
1. Duration never married, children (NM1C)	–2602.75	–3569.53	–3993.75	–3825.21
	(675.08)	(720.88)	(739.50)	(742.86)
	[–3926.18, –1279.33]	[–4982.72, –2156.33]	[–5443.44, –2544.05]	[–5281.50, –2368.92]
2. Duration never married, no children (NMNC)	823.21	–413.98	–1092.79	–1245.74
	(520.93)	(610.19)	(656.42)	(666.68)
	[–198.02, 1844.44]	[–1610.19, 782.22]	[–2379.63, 194.05]	[–2552.69, 61.20]
3. Number of M2C–M3C–DSC sub-sequences (ref.: none)				
once		–78687.68	–47043.18	–44915.87
		(16142.11)	(20394.38)	(20204.33)
		[–110332.38, –47042.99]	[–87023.94, –7062.41]	[–84524.06, –5307.69]
4. Transition to MNC at ages 20–24		–52402.70	–55494.99	–58657.61
		(10739.29)	(11469.19)	(11750.61)
		[–73455.79, –31349.60]	[–77978.98, –33011.00]	[–81693.29, –35621.93]
5. Transition to M2C at ages 20–24		–48965.93	–55427.32	–43255.35
		(12030.46)	(12845.21)	(13118.48)
		[–72550.22, –25381.65]	[–80608.84, –30245.81]	[–68972.58, –17538.11]
6. Duration divorced/separated, children (DSC)			–3326.73	–3027.54
			(1490.79)	(1491.81)
			[–6249.24, –404.22]	[–5952.06, –103.02]
7. Duration married, 1 child (M1C)			–422.98	–389.14
			(661.93)	(662.37)
			[–1720.62, 874.67]	[–1687.64, 909.36]
8. Number of DSC periods (ref.: none)				
once			–7028.33	–9767.13
			(22981.51)	(23008.26)
			[–52080.87, 38024.20]	[–54872.10, 35337.84]
twice			–113291.97	–115529.51
			(38471.00)	(38602.50)
			[–188709.80, –37874.14]	[–191205.14, –39853.89]
9. Duration remarried, children (RMC)			193.78	386.73
			(2487.30)	(2492.26)
			[–4682.27, 5069.84]	[–4499.06, 5272.52]
10. Transition to M1C at ages 15–19				–21407.68
				(14403.19)
				[–49643.44, 6828.07]
11. Transition to NM1C at ages 15–19				–43668.06
				(13075.04)
				[–69300.13, –18035.99]
N Individuals	5,702	5,702	5,702	5,702

Note: NMNC = never-married childless, NM1C = never-married with child(ren), MNC = married and childless, M1C = married with one child, M2C = married with two children, M3C = married with three children, M4C = married with four or more children, DSNC = divorced/separated and childless, DSNC = divorced/separated with child(ren), RMNC = remarried and childless, RMC = remarried with child(ren). All regressions account for respondent's age, marital dissolution at or after age 50, and widowhood at or after age 50m, migration background, parental education, number of siblings. Weighted.
Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).

Table A-5: Cell sizes for top 11 variables with continuous variables dichotomised by gender

Features	Women	Men
1. Duration never married, children (NM1C)		
No	2,130	2,152
Yes	721	699
2. Duration never married, no children (NMNC)		
No	5	2
Yes	2,846	2,849
3. Number of M2C–M3C–DSC sub-sequences		
Never	2,759	2,797
Once	92	54
4. Transition to MNC at ages 20–24		
No	2,234	2,566
Yes	617	285
5. Transition to M2C at ages 20–24		
No	2,527	2,760
Yes	324	91
6. Duration divorced/separated, children (DSC)		
No	2,262	2,453
Yes	589	398
7. Duration married, 1 child (M1C)		
No	740	803
Yes	2,111	2,048
8. Number of DSC periods		
Never	2,262	2,453
Once	532	368
Twice	57	30
9. Duration remarried, children (RMC)		
No	2,609	2,613
Yes	242	238
10. Transition to M1C at ages 15–19		
No	2,732	2,844
Yes	119	7
11. Transition to NM1C at ages 15–19		
No	2,702	2,817
Yes	149	34

Note: Duration variables were dichotomised so that 0 reflected no time spent in this state and 1 reflected any time spent in this state. Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).

Table A-6: Multiple regression: Adding features as explanatory variables in a stepwise fashion to predict personal net wealth in late working age. Disaggregated analyses by gender

	Women: Top 2	Men: Top 2	Women: Top 5	Men: Top 5	Women: Top 9	Men: Top 9	Women: Top 11
	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]
1. Duration never married, children (NM1C)	–2696.84	–2699.79	–3105.94	–3819.73	–3567.95	–4098.21	–3480.37
	(677.17)	(1269.11)	(722.14)	(1337.76)	(765.05)	(1341.68)	(771.20)
	[–4024.63, –1369.04]	[–5188.26, –211.31]	[–4521.92, –1689.96]	[–6442.81, –1196.64]	[–5068.05, –2067.85]	[–6728.98, –1467.44]	[–4992.55, –1968.20]
2. Duration never married, no children (NMNC)	899.47	–202.91	261.14	–1608.07	–478.14	–2023.69	–641.47
	(582.60)	(810.48)	(678.43)	(926.84)	(721.00)	(991.52)	(742.70)
	[–242.89, 2041.83]	[–1792.10, 1386.28]	[–1069.13, 1591.41]	[–3425.42, 209.28]	[–1891.87, 935.59]	[–3967.86, –79.53]	[–2097.76, 814.81]
3. Number of M2C–M3C–DSC sub-sequences (ref.: none)							
Once			–83086.20	–77585.09	–40110.85	–60642.22	–38154.73
			(19314.45)	(26016.30)	(23929.90)	(34290.69)	(23791.41)
			[–120957.99, –45214.41]	[–128597.86, –26572.31]	[–87032.64, 6810.94]	[–127879.46, 6595.02]	[–84804.98, 8495.52]
4. Transition to MNC at ages 20–24			–12257.65	–90770.85	–16849.97	–91285.46	–19964.39
			(12105.28)	(18017.76)	(12058.46)	(21183.34)	(12446.50)
			[–35993.68, 11478.38]	[–126100.08, –55441.63]	[–40494.21, 6794.27]	[–132821.78, –49749.15]	[–44369.50, 4440.73]
5. Transition to M2C at ages 20–24			–23626.26	–76948.69	–28370.96	–84029.16	–20553.12
			(14353.36)	(21739.92)	(15042.08)	(23312.16)	(15653.89)
			[–51770.34, 4517.81]	[–119576.33, –34321.05]	[–57865.50, 1123.58]	[–129739.67, –38318.64]	[–51247.32, 10141.07]
6. Duration divorced/separated, children (DSC)					–4321.12	–2488.07	–4152.63
					(2110.30)	(2585.71)	(2107.61)
					[–8458.99, –183.24]	[–7558.13, 2581.99]	[–8285.24, –20.03]
7. Duration married, 1 child (M1C)					27.72	–724.44	66.07
					(694.58)	(1230.46)	(694.54)
					[–1334.22, 1389.65]	[–3137.12, 1688.24]	[–1295.78, 1427.92]
8. Number of DSC periods (ref.: none)							
Once					–1992.74		–3636.07
					(35016.59)		(35002.35)
					[–70653.34, 66667.86]		[–72268.76, 64996.62]
Twice					–51281.99		–51856.00
					(42445.26)		(42512.36)
					[–134508.72, 31944.74]		[–135214.33, 31502.32]

Table A-7: (Continued)

	Women: Top 2	Men: Top 2	Women: Top 5	Men: Top 5	Women: Top 9	Men: Top 9	Women: Top 11
	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]	B/(SE)/[CIs]
9. Duration remarried, children (RMC)					–1460.42	141.74	–1255.03
					(1878.01)	(4081.28)	(1865.65)
					[–5142.84, 2221.99]	[–7860.85, 8144.32]	[–4913.20, 2403.13]
Once/twice						–5647.26	
						(31045.20)	
						[–66520.72, 55226.20]	
10. Transition to M1C at ages 15–19							–18305.34
							(15473.80)
							[–48646.40, 12035.73]
11. Transition to NM1C at ages 15–19							–22833.41
							(13010.80)
							[–48345.01, 2678.20]
N Individuals	2,851	2,851	2,851	2,851	2,851	2,851	2,851

Note: NMNC = never-married childless, NM1C = never-married with child(ren), MNC = married and childless, M1C = married with one child, M2C = married with two children, M3C = married with three children, M4C = married with four or more children, DSNC = divorced/separated and childless, DSNC = divorced/separated with child(ren), RMNC = remarried and childless, RMC = remarried with child(ren). All regressions account for respondent's age, marital dissolution at or after age 50, and widowhood at or after age 50, migration background, parental education, number of siblings. Regressions for the top 11 features are only conducted for women because of the small cell sizes for men for features 10 and 11. Weighted.
Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).

Table A-8: Descriptive statistics for top 11 categorical variables by gender and over the pooled sample

Feature	Men mean/(SE)/min/ max	Women mean/(SE)/min/ max	Total mean/(SE)/min/ max
1. Duration never married, children (NM1C)	1.23	1.43	1.32
	(4.17)	(4.76)	(4.46)
	0.00	0.00	0.00
	32.00	34.00	34.00
2. Duration never married, no children (NMNC)	16.68	12.02	14.49
	(9.65)	(8.71)	(9.51)
	0.00	0.00	0.00
	36.00	36.00	36.00
3. Number of M2C–M3C–DSC sub-sequences			
Never	0.98	0.97	0.97
Once	0.02	0.03	0.03
4. Transition to MNC at ages 20–24	0.10	0.22	0.16
5. Transition to M2C at ages 20–24	0.04	0.13	0.08
6. Duration divorced/separated, children (DSC)	1.21	2.02	1.59
	(3.53)	(5.04)	(4.33)
	0.00	0.00	0.00
	25.00	30.00	30.00
7. Duration married, 1 child (M1C)	4.12	5.12	4.59
	(6.19)	(7.42)	(6.81)
	0.00	0.00	0.00
	31.00	34.00	34.00
8. Number of DSC periods			
Never	0.84	0.80	0.82
Once	0.14	0.18	0.16
Twice	0.01	0.02	0.02
9. Duration remarried, children (RMC)	0.91	0.92	0.92
	(3.52)	(3.68)	(3.59)
	0.00	0.00	0.00
	26.00	29.00	29.00
10. Transition to M1C at ages 15–19	0.00	0.05	0.02
11. Transition to NM1C at ages 15–19	0.02	0.06	0.04
N Individuals	2,851	2,851	5,702

Note: Data are from the Socio-Economic Panel Survey v38 (2002, 2007, 2012, 2017). Weighted.

Table A-9: Boruta feature selection ranks for the pooled sample and by gender

Feature	rank Main	rank Men	rank Women
Duration never married, children (NM1C)	1*	9*	2*
Duration never married, no children (NMNC)	2*	1*	3*
Number of M2C–M3C–DSC sub-sequences	3*	11*	11*
Transition to MNC at ages 20–24	4*	8*	23
Transition to M2C at ages 20–24	5*	15	15
Duration divorced/separated, children (DSC)	6*	17	1*
Duration married, 1 child (M1C)	7*	3*	5*
Number of DSC periods	8*	16	7*
Duration remarried, children (RMC)	9*	2*	4*
Transition to M1C at ages 15–19	10*	25	12*
Transition to NM1C at ages 15–19	11*	24	9*
Transition to M1C at ages 30–34	12*	18	20
Number of M1C periods	13*	4*	18
Transition to MNC at ages 25–29	14*	20	17
Number of M1C–M2C–DSC sub-sequences	15*	14	14
Duration married, no children (MNC)	16*	5*	21
Transition to M1C at ages 20–24	17*	19	24
Number of NM1C –DSC sub-sequences	18*	7*	13*
Number of M1C–M2C–M3C–DSC sub-sequences	19*	12*	16
Number of M1C–DSC sub-sequences	20*	22	10*
Duration married, 4 children or more (M4C)	21*	21	22
Duration married, 3 children (M3C)	22*	6*	6*
Duration married, 2 children (M2C)	23 *(–)*	10*	8*
Transition to NM1C at ages 20–24	24 *(–)*	13*	25
Transition to DSC at ages 35–39	25* *(23)*	23	19

Note: Depicts all features that were either selected in the main or gender-specific models. ‘Main’ refers to the main model for the whole sample, ‘men’ to the model restricted to men only, and ‘women’ to the model restricted to women only. Ranked by mean importance in the main model measured as z-scores. * indicates features that were confirmed as important in the specific model.
(–) refers to features that were not selected as important in the main model but were confirmed in the gender-specific models (‘Transition to NM1C at ages 20–24’ and ‘Transition to DSC at ages 35–39). *(23)* indicates that feature ‘Transition to DSC at ages 35–39’ is ranked 23rd in the main model, when not considering features that were confirmed in the gender-specific model only.
Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).

Additional figures

Figure A-2: Sample selection steps

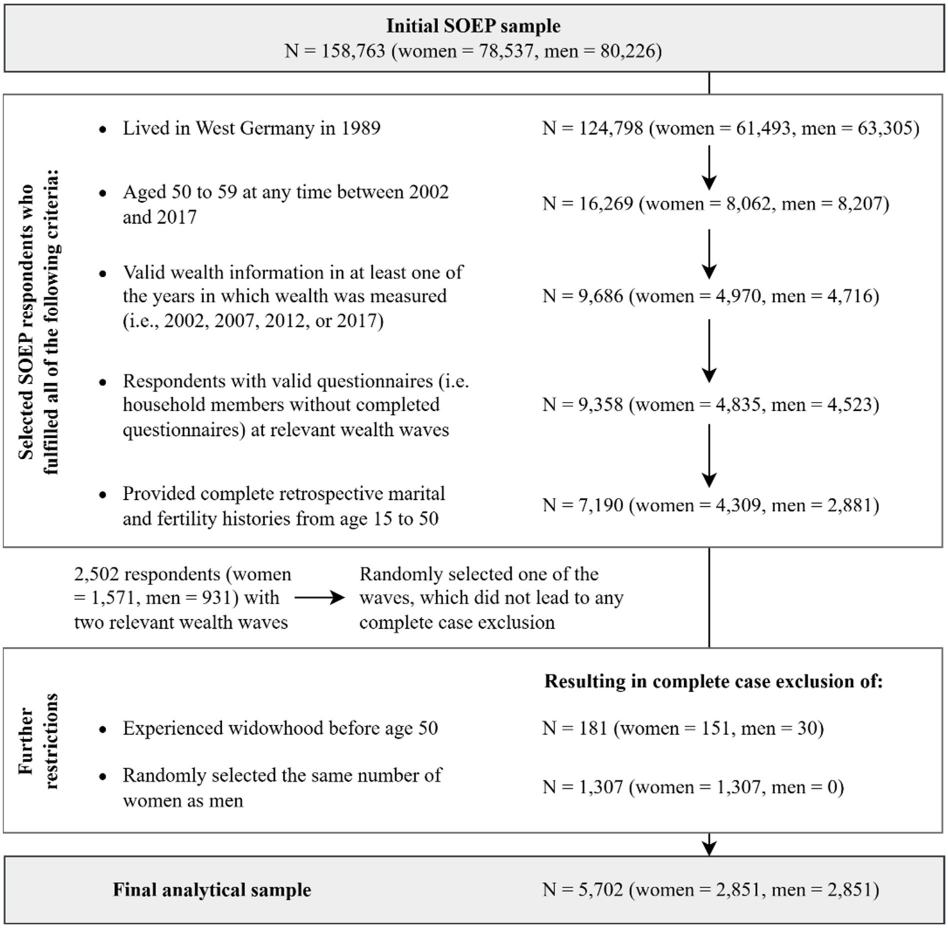
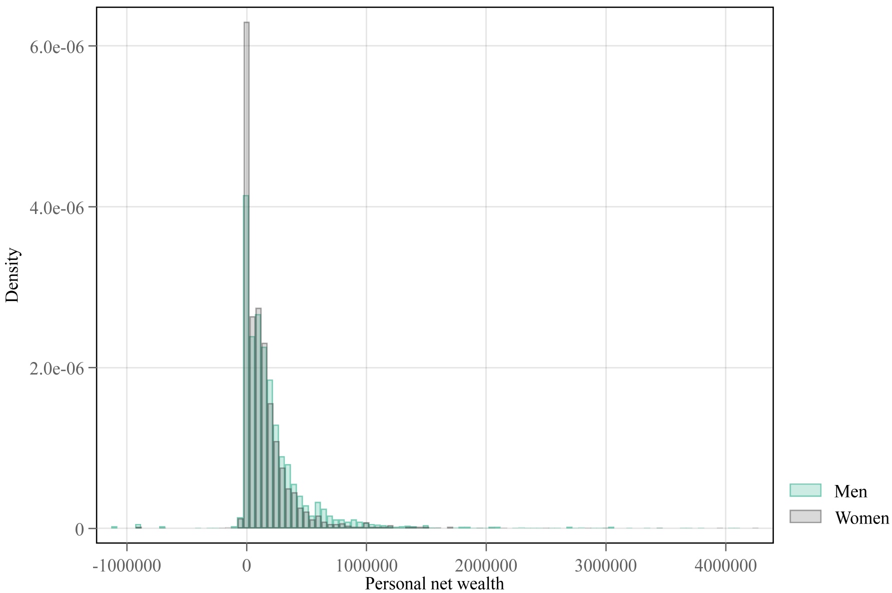
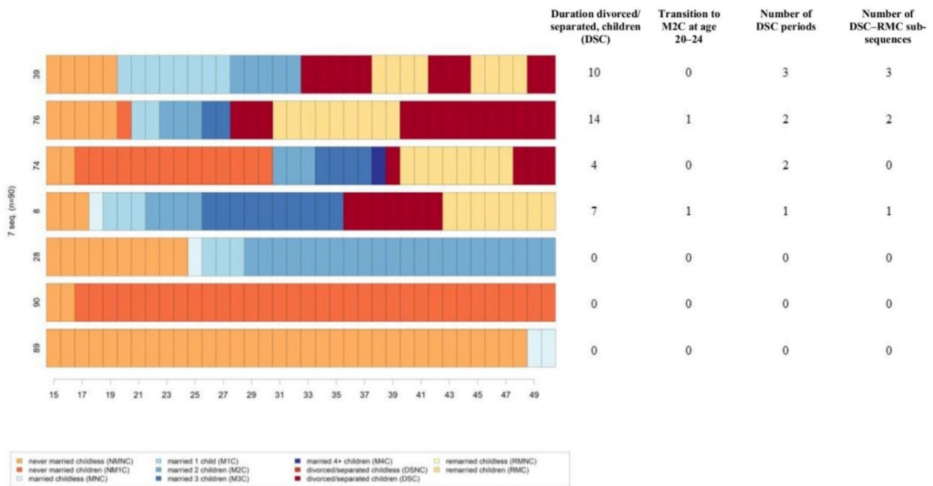


Figure A-3: Density plot of personal net wealth by gender for the analytical sample



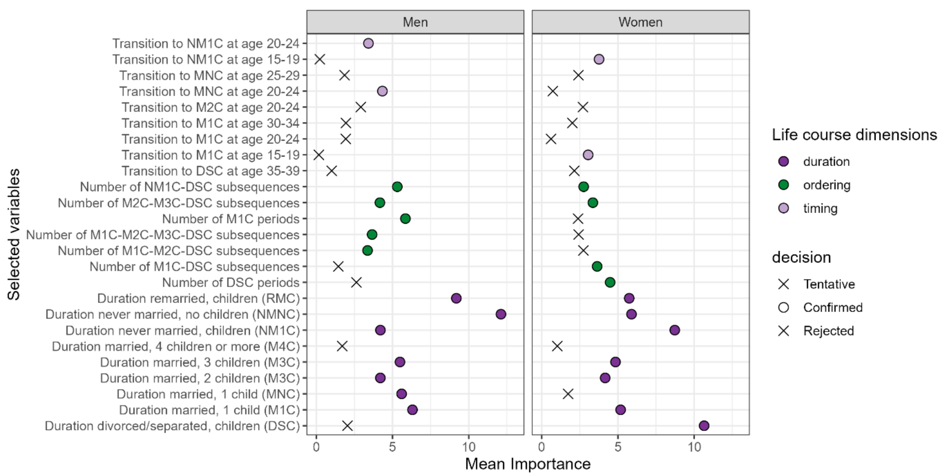
Note: Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).

Figure A-4: Example trajectories and life-course features



Notes: Data are from the Socio-Economic Panel Survey v38 (2002, 2007, 2012, 2017).

Figure A-5: Boruta feature selection results considering women and men in separate samples



Note: Depicts the mean importance of all features that were confirmed important in the pooled model, in the model for wealth inequality among men only, or among women only. Feature importance is measured as z-scores.
Data are from the Socio-Economic Panel v38 (2002, 2007, 2012, 2017).