NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Assessing the Impact of Parental Allowance on Fertility

A Synthetic Control Study of Germany

Manuel Esteves Pereira Gonçalves

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Assessing the Impact of Parental Allowance on Fertility
A Synthetic Control Study of Germany

by

Manuel Esteves Pereira Gonçalves

Master Thesis presented as a partial requirement for obtaining the Master’s degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Bruno Damásio, PhD, Nova IMS

July, 2025

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Barreiro, 2025

Manuel Esteves Pereira Gonçalves

Abstract

Declining fertility rates in advanced economies have led governments to adopt financial incentives to support family formation. This thesis evaluates the causal impact of Germany's 2007 Elterngeld parental allowance policy on national fertility using the Augmented Synthetic Control Method (ASCM), which improves pre-treatment fit through ridge regression. Using World Bank data from 1995–2014, we construct a counterfactual for Germany's Total Fertility Rate (TFR) based on a weighted combination of eight comparable European countries. Across eight model specifications, the most balanced model estimates an average policy effect of +0.043 births per woman during 2007–2014. A migration-adjusted model yields a smaller ATT of +0.018, suggesting partial mediation by demographic shifts. While the placebo-based p-value (0.111) falls short of conventional significance, the model shows strong pre-treatment fit and post-treatment divergence. These findings suggest a modest positive impact of Elterngeld on national fertility, though not large enough to reverse long-term trends. Methodologically, this study demonstrates the application of ASCM in a national demographic policy context. Substantively, it underscores that financial incentives alone are insufficient to counter fertility decline without complementary measures such as accessible childcare, flexible work options, and gender-equitable leave policies.

Keywords

Fertility Rates ; Synthetic Control Method ; Augmented SCM ; Causal Inference; Parental Allowance ; Germany

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations and Acronyms

ASCM	Augmented Synthetic Control Method
ATT	Average Treatment Effect
MSPE	Mean Squared Prediction Error
SCM	Synthetic Control Method
TFR	Total Fertility Rate

1. INTRODUCTION

Across the world a significant decline in the existing birth rates has been noted. Particularly in western countries, such as Germany, the total fertility rate which is the average measure of children born per woman, has declined well below the replacement level of 2.1 children per woman, and in several cases, below the critical threshold of 1.5 (Lutz et al., 2003; Balbo et al., 2013). This demographic trend presents pressing economic and social challenges, such as aging populations, labor market imbalances, and pressures on social security systems (Lee et al., 2014). Understanding the drivers of fertility decline and evaluating policy effectiveness is crucial.

One of the key drivers behind low fertility rates is the economic and social cost of childbearing. High childcare expenses, career interruptions, work-life balance difficulties, and financial insecurity contribute to couples postponing or forgoing parenthood (Mcdonald, 2006). Governments have increasingly turned to family policies as a means to support fertility rates, using financial incentives, parental leave provisions, and childcare support to alleviate these burdens (Lutz et al., 2007). Among such policies, parental allowance (Elterngeld) , which is a direct money transfer payment dependent on net income, has emerged as a major intervention designed to mitigate the economic disadvantages associated with having children.

This paper examines the impact of Germany's parental allowance (Elterngeld) on fertility rates between 2007 and 2014, time period between introduction of original policy and further policy expansion (Elterngeld Plus). To estimate the policy's causal impact, this study applies the Augmented Synthetic Control Method (ASCM), an extension of the original Synthetic Control Method (SCM) that incorporates ridge regression to improve pre-treatment fit. This methodology is particularly suited for evaluating national-level interventions like Elterngeld, where constructing a credible counterfactual is challenging due to the absence of a natural control group. Introduced in 2007, Elterngeld aimed to encourage higher fertility rates by replacing a proportion of parents' income during the first year after childbirth, providing financial security and incentivizing family expansion (Spiess & Wrohlich, 2006). By employing the ASCM, this study constructs a counterfactual Germany without the policy to assess whether Elterngeld had a causal effect on fertility rates.

Prior research on fertility policies has explored various family support measures, with studies finding mixed results regarding the effectiveness of financial incentives on birth rates. While some studies suggest that monetary benefits can positively influence short-term fertility decisions (Fitzenberger & Seidlitz, 2024), others argue that long-term demographic shifts require broader structural changes, including gender equality in labor markets and affordable childcare solutions (Bloom et al., 2011).

This paper contributes to the literature in three key ways, firstly by exploring methodological innovation which is derived by applying an improvement upon the original Synthetic Control Method, the Augmented Synthetic Control Method, since it enhances the causal inference of policy evaluation by offering a more robust alternative to traditional difference-in-differences approaches (Abadie, 2021; Abadie et al., 2010) and corrects for imperfect pretreatment fit (Ben-Michael et al., 2021). Secondly by exploring a more in-depth policy-specific analysis since unlike most other studies on family policies, this research isolates the impact of Elterngeld, providing a clearer understanding of its effectiveness in influencing fertility behavior. And lastly by providing a time-sensitive evaluation since the study focuses on the period from 2007 to 2014, this enables the capturing of both the immediate and medium-term effects of the policy, offering insights into whether the policy's fertility effects were sustained over multiple years and what this reveals about its effectiveness as a long-term demographic intervention.

2. Literature review

2.1. Determinants of Fertility Decline: Economic & Social Factors

Fertility trends have long been a subject of interdisciplinary research, spanning demography, economics, sociology, and public policy (Morgan & Taylor, 2006; Balbo et al., 2013). As developed countries face sustained declines in fertility rates, understanding and effectively breaking down the main drivers behind reproductive planning decisions has become increasingly important for affected societies.

One of the most comprehensive overviews of fertility determinants in advanced societies is provided by Balbo et al. (2013). Their review synthesizes a wide range of empirical studies and highlights the complex interplay between individual-level factors (e.g., educational attainment, partnership dynamics), structural constraints (e.g., labor market uncertainty, gender inequality), and institutional support (e.g., family policies, childcare infrastructure). They emphasize that while economic conditions influence fertility decisions, the cultural and societal context also play a pivotal role, particularly in shaping fertility timing. This conclusion is directly relevant to this study's focus, as strictly economics conditions and incentives aimed at alleviating difficulties in this spectrum, such as the Elterngeld policy, may only play a limited role in the overall birth planning decision scenario.

While this cultural framing may initially appear broad or imprecise, if we break it down and examine key contributing factors a clearer picture emerges. Esping-Andersen (2001) evidence that a more neoliberal job market that provides more job security and competitiveness delays family formation particularly among young adults. At the same time the changing of gender norms have led to higher female labour market participation, but without proper institutional support, women are often forced to choose between their careers or family (McDonald, 2000). These dynamics consequently lead to, as Hochschild (1997) argues, that children are ever more seen as a high-cost choice, in a society ever more driven by consumption trends, individualism and lifestyle optimization.

This leads us to evaluate the implications of low-fertility values, which have been addressed by researchers such as Lee et al. (2014) and Bloom, Canning, and Fink (2011). These studies emphasize the more macroeconomic side of the consequences posed by having sustained fertility below replacement level, such as population aging, increased dependency ratios, and potential declines in economic productivity. In particular, Lee et al. argue that fertility below 1.5 children per woman, the so-called "low-fertility trap", may create a self-reinforcing cycle through altered intergenerational transfers and public finance pressures. This concept of a "low-fertility trap" was formally theorized by Lutz, Skirbekk, and Testa (2007), who proposed that certain societal dynamics, such as negative population momentum, altered life-course preferences, and pessimistic fertility expectations, can push fertility rates further downward. Their model suggests that once a society's fertility rate drops below a critical threshold, reversing this trend becomes significantly more difficult without substantial institutional or cultural shifts.

Figure 1 illustrates this long-term downward trajectory in fertility. Since the 1970s, Germany's TFR has consistently remained below the EU average, with especially pronounced declines in the decades leading up to the introduction of Elterngeld in 2007. This persistent fertility gap contextualizes both the urgency and the constraints of policies aimed at reversing demographic decline.

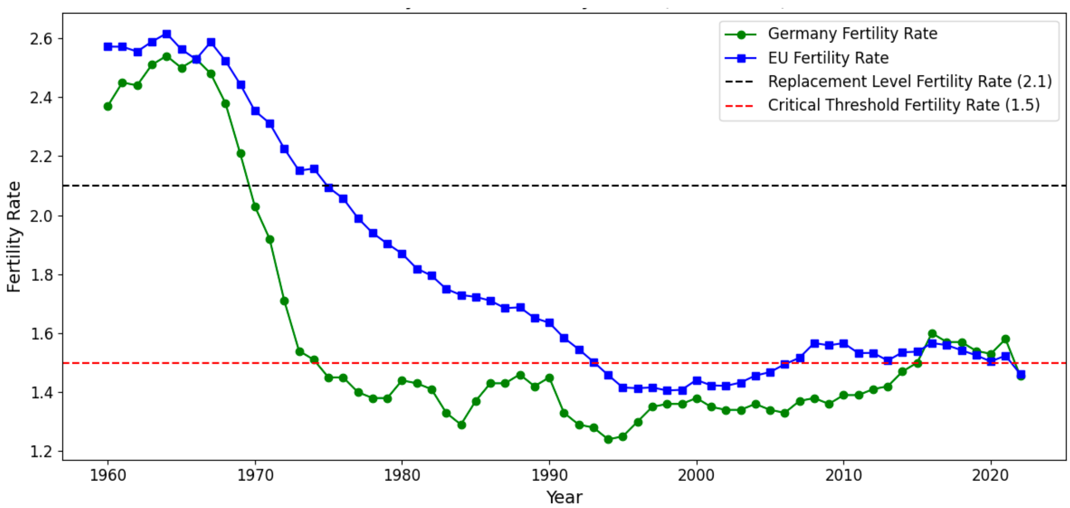

Figure 1: TFR in Germany vs EU (1960-2022)

Taken together, these foundational studies suggest that fertility behavior results from a complex decision process, influenced not only by personal circumstances and economic incentives, which Elterngeld aims to address, but also by broader institutional structures and cultural norms.

2.2. Policy Interventions & Fertility: Global Evidence

From a policy perspective, McDonald (2006) discusses the state's role in shaping fertility outcomes. He argues that policy measures can be effective when they align with the needs and expectations of potential parents, particularly in reconciling work and family life. However, McDonald warns that policies must be both comprehensive and sustained to have a meaningful impact, noting that fragmented or one-dimensional interventions often fail to produce long-term fertility changes. They also lay the groundwork for evaluating the effectiveness of targeted policy interventions, such as parental allowance, within this broader socio-demographic framework. McDonald also provides a foundational framework for assessing the impact of family policies, arguing that fertility can be positively influenced when state interventions address the opportunity costs of childbearing and facilitate the compatibility of having children with the corresponding estimated foregone labor market participation. According to McDonald, policies that are broad, consistent, and gender-equal in approach are more likely to yield long-term demographic change than short-term or fragmented programs.

Empirical evidence from country case studies offers mixed but nonetheless valuable insights. The Nordic countries, often seen as leaders in family policy, provide further insight into the long-term effects of generous, gender-balanced systems. Hellstrand et al. (2020) shows that countries like Sweden and Norway which have been experiencing persistently low fertility rates, especially in the last decade, aimed to address this in part with the availability of extended paid parental leave, subsidized childcare, and strong support for maternal employment. However, despite their generous efforts aimed at making family policy more robust, the fertility rate values in the Nordics still persist well below replacement level despite possible arguments for moderate positive effect of these policies.

Other governments across the globe have also implemented and tested a variety of policy measures aimed at supporting family formation and increasing birth rates. These interventions range from direct financial incentives and paid parental leave, to childcare provision, tax benefits, and broader work-family reconciliation measures. The effectiveness of such policies remains a subject of ongoing debate, with different outcomes often depending on the policy design, duration, and contextual fit of the actual interventions.

Beyond financial incentives, scholars have emphasized the importance of social infrastructure and normative frameworks. Lutz, O'Neill, and Scherbov (2003) argue that long-term demographic resilience depends not only on policy generosity but also on its integration with labor market structures, gender norms, and institutional trust. This perspective aligns with findings by Balbo et al. (2013), who stress that policy success is often mediated by how effectively it addresses the underlying causes of fertility postponement, such as employment insecurity, housing instability, and gendered care burdens.

In sum, while policy interventions can shape fertility behavior, their impact is not uniform across settings. These international experiences highlight both the potential and the limits of family policy design, questions that become particularly relevant when considering Germany's 2007 Elterngeld reform.

2.3. The Elterngeld Policy in Germany

In an effort to address persistently low fertility rates and modernize family policy, the German government introduced Elterngeld in January 2007. This parental allowance scheme replaced a share of a parent's previous income, up to 67%, capped at 1800€, for a duration of up to 14 months following childbirth. The policy marked a significant shift away from the former *Erziehungsgeld* system, which offered a flat-rate payment with weaker labor market incentives. Elterngeld aimed not only to provide financial support during early parenthood but also to encourage earlier family formation, increase birth rates, and promote gender equality by incentivizing paternal involvement through a two-month bonus for the non-primary caregiver.

While the labor market implications of Elterngeld have been studied extensively, its national level demographic impact remains underexplored. Spiess and Wrohlich (2006) discussed the policy's potential to promote a more Scandinavian model of work-family reconciliation, particularly by supporting maternal labor force re-entry. Fitzenberger and Seidlitz (2024) further examined heterogeneous effects on motherhood decisions and employment, highlighting behavioral differences across income and education groups. These contributions emphasize Elterngeld's social and economic reach, but offer limited insight into its effectiveness as a fertility-enhancing instrument. Research explicitly targeting the policy's consequences is primarily focused on using micro-level data and quasi-experimental techniques to estimate its impact on fertility behavior. Hart et al. (2023) evaluated the causal effect of family policies on fertility in various countries, including Germany. However, their work encountered difficulties in identifying a credible control group for Germany, limiting the strength of their causal claims. This challenge is common in national-level policy evaluations, where randomized experiments are infeasible, and suitable counterfactuals are hard to construct.

Despite these challenges, several quasi-experimental studies have attempted to identify causal effects at the individual level using eligibility thresholds and natural policy variation. Raute (2019) applied a differences-in-differences design to compare women just eligible versus ineligible for the 2007 reform, and found that Elterngeld led to a 23% increase in the fertility of tertiary-educated women. The study suggests that income replacement incentives can significantly affect fertility timing and planning for career-oriented women. Similarly, Kluve and Tamm (2013) used a natural experiment approach based on DiD to assess broader behavioral impacts. They find increased second birth probabilities, greater maternal labor market continuity, and enhanced paternal childcare involvement, emphasizing the policy's multi-dimensional effects. However, these studies estimate local average treatment effects and do not assess aggregate national-level changes.

This study builds on these micro-level insights by offering a macro-level evaluation of Elterngeld's national impact using the Augmented Synthetic Control Method (ASCM). By constructing a credible counterfactual for Germany's fertility trajectory, it addresses the identification gap in existing literature and contributes new evidence on the policy's effectiveness at the population level.

2.4. Synthetic Control Method in Demographic Policy Evaluation

The SCM has emerged as a powerful tool for policy evaluation, particularly in cases where randomized experiments are infeasible (Abadie, Diamond, & Hainmueller, 2010). Unlike traditional methods of causal inference such as Difference-in-Differences (DiD) or fixed effects models, which rely on the parallel-trends assumption, SCM constructs a counterfactual by assigning data-driven weights to untreated units to closely match the treated unit's pre-treatment trajectory. This approach reduces extrapolation bias and improves causal inference in settings where finding a comparable control group is challenging (Abadie, 2021).

There's extensive literature on the SCM application for demographic policy evaluation, Gietel-Basten et al. (2019) assesses the impact of the "one-child policy" in China. By building a synthetic China it was found that the "one-child policy" (1979) had no statistically significant impact on the observed fertility decline compared with the outcome in the counterfactual. The synthetic trajectory mirrored earlier demographic shifts under the previous "later-longer-fewer" policy (1973) and normal economic development. The study finally notes that the "one-child policy" did not yield its demographic benefits pertaining to population control but instead created serious unintended consequences like skewed sex-ratios (due to son preferences) and social challenges (pressure on family structures, aging population).

Reich (2024) applies SCM to study the impact of the 2004 Baby Bonus policy on Australia's fertility rates. The results show a positive statistically significant impact of 6.82% between 2005 and 2012. The policy was based on a universal cash transfer of 3000 AUD per birth after July 1st 2004, suggesting that even financial incentives can modestly alter fertility trajectories in high-income countries. Similarly, Podör et al. (2024) employs synthetic control to evaluate how Hungary's CSOK housing subsidy policy, aimed at reversing low fertility rates performed. This policy consisted of a housing subsidy for families with children, and their findings suggest that post implementation (2015-2021) no statistically significant divergence between Hungary's actual fertility trajectory and the counterfactual after the policy was introduced. This finding underscores its cost-ineffectiveness (€1.9 billion) and the broader limitations of standalone financial incentives, suggesting that effective policy design must account for overall economic stability as well as holistic family support (childcare support, gender equity).

These studies highlight SCM's advantage in capturing treatment effects in complex policy environments. However, its effectiveness also relies on selecting an appropriate donor pool and ensuring that pre-treatment trends align. SCM also has technical limitations, including sensitivity to extrapolation (when the treated unit lies far from the donor pool), the requirement for a long pre-treatment period, and a risk of overfitting if too many predictors are used.

To address these issues, Ben-Michael et al. (2021) introduce the ASCM which incorporates outcome modeling, such as ridge regression, to improve pre-treatment balance and reduce bias. In Germany's case, constructing a valid counterfactual is particularly challenging, ASCM helps overcome potential bias in donor pool weighting by incorporating outcome modeling, thereby improving fit and robustness.

Other common challenges in SCM applications, such as biased estimates due to inappropriate donor pool selection and policy spillovers, are addressed later in this study by employing a rigorous donor pool selection process, conducting robustness checks, and ensuring valid inference.

3. EMPIRICAL STRATEGY

3.1. Data Sources

The data used in this study is sourced from the World Bank Open Data repository to model both the fertility trajectory of the treated unit (Germany) and to construct the synthetic counterfactual from the donor pool data. The TFR is set as the dependent variable, covering the period from 1995-2014, with the Elterngeld policy being implemented in 2007 marking the beginning of the post-treatment period.

For the donor pool specification, the candidates were selected on their structural similarity compared to Germany (EU countries), that their fertility trends pre-2007 match that of the treated unit and that the risk of policy contamination was mitigated by excluding countries that possess similar parental allowance schemes. Therefore, the full donor pool comprises: Portugal, Switzerland, Belgium, Spain, Italy, the Netherlands, Ireland, and Slovenia.

The time period of analysis for pre-treatment is from 1995-2006 to ensure a time frame of at least 10 years to validate matching fertility trends. The Elterngeld policy was established in 2007 and the post-treatment period ended in 2014 due to the expansion of the policy in 2015, known as Elterngeld Plus (BMFSFJ, 2015) which would bias the study of the impact of the original policy.

The set of socio-economic predictor (covariate) variables and special predictors was guided by relevance to fertility planning decisions and practice in the synthetic control literature (Reich, 2024). These variables are then used to ensure pre-treatment similarity between the treated and synthetic units.

Table 1 – Variables for model's specifications

Variable	**Unit**	**Notes**
Total Fertility Rate (TFR)	Births per woman	Dependent variable
GDP per capita (PPP, constant $)	USD	Predictor
Female Labor Force Participation	% of female population aged 15+	Predictor
Unemployment Rate	% of total labor force	Predictor
Urban Population Rate	% of total population	Predictor
Net Migration	People (net annual)	Predictor
Inflation (consumer prices)	Annual %	Predictor
Average Fertility Rate	Births per woman (mean)	Special Predictor

3.2. Augmented Synthetic Control Method: Concept & Implementation

The SCM, introduced by Abadie and Gardeazabal (2003), is a comparative case study approach for estimating the causal impact of an intervention. It constructs a synthetic control unit, which is a weighted combination of untreated units from a donor pool, designed to closely match the pre-intervention trajectory of the treated unit's outcome variable. By optimizing the fit using both predictor variables and outcome values in the pre-treatment period, SCM provides a data-driven counterfactual to estimate what would have happened in the absence of the treatment.

However, SCM has its limitations, such as sensitivity to poor pre-treatment fit and risk of overfitting. The ASCM, introduced by Ben-Michael, Feller, and Rothstein (2021), aimed to address these limitations by incorporating outcome modeling through ridge regression which directly controls the pre-treatment fit.

The base SCM objective selects weights W to ensure that a weighted average of the control units' predictor variables and pre-intervention TFR closely resembles that of Germany. Formally SCM minimizes the following expression:

$$min_W \parallel X_1 - X_0 W \parallel_V \quad (1)$$

where $\parallel x \parallel V = \sqrt{x^T V x}$ and V is a positive semidefinite predictor weight matrix that penalizes mismatch in important predictors, X_1 represents the vector predictors for the treated unit and X_0 the matrix of predictors for the control units.

ASCM then applies a ridge regression correction to the SCM weights which contrary to SCM, admits negative weights to improve pre-treatment fit and is represented by:

$$W_{aug} = W + W_{ridge} \quad (2)$$

where the ridge correction is given by:

$$W_{ridge} = (A - BW)(BB^\top + \lambda I)^{-1} B \quad (3)$$

Considering λ as the ridge regularization parameter that penalizes the distance from SCM weights in order to directly parameterize the degree of extrapolation (Ben-Michael et al., 2021), and with the matrices defined as:

- $A = \tilde{X}_1 \cup \tilde{Z}_1$: stacked and normalized treated outcomes and covariates.
- $B = \tilde{X}_0 \cup \tilde{Z}_0$: stacked and normalized donor outcomes and covariates.

Where Z_0, Z_1 represent any additional covariates. Normalization is then carried out before the ridge correction to ensure comparability and then the estimated synthetic Germany is represented as:

$$\hat{Y}_{1t}^{aug} = \sum_{j=1}^{J} W_{aug,j} Y_{jt} \quad (4)$$

Therefore $\hat{Y}_{1t}^{aug}$ represents the outcome for the treated unit (Germany) at time t in the absence of treatment and Y_{jt} the outcome for control unit *j* at time *t.* And so the treatment effect at time $t \geq T_0$ is:

$$\tau_t = Y_{1t} - \hat{Y}_{1t}^{aug} \quad (5)$$

The application of the ASCM was conducted using the open-source Python library *pysyncon* (Fordham, 2022), which provides tools for constructing and evaluating synthetic controls with both traditional and augmented estimators. The library follows the methodology introduced by Ben-Michael, Feller, and Rothstein (2021), and allows for flexible specification of predictor sets, special predictors, and regularization parameters.

In implementation, the donor pool consists of European countries that did not enact major fertility-related reforms around 2007, ensuring policy exogeneity. Predictor variables *X* and covariates *Z* include the variables specified in Table 1. These are selected for their theoretical relevance to fertility behavior and data availability across countries.

Together, the ASCM framework provides a robust, data-driven approach to estimate the causal effect of Elterngeld on Germany's fertility trajectory by leveraging a transparent and interpretable counterfactual construction method.

3.3. Selection of Donor Pool & Predictor Variables for Synthetic Germany

To ensure the credibility and robustness of the estimated treatment effects, multiple model specifications were tested and compared using standard SCM evaluation metrics (a detailed discussion of these metrics follows in the next chapter). The goal of these variations was to assess the sensitivity of the results to different methodological choices inherent in the ASCM framework. In particular, the specifications were designed to address three potential sources of bias or misestimation: (i) contamination risk arising from the inclusion of donor countries with overlapping family policy reforms, (ii) predictor overrepresentation given Germany's migration geopolitical context, leading to overfitting in the pre-treatment period, and (iii) the influence of special predictor construction strategies, specifically the trade-off between using disaggregated year-by-year pre-treatment outcomes versus a single aggregated mean as outcome predictors.

The donor pool selection process requires a high degree of rigor, as policy spillover effects from poor selection can bias the outcome of the analysis. The inclusion criteria for the donor pool to construct synthetic Germany were: are structurally similar to Germany, have similar pre-2007 fertility trends and didn't have similar parental allowance schemes with a policy design like Germany's Elterngeld. This selection process resulted in the selection of 8 countries: Portugal, Switzerland, Belgium, Spain, Italy, the Netherlands, Ireland, and Slovenia. This number of units in the donor pool is in line with other demographic studies that leveraged

SCM (Podör et al., 2024). To further mitigate risk and following prior work in demographic policy evaluation (Reich, 2024), we classify donor countries based on their perceived policy risk of biasing the synthetic control. Starting with the low-risk countries group, these countries had stable fertility policies, no large-scale expansions of family or childcare benefits, and similar demographic profiles. They form the basis of the Strict donor pool: Portugal, Spain, Switzerland and Italy. For the moderate-risk countries, these have either generous existing policies (possibly biasing the counterfactual TFR), or underwent minor reforms or macroeconomic transitions: Belgium, Netherlands, Ireland and Slovenia. The Full donor pool includes both low- and moderate-risk countries, while the Strict donor pool includes only the low-risk subset. Comparing results across both pools tests the robustness of treatment effects to donor selection.

Figure 2 illustrates the TFR trajectories of Germany and all potential donor pool countries between 1995 and 2014. Prior to the 2007 Elterngeld reform (marked by the vertical dashed line), Germany's fertility rate remained persistently low but stable, in contrast to wider variation across candidate countries. Nations such as Ireland (IE) and the Netherlands (NL) consistently recorded higher TFRs, while others like Italy (IT) and Portugal (PT) mirrored Germany more closely. This visual assessment supports the rationale for constructing a donor pool composed of countries with comparable fertility dynamics in the pre-treatment period.

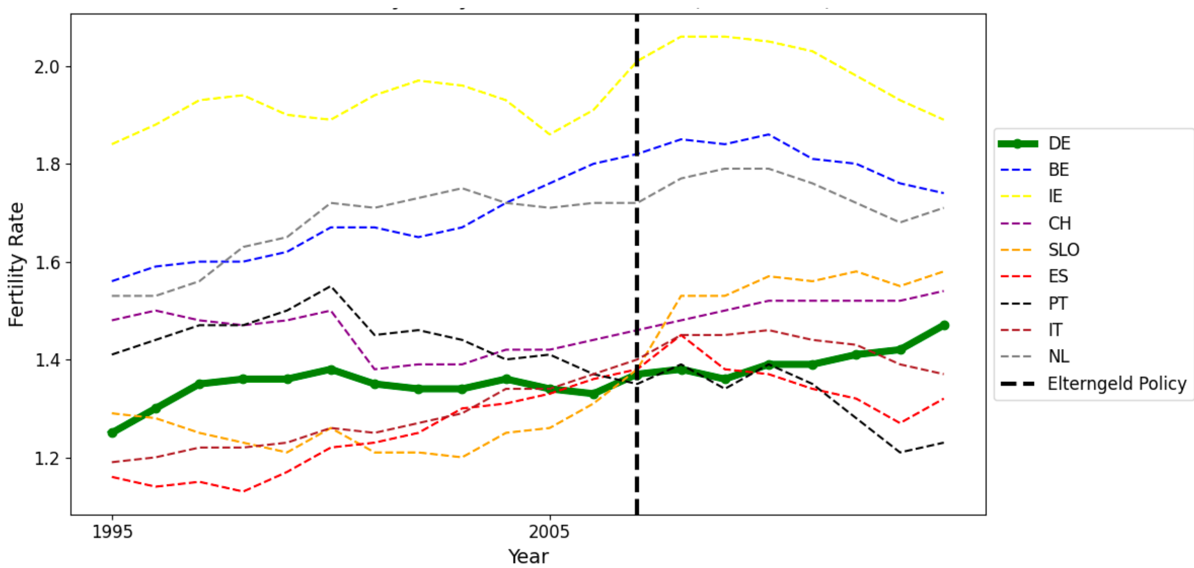

Figure 2: TFR in Germany vs Synthetic Candidates (1995-2014)

Given Germany's geopolitical context during the study period, it is essential to account for migration when analyzing fertility trends. According to OECD (2024), Germany experienced a sharp increase in immigration during the early 2000's, primarily driven by mobility from other EU countries following the 2004 and 2007 enlargements. This trend substantially increased net migration into Germany during both the pre- and post-treatment periods. Thus, it's relevant to discern the effect of migration on the effectiveness of the Elterngeld's policy, since generally migrant women have different fertility norms and their decisions may not be influenced by national policy in the same way as the native-born population. In a practical setting, if Germany's fertility increases due to increased share of migrants with high fertility norms, and this is not mirrored in donor countries, then the synthetic control will misestimate the treatment effect. To account for this potential bias, the analysis includes two predictor specifications: a Baseline set (including all variables from Table 1 excluding Net Migration), and an Adjusted set (including Net Migration), allowing us to assess the sensitivity of the estimated treatment effect to demographic composition changes stemming from migration flows.

Special predictors are used in addition to predictors in synthetic control methods to improve the pre-treatment fit of the counterfactual on the treated unit's TFR trajectory. These often include using pre-treatment lagged values of the outcome variable itself. Ben-Michael, Feller, and Rothstein (2021) show that the structure of special predictors such as the number and form of lagged outcomes included can significantly influence inference in synthetic control methods. Therefore, we include two special predictor specifications, the first includes 11 year-specific lagged values of the TFR from 1995 to 2006, while the second uses a single aggregated mean of TFR over the same period.

Combining the variations in donor pool, predictors, and special predictor configurations, we estimate a total of eight model specifications, whose robustness and inference quality are evaluated using placebo tests and fit diagnostics described in the next section.

3.4. Placebo Tests & Robustness Checks

Placebo tests are widely used in the SCM framework as a method of result validation. Following Abadie (2021), we construct a permutation distribution by reassigning treatment to each donor unit and estimating placebo effects, where the true effect's significance is assessed by its quantile in this distribution. The null hypothesis tested is that the Elterngeld reform had no effect on Germany's TFR trajectory. The p-value is expressed as:

$$p - value = \frac{\left\{\# \; j: \frac{MSPE_{post,j}}{MSPE_{pre,j}} \geq \frac{MSPE_{post,Germany}}{MSPE_{pre,Germany}}\right\}}{J} \quad (6)$$

The numerator considers the count of donor units with more "unusual" divergence post-treatment than Germany, in Mean Squared Prediction Error (MSPE) terms, and the denominator the total number of valid control units (excluding Germany).

The MSPE is a standard diagnostic used in synthetic control methods to evaluate the quality of the synthetic fit (Abadie, Diamond, & Hainmueller, 2010). The pre-treatment MSPE quantifies how well the model fits Germany before the Elterngeld reform (1995–2006), while the post-treatment MSPE evaluates the discrepancy between observed and synthetic outcomes after the policy implementation (2007-2014). The ratio of post- to pre-treatment MSPE is a key indicator of signal strength. A high MSPE ratio suggests that the synthetic control closely matched Germany's TFR before the intervention, but diverged notably afterward, consistent with a genuine treatment effect. Conversely, a low ratio implies poor fit or weak treatment signal.

Gap plots are leveraged to observe the evolution of the treatment effect across the different placebo tests permutations. A flat gap in the pre-treatment period supports the validity of the ASCM counterfactual, while a large gap post-treatment indicates the strength of the policy's causal effect on Germany's TFR.

Given the different model specifications across donor pool, predictors, and special predictors, applying these tests allows us to evaluate the stability of estimated treatment effects and to identify models that are well-balanced, minimally biased, and consistent with theoretical expectations. Looking specifically at the predictor specification of Net Migration, we informally test whether migration-related demographic shifts mediate the observed policy effect.

4. RESULTS AND DISCUSSION

4.1. Model Comparison & Selection

Table 2 summarizes the results across these eight different model specifications, including the Average Treatment Effect on the Treated (ATT), standard error (SE), placebo-based p-values, and the MSPE ratio.

Table 2 – Models Specifications Metrics

Donor Pool	Predictors	Special Predictor	ATT	SE	p-value	MSPE Ratio
Full	Baseline	11 years	0.046	0.017	0.222	17.506
Full	Baseline	1 mean	0.043	0.017	0.111	14.963
Full	Adjusted	11 years	0.027	0.013	0.222	9.211
Full	Adjusted	1 mean	0.018	0.011	0.222	5.065
Strict	Baseline	11 years	0.002	0.011	0.800	3.629
Strict	Baseline	1 mean	-0.015	0.009	0.800	1.684
Strict	Adjusted	11 years	-0.003	0.010	0.800	3.962
Strict	Adjusted	1 mean	-0.022	0.009	0.800	2.344

The model using the Full donor pool, Baseline predictor set, and 1-mean special predictor structure emerges as the most balanced specification across the key metrics. This model yields an ATT estimate of 0.043, indicating an average increase of approximately 0.043 children per woman in Germany's TFR following the introduction of Elterngeld. Among all specifications, it also produces the lowest standard error (0.017) and the lowest p-value (0.111), suggesting that Germany's post-treatment divergence from its synthetic counterpart is more accentuated than in most placebo cases. The MSPE ratio of 14.96 indicates a strong signal, where the model fit Germany well in the pre-treatment period, and the post-2007 deviation is substantial by comparison. Although the model with the Full donor pool, Baseline predictor set, and a 11-year-mean special predictor structure presents strong results in metrics such as the highest ATT (0.046) and MSPE ratio (17.506), it's exceedingly large p-value (0.222) discards it from being a more balanced foundation model choice.

To explore whether the estimated effect is robust to migration-based demographic changes, the Adjusted predictor set (including net migration) was tested using the same special predictor structure as the main model choice. The corresponding model (Full, Adjusted, 1-mean) produces a lower ATT of 0.018, with a slightly smaller standard error (0.011), an MSPE ratio of 5.07, however a higher p-value (0.222) which casts doubt on the statistical significance of the model's results. This reduction in effect magnitude modestly suggests that part of the observed increase in fertility may be attributed to demographic shifts, particularly the rising share of foreign-born women, whose fertility planning behavior may differ from the native-born population and may not respond to national policy in the same way.

In contrast, specifications using the Strict donor pool consistently produce near-zero or even slightly negative ATT estimates, with high p-values (0.800) and low MSPE ratios (below 4). These results indicate weak or no treatment signal and reinforce the importance of using a larger donor pool to capture relevant synthetic counterfactual trajectories.

Based on these criteria, the Full–Baseline–1 mean model is selected as the main specification for reporting the estimated impact of Elterngeld. The Full–Adjusted–1 mean model is considered as a robustness specification to assess the potential mediating role of migration. The following sections examine the quality of model fit and discuss the substantive interpretation of these results in greater detail.

4.2. Synthetic Germany: Fit & Validation

Taking a closer look at the Full–Baseline–1 mean model, Figure 3 depicts the evolution trajectory of TFR for both synthetic and real Germany. The strong pre-treatment fit is visually corroborated by the figure and by the low pre-MSPE value of 0.001, evidencing a strong overlap between the two before the treatment. For the post-treatment fit, a clear divergence after the implementation of the Elterngeld policy is noted after 2007 which explains the higher post-MSPE of 0.016. The MSPE ratio of 14.96 further indicates a strong post-treatment signal relative to the pre-treatment fit, strengthening confidence that the observed effect is not a product of random variation or model misfit. The estimated ATT of 0.043 suggests that the introduction of Elterngeld led to an average increase of approximately 0.043 births per woman in Germany's TFR.

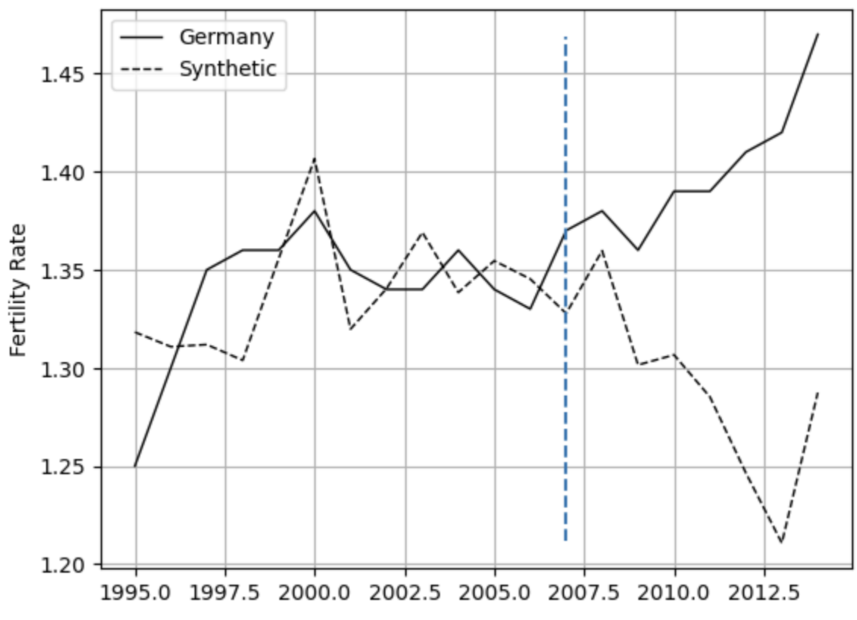

Figure 3: Main model Synthetic vs Real German TFR trajectory (1995-2014)

To validate the synthetic control's pre-treatment fit, Table 3 compares mean values (1995–2006) of the key predictors between Germany and its synthetic counterpart. The synthetic Germany closely matches Germany's fertility rate (1.338 vs. 1.341) and economic indicators (GDP per capita within 5%), though slight discrepancies exist in labor market variables (e.g., unemployment: Germany = 9.30%, synthetic = 7.28%). Notably, net migration differs substantially (59,979 vs. 98,167), justifying its inclusion in robustness checks. This alignment confirms the ASCM's ability to replicate Germany's pre-policy demographic and economic profile, strengthening causal claims despite minor imbalances in urban population and inflation rates.

Table 3 - Pre-Treatment Balance of Predictors (1995–2006)

Variable	**Germany**	**Synthetic Germany**
GDP per capita (PPP, constant $)	28,263.72	26,826.92
Female Labor Force Participation Rate	73.83	71.75
Unemployment Rate	9.30	7.28
Urban Population Rate	75.06	76.30
Net Migration	59,979.33	98,166.70
Inflation (consumer prices)	1.44	1.60
Average Fertility Rate	1.338	1.341

The weights behind the construction of the counterfactual as represented in Table 4, show that weights are concentrated in low-risk countries (Spain, Switzerland, Portugal). The potentially contaminated ones (e.g., Belgium, Netherlands, Ireland) have low or negative weights, so their influence is minimal. Negative weights reflect ASCM’s flexibility in outcome modeling and are used to improve pre-treatment fit by adjusting extrapolation bias (Ben-Michael et al., 2021) and can be interpreted as part of the ridge-augmented adjustment to minimize pre-treatment error, particularly when extrapolating from countries with otherwise dissimilar trends.

Table 4 – Model Weights

Country	Weight
Portugal	0.325
Switzerland	0.581
Belgium	-0.177
Spain	0.839
Italy	-0.469
Netherlands	0.172
Ireland	-0.095
Slovenia	-0.177

The statistical credibility of this model's estimate is therefore supported by a low standard error (0.0166) and the lowest placebo-based p-value (0.111) among all specifications. Following Reich (2024) interpretation of the gaps plot applied in SCM, the black line in Figure 4 evidences the difference in TFR between Germany and its synthetic counterfactual while the grey lines show the discrepancy in TFR between each donor country and its respective synthetic version in each placebo run. By 2014, the gap reached nearly +0.2, meaning the observed fertility rate is 0.2 units higher than the synthetic Germany. Only one of the eight donor countries exhibited a post-treatment fertility change larger than Germany's, leading to the p-value of 0.111. Although this does not meet the conventional 5% significance level, it suggests that the observed treatment effect is relatively rare among comparable countries, strengthening the argument that the Elterngeld policy may have contributed to the fertility increase in Germany.

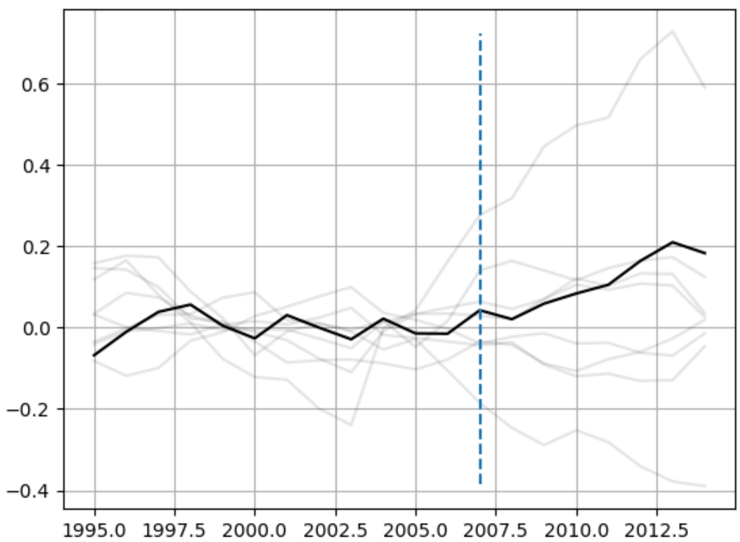

Figure 4: Placebo Plot

To explore whether the estimated effect is robust to compositional changes in the population due to migration, the Adjusted predictor set (including net migration) was tested using the same special predictor structure and is compared against the main chosen model specification. The Full–Adjusted–1 mean model produces a lower ATT of 0.018, with a slightly smaller standard error (0.011) and an MSPE ratio of 5.07.

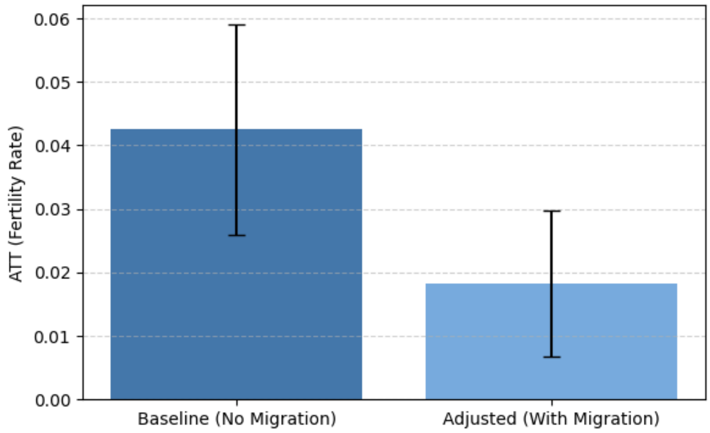

Figure 5: Estimated Policy Effect of Elterngeld

The higher p-value (0.222) indicates that the observed treatment effect is less distinct relative to placebo units. In this setting at least two of the eight donor countries exhibited a post-treatment fertility change as large or larger than Germany's, which questions its overall statistical significance. Given Germany's geopolitical context pertaining to migration policies, despite the questionable significance of the results, the notable reduction in effect size, from 0.043 to 0.018 as evidenced in Figure 5, suggests that migration likely mediated part of the fertility increase observed after 2007.

4.3. Policy Implications & Limitations

Our analysis indicates that the Elterngeld policy had a modest yet positive impact on Germany's TFR suggesting an average yearly increase of approximately 0.043 additional births per woman in the main model specification during the study period (2007-2014). Our findings are aligned with previous literature, such as McDonald (2006), evidencing that financial incentives can influence fertility behavior, though typically to a limited extent. However, fully interpreting this result requires careful attention to both demographic theory and broader methodological constraints.

A key limitation concerns the tempo versus quantum distinction in fertility analysis. Period fertility indicators like the TFR, may reflect shifts in the timing of births (tempo effects) rather than a sustained increase in the total number of children born per woman (quantum effects). The estimated increase following Elterngeld may partly represent birth postponement reversal or timing adjustments by parents seeking to maximize benefits, rather than a structural shift in completed fertility meaning that such policies can mask underlying low fertility trends by creating temporary spikes. Moreover, the study focuses on period fertility trends and does not account for cohort fertility (lifetime births of specific groups of people), which would provide a more accurate measure of long-term reproductive behavior. This is a notable limitation, as some of the effects of family policies like Elterngeld may only be observable over a longer time horizon.

From a methodological standpoint the ASCM helps mitigate biases through careful donor pool selection and ridge-augmented weight estimation, however it cannot fully eliminate the risk to causal inference of omitted variable bias. Unobserved or difficult to quantify factors such as cultural, economic, or policy shifts that differ between Germany and donor countries may still confound the overall estimates. The inclusion of net migration as a predictor in robustness checks reveals some attenuation of the treatment effect, suggesting that demographic shifts related to migration may partially mediate the observed increase in fertility. This highlights the importance of considering population composition and broader geopolitical contexts when evaluating fertility-related interventions.

The external validity of the findings should be treated with caution as Germany's institutional, economic, and demographic context makes it difficult to generalize results to other developed countries with different family support systems or cultural norms around family structures. However, from a policy perspective, our findings reinforce the position that financial incentives alone are unlikely to reverse long-term fertility decline in developed nations. While we show that Elterngeld appears to have had an impact, its magnitude is modest. To reverse persistently downward fertility trends, policies likely need to be considered in combination within a broader framework of structural supports, including affordable and accessible childcare, flexible work arrangements, and more generous gender-equal parental leave. These measures not only reduce the opportunity cost of having children but also promote more equitable sharing of caregiving responsibilities, factors shown to be critical in enabling family formation in modern societies (Lutz, O'Neill, and Scherbov, 2003; Balbo et al., 2013). Nevertheless, aspects of the Elterngeld policy, such as income-based parental leave and incentives for shared caregiving, may offer transferable lessons for countries with similar labor market structures and welfare regimes. In more traditional or less generous systems, however, such effects are unlikely to emerge in isolation from broader family support reforms.

In summary, while the Elterngeld policy appears to have had a small positive effect on Germany's fertility, its broader impact may depend on integration with a more comprehensive package of family-oriented social policies.

5. Conclusions and future works

This study set out to evaluate the causal impact of Germany's 2007 Elterngeld policy reform on national fertility trends using the ASCM. Drawing on World Bank Open Data, the analysis modeled Germany's TFR from 1995 to 2014, comparing it to a synthetic counterfactual.

The ASCM framework was applied across eight unique model specifications, where they varied in either donor pool composition, predictor sets, or special predictor structures. Placebo tests and robustness checks were employed to further assess the validity and stability of results. The main model specification (Full donor pool, Baseline predictors, 1-mean special predictor) estimated an ATT of approximately 0.043 births per woman, with strong pre-treatment fit and a relatively low placebo-based p-value (0.111). Adjusted models controlling for migration showed a reduced ATT of 0.018 but a higher p-value (0.222), suggesting that demographic shifts, particularly increased immigration levels, may partially mediate the observed effect. While the results stop short of definitive statistical significance at conventional thresholds, the consistency of direction and magnitude across the multiple distinct specifications strengthens the argument that Elterngeld had a modest positive effect on Germany's TFR during the post-treatment period.

There are several possible avenues pertaining to future research. First, a formal mediation analysis could better disentangle the role of migration in moderating policy impacts, given that our synthetic control-based analysis is limited in this sense. Second, subgroup analyses by age, education, or nationality could reveal heterogeneous treatment effects that are masked in aggregate TFR trends. Third, extending the analysis to cohort fertility trends and incorporating more recent reforms, such as Elterngeld Plus introduced in 2015, would provide a richer and more long-term picture of the actual policy effectiveness of parental allowance.

In conclusion, this study contributes to the growing body of demographic policy evaluation using modern causal inference tools, offering both methodological rigor and relevant insights into the effectiveness of fertility-boosting policies in low-fertility contexts for developed nations. While financial incentives like Elterngeld appear to play a role, the evidence suggests they must be embedded within a broader system of family support and addressing of deeper societal structure issues to produce lasting demographic positive change.

BIBLIOGRAPHICAL REFERENCES

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature, 59*(2), 391–425. https://doi.org/10.1257/jel.20191450

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

Balbo, N., Billari, F. C., & Mills, M. (2013). Fertility in advanced societies: A review of research. *European Journal of Population, 29*(1), 1–38. https://doi.org/10.1007/s10680-012-9277-y

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *NBER Working Paper No. 27738*. National Bureau of Economic Research. https://doi.org/10.3386/w27738

Bloom, D. E., Canning, D., & Fink, G. (2011). Implications of population aging for economic growth. *NBER Working Paper No. 16705*. National Bureau of Economic Research. http://www.nber.org/papers/w16705

BMFSFJ. (2015). *Elterngeld Plus: Für Mütter und Väter*. Federal Ministry for Family Affairs, Senior Citizens, Women and Youth. https://www.bmfsfj.de/

Esping-Andersen, G. (2001). *The social foundations of post-industrial economies*. Oxford University Press.

Fitzenberger, B., & Seidlitz, A. (2024). Changing fertility and heterogeneous motherhood effects: Revisiting the effects of a parental benefits reform. *SSRN*. https://ssrn.com/abstract=4820806

Fordham, S. (2022). *pysyncon: A Python package for the Synthetic Control Method* [Computer software]. GitHub. https://github.com/sdfordham/pysyncon

World Bank. (2024). *World Development Indicators* [Data set]. https://data.worldbank.org/

Gietel-Basten, S., Han, X., & Cheng, Y. (2019). Assessing the impact of the "one-child policy" in China: A synthetic control approach. *PLoS ONE, 14*(11), e0220170. https://doi.org/10.1371/journal.pone.0220170

Hart, R. K., Bergsvik, J., Fauske, A., & Kim, W. (2023). Causal analysis of policy effects on fertility. *RePEc Working Paper*.

https://www.researchgate.net/publication/374916474_Causal_Analysis_of_Policy_Effects_on_Fertility

Hellstrand, J., Nisén, J., Miranda, V., Fallesen, P., Dommermuth, L., & Myrskylä, M. (2020). Not just later, but fewer: Novel trends in cohort fertility in the Nordic countries. *MPIDR Working Paper WP-2020-007*. https://doi.org/10.4054/MPIDR-WP-2020-007

Hochschild, A. R. (1997). *The time bind: When work becomes home and home becomes work*. Metropolitan Books.

Kluve, J., & Tamm, M. (2013). Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: Evidence from a natural experiment. *Journal of Population Economics, 26*(3), 983–1005. https://doi.org/10.1007/s00148-012-0441-1

Lee, R., Mason, A., Amporfu, E., An, C. B., Bixby, L. R., Bravo, J., ... Zhang, Q. (2014). Is low fertility really a problem? Population aging, dependency, and consumption. *Science, 346*(6206), 229–234. https://doi.org/10.1126/science.1250542

Lutz, W., O'Neill, B. C., & Scherbov, S. (2003). Europe's population at a turning point. *Science, 299*(5615), 1991–1992. https://doi.org/10.1126/science.1080316

Lutz, W., Skirbekk, V., & Testa, M. R. (2007). The low-fertility trap hypothesis: Forces that may lead to further postponement and fewer births in Europe. *Vienna Yearbook of Population Research, 2006*, 167–192. https://doi.org/10.1553/populationyearbook2006s167

McDonald, P. (2000). Gender equity in theories of fertility transition. *Population and Development Review, 26*(3), 427–439. https://doi.org/10.1111/j.1728-4457.2000.00427.x

McDonald, P. (2006). Low fertility and the state: The efficacy of policy. *Population and Development Review, 32*(3), 485–510. https://doi.org/10.1111/j.1728-4457.2006.00134.x

Morgan, S. P., & Taylor, M. G. (2006). Low Fertility at the Turn of the Twenty-First Century. *Annual Review of Sociology, 32,* 375–399. https://doi.org/10.1146/annurev.soc.31.041304.122220

OECD. (2024). *State of immigrant integration - Germany*. https://www.oecd.org/migration/integration-indicators/

Podör, A., Skracic, A., & Garz, M. (2024). Addressing Europe's declining fertility rates: A case study of housing support policy in Hungary. *SSRN*. https://doi.org/10.2139/ssrn.4824144

Raute, A. (2019). The impact of income and employment transfers on fertility: Evidence from the German Elterngeld reform. *Journal of Public Economics, 172*, 124–150. https://doi.org/10.1016/j.jpubeco.2018.12.002

Reich, G. (2024). Determining the impact of the 2004 Australian Baby Bonus on fertility rates using a synthetic control analysis. *Australian Economic Papers, 63*(S1), 23–32. https://doi.org/10.1111/1467-8454.12345

Spiess, C. K., & Wrohlich, K. (2006). *The parental leave benefit reform in Germany: Costs and labour market outcomes of moving towards the Scandinavian model* (Discussion Paper No. SP I 2006-10). DIW Berlin. https://www.diw.de/documents/publikationen/73/54399/dp06010.pdf

NOVA IMS
Information Management School