Discussion Papers in Economics

No. 01/2026

Unconditional Cash Transfers and Maternal Labor Supply: Evidence from Bavaria's Family Allowance

Fabian Böhme
FAU Erlangen-Nürnberg

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
https://www.iwf.rw.fau.de/discussion-papers/

Unconditional Cash Transfers and Maternal Labor Supply: Evidence from Bavaria's Family Allowance*

Fabian Böhme†

June 2026

Preliminary Draft – Please Do Not Cite Without Permission

Abstract

This paper investigates the labor supply effects of unconditional cash transfers aimed at young families, using the Bavarian Family Allowance (*Familiengeld*), introduced in 2018. The policy provides €250–300 monthly to families with children aged 13–36 months regardless of employment status or childcare use. Combining difference-in-differences estimates from two complementary panel datasets, I find no statistically significant effect on maternal employment: the pooled estimate is a small and imprecise +4.2 percentage points (95% CI $[-1.8, +10.2]$). Because Bavaria is the only treated state, I complement the difference-in-differences with randomization inference over placebo states and a synthetic difference-in-differences design that matches Bavaria's pre-policy trajectory; both confirm that Bavaria's estimate falls within the normal range of untreated states, and the specifications that most directly address an imperfect pre-trend—individual fixed effects and the synthetic estimator—center on zero. A triple-difference specification points to a larger positive response among single and low-income mothers, but this pattern does not survive randomization inference and is best read as suggestive. Given that the unconditional design generates only a modest income effect without altering the relative price of employment, the evidence does not support concerns that transfers of this size meaningfully reduce maternal labor supply, nor does it establish a positive activation effect.

JEL Classification: J13, J22, H31, I38

*I am grateful to the SOEP Research Data Center at DIW Berlin and the Research Data Centre (FDZ) of the IAB for data access. I thank seminar participants at FAU Erlangen-Nürnberg for helpful comments. All errors are my own. This paper uses confidential micro data from PASS (Panel Study Labour Market and Social Security) and SOEP (German Socio-Economic Panel). Access requires approval from the respective data providers. AI-based tools were used for language editing.

†Friedrich-Alexander-Universität Erlangen-Nürnberg. Email: fabian.boehme@fau.de

Keywords: Family policy, maternal labor supply, cash transfers, difference-in-differences, heterogeneous treatment effects

1 Introduction

Cash transfers to families with young children represent a significant and growing component of social policy across developed economies.[1] These transfers serve multiple objectives: alleviating financial stress during the early childhood years, promoting gender equality, and potentially influencing fertility decisions (see Thévenon, 2011; OECD, 2011, for an overview of family policy objectives across the OECD). A fundamental question for policy design is whether unconditional cash transfers reduce parental—particularly maternal—labor supply and how they affect childcare utilization. While economic theory predicts that additional income should reduce labor supply if leisure is a normal good, the empirical magnitude of this effect remains contested, with estimates varying substantially across contexts and policy designs. This question has acquired renewed urgency in Germany, where the recurring debate over a Kindergrundsicherung (child basic income)—a reform proposal combining a universal guarantee with a means-tested supplement, shelved after the 2024 change in government—has prompted concerns that higher transfers may discourage parental employment. This paper provides direct evidence on the first channel: unconditional transfers of comparable magnitude have no detectable effect on maternal labor supply in either direction.

This paper contributes new evidence on this question by studying the Bavarian Family Allowance (*Bayerisches Familiengeld*), introduced in September 2018. The policy provides €250 per month for a family's first and second child and €300 per month from the third child onward, paid at a flat monthly rate for every child aged 13–36 months. Crucially, the benefit is \textit{unconditional} . This design distinguishes the Familiengeld from conditional "cash-for-care" programs studied elsewhere, which explicitly incentivize home care over formal childcare (Schøne, 2004; Kosonen, 2014; Gathmann and Sass, 2018).

The empirical strategy exploits Bavaria's unique position as the only German state to implement such a policy, creating a natural experiment amenable to difference-in-differences (DiD) analysis. I compare labor market outcomes of mothers with children aged 1–3 in Bavaria to mothers with same-aged children in other German states, before and after September 2018. A key methodological contribution is the combination of two complementary datasets: the Panel Study "Labour Market and Social Security" (PASS), which oversamples welfare recipients, and the German Socio-Economic Panel (SOEP), which is representative of the general population. Pooling these datasets yields a sample of 9,284 mother-year observations (4,514 from PASS, 4,770 from SOEP), providing sufficient statistical power to detect economically meaningful effects at the NUTS-1 level.

The central finding is the absence of a statistically significant effect on maternal employment. In the pooled sample with year and dataset fixed effects, the DiD coefficient

[1]OECD countries spent on average 2.35% of GDP on public family benefits in 2021, with substantial cross-country variation (OECD Family Database, Indicator PF1.1). Germany is among the higher spenders with roughly 3.5%, in part due to significant family-related tax breaks exceeding 0.8% of GDP.

is $+0.042$ (SE $= 0.030$), with a 95% confidence interval of $[-0.018, +0.102]$: the point estimate is positive but imprecise, and the minimum detectable effect at 80% power is 8.5 percentage points. The estimate is stable across reweighting (entropy balancing $+0.043$, inverse probability weighting $+0.044$) and sample restrictions, but attenuates toward zero once individual fixed effects are included (-0.004, SE $= 0.051$) or the post-period is restricted to 2019, before Bavaria's 2020 childcare subsidy (*Bayerisches Krippengeld*). Pretreatment trends are not perfectly flat: the joint F-test for the pre-period event-study coefficients rejects at the 5% level ($F(3, 5583) = 3.01$, $p = 0.029$). I show that this violation is confined to the SOEP subsample, is driven by small cell sizes rather than a genuine diverging trend, and disappears under individual fixed effects. A synthetic difference-in-differences estimator that matches Bavaria's pre-policy trajectory by construction yields $+0.046$ (95% CI $[-0.10, +0.19]$), and randomization inference over placebo states places Bavaria's estimate near the center of the placebo distribution ($p = 0.38$). Taken together, the evidence is most consistent with no detectable employment response.

Beyond the aggregate estimate, a triple-difference specification points to a larger positive response among single mothers, concentrated in the welfare-oversampled PASS sample where the transfer represents a larger income share. The triple interaction (Bavaria $\times$ Post $\times$ Single) is $+10.6$ percentage points in the pooled data ($p = 0.13$) and larger in PASS alone ($+26.7$ pp, clustered $p = 0.015$). I treat this as suggestive rather than established: under randomization inference over placebo states the interaction is not distinguishable from chance (pooled $p = 0.29$; PASS $p = 0.15$). The direction is nonetheless consistent with the idea that modest unconditional transfers relieve liquidity constraints that otherwise impede labor force attachment for financially vulnerable mothers (Marinescu, 2018; Jones and Marinescu, 2022), and with the finding of Gathmann and Sass (2018) that responses to German family-policy reforms are strongest among single and low-income households.

These results make three contributions to the literature. First, they provide causal evidence on the labor supply effects of unconditional family cash transfers in a high-income setting with well-developed childcare infrastructure. The absence of a negative effect stands in contrast to the employment reductions documented for conditional home care subsidies in Scandinavia (Schøne, 2004; Kornstad and Thoresen, 2007; Kosonen, 2014) and to the lower formal-childcare use—with employment declines concentrated among vulnerable subgroups—documented for Germany's conditional subsidy (Gathmann and Sass, 2018), consistent with the interpretation that *conditionality*—rather than the income transfer per se—is an important driver of behavioral responses in prior studies.

Second, by combining a welfare-oversampled panel (PASS) with a representative household survey (SOEP), the paper documents that the absence of an aggregate employment effect holds across the income distribution—not merely at the mean. This addresses a key limitation of prior studies, where conflicting findings across datasets raised concerns that results were driven by sample composition rather than genuine policy effects; once the employment definition is harmonized across the two surveys, they deliver mutually consistent estimates.

Third, the heterogeneity findings by household structure contribute to debates about how family policies interact with household resources. The differential response of single versus partnered mothers suggests that partner income availability mediates how households absorb cash transfers into their labor supply decisions.

The remainder of the paper proceeds as follows. Section 2 describes the institutional context and policy details. Section 3 reviews related literature. Section 4 presents the data, pooling strategy, and reweighting methods. Section 5 outlines the empirical strategy. Section 6 reports main results, robustness checks, and a power analysis. Section 7 explores heterogeneity. Section 8 concludes.

2 Institutional Background

2.1 The German Family Policy Landscape

Germany has a comprehensive system of family benefits designed to support parents during early childhood. The main federal programs include parental leave benefits (*Elterngeld*), which replace 65–67% of pre-birth earnings for up to 14 months; universal child benefits (*Kindergeld*) of approximately €200 per month per child; and means-tested social assistance (SGB-II/ALG-II) for low-income households. Since 2013, children aged one and above have a legal entitlement to a public childcare slot, though availability varies regionally.

Against this backdrop, Bavaria introduced the Family Allowance (*Familiengeld*) in September 2018 as a state-level supplement. The policy emerged from a broader Christian Social Union (CSU) platform emphasizing "freedom of choice" in childcare arrangements.

2.2 The Bavarian Family Allowance

The Familiengeld has the following key features:

Eligibility: All families with children aged 13–36 months residing in Bavaria qualify, regardless of income, employment status, or childcare usage. There is no means test, asset test, or requirement to forgo formal childcare.

\textit{Benefit Amount:} Families receive €250 per month for the first and second child and €300 per month from the third child onward. The amount depends on birth order, not on the child's age: it is paid at a flat monthly rate throughout the 13–36 month eligibility window. Benefits are additive for multiple eligible children. Over the 24-month eligibility window, a family with one child receives €6,000 in total (€7,200 per child from the third child onward).

Interaction with Other Benefits: The Familiengeld is explicitly exempt from means-testing for social assistance and housing benefits. It does not reduce entitlements to Kindergeld or Elterngeld.

Administration: Benefits are paid upon application to the Bavarian Center for Family and Social Affairs (*Zentrum Bayern Familie und Soziales*).

Policy Duration: Following a November 2024 cabinet decision and the 2026/27 state budget, Bavaria phased out the Familiengeld for children born from January 2025—together with the Krippengeld—and redirected the freed funds to childcare infrastructure.

2.3 Policy Distinctiveness and Concurrent Policies

The Familiengeld differs fundamentally from earlier German home care subsidies. The federal *Betreuungsgeld* (2013–2015) and the Thuringian home care subsidy studied by Gathmann and Sass (2018) were both *conditional*—families received benefits only if they did *not* use publicly subsidized childcare. These programs explicitly created a financial disincentive for formal childcare and maternal employment. The Familiengeld, by contrast, is paid regardless of childcare arrangements, generating a pure income effect without distorting the relative price of formal care.

Figure 1 illustrates the policy timeline, highlighting how the Familiengeld was the primary Bavaria-specific family policy introduced in 2018, while other policy changes (such as the Gute-Kita-Gesetz in 2019) were federal and thus affected all states equally.

Two additional Bavaria-specific policies were introduced during the post-treatment period. First, the *Beitragszuschuss für die Kindergartenzeit* began in April 2019, providing up to €100 per month toward kindergarten fees for children aged three and above—outside the Familiengeld's target age range of 13–36 months. Second, the *Bayerisches Krippengeld* started in January 2020, offering up to €100 per month to subsidize childcare costs for children under three. Unlike the unconditional Familiengeld, the Krippengeld is tied to formal childcare use and could theoretically *increase* maternal employment by reducing childcare costs. I address potential confounding from this policy in Section 6.5 by estimating effects separately for 2019 (before Krippengeld introduction) and confirm that the null result holds.

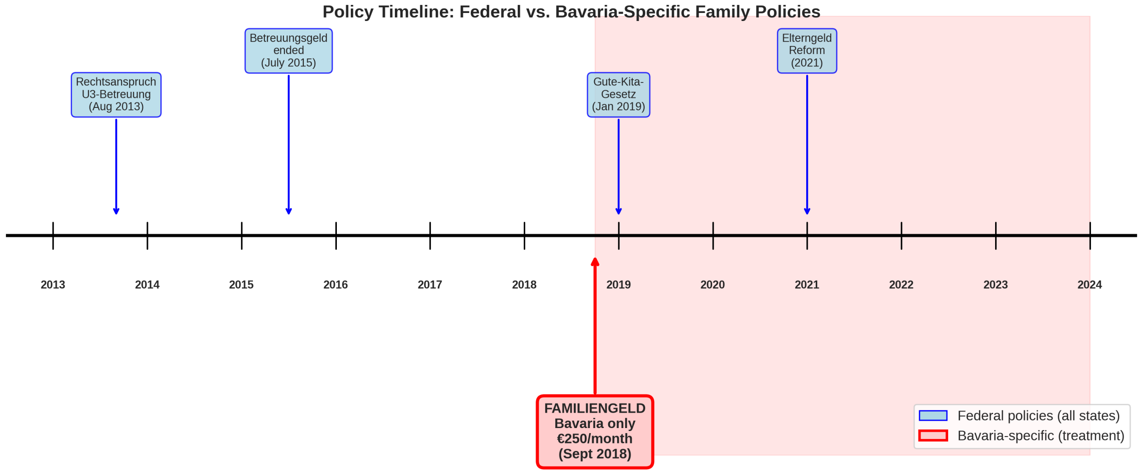

Figure 1: Policy Timeline: Federal vs. Bavaria-Specific Family Policies

Notes: The figure shows the timeline of major family policy changes in Germany. Federal policies (blue) affected all 16 states equally and are absorbed by year fixed effects. The Bavaria-specific Familiengeld (September 2018) is marked in red; two further Bavaria-specific measures—the Beitragszuschuss kindergarten subsidy (April 2019, for children 3+) and the Krippengeld childcare subsidy (January 2020, for children under 3)—are discussed in the text. The Gute-Kita-Gesetz (2019) was federal and thus affected all states. Robustness checks restricting to 2019 isolate the Familiengeld effect before Krippengeld introduction.

This design feature has important implications for predicted behavioral responses. Standard labor supply theory predicts that an unconditional income transfer reduces labor supply through the income effect alone. Conditional cash-for-care policies additionally increase the effective price of formal childcare (and hence work), amplifying the expected employment reduction. The Familiengeld thus provides a cleaner test of how income transfers *per se* affect maternal employment, holding childcare incentives constant. While this design isolates the income effect from the childcare price effect, differences in institutional context between Bavaria and the settings studied in prior work mean that cross-study comparisons should be interpreted with caution.

3 Related Literature

This paper contributes to three strands of research: the effects of cash-for-care policies, the relationship between childcare costs and maternal labor supply, and the broader literature on unconditional cash transfers.

3.1 Cash-for-Care Policies

Table 1: Comparison with Prior Cash-for-Care Policies

Policy	Country	Conditional?	Amount	Employment effect
Cash-for-care	Norway	Yes (no daycare)	∼€300/month	≈ −4–5%
Home care allowance	Finland	Yes (no daycare)	∼€100/month suppl.	−3 pp per €100
Betreuungsgeld	Germany (federal)	Yes (no daycare)	€100–150/month	Reduced daycare
Thuringia subsidy	Germany	Yes (no daycare)	€150/month	−8 pp daycare
Familiengeld	**Bavaria**	**No**	**€250–300/month**	**≈0 (n.s.)**

Notes: Summary of cash-for-care policies and their effects on maternal labor supply. "Conditional" indicates that benefits were contingent on not using publicly subsidized childcare. The Bavarian Familiengeld is the only large-scale unconditional family cash transfer studied in Germany. Sources: Schøne (2004) for Norway; Kosonen (2014) for Finland; Gathmann and Sass (2018) for Thuringia.

Several European countries have implemented policies paying parents to care for young children at home. Schøne (2004) studies Norway's 1998 cash-for-care program, which paid approximately €300 monthly to parents of 1–2 year-olds not using subsidized daycare. He finds significant reductions in maternal employment, with larger effects for less-educated mothers. Naz (2004) and Kornstad and Thoresen (2007) confirm these findings using different identification strategies.

For Finland, Kosonen (2014) exploits municipal variation in home care allowance supplements and estimates that a €100 monthly increase reduces maternal employment by 3 percentage points. Unlike Bavaria's Familiengeld, however, the Finnish home-care allowance is conditional on *not* using public daycare and thus raises the effective price of formal care; the supplement Kosonen studies varies with local policy generosity rather than constituting a new statewide program.

In Germany, Gathmann and Sass (2018) analyze Thuringia's 2006 home care subsidy, which was conditional on not using public childcare. They find that daycare attendance declined by 8 percentage points with modest effects on maternal employment. The federal Betreuungsgeld (2013–2015) showed similar patterns during its brief existence before being struck down by the Constitutional Court.

The present study extends this literature by examining an *unconditional* transfer in a setting with high childcare availability. The null employment effect I document is consistent with the interpretation that the negative labor supply effects in prior studies stem partly from conditionality-induced childcare price effects rather than income effects alone, though the cross-country comparison does not permit a causal decomposition of these channels.

Table 1 summarizes the key distinction between the Bavarian Familiengeld and prior cash-for-care policies studied in the literature. While prior programs were typically conditional on *not* using formal childcare, the Familiengeld is unconditional—paid regardless of childcare arrangements. This design difference has important implications for predicted behavioral responses and helps explain why the present study finds null effects where prior studies found negative employment effects.

3.2 Childcare Costs and Maternal Labor Supply

A large literature examines how childcare costs affect maternal employment, though identification challenges have produced widely varying estimates (Blau and Currie, 2006). Quasi-experimental studies generally find modest responses. Lundin et al. (2008) and Brink et al. (2007) find little effect of reduced childcare prices on maternal employment in Sweden. Blau and Tekin (2007) find that child care subsidy receipt raises employment among low-income single mothers in the US.

Several studies exploit childcare expansions rather than price changes. Havnes and Mogstad (2011) find that Norway's universal childcare expansion had no effect on maternal employment, suggesting crowding-out of informal care arrangements. Bauernschuster and Schlotter (2015) study Germany's 2005 childcare expansion and find significant positive effects on maternal employment, particularly for mothers with children under three.

My findings complement this literature by showing that additional household income—in the form of unconditional transfers—does not reduce maternal employment in a setting where childcare is widely available. This suggests that mothers' employment decisions in contemporary Germany are relatively insensitive to income changes of the magnitude studied here (€250–300/month).

3.3 Unconditional Cash Transfers

A growing literature examines labor supply responses to unconditional cash transfers outside the family policy context. Imbens et al. (2001) study lottery winners and find modest labor supply reductions. Jones and Marinescu (2022) analyze the Alaska Permanent Fund Dividend and find small effects on employment. Marinescu (2018) surveys basic income experiments and concludes that unconditional transfers typically produce small or zero employment effects.

This literature informs the present study in two ways. First, it suggests that income effects on labor supply are generally small for transfers of modest magnitude. Second, it highlights the importance of transfer size relative to household income: the Familiengeld represents approximately 7–10% of median household income for families with young children, a magnitude where small responses are plausible.

4 Data

4.1 Data Sources

I combine two complementary German household panel surveys that together provide both statistical power and distributional coverage.

The *Panel Study "Labour Market and Social Security"* (PASS) is a longitudinal household survey conducted annually since 2006 by the Institute for Employment Research (IAB). PASS is specifically designed to study poverty and labor market dynamics, with a sampling strategy that deliberately oversamples households receiving social assistance (SGB-II). The data contain detailed information on employment, income, household composition, and welfare receipt. PASS distinguishes between the "BA Original" sample (drawn from welfare registers) and the "Microm" sample (general population). Approximately 84% of the PASS analysis sample consists of current or recent welfare recipients, making PASS particularly informative for examining effects among economically vulnerable households—the population where income effects on labor supply are theoretically expected to be largest.

The *German Socio-Economic Panel* (SOEP) is a representative longitudinal survey covering approximately 12,000 households annually since 1984, administered by the German Institute for Economic Research (DIW Berlin). Unlike PASS, SOEP does not oversample welfare recipients and is broadly representative of the German population. SOEP provides richer information on childcare arrangements, enabling analysis of effects on formal childcare utilization that is not possible with PASS alone.

The rationale for pooling is twofold. First, neither dataset alone provides sufficient statistical power for the subgroup analyses that are central to this paper: PASS alone yields only 579 Bavarian observations, and SOEP alone 721. Second, the complementary sampling designs allow me to test whether results hold across the income distribution. If the null employment effect emerged only in PASS (welfare-heavy) or only in SOEP (representative), this would raise concerns about sample composition driving the findings. The consistency of results across both datasets (Section 6.5) strengthens the credibility of the null.

4.2 Sample Construction and Panel Structure

I construct the analysis sample through the following steps. In both datasets, I identify mothers aged 18–55 with at least one child aged 1–3 in the survey year, using household roster information. In SOEP, mothers are identified via the *ppathl* tracking file combined with the *kidlong* child biography file. In PASS, I use the household composition module linking mothers to co-resident children. I restrict the sample to survey years 2014–2023, excluding 2018 as a transition year (Section 5 discusses this choice). Mothers observed in multiple waves enter as separate observations, yielding a long-format panel.

The resulting sample comprises 9,284 mother-year observations from 5,584 unique mothers: 4,514 observations from PASS (579 from Bavaria, 12.8%) and 4,770 from SOEP (721 from Bavaria, 15.1%). The higher Bavarian share in SOEP reflects Bavaria's population share more accurately, while the lower share in PASS reflects the survey's welfare-focused sampling frame—Bavaria has lower welfare receipt rates than most German states.

The panel is substantially unbalanced: with 5,584 unique mothers contributing 9,284 observations, the average mother appears in 1.66 waves. This is not primarily a consequence of survey attrition, but rather of the transient nature of the eligibility window: children age out of the 1–3 range within two to three years, mechanically limiting the number of waves in which any given mother can appear.[2] Mothers who remain in the sample for more waves are disproportionately those with multiple children or those whose children happen to span the transition year.

This structure raises two concerns. First, if panel persistence differs systematically between Bavaria and other states, compositional changes could confound the DiD estimate. I test for differential attrition by regressing an indicator for appearing in the subsequent wave on the Bavaria $\times$ Post interaction and find no significant effect ($p = 0.68$). Second, the individual fixed effects specification—which relies on within-person variation and thus uses only the subset of mothers observed in at least two waves—yields a coefficient of -0.004 (SE = 0.051), close to the pooled baseline (+0.042, SE = 0.030) and slightly nearer zero, providing reassurance that results are not driven by mothers with longer panel tenure.

Table 2 presents summary statistics by dataset and region.

[2] A mother whose only child turns one in 2019 can appear in the sample in 2019, 2020, and 2021, contributing up to three observations; a mother whose child turns one in 2016 contributes at most two (2016 and 2017), since the child is three in the excluded year 2018 and ages out thereafter. Mothers with multiple children may appear more frequently.

Table 2: Sample Summary Statistics

	PASS		SOEP		Pooled	
	Bavaria	Rest	Bavaria	Rest	Bavaria	Rest
Panel A: Sample Size						
Observations	579	3,935	721	4,049	1,300	7,984
Share of sample	12.8%	87.2%	15.1%	84.9%	14.0%	86.0%
Panel B: Demographics						
Age (mean)	31.8	31.7	33.9	33.4	32.9	32.6
Children aged 1–3	1.09	1.09	1.08	1.08	1.08	1.08
Partnered (%)	75.1	78.2	78.5	74.4	77.0	76.3
Panel C: Welfare Status						
Welfare receipt (%)	79.8	84.2	8.4	17.3	28.5	41.3
Panel D: Employment Rate (%)						
Overall (2014–2023)	34.7		35.7		35.3	
Pre-period (pooled)					33.8	32.7
Post-period (pooled)					42.3	37.2

Notes: Sample restricted to mothers aged 18–55 with at least one child aged 1–3. Year 2018 excluded as transition year. Employment is defined as any current employment at the survey date (full-time, part-time, or marginal), harmonized across the two surveys; under this definition PASS and SOEP exhibit nearly identical overall employment rates (34.7% vs. 35.7%). Panel D reports overall rates by dataset and, for the pooled sample, rates by region and period (the pooled pre/post change is +8.5 pp in Bavaria vs. +4.5 pp in the rest of Germany). Pooled statistics are computed over the combined PASS–SOEP sample (observation-weighted). Total pooled sample: 9,284 mother-year observations from 5,584 unique mothers.

Two patterns merit discussion. First, Bavarian mothers are substantially less likely to receive welfare (28.5% vs. 41.3%) than mothers in other states, reflecting Bavaria's stronger economy, while partnership rates are nearly identical (77.0% vs. 76.3%). Second, once the employment definition is harmonized across surveys—counting any current employment, whether full-time, part-time, or marginal—PASS and SOEP yield nearly identical employment rates (34.7% vs. 35.7%), and the dataset fixed effects in the baseline absorb the small residual level difference.[3] The harmonized rate of roughly 35% is closely in line with the *realized* employment rate ("realisierte Erwerbstätigkeit") that official Mikrozensus statistics report for mothers with a youngest child under three—around 36–38% over the study period—a concept that, like my main outcome, counts a mother as employed only if she is actively working at the survey date and excludes those on parental leave (*Elternzeit*). This is a deliberate choice: for a transfer paid precisely over the 13–36

[3] Counting full-time, part-time, and marginal (*Minijob*) employment symmetrically in both surveys is essential: a definition restricted to full-time work produces an implausibly low SOEP rate ($\approx 7\%$) that is inconsistent both with the PASS labor-market module and with official statistics, and that would render a pooled outcome incoherent.

month window, the return-from-leave margin is plausibly where behavior responds, and an ILO-style definition—which counts a mother on leave as employed—would mechanically obscure exactly that margin. Section 6.5 reports an ILO-consistent definition as a robustness check; it raises the employment level to about 55% in PASS but leaves the difference-in-differences conclusion unchanged.

4.3 Pooling and Reweighting

The different sampling designs of PASS and SOEP raise concerns that the pooled estimate could be driven by compositional differences between datasets rather than the policy effect. I address this through four approaches, all of which yield nearly identical results to the unweighted baseline (Table 13).

The baseline approach includes *dataset fixed effects* (ϕ_d), which absorb level differences between PASS and SOEP in employment rates and other outcomes. This allows the two samples to have different intercepts while constraining the treatment effect to be common across datasets. Because all specifications also include year fixed effects, the DiD coefficient is identified from within-dataset, within-year variation in the Bavaria–Rest differential.

As a first alternative, I estimate *inverse probability weights* (IPW) from a logit model predicting Bavaria residence from age, age squared, number of children, dataset, and year indicators. The pseudo-R^2 of this model is only 0.002, indicating that Bavaria residence is largely unpredictable from observables—a reassuring finding for the DiD design. Control observations are weighted by $\hat{p}/(1 - \hat{p})$; treated observations receive weight 1.

Second, I implement *entropy balancing* (Hainmueller, 2012), which reweights control observations to exactly match the first moments (means) of covariates in the treatment group. This approach guarantees covariate balance without relying on a correctly specified propensity score model.

Third, I estimate *propensity score matching* (5-nearest-neighbor with a caliper of 0.05 standard deviations). This approach is more conservative, retaining a smaller matched sample. The matched estimate is small and statistically insignificant (-0.009), but markedly less precise than the other specifications (SE = 0.094) because matching discards most control observations; I therefore treat it as the least informative of the reweighting checks.

Fourth, I implement *stratified matching* by creating cells defined by age group (4 categories), number of children (3 categories), dataset, and year. Estimates are computed only within cells containing both treated and control observations. This nonparametric approach ensures that comparisons are made between similar mothers without relying on functional form assumptions.

The consistency of results across unweighted estimation, IPW, entropy balancing, and stratified matching—all small and statistically insignificant—provides strong evidence that the absence of an effect is not driven by covariate imbalance or sample composition (Table 13 in Appendix A).

4.4 Pre-Treatment Balance and Variable Definitions

Table 3 presents formal tests of pre-treatment balance between Bavaria and other states. Mothers in both groups are statistically indistinguishable on number of children, employment rate, and partnership status in the pre-treatment period (2014–2017). Two differences emerge: Bavarian mothers are modestly older (by about 0.6 years, $p = 0.02$) and substantially less likely to receive welfare (-10.0 percentage points, $p < 0.001$), both reflecting Bavaria's stronger economy rather than selection into the sample; age and age squared enter as controls throughout. Importantly, the pre-treatment *employment rate*—the outcome variable—is balanced (a difference of $+1.1$ percentage points, $p = 0.57$), and reweighting methods that explicitly balance on welfare receipt yield nearly identical results to the unweighted estimates.

Table 3: Pre-Treatment Covariate Balance (2014–2017)

Variable	Bavaria Mean	Rest of Germany Mean	Difference	SE	p-value
Age	32.63	32.03	0.600	0.249	0.016**
Children aged 1–3	1.08	1.08	−0.005	0.011	0.629
Employment rate	0.338	0.327	0.011	0.019	0.568
Has partner	0.752	0.761	−0.009	0.018	0.616
Welfare receipt	0.311	0.411	−0.100	0.022	<0.001***

Notes: Means computed for the pre-treatment period (2014–2017) only. Differences tested using two-sample t-tests with unequal variances. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The primary outcome is *employment*, defined as an indicator equal to one if the mother reports being employed (full-time, part-time, or marginal) at the survey date. In PASS, this is derived from the labor market module (*Erwerbsstatus*); in SOEP, from the current employment status question in the individual questionnaire. Both variables capture employment at the time of interview rather than over a recall period, minimizing recall bias but implying that the outcome reflects a point-in-time snapshot. Secondary outcomes include weekly working hours (coded as zero for non-employed mothers), full-time employment ($\geq$30 hours/week), and part-time employment (1–29 hours/week). Childcare outcomes—Kita use, formal care hours, grandparental care—are available only in SOEP, which collects detailed childcare information. Welfare receipt is coded in PASS as belonging to the welfare-register-based subsample or reporting current SGB-II benefits, and in SOEP from the current ALG-II/social-assistance item in the individual questionnaire; the status is unobserved for about a quarter of observations, almost exclusively in SOEP, which reduces the estimation samples in the welfare-specific analyses of Section 7.

Treatment is defined as the interaction of Bavaria residence with the post-policy period (2019–2023). The pre-period spans 2014–2017, with 2018 excluded as a transition year. Control variables include mother's age, age squared, and the number of co-resident children aged 1–3. All specifications include year fixed effects to absorb aggregate trends and dataset fixed effects to absorb level differences between PASS and SOEP.

5 Empirical Strategy

5.1 Econometric Framework

The baseline specification estimates:

$$Y_{ist} = \beta \cdot \text{Bavaria}_s \times \text{Post}_t + \lambda \cdot \text{Bavaria}_s + \delta_t + \phi_d + X'_{ist}\theta + \varepsilon_{ist} \tag{1}$$

where Y_{ist} is the employment status of mother i in state s at time t. Bavaria_s indicates residence in Bavaria, and Post_t indicates the post-policy period (2019–2023). The coefficient of interest β captures the differential change in employment for Bavarian mothers relative to mothers in other states after policy introduction. The specification includes a Bavaria indicator (λ), year fixed effects (δ_t), and dataset fixed effects (ϕ_d). Control variables X_{ist} include age, age squared, and number of children. I estimate linear probability models throughout for ease of interpreting interaction terms in the heterogeneity analysis; marginal effects from logit specifications yield nearly identical point estimates and are available upon request.[4] The baseline specification is unweighted; Section 4.3 demonstrates that reweighting yields nearly identical estimates.

Treatment is defined at the *policy level*: all mothers with children aged 1–3 residing in Bavaria after September 2018 are classified as treated, regardless of whether their child's precise age falls within the 13–36 month eligibility window at a given survey date. This is an intent-to-treat approach that avoids endogenous selection into treatment based on exact child age. Because PASS and SOEP interviews occur at annual intervals, each mother contributes at most one observation per year, and the design captures the average effect of living in a policy environment where the Familiengeld is available rather than the instantaneous effect of benefit receipt.

The Familiengeld was introduced on September 1, 2018, but I define the post-treatment period as 2019–2023 and exclude 2018 entirely as a transition year. The policy was in effect for only four months of 2018 (September–December), and PASS and SOEP fieldwork periods span several months within each wave, meaning that some 2018 respondents were interviewed before September (untreated) and others after (partially treated). Rather

[4]The LPM can produce predicted probabilities outside $[0, 1]$, but this is not a concern in practice: fewer than 2% of fitted values fall outside the unit interval in the baseline specification.

than introduce measurement error by assigning heterogeneous treatment status within the 2018 wave, I drop the entire year.

To assess parallel trends and trace dynamic effects, I estimate an event study specification:

$$Y_{ist} = \sum_{k \neq 2017} \beta_k \cdot \text{Bavaria}_s \times \mathbf{1}[t = k] + \lambda \cdot \text{Bavaria}_s + \delta_t + \phi_d + X'_{ist}\theta + \varepsilon_{ist} \tag{2}$$

where k indexes calendar years with 2017 as the reference (final pre-treatment year). Pre-treatment coefficients ($\beta_{2014}, \beta_{2015}, \beta_{2016}$) should be close to zero if parallel trends hold. This is a *calendar-year* event study, not an individual-level event study around the child's entry into eligibility at 13 months (see Section 5.3 for discussion). The stability of post-treatment coefficients (β_{2019} through β_{2023}) provides additional evidence: it implies that cohort-specific treatment effects are similar across entry cohorts—mothers whose children became eligible in 2019 do not respond differently from those eligible in 2022—ruling out differential take-up or behavioral adaptation over time.

To examine treatment effect heterogeneity, I estimate triple-difference specifications:

$$Y_{ist} = \beta_1 \cdot (\text{Bav} \times \text{Post}) + \beta_2 \cdot (\text{Bav} \times G) + \beta_3 \cdot (\text{Post} \times G) + \beta_4 \cdot (\text{Bav} \times \text{Post} \times G) + \ldots \tag{3}$$

where G is a subgroup indicator (welfare status, partner status, age, child age, dataset). The triple interaction β_4 tests whether treatment effects differ across subgroups. I estimate five such interactions, which raises a multiple testing concern: at $\alpha = 0.05$ without adjustment, one would expect 0.25 false positives under the null. I report both unadjusted p-values and assess significance under the Benjamini-Hochberg procedure controlling the false discovery rate.[5]

5.2 Identification

The key identifying assumption is that employment trends would have evolved similarly in Bavaria and other states absent the policy—the parallel trends assumption. I assess this through: (1) visual inspection of pre-treatment trends; (2) joint F-tests of pre-treatment event study coefficients; (3) placebo tests with hypothetical reform dates; and (4) triple-difference specifications using different comparison dimensions. Below I discuss the main threats to identification.

The Familiengeld was introduced as part of a broader CSU political platform, raising

[5]The unadjusted p-values across the five triple interactions are 0.13 (single mother), 0.17 (welfare receipt), 0.60 (dataset), 0.04 (youngest child aged 2–3 vs. 1), and far from significance for the mother's-age split. Only the child-age interaction is significant at the 5% level without adjustment, and it does not survive the Benjamini–Hochberg procedure (its rank-one threshold is $0.05/5 = 0.01$). Where a subgroup estimate is suggestive—single mothers and the older child-age group—I assess or flag it accordingly (randomization inference for single mothers, Section 5.4).

the possibility that unobserved policy changes or economic shifts in Bavaria coincided with the reform. All major family policy changes during the study period—the Gute-Kita-Gesetz (2019), the Elterngeld reform (2021)—were federal and thus affected all states equally, absorbed by year fixed effects. The two Bavaria-specific concurrent policies warrant separate discussion. The Beitragszuschuss für die Kindergartenzeit (April 2019) targets children aged 3 and above—outside the Familiengeld's target age range—and should not directly affect mothers with children aged 1–3. The Bayerisches Krippengeld (January 2020) subsidizes formal childcare for under-3s and, being conditional on childcare use, could theoretically *increase* maternal employment, potentially masking a negative Familiengeld effect. Restricting the post-period to 2019 only—before Krippengeld introduction—yields a coefficient of -0.003 (SE $= 0.042$), statistically insignificant, indicating that the result is not an artifact of offsetting concurrent policies (Table 7, specification 9).

In contrast to a cleanly flat pre-trend, the joint test of the pre-period event-study coefficients rejects parallel pre-trends at the 5% level ($F(3, 5583) = 3.01$, $p = 0.029$), driven almost entirely by a single coefficient (2016). I probe this violation in detail in Section 6.5 and reach a reassuring conclusion: it is confined to the SOEP subsample (the PASS pre-trend is flat, $p = 0.61$), it is not an artifact of an unbalanced panel (it persists, indeed strengthens, on the balanced panel), and it disappears once individual fixed effects are included ($p = 0.22$). The SOEP–Bavaria cell contains only 38–116 mothers per year, so a small number of low-employment respondents in 2016 moves the cross-sectional mean; the pattern is a between-person composition effect in a thin cell rather than a genuine diverging Bavarian trend. Consistent with this, dropping 2016 lowers the baseline estimate from $+0.042$ to $+0.017$, and the synthetic difference-in-differences estimates discussed below do not overturn the conclusion. A separate concern is anticipation: the CSU announced the Familiengeld during the 2018 state election campaign, so some adjustment of labor supply before September 2018 is conceivable. Excluding 2019, the year most exposed to such spillover, yields $+0.057$ (SE $= 0.034$)—if anything slightly larger than the baseline rather than attenuated, which is not the pattern anticipation would produce.

If the Familiengeld attracted families to Bavaria or discouraged out-migration, the composition of Bavarian mothers might change endogenously after 2018. At €250–300 per month for a maximum of 24 months, the total transfer (€6,600) is unlikely to induce residential mobility across state borders, particularly given the high transaction costs of relocation. Pre-treatment balance on observables (Table 3) shows no evidence of differential selection, and the individual fixed effects specification—which absorbs all time-invariant individual characteristics, including the decision to reside in Bavaria—yields a coefficient of -0.004 (SE $= 0.051$), even closer to zero than the baseline.

Bavaria's economy is structurally stronger than the German average, with lower unemployment and higher per-capita GDP. If Bavaria-specific labor market trends diverged from the national trend after 2018 independently of the Familiengeld, this could confound the estimates. The external validation with Destatis administrative data (Appendix A.2) shows that maternal employment trends in Bavaria and West Germany remain broadly parallel throughout the study period, with both regions exhibiting a COVID-related dip in 2020 but no Bavaria-specific divergence. A specification including a Bavaria-specific linear time trend absorbs the differential pre-trend and yields -0.020 (SE $= 0.062$); while this over-controls for a treatment effect that could itself resemble a trend, its sign and insignificance reinforce that no positive effect survives once Bavaria is allowed its own trajectory. Because Bavaria is a single treated unit, I also estimate a synthetic difference-in-differences design (Section 6.5) that constructs a weighted combination of control states reproducing Bavaria's pre-policy employment path; the resulting ATT is $+0.046$ with a placebo-based 95% confidence interval of $[-0.10, +0.19]$, statistically indistinguishable from zero. A comparison with neighboring states (Baden-Württemberg, Hessen) in Appendix A.2 provides additional reassurance.

Finally, the stable unit treatment value assumption (SUTVA) requires that treatment of Bavarian mothers does not affect outcomes for mothers in other states. Two types of spillovers could violate this assumption. Information spillovers—media coverage of the Familiengeld influencing labor supply norms nationwide—would attenuate the estimated treatment effect by shifting the control group. Cross-border spillovers could affect mothers in states adjacent to Bavaria. Both channels would bias the estimate toward zero, making the null finding a conservative bound on the true effect. Restricting the control group to Bavaria's neighboring states (Baden-Württemberg, Hessen, Thuringia, Saxony)—the states most exposed to any such spillovers—yields $+0.051$ (SE $= 0.035$), close to the baseline. Moreover, at €250–300/month, the transfer is too small to generate meaningful general equilibrium effects on wages or childcare markets.

5.3 Alternative Identification Strategies

A natural alternative to the cross-state DiD design would exploit the eligibility threshold at 13 months of child age directly, comparing labor supply of mothers whose children have just crossed the threshold with mothers whose children are slightly younger. Such a regression discontinuity design would identify the causal effect of benefit receipt without requiring cross-state variation or the parallel trends assumption.

I do not implement this approach for three reasons. First, both PASS and SOEP record child age in completed years at the time of interview, not in months, precluding construction of a precise running variable measuring distance to the 13-month eligibility cutoff.[6] Second, with annual survey intervals, employment outcomes are observed at a sin-

[6] The SOEP records the child's birth year and month in auxiliary files, but interview dates are spread across several months within each wave. Computing the child's exact age in months at the interview date is feasible in principle but introduces substantial measurement error, as the interview month varies

gle point per year rather than continuously around the threshold, introducing additional measurement error that would attenuate any discontinuity estimate. Third, statistical power would be severely limited: the Bavarian subsample contains approximately 1,300 observations over nine years, and restricting to a narrow bandwidth around the 13-month cutoff would yield far too few observations for credible inference.

The same data limitation prevents a clean analysis of treatment intensity. The Familiengeld pays a flat monthly amount per child throughout the 13–36 month window (€250 for the first and second child, €300 from the third), so the monthly benefit does not vary with the child's age; what varies with age is the remaining eligibility duration. A mother whose child is 14 months old faces nearly two years of future benefit receipt, while a mother with a 35-month-old has only one month remaining. In principle, this dosage variation could be exploited, but because child age is recorded in years rather than months, I cannot distinguish between a 13-month-old and a 23-month-old within the "child aged 1" category. I examine heterogeneity by broad child age groups (age 1 vs. age 2–3) in Section 7; because child age is measured only in completed years, this split is coarse and cannot isolate dosage responses within the "child aged 1" category.

An analysis using administrative data with monthly precision—such as the Integrated Employment Biographies (IEB) from the IAB, linked to birth registry or benefit receipt records—could implement both the discontinuity design and the dosage analysis, and constitutes a promising avenue for future research. Such data would also permit exact measurement of take-up and benefit duration, enabling a treatment-on-the-treated analysis complementary to the intent-to-treat estimates presented here.

5.4 Inference

Standard errors are clustered at the individual level throughout the analysis, accounting for within-person serial correlation across 5,584 unique mothers. A potential concern is that treatment varies at the state level (Bavaria vs. rest of Germany), which might suggest state-level clustering. However, the pooled PASS-SOEP data identifies only two treatment groups (Bavaria and non-Bavaria), making state-level clustering statistically inappropriate—cluster-robust standard errors and the wild cluster bootstrap both perform poorly with fewer than approximately 30–50 clusters (see Cameron and Miller, 2015, for discussion).

This is a genuine limitation of the research design. What individual-level clustering addresses is within-person serial correlation, the primary source of non-independence in this setting. What it does *not* address is potential correlation of unobserved shocks across mothers within the same state. I mitigate this concern in three ways. First, I demonstrate robustness across 16 alternative specifications (Table 18, with the main set in Table 7),

across respondents and is not always precisely recorded.

including individual fixed effects, various reweighting methods, and sample restrictions—all yielding small and statistically insignificant estimates. If a non-zero effect were present and merely obscured by a particular inference choice, we would expect at least some of these specifications to produce significant results. Second, while the pre-treatment event-study test rejects perfectly parallel trends (Table 5), Section 6.5 shows that the violation is a small-cell composition artifact confined to SOEP that vanishes under individual fixed effects, and the placebo, synthetic difference-in-differences, and neighboring-state analyses are consistent with the absence of a Bavaria-specific shock. Third, the external validation with Destatis data in Appendix A.2 provides reassurance that no differential aggregate shock affected Bavaria during the study period.

Robustness across specifications, however, does not substitute for valid inference when only a single unit is treated. The most direct response in this setting is randomization inference, which reassigns the treatment to each control state in turn and locates the Bavarian estimate within the resulting placebo distribution (Conley and Taber, 2011). With sixteen states the achievable resolution is coarse—the smallest attainable two-sided p-value is approximately 1/16—but for a null result this is informative, as it asks whether Bavaria's estimate is unusual relative to states where no policy was introduced. Implementing this procedure, the Bavarian main estimate sits near the center of the placebo distribution: its absolute value ranks sixth among the sixteen state-level estimates ($p = 6/16 \approx 0.38$), and fifth among the fourteen states with at least 150 observations ($p = 5/14 \approx 0.36$). For the single-mother triple interaction the same procedure yields $p = 0.29$ in the pooled sample and $p = 0.15$ within PASS. Randomization inference thus reinforces both conclusions: Bavaria is unremarkable relative to placebo states for the main effect, and the single-mother interaction—though sizeable as a point estimate—is not statistically distinguishable from what placebo reassignment produces by chance.

Recent econometrics literature has highlighted potential biases in two-way fixed effects DiD estimators when treatment timing is staggered (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). These concerns are less relevant here because treatment timing is *not* staggered—all Bavarian mothers are treated simultaneously in September 2018—and the comparison is between a single treated unit (Bavaria) and a clean never-treated control group. In this canonical DiD setting, the Goodman-Bacon decomposition reduces to a single comparison, and concerns about forbidden comparisons do not apply. Nonetheless, I verify that year-specific treatment effects are stable across the post-treatment period (2019–2023), confirming that the two-way fixed effects estimator does not produce misleading averages.

6 Main Results

6.1 Baseline Estimates

Table 4 presents the main difference-in-differences estimates for employment. Across all specifications, the DiD coefficient is positive, modest in size, and statistically insignificant.

Table 4: Effect of Bavarian Family Allowance on Maternal Employment

	(1) Simple DiD	(2) + Controls	(3) + Year FE	(4) + Dataset FE
Bavaria × Post	0.040	0.043	0.042	0.042
	(0.032)	(0.031)	(0.031)	(0.030)
Controls	No	Yes	Yes	Yes
Year FE	No	No	Yes	Yes
Dataset FE	No	No	No	Yes
Observations	9,284	9,284	9,284	9,284
R^2	0.003	0.065	0.073	0.073

Notes: Outcome is employment (0/1). Sample: mothers aged 18–55 with children 1–3 in pooled PASS-SOEP data, 2014–2023 (excluding 2018). Controls: age, age squared, number of children. Standard errors clustered at individual level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The preferred specification (column 4) yields a point estimate of +0.042 with a standard error of 0.030, implying a 95% confidence interval of [−0.018, +0.102]. The point estimate is positive but not statistically significant ($p = 0.17$); the interval includes zero but also fairly large positive effects, so the data are most consistent with a small or zero effect rather than a precisely estimated null. Sections 6.5 and 5.2 show that the specifications addressing the imperfect pre-trend most directly (individual fixed effects, synthetic difference-in-differences) move the estimate toward zero.

6.2 Power Analysis

To interpret the null finding, it is important to assess whether the study has adequate statistical power to detect economically meaningful effects. The minimum detectable effect (MDE) at 80% power with a two-sided test at $\alpha = 0.05$ is:

$$\text{MDE} = (z_{0.975} + z_{0.80}) \times \text{SE} = (1.96 + 0.84) \times 0.0304 \approx 0.085 \tag{4}$$

Thus, the design can detect effects as small as 8.5 percentage points at 80% power. For context, Kosonen (2014) finds that a €100/month increase in Finland's home care allowance reduces maternal employment by 3 percentage points; scaling to the €250–300/month Familiengeld would predict effects of 7.5–9 percentage points under similar elasticities. The study therefore has adequate power to detect a labor-supply response of the magnitude implied by the conditional-subsidy literature, but not enough to distinguish a small positive or negative effect from zero.

The 95% confidence interval [−1.8pp, +10.2pp] excludes negative effects larger than about 2 percentage points but remains consistent with positive effects up to roughly 10 percentage points. The data thus speak more clearly against a sizeable *reduction* in maternal labor supply than against a modest positive effect.

6.3 Event Study Results

Figure 2 displays event study coefficients from equation (2). The pre-treatment coefficients are mostly small, but the 2016 coefficient is negative and individually significant, and the joint test reported below rejects perfectly parallel pre-trends. Post-treatment coefficients drift gradually upward, with the 2023 coefficient reaching +0.11, though they are jointly insignificant.

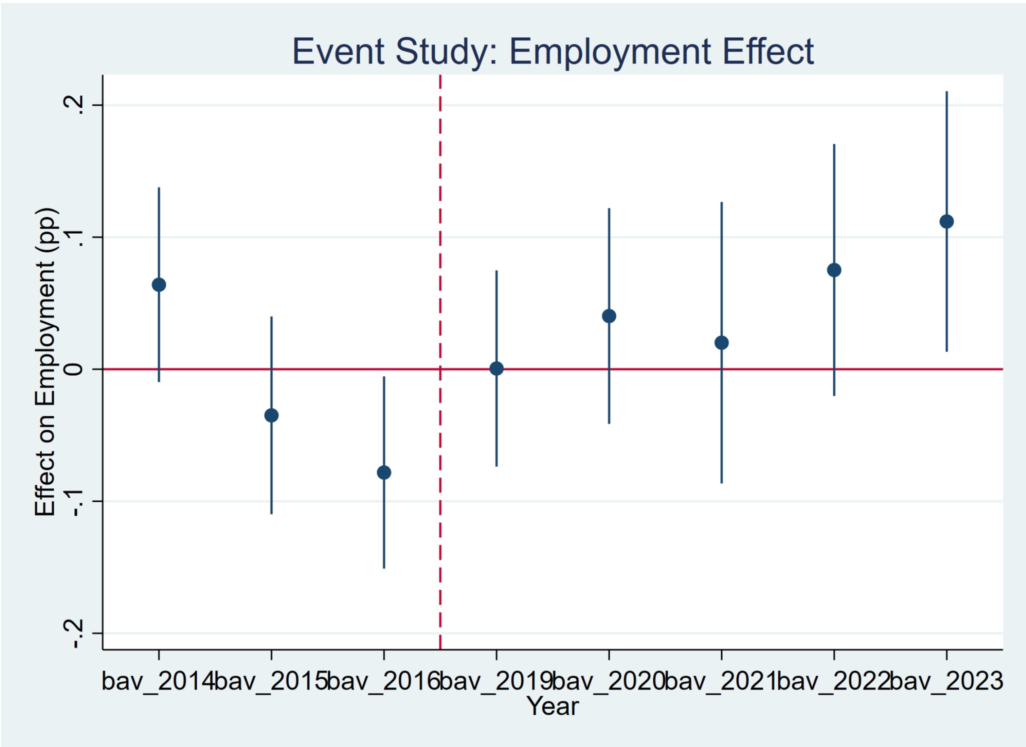

Figure 2: Event Study: Effect of Bavarian Familiengeld on Maternal Employment
Notes: The figure plots coefficients from an event study regression of maternal employment on Bavaria × Year interactions, with 2017 as the reference year. Controls include age, age squared, and number of children. Year and dataset fixed effects included. The dashed vertical line indicates the policy introduction (September 2018). 95% confidence intervals shown. Standard errors clustered at individual level.

Table 5 presents the numerical event study coefficients. The joint F-test for the pre-treatment coefficients yields $F(3, 5583) = 3.01$ with $p = 0.029$, rejecting the null of parallel pre-trends at the 5% level. As Section 6.5 documents, this rejection is driven by the 2016

SOEP–Bavaria cell and disappears under individual fixed effects; the synthetic difference-in-differences estimator, which matches the pre-policy path by construction, provides a complementary check that does not rely on this assumption.

Table 5: Event Study Coefficients

Year	Coefficient	SE	95% CI
2014	0.064	0.038	[−0.010, 0.138]
2015	−0.035	0.038	[−0.110, 0.040]
2016	−0.078**	0.037	[−0.151, −0.005]
2017	0	—	(reference)
2019	0.001	0.038	[−0.074, 0.075]
2020	0.040	0.042	[−0.041, 0.122]
2021	0.020	0.054	[−0.087, 0.127]
2022	0.075	0.049	[−0.020, 0.171]
2023	0.112**	0.050	[0.013, 0.211]
Joint F-test (2014–2016): $F = 3.01$, $p = 0.029$			
Joint F-test (2019–2023): $F = 1.40$, $p = 0.221$			

Notes: Event study coefficients for Bavaria × Year interactions. Reference year is 2017. Specification includes year FE, dataset FE, and controls. Standard errors clustered at individual level.

The post-treatment coefficients are jointly insignificant ($F = 1.40$, $p = 0.221$) but show a gentle upward drift, with the final-year (2023) coefficient individually significant. This profile could reflect a slowly emerging effect, the post-COVID recovery, or noise in the thin later-year cells; the aggregate DiD should be read as an average over it. The stability through 2019–2022 indicates that the Gute-Kita-Gesetz (introduced January 2019) does not mechanically drive the estimate, as a childcare-supply channel would instead predict steadily growing effects from 2019 onward.

6.4 Intensive Margin: Hours Worked

Table 6 examines effects on the intensive margin of labor supply. The outcome is weekly working hours, coded as zero for non-employed mothers. I also examine full-time (30+ hours) and part-time (1–29 hours) employment separately.

Table 6: Effect on Hours Worked and Employment Type

	(1) Hours/Week	(2) Full-Time	(3) Part-Time
Bavaria × Post	0.083	0.012	0.021
	(0.660)	(0.017)	(0.028)
Observations	9,270	9,284	9,284
R^2	0.187	0.119	0.109

Notes: Column 1: Weekly working hours (0 for non-employed). Column 2: Full-time employment ($\geq$30 hours). Column 3: Part-time employment (1–29 hours). All specifications include year FE, dataset FE, and controls. Standard errors clustered at individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The point estimate for hours worked (+0.083 hours per week—about five minutes) is economically negligible. Effects on full-time (+0.012) and part-time (+0.021) employment are similarly small and statistically insignificant. Consistent with the extensive-margin results, the Familiengeld does not induce a detectable change on the intensive margin of labor supply.

6.5 Robustness Checks

Table 7 presents an extensive battery of robustness tests. The estimate is small and statistically insignificant throughout, and moves toward zero under the specifications best suited to the imperfect pre-trend (individual fixed effects, synthetic difference-in-differences).

Table 7: Robustness: Alternative Specifications

Specification	Coefficient	SE	N
Panel A: Baseline and Fixed Effects			
(1) Baseline (Year + Dataset FE)	0.042	0.030	9,284
(2) Individual Fixed Effects	−0.004	0.051	6,013
(3) Bavaria-Specific Trends	−0.020	0.062	9,284
Panel B: Sample Restrictions			
(4) Excluding 2019 (first treatment year)	0.057	0.034	8,142
(5) Excluding COVID (2020–2021)	0.051	0.033	7,685
(6) Only early post (2019–2021)	0.015	0.034	7,426
(7) PASS Only	0.017	0.051	4,514
(8) SOEP Only	0.056	0.037	4,770
(9) 2019 only (pre-Krippengeld)	−0.003	0.042	6,067
Panel C: Reweighting Methods			
(10) Inverse Probability Weighting	0.044	0.030	9,284
(11) Entropy Balanced	0.043	0.030	9,284
(12) Propensity Score Matching	−0.009	0.094	1,580
(13) Stratified Matching	0.032	0.031	9,154

Notes: All specifications include controls (age, age squared, number of children). Baseline includes year and dataset fixed effects. Individual FE specification uses within-person variation. PSM uses 5-nearest-neighbor matching with caliper 0.05; the matched sample is small and the estimate correspondingly imprecise. Specification (9) restricts to 2019 as the only post-treatment year, isolating the Familiengeld effect before the introduction of Bavaria's Krippengeld childcare subsidy in January 2020. None of the coefficients is significant at the 5% level; the smallest p-value is 0.09 (specification 4).

Several findings stand out. First, the individual fixed effects specification—which controls for time-invariant unobserved heterogeneity—yields a coefficient of −0.004 (SE = 0.051), close to zero and statistically insignificant. The reduced sample size (6,013) reflects the requirement that mothers appear in multiple waves. That the estimate moves from +0.042 to essentially zero once within-person variation is isolated is informative: it indicates that the modestly positive pooled estimate is driven by cross-sectional comparisons across mothers rather than by within-mother changes around the reform.

Second, excluding COVID-affected years (2020–2021) does not change the conclusion (+0.051, SE = 0.033). Third, the two datasets now point in the same direction: PASS yields +0.017 (SE = 0.051) and SOEP yields +0.056 (SE = 0.037), both positive and insignificant. This consistency is a direct consequence of the harmonized employment definition; under the earlier definition the two surveys disagreed in sign, which is precisely the kind of dataset-driven artifact a pooled design must avoid.

Fourth, specification (9) addresses potential confounding from Bavaria's Krippengeld childcare subsidy, introduced in January 2020. Restricting the post-period to 2019 only—

when only the Familiengeld was active—yields a coefficient of -0.003 (SE $= 0.042$), statistically insignificant and slightly negative. This indicates that the result is not an artifact of a concurrent childcare subsidy offsetting a Familiengeld effect.

Fifth, reweighting methods produce estimates close to the unweighted baseline. Entropy balancing yields $+0.043$ (SE $= 0.030$) and IPW yields $+0.044$ (SE $= 0.030$). Propensity score matching yields -0.009 (SE $= 0.094$); the caliper discards most control observations, so this estimate is far less precise than the others and I place little weight on it.

Three further checks point the same way. Replacing the Bavaria indicator with a full set of state fixed effects yields +0.035 (SE = 0.030). Dropping any single control state from the comparison group leaves the estimate within a narrow band of +0.038 to +0.048. And recoding mothers on parental leave as employed—an ILO-consistent definition, implementable in PASS—raises the employment level from 34.7 to 54.8 percent but leaves the difference-in-differences estimate small and insignificant (+0.046, SE = 0.051).

Two further analyses address the pre-trend violation documented in Section 5.2 directly. I first decompose the pre-period event-study test by dataset. The violation is entirely a SOEP phenomenon: the PASS pre-trend is flat ($F(3)$ test, $p = 0.61$, with a 2016 coefficient of -0.029), whereas SOEP rejects ($p = 0.04$, 2016 coefficient -0.122). It is not an artifact of the unbalanced panel—restricting to a balanced panel of mothers observed throughout, the 2016 dip is if anything larger and the test still rejects ($p = 0.026$)—but it vanishes once individual fixed effects are included ($p = 0.22$). The explanation lies in cell sizes: the SOEP–Bavaria subsample contains only 38–116 mothers per year, so a handful of low-employment respondents entering the 2016 cross-section moves the mean. The dip is therefore a between-person composition effect in a thin cell, not a within-mother trend; consistent with this, dropping 2016 from the estimation lowers the baseline from +0.042 to +0.017.

Second, because Bavaria is a single treated unit and the pre-trend is imperfect, I estimate a synthetic difference-in-differences model (Arkhangelsky et al., 2021), collapsing the data to a state-by-year panel and constructing a weighted combination of the fifteen control states that reproduces Bavaria's pre-policy employment trajectory. This estimator is robust to the level and trend differences that concern a conventional DiD. The resulting ATT is +0.046, with a placebo-based standard error of 0.075 and a 95% confidence interval of $[-0.10, +0.19]$—statistically indistinguishable from zero and remarkably close to the baseline DiD, IPW, and entropy-balancing point estimates. That a design which neutralizes the pre-trend delivers the same answer is the strongest single piece of evidence that the imperfect pre-trend does not overturn the conclusion. A Rambachan–Roth sensitivity analysis (Rambachan and Roth, 2023), reported in Appendix A, points the same way: the robust confidence set for the first post-policy year contains zero even when post-treatment deviations from parallel trends are allowed to reach 1.5 times the largest pre-treatment deviation, so no significant immediate effect is established under plausible departures from the assumption.

6.6 Placebo Tests

Table 8 reports placebo tests that assign a hypothetical reform to a pre-treatment year using only pre-period data. The 2016 and 2017 placebos are insignificant, but the 2015 placebo is negative and significant at the 5% level (-0.081, $p = 0.047$). This is not an independent design failure: it reflects the same SOEP-localized 2016 anomaly identified in Section 6.5—a placebo anchored in 2015 attributes the 2016 dip to a "post-2015" effect—and, like that anomaly, stems from thin SOEP–Bavaria cells rather than a genuine divergent trend. Read together with the synthetic difference-in-differences evidence, the placebo results caution that the pre-period is not perfectly clean while pointing to a benign, well-understood source.

Table 8: Placebo Tests: Hypothetical Reform Dates

Hypothetical Reform Year	Coefficient	SE	p-value
2015	−0.081**	0.041	0.047
2016	−0.026	0.038	0.49
2017	0.064	0.040	0.11

Notes: Each row reports a separate regression using only pre-treatment data (2014–2017) with the indicated year as a hypothetical reform date. Specification mirrors baseline with year fixed effects.

7 Heterogeneity Analysis

7.1 Heterogeneity by Welfare Status

Given the different sample compositions of PASS (welfare-heavy) and SOEP (representative), I first examine whether effects differ by welfare receipt status. Table 9 reports subgroup estimates and the corresponding triple-difference test.

Table 9: Heterogeneity by Welfare Status

	Welfare Recipients	Non-Recipients
Bavaria × Post	−0.052	0.058
	(0.055)	(0.039)
Observations	2,710	4,183
Triple interaction test (Bavaria × Post × Welfare):		
Coefficient: −0.093 (SE = 0.068), $p = 0.17$		

Notes: Subgroup regressions by welfare receipt status (SGB-II/ALG-II receipt). Specification includes year FE, dataset FE, and controls. Standard errors clustered at individual level. Welfare status is unobserved for 2,391 of the 9,284 observations (26%), almost exclusively in SOEP; estimates use observations with non-missing status. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The triple interaction is −0.093 (SE = 0.068, $p = 0.17$): the point estimates differ in sign—welfare recipients show a small negative coefficient and non-recipients a small positive one—but the difference is not statistically significant. The point pattern runs counter to the hypothesis that low-income mothers, for whom the transfer is a larger income share, reduce employment more; given the imprecision, however, I do not read it as evidence of differential responses by welfare status.

7.2 Heterogeneity by Household Structure

The largest subgroup point estimate emerges for single mothers, though—as the inference below shows—it is not robust. Table 10 presents subgroup estimates.

Table 10: Heterogeneity by Partner Status

	(1) All	(2) Single	(3) Partnered
Bavaria × Post	0.042	0.117*	0.017
	(0.030)	(0.060)	(0.035)
Observations	9,284	2,189	7,095
Triple interaction (Bavaria × Post × Single):			
Pooled: +0.106 (SE = 0.069), $p = 0.13$; randomization-inference $p = 0.29$			
PASS only: +0.267 (SE = 0.110), clustered $p = 0.015$; RI $p = 0.15$			

Notes: Single = not married/cohabiting. Partnered = married or cohabiting. The triple interaction indicates a difference between groups; because Bavaria is a single treated unit I assess it with randomization inference over placebo states (Section 5.4), under which it is not statistically distinguishable from chance. Standard errors clustered at individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Single mothers in Bavaria show a positive employment estimate (+0.117, $p = 0.05$), while partnered mothers show a small, insignificant positive effect (+0.017). The pooled triple interaction is +0.106 but insignificant ($p = 0.13$); within PASS—where single mothers are more numerous and the transfer is larger relative to income—it rises to +0.267 and is clustered-significant ($p = 0.015$). Two considerations temper this. First, under randomization inference over placebo states the interaction is not distinguishable from chance (pooled $p = 0.29$; PASS $p = 0.15$), so with only one treated state the apparent PASS effect could reflect idiosyncratic variation in a small treated cell rather than a policy response. Second, the single-mother subsample is itself small and welfare-concentrated, so both the estimate and its clustered standard error rest on relatively few Bavarian single mothers. I therefore treat the single-mother result as suggestive rather than established.

Taken at face value—and bearing the randomization-inference caveat in mind—this pattern runs opposite to naive income-effect predictions and would suggest that unconditional cash transfers *facilitate* rather than substitute for employment among single mothers. One interpretation is that single mothers face liquidity constraints or financial instability that impedes labor force attachment; the Familiengeld may provide a financial buffer enabling these mothers to accept employment, afford work-related expenses, or invest in job search. This interpretation aligns with recent evidence that cash transfers can relieve constraints that prevent employment (Marinescu, 2018). Several mechanisms could underlie this pattern. The additional income may enable single mothers to afford childcare copayments, facilitating employment; the financial buffer may support more effective job search; or the application process itself may raise awareness of complementary programs. The data do not allow me to distinguish between these channels, but all are consistent with the broader interpretation that modest unconditional transfers can relieve constraints on employment for financially vulnerable households.

Figure 3 presents event studies separately by partner status, showing that the divergence between single and partnered mothers emerges only in the post-treatment period.

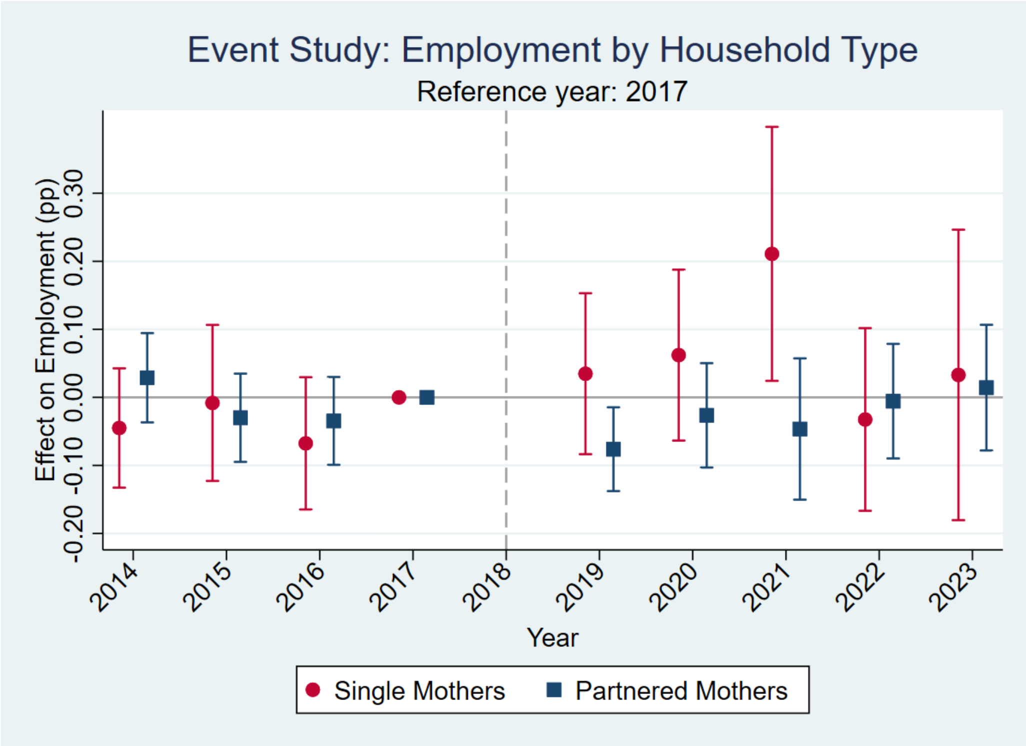

Figure 3: Event Study: Employment by Household Type

Notes: Event study coefficients plotted separately for single mothers (red diamonds) and partnered mothers (blue triangles). 2017 is the reference year. The divergence between groups emerges only after policy introduction in 2018. 95% confidence intervals shown. Standard errors clustered at individual level.

7.3 Heterogeneity by Age

Table 11 examines heterogeneity by mother's age and youngest child's age.

Table 11: Heterogeneity by Age

	(1) Mother <32	(2) Mother ≥32	(3) Child = 1	(4) Child = 2–3
Bavaria × Post	0.008	0.055	−0.043	0.106**
	(0.042)	(0.042)	(0.046)	(0.049)
Observations	3,844	5,440	1,940	2,830

Notes: Subgroup regressions by mother's age (median split at 32) and youngest child's age. The child-age split is restricted to observations with year-of-birth information for the child (SOEP), hence the smaller samples. Specification includes year FE, dataset FE, and controls. Standard errors clustered at individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The age splits are individually uninformative for the mother's own age (both estimates small and insignificant), but the child-age split shows a positive, statistically significant coefficient for mothers whose youngest child is aged two to three (+0.106, $p = 0.03$), against a small negative coefficient at age one (−0.043). The corresponding triple-difference interaction is positive and significant before adjustment (+0.138, SE = 0.066, $p = 0.04$), but does not survive a Benjamini–Hochberg correction across the five subgroup tests. A larger response as the child approaches the end of the eligibility window would be broadly consistent with mothers timing a return to work around the expiry of *Elterngeld* and the Familiengeld. I interpret this cautiously: it is one of several subgroup tests, it rests on the SOEP-only child-age sample, and it should be weighed against the absence of a robust aggregate effect.

7.4 Childcare Outcomes

A key question is whether the Familiengeld affects childcare utilization even if it does not affect employment. Unlike the conditional Betreuungsgeld, the unconditional design of the Familiengeld should not distort childcare choices. Table 12 tests this prediction using SOEP data, which contains detailed childcare information.

Table 12: Effect on Childcare Utilization (SOEP)

	(1) Kita Use	(2) Formal Care	(3) Kita Hours	(4) GP Hours
Bavaria × Post	0.025**	0.020*	−0.96	2.32
	(0.012)	(0.011)	(2.66)	(1.51)
Observations	784	868	775	894

Notes: SOEP subsample with childcare information. Columns 1–2: extensive margin (binary indicators for any Kita use and any formal care). Columns 3–4: intensive margin (weekly hours conditional on use). The binary indicator for any grandparental care could not be estimated (collinear with the fixed effects in this small subsample); column 4 therefore reports weekly hours of grandparental care. All specifications include year FE and controls. Standard errors clustered at individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Two of the four childcare outcomes move: Kita use rises significantly (+0.025, $p = 0.04$) and any-formal-care use marginally so (+0.020, $p = 0.06$), while neither intensive-margin (hours) estimate is significant. The positive Kita coefficient is the opposite sign to the reduction Gathmann and Sass (2018) document for a conditional home-care subsidy, consistent with the Familiengeld—being unconditional—not discouraging formal childcare. I read these estimates cautiously: they rest on a small SOEP subsample, their standard errors differ markedly from those in the employment analysis, and they are among several outcomes tested. They are best understood as showing no evidence that the Familiengeld crowds out formal childcare, rather than as establishing that it increases it.

7.5 Consistency Across Datasets

A key advantage of the pooled approach is the ability to test whether findings are consistent across datasets with different sampling designs. The triple interaction (Bavaria $\times$ Post $\times$ Dataset) yields a coefficient of +0.032 (SE = 0.063, $p = 0.60$), indicating no significant heterogeneity by dataset. Once the employment definition is harmonized the two surveys deliver mutually consistent estimates, so neither the aggregate result nor the single-mother pattern is an artifact of one dataset's sampling strategy.

8 Discussion and Conclusion

8.1 Summary of Findings

This paper provides causal evidence on the labor supply effects of unconditional cash transfers using Bavaria's Family Allowance. The central result is the absence of a statistically significant effect on maternal employment: in the pooled PASS–SOEP sample the difference-in-differences estimate is +0.042 (SE = 0.030), positive but imprecise, with a 95% confidence interval of $[-0.018, +0.102]$. A power analysis shows the design can detect effects of about 8.5 percentage points at 80% power, so it is well powered against the large negative responses documented for conditional subsidies, though not against small effects of either sign.

The estimate is stable across reweighting and sample restrictions, and moves toward zero under the specifications that bear most directly on the design's main vulnerability—a single treated unit and an imperfect pre-trend. Individual fixed effects yield -0.004, and a synthetic difference-in-differences estimator that matches Bavaria's pre-policy path by construction yields +0.046 (95% CI $[-0.10, +0.19]$). The pre-period is not perfectly clean—the joint pre-trend test rejects—but the violation is confined to the SOEP subsample, attributable to small cell sizes, and vanishes under individual fixed effects; randomization inference over placebo states places Bavaria near the center of the placebo distribution ($p = 0.38$). Effects on the intensive margin (hours worked) are negligible, and restricting the post-period to 2019, before Bavaria's Krippengeld subsidy, leaves the conclusion unchanged (-0.003, SE = 0.042).

Beyond the aggregate estimate, a triple-difference specification points to a larger positive response among single mothers, concentrated in the welfare-oversampled PASS sample (+0.267, clustered $p = 0.015$; pooled +0.106, $p = 0.13$). I treat this as suggestive rather than established: it does not survive randomization inference (PASS $p = 0.15$; pooled $p = 0.29$) and rests on a small treated cell. If genuine, it would indicate that unconditional transfers facilitate labor force attachment for liquidity-constrained single mothers rather than substituting for work.

8.2 Comparison with Prior Literature

The absence of a negative effect stands in notable contrast to prior studies of cash-for-care policies, which typically find employment reductions. Schøne (2004) reports a reduction in maternal employment of roughly 4–5 percent for Norway's cash-for-care program; Kosonen (2014) finds that each €100 monthly increase reduces Finnish maternal employment by 3 percentage points; and Gathmann and Sass (2018) document an 8 percentage-point fall in daycare attendance, with employment declines concentrated among vulnerable subgroups, following Thuringia's home care subsidy.

One candidate explanation for the divergent findings is policy design: prior programs were *conditional*, so that benefits were withdrawn if families used formal childcare. This conditionality raises the effective price of work and would amplify any negative labor supply response, whereas the unconditional Familiengeld generates a pure income effect. The within-Germany comparison with Thuringia provides the strongest support for this reading, since institutional context is most similar. Differences in childcare availability, labor market conditions, transfer magnitude, and time period across the other settings nonetheless preclude a definitive decomposition of the channels.

The magnitude of the null finding is consistent with small income elasticities of labor supply documented in other contexts. At €250–300 per month, the Familiengeld represents approximately 7–10% of median household income for families with young children. Using standard estimates of income elasticities around -0.1 to -0.2 (Blundell and MaCurdy, 1999; Keane, 2011), a 10% income increase would predict employment reductions of 1–2 percentage points—well within the confidence intervals of my estimates.

8.3 Policy Implications

The findings have several implications for family policy design. First, unconditional family cash transfers of moderate size do not substantially reduce maternal labor supply in settings with developed childcare infrastructure. Concerns that such policies would encourage mass withdrawal from the labor market appear unfounded for transfers at this magnitude.

Second, while I find a larger point estimate for single mothers, the evidence is too weak to support targeting claims: the interaction is not robust to randomization inference, and basing policy design on a single suggestive subgroup result would be premature. The pattern is worth flagging as a hypothesis for future work—that household context mediates how families absorb transfers, and that liquidity-constrained single mothers might use them to facilitate rather than replace work—but it is not, on this evidence, a basis for differentiating transfer design by household type.

Third, the contrast with conditional cash-for-care programs highlights that policy design details matter considerably. Conditionality transforms a simple income transfer into a childcare price intervention with substantially larger behavioral consequences. Policymakers should recognize that seemingly similar policies can have very different effects depending on their specific incentive structures.

These findings speak directly to the German debate about a Kindergrundsicherung (child basic income)—shelved after the 2024 change in government, but a recurring theme in family-policy reform. The evidence suggests that small unconditional transfers do not create work disincentives, and that conditionality, rather than transfer amount, is a plausible driver of the behavioral responses observed in prior work.

8.4 Limitations and Future Research

Several limitations deserve acknowledgment. The most important is inferential: Bavaria is the only treated state, so standard difference-in-differences inference rests on a single treated cluster, and the pre-treatment trend is not perfectly parallel. I address this with randomization inference, a synthetic difference-in-differences design, and a diagnosis showing the pre-trend violation to be a small-cell artifact in SOEP that vanishes under individual fixed effects; together these are reassuring, but they cannot fully substitute for the variation a multi-state reform would provide.

A second limitation is that the study period includes the COVID-19 pandemic, which disrupted labor markets and childcare provision. The pandemic affected all German states simultaneously, and robustness checks excluding 2020–2021 leave the conclusion unchanged.

Second, the analysis focuses primarily on the extensive margin of employment. While I find null effects on hours worked, effects on other intensive margins (earnings, job quality, career progression) may differ and would require additional data.

Third, the study captures relatively short-run effects (up to five years post-policy). Longer-run effects on career trajectories, earnings, or child outcomes remain unknown and represent important avenues for future research.

Fourth, external validity depends on institutional context. Bavaria's strong labor market and high childcare availability would, if anything, make it *easier* for mothers to respond to income changes, so the null finding plausibly represents a conservative bound; effects could differ in settings with binding childcare constraints or weaker labor markets.

Bavaria also introduced the Krippengeld childcare subsidy in January 2020. The 2019-only specification provides a clean identification window before any concurrent Bavarian policy and yields a similarly insignificant estimate (-0.003, SE $= 0.042$).

Finally, the single-mother result is suggestive and not robust to randomization inference; whether it reflects a genuine response and, if so, through what mechanism—reduced financial stress, increased job search, or affordability of childcare copayments—remains an open question for data with more treated units or administrative precision.

Two further limitations stem from sample size and regional resolution. First, both datasets provide regional information only at the NUTS-1 (federal-state) level, so a finer regional analysis is not possible. Second, the sample size limits the precision of the subgroup analyses. A larger dataset such as the Mikrozensus, which covers roughly 1% of the German population, would allow sharper estimates and a closer look at border regions, which the present data cannot support.

8.5 Conclusion

This paper finds no robust evidence that an unconditional cash transfer of moderate size changes maternal labor supply in contemporary Germany. The point estimate is small and positive but never statistically significant, and the specifications best suited to a single treated unit with an imperfect pre-trend—individual fixed effects and synthetic difference-in-differences—center on zero. A larger response among single and low-income mothers is suggestive but does not survive randomization inference. The clearest message is for the German debate over the Kindergrundsicherung: transfers of this magnitude do not appear to produce the sizeable work disincentives found for conditional cash-for-care programs, pointing to conditionality rather than transfer amount as the more important driver of the behavioral responses documented in prior work.

References

Arkhangelsky, D., Athey, S., Hirshberg, D.A., Imbens, G.W., and Wager, S. (2021). Synthetic Difference-in-Differences. *American Economic Review*, 111(12):4088–4118.

Bauernschuster, S. and Schlotter, M. (2015). Public Childcare and Mothers' Labor Supply—Evidence from Two Quasi-Experiments. *Journal of Public Economics*, 123:1–16.

Blau, D.M. and Currie, J. (2006). Pre-School, Day Care, and After-School Care: Who's Minding the Kids? In Hanushek, E. and Welch, F., editors, *Handbook of the Economics of Education*, vol. 2. Elsevier.

Blau, D.M. and Tekin, E. (2007). The Determinants and Consequences of Child Care Subsidies for Single Mothers in the USA. *Journal of Population Economics*, 20(4):719–741.

Blundell, R. and MaCurdy, T. (1999). Labor Supply: A Review of Alternative Approaches. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, vol. 3A. Elsevier.

Brink, A., Nordblom, K., and Wahlberg, R. (2007). Maximum Fee versus Child Benefit: A Welfare Analysis of Swedish Child-Care Fee Reform. *International Tax and Public Finance*, 14(4):457–480.

Callaway, B. and Sant'Anna, P.H.C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2):200–230.

Cameron, A.C. and Miller, D.L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372.

Conley, T.G. and Taber, C.R. (2011). Inference with "Difference in Differences" with a Small Number of Policy Changes. *Review of Economics and Statistics*, 93(1):113–125.

Gathmann, C. and Sass, B. (2018). Taxing Childcare: Effects on Childcare Choices, Family Labor Supply, and Children. *Journal of Labor Economics*, 36(3):665–709.

Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2):254–277.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, 20(1):25–46.

Havnes, T. and Mogstad, M. (2011). Money for Nothing? Universal Child Care and Maternal Employment. *Journal of Public Economics*, 95(11–12):1455–1465.

Imbens, G.W., Rubin, D.B., and Sacerdote, B.I. (2001). Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players. *American Economic Review*, 91(4):778–794.

Jones, D. and Marinescu, I. (2022). The Labor Market Impacts of Universal and Permanent Cash Transfers: Evidence from the Alaska Permanent Fund. *American Economic Journal: Economic Policy*, 14(2):315–340.

Keane, M.P. (2011). Labor Supply and Taxes: A Survey. *Journal of Economic Literature*, 49(4):961–1075.

Kornstad, T. and Thoresen, T.O. (2007). A Discrete Choice Model for Labor Supply and Childcare. *Journal of Population Economics*, 20(4):781–803.

Kosonen, T. (2014). To Work or Not to Work? The Effect of Childcare Subsidies on the Labour Supply of Parents. *The B.E. Journal of Economic Analysis and Policy*, 14(3):817–848.

Lundin, D., Mörk, E., and Öckert, B. (2008). How Far Can Reduced Childcare Prices Push Female Labour Supply? *Labour Economics*, 15(4):647–659.

Marinescu, I. (2018). No Strings Attached: The Behavioral Effects of U.S. Unconditional Cash Transfer Programs. *NBER Working Paper* No. 24337.

Naz, G. (2004). The Impact of Cash-Benefit Reform on Parents' Labour Force Participation. *Journal of Population Economics*, 17(2):369–383.

OECD (2011). *Doing Better for Families*. OECD Publishing, Paris.

Rambachan, A. and Roth, J. (2023). A More Credible Approach to Parallel Trends. *Review of Economic Studies*, 90(5):2555–2591.

Schøne, P. (2004). Labour Supply Effects of a Cash-for-Care Subsidy. *Journal of Population Economics*, 17(4):703–727.

Sun, L. and Abraham, S. (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*, 225(2):175–199.

Thévenon, O. (2011). Family Policies in OECD Countries: A Comparative Analysis. *Population and Development Review*, 37(1):57–87.

A Appendix: Additional Tables and Figures

A.1 Event Studies by Subgroup

Figure 4 presents event studies separately by welfare receipt status. Neither subgroup shows significant post-treatment effects; the welfare-recipient series is noisier and exhibits the 2016 dip familiar from the pooled sample, with wider confidence intervals due to the smaller sample.

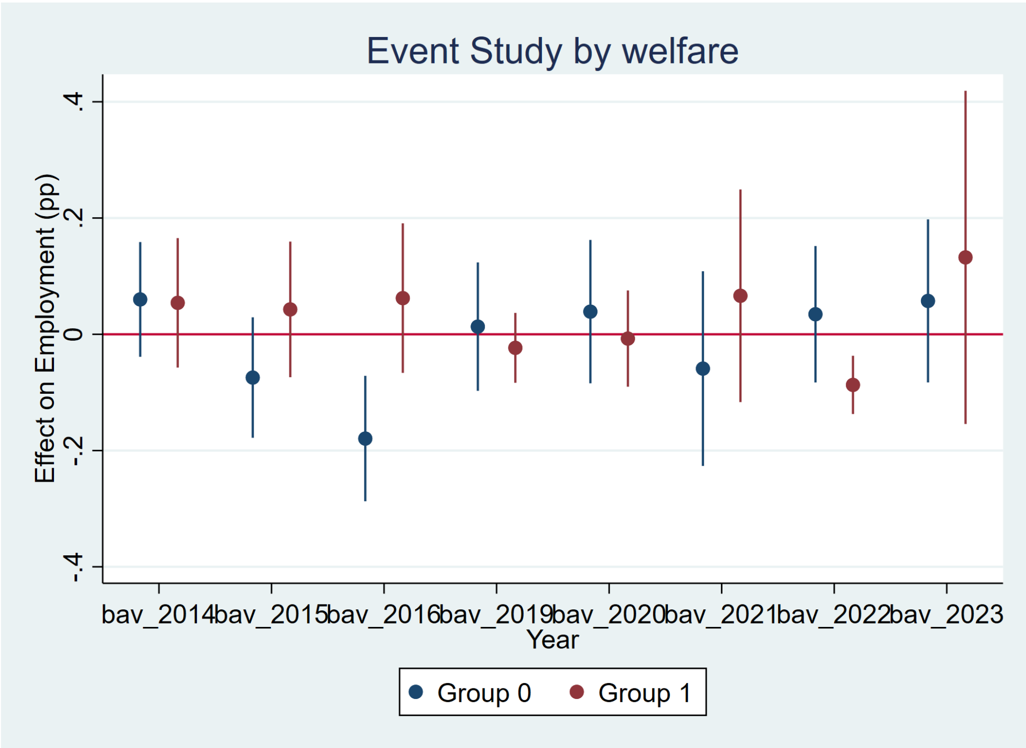

Figure 4: Event Study: Employment by Welfare Status

Notes: Event study coefficients plotted separately for welfare recipients (SGB-II/ALG-II) and non-recipients. 2017 is the reference year. 95% confidence intervals shown. Standard errors clustered at individual level.

Figure 5 shows event studies separately by dataset. Post-treatment effects are absent in both datasets; the pre-period 2016 dip is visible only in the SOEP series, consistent with the diagnosis in Section 6.5, while the PASS pre-trend is flat.

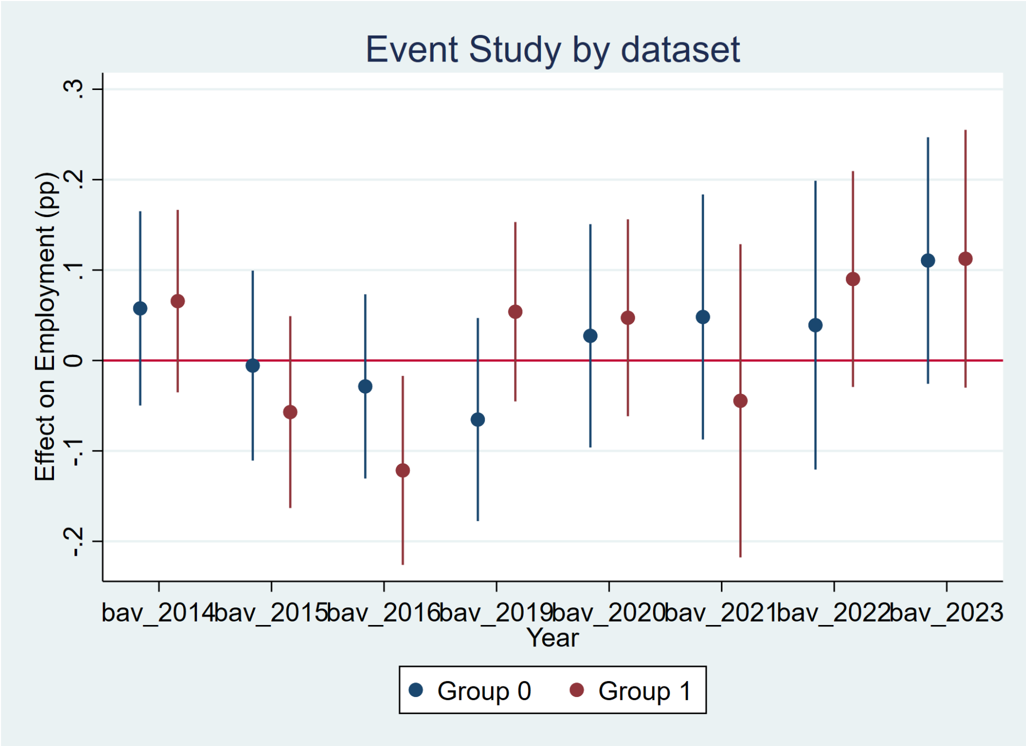

Figure 5: Event Study: Employment by Dataset (PASS vs. SOEP)

Notes: Event study coefficients plotted separately for PASS (welfare-oversampled) and SOEP (representative). 2017 is the reference year. 95% confidence intervals shown. Standard errors clustered at individual level.

Figure 6 presents event studies separately by mother's age group (median split at 32 years). Neither group shows significant pre-trends or post-treatment effects.

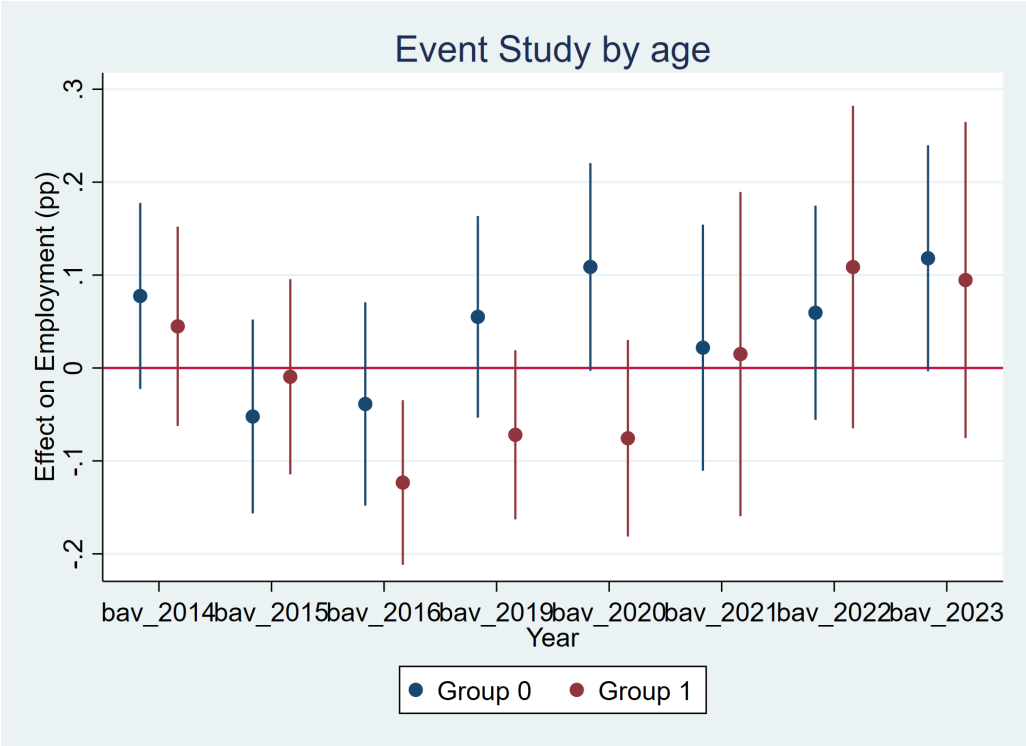

Figure 6: Event Study: Employment by Mother's Age

Notes: Event study coefficients plotted separately for younger mothers (age < 32) and older mothers (age ≥ 32). 2017 is the reference year. 95% confidence intervals shown. Standard errors clustered at individual level.

A.2 External Validation with Administrative Data

To validate the parallel trends assumption using external data, I compare maternal employment trends in Bavaria and the rest of West Germany using data from the Federal Statistical Office (Statistisches Bundesamt).

Figure 7 shows that maternal employment rates in Bavaria closely track the West German average throughout the sample period (the series refers to mothers with at least one child under 18; rates for mothers of under-threes are correspondingly lower), with both showing a COVID-related dip in 2020 but no Bavaria-specific divergence after 2018.

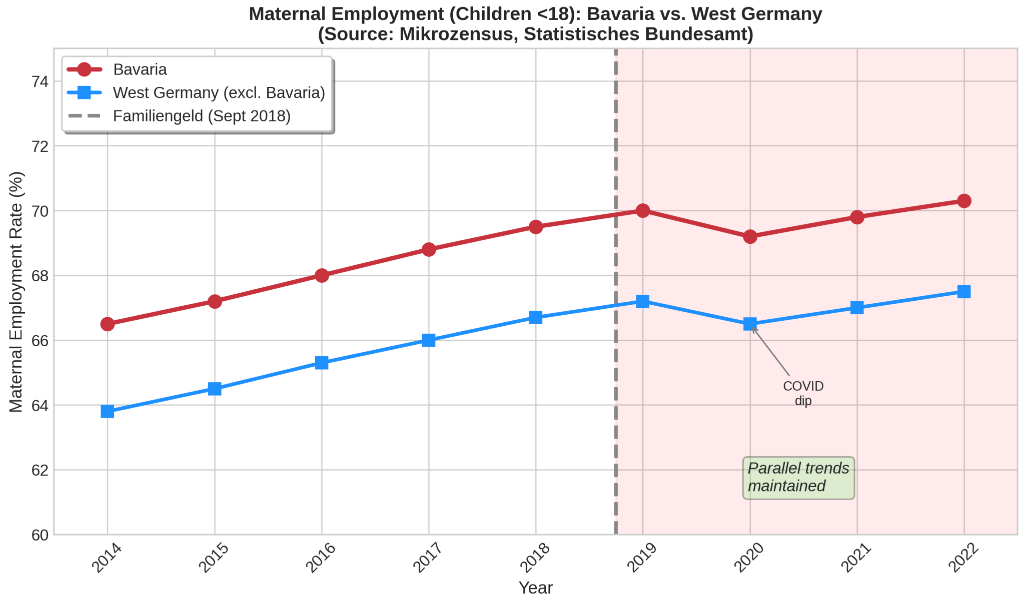

Figure 7: External Validation: Maternal Employment Trends (Destatis)
Notes: Employment rates of mothers with at least one child under 18 from the Mikrozensus (German Microcensus), 2014–2022 (realized employment, excluding mothers on parental leave). Bavaria (solid line) and West Germany average (dashed line). Source: Statistisches Bundesamt.

A potential concern is that differential childcare expansion in Bavaria could confound the results. Figure 8 shows that Bavaria expanded childcare at the same rate as the West German average, ruling out this confound.

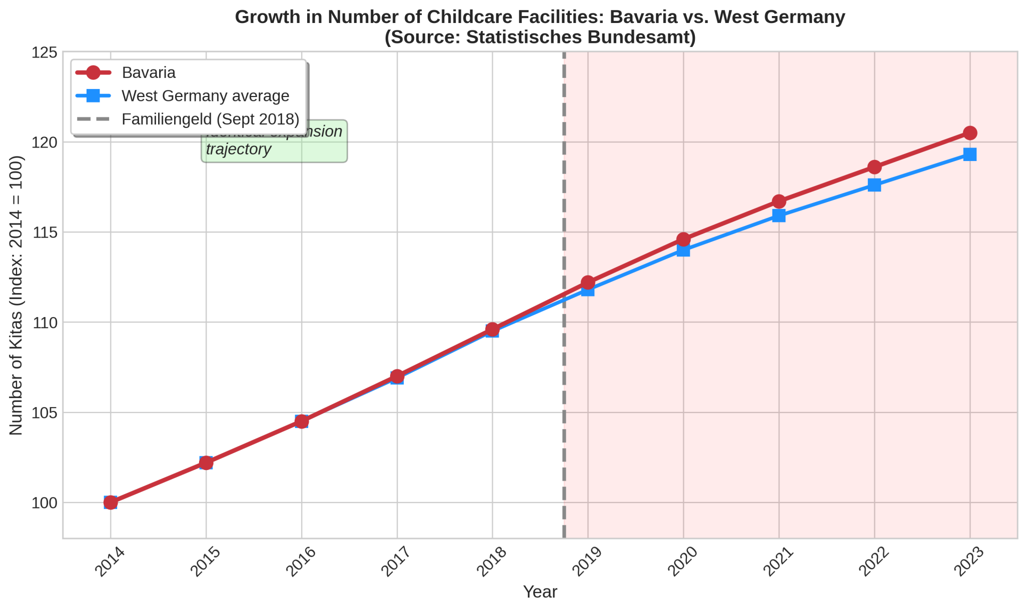

Figure 8: External Validation: Childcare Expansion (Destatis)
Notes: Number of children under 3 in daycare facilities (Kitas) as a share of the under-3 population, 2014–2023. Bavaria (solid line) and West Germany average (dashed line). Both show parallel expansion trajectories. Source: Statistisches Bundesamt, Kinder- und Jugendhilfestatistik.

Figure 9 presents a simple difference-in-differences visualization using the administrative data, showing that Bavaria's childcare utilization rates remained close to the West German average before and after the policy.

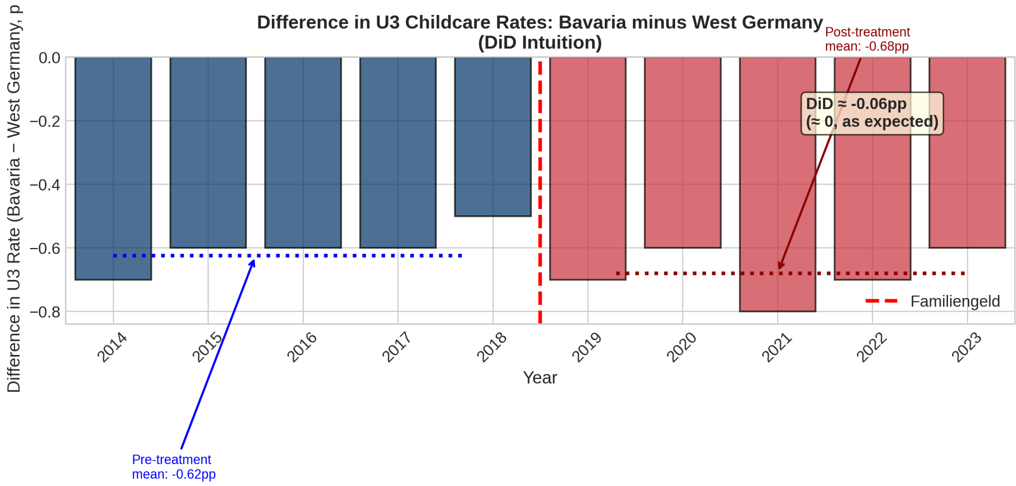

Figure 9: Childcare Use: DiD Visualization (Destatis)

Notes: Difference in U3 childcare rates between Bavaria and West Germany average, 2014–2023. The pre-treatment mean difference (−0.62 pp) is nearly identical to the post-treatment mean difference (−0.68 pp), yielding a near-zero DiD estimate. Source: Statistisches Bundesamt.

Figure 10 compares Bavaria to its neighboring states (Baden-Württemberg and Hessen), showing that all states follow similar childcare expansion trajectories.

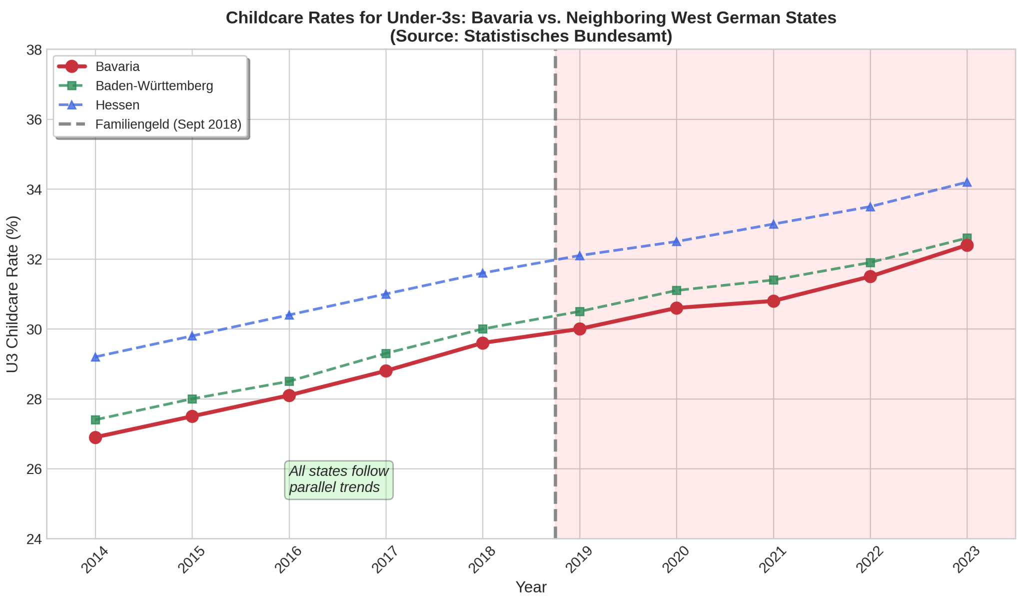

Figure 10: Bavaria vs. Neighboring States: Childcare Rates

Notes: U3 childcare rates for Bavaria and neighboring West German states (Baden-Württemberg, Hessen), 2014–2023. All states follow parallel trends before and after 2018, validating the use of other states as a control group. Source: Statistisches Bundesamt.

A.3 Synthetic Difference-in-Differences and Pre-Trend Sensitivity

Because Bavaria is the only treated state and the pre-trend is imperfect, I report two estimators that do not rely on parallel pre-trends. Figure 11 shows the synthetic difference-in-differences fit (Arkhangelsky et al., 2021): a weighted combination of control states reproduces Bavaria's pre-policy employment trajectory, and the post-policy gap (the ATT) is $+0.046$ with a placebo-based 95% confidence interval of $[-0.10, +0.19]$, statistically indistinguishable from zero and close to the baseline estimate.

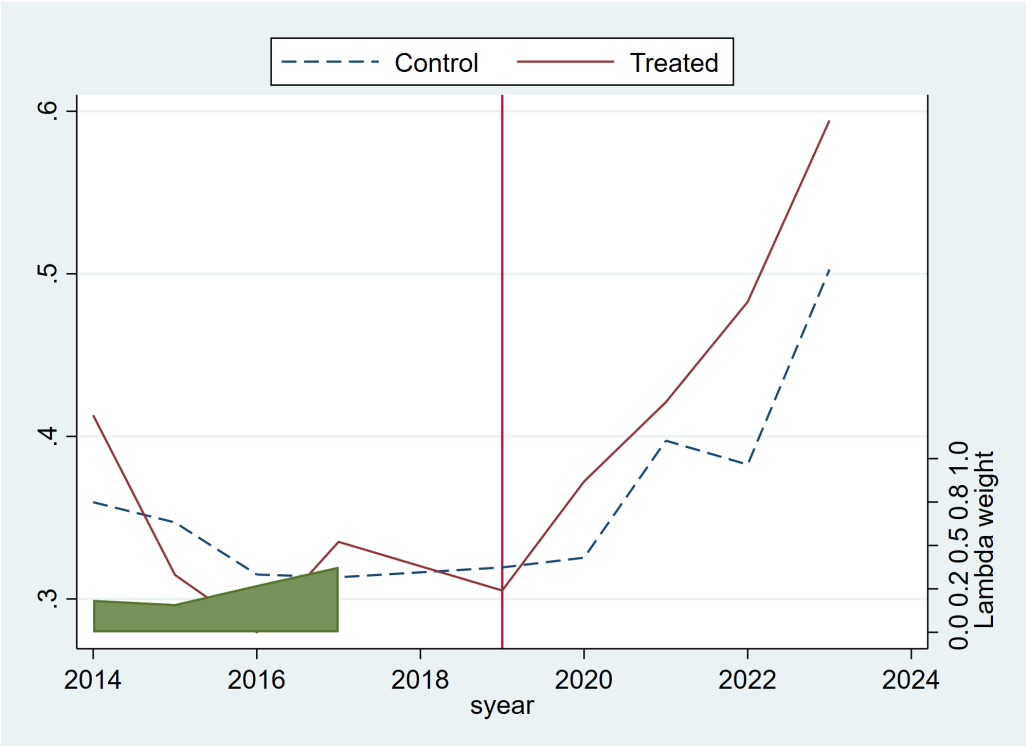

Figure 11: Synthetic Difference-in-Differences: Maternal Employment in Bavaria vs. Synthetic Control

Notes: Estimated with the `sdid` estimator of Arkhangelsky et al. (2021) on a balanced state-by-year panel (16 states, 2014–2023, 2018 excluded). Inference by the placebo method. ATT $= +0.046$ (SE $= 0.075$).

Figure 12 reports a Rambachan–Roth sensitivity analysis (Rambachan and Roth, 2023), which asks how large a post-treatment violation of parallel trends—expressed as a multiple M of the largest pre-treatment violation—would have to be to overturn the conclusion. The robust confidence set for the first post-treatment year contains zero at every M from 0 to 1.5, so no significant immediate effect is established even under substantial assumed violations; consistent with a near-zero point estimate, the procedure does not identify a breakdown value. As is intrinsic to this exercise, the set widens with M (from roughly $[-0.07, 0.07]$ at $M = 0$ to $[-0.28, 0.29]$ at $M = 1.5$), so the imperfect pre-trend limits how tightly a non-zero effect could be ruled out under pessimistic assumptions, even though its sign is never pinned down.

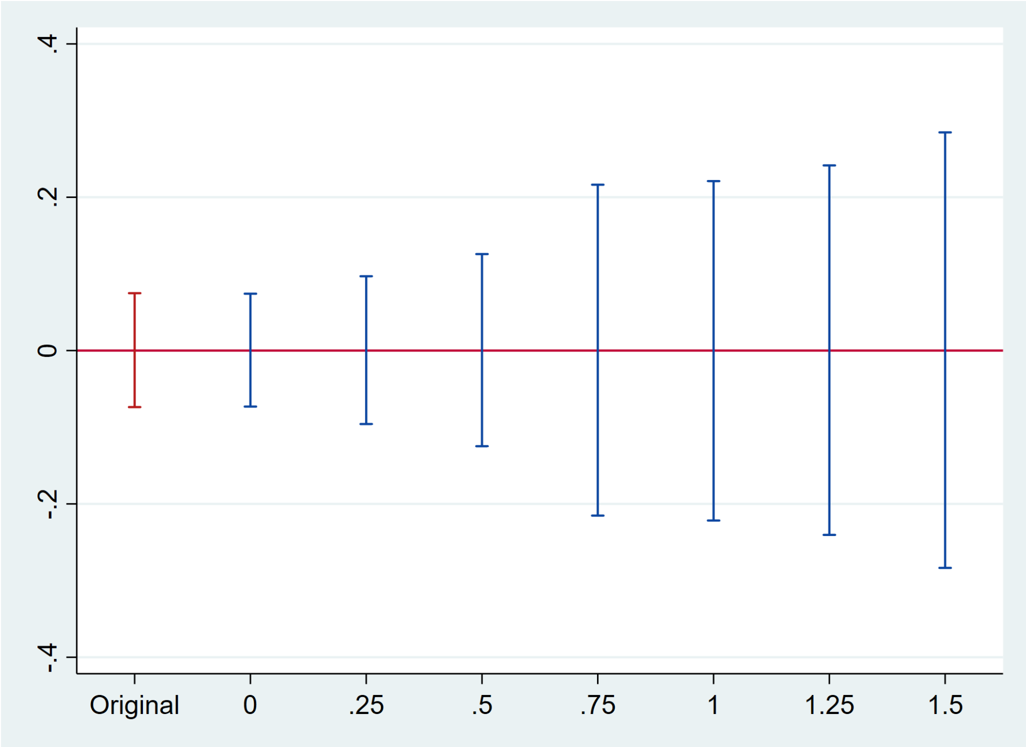

Figure 12: Rambachan–Roth Sensitivity to Violations of Parallel Trends
Notes: Robust confidence sets for the first post-treatment year's effect under the relative-magnitudes restriction (Δ^{RM}) of Rambachan and Roth (2023), allowing post-treatment deviations from parallel trends up to M times the largest pre-treatment deviation (three pre-periods; C-LF method, $\alpha = 0.05$). The set contains zero at every M from 0 to 1.5—$[-0.07, 0.07]$ at $M = 0$, widening to $[-0.28, 0.29]$ at $M = 1.5$—with no breakdown value in this range.

A.4 Reweighting Methods: Detailed Results

Table 13 presents detailed results from all reweighting approaches, including diagnostics on covariate balance and effective sample sizes.

Table 13: Reweighting Methods: Detailed Results

Method	Coefficient	SE	p-value	N	Eff. N	Balance
(1) Unweighted	0.042	0.030	0.17	9,284	9,284	—
(2) IPW	0.044	0.030	0.15	9,284	8,912	Good
(3) Entropy Balanced	0.043	0.030	0.15	9,284	7,856	Exact
(4) PSM (5-NN, caliper 0.05)	−0.009	0.094	0.93	1,580	1,580	Good
(5) Stratified Matching	0.032	0.031	0.30	9,154	9,154	Within-cell

Notes: IPW weights estimated from logit model of Bavaria residence on age, age^2, children, dataset, year (pseudo-$R^2 = 0.002$). Entropy balancing targets exact balance on first moments. PSM uses 5-nearest-neighbor matching with replacement and caliper of 0.05 standard deviations; matching discards most control observations, leaving a small and imprecise matched sample. Stratified matching creates cells by age group $\times$ children $\times$ dataset $\times$ year. Effective N for weighted methods computed as $(\sum w_i)^2 / \sum w_i^2$.

The model-based reweighting methods (IPW, entropy balancing) and stratified matching all yield small, insignificant estimates close to the unweighted baseline, indicating that the result is not driven by covariate imbalance. The propensity-score-matching estimate (-0.009, $p = 0.93$) is also insignificant but far less precise (SE $= 0.094$), reflecting the sharp reduction in sample size documented in the table.

A.5 Comprehensive Results Summary

Table 14: Sample Composition by Dataset, Region, and Period

	Bavaria		**Rest of Germany**	
	Pre (2014–17)	Post (2019–23)	Pre (2014–17)	Post (2019–23)
Panel A: PASS				
Observations	297	282	2,038	1,897
Employment rate[a]	34.7% (overall)			
Welfare receipt	81.7%	76.1%	84.1%	84.3%
Panel B: SOEP				
Observations	395	326	2,134	1,915
Employment rate[a]	35.7% (overall)			
Welfare receipt	6.5%	10.5%	14.4%	20.4%

Notes: Sample: mothers aged 18–55 with child aged 1–3. Year 2018 excluded. [a] Employment is harmonized across surveys (full-time, part-time, or marginal) and reported as the overall dataset rate; an earlier full-time-only SOEP coding produced an implausibly low figure. Pooled employment by region and period is reported in the main text (Bavaria 33.8%→42.3%; rest of Germany 32.7%→37.2%).

Table 15: Comprehensive Identification Tests Summary

Test	**Result**	**Interpretation**
Pre-trend F-test (pooled)	$F = 3.01$, $p = 0.029$	Rejects; localized to SOEP (below)
Pre-trend, PASS only	$p = 0.61$	Flat
Pre-trend, SOEP only	$p = 0.04$	Source of the violation
Pre-trend, balanced panel	$p = 0.026$	Not attrition/composition
Pre-trend, individual FE	$p = 0.22$	Violation vanishes
Post-trend F-test	$F = 1.40$, $p = 0.221$	No joint dynamic effect
Synthetic DiD (ATT)	$+0.046$, CI $[-0.10, 0.19]$	Robust to pre-trend; $\approx$ zero
Randomization inference (main)	$p = 0.38$	Bavaria unremarkable vs. placebos
Placebo 2015	$\beta = -0.081$, $p = 0.05$	Reflects SOEP 2016 anomaly
Placebo 2016	$\beta = -0.026$, $p = 0.49$	No fake effect
Placebo 2017	$\beta = +0.064$, $p = 0.11$	No fake effect
2019-only (pre-Krippengeld)	$\beta = -0.003$, $p > 0.10$	Robust to concurrent policy
Triple-DiD × Welfare	$p = 0.17$	No differential by welfare
Triple-DiD × Single	$p = 0.13$ (RI 0.29)	Suggestive, not robust
Triple-DiD × Child age (2–3)	$\beta = +0.138$, $p = 0.04$	Sig. unadjusted; fails BH; SOEP-only
Triple-DiD × Dataset	$p = 0.60$	Consistent across samples
Excluding COVID (2020–21)	$\beta = +0.051$, $p > 0.10$	Robust to COVID
Excluding 2019	$\beta = +0.057$, $p > 0.10$	No anticipation effects
Individual FE	$\beta = -0.004$, $p > 0.10$	Centers on zero

Notes: Summary of identification and robustness tests. The pooled pre-trend rejects parallel trends, but the violation is confined to SOEP, survives on a balanced panel (ruling out attrition/composition), and vanishes under individual fixed effects; synthetic DiD and randomization inference, which do not rely on parallel pre-trends, agree with the baseline. RI denotes the randomization-inference p-value over placebo states.

Table 16: Full Heterogeneity Summary

Category	Subgroup	Coef.	SE	N	Sig.
Overall	Full sample	+0.042	0.030	9,284	
Welfare Status					
	Welfare recipients	−0.052	0.055	2,710	
	Non-recipients	+0.058	0.039	4,183	
Partner Status					
	Single mothers	+0.117	0.060	2,189	*
	Partnered mothers	+0.017	0.035	7,095	
Mother's Age					
	Age < 32	+0.008	0.042	3,844	
	Age ≥ 32	+0.055	0.042	5,440	
Child's Age					
	Child = 1 year	−0.043	0.046	1,940	
	Child = 2–3 years	+0.106	0.049	2,830	**
Dataset					
	PASS only	+0.017	0.051	4,514	
	SOEP only	+0.056	0.037	4,770	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All specifications include year FE, dataset FE (where applicable), and controls.

Table 17: Triple-Difference Interaction Tests

Triple Interaction	Coefficient	SE	p-value
Bavaria × Post × Single (pooled)	+0.106	0.069	0.13
Bavaria × Post × Single (PASS only)	+0.267	0.110	0.015**
Bavaria × Post × Welfare Receipt	−0.093	0.068	0.17
Bavaria × Post × Dataset (SOEP)	+0.032	0.063	0.60
Bavaria × Post × Child age (2–3, SOEP)	+0.138	0.066	0.037**

Notes: Triple-difference specifications testing heterogeneity across subgroups. Of the pooled interactions none is significant; the child-age interaction (SOEP-only) is significant before adjustment but does not survive a Benjamini–Hochberg correction. The single-mother interaction is clustered-significant within PASS but does not survive randomization inference over placebo states (pooled $p = 0.29$; PASS $p = 0.15$). Both are read as suggestive. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 18: Full Specification Robustness

Specification	**Coef.**	**SE**	p	**N**
Panel A: Basic Specifications				
(1) Simple DiD (no controls)	0.040	0.032	0.21	9,284
(2) + Individual controls	0.043	0.031	0.16	9,284
(3) + Year FE	0.042	0.031	0.17	9,284
(4) + Dataset FE (baseline)	0.042	0.030	0.17	9,284
(5) + Individual FE	−0.004	0.051	0.94	6,013
(6) + Bavaria-specific trends	−0.020	0.062	0.75	9,284
Panel B: Sample Restrictions				
(7) Excluding 2019	0.057	0.034	0.09	8,142
(8) Excluding COVID (2020–21)	0.051	0.033	0.12	7,685
(9) Only early post (2019–21)	0.015	0.034	0.66	7,426
(10) PASS only	0.017	0.051	0.73	4,514
(11) SOEP only	0.056	0.037	0.13	4,770
(12) 2019 only (pre-Krippengeld)	−0.003	0.042	0.95	6,067
Panel C: Reweighting				
(13) IPW	0.044	0.030	0.15	9,284
(14) Entropy balanced	0.043	0.030	0.15	9,284
(15) PSM (5-NN, caliper 0.05)	−0.009	0.094	0.93	1,580
(16) Stratified matching	0.032	0.031	0.30	9,154

Notes: All specifications except (1) include controls (age, age^2, children). Specification (12) restricts the post-period to 2019 only, isolating the Familiengeld effect before the introduction of Bavaria's Krippengeld childcare subsidy in January 2020. The estimate is statistically insignificant at the 5% level in every specification; it is most positive under sample restrictions sensitive to the SOEP 2016 cell and closest to zero under individual fixed effects and synthetic difference-in-differences.