DIW BERLIN

2171

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2026

Combining a Survey Experiment with Lifecycle Models to Evaluate Pronatalist Policies

Joshua Goldstein, Christos Koulovatianos, Jian Li, Carsten Schroeder

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Anton-Wilhelm-Amo-Str. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and Econstor:
http://ideas.repec.org/s/diw/diwwpp.html
https://www.econstor.eu/handle/10419/10

Combining a Survey Experiment with Lifecycle Models to Evaluate Pronatalist Policies*

Joshua Goldstein † Christos Koulovatianos ‡§ Jian Li ¶
Carsten Schroeder ‖**

This draft: July 2, 2026

*We thank the Editor, Petra Todd, and four anonymous referees for their insightful suggestions. We are especially grateful to Hanna Wang for advice on estimating Model 2. We have benefitted from comments and remarks from Jerome Adda, Alexander Bick, Yan Bai, Yongsung Chang, Kaiji Chen, Matthias Doepke, James Heckman, Dun Jia, Scott Kostyshak, Fabian Kindermann, Ilyana Kuziemko, Yulei Luo, Chang Ma, Charles Manski, Costas Meghir, Juan Pantano, Jun Nie, Frederic Vermeulen, Katharina Wrohlich, Giulio Zanella, Donghai Zhang, Vivian Yue, seminar participants at DIW Berlin, LISER, U Luxembourg, U Munich, as well as conference participants at AMES 2017 CUHK, Hong Kong, NASMES 2017 WUSTL, St. Louis, IAER Macroworkshop 2024, Dalian, AMES 2024 ZJU, Hangzhou, and the Econometric Society World Congress, Seoul 2025. Koulovatianos thanks the Luxembourg National Research Fund (grant number FNR CORE POPULISM C23/SC/18098569), and the Research Office of U Luxembourg for financial support (grant number F2R-CRE-PDE-13KOUL). Li acknowledges financial support from the Natural Science Foundation of Zhejiang Province, China (contract LQ20G030006) and the China Scholarship Council. Schroeder acknowledges financial support for the project *Lifetime Inequality Dynamics* (LINDY) from the German Research Foundation & French National Research Agency (contract ANR-19-FRAL-0006-01), *Wealth at the Top* (WATT) from the German Research Foundation (contract 430972113), and SOEP-LEE2 project from dtec.bw—Digitalization and Technology Research Center of the German Bundeswehr. The views expressed and any remaining errors are solely those of the authors.

†UC Berkeley.
‡University of Luxembourg.
§Center for Financial Studies (CFS), Goethe University Frankfurt.
¶Institute for Advanced Economic Research, Dongbei University of Finance and Economics.
‖Department of Economics, Freie Universität Berlin.
**German Institute for Economic Research, German Socio-Economic Panel, Berlin.

Abstract

We develop a three-step procedure for ex-ante policy evaluation. It combines (i) a vignette survey that randomizes out-of-sample policies (child allowances and external childcare subsidies) to infer fertility responses; (ii) two applied-theory lifecycle models estimated on external microdata; and (iii) the inserting of the randomized policies into the models' estimated fertility decision rules to obtain model-based fertility predictions for the vignette respondents, thus cross-validating these predictions with their fertility responses. These steps, applied to the United States and Germany, enable comparable out-of-sample policy evaluation across surveys and models, suggesting how to use vignettes as a compass for model specification.

Keywords: Childcare, fertility, labor supply, vignette, lifecycle models, pronatalist policies

JEL classification: J13, J18, J38, D91, C83, D10

Contents

1 Introduction

Fertility rates in many advanced economies remain well below replacement levels, posing risks to population stability, public pension systems, and economic competitiveness.[1] Pronatalist policies, such as child allowances and childcare subsidies, may need to counteract a strong shift in social norms that discourages childbearing.[2] Such a challenge would require "large" pronatalist policy reforms, calling for a demanding out-of-sample policy evaluation. Here we propose a three-step procedure for addressing this challenge: It combines (i) a vignette survey that randomizes out-of-sample policies (child allowances and external childcare subsidies) to infer fertility responses; (ii) two applied-theory lifecycle models estimated on external microdata; and (iii) for serving the purpose of immediate comparability, the inserting of the randomized policies into the estimated fertility decision rules of the models to obtain model-based fertility predictions for the vignette respondents, thus cross-validating these fertility predictions with their responses.

The vignette survey seeks to elicit how pronatalist policies shape fertility/career decisions by simultaneously asking respondents to state desired fertility and labor supply under alternative out-of-sample pronatalist policies. The vignettes differ along two general dimensions: (a) household economic features: wage, partner income, and partner working hours; and (b) family policies: level of child allowances and/or external childcare-time subsidies. Each feature — randomly assigned to respondents in four rounds — mimics controlled experiments.[3] Two well-established large-scale representative household panel studies, "UAS" for the US, and "Pairfam" in Germany,[4] implemented the vignette module. The integration in existing panels has two advantages. It allows us (a) to control for socio-demographics and (b) to show that "people mean what they say" [Bertrand and Mullainathan, 2001], i.e., whether stated desired fertility and labor in the vignette have strong predictive power for actual working hours and fertility choices in respondents' future lives years later (see

[1] Source: World Bank [2023].

[2] According to Kearney and Levine [2025], "[. . .] the evidence points to a broad reordering of adult priorities with parenthood occupying a diminished role." Fertility effects also seem to work in the opposite direction, leading to cultural and political effects, e.g. influencing populism [see Gavresi et al., 2023]. In addition, abortion and reproduction rights are a major aspect of populist politics [see Rohrmeier, 2025, Patavardhan et al., 2024].

[3] The Appendix provides the complete vignette module. Krueger and Stone [2014] refer to vignettes as a key new advance toward the improvement of well-being evaluations in economic policy. The website of Gary King (http://gking.harvard.edu/vign/eg/) provides a comprehensive overview of vignettes in social sciences and public-health studies. In economics, examples include: (a) Kapteyn et al. [2007] studying work disability in the labor market; (b) Koulovatianos et al. [2005, 2009, 2019] estimating equivalence scales and time costs of children; and (c) Bertoni [2015] linking early life experiences and future well-being evaluations; Alesina et al. [2018] studying intergenerational mobility and preferences for redistribution.

[4] UAS is the Understanding America Study (see https://uasdata.usc.edu/UAS-27) and Pairfam is a satellite of the German Family Panel (see http://www.pairfam.de/en/study/satellite-projects/).

Section 2.2.3).

We develop two structural lifecycle models that differ in complexity. Model 1 is aggregative and parsimonious, with twelve parameters, targeting an assessment of fundamental incentives of pronatalist policies. It jointly matches fertility, labor supply, asset accumulation, adult and child consumption, and childcare choices, simplifying Bick [2016] and Adda et al. [2017]. Model 1's advantage is its ability to uncover mechanisms through analytical characterizations. Model 2, following Adda et al. [2017] and Wang [2026], respects cross-sectional heterogeneities and offers a richer treatment of female labor-market intermittency and fertility. It captures cross-sectional and age-related variation in work experience, the age of the youngest child, marital status, fecundity, and unemployment spells. Both models distinguish between internal (parental) and external childcare time. We estimate both models using independent representative household micro data from the US and Germany.

The vignette sample was not used for model estimation. Hence, it can serve as an independent "test" dataset for model validation. This validation takes each model's fertility-decision rules and parameter estimates to predict fertility for every single vignette respondent – given the randomly assigned four sets of vignette stimuli, and respondents' individual characteristics (e.g., age and gender). Then, for every respondent and set of stimuli, we compare the model-based prediction with the corresponding desired fertility statement from the vignette. This procedure has two advantages: First, it secures "apples-to-apples" comparability by inserting the vignette's policies and household economic characteristics that have exogenous variation into the models. Second, since the comparison in both countries relies on an independent test dataset, the comparison can serve as a navigator for cross-validating model-based predictions.

Our findings are as follows. First, the vignette survey turns out to be suitable for assessing policy responses in both countries: fertility intentions according to the vignette have strong predictive power for actual fertility years after the vignette was implemented.[5] This finding suggests that the module constitutes a timely and cost-effective source of ex-ante information for assessing out-of-sample pronatalist policy effects.

Second, the estimation of Models 1 and 2 in two countries (four applications) provides "deep" and potentially policy-invariant parameters that fit well, most importantly, age profiles of fertility, labor supply and parental childcare time. This finding is not surprising for heterogeneous-agent Model 2, since similar model versions have been successfully fitted to data before [see Adda et al., 2017, Wang, 2026]. Yet, the more parsimonious Model 1, which,

[5]The vignette module was run in Germany in 2013/14 and in the US in 2015/16, giving us enough time to perform this evaluation. Other papers reporting that respondents understand vignette environments and that they provide credible information include McFadden [1998], Manski [1999], and Morgan and Rackin [2010].

in a deliberately reductionist style restricted to twelve parameters, also fits the aggregate multifaceted age profiles well.

Third, the cross-validation shows that vignette-implied and model-implied fertility responses for the test datasets (the vignette samples) agree. In both countries, both policies raise total fertility, while their relative cost-effectiveness is country-specific: for the United States, which policy is more cost-effective depends on the modeling lens, with the vignette estimates lying between the two models' predictions; for Germany, the three approaches are consistent and the daycare subsidy is the more cost-effective policy. A back-of-the-envelope calculation for a \$100 (€100) per-month-per-child policy implies an annual fiscal cost of between 0.07% and 0.24% of GDP across the two models and both countries, well within a feasible budget.

Our study contributes to two strands of literature.

Conceptually, our three-step procedure follows the ideas emphasized by Todd and Wolpin [2023], who argue that combining the forecasting power of structural models with experimental rigor delivers the "best of both worlds." External validation is particularly important given the requirement of structural invariance [Marschak, 1953, Lucas, 1976].[6] By leveraging the comparative strengths of different types of models and data, evaluation designs such as ours lead to better ex-ante assessments of policy effectiveness. They disentangle effective from ineffective policies—their immediate and (unintended) side effects—before they are implemented. This consideration is particularly salient for policies whose effects unfold only over the long term and which, if ineffective, can be reversed only after a long time.

Regarding the practical application, the evaluation of pronatalist policies, studies using applied-theory models include Laroque and Salanié [2014], Bick [2016], Doepke and Kindermann [2019], Adda et al. [2017], and Wang [2026].[7] However, none of these studies externally validate the model implications with independent experimental data (such as a vignette field experiment). Other studies, including Bauernschuster and Schlotter [2015], Bauernschuster et al. [2016], and Gathmann and Sass [2018] for Germany, evaluate actual policy reforms using econometric causal research designs. By design, our study serves a complementary

[6] See also Low et al. [2025] for a quasi-experiment methodology on ex-ante within-sample policy evaluation. A seminal application combining a structural model with a randomized experiment is Attanasio et al. [2012] (schooling choices and PROGRESA).

[7] Other related studies include: (a) Manski and Mayshar [2003] on the interaction between child allowances, social norms, and fertility patterns in Israel; (b) Fort et al. [2020] on the effects of daycare attendance on children's cognitive outcomes; (c) Haan and Wrohlich [2011] on employment-conditional childcare subsidies, female labor supply, and fertility; and (d) Gayle and Miller [2012] on how public expenditures on children and childcare affect fertility and female labor supply over the lifecycle using a dynamic model that incorporates maternal time and financial costs.

objective, i.e., to provide ex-ante out-of-sample policy assessments.

Differences in effect sizes estimated in the present study for the United States and Germany also speak to Manski and Mayshar [2003]. Their central message is that complex fertility patterns may arise from the interaction of economic incentives and social norms, but empirical data often cannot uniquely identify which mechanism is responsible. Following their rationale, the differences in fertility responses that we find across the two countries may be partly attributed to cultural differences.

The remainder of the article is organized as follows. Section 2 describes the data sources used for model estimation, presents the vignette, and discusses its suitability for policy evaluation. Sections 3 and 4 introduce the aggregative Model 1 and the non-aggregative Model 2, respectively, and assess their fit to the data. Section 5 presents vignette-based estimates of intended fertility and labor supply under alternative pronatalist policies. Section 6 uses the vignette responses to conduct both an out-of-sample qualitative validation of the models' predictions and a comparison of policy-effectiveness across the vignette and the two models. Section 7 reports model-implied aggregate elasticities at the structural estimates and two cost-equivalent policy experiments comparing the effectiveness of child allowances and childcare subsidies. Section 8 concludes.

2 Overview of data sources

The following two subsections briefly describe the data used for matching the dynamic models and the vignette data, which are part of two ongoing panel studies in the United States and Germany. Further details regarding our samples and the construction of variable-specific age trajectories appear in Appendix A.[8]

2.1 Data for model estimation

The two models are matched to different but complementary sets of empirical targets. Model 1, which follows couples, is fit to the average age profiles of seven household outcomes—the number of children, labor supply, parental and external childcare time, adult and child consumption, and household assets;[9] the age profiles of the household head's wage and partner income enter the model as exogenous inputs, mirroring the "wage" and "partner

[8]In Appendix A, we construct age profiles using a version of the Carroll and Samwick [1997] methodology. For better treatment of cohort effects, an alternative construction follows Gourinchas and Parker [2002]. We thank an anonymous referee for this remark.

[9]In the German database, household consumption data is not available (see Appendix A). We use the difference between income and savings as a proxy, taking capital gains into account.

income" attributes randomized in the vignette survey. Because Model 2 also includes single women, it is fit to a richer set of moments that summarize means, dispersion, and marital composition. As reported in Table 7, these are the mean log wage; the mean and variance of labor supply, its full- and part-time rates, and the mean labor supply by number of children; mean parental time; the mean number of children, the fraction childless, the birth probability, and the covariance between labor supply and the number of children; and the shares of single women with and without children.

Table 1 summarizes the US and German data sources as well as the sample years used for deriving the above age trajectories. For the United States, we use the Panel Study of Income Dynamics (PSID), the American Time Use Survey (ATUS), and the Survey of Consumer Finances (SCF), while for Germany we use the German Socio-Economic Panel (SOEP), the German Family Panel (Pairfam), and the German subset of the Eurosystem Household Finance and Consumption Survey (HFCS). These surveys underpin the Model 1 estimation (Panel A of Table 1); the estimation of Model 2 (Panel B) for the US additionally draws on the Current Population Survey (CPS) for labor supply, fertility, and marital structure.

Table 1: Summary of data sources

Panel A: Model 1

	Consumption		No.	Labor	External	Parental	Total Assets	Wage	Partner
	Adults	Children	Children	Supply	Childcare	Childcare			Income
US	PSID 1999-2013					Atus 2014	SCF 2007	PSID 1999-2013	
DE	SOEP 1995-2013					Pairfam 2014	SOEP 02/07/12	SOEP 1995-2013	

Panel B: Model 2

	Labor Supply	Fertility	Marital Structure	Parental Childcare	Wages
US	CPS 2000-2025	CPS/PSID	CPS 2000-2025	ATUS 2014	PSID 1999-2013
DE	SOEP 1995-2013			Pairfam 2014	SOEP 1995-2013

Notes: See Appendix A for the samples, the full set of moments, and construction details for both models.

2.2 The vignette survey

2.2.1 Illustration of vignettes

In both theoretical models, households are forward-looking and form expectations about future circumstances, such as marital stability, divorce risk, and the evolution of earnings profiles. Given that the vignette survey cannot capture the full set of information underlying these expectations, it necessarily simplifies the decision-making environment. In the survey, respondents are asked to assume that they "are in the situation of planning a family," and the vignette stimuli—such as wages, partner's earnings, and fertility policies—refer to this context. Thus, the vignette responses can be interpreted as choices in a stylized version of the model's decision problem, conditional on the information provided. Reported fertility and labor supply choices therefore reflect the information in the vignette together with respondents' expectations about the future employment and family biographies of the vignette households. These expectations may depend on the vignette stimuli and on respondents' own experiences.[10]

Here we provide an intuitive description of the vignette survey on fertility/career decisions used for the cross-validation of the model-based predictions. Section 5 and Appendix B provide further details on the randomization and definition of vignette stimuli, etc. Four out of 80 vignettes are randomly assigned to each respective respondent. One such vignette is illustrated in Figure 1. A vignette characterizes a particular environment, defined by three economic characteristics of a hypothetical household and two pronatalist policies. The former are working hours of the partner; earnings of the partner; and hourly wage rate. The latter are the level of subsidization of daycare and child allowances.

We provide four randomized environments to each respondent. In each environment, the respondent is asked to state the desired number of children and working hours. This gives four sets of answers per respondent. Comparing stated desired fertility and labor-supply choices across environments allows us to systematically examine how pronatalist policies alter desired fertility and working hours.

2.2.2 Implementation and sampling

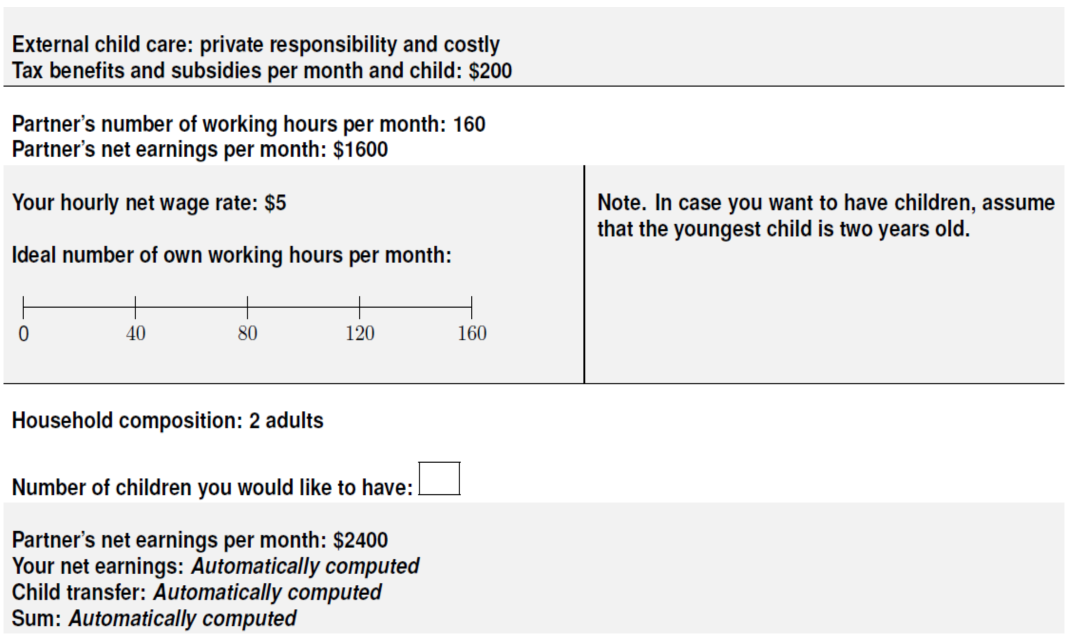
The fertility/career vignettes were embedded in established representative-household surveys: a module of the Understanding America Study[11] and a satellite of the German Family Panel.[12] The target population for the vignette module in both UAS and Pairfam was respondents aged 18 to 50, i.e., the population within an active and fertile age. In UAS, the survey was made available to 2,639 respondents, with 1,872 people completing it, giving

[10] In fact, our survey instrument does not specify whether vignette-based wage variations are permanent or temporary; addressing this distinction is a direction for future work; we thank two anonymous referees for raising this point. Nevertheless, one can gain insight into this issue through forward-looking models, by solving models for fixed respondent characteristics, and by comparing model-implied labor-supply and fertility responses to temporary vs. permanent wage changes. In Sections 6 and 7, we present such model comparisons.

[11] For UAS, see `https://uasdata.usc.edu/UAS-27`.

[12] For Pairfam, see `http://www.pairfam.de/en/study/satellite-projects/`.

Figure 1: Illustration of the vignette module


External child care: private responsibility and costly
Tax benefits and subsidies per month and child: $200

Partner's number of working hours per month: 160
Partner's net earnings per month: $1600

Your hourly net wage rate: $5

Ideal number of own working hours per month:

0 40 80 120 160

Note. In case you want to have children, assume that the youngest child is two years old.

Household composition: 2 adults

Number of children you would like to have:

Partner's net earnings per month: $2400
Your net earnings: *Automatically computed*
Child transfer: *Automatically computed*
Sum: *Automatically computed*

Note: This is the actual visual design from the US vignette.

a response rate of about 71%.[13] In Pairfam, respondents are provided with information on recent activities and results once per year. An invitation letter to participate in the fertility/career vignette was attached to this regular letter for 7,500 participants. This invitation letter included a brief description of the study topic, a link to the survey homepage, and an individual password for login. A total of 593 Pairfam respondents participated in the vignette survey, giving a response rate of about 8%.[14]

2.2.3 Testing if vignette respondents meant what they said

An important condition for the suitability of vignette data for external validation of the model-based fertility responses is that respondents understand the vignettes and provide credible information on intended fertility.[15] For each respondent, we draw on three distinct measures of fertility: (a) the desired number of children stated in the hypothetical vignette scenarios; (b) the desired number of children in the respondent's current situation, from a separate auxiliary question; and (c) the actual number of children recorded in the panel. As a first check, Table 2 shows a high and significant positive correlation between the current-situation desired number and the actual number in both countries.[16] This correlation supports the idea that model-based pronatalist policy evaluation can be combined with policy evaluation by randomized vignette-policy experiments.

The difference between the actual number of children vs. the intended (desired) number of children originates from the fact that databases in both countries contain a large fraction of households headed by younger people. Most likely, for many of these respondents, fertility is incomplete. In particular, in Germany about one-third of the sample consists of individuals in the 18 to 22 age group.

[13] In UAS, the LA County sample and the Spanish speaking population did not receive the questionnaire.

[14] In UAS, our module was fully integrated into an established panel, yielding a response rate of 71%. The lower response rate in Germany's Pairfam likely reflects that the survey invitation was embedded in a standard information letter and required respondents to manually enter the survey link, creating additional barriers. Response rates in online surveys vary widely, from below 1% to 100% [Wu et al., 2022]. While higher response rates reduce sampling variance and improve estimate reliability, they do not guarantee a random or fully generalizable sample. We thank an anonymous referee for raising this concern.

[15] Credibility requires that stated desired fertility in the vignettes positively correlates with actual fertility, as conceptually similar items should be related (see Table 3 for a detailed analysis). See Meade and Craig [2012] for this and other indicators of careless survey responses. For the United States, two additional indicators suggest that respondents took the survey seriously and understood the tasks. First, completion times were reasonable: respondents spent 12 minutes on average; 1,431 completed the survey in 6–15 minutes, 289 in 16–25 minutes, and only 126 in less than 5 minutes. Second, respondents reported substantial interest in the survey: about 30% found it "very interesting," 46% "interesting," and 17% "neither interesting nor uninteresting," while less than 7% found it "uninteresting" or "very uninteresting." Comparable information on completion time and survey interest is unavailable for Germany.

[16] The precise wording of the question is: "In your current situation, how many children would you like to have altogether?" We also use the Spearman correlation for categorical values, and the results stay almost the same.

Table 2: Correlation of stated desired fertility and actual current number of children

	Age	Frequency				Mean	Corr.
		US (A)					
Number of children		0	1	2	3+		
Real world	18-42	49	20	19	12	0.95	**0.5035**
Desired altogether (current situation)	18-42	22	15	35	28	1.69	(p value = 0.000)
		Germany (B)					
Number of children		0	1	2	3+		
Real world	18-42	55	16	20	9	0.82	**0.6239**
Desired altogether (current situation)	18-42	31	15	37	17	1.40	(p value = 0.000)

(A) Cross-validation of UAS and vignette module (2016), US.

(B) Cross-validation of Pairfam (2014) and vignette module (2014), Germany.

As an additional validation exercise, we examine whether vignette responses predict respondents' actual future fertility. Because the vignette was embedded in ongoing panel surveys, we can follow the *same* respondents over nearly a decade, from the vignette (2015/16 in the United States and 2013/14 in Germany) to the most recent panel waves (2025 in the United States and the latest Pairfam wave, 2021/22, in Germany), and test whether their stated fertility intentions predict their realized fertility.

We estimate,

$$\Delta_i = \alpha + \beta n^*_{i,e} + \varepsilon_{i,e}, \tag{1}$$

where Δ_i is the change in respondent i's actual number of children between the year the vignette was implemented and the respondent's latest panel wave (that is, the children realized over the follow-up window). $n^*_{i,e}$ is stated fertility in the environment e, and $\varepsilon_{i,e}$ is an error term, clustered at the individual level.

Column 1 of Table 3, Panel A (United States) and Panel B (Germany), shows that in both countries, $\hat{\beta}$ is positive and highly significant, indicating that vignette responses meaningfully predict future fertility. Including the vignette stimulus variables does not change this result (column 2); further, restricting the sample to respondents aged 42 and below leaves the conclusions unchanged, with the association even somewhat stronger among younger respondents (columns 3 and 4).

These predictive findings support the validity of the vignette data, suggesting that responses reflect real-world fertility intentions. Crucially, because the comparison follows the same respondents over nearly a decade, it reflects genuine individual-level prediction—stated vignette fertility forecasts each respondent's own realized fertility, and if anything more strongly for the young—rather than a cross-sectional pattern across different people.

Table 3: Predictive power of stated desired fertility for future fertility decisions

	Full sample		Sample below age 42	
	Baseline	Extended	Baseline	Extended
Panel A: United States				
n^*	0.094***	0.047*	0.109***	0.073**
	(0.021)	(0.020)	(0.026)	(0.025)
Controls		✓		✓
N	7,269	4,923	5,056	3,446
R^2	0.022	0.080	0.027	0.114
Panel B: Germany				
n^*	0.098***	0.092***	0.154**	0.146**
	(0.028)	(0.027)	(0.049)	(0.045)
Controls		✓		✓
N	2,420	2,420	1,488	1,488
R^2	0.029	0.066	0.055	0.190

Notes: n^* is the desired number of children. N is the number of respondents times number of vignette responses.

The vignette also elicits desired labor supply under different policy environments. Evidence from Beckmannshagen and Schröder [2022] shows that stated desired working hours in survey data have substantial predictive power for subsequent labor market behavior: gaps between desired and actual working hours explain later adjustments in labor supply. Together, these findings support the interpretation that vignette responses capture meaningful information about respondents' fertility and labor supply preferences.

3 Model 1 of career-fertility decisions: aggregative

Albanesi and Sahin [2018] shows that gender convergence in labor force participation is the main factor explaining the decline in the gender unemployment gap in the US. This channel is also supported by evidence from OECD countries. Goldin [1990] and Royalty [1998] provide empirical evidence demonstrating that, historically, females' high incidence of non-participation spells in relation to childbirth has substantially declined. In terms of educational background and age, since 2003, on average, female workers have either caught up or even surpassed male workers. Given this upgraded role of females in the labor market and following Bick [2016] and Jones et al. [2010], in our analysis, the female is assumed to be the sole decision maker.[17]

[17] Model 1 is a simplified unitary-household model, where female decisions coincide with those of males, but

We analyze a part of every household's lifecycle, the period during which adults are building their careers and having children, a time interval denoted by $[0, T]$ representing the age of the head of a household. We split this period into two phases:[18]

> *Phase 1*: The phase before having children, when adults in the household start their working career; the subinterval denoted by $[0, \bar{t}]$, an "infertile" phase.
>
> *Phase 2*: The phase of producing and raising children, while adults in the household also pursue a working career; the subinterval denoted by $[\bar{t}, T]$, a "fertile" phase.

We also assume that, beyond time T the household enters another nonfertile phase that we do not examine. To keep the analysis tractable, we assume that times, $\bar{t}$ and T, distinguishing an infertile phase from a fertile phase, are both exogenous.[19] In addition, we assume that, during the fertile phase, the chosen number of children is a continuous variable. This rests on the existence of approximate aggregation across households of the same cohort regarding all goods, including children.[20] The analytical tractability from these two assumptions lowers the computational cost of estimation and facilitates the matching of target moments.

---

we relax this assumption in Model 2. To explain long-term gender-specific labor market dynamics, Eckstein et al. [2019] set up and estimate a lifecycle model incorporating individual and household decisions about education, employment, marriage/divorce, and fertility. They show that improved contraception control is the most important explanation for the increase in female labor market participation in the United States across the birth cohorts 1935 to 1975, followed by higher female education and changing marriage market conditions.

[18]Model 1 is deterministic, which is a major simplification, together with the assumption of having a non-fertile period in early life. Model 2 in the next section corrects for these oversimplifications.

[19]In our empirical analysis in the US and Germany, time 0 corresponds to age 22 (referring to the age of the household head), time $\bar{t}$ corresponds to age 28, and time T corresponds to age 45. Our assumption, that households headed by persons aged 22 to 27 do not have any children, is motivated by the fact that the average number of children in our samples is small for these households. At age 28 the average number of children is approximately 1, which motivates why age 28 is the cutoff age for entering fertile Phase 2 in life. Similarly, 45 is the cutoff age for entering the second infertile phase since few households give birth after that age.

[20]Despite the fact that the choice of children is a discrete variable, aggregation allows the average household (the representative consumer) to make choices of children from a continuum. See Caselli and Ventura [2000] and Koulovatianos et al. [2019] for definitions of the dynamic representative consumer.

3.1 Phase 1

The momentary utility function of an adult living in a household without children (during the time interval $[0, \bar{t}]$) is,

$$u^{\text{no children}}\left(C_A\left(t\right), \ell\left(t\right)\right) = \frac{\left\{\left[\frac{C_A(t)}{\sqrt{N_1}}\right]^{\theta}\left(1 - \underline{\ell} - \ell\left(t\right)\right)^{1-\theta}\right\}^{1-\gamma}}{1-\gamma} \ , \ \ t \in [0, \bar{t}] \ , \tag{2}$$

in which $\gamma, \theta, \underline{\ell} > 0$, $\ell\left(t\right)$ denotes supplied labor hours in the market by the household head, while $C_A\left(t\right)$ denotes the total consumption of all adults in the household. Parameter N_1 is the exogenous number of adults in the household in Phase 1. So, $C_A\left(t\right)/\sqrt{N_1}$ is the adult-equivalent consumption in the household. The momentary utility function given by (2) also takes into account that modern careers oblige workers to regularly commit a minimum number of hours from their time endowment (e.g., acquiring additional career-related skills at home). This career time-commitment is captured by the reduction in the available time endowment of the household by the constant $\underline{\ell}$, which denotes the continuous flow of committed time.

The budget constraint of a household without children is,

$$\dot{a}\left(t\right) = ra\left(t\right) + y_p\left(t\right) + w\left(t\right)\ell\left(t\right) - C_A\left(t\right) \ , \ \ t \in [0, \bar{t}] \ , \tag{3}$$

in which $a\left(t\right)$ denotes the accumulated household assets, $\dot{a} \equiv da/dt$, $y_p\left(t\right)$ is the partner's labor income and also any household windfall income, and $w\left(t\right)$ is the hourly wage rate.

The optimization problem described by maximizing utility given by (2), subject to the budget constraint given by (3), implies that the decision-maker decides independently on how much to work as an individual. Yet, regarding consumption, we assume that chosen equivalent-adult levels of consumption, $C_A\left(t\right)/\sqrt{N_1}$, are in line with the corresponding optimal consumption choices of her/his partner.[21] Importantly, variations of $y_p\left(t\right)$ are treated as exogenous in Model 1 in order to be in line with the vignette module design: in the vignette survey, partner's income, $y_p\left(t\right)$, is randomized.[22]

[21] Bick [2016] makes a similar modeling choice, using adult-equivalent consumption in the utility function and assigning females full (100%) bargaining power over fertility while preserving within-household economies of scale. Our setup is equivalent: the sole decision maker either has full bargaining power or respects a commonly-agreed sharing rule.

[22] Aligning the model with the desired individual decisions of our vignette respondents is the key reason we do not extend our model to a collective-household setup with household production, as suggested by Chiappori [1997]. Given the tight fitting of our model to the data, our results should not be sensitive to introducing bargaining to our model in future work.

3.2 Phase 2

The momentary utility function of households with children (during the time interval $[\bar{t}, T]$) is,

$$u^{\text{with children}}\left(C_A(t), C_C(t), \ell(t), m(t), m_b(t), n(t)\right) =$$

$$= \frac{\left[F\left(H\left(C\left(\frac{C_A(t)}{\sqrt{N_2}}, \frac{C_C(t)}{\sqrt{n(t)}}\right) \ , \ L_2(t)\right) \ , \ Q\left(m(t), m_b(t), n(t)\right)\right)\right]^{1-\gamma}}{1-\gamma} \ , \ \ t \in [\bar{t}, T] \ . \quad (4)$$

In equation (4), function F is an aggregator of two composite goods, H, a composite home-production good, and Q, a composite good proxying the quantity/quality of produced children in the household. Function F is given by,

$$F(H, Q) = H^{\alpha} Q^{1-\alpha} \ , \ \ \alpha \in (0,1) \ . \quad (5)$$

Function H is a home-production function depending on consumption, C, and leisure, L_2, given by,

$$H(C, L_2) = C^{\theta} L_2^{1-\theta} \ , \ \ \theta \in (0,1) \ . \quad (6)$$

Consumption, C, in equation (6) is a composite good, given by the function,

$$C\left(\frac{C_A(t)}{\sqrt{N_2}}, \frac{C_C(t)}{\sqrt{n(t)}}\right) = \left(\frac{C_A(t)}{\sqrt{N_2}}\right)^{\theta_1} \left(\frac{C_C(t)}{\sqrt{n(t)}}\right)^{1-\theta_1} \ , \ \ \theta_1 \in (0,1) \ , \quad (7)$$

in which $C_A(t)$ and $C_C(t)$ denote the household-level amount of adult and child consumption.[23] Leisure, L_2, in equation (6) is given by,

$$L_2(t) = 1 - \underline{\ell} - \underline{m} - m(t) - \ell(t) \ , \quad (8)$$

in which $m(t)$ denotes time spent by the decision-maker for childcare. Parameter $\underline{m}$ can be a positive number, capturing, e.g., the time commitment of childrearing if it is a positive number, or it can be a negative number capturing, e.g., childcare that grandparents offer for free. In equation (4), Q is a production function of both quality and quantity of children,

[23] Children's consumption is not interpreted as an expenditure on children. Nevertheless, the Cobb-Douglas functional forms in the utility function imply a preference-driven complementarity between child quality, Q, and consumption expenditure of children. We thank an anonymous referee for this remark.

given by,

$$Q(m(t), m_b(t), n(t)) = n(t)^{\psi}\left[m(t) + \zeta m_b(t)^{\eta}\right]^{1-\psi} \ , \ \ \psi \in (0,1) \ , \ \ \zeta, \eta > 0 \ , \tag{9}$$

in which $n(t)$ denotes the number of chosen children and $m_b(t)$ is the external childcare time by babysitters or instructors per child.[24]

The budget constraint of a household with children is,

$$\dot{a}(t) = ra(t) + y_p(t) + w(t)\ell(t) + \left[A - \kappa w(t)^{\phi} - \tilde{p}_b m_b(t)\right] n(t) - C_C(t) - C_A(t) \ , \ \ t \in [\bar{t}, T] \ , \tag{14}$$

in which $\kappa > 0$, $\phi \geq 0$, $A \geq 0$ is the state-funded child allowance per child, while $\tilde{p}_b$ is the price per unit of childcare time in the market (external childcare), after eventually subtracting a subsidy for this external childcare,

$$\tilde{p}_b = (1 - s) p_b \ , \tag{15}$$

in which $s \geq 0$ is the percentage of the price subsidy by the state. The dependence of child costs on wages, $w(t)$, captures the additional age-dependent variations of providing education to children, an additional monetary cost spent in the process of domestic childcare, which

[24] The production function Q given by equation (9) is based on an aggregate production function for children of the form,

$$\bar{Q}(M, M_b, n) = n^{\bar{\psi}}\left(M + Z M_b^{\eta}\right)^{\bar{\zeta}} \ , \tag{10}$$

in which M is time devoted by both parents to all children, while M_b is the total external time for childcare. We assume that the relationship between M and m is given by,

$$M = m N_2^{\bar{\xi}} n^{\bar{\xi}_1} \ , \tag{11}$$

in which the term $N_2^{\bar{\xi}}$ captures economies of scale regarding the parental input in childcare (e.g., times that parents contribute simultaneously to childcare), while the term $n^{\bar{\xi}_1}$ captures economies of scale stemming from the presence of more than one child (e.g., taking care of two children simultaneously, or spillover effects from older siblings giving advice to younger siblings), with $\bar{\xi}, \bar{\xi}_1 > 0$. In addition, we assume that the relationship between M_b and m_b is given by,

$$M_b = m_b n^{\bar{\eta}} \ , \tag{12}$$

in which the term $n^{\bar{\eta}}$ ($\bar{\eta} > 0$) captures economies of scale due to the presence of more than one child (e.g., older siblings playing with younger siblings). Substituting (11) and (12) into (10) and setting $\bar{\xi} = \bar{\eta}\eta$, $\psi = \bar{\psi}\bar{\xi}_1$, $\bar{\zeta} = 1 - \bar{\psi}\bar{\xi}_1$, and $\zeta = Z N_2^{-\bar{\xi}}$, we obtain,

$$\bar{Q}(M, M_b, n) = N_2^{\bar{\xi}\bar{\zeta}} n^{\psi}\left(m + \zeta m_b^{\eta}\right)^{1-\psi} = N_2^{\bar{\xi}\bar{\zeta}} Q(m, m_b, n) \ . \tag{13}$$

Since function (10) enters a utility function, we can omit the constant $N_2^{\bar{\xi}\bar{\zeta}}$ through renormalization of utility weights in Phases 1 and 2.

probably varies with the age of the child. An example of this additional cost would be the tendency of more educated parents to want more educated children, being willing to spend more on education, foreign languages, physical training, and healthy food for children.[25]

3.3 Solving and simulating the model

The optimization problem of the household is,

$$\max_{(C_A(t),C_C(t),\ell(t),m(t),m_b(t),n(t))_{t\geq 0}} \int_0^{\bar{t}} e^{-\rho t} u^{\text{no children}}\left(C_A\left(t\right), \ell\left(t\right)\right) dt$$

$$+e^{-\rho \bar{t}} \int_{\bar{t}}^{T} e^{-\rho(t-\bar{t})} u^{\text{with children}}\left(C_A\left(t\right), C_C\left(t\right), \ell\left(t\right), m\left(t\right), m_b\left(t\right), n\left(t\right)\right) dt$$

subject to constraints (3) and (14). Appendix C explains how we solve the model in a way that facilitates minimum-distance estimation to data.

A final remark is about focusing on modeling only Phases 1 and 2, and ignoring later stages in life. Explicitly modeling decisions beyond time T should not affect our conclusions substantially. Our model has asset/wealth accumulation and the ultimate target is to match accumulated assets at time T. Matching that asset value should be a sufficient proxy of discounted continuation utility at later stages in life. In a similar manner and for similar reasons, Adda et al. [2017] match their lifecycle model for households up to age 55.

3.4 Fitting the model to the data

For women in the United States and Germany, respectively, Figure 2 depicts the empirical age profiles of numbers of children, labor supply, and parental childcare time, together with the Model 1 fit obtained by Method of Simulated Moments (MSM). We have derived the empirical profiles for all women in couples of a certain age, including employed women, registered unemployed women, and inactive women. Labor supply is measured as the fraction of a 24-hour day devoted to work.[26] While labor supply typically declines for women following a birth, this effect is partly offset in the aggregate by women who have not yet had

[25] Cross-sectional differences in parental education are captured by cross-sectional differences in wages. By setting parameter ϕ equal to zero, such parental education effects on child costs and their impact on average age trajectories disappear.

[26] We measure labor supply ℓ as a three-level discrete variable (not employed, part-time, full-time), with country-specific bin values calibrated to each country's labor-market convention: The US uses $\ell \in \{0, 0.18, 0.36\}$ (annual-hours thresholds 0 / 1820); Germany uses $\ell \in \{0, 0.15, 0.30\}$ (SOEP continuous labor-supply thresholds 0.075 / 0.225). In our couples-only sample, the empirical mean ℓ at age 22 is 0.30 (US) and 0.22 (Germany).

children and by women who remain childless, so the average profile for couples-only women does not necessarily display a pronounced decline around the typical age of first birth.

Although Model 1 is parsimonious, in both countries estimated and actual profiles for almost all variables overlap closely, demonstrating its capability of coping with this demanding data-matching exercise. It is more successful in fitting the US number-of-children profile, while in Germany the fit is better for labor supply, consumption, and assets.[27]

Table 4: Model 1 parameter estimates with Jacobian standard errors

	United States		Germany	
Parameter	Estimate	SE	Estimate	SE
Discount factor				
ρ Time preference	0.010	Calibrated	0.020	Calibrated
Consumption preferences				
θ Weight, consumption	0.450	(0.002)	0.421	(0.002)
α Weight, consumption and leisure	0.867	(0.013)	0.874	(0.021)
γ Relative risk aversion	5.943	(1.662)	1.510	(1.072)
Child preferences				
ψ Weight, quantity and quality of children	0.500	(0.071)	0.306	(0.293)
θ_1 Weight, adult consumption to children	0.810	(0.004)	0.847	(0.003)
η Weight, external childcare on child quality	1.050	(0.210)	1.070	(0.619)
Cost of children				
κ Cost of children, linear par.	3.800	(0.009)	1.001	(0.087)
ϕ Cost of children, nonlinear par.	0.402	(0.312)	0.495	(1.130)
ζ Scaling, external childcare quality	0.592	(0.192)	0.932	(0.990)
Minimum levels				
ℓ_{low} Minimum level, labor supply	0.075	Calibrated	0.097	Calibrated
m_{low} Minimum level, parental childcare supply	−0.068	(0.009)	−0.117	(0.012)

Notes: Two-step limited-memory BFGS (L-BFGS) MSM estimates of Model 1. Both countries: couples-only sample, 124 targeted moments (ages 22–45; see Appendix A for the moment-group structure). Step 1 priority subset $\{\ell_{\text{age}}, n_{\text{age}}, m_{\text{age}}, m_b\}$, with ℓ_{age} upweighted by $k_{\ell_{\text{age}}} = 20$ (US), 10 (DE) and m_b upweighted by $k_{m_b} = 10$ (US), 100 (DE). Jacobian-based standard errors at step $= 5\% \cdot |\theta|$, in parentheses. Pinned parameters report "Calibrated". US: 9/10 free parameters satisfy $|t| > 1.96$ ($\text{cond}(J'WJ) = 1.09\times10^{11}$); Germany: 5/10 ($\text{cond}(J'WJ) = 1.21\times10^{11}$). At the estimate, $\hat{m}_b = 0.084$ (US) and 0.072 (DE) versus the calibration target $1/12 \approx 0.0833$.

Table 4 reports the underlying estimated structural parameters of both countries, while Table 5 reports the cross-cell mean of each of the 7 moment groups at the Model 1 esti-

[27] In both countries, labor supply drops between ages 28 and 29 — the model threshold $\overline{a} = 28$ at which parental-time costs begin to bind. At the structural estimate, the decline is about 25% in the United States ($0.37 \rightarrow 0.28$) and 22% in Germany ($0.27 \rightarrow 0.22$). The 5-year-bin aggregation in Figure 2 smooths this transition. The decline is consistent with extensive evidence that having young children reduces female labor supply. Early contributions include Mincer [1962] and Cain [1966]; surveys include Lehrer and Nerlove [1986], Nakamura and Nakamura [1992], and Blundell and MaCurdy [1999]; see Browning [1992] for a critical review. To identify causal effects, instrumental-variable approaches commonly use twin births or children's sex composition as instruments [Rosenzweig and Wolpin, 1980b,a, Angrist and Evans, 1998].

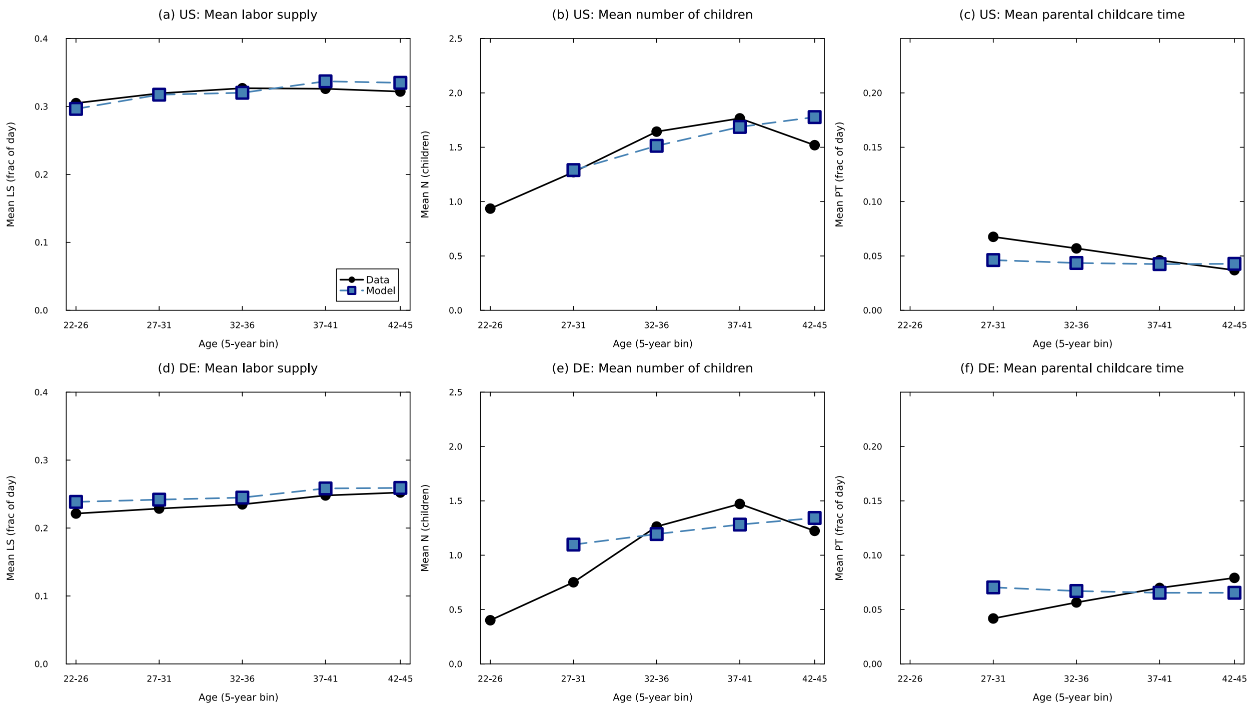


Figure 2: Fit of Model 1

Notes: Data (black solid) vs. Model 1 (steelblue dashed). Top row: United States. Bottom row: Germany. Panels (a,d): mean labor supply as fraction of day. Panels (b,e): mean number of children. Panels (c,f): mean parental childcare time as fraction of day. Each panel plots the 5-year age-bin averages (22–26, 27–31, 32–36, 37–41, 42–45). Sample restricts to couples only.

mate, with rows ordered by overall fit quality (best at top, by average |%err| across the two countries); Appendix D provides the details on the structural estimation.

Table 5: Per-group fit summary: Model 1 vs. data, group-mean levels in both countries

	United States		Germany	
Moment group	data	model	data	model
Mean log adult consumption by age	10.844	10.830	10.217	10.129
Mean log assets by age	11.799	11.605	11.161	11.230
Mean log child consumption by age	9.374	9.468	8.457	8.457
Mean labor supply by age	0.320	0.321	0.236	0.248
Mean number of children by age	1.600	1.592	1.240	1.240
Mean parental childcare time by age	0.051	0.043	0.064	0.067
Mean external childcare time†	0.083	0.084	0.083	0.072

Notes: Data and model columns are the cross-cell mean of the corresponding moment block at the Model 1 estimate. Rows ordered by average $|\%\text{err}|_g = \frac{1}{|g|}\sum_{k\in g}|m_k(\theta^*) - M_k^d|/\max(|M_k^d|, 10^{-3}) \times 100$ across the two countries (best at top). Total: 124 targeted moments across the 7 groups (ages 22–45). †Outsourced childcare time: the data column is the calibration target $1/12 \approx 0.0833$ (same in both countries by construction), and the model column is the post-$\bar{t}$ mean of m_b at $\hat{\theta}$.

4 Model 2 of career-fertility decisions: non-aggregative

In Model 1, children are not a state variable. Model 1 is an aggregative model, in the sense that it pre-aggregates its state variables as inputs to the model and then analyzes all choices through trade-offs of capital-accumulation expansion paths (see, for example, Barro and Sala-i Martin [2004, 2nd edition, pp. 135–137]). Model 1 reflects a purist neoclassical approach, assuming that the number of children is also pre-aggregated (not restricting the numbers of children to be natural numbers). By contrast, Becker and Barro [1988] and Barro and Becker [1989] also let the number of children be continuous, yet still carry it as an irreversible state variable; Model 1's pre-aggregation of children into a flow is thus a further simplification. To overcome these oversimplifications, Model 2 breaks these assumptions and is based on the seminal model by Adda et al. [2017], paying attention to fitting individual decisions at the micro level.[28] The purpose of Model 2 is to examine if and how a more descriptive model at the micro level can match the policy implications of our vignette module on out-of-sample policy interventions. Given that one of the countries we examine is Germany, we follow Wang [2026], abstracting, for simplicity, from asset accumulation.[29]

[28] Moffitt [1984], Francesconi [2002] and Keane and Wolpin [2010] are important earlier works on the joint dynamic modeling of female labor supply and fertility.

[29] In our Model 2, which respects heterogeneity, policies that may have different impacts on families and children with different socio-economic status can be studied. An example of such policies is the universal

The lifespan of a female ranges from age 22 to 62. As in Wang [2026], we assume that only the period between ages 22 and 45 is the fertile age relevant for estimation. We start with a backward-induction solution from age 62, during which the only choice is the labor-supply choice. Moving backward from age 62, once the age of 45 is reached, the fertility choice is added but with different timing from the labor-supply choice within a period. Specifically, the timing of actions when $age_t \in \{22, ..., 45\}$ is split into two stages, 1 and 2: the fertility choice takes place in stage 1, and the labor choice in stage 2. Therefore, within period $t \in \{22, ..., 45\}$ the two decisions are solved backward as well, as a two-stage decision-making equilibrium within period t.

The set of state variables of Model 2 is summarized by,

$$\Omega_t = (x_t, n_t, a_t, \text{mar}_t, y_{p,t}, job_t, \xi_t) \ , \tag{16}$$

where x_t denotes work experience, $n_t \in \{0, 1, 2, 3\}$ is the number of children, a_t is the age of the youngest child in the household (if $n_t > 0$), $\text{mar}_t \in \{0, 1\}$ describes the state $\text{mar}_t = 1$ "married" in t, versus "single" ($\text{mar}_t = 0$), $y_{p,t}$ is the partner's income defined as in Model 1 above, while $job_t \in \{0, 1\}$ is a job-availability shock. If the female receives the shock $job_t = 0$, then in period t there is no job availability, and, in this case, there is no labor-supply option (she must wait until a future period $s > t$ where $job_s = 1$), i.e., this is an involuntary-unemployment spell. Apart from children being a discrete-choice variable, labor choice is a discrete-choice variable as well, with $\ell_t \in \left\{0, \ell^p, \ell^f\right\}$, where ℓ^p and ℓ^f denote part-time and full-time employment, respectively.[30]

As in Adda et al. [2017], in order to capture the impact of labor-supply choices on career progression, the evolution of experience, denoted by x_t, is governed by,

$$x_{t+1} = \left(x_t + \lambda_p \mathbb{I}_t^{\ell^p} + \mathbb{I}_t^{\ell^f}\right) \cdot \left[\mathbb{I}_t^{\ell^p} + \mathbb{I}_t^{\ell^f} + \delta \cdot \left(1 - \mathbb{I}_t^{\ell^p} - \mathbb{I}_t^{\ell^f}\right)\right] \ , \tag{17}$$

where $\mathbb{I}_t^{\ell^p}$ and $\mathbb{I}_t^{\ell^f}$ are indicator functions when ℓ_t equals ℓ^p or ℓ^f. Equation (17) implies that a woman's experience in period $t + 1$ increases by 1 if she works full time in period t, by $\lambda_p \in (0, 1)$ if she works part time, and depreciates by $\delta \in (0, 1)$ if she does not work (i.e., when $\mathbb{I}_t^{\ell^p} = \mathbb{I}_t^{\ell^f} = 0$). The wage of a woman is given by,

$$w_t = e^{\phi_0 + \phi_1 x_t + \phi_2 x_t^2 + \xi_t} \ ,$$

childcare policy in Quebec [e.g., see Ding et al., 2020].

[30] Recall that labor supply is measured as the fraction of a 24-hour day devoted to work. For a full-time worker (8 hours), this corresponds to 0.33.

where $\xi_t \sim N(0, \sigma^2)$, is an i.i.d. productivity shock.

The probability of unemployment, $1 - p_t^j$, governing the shock $job_t \in \{0, 1\}$ depends on age and work experience, x, through the regression-based equation,

$$\ln\left(\frac{p_t^j}{1 - p_t^j}\right) = \pi_1 + \pi_2 \cdot age_t + \pi_3 \cdot age_t^2 + \pi_4 \cdot x_t + \pi_5 \cdot x_t^2 \ , \tag{18}$$

where p_t^j is the fraction of employed females out of the female labor force in year t in the sample. Equation (18) follows Wang [2026, p. 10].[31]

The momentary utility function differs from the corresponding function in Wang [2026] only with respect to the formulation of external childcare/daycare time, which is important in order to evaluate daycare subsidies. Specifically, the external childcare time is a continuous choice variable. The momentary utility function, where $\tilde{c}_t = c_t + n_t (1 - s) p_{b,t} m_{b,t} - n_t \cdot A$, with c_t denoting consumption, $p_{b,t}$ and $m_{b,t}$ denoting the daycare price and the daycare time per child, is,

$$\max_{m_{b,t},\ \text{if } n_t > 0} u_t (\ell_t, \tilde{c}_t, n_t, \text{mar}_t, a_t) =$$

$$\max_{m_{b,t},\ \text{if } n_t > 0} \Bigg\{$$

$$\alpha_1 \left\{ \ln \left[\underbrace{\tilde{c}_t - n_t (1 - s) p_{b,t} m_{b,t} + n_t \cdot A}_{c_t} + \alpha_3 (n_t) \right] - \ln \left(\underbrace{1 + 0.5\, \text{mar}_t + 0.3\, n_t}_{\text{OECD equivalence scale}} \right) \right\}$$

$$+ (1 - O_1 \alpha_6) \left\{ \sum_{k=1}^{3} \alpha_{2k} n_t \mathbb{I}_t^{\{n_t = k\}} + (1 - \text{mar}_t) \left[\alpha_4 \ln \left(\sum_{k=1}^{3} n_t \mathbb{I}_t^{\{n_t = k\}} + 1 \right) + \alpha_5 \mathbb{I}_t^{\{n_t = 1\}} \right] \right\}$$

$$- \left\{ O_3^L \cdot \left\{ \mathbb{I}_t^{\{\ell_t = \ell^f\}} + \gamma_4\, \mathbb{I}_t^{\{\ell_t = \ell^p\}} \right\} + O_3^C \cdot \underbrace{\sum_{j=1}^{5} \zeta_j \mathbb{I}_t^{\{a_t(j) = j\}}}_{\Gamma_1} \left\{ \left[1 - \text{mar}_t \cdot (1 - \phi_b) \right] \cdot \bar{\ell}^c (a_t, n_t) - n_t m_{b,t} \right\} \right\}$$

$$\times \underbrace{\left[\gamma_1 + (1 - O_2 \gamma_5) \left(\sum_{k=1}^{3} \gamma_{2k} n_t \mathbb{I}_t^{\{n_t = k\}} \right) \left(\sum_{j=1}^{5} \gamma_{3j} \mathbb{I}_t^{\{a_t(j) = j\}} \right) \right]}_{\Gamma} \Bigg\} \ , \tag{19}$$

where $\mathbb{I}_t^{\{n_t = k\}}$ is a dummy variable that takes the value 1 if $n_t = k$, $\mathbb{I}_t^{\{a_t(j) = j\}}$ is a dummy

[31] The estimated parameter values of $(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$ for the US and Germany (DE) appear in Table D.1 in Appendix D.

variable that takes the value 1 if $a_t\,(j) = j$, where $a_t\,(j)$ is a branch function given by,

$$a_t\,(j) = \begin{cases} 1 & , \quad \text{if } a_t \in [0,3) \\ 2 & , \quad \text{if } a_t \in [3,6) \\ 3 & , \quad \text{if } a_t \in [6,9) \\ 4 & , \quad \text{if } a_t \in [9,12) \\ 5 & , \quad \text{if } a_t \geq 12 \end{cases} . \tag{20}$$

Term $\alpha_3\,(n)$ denotes a household-level subsistence consumption level that depends on children in the household, which can be calibrated separately based on cross-country estimates in Koulovatianos et al. [2019], and which is adjusted for the OECD equivalence scale. The childcare-time function, denoted by $\bar{\ell}^c\,(a_t, n_t)$, is a discrete-valued (branch) function specified in Appendix D.[32]

Following Wang [2026] on the structure of the utility function, equation (19) depends on three components. The first component focuses on household consumption and its budget-constraint determinants. The second component captures fertility preferences, modulated by $(1 - O_1\alpha_6)$. The third component refers to the disutility of effort (market labor supply and parental childcare), scaled by separate parameters O_3^L (labor) and O_3^C (parental childcare). Parameter Γ, which is affected by family demographics and notably the age of the youngest child, drives the disutility of both types of effort. Parameter Γ_1, affected by the age of the youngest child, drives the preference tradeoff between market time and parental childcare time. Importantly, policy parameter s denotes the subsidy percentage of external childcare costs, and policy parameter A denotes the allowance amount per child.

Focusing on the case where $n_t > 0$, the first-order conditions (FOC) for maximizing with respect to $m_{b,t}$ imply total external childcare time, subject to a non-negativity bound and an outsourcing-cap upper bound. Let $\phi(\text{mar}_t, a_t) \equiv \alpha(a_t)\big[1 - \text{mar}_t(1-\phi_b)\big]$ collapse the marriage adjustment $[1-\text{mar}_t(1-\phi_b)]$ (which scales the focal parent's share of total parental childcare time) and the time-varying outsourcing cap $\alpha(a_t) = \max\{0,\ 0.856 - 0.044\,a_t\}$ (which is decreasing in the age of the youngest child)[33] into a single composite factor. The optimal

[32] Most parameter values in (19) are given by Table 6, with the remaining appear in Appendix D.

[33] Economically, as the youngest child ages, an increasing share of parental childcare time involves activities that are difficult to outsource to formal care – school supervision, homework assistance, emotional and developmental support during adolescence – so the non-outsourceable share $1 - \alpha(a_t)$ grows. Practical formal-care availability also narrows: daycare ends, after-school care has fewer hours and higher per-hour costs. The reduced-form cap $\alpha(a_t)$ captures both effects: at infancy ($a_t = 0$), up to 85.6% of parental childcare time can be outsourced; by $a_t = 12$ the cap falls to $0.856 - 0.044 \cdot 12 = 32.8\%$.

external childcare time then satisfies $0 \leq n_t m_{b,t} \leq \phi(\text{mar}_t, a_t)\, \bar{\ell}^c(a_t, n_t)$, so that

$$n_t m_{b,t} = \begin{cases} \min\left\{\max\left\{\frac{\tilde{c}_t + \alpha_3(n_t) + n_t \cdot A}{(1-s)p_{b,t}} - \frac{\alpha_1}{O_3^C \cdot \Gamma \cdot \Gamma_1}\ ,\ 0\right\}\ ,\ \phi(\text{mar}_t, a_t)\, \bar{\ell}^c\,(a_t, n_t)\right\} & ,\quad \text{if } n_t > 0 \\ 0 & ,\quad \text{if } n_t = 0 \end{cases} . \tag{21}$$

Equation (21) determines total purchased external childcare time. First, external childcare time is related to labor supply. As labor supply (and wages) determine optimal consumption and the demand for external childcare is also a function of consumption (see term $\tilde{c}$ in equation (21)), the two vary with consumption choices. If optimal consumption increases, external childcare time also increases. Second, outsourcing childcare increases with more dislike for any type of effort (driven by parameter Γ) and furthermore by dislike for parental childcare (driven by the childcare-disutility scale O_3^C and parameter Γ_1). Here, equations (21) and (19) imply that the relationship between total purchased external childcare time and the age of the youngest child is an empirical question, determined by the relative sizes of the estimates of parameters ζ_k and γ_{3k}. Third, and importantly, both pronatalist policies (A and s) directly increase purchased external childcare time.

Given a specific constellation of state variables, the individual parental time is,

$$\text{individual parental time} = \left[1 - \text{mar}_t(1-\phi_b)\right] \bar{\ell}^c\,(a_t, n_t) - n_t m_{b,t}\,, \tag{22}$$

where $n_t m_{b,t}$ is given by the closed-form expression in eq. (21).

4.1 Bellman equation and backward induction during the nonfertile period

As in Wang [2026, p. 9, footnote 16] we eliminate fertility choices for women aged 46-62, and we set terminal values at age 63 equal to zero. Nevertheless, unlike her, we keep marital transitions beyond age 45, as they may play a role regarding labor-supply choices. As in Wang [2026], during this age interval, from 46-62, for simplicity we assume that there are no children around in the household. Therefore, the state space described by Ω_t in (16) is reduced to,

$$\Omega_t^0 = (x_t, n_t = 0, a_t = \emptyset, \text{mar}_t, y_{p,t}, job_t, \xi_t)\ \ , \tag{23}$$

where $n_t = 0$ for all ages $age_t \in \{46, ..., 62\}$. The age of the youngest child is irrelevant, being essentially equal to the null set ($a_t = \emptyset$), i.e., one dimension is eliminated. The same functional form for the utility function given by (19), accommodates this setting of Ω_t^0, since

(19) implies,

$$u_t\left(\ell_t, \tilde{c}_t, n_t = 0, \mathrm{mar}_t, a_t = \emptyset\right) = \alpha_1 \ln\left[\underbrace{\tilde{c}_t}_{c_t} + \ \alpha_3\left(n_t = 0\right)\right] - O_3^L\left\{\mathbb{I}_t^{\{\ell_t = \ell^f\}} + \gamma_4\, \mathbb{I}_t^{\{\ell_t = \ell^p\}}\right\}\gamma_1\ . \tag{24}$$

Notice that the functional form of (19) automatically switches off the dimension a_t once we set $n_t = 0$.

The Bellman equation during the age period 46-62 is given by[34]

$$V_t\left(\Omega_t^0\right) = p_t^j \max_{\ell_t}\left\{u_t\left(\ell_t, \tilde{c}_t, n_t = 0, \mathrm{mar}_t, a_t = \emptyset\right) + \beta E\left[V_{t+1}\left(\Omega_{t+1}^0 \mid \ell_t, \Omega_t^0\right)\right]\right\} +$$

$$+\left(1 - p_t^j\right)\left\{u_t\left(\ell_t = 0, \tilde{c}_t, n_t = 0, \mathrm{mar}_t, a_t = \emptyset\right) + \beta E\left[V_{t+1}\left(\Omega_{t+1}^0 \mid \ell_t = 0, \Omega_t^0\right)\right]\right\}\ . \tag{25}$$

In equation (25), the main distinction is about the probability that, in the beginning of period t, the female receives the shock $job_t = 0$. In this case, there is no labor-supply option (she must wait until a future period $s > t$ where $job_s = 1$), i.e., this is an involuntary-unemployment spell. When $job_t = 0$, according (17),

$$x_{t+1} = \delta x_t\ . \tag{26}$$

In case $job_t = 1$, the female has the option to supply labor. If the optimal solution to (25) is $\ell_t^* = 0$, then this is a case of voluntary unemployment, in which case the penalizing outcome of (26) is part of the optimal solution.

Since, unlike Wang [2026], we keep marital transitions beyond age 45, i.e., for all ages $age_t \in \{46, ..., 62\}$, the expectations operator of (25) contains marital transitions, with transition probabilities as period-index logits, according to,

$$\ln\left(\frac{p_t^m}{1 - p_t^m}\right) = \theta_{0m} + \theta_{1m}\cdot t + \theta_{2m}\cdot t^2 + \theta_{3m}\cdot t^3\ , \tag{27}$$

where[35] $p_t^m = \Pr\left(\mathrm{mar}_t = 1 \mid \mathrm{mar}_{t-1} = 0\right)$, $\mathrm{mar}_t = 1$ is a marriage status and $\mathrm{mar}_t = 0$ is a single status, i.e., p_t^m denotes a transition to getting married from a status of being single. Similarly, the expectations operator of (25) contains divorce transitions, with transition

[34] The estimated specification smooths the labor-supply choice with an i.i.d. Type-I extreme-value (Gumbel) preference shock of estimated scale σ_ε; the deterministic $\max_{\ell_t}$ displayed in equation (25) is the $\sigma_\varepsilon \to 0$ limit. See Appendix D.

[35] $t = age_t - 22 + 1$ is the model period index, with $t = 1$ corresponding to age 22 and $t = 24$ to age 45.

probabilities as period-index logits, according to,

$$\ln\left(\frac{p_t^d}{1-p_t^d}\right) = \theta_{0d} + \theta_{1d}\cdot t + \theta_{2d}\cdot t^2 \ , \tag{28}$$

where $p_t^d = \Pr(\text{mar}_t = 0 \mid \text{mar}_{t-1} = 1)$, i.e., p_t^d denotes a transition to getting divorced (single) from a status of being married.

4.2 Bellman equation and backward induction during the fertile period

As in Wang [2026, p. 13], the timing of actions for all ages $age_t \in \{22, ..., 45\}$ is split into two stages, 1 and 2: the fertility choice takes place in stage 1, and the labor choice in stage 2. Therefore, within a period t referring to $age_t \in \{22, ..., 45\}$, the two decisions are solved backward as well, as a two-stage decision-making equilibrium within that period t.

The set of state variables in stage 1 is described by, $\Omega_t = (x_t, n_t, a_t, \text{mar}_t, y_{p,t}, job_t, \xi_t)$. The set of state variables in stage 2, after fertility choices have been made, is Ω_t^*,

$$\Omega_t^* = (x_t, n_t^*, a_t^*, \text{mar}_t, y_{p,t}, job_t, \xi_t) \ \ , \tag{29}$$

where,

$$n_t = n_t^* + f_t \tag{30}$$

and

$$a_t = (1-f_t)\cdot a_t^* \ , \tag{31}$$

with $f_t \in \{0,1\}$ being the fertility decision ($f_t = 1$, means that there will be a newborn). In (31), setting $f_t = 1$ immediately sets the age of the youngest child in the household equal to 0, because of the arrival of the newborn. Therefore, the only difference between the sets Ω_t and Ω_t^*, is that Ω_t is an extension, a superset of Ω_t^*, containing the "free" decision $f_t \in \{0,1\}$, as (16) combined with (30) and (31) imply,

$$\Omega_t = (x_t, n_t^* + f_t, a_t^*\cdot(1-f_t), \text{mar}_t, y_{p,t}, job_t, \xi_t) \ \ . \tag{32}$$

Importantly, after optimization in stage 1 is performed, Ω_t^* will be restricted to containing the optimal fertility choices, $f\ _t^*$.

Therefore, in stage 2, the Bellman equation focusing on the labor-supply choice alone, is similar to that in (25), namely,

$$V_t(\Omega_t) = p_t^j \max_{\ell_t} \left\{ u_t(\ell_t, \tilde{c}_t, n_t, \text{mar}_t, a_t) + \beta E\left[V_{t+1}^*\left(\Omega_{t+1}^* \mid \ell_t, \Omega_t \right)\right] \right\} +$$

$$+ \left(1 - p_t^j\right) \left\{ u_t(\ell_t, \tilde{c}_t, n_t, \text{mar}_t, a_t) + \beta E\left[V_{t+1}^*\left(\Omega_{t+1}^* \mid \ell_t = 0, \Omega_t \right)\right] \right\} \ , \tag{33}$$

whereas in stage 1, the optimization procedure is not through a Bellman equation, but through maximizing the left-hand side of (33), yet after taking into account an exogenous fecundity probability p_t^f (a biological chance of not conceiving, increasing with age, see Appendix D)[36] namely,

$$V^*(\Omega_t^*) = p_t^f \max_{f_t} \left[V_t(\Omega_t \mid f_t, \Omega_t^*) \right] + \left(1 - p_t^f\right) V_t(\Omega_t \mid f_t = 0, \Omega_t^*) \ . \tag{34}$$

In order to execute this step, we need to define the age of the youngest child not into categories but on a grid representing the set of natural numbers $\{0, 1, ..., 13\}$, as explained in Appendix C.[37] Finally, the expression $\tilde{c}_t$ in the utility function given by (19) is a function that describes much of the structure of personal incomes and their relation to policies in Germany, as detailed in Appendix C.

To solve the model, first we used backward induction based on the Bellman equations (25), (33), and (34), and then used the cross-sectional data of the state variables as inputs in order to derive aggregates that match target means. The details of the calibration are explained in Appendix D.

4.3 Heterogeneous populations

Model 2 admits a four-type unobserved-heterogeneity structure through the binary attitudinal switches $O_1, O_2 \in \{0, 1\}$ in (19). Setting $O_1 = 1$ scales child-related utility by $(1 - \alpha_6)$, and setting $O_2 = 1$ scales the disutility of effort (market work and parental child-care) by $(1 - \gamma_5)$. Each (O_1, O_2) combination is solved by its own dynamic program, and simulated moments are a mixture with independent type probabilities $\pi_{O_1} = p(O_1{=}1)$ and $\pi_{O_2} = p(O_2{=}1)$. We estimate (π_{O_1}, π_{O_2}) within narrow, literature-informed bounds.[38]

[36] The estimated specification also smooths the stage-1 fertility decision with an i.i.d. Type-I extreme-value (Gumbel) preference shock of estimated scale σ_ε^f; the deterministic $\max_{f_t}$ displayed in equation (34) is the $\sigma_\varepsilon^f \to 0$ limit. See Appendix D.

[37] Apparently, in all Bellman equations, representing the dynamics of x_t given by equation (17) necessitated the extrapolating of the grid for x_t into a special grid for x_{t+1}. This special grid for x_{t+1} captures the extension of experience when working years are added and the depreciation of experience when years of non-employment are added.

[38] For Germany, the bounds for π_{O_1} and π_{O_2} are taken from Wang [2026] at $\pm 20\%$ around her central values, giving $\pi_{O_1} \in [0.60, 0.90]$ and $\pi_{O_2} \in [0.272, 0.408]$, with point estimates $(\pi_{O_1}^*, \pi_{O_2}^*) = (0.759, 0.344)$.

The four-type mixture serves two purposes: it captures permanent population heterogeneity in fertility and caregiving preferences documented in PSID/NLSY79 (US) and SOEP/Pairfam (DE), and it enables joint identification of variance moments—most notably the age-specific variance of labor supply $\mathrm{Var}(LS)_t$, which a representative-agent formulation mechanically drives to zero at older ages.

4.4 Fitting Model 2 to the data

We estimate Model 2 via priority-weighted MSM for both the United States and Germany. Since Model 2 takes into account both married and single females, the data targets differ from those of Model 1. Table 6 reports the estimated parameters with Jacobian-based standard errors (5% step, $N_{\mathrm{sim}} = 60{,}000$). All 37 parameters satisfy $|t| > 1.96$ in both countries. Notable cross-country differences include the single-mother cost parameter α_4 (-6.83 in the US vs. -7.17 in Germany), reflecting differences in institutional support structures. Figure 3 shows the model fit for the three priority moment groups in both countries, plotted in 5-year age bins from 22 to 45.

Table 7 reports the cross-cell mean of the corresponding moment block at the Model 2 estimate, in both data and the model, for each of the 13 moment groups. Rows are ordered by overall fit quality (best at top, by average |%err| across the two countries).

4.5 Comparison with Model 1

Both models are structurally estimated for the United States and Germany, providing a basis for comparing their fit and policy implications. Model 1 (Section 3) is estimated via Method of Simulated Moments with 10 free parameters matching 124 moments organized into seven groups by age (Appendix A) — labor supply, number of children, parental childcare time, log adult consumption, log child consumption, log assets, and external childcare time (as a calibration target); Figure 2 shows that the model and data lines overlap closely in almost all cases, with the US fit being more successful for the number of children while the German fit is tighter for labor supply, consumption, and assets. Model 2 is estimated via priority-weighted MSM with 37 parameters matching 292 moments organized into 13 groups for both countries: wages (mean log wage), labor supply (mean and variance, full-time rate, part-time rate, mean labor supply by number of children), parental child-caring time, fertility (mean number of children, fraction childless, birth probability, covariance(LS, N)), and marital-status (share single without children, share single with children). Figures 2/3

For the United States, the bounds are informed by attitudinal-question shares in the NLSY79, giving $\pi_{O_1} \in [0.40, 0.70]$ and $\pi_{O_2} \in [0.25, 0.55]$, with point estimates $(\pi^*_{O_1}, \pi^*_{O_2}) = (0.600, 0.495)$.

Table 6: Model 2 parameter estimates with Jacobian standard errors

	United States		Germany	
Parameter	Estimate	SE	Estimate	SE
Discount factor				
β	0.9618	(0.0071)	0.9308	(0.0119)
Consumption and children				
α_1 Consumption slope	13.450	(0.439)	10.083	(0.719)
α_{21} First child	0.770	(0.063)	17.090	(2.037)
α_{22} Second child	10.376	(0.320)	13.140	(0.847)
α_{23} Third child	12.926	(0.422)	14.999	(1.196)
α_4 Single mother cost	−6.834	(0.422)	−7.173	(0.860)
α_5 Single mother shifter	18.128	(0.759)	22.171	(2.238)
α_6 Low-fertility-type pref. scaling	0.714	(0.010)	0.373	(0.044)
Work disutility				
γ_1 Constant	2.548	(0.171)	2.645	(0.411)
γ_{21} First child	1.447	(0.104)	9.982	(0.567)
γ_{22} Second child	0.095	(0.007)	0.641	(0.083)
γ_{23} Third child	0.631	(0.033)	9.927	(1.400)
γ_{31} Youngest child 0–2	7.499	(0.440)	6.103	(0.594)
γ_{32} Youngest child 3–5	0.011	(0.001)	4.227	(0.644)
γ_{33} Youngest child 6–8	19.986	(0.654)	3.696	(0.513)
γ_{34} Youngest child 9–11	0.792	(0.056)	2.121	(0.352)
γ_{35} Youngest child 12–13	7.394	(0.367)	1.878	(0.251)
γ_4 Part-time coef.	0.202	(0.014)	0.319	(0.027)
γ_5 Married-type pref. scaling	0.586	(0.029)	0.334	(0.043)
Parental childcare time cost				
ζ_1 Youngest child 0–2	6.192	(0.386)	1.199	(0.077)
ζ_2 Youngest child 3–5	6.566	(0.499)	1.094	(0.145)
ζ_3 Youngest child 6–8	11.309	(0.784)	6.788	(0.909)
ζ_4 Youngest child 9–11	0.566	(0.042)	10.730	(1.512)
ζ_5 Youngest child 12–13	11.990	(0.479)	10.352	(0.941)
Disutility scales				
O_3^L Labor disutility scale	0.249	(0.014)	0.173	(0.019)
O_3^C Childcare disutility scale	0.502	(0.024)	3.679	(0.389)
Wage process				
ϕ_0 Constant	2.203	(0.003)	2.028	(0.010)
ϕ_1 Experience slope	0.183	(0.001)	−0.034	(0.003)
ϕ_2 Experience slope sq.	−0.011	(0.000)	0.003	(0.000)
Other				
λ_p PT experience accumulation	0.333	(0.012)	0.557	(0.052)
σ_ε Labor pref. shock scale	3.027	(0.117)	2.597	(0.224)
π_1 Job-finding constant	1.206	(0.009)	−0.494	(0.027)
σ_η Fecundity heterogeneity scale	2.916	(0.181)	3.177	(0.156)
Anchor-introduced				
σ_ε^f Fertility pref. shock scale	4.241	(0.217)	1.400	(0.242)
$\overline{lc}$-scale Childcare-time scale	0.771	(0.029)	0.500	(0.017)
π_{O_1} Type-mix prob. (O_1=1)	0.600	(0.008)	0.759	(0.053)
π_{O_2} Type-mix prob. (O_2=1)	0.495	(0.041)	0.344	(0.068)

Notes: Priority-weighted MSM estimates (37 parameters) at $N_{\text{sim}} = 60{,}000$, seed = 123. Jacobian-based standard errors at step $= 5\% \cdot |\theta|$, in parentheses. Both countries: 292 targeted moments (ages 22–45). All 37 parameters satisfy $|t| > 1.96$.

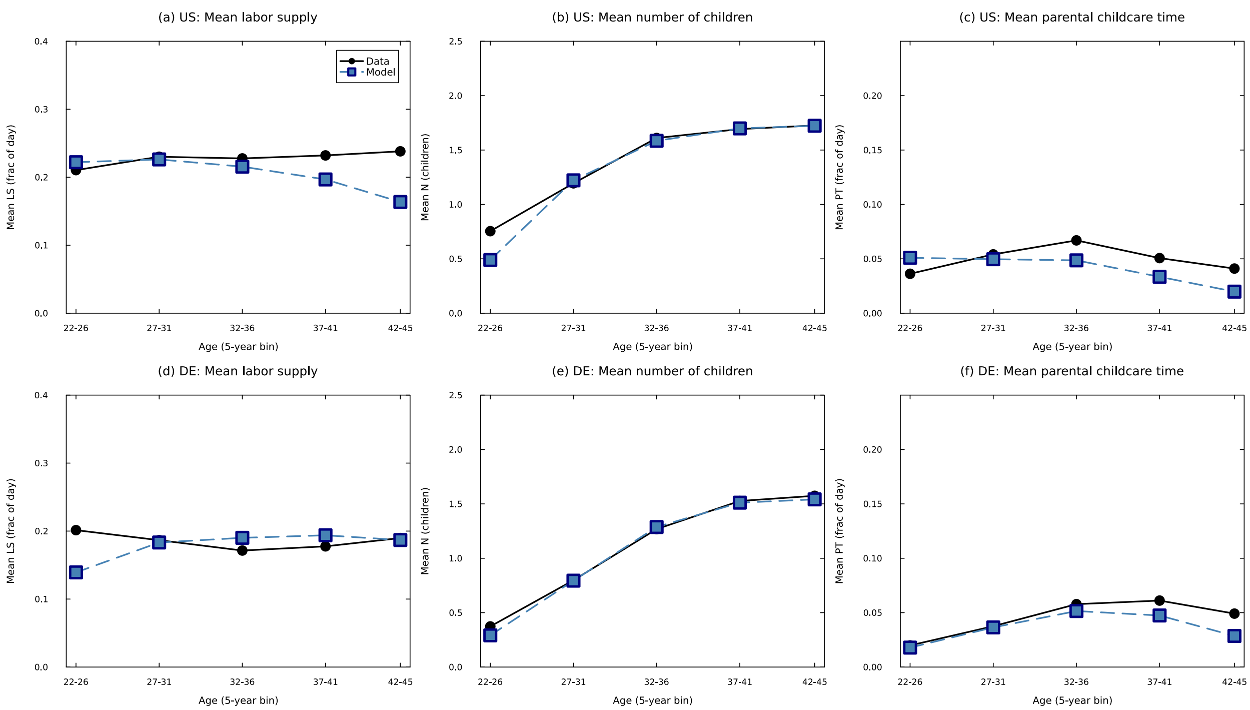


Figure 3: Fit of Model 2

Notes: Data (black solid) vs. Model 2 (steelblue dashed). Top row: United States. Bottom row: Germany. Panels (a,d): mean labor supply as fraction of day. Panels (b,e): mean number of children. Panels (c,f): mean parental childcare time as fraction of day. Each panel plots the 5-year age-bin averages (22–26, 27–31, 32–36, 37–41, 42–45).

Table 7: Model 2 fit of group means

	United States		Germany	
Moment group	data	model	data	model
Mean log wage	2.779	2.778	2.008	2.011
Variance of labor supply	0.026	0.025	0.017	0.016
Mean number of children	1.382	1.328	1.089	1.066
Mean labor supply	0.227	0.207	0.185	0.178
Full time rate	0.561	0.479	0.518	0.478
Fraction childless	0.400	0.483	0.460	0.410
Mean labor supply by number of children	0.220	0.206	0.155	0.120
Mean parental time	0.050	0.041	0.045	0.037
Part time rate	0.140	0.190	0.198	0.232
Birth probability	0.058	0.072	0.055	0.065
Share single without children	0.277	0.104	0.189	0.217
Covariance of labor supply and number of children	−0.036	−0.003	−0.052	−0.001
Share single with children	0.184	0.033	0.061	0.112

Notes: Data and model columns are the cross-cell mean of the corresponding moment block at the Model 2 estimate ($N_{\text{sim}} = 10{,}000$, seed = 123). Rows ordered by average $|\%\text{err}|_g = \frac{1}{|g|}\sum_{k\in g} |m_k(\theta^*) - M_k^d| / \max(|M_k^d|, 10^{-3}) \times 100$ across the two countries (best at top).

and Tables 4/6 present the model fit and estimated parameters with Jacobian-based standard errors for both models and countries.

An important difference between the two models is that Model 1 targets age profiles for couples only, while Model 2 includes both married and single females as distinct states. Model 1 is fundamentally an aggregative framework: the number of children N enters as a continuous (real-valued) population average per age cell, so choices are analyzed at the level of a representative couple. Model 2, in contrast, embeds cross-sectional heterogeneity as a core structural feature: women differ by marital status (married vs. single), by integer parity (number of children currently: 0, 1, 2, 3+), and by unobserved type, with each combination tracked as a distinct decision node. Despite these differences in data targets and quantitative modeling choices, both models produce a good fit to their respective moments.

5 Randomized vignette survey

5.1 Design of the vignette survey

Our vignette survey implemented in UAS (year 2016) and Pairfam (year 2014) asks about fertility/career decisions. The vignette defines environments that are characterized along five dimensions: availability of childcare facilities; level of child allowances; working hours of the partner; earnings of the partner; and individual hourly wage rate. A summary of the attributes of the environments is given in Table 8. From all possible environments, each respondent was randomly assigned four environments and was asked, for each, to state their desired working hours and number of children.

Table 8: Stimuli of environments

Stimulus	Unit	Attributes
Childcare facilities		guaranteed & costless during working hours
		private responsibility and costly
Child allowance	USD per month	200; 400
	€ per month	250; 500
Working hours of partner	per month	0; 80; 160
Net earnings of partner	USD per month	0; 800; 1600; 2400; 4800
	€ per month	0; 800; 1600; 3200
Individual wage rate	USD per hour	5; 10; 20; 30
	€ per hour	5; 10; 30; 50

By varying the attributes along the five dimensions of environments, the vignettes allow us to evaluate the joint decision of fertility and labor-market participation as a function of key background variables, i.e., pronatalist policies.

5.2 Sample compositions

The socio-demographics of the US and German vignette samples are described in Table 9. All numbers are unweighted. In both countries, we have a slight over-representation of female respondents. Consistent with the targeted population of the survey (active population at fertile age), the sample is relatively younger than the country-specific average. The average household size is about three persons in both countries, but there is a higher fraction of married respondents in the United States (about 60% vs. 43% in Germany). The US sample is highly educated: about 78% have a college degree. For Germany, the share is about 46%. The majority in both countries is employed: about 76% in the United States and about 67% in Germany. Native born US citizens make up 93% of the US respondents. In Germany, about 70 percent are resident in the former West Germany, which is close to the official statistics. The incomes of the surveyed populations in both countries encompass a wide range of the income distribution, including households with very low and high incomes.

Table 9: Description of vignette samples

	US		DE	
	mean	sd	mean	sd
Fraction of male	0.404	0.491	0.359	0.480
Age	37.051	8.459	31.403	8.083
Hhsize	3.067	1.401	2.983	1.353
Married	0.597	0.491	0.432	0.496
College	0.783	0.412	0.460	0.499
Employed	0.761	0.427	0.675	0.469
USborn/DEwest	0.928	0.258	0.705	0.456
Income category	Fraction (%)		Fraction (%)	
<10K	8.495		8.094	
10-30K	17.305		26.476	
30-50K	18.039		34.739	
50-100K	33.351		20.742	
>100K	22.811		9.949	
N (respondents)	1,907		593	

Notes: UAS 2015/2016 and Pairfam 2013/2014. Income in USD respectively EUR.

5.3 Results from vignette policy experiments

For the external validation of the model-based policy predictions, it is essential to study how stated desired numbers of children and working hours vary with changes in the two policy variables, child allowances and external childcare facilities.

Table 10 gives a descriptive overview of the results from the vignette, reporting averages of desired numbers of children and working hours for the four possible policy combinations. For the most generous policy constellation (high child allowance; guaranteed and costless childcare during working hours), the desired average number of children is about 2.02 in the United States and 1.99 in Germany; the desired average working hours per month is about 124 in the United States and 100 in Germany. Departing from the most generous policy constellation, reducing the child allowance by 50% lowers the desired number of children to about 1.87 in the US and 1.69 in Germany. Average desired working hours remain the same in the United States and slightly increase in Germany.

Similarly, abolishing guaranteed and costless childcare during working hours also lowers the desired number of children (to about 1.8 in both countries), but in both countries it seems to have no effect for the desired working hours. Simultaneously cutting the allowance and abolishing costless childcare has the strongest effect on desired children, reducing it to about 1.73 in the United States and 1.55 in Germany. The effects on desired working hours are again moderate: In the United States we find a slight decrease compared to the most generous policy constellation, in Germany a slight increase.

Table 10: Desired fertility and working hours under different pronatalist policies

Policy Mix		**High allowance**				**Low allowance**			
	Stat.	US		DE		US		DE	
		$n^\star$	$l^\star$	$n^\star$	$l^\star$	$n^\star$	$l^\star$	$n^\star$	$l^\star$
Free childcare	Mean	2.018	124.4	1.993	99.63	1.873	124.2	1.694	107.3
Costly childcare	Mean	1.788	121.9	1.768	103.9	1.725	117.3	1.554	104.3

Note. UAS 2015/2016 and Pairfam 2013/2014. Income in USD respectively €.

The above assessment of desired fertility and labor supply responses to policy changes are descriptive, neither controlling for individual fixed effects nor for variations in the attributes defining the environments, i.e., individual wage, partner's working time, or earnings. To econometrically control for these aspects, we estimate two fixed-effects regression models.[39] Because desired children is a discrete variable in the survey, we estimate a Poisson Model using the specification,

[39] Fixed effects model is appropriate because each respondent was asked to report desired children and working hours for a series of four randomly assigned environments.

$$n_{i,e}^* = \alpha_n + \beta_n l_e^* + \sum_w \gamma_{w,n} D_{w,e} + \delta_n l_{partner,e} + \rho_n w_{partner,e} + \theta_n T_e + \chi_n D_{c,e} + u_{i,n} + \nu_{i,e}, \quad (35)$$

with the following notation:

1. $n_{i,e}^*$: desired number of children of respondent i in environment e;
2. $l_{i,e}^*$: desired number of working hours of i in environment e;
3. $D_{w,e}$: three dummies for individual wage according to environment e, such that the lowest wage level serves as the benchmark;
4. $l_{partner,e}$: partner's working hours according to environment e;
5. $w_{partner,e}$: partner's wage according to environment e;
6. T_e: child-specific level of allowances according to environment e;
7. $D_{c,e}$: dummy variable defining childcare facilities. It is equal to one if external childcare is in private responsibility and costly in environment e;
8. $u_{i,n}$: individual fixed effect;
9. $\nu_{i,e}$: error term in fertility regression.

For desired working hours, we run a linear model of the form,

$$l_{i,e}^* = \alpha_l + \beta_l n_e^* + \sum_w \gamma_{w,l} D_{w,e} + \delta_l l_{partner,e} + \rho_l w_{partner,e} + \theta_l T_e + \chi_l D_{c,e} + u_{i,l} + \epsilon_{i,e}, \quad (36)$$

with $\epsilon_{i,e}$ denoting the error term.

Table 11 summarizes the estimation results for the desired number of children for both countries. The first three columns relate to the US, providing the estimates for all respondents, male respondents, and female respondents. The subsequent three columns provide the corresponding estimates for Germany. In general, for both countries, we find consistent empirical patterns: Desired working hours and working hours of the partner have a mild negative or no effect, while own and partner wage have a positive effect on the desired number of children.[40] The two policy variables have significant effects: higher child allowances increase the desired number of children,[41] while switching from guaranteed and costless external childcare to a costly one reduces the desired number of children.In Germany, the effects are quantitatively larger than in the US.[42]

Table 12 has the same structure as Table 11 and summarizes the results for desired working hours. Patterns are again rather consistent across the two countries: Desired working hours drop with the number of desired children. Further, compared to the lowest wage level, desired working hours are higher for the three higher wage levels, but there is no further increase over the higher wage levels. In all samples, desired working hours decline with the labor supply of the partner and in the partner's wage rate suggesting a specialization of spouses in labor-market and household activities or a substitution pattern. Pertaining to the two policy variables, an increase of child-related allowances negatively affects desired working hours,[43] but the effect is insignificant in the US. Switching from guaranteed and costless external childcare to costly care reduces desired working hours, but the effect is insignificant for Germany.

Running two separate fixed-effects estimations, one for desired children and one for working hours, does not take into account the simultaneity of the decisions. To cope with this simultaneity, as a robustness check, we re-analyze the data using a dynamic bivariate probit model. The dynamic model supports the general results from the previous estimations.[44]

[40]Empirical studies generally find that fertility increases with household income [Rosenzweig and Schultz, 1985, Hotz and Miller, 1988, Heckman and Walker, 1987, 1990b,a]. Consistent with this finding, Cohen et al. [2013] show that fertility responses to child subsidies in Israel are weakest among high-income families. Regarding methodology, Hotz and Miller [1993] develop an estimation approach for dynamic fertility models, while Francesconi [2002] – who estimates a finite-horizon dynamic model of fertility and labor supply under uncertainty for married women – finds fertility elasticities with respect to women's wages between -1.0 and -1.5 in the United States.

[41]Empirical evidence on the fertility effects of welfare and child transfers is mixed. Hoynes [1997] concludes that the literature provides no compelling evidence of a welfare effect on fertility, while Moffitt [1998] argues that the evidence supports such a relationship. Studying a lump-sum child transfer in Quebec, Milligan [2005] finds a strong fertility response, while Kim [2014], using the same reform, finds little or no permanent effect. Likewise, Parent and Wang [2007] show that long-run fertility responses are substantially smaller than short-run responses.

[42]Table 11 uses wage dummies as control variables (with the lowest wage level as reference value, which is left out). Appendix F, Tables F.2 and F.4 differ slightly, because the underlying estimations do not use dummies but the wage rate as the control variable. This is because we must insert the wage into the model equation to ensure comparability of the model-based simulated results and the vignette results.

[43]Laroque and Salanié [2014] estimate that a €150 monthly increase in child subsidies would reduce female labor supply in France by about 0.5%. Using variation in Israeli child subsidies, Cohen et al. [2013] find that higher subsidies increase fertility, although the effect is weakest among high-income families, for whom the subsidy is less economically relevant.

[44]Details on the dynamic bivariate probit model and estimation results are provided in Appendix E.

Table 11: Desired numbers of children from vignette survey

	US			DE		
	Female	Male	All	Female	Male	All
Desired working hours	-0.0006	-0.0005	-0.0005	-0.0012	-0.0014	-0.0014*
	(0.0004)	(0.0006)	(0.0004)	(0.0009)	(0.0011)	(0.0007)
Hourly wage (10 EUR / 10 USD)	0.0408	-0.0332	0.0113	0.1662*	0.1482	0.1587**
	(0.0464)	(0.0576)	(0.0361)	(0.0859)	(0.1068)	(0.0669)
Hourly wage (20 EUR / 30 USD)	0.2099***	0.1836***	0.2004***	0.2035**	0.2639**	0.2250***
	(0.0533)	(0.0657)	(0.0413)	(0.0943)	(0.1146)	(0.0727)
Hourly wage (30 EUR / 50 USD)	0.3423***	0.3001***	0.3261***	0.4562***	0.5469***	0.4771***
	(0.0757)	(0.0930)	(0.0587)	(0.1367)	(0.2006)	(0.1122)
Partner working hours	0.0005	0.0004	0.0005*	0.0009	-0.0005	0.0003
	(0.0004)	(0.0004)	(0.0003)	(0.0007)	(0.0008)	(0.0005)
Partner wage	0.0065***	0.0068***	0.0066***	0.0119***	0.0123*	0.0116***
	(0.0016)	(0.0018)	(0.0012)	(0.0045)	(0.0063)	(0.0036)
Child allowance (low / high)	0.0003	0.0003	0.0003*	0.0008**	0.0005	0.0006***
	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0004)	(0.0002)
Costly care (0 / 1)	-0.1118***	-0.1319***	-0.1208***	-0.1385***	-0.0836	-0.1223***
	(0.0321)	(0.0378)	(0.0244)	(0.0519)	(0.0710)	(0.0417)
N	3928	2671	6599	1436	788	2224
Pseudo-R^2	0.011	0.013	0.012	0.019	0.015	0.016

Notes: Vignette survey from UAS-27 (US, 2015/16) and Pairfam wave 5 (Germany, 2013/14). Standard errors in parentheses. Estimates from Poisson fixed-effects model. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Desired numbers of working hours from vignette survey

	US			DE		
	Female	Male	All	Female	Male	All
Desired no. of children	-1.7485**	-0.6969	-1.3612**	-3.3402**	-7.1666***	-4.5937***
	(0.7418)	(0.7911)	(0.5463)	(1.3454)	(1.9814)	(1.1086)
Hourly wage (10 EUR / 10 USD)	3.8406**	10.1077***	6.1596***	13.5058***	9.8789***	12.0328***
	(1.7848)	(1.9531)	(1.3277)	(2.8429)	(3.3634)	(2.1731)
Hourly wage (20 EUR / 30 USD)	4.0339*	10.1739***	6.4360***	8.2402**	7.9809**	7.9550***
	(2.1866)	(2.3497)	(1.6154)	(3.2465)	(4.0340)	(2.5238)
Hourly wage (30 EUR / 50 USD)	3.0806	5.6184*	4.0187*	-8.6976*	4.0741	-5.3560
	(2.9845)	(3.2637)	(2.2205)	(4.5711)	(6.6196)	(3.7306)
Partner working hours	-0.0541***	-0.0162	-0.0389***	-0.2170***	-0.1527***	-0.1918***
	(0.0140)	(0.0149)	(0.0103)	(0.0214)	(0.0269)	(0.0167)
Partner wage	-0.4118***	-0.2600***	-0.3480***	-0.8013***	-0.7814***	-0.7980***
	(0.0637)	(0.0654)	(0.0462)	(0.1555)	(0.2185)	(0.1266)
Child allowance (low / high)	-0.0118	-0.0092	-0.0105	-0.0227**	-0.0307**	-0.0251***
	(0.0092)	(0.0095)	(0.0067)	(0.0107)	(0.0133)	(0.0083)
Costly care (0 / 1)	-0.9834	-4.2644***	-2.3326**	-0.1154	-2.8029	-1.0508
	(1.2961)	(1.3418)	(0.9434)	(1.8535)	(2.5015)	(1.4871)
N	4385	2972	7357	1520	852	2372
R^2 (within)	0.028	0.033	0.026	0.183	0.155	0.168

Notes: Vignette survey from UAS-27 (US, 2015/16) and Pairfam wave 5 (Germany, 2013/14). Standard errors in parentheses. Estimates from linear fixed-effects model. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Cross-validation strategy and results

We call a *simulated panel* the synthetic dataset obtained by feeding each female vignette respondent's observed characteristics and four randomized vignette environments into our structural models, and recording the model's predicted desired fertility (and labor supply). In this exercise, we hold the model's structural parameters fixed at the country-specific estimated values and let the model's decision rules generate one prediction per (respondent $\times$ environment) cell. The resulting simulated panel mirrors the vignette panel exactly in structure — same respondents, same set of randomized environments, same dimensions — but with stated answers replaced by model-implied predictions, one for Model 1 and one for Model 2. This parallel structure enables the "apples-to-apples" cross-validation, because the population is held fixed and the policy variations for the predictions match the vignette environments.

The cross-validation strategy relies on the following idea: take the test population of female vignette respondents and compare their stated desired fertility responses for the different vignette environments with the model-based predictions of completed fertility.

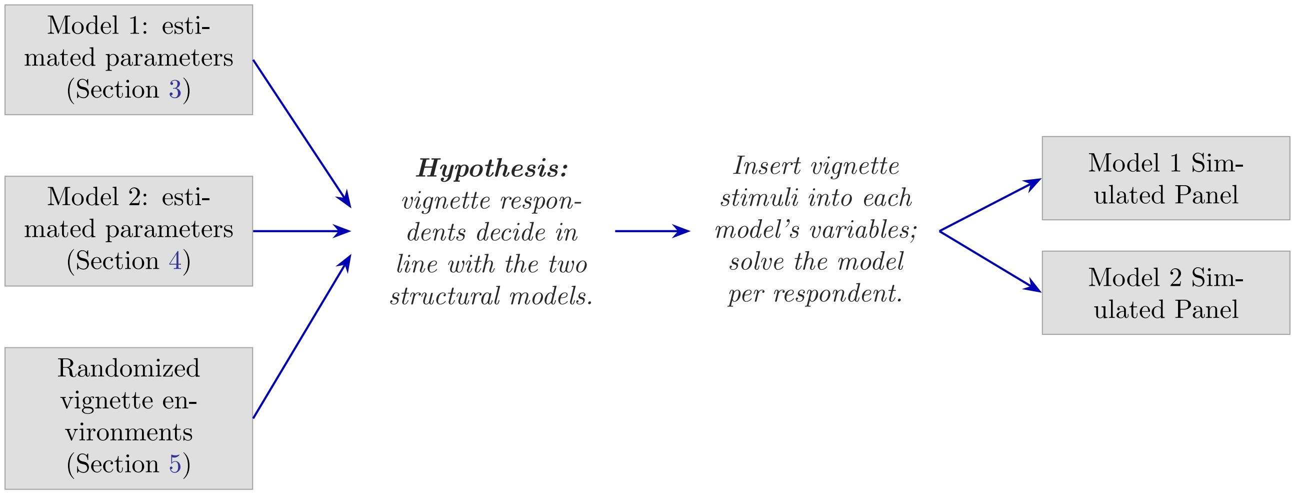
For each model, fertility predictions are a function of (a) the model's structure and estimated parameters, (b) the vignette environments, denoted by "e" (pronatalist policies, individual wage, and partner's wage and working hours),[45] and (c) vignette respondents' characteristics, denoted by "X" (age, marital status, number of children). Formally, the prediction is $\widetilde{F}(X_i, e_i)$, where X_i are respondent i's characteristics and e_i are respondent i's randomly assigned environment characteristics (or vignette stimuli). Therefore, model-specific predictions vary because of variations in environments and respondents' individual characteristics. The resulting database contains all the information from the vignettes and also the model predictions. To derive the latter, one needs the whole sequence of lifecycle decisions. This is why we call the output of this procedure a "simulated panel" (see Figure 4).

The simulated panel permits two cross-validations. *First*, to examine whether the vignette environments have the same qualitative implications for stated desired fertility (according to the vignette) and predicted fertility (according to the models). To this end, we regress stated and model-predicted fertility on the same set of control variables, defined by the vignette stimuli. This delivers an assessment of whether the model-based stimulus responses align with the stated fertility-preference responses. *Second*, importantly for policy evaluation, to evaluate whether the policy elasticities predicted by the models align with direct desired fertility responses from the vignette.

[45]Besides the policy variables, the vignettes comprise additional stimulus variables: the own wage, the wage of the partner, etc. The latter allow us to explore heterogeneities in fertility/labor supply responses. We thank an anonymous referee for pointing out this potential to us.

6.1 Construction of the simulated panels

Figure 4: Construction of the simulated panels for cross validation of vignette and models



We construct a *simulated panel* for each of the two structural models, using the information from the vignette respondents as initial conditions and the randomized vignette stimuli as model inputs. As detailed in Figure 4 and Table 13, the simulated panels for both models have a common structure.

Since the vignette samples played no role in the estimation, the model-implied fertility predictions are genuinely out-of-sample. The simulated panel is restricted to female respondents aged 22–42 to match the age range of the vignette survey. In the US, this leaves us with 3,056 observations from 764 respondents, and 1,128 observations from 282 respondents in Germany.

Each vignette environment is characterized by five randomized attributes: the household head's hourly net wage (w_e), the partner's working hours ($h_{p,e}$) and hourly net wage ($w_{p,e}$) (with partner labor income $y_{p,e} = h_{p,e} \cdot w_{p,e}$ as a derived quantity), the availability of childcare facilities ($s_e \in \{\text{costly, subsidized}\}$), and the allowance per child (A_e); we insert the exact values of these attributes per survey respondent, into each model's budget constraint as $w(t)$, $y_p(t)$, $\tilde{p}_b$, and A, while holding the country-specific structural parameters $\hat{\theta}$ fixed at the values reported in Section 3 (Model 1) and Section 4 (Model 2). The description of the environments in the vignette referred to the life stage of "family planning." This includes the attribute "wage." It is not clear if respondents interpret the wage as permanent or temporary. For this reason, we report two perception treatments that span the natural range of interpretations: a *permanent-wage* treatment, in which w_e replaces the wage process for all remaining periods of the lifecycle, and a *current-period-wage* treatment, in which w_e applies only at the respondent's current age; the policy stimuli (A_e, s_e) and the partner-income stimulus ($y_{p,e}$) are applied identically under both treatments.[46]

For the aggregative Model 1, for each respondent in every environment, we solve the model using $\hat{\theta}_{\text{M1}}$ from Section 3, taking the country's estimated wage per age and partner-income profiles as the baseline. The vignette stimuli enter Model 1 as time paths and scalar policies. Under the *permanent-wage* treatment, the wage and partner-income paths are replaced by the vignette values $w(t) = w_e$ and $y_p(t) = y_{p,e} = h_{p,e} \cdot w_{p,e}$ from the age the respondent answers the vignette ($t_0 = \text{age} - 21$) through the end of the lifecycle. Under the *current-period-wage* treatment, the vignette wage and partner income apply only at t_0; the estimated wage and partner income are restored for all later periods. The child allowance is $A = A_e$; and the daycare subsidy rate is $s = 0.95$ under subsidized care or $s = 0$ under costly care, applied to the baseline price as $\tilde{p}_b = (1 - s) \cdot \tilde{p}_b^{\text{baseline}}$. The model output for each respondent in every environment is the mean number of children over ages 28 to 45, Model 1's fertility phase.

For the heterogeneous-agent dynamic-programming (DP) Model 2, for each respondent and environment, we initialize it at the respective observed state ($t_0 = \text{age} - 21$, $n_0 = \min(\text{observed children}, 3)$, marital status as observed, experience $x_0 = \text{age} - 22$, age of youngest child fixed at 3 for everyone) and solve the dynamic program at $\hat{\theta}_{\text{M2}}$ from Section 4. The vignette stimuli enter Model 2 through its wage process and budget constraint: the wage-process intercept is set to $\phi_0 = \log w_e$ with $\phi_1 = \phi_2 = 0$ (so the deterministic component of $w_t = \exp(\phi_0 + \phi_1 x_t + \phi_2 x_t^2 + \xi_t)$ equals w_e, with the transitory productivity shock $\xi_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$ drawn independently each period (discretized onto a fixed grid of size S); the substitution does not alter the shock distribution); the annualized partner income is $y_{p,t} = 12 \cdot y_{p,e}$; the annualized allowance is $A = 12 \cdot A_e$; and the daycare subsidy rate is $s = 0.95$ under subsidized care and $s = 0$ under costly care. Again, we consider a *permanent-* and *current-period-wage* treatment.[47]

The model output for each respondent and environment is completed fertility $r^{\text{sim, M2}}_{\text{children}} = \overline{n_{45}}$, defined as the mean over $N_{\text{SIM}} = 10{,}000$ stochastic simulations and over Model 2's four heterogeneity types $(O_1, O_2) \in \{0, 1\}^2$, probability-weighted by the anchor heterogeneity distribution.

[46]We thank an anonymous referee for making both of these remarks, on (a) the permanent vs. temporary nature of the vignette wage stimuli, and (b) their application to the respondent's age at the time of the survey.

[47]Computationally, Model 2 implements the *current-period-wage* treatment as a two-DP-solution construction: at t_0 she uses the policy from DP_{Perm} (the dynamic program solved under the vignette wage permanent), and from $t_0 + 1$ onward she uses the policy from DP_{N} (solved under her estimated wage process). The non-wage stimulus variables (A_e, s_e, $y_{p,e}$) are applied identically across both DPs. See Appendix F, Section F.2.1 for details.

Table 13: Construction consistency: Model 1 and Model 2 simulated panels

Dimension	**Specification (shared by Model 1 and Model 2)**
Vignette respondents	female, age 22–42; US $n = 3{,}056$, DE $n = 1{,}128$
Randomized vignette stimuli	w_e, $h_{p,e}$, $w_{p,e}$, s_e, A_e
Wage treatment	Permanent and Current-period
Daycare subsidy rate	$s = 0.95$ (subsidized), $s = 0$ (costly)
Support of n	$\{0, 1, 2, 3\}$ (Model 2: discrete state grid; Model 1: per-age cap $\min(n(t), 3)$)

6.2 Cross-validation 1: qualitatively reproduce drivers of vignette responses

This section examines whether attributes of the environments imply similar model-based fertility predictions and stated fertility in the vignette. To this end, for the test sample, Table 14 reports a linear regression of the model-predicted and stated number of children on the vignette stimuli — hourly wage, partner earnings, child allowance, and the costly-care indicator — estimated separately for the vignette panel and for each of the four columns (Model 1 / Model 2 × Permanent / Current-period). The specification includes respondent fixed effects (FE) with cluster-robust standard errors at the respondent level. Because each respondent rates four vignette environments, the four stimulus coefficients are identified from within-respondent variation across vignette environments.

In both countries, the costly-care coefficient is negative and highly significant in all five regressions (see Table 14), indicating that lower external childcare costs increase fertility regardless of the wage treatment. Higher child allowances imply higher fertility, with 8 out of 10 coefficients being significant. The hourly-wage coefficient is always positive and significant in the United States. The same holds for the vignette panel and the simulated panel of Model 1. For Model 2, it is negative but only significant under the permanent-wage treatment. The partner-earnings coefficient is positive and significant in the vignette panel for both countries; in the structural models, the responses are smaller in magnitude and mixed in sign.

In sum, the first cross-validation is successful, with the two structural models qualitatively reproducing the drivers behind fertility responses in the vignettes. Quantitative differences between Models 1 and 2 reflect modeling differences between more and less parsimonious specifications, as anticipated in Section 6.1.

Table 14: Fertility predictions from vignette and simulated panels

	Vig. panel	Model 1 Sim. panel Permanent	Model 1 Sim. panel Current-period	Model 2 Sim. panel Permanent	Model 2 Sim. panel Current-period
US					
hourly wage	0.013***	0.017***	0.002***	0.002***	0.002***
	[0.010, 0.017]	[0.015, 0.018]	[0.001, 0.002]	[0.001, 0.002]	[0.001, 0.002]
partner: earnings (per $1,000)	+0.118***	-0.003	-0.001	+0.006	+0.011**
	[+0.085, +0.152]	[-0.013, +0.007]	[-0.002, +0.000]	[-0.002, +0.014]	[+0.003, +0.020]
transfer (per $100)	+0.032	+0.023***	+0.012***	+0.185***	+0.182***
	[-0.014, +0.079]	[+0.009, +0.037]	[+0.011, +0.013]	[+0.154, +0.217]	[+0.151, +0.213]
costly care (0,1)	-0.189***	-1.321***	-1.372***	-0.108***	-0.116***
	[-0.268, -0.110]	[-1.387, -1.256]	[-1.378, -1.366]	[-0.122, -0.093]	[-0.132, -0.100]
R^2	0.071	0.715	0.997	0.359	0.363
N	3,001	3,056	3,056	3,056	3,056
DE					
hourly wage	0.020***	0.029***	0.003***	-0.003***	-0.001
	[0.011, 0.029]	[0.024, 0.035]	[0.003, 0.004]	[-0.004, -0.002]	[-0.001, 0.000]
partner: earnings (per €1,000)	+0.240***	-0.019	-0.003**	-0.003	-0.010***
	[+0.166, +0.314]	[-0.043, +0.004]	[-0.006, -0.000]	[-0.008, +0.001]	[-0.016, -0.004]
transfer (per €100)	+0.132***	+0.014	+0.019***	+0.012***	+0.013***
	[+0.064, +0.201]	[-0.006, +0.033]	[+0.017, +0.021]	[+0.008, +0.016]	[+0.009, +0.017]
costly care (0,1)	-0.250***	-1.229***	-1.661***	-0.043***	-0.049***
	[-0.340, -0.160]	[-1.314, -1.145]	[-1.670, -1.651]	[-0.056, -0.031]	[-0.062, -0.036]
R^2	0.154	0.758	0.998	0.244	0.216
N	1,128	1,128	1,128	1,128	1,128

95% confidence intervals in brackets, cluster-robust at the respondent

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Data source: vignette panel from UAS-27 (US, 2015/16) and Pairfam wave 5 (Germany, 2013/14); simulated panels generated at $\hat{\theta}_{\mathrm{M1}}$ and $\hat{\theta}_{\mathrm{M2}}$ from Sections 3 and 4. Female participants only, age 22–42. *Permanent*: vignette wage applied for the full remaining lifecycle. *Current-period*: vignette wage applied for one period only. The Vignette panel reports $N = 3{,}001$ (US) and $N = 1{,}128$ (DE) because 55 US female respondents did not state r_{children} and are dropped from the regression sample; the simulated-panel columns use the model-predicted r_{children} and retain all 3,056 (US) / 1,128 (DE) vignette-panel observations.

6.3 Cross-validation 2: policy-effectiveness according to models and vignette

In structural models, estimated policy responses depend on, among others, the stipulated behavioral framework, functional forms, and other technical assumptions (see Todd and Wolpin, 2023, p. 44). Potentially, many model structures could match the same estimation targets equally well, but may have different policy implications. Thus, we cross-validate the policy implications from Models 1 and 2 with the implications derived from the vignette survey. This section conducts the cross-validation under a unified linear fixed-effects specification, which delivers a per-respondent *elasticity* that is invariant to the magnitude of the policy change (see below).

Building on the linear fixed-effects regressions of Section 6.2, we compute a per-respondent *elasticity* of children with respect to (i) the child allowance A and (ii) the daycare price p_b, separately for each of the five panels (vignette, Model 1 permanent / current-period, Model 2 permanent / current-period). For each panel and each respondent i, the FE model is predicted at two counterfactual values of the policy, X_{low} and X_{high}, yielding the respondent-specific fitted values $\hat{n}_{i,\text{low}}$ and $\hat{n}_{i,\text{high}}$. The per-respondent elasticity is

$$\varepsilon_{X,i} \;=\; \frac{\big(\hat{n}_{i,\text{high}} - \hat{n}_{i,\text{low}}\big)/\hat{n}_{i,\text{low}}}{\big(X_{\text{high}} - X_{\text{low}}\big)/X_{\text{low}}}\,, \qquad X \in \{A,\ p_b\},$$

indexed by respondent i across the 3,056 US (1,128 DE) vignette-panel observations.[48] Table 15 reports the means as well as the minimum and maximum of the estimates of the elasticity distribution $\{\varepsilon_{X,i}\}_{i=1}^{N}$, while Figure 5 displays the density histograms. Because the elasticities are invariant to the magnitude of the policy change and all estimates come from the same test sample, the estimates are directly comparable.

In both countries and across all five estimates from Models 1, 2 and vignette, the ranges of individual elasticities lie above zero for the allowance and below zero for daycare price, the latter indicating that daycare subsidies increase fertility. At the same time, the ranges of elasticities from Model 1 and Model 2 largely do not overlap. This is unsurprising because model estimates are conditional upon model structure and therefore self-referential. Yet, the usage of the same test dataset ensures comparability ("apples to apples"). In three out of four cases, the ranges of vignette-implied elasticities fall between the two models. For the allowance in Germany, the range of vignette elasticities lies beyond those of the two models. Taking the vignette findings as a guide, this calls for model refinement in future research.

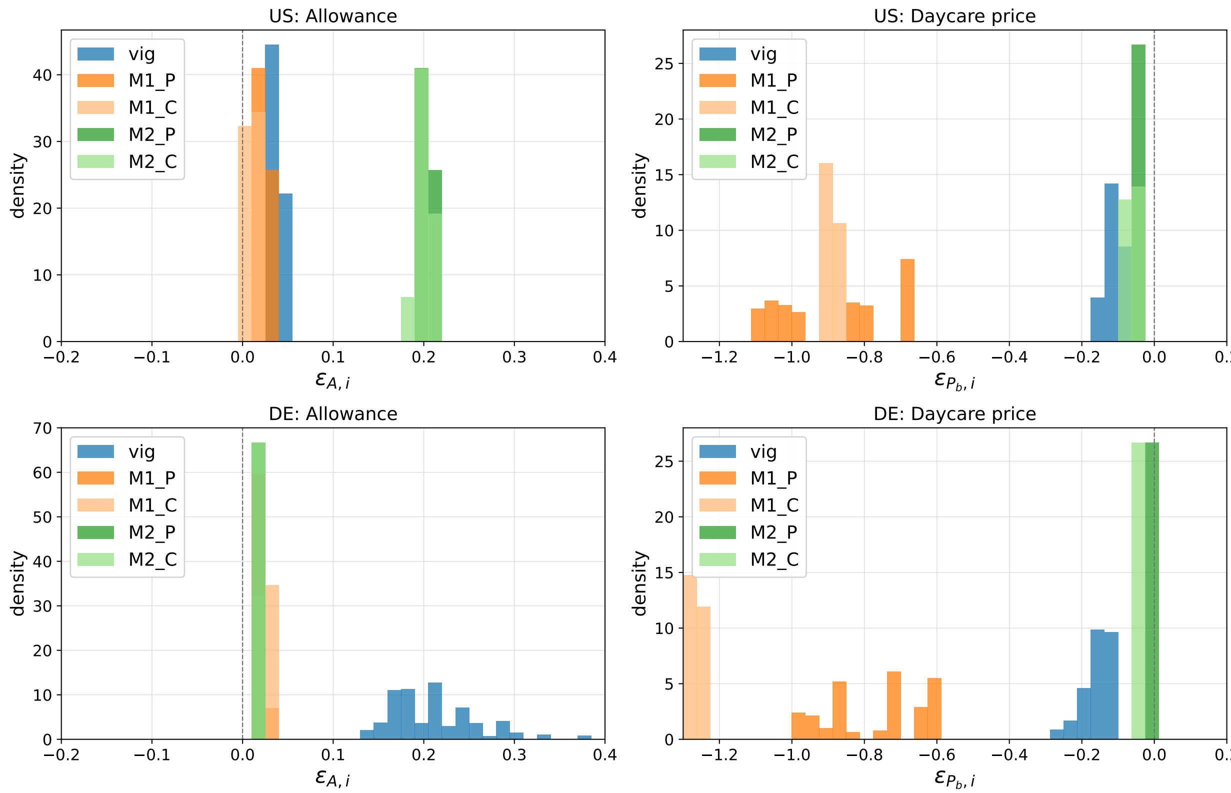
[48] Section 7 reports model-implied structural elasticities at $\hat{\theta}^*$ and places them within the empirical literature.

Table 15: Regression-based predicted individual fertility elasticities from vignette and simulated panels

	Vig. panel	Model 1 Sim. panel		Model 2 Sim. panel	
		Permanent	Current-period	Permanent	Current-period
US					
Allowance ε_A	$+0.037$	$+0.023$	$+0.012$	$+0.203$	$+0.200$
	$[+0.025, +0.054]$	$[+0.014, +0.037]$	$[+0.008, +0.016]$	$[+0.190, +0.214]$	$[+0.186, +0.212]$
Daycare price ε_{p_b}	-0.112	-0.887	-0.891	-0.055	-0.060
	$[-0.156, -0.077]$	$[-1.108, -0.674]$	$[-0.916, -0.864]$	$[-0.062, -0.049]$	$[-0.067, -0.053]$
N (respondent–vignette obs.)	3,056	3,056	3,056	3,056	3,056
DE					
Allowance ε_A	$+0.212$	$+0.017$	$+0.027$	$+0.017$	$+0.018$
	$[+0.139, +0.377]$	$[+0.010, +0.026]$	$[+0.016, +0.038]$	$[+0.016, +0.017]$	$[+0.018, +0.018]$
Daycare price ε_{p_b}	-0.155	-0.768	-1.316	-0.024	-0.027
	$[-0.285, -0.101]$	$[-0.992, -0.608]$	$[-1.390, -1.254]$	$[-0.025, -0.023]$	$[-0.028, -0.026]$
N (respondent–vignette obs.)	1,128	1,128	1,128	1,128	1,128

Note: Data source: vignette panel from UAS-27 (US, 2015/16) and Pairfam wave 5 (Germany, 2013/14); simulated panels generated at $\hat{\theta}_{\mathrm{M1}}$ and $\hat{\theta}_{\mathrm{M2}}$ from Sections 3 and 4. Per-respondent elasticities computed from a linear fixed-effects regression of the (stated or model-predicted) number of children on the four randomized policy stimulus variables, then predicted at counterfactual policy values and aggregated by the per-respondent elasticity formula $\varepsilon_{X,i} = \big[(\hat{n}_{i,\mathrm{high}} - \hat{n}_{i,\mathrm{low}})/\hat{n}_{i,\mathrm{low}}\big] \big/ \big[(X_{\mathrm{high}} - X_{\mathrm{low}})/X_{\mathrm{low}}\big]$. Under the linear FE specification, the elasticity is invariant to the magnitude of the policy change. Cells report the cross-respondent mean and $[\min, \max]$ over the indicated sample (female participants, age 22–42; transfer values: the allowance increases from USD200 to \$400 (DE €250 to €500); the daycare subsidy $s_{\mathrm{costly\_care}}$ varies from 0.05 (subsidized) to 1 (costly)).

Figure 5: Distribution of regression-based predicted individual fertility elasticities from vignette and simulated panels



Note: density histograms of the individual fertility elasticities $\{\varepsilon_{A,i}\}$ (left column) and $\{\varepsilon_{p_b,i}\}$ (right column) over female respondents aged 22–42, computed as in Table 15. Top row: US; bottom row: DE. Each subplot overlays five distributions: *vig* = vignette panel; *M1_P* = Model 1 with permanent-wage treatment; *M1_C* = Model 1 with current-period-wage treatment; *M2_P* = Model 2 with permanent-wage treatment; *M2_C* = Model 2 with current-period-wage treatment. Bin count: 40; X-axis range $[-0.20, +0.40]$ for the Allowance subplots and $[-1.30, +0.20]$ for the Daycare-price subplots; dashed line at $\varepsilon = 0$.

7 Literature comparison and quantitative model-based policy experiments

This section has two purposes. First, it compares the policy fertility elasticities from Section 6 with those from previous literature. Second, it performs a typical model-based back-of-envelope calculation, i.e., how policy costs translate into additional births.

Regarding the first purpose, Figure 6 shows the individual elasticities from the vignette and the two models (top six rows) then compares them to estimates from previous studies. These studies rely on estimated econometric models, applied theory models (such as ours), and causal econometric designs. While elasticities for the allowance and subsidy point in the same direction, they quantitatively differ across studies and countries. Furthermore, except for one case (the vignette-based elasticity for the allowance), our estimates are in the range of previous studies. This literature comparison is another indication that the vignette is a useful tool for ex-ante policy evaluation.

The structural elasticity of the model-implied number of children n with respect to a pronatalist policy $X \in \{A, p_b\}$, at the MSM estimate θ^* and age a, is

$$\varepsilon_A^{\text{str}}(\theta^*, a) = \frac{n\big(\theta^*,\, A_{\text{base}}(1+v_A),\, p_b^{\text{base}}\big) - n\big(\theta^*,\, A_{\text{base}},\, p_b^{\text{base}}\big)}{n\big(\theta^*,\, A_{\text{base}},\, p_b^{\text{base}}\big)} \Big/ v_A,$$

$$\varepsilon_{p_b}^{\text{str}}(\theta^*, a) = \frac{n\big(\theta^*,\, A_{\text{base}},\, p_b^{\text{base}}(1+v_{p_b})\big) - n\big(\theta^*,\, A_{\text{base}},\, p_b^{\text{base}}\big)}{n\big(\theta^*,\, A_{\text{base}},\, p_b^{\text{base}}\big)} \Big/ v_{p_b},$$

with $v_A = +0.20$ and $v_{p_b} = -0.20$. The other pronatalist policy is held at its baseline value in each case.[49] The same MSM estimate θ^* enters both the numerator and the denominator (own-θ baseline), so ε^{str} measures the pure policy response, free of parameter-shift noise.

The aggregate Model 1 and Model 2 policy elasticities at the estimated θ^* (per 20% policy variation) are reported in Table 16. Two outcomes are reported per (country, policy) cell: ε_n uses mean fertility (Model 1: post-$\bar{t}$ mean; Model 2: mean cumulative fertility), and ε_ℓ uses mean labor supply over the model's lifecycle window.[50] Both Model 1 and Model 2 report elasticities for a 20% reduction in the effective daycare cost (Model 2: subsidy raised from 0 to 0.20; Model 1: p_b reduced by 20%). Signs are aligned so that fertility responses to cost-reducing interventions appear positive. Model 1 represents children as an aggregate stock, without explicitly modeling their ages, while Model 2 tracks each child's age explicitly

[49] Country-specific baselines: p_b^{base} = \$8/hour (US) and €7/hour (DE), as set in Appendix A.1; A_{base} = \$89/month (US) and €200/month (DE), matching the calibrated allowance levels at which θ^* was estimated.

[50] We thank an anonymous referee for suggesting that we report labor-supply elasticities alongside fertility elasticities.

and embeds an age-specific demand for daycare: young children require formal care, older children do not. Thus, a daycare-cost reduction in Model 1 mechanically applies to *all* children in the aggregate count, while in Model 2 it benefits only those currently of daycare age. The two specifications represent the same economic intervention at different levels of granularity, and the mechanical aggregation in Model 1 amplifies the fertility response: Model 1's ε_n for the daycare subsidy is larger than Model 2's by a factor of roughly 1.6× in Germany and 7× in the United States — the cross-country range broadly reflects the inverse of the share of dependent-children years that actually require *intensive* daycare. The gap reflects Model 1's parsimonious aggregative structure, not a disagreement between the two models on the direction or qualitative pattern of fertility responses to daycare cost. Labor-supply elasticities, which respond to the presence of children of any age, are of comparable order of magnitude across the two models.

Table 16: Model 1 and Model 2 policy elasticities based on averaged cross-sectional model responses at the estimated θ^* (per 20% policy variation)

	US		**DE**	
Experiment	ε_n	ε_ℓ	ε_n	ε_ℓ
Model 1				
Child allowance (A) ×1.20 (+20%)	+0.013	+0.021	+0.054	+0.101
Subsidy rate (s): 0 → 0.20 (price −20%)	+1.262	+0.016	+1.326	+0.114
Model 2				
Child allowance (A) ×1.20 (+20%)	+0.266	−0.016	+0.348	−0.067
Subsidy rate (s): 0 → 0.20 (price −20%)	+0.174	−0.008	+0.835	−0.157

In Model 2 the labor supply is a discrete choice among three states, non-employment, part-time work, and full-time work, which mirrors the institutional structure of female labor markets and allows the model to capture the major intensive-margin shifts induced by changing childcare and allowance policies. Model 1's labor supply is a continuous fraction of available time and does not separate participation from hours, so its labor-supply response to policy reflects a marginal adjustment along a smooth continuum rather than discrete transitions across employment states. As a result, Model 2's labor-supply elasticities are systematically negative and economically interpretable (more children imply that parents are more time constrained), while Model 1's are small in absolute value and positive in all four cells (between 0.02 and 0.11 at the structural estimate), reflecting Model 1's simpler labor-supply specification rather than a substantive economic disagreement with Model 2.

Figure 6 places these model-implied elasticities within the empirical literature on child-allowance and daycare-price fertility elasticities. Our Model 2 estimates and the vignette estimates lie within the central range of values for both countries and both policies; Model 1's daycare-price elasticities sit at the upper end of the band in magnitude, because Model 1 applies the daycare-cost reduction to all children, whereas Model 2 applies it only to those of daycare age (decomposed in Appendix C.3).

Turning to the second purpose, we now conduct two model-based policy experiments using the structural estimates of the two models ($\hat{\theta}_{\mathrm{M1}}$ in Table 4 and $\hat{\theta}_{\mathrm{M2}}$ in Table 6). In the experiments, the allowance and the childcare subsidy imply a monthly cost of \$100 (€100) per family, so that both become fiscally comparable:[51]

1. Increase of child allowance equivalent to \$100 (€100) per month per household with children.

2. Subsidization of the daycare price equivalent to \$100 (€100) per month per household with children.

Table 17 summarizes the resulting fertility responses for both models in the United States and Germany, while Figure 7 shows the corresponding $N(t)$ age profiles.

Table 17: Fertility policy responses

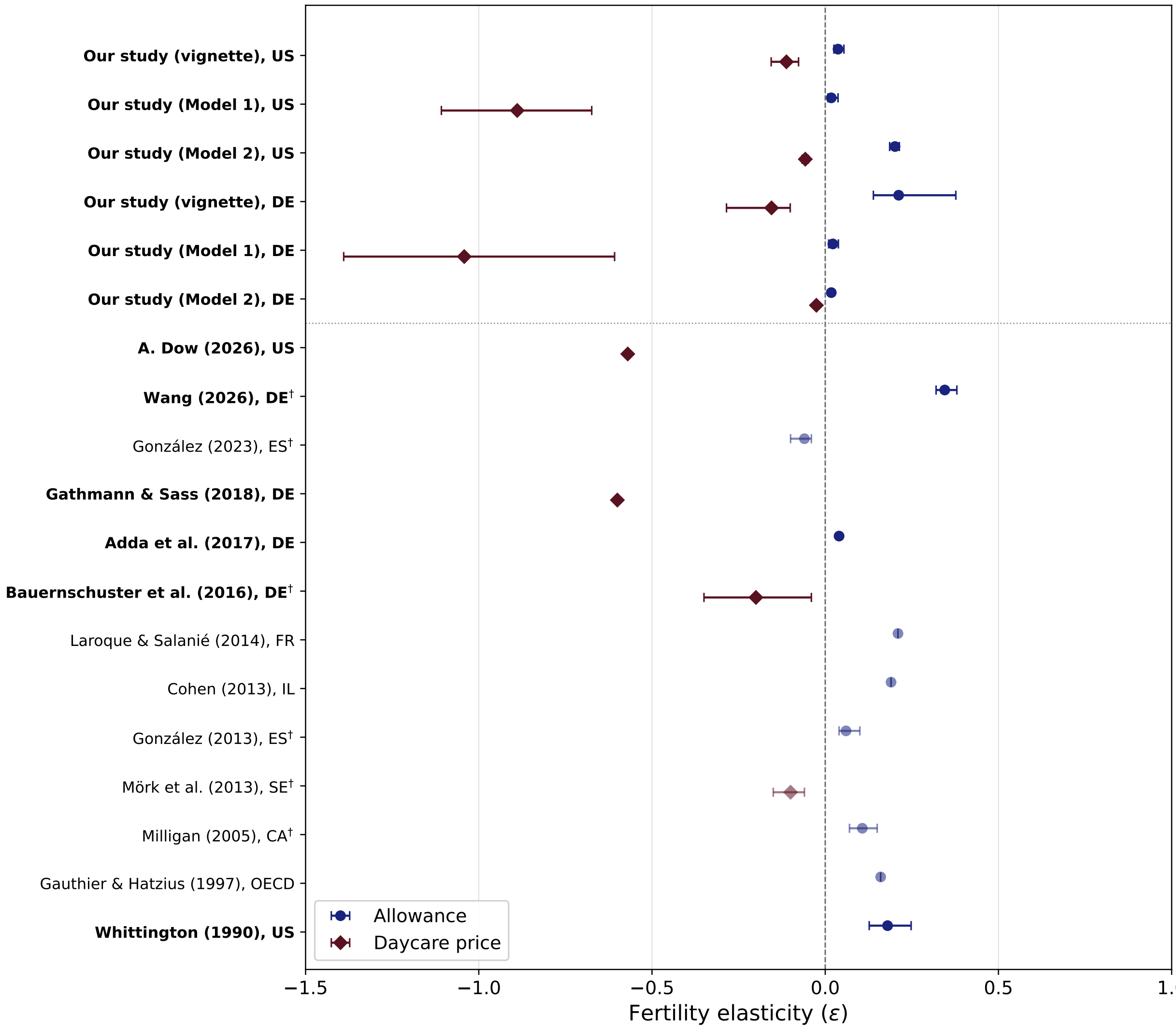
	Baseline N	After allowance	After daycare subsidy
Model 1 (average number of children, post-childbearing window)			
United States	1.571	1.585 (+0.9%)	1.956 (+24.5%)
Germany	1.230	1.255 (+2.1%)	1.898 (+54.3%)
Model 2 (completed fertility at age 45)			
United States	1.729	1.922 (+11.2%)	1.802 (+4.2%)
Germany	1.557	1.651 (+6.0%)	1.683 (+8.1%)

Notes: Model 1 reports the average number of children over the post-childbearing ages at the estimated model parameters. Model 2 reports completed fertility, i.e., the cumulative number of children ever born by age 45, at the estimated model parameters, averaged over 10,000 simulated lifecycles.

Three features stand out: (i) both models in both countries consistently indicate that both policies raise fertility; (ii) in Germany, both models consistently indicate that the daycare subsidy is more cost-effective than child allowances; (iii) in the United States, the

[51]In the budget constraints of the two models, allowances and daycare subsidies are specified per child; for cross-policy comparison, we convert each intervention into its household monetary-equivalent given the baseline number of children N_{base}: $\Delta A = \$100\,(€\,100)/N_{\mathrm{base}}$ per child per month for the allowance, and $\Delta\tilde{p}_b = -\$100\,(€\,100)/(N_{\mathrm{base}}\, m_{b,\mathrm{base}}\, H_{\mathrm{mo}})$ per hour for the daycare subsidy in Model 1 (with $H_{\mathrm{mo}} = 24\times 20 = 480$ work-hours per month). Model 2's daycare cost is structural (varying by child age and parent labor-supply choice) and has no scalar $\tilde{p}_b$, so we apply the equivalent fractional cost reduction via a subsidy rate s matched to Model 1's percentage Pb reduction at the same \$100 (€100) magnitude ($s = 0.196$ in the US, $s = 0.337$ in Germany).

Figure 6: Comparison with fertility elasticities from literature



† Elasticity approximated from reported birth-percentage effect and policy change.

Note: Coefficient plot of fertility elasticities with respect to the child allowance (ε_A, blue circles) and the daycare price (ε_{p_b}, dark-red diamonds). Our study is shown above the dotted separator (vignette, Model 1, Model 2 separately for the US and Germany); literature is shown below, sorted by publication year (most recent at top). United States and German studies (our focal countries) are highlighted in bold black; other-country studies are shown in regular weight with lower-opacity markers. The dagger (†) flags elasticities approximated from reported birth-percentage effects and policy changes. Markers indicate point estimates. For our study, bars span the $[\min, \max]$ of the per-respondent elasticity distribution from Table 15 (cross-respondent range). For Whittington et al. [1990], bars span the across-specification range reported in the original paper. For dagger-flagged entries, bars are approximate ranges consistent with the reported birth-percentage effect. Studies without bars report a single point estimate. Adda et al. [2017] reports the short-run elasticity (≈ 0 long-run). Gathmann and Sass [2018] report a childcare-attendance price elasticity (-0.6); the fertility effect itself is statistically insignificant overall (positive only for higher-parity families). The reduced-form childcare-price studies (Mörk et al. 2013, Bauernschuster et al. 2016, Dow 2026) report effects on period birth rates or birth timing, not completed fertility. Bauernschuster et al. [2016] study a childcare availability (coverage) expansion rather than a price change, so the plotted ε_{p_b} is an approximate price-equivalent conversion.

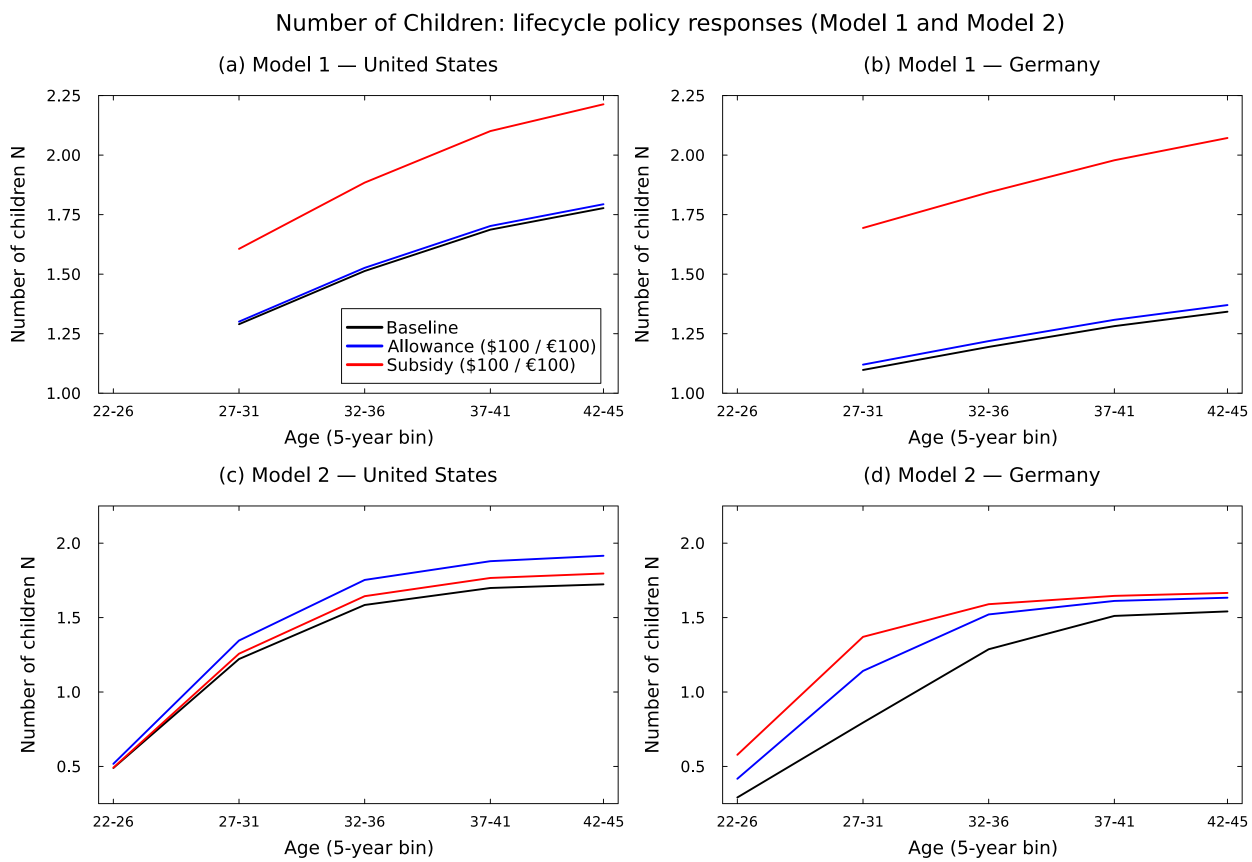


Figure 7: Age profiles of fertility under different policies

Notes: Three scenarios: baseline (black solid), +\$100 (+€100) per month increase of allowance (solid blue), and −\$100 (−€100) per month daycare subsidy (solid red). Top row: Model 1, panels (a) US, (b) Germany; bottom row: Model 2 with $N_{\text{SIM}} = 10{,}000$, panels (c) US (subsidy rate $s = 0.196$), (d) Germany (subsidy rate $s = 0.337$). The Model 1 N is $n(t)$ at each age; the Model 2 N is cumulative fertility (number of children ever born by age t).

models yield inconsistent policy ranking[52] with Model 1 assessing the daycare subsidy and Model 2 the allowance as more cost-effective.[53]

To gauge the macroeconomic scale, we conduct a back-of-the-envelope fiscal-cost calculation for 2020 using the Model-selected cost-effective policies. Two ingredients enter the calculation. First, the *per-child equivalent* of the \$100 (€100) per month policy is $\Delta A = \$100/N_{\text{base}}$ per child per month (annualized: $12\Delta A$ per child per year), applied to the eligible pool of children (under five for the daycare subsidy; under eighteen for the allowance). Second, the policy also generates *additional births* in the rollout cohort, equal to the policy's fractional response rescaled to the actual 2020 birth count:

$$\text{additional births} \;=\; \underbrace{\frac{N_{\text{policy}} - N_{\text{base}}}{N_{\text{base}}}}_{\%\,\Delta N \text{ from Table 17}} \times \underbrace{\text{births}_{2020}}_{\text{actual cohort size}} .$$

Adding the additional births to the baseline eligible pool and multiplying by $12\Delta A$ gives the annual fiscal cost.[54]

In the United States, Model 1 ranks the daycare subsidy as more cost-effective. The per-child equivalent is \$100/1.571 = \$63.7 per month (\$764 per year), and the additional births from the policy are $0.385/1.571 \times 3{,}605{,}201 \approx 883{,}500$. The implied annual fiscal cost is $(18{,}500{,}000 + 883{,}500) \times \$764 \approx \$14.8$ billion, or 0.07% of GDP. Model 2 favors the allowance, with $N_{45} = 1.729$ and a per-child equivalent of \$57.8 per month (\$694 per year). With ≈ 73.5 million eligible children under eighteen, the implied annual fiscal cost is ≈ \$51.0 billion, or 0.24% of GDP.

In Germany, both models favor the daycare subsidy. Model 1 has $N_{\text{base}} = 1.230$ and a per-child equivalent of €81.3 per month (€976 per year); with ≈ 3.8 million children under five plus $0.668/1.230 \times 773{,}144 \approx 419{,}900$ additional births, the annual fiscal cost is ≈ €4.1 billion (0.12% of GDP). Model 2 has $N_{45} = 1.557$ and a per-child equivalent of €64.2 per month (€771 per year), implying ≈ €3.0 billion or 0.09% of GDP. In all four (model, country) cells, the implied annual fiscal cost is in the range 0.07–0.24% of GDP at the policy magnitudes considered here.

Three policy-relevant messages emerge. First, both interventions raise fertility in both

[52]The architectural reasons — Model 1's aggregate-count daycare-cost mechanism and the channel-cancellation in Model 1's allowance response — are decomposed in Appendix C.3.

[53]For Norway, Chan and Liu [2018] find significant fertility effects of cash transfers, particularly among less-educated women. The model-specific pattern reported here suggests that cross-country differences may depend on the structural specification used. Chan and Liu [2018] is omitted from Figure 6 because it does not report directly comparable elasticity estimates.

[54]2020 cohort and stock data used below: US births 3,605,201, children under five ≈ 18.5 million, children under eighteen ≈ 73.1 million, GDP \$21.060 trillion; DE births 773,144, children under five ≈ 3.8 million, children under eighteen ≈ 13.8 million, GDP €3.329 trillion.

models and both countries, so the qualitative case for pronatalist policy is robust to the choice of structural model. Second, the implied annual fiscal cost ranges from 0.07% to 0.24% of GDP across the four (model, country) cells—well within a politically feasible budget envelope. Third, the daycare subsidy operates on a meaningfully calibrated margin in both models: Model 1's post-$\bar{t}$ baseline level of external childcare time per child, $\bar{m}_b = 0.085$ in the US (≈ 2.0 hours per work-day per child) and 0.072 in Germany (≈ 1.7 hours per work-day per child), is within rounding of the calibration target $1/12 \approx 0.0833$ (Table 5); Model 2's equivalent subsidy rate, $s = 0.196$ in the US and $s = 0.337$ in Germany, lies within the empirical range of childcare subsidies in the two countries. Together, these findings give policymakers a meaningfully bounded set of expected outcomes, even where the two structural models disagree on the precise ranking of allowance versus subsidy in the United States.

8 Conclusion

Two broad approaches are commonly used to assess the effectiveness of policy interventions. The first relies on estimated or calibrated structural models. A key advantage of this approach is that it can provide ex ante estimates of the effects of pronatalist policies before they are implemented, thereby informing policy design and evaluation. At the same time, multiple models may be consistent with the same observed data while yielding conflicting policy implications. The second approach evaluates implemented policies using causal econometric methods applied to field data. Its primary advantage is that it provides direct empirical evidence on policy effects. Its limitation, however, is that policies can only be evaluated ex post, once they have been implemented.

Building on the ideas of Todd and Wolpin [2023], we designed and implemented a three-stage framework that combines the strengths of both approaches. The framework uses direct empirical evidence to discipline structural analysis while retaining the ability to conduct ex-ante evaluations of policies that have not yet been implemented: (a) for direct experimental evidence, we developed and implemented a vignette survey that randomizes out-of-sample policies; (b) for structural analysis, we solved and estimated two competing applied-theory models (one aggregative and parsimonious; another richer and respecting cross-sectional heterogeneity); (c) for cross-validation, we used the vignette sample as an independent test database for comparing model-predicted policy responses with direct estimates from the vignette.

We applied this framework to the evaluation of out-of-sample pronatalist policies. Such policies are particularly difficult to assess because their effects often materialize only with substantial delays, and because fertility responses are shaped by social norms and other equilibrium forces that evolve over time. The application provided respondents with two types of randomized stimuli: (a) pronatalist policy variables (levels of child allowances and daycare subsidies), and (b) economic household features (individual wages and partner's earnings). We asked respondents to state desired fertility and working hours. Hence, by design, the vignette addressed career–fertility tradeoffs.

The vignette was integrated into two existing representative panels, one in the US and another in Germany. The longitudinal component of the panels allowed us to explore whether respondents' statements agreed with their actual-life fertility and labor-supply choices about a decade later, a necessary condition for whether they "meant what they said."

The estimation of the competing structural models relied on independent, representative micro-data from both countries. As in any estimated model, goodness of fit to data relies on model specification, i.e., the models are self-referential, and policy implications can be sensitive to model specification, leading to "apples-to-oranges" policy-evaluation comparisons.

The third step, which is the heart of our method, involved using the vignette samples as independent test datasets. In the test datasets, (i) the personal characteristics of the respondents were fixed, and (ii) the variation in the randomized vignette-environment stimuli was fixed. Inserting both into the estimated models generated policy predictions through model simulations. These individual-level predictions could be compared directly with the stated responses from the vignette. In this way, we could achieve direct "apples-to-apples" comparability of policy implications across both models and the vignette.

Regarding the policy assessment, our main findings indicated that stated fertility responses of the vignette were not just in the ballpark of the directly comparable policy predictions of our two models, but also with evidence from previous literature (which are less directly comparable). Daycare subsidies seemed more cost-effective than child allowances. Methodologically, the main message from our study was that the three-step procedure achieves four goals: (1) comparability between vignette-survey policy implications and model implications, (2) comparability across models, (3) guidance from the vignette survey regarding the direction of economic-behavior mechanisms that one should pursue in building applied-theory models, and (4) guidance from our vignette survey regarding the direction of policy implications of applied-theory models. Further developing vignette-survey instruments can circumvent the cost of designing and executing field experiments. Using them as navigation guides on how to specify applied-theory models for out-of-sample policy evaluations is a direction for future research on externally validating policy implications of model-based research.

References

J. Adda, C. Dustmann, and K. Stevens. The Career Costs of Children. *Journal of Political Economy*, 125(2):293–337, 2017.

S. Albanesi and A. Sahin. The Gender Unemployment Gap. *Review of Economic Dynamics*, 30:47–67, 2018.

A. Alesina, S. Stantcheva, and E. Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–554, 2018.

R. Alessie, S. Hochguertel, and A. van Soest. Ownership of Stocks and Mutual Funds: A Panel Data Analysis. *Review of Economics and Statistics*, 86:783–796, 2001.

J. G. Altonji and L. M. Segal. Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business and Economic Statistics*, 14:353–366, 1996.

J. D. Angrist and W. N. Evans. Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*, 88:450–477, 1998.

O. Attanasio, H. Low, and V. Sánchez-Marcos. Explaining Changes in Female Labor Supply in a Life-Cycle Model. *American Economic Review*, 98(4):1517–1552, 2008.

O. P. Attanasio, C. Meghir, and A. Santiago. Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. *The Review of Economic Studies*, 79(1):37–66, 2012.

Australian Government. Interim Home Based Carer Subsidy Programme (Nanny Pilot Programme). Technical report, Department of Education and Training, May 2016. https://docs.education.gov.au/system/files/doc/other/nanny_pilot_programme_-_document_-_guidelines_and_addendum.pdf.

R. J. Barro and G. S. Becker. Fertility Choice in a Model of Economic Growth. *Econometrica*, 57:481–501, 1989.

R. J. Barro and X. Sala-i Martin. *Economic Growth.* MIT Press, Cambridge, MA, 2nd edition, 2004.

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

S. Bauernschuster, T. Hener, and H. Rainer. Children of a (Policy) Revolution: The Introduction of Universal Child Care and Its Effect on Fertility. *Journal of the European Economic Association*, 14(4):975–1005, 2016.

G. S. Becker and R. J. Barro. A Reformulation of the Economic Theory of Fertility. *Quarterly Journal of Economics*, 103:1–25, 1988.

M. Beckmannshagen and C. Schröder. Earnings Inequality and Working Hours Mismatch. *Labour Economics*, 76:102184, 2022.

M. Bertoni. Hungry Today, Unhappy Tomorrow? Childhood Hunger and Subjective Well-being Later in Life. *Journal of Health Economics*, 40:40–53, 2015.

M. Bertrand and S. Mullainathan. Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review*, 91(2):67–72, 2001.

A. Bick. The Quantitative Role of Child Care for Female Labor Force Participation and Fertility. *Journal of the European Economic Association*, 14:639–668, 2016.

D. M. Blau and P. K. Robins. Child-Care Costs and Family Labor Supply. *Review of Economics and Statistics*, 70:374–381, 1988.

R. Blundell and T. MaCurdy. Labor Supply: A Review of Alternative Approaches. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 3, chapter 27, pages 1559–1695. Elsevier, 1999.

R. Borck. Adieu Rabenmutter – Culture, Fertility, Female Labour Supply, the Gender Wage Gap and Childcare. *Journal of Population Economics*, 27:739–765, 2014.

Y. Brilli, D. Del Boca, and C. D. Pronzato. Does Child Care Availability Play a Role in Maternal Employment and Children's Development? Evidence from Italy. *Review of Economics of the Household*, 14:27–51, 2016.

M. Browning. Children and Household Economic Behavior. *Journal of Economic Literature*, 30:1434–1475, 1992.

G. Burtless and J. A. Hausman. The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment. *Journal of Political Economy*, 86:1103–1130, 1978.

G. G. Cain. *Married Women in the Labor Force: An Economic Analysis*. University of Chicago Press, 1966.

A. C. Cameron and P. K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005.

C. D. Carroll and A. A. Samwick. The Nature of Precautionary Wealth. *Journal of Monetary Economics*, 40(1):41–71, 1997.

F. Caselli and J. Ventura. A Representative Consumer Theory of Distribution. *American Economic Review*, 90:909–926, 2000.

D. Cesarini, E. Lindqvist, R. Östling, and B. Wallace. Wealth, Health, and Child Development: Evidence from Administrative Data on Swedish Lottery Players. *Quarterly Journal of Economics*, 131:687–738, 2016.

M. K. Chan and K. Liu. Life-cycle and Intergenerational Effects of Child Care Reforms. *Quantitative Economics*, 9:659–706, 2018.

P.-A. Chiappori. Introducing Household Production in Collective Models of Labor Supply. *Journal of Political Economy*, 105:191–209, 1997.

CIA World Factbook. Country Comparison: Total Fertility Rate, 2016. https://www.cia.gov/library/publications/the-world-factbook/rankorder/2127rank.html.

A. Cohen, R. Dehejia, and D. Romanov. Financial Incentives and Fertility. *Review of Economics and Statistics*, 95:1–20, 2013.

D. de la Croix and M. Doepke. Inequality and Growth: Why Differential Fertility Matters. *American Economic Review*, 93:1091–1113, 2003.

A. Delavande and C. F. Manski. Using Elicited Choice Probabilities in Hypothetical Elections to Study Decisions to Vote. *Electoral Studies*, 38:28–37, 2015.

W. Ding, M. J. Kottelenberg, and S. F. Lehrer. Anticipating the (Un)expected: Evidence from Introducing a Universal Childcare Policy with a Shortage of Spaces. Working paper, Queen's University, 2020.

M. Doepke. Accounting for Fertility Decline During the Transition to Growth. *Journal of Economic Growth*, 9:347–383, 2004.

M. Doepke and F. Kindermann. Bargaining over Babies: Theory, Evidence, and Policy Implications. *American Economic Review*, 109(9):3264–3306, 2019.

D. B. Dunson, B. Colombo, and D. D. Baird. Changes with Age in the Level and Duration of Fertility in the Menstrual Cycle. *Human Reproduction*, 17(5):1399–1403, 2002.

Z. Eckstein and K. I. Wolpin. The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey. *Journal of Human Resources*, 24(4):562–598, 1989.

Z. Eckstein, M. Keane, and O. Lifshitz. Career and Family Decisions: Cohorts Born 1935–1975. *Econometrica*, 87(1):217–253, 2019.

M. Fort, A. Ichino, and G. Zanella. Cognitive and Noncognitive Costs of Day Care at Age 0–2 for Children in Advantaged Families. *Journal of Political Economy*, 128(1):158–205, 2020.

M. Francesconi. A Joint Dynamic Model of Fertility and Work of Married Women. *Journal of Labor Economics*, 20:336–380, 2002.

S. Galiani and J. Pantano. Structural Models. In K. F. Zimmermann, editor, *Handbook of Labor, Human Resources and Population Economics*. Springer, Cham, 2023.

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

D. Gavresi, A. Irmen, and A. Litina. Population Aging and the Rise of Populist Attitudes in Europe. DEM Discussion Paper Series 2023-10, Department of Economics and Management, University of Luxembourg, 2023.

G.-L. Gayle and R. A. Miller. Life-Cycle Fertility and Human Capital Accumulation. Working paper, Carnegie Mellon University, 2012.

German Federal Statistical Office. Social Statistics, 2016. https://www.destatis.de/EN/FactsFigures/SocietyState/SocialStatistics/SocialStatistics.html.

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

C. Goldin. *Understanding the Gender Gap: An Economic History of American Women*. Oxford University Press, Oxford, 1990.

J. R. Goldstein, C. Koulovatianos, J. Li, and C. Schröder. UAS 27: Vignette Survey Exploring Fertility and Labor Market Participation. Understanding America Study, Center for Economic and Social Research, University of Southern California, 2016. https://uasdata.usc.edu/survey/UAS+27.

E. Gould, T. Cooke, A. Davis, and W. Kimball. Economic Policy Institute 2015 Family Budget Calculator: Technical Documentation. EPI Working Paper 299, Economic Policy Institute, 2015. URL https://www.epi.org/publication/family-budget-calculator-technical-documentation/.

P.-O. Gourinchas and J. A. Parker. Consumption over the Life Cycle. *Econometrica*, 70: 47–89, 2002.

J. Greenwood, N. Guner, and J. A. Knowles. More on Marriage, Fertility, and the Distribution of Income. *International Economic Review*, 44:827–862, 2003.

P. Haan and K. Wrohlich. Can Child Care Policy Encourage Employment and Fertility? Evidence from a Structural Model. *Labour Economics*, 18(4):498–512, 2011.

L. P. Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, 1982.

J. J. Heckman. The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process. In C. F. Manski and D. L. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 179–195. MIT Press, London, 1981a.

J. J. Heckman. Statistical Models for Discrete Panel Data. In C. F. Manski and D. L. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 114–178. MIT Press, London, 1981b.

J. J. Heckman and J. R. Walker. Using Goodness of Fit and Other Criteria to Choose Among Competing Duration Models: A Case Study of Hutterite Data. *Sociological Methodology*, 17:248–307, 1987.

J. J. Heckman and J. R. Walker. The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data. *Econometrica*, 58:1411–1441, 1990a.

J. J. Heckman and J. R. Walker. The Third Birth in Sweden. *Journal of Population Economics*, 3:235–275, 1990b.

V. J. Hotz and R. A. Miller. An Empirical Analysis of Life Cycle Fertility and Female Labor Supply. *Econometrica*, 56:91–118, 1988.

V. J. Hotz and R. A. Miller. Conditional Choice Probabilities and the Estimation of Dynamic Models. *Review of Economic Studies*, 60:497–529, 1993.

Household Finance and Consumption Network. The Eurosystem Household Finance and Consumption Survey: Results from the First Wave. Statistics Paper Series 2, European Central Bank, Apr. 2013. URL https://www.ecb.europa.eu/pub/pdf/scpsps/ecbsp2.en.pdf.

H. M. Hoynes. Work, Welfare and Family Structure. In A. J. Auerbach, editor, *Fiscal Policy: Lessons from Economic Research*. MIT Press, Cambridge, MA, 1997.

D. E. Hyatt and W. J. Milligan. Can Public Policy Affect Fertility? *Canadian Public Policy*, 17(1):77–85, 1991.

L. E. Jones, A. Schoonbroodt, and M. Tertilt. Fertility Theories: Can They Explain the Negative Fertility-Income Relationship? In J. B. Shoven, editor, *Demography and the Economy*, pages 43–100. University of Chicago Press, 2010.

A. Kapteyn, J. P. Smith, and A. van Soest. Vignettes and Self-Reports of Work Disability in the United States and the Netherlands. *American Economic Review*, 97:461–473, 2007.

M. P. Keane and K. I. Wolpin. The Career Decisions of Young Men. *Journal of Political Economy*, 105(3):473–522, 1997.

M. P. Keane and K. I. Wolpin. Exploring the Usefulness of a Non-Random Holdout Sample for Model Validation: Welfare Effects on Female Behavior. *International Economic Review*, 48:1351–1378, 2007.

M. P. Keane and K. I. Wolpin. The Role of Labor and Marriage Markets, Preference Heterogeneity and the Welfare System in the Life Cycle Decisions of Black, Hispanic and White Women. *International Economic Review*, 51:851–892, 2010.

M. P. Keane, P. E. Todd, and K. I. Wolpin. The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications. In *Handbook of Labor Economics*, volume 4, Part A, pages 331–461. Elsevier, 2011.

M. S. Kearney and P. B. Levine. Why Is Fertility So Low in High Income Countries? NBER Working Paper 33989, National Bureau of Economic Research, 2025.

Y.-I. A. Kim. Lifetime Impact of Cash Transfer on Fertility. *Canadian Studies in Population*, 41:97–110, 2014.

S. Kimberlin. Documentation: Calculating Federal and State Income Taxes Using NBER's Internet TAXSIM Version 9 with PSID Data. Technical documentation, Institute for

Social Research, University of Michigan, 2015. URL https://simba.isr.umich.edu/help/taxsim.aspx.

G. King and J. Wand. Comparing Incomparable Survey Responses: New Tools for Anchoring Vignettes. *Political Analysis*, 15:46–66, 2007.

C. Koulovatianos and C. Schröder. Income-dependent Equivalence Scales and Choice Theory: Implications for Poverty Measurement. In J. Silber, editor, *Research Handbook on Measuring Poverty and Deprivation*. Edward Elgar Publishing, 2023. ISBN 978-1-80088-345-1.

C. Koulovatianos, C. Schröder, and U. Schmidt. On the Income Dependence of Equivalence Scales. *Journal of Public Economics*, 89:967–996, 2005.

C. Koulovatianos, C. Schröder, and U. Schmidt. Nonmarket Household Time and the Cost of Children. *Journal of Business and Economic Statistics*, 27:42–51, 2009.

C. Koulovatianos, C. Schröder, and U. Schmidt. Do Demographics Prevent Consumption Aggregates from Reflecting Micro-level Preferences? *European Economic Review*, 111: 166–190, 2019.

A. B. Krueger and A. A. Stone. Progress in Measuring Subjective Well-being. *Science*, 346 (6205):42–43, 2014.

G. Laroque and B. Salanié. Identifying the Response of Fertility to Financial Incentives. *Journal of Applied Econometrics*, 29:314–332, 2014.

E. L. Lehrer and M. Nerlove. Female Labor Force Behavior and Fertility in the United States. *Annual Review of Sociology*, 12:181–204, 1986.

M. F. Lovenheim and K. J. Mumford. Do Family Wealth Shocks Affect Fertility Choices? Evidence from the Housing Market. *Review of Economics and Statistics*, 95:464–475, 2013.

H. Low, C. Meghir, L. Pistaferri, and A. Voena. Marriage, Labor Supply and the Dynamics of the Social Safety Net. Working Paper 2025-27, Federal Reserve Bank of Chicago, 2025.

R. E. Lucas. Econometric Policy Evaluation: A Critique. In *Carnegie-Rochester Conference Series on Public Policy*, volume 1, pages 19–46. North-Holland, 1976.

C. F. Manski. Analysis of Choice Expectations in Incomplete Scenarios. *Journal of Risk and Uncertainty*, 19(1–3):49–65, 1999.

C. F. Manski and J. Mayshar. Private Incentives and Social Interactions: Fertility Puzzles in Israel. *Journal of the European Economic Association*, 1:181–211, 2003.

J. Marschak. Economic Measurements for Policy and Prediction. In W. C. Hood and T. C. Koopmans, editors, *Studies in Econometric Method*, pages 1–26. Wiley, New York, 1953.

D. McFadden. Measuring Willingness-To-Pay for Transportation Improvements. In T. Gärling, T. Laitila, and K. Westin, editors, *Theoretical Foundations of Travel Choice Modeling*, pages 339–364. Elsevier Science, Amsterdam, 1998.

A. W. Meade and S. B. Craig. Identifying Careless Responses in Survey Data. *Psychological Methods*, 17:437–455, 2012.

K. Milligan. Subsidizing the Stork: New Evidence on Tax Incentives and Fertility. *Review of Economics and Statistics*, 87:539–555, 2005.

J. Mincer. Labor Force Participation of Married Women: A Study of Labor Supply. In H. G. Lewis, editor, *Aspects of Labor Economics*, pages 63–97. Princeton University Press, Princeton, 1962.

R. A. Moffitt. The Estimation of a Joint Wage-Hours Labor Supply Model. *Journal of Labor Economics*, 2:550–566, 1984.

R. A. Moffitt. The Effect of Welfare on Marriage and Fertility. In R. A. Moffitt, editor, *Welfare, the Family, and Reproductive Behavior*. National Academy Press, Washington, DC, 1998.

S. P. Morgan and H. Rackin. The Correspondence Between Fertility Intentions and Behavior in the United States. *Population and Development Review*, 36(1):91–118, 2010.

A. Nakamura and M. Nakamura. The Econometrics of Female Labor Supply and Children. *Econometric Reviews*, 11:1–71, 1992.

W. K. Newey and D. McFadden. Large Sample Estimation and Hypothesis Testing. In R. F. Engle and D. L. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier, 1994.

D. Parent and L. Wang. Tax Incentives and Fertility in Canada: Quantum vs. Tempo Effects. *Canadian Journal of Economics*, 40:371–400, 2007.

R. Patavardhan et al. Looming Battle for Reproductive Rights: Discourses of Right-wing Populism and Abortion Laws in USA and India. In J. Chacko Chennattuserry, M. Deshpande, and P. Hong, editors, *Encyclopedia of New Populism and Responses in the 21st Century*. Springer, Singapore, 2024. doi: 10.1007/978-981-16-9859-0_377-1.

J. Rodriguez. Understanding the Effects of Workfare Policies on Child Human Capital. *Journal of Labor Economics*, 41(1):39–75, 2023.

M. Rohrmeier. *Populism, abortion and human rights: a legal analysis of populist tactics to restrict reproductive rights across the US and Europe*. PhD thesis, University Pompeu Fabra, 2025.

M. R. Rosenzweig and T. P. Schultz. The Demand for and Supply of Births: Fertility and Its Life Cycle Consequences. *American Economic Review*, 75:992–1015, 1985.

M. R. Rosenzweig and K. I. Wolpin. Life-Cycle Labor Supply and Fertility: Causal Inference from Household Models. *Journal of Political Economy*, 88:328–348, 1980a.

M. R. Rosenzweig and K. I. Wolpin. Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment. *Econometrica*, 48:227–240, 1980b.

A. B. Royalty. Job-to-Job and Job-to-Nonemployment Turnover by Gender and Education Level. *Journal of Labor Economics*, 16(2):392–443, 1998.

A. Schoonbroodt and M. Tertilt. Property Rights and Efficiency in OLG Models with Endogenous Fertility. *Journal of Economic Theory*, 150:551–582, 2014.

F. Schorfheide and K. I. Wolpin. On the Use of Holdout Samples for Model Selection. *American Economic Review: Papers and Proceedings*, 102:477–481, 2012.

F. Schorfheide and K. I. Wolpin. To Hold Out or Not to Hold Out. *Research in Economics*, 70:332–345, 2016.

Statistisches Bundesamt (Destatis). KORREKTUR: Arbeitskräfteangebot: Erwerbsbeteiligung unter den 25- bis 54-Jährigen, 2023. https://www.destatis.de/DE/Presse/Pressemitteilungen/2023/04/PD23_N026_13.html.

M. Tertilt. Polygyny, Fertility, and Savings. *Journal of Political Economy*, 113:1341–1371, 2005.

C. Tietze. Reproductive Span and Rate of Reproduction Among Hutterite Women. *Fertility and Sterility*, 8(1):89–97, 1957.

P. E. Todd and K. I. Wolpin. Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility. *American Economic Review*, 96:1384–1417, 2006.

P. E. Todd and K. I. Wolpin. The Best of Both Worlds: Combining Randomized Controlled Trials with Structural Modeling. *Journal of Economic Literature*, 61(1):41–85, 2023.

A. Tversky and D. Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, 1974.

H. Wang. Fertility and Family Leave Policies in Germany: Optimal Policy Design in a Dynamic Framework. Job market paper, Universitat Autònoma de Barcelona, 2022.

H. Wang. Fertility and Family Leave Policies in Germany: Optimal Policy in a Dynamic Framework. *The Review of Economic Studies*, 2026. Accepted.

L. A. Whittington, J. Alm, and H. E. Peters. Fertility and the Personal Exemption: Implicit Pronatalist Policy in the United States. *American Economic Review*, 80(3):545–556, 1990.

K. I. Wolpin. An Estimable Dynamic Stochastic Model of Fertility and Child Mortality. *Journal of Political Economy*, 92:852–874, 1984.

J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2001.

World Bank. Fertility Rate, Total (Births per Woman). World Development Indicators, 2023. http://data.worldbank.org/indicator/SP.DYN.TFRT.IN.

M.-J. Wu, K. Zhao, and F. Fils-Aime. Response Rates of Online Surveys in Published Research: A Meta-analysis. *Computers in Human Behavior Reports*, 7:100206, 2022.

S. Yamaguchi. Effects of Parental Leave Policies on Female Career and Fertility Choices. *Quantitative Economics*, 10:1195–1232, 2019.

A Data for model estimation

Building on **Model 1**, we estimate lifecycle profiles from age 22 to 45 for seven moment groups: adult consumption, child consumption, number of children, household labor supply, total household assets, parental childcare time, and external childcare time. These seven groups together yield 124 targeted moments. The 124 moments comprise age-indexed profiles for six of the seven groups (adult consumption, child consumption, number of children, household labor supply, total household assets, and parental childcare time) over the relevant lifecycle windows, plus a single estimated target for external childcare time. The per-group moment counts and age windows are detailed in Sections A.1.1 and A.1.2. The two childcaring profiles enter through the model's time-allocation equations (9): parental childcare time is matched to its data age profile, while external childcare time is matched to an estimated target. The household head's hourly wage and the spouse's labor income enter Model 1 as exogenous inputs rather than estimation targets.

Model 2 targets a richer moment set that exploits the woman-level cross-section: 292 moments organized into 13 blocks. Twelve blocks are indexed by age 22 to 45 (24 ages each); the thirteenth is the pooled mean of labor supply by capped number of children (4 cells, $n_{\text{cap}} \in \{0, 1, 2, 3\}$). The age-indexed blocks cover mean labor supply, mean number of children, and mean parental time; the employment composition (full-time and part-time rates, cross-sectional within-age-group variance of labor supply); the fertility margin (fraction childless, birth probability, covariance of labor supply and number of children); the marital-fertility joint distribution (shares of single-without-child and single-with-child); and the mean log hourly wage.

Both models draw on the same underlying surveys: PSID, ATUS, CPS, and SCF for the United States; SOEP, Pairfam, and HFCS for Germany[55]. Estimation samples, however, differ:

- **Model 1 sample** (couples-only): Because Model 1 is a household-level model, all variables are at the *household-head* level: the wage and labor-supply profiles target the head's variables, while the partner-earning profile targets the spouse's labor income. The household head is thus the empirical counterpart of Model 1's single decision-maker.

- **Model 2 sample** (woman-level, pooling singles and women in couples): Because Model 2 takes the perspective of women, we exclude all men and, instead, include

[55]Throughout this appendix, unless otherwise noted, US dollar amounts are deflated to 2011 prices and German euro amounts to 2010 prices, and all reported means are survey-weighted. In all profile figures, shaded bands are 95% bootstrap confidence intervals ($B = 300$ draws).

unpartnered and partnered females. It is these women's characteristics that enter the moment computations, grounding Model 2's profiles on the woman's lifecycle.

The remainder of this appendix is organized model-by-model and country-by-country. Sections A.1.1 and A.1.2 describe the lifecycle data underlying the couples-only Model 1 estimation for the United States and Germany, summarized in Figures A.7 and A.14. Sections A.2.1 and A.2.2 describe the woman-level data underlying the Model 2 estimation, summarized in Figures A.28 and A.41.

A.1 Model 1

A.1.1 Model 1 for the United States

Our data infrastructure comprises the Panel Study of Income Dynamics (PSID) for the lifecycle panel, the American Time Use Survey (ATUS) for childcare time, and the Survey of Consumer Finances (SCF) for household net wealth.

The PSID is a longitudinal survey that follows a nationally representative random household sample. It details economic and demographic information annually from 1968 to 1997 and biannually after 1997. Most of our variables are constructed from PSID data from 1999 to 2013, as the time window before our vignette surveys were fielded. ATUS provides nationally representative data of how, where, and with whom Americans spend their time. It is the only federal survey providing micro-level data on the full range of non-market activities, from childcare to volunteering. We use the 2014 ATUS to compute parental childcare time of the household head. The SCF is a cross-sectional household survey on families' balance sheets, pensions, income, and demographic characteristics. We use SCF to derive households' wealth accumulation profiles.

For all three datasets, we focus on couples with the household head aged 22 to 45.

Consumption

We construct the consumption-age profile from PSID's consumption modules from 2005 to 2013. Total household consumption is the sum across all PSID consumption subcategories. To adjust for differences in needs across household types, we apply the modified OECD equivalence scale,

$$S_{\text{OECD}} = 1 + 0.5\,(n_a - 1) + 0.3\,n_c,$$

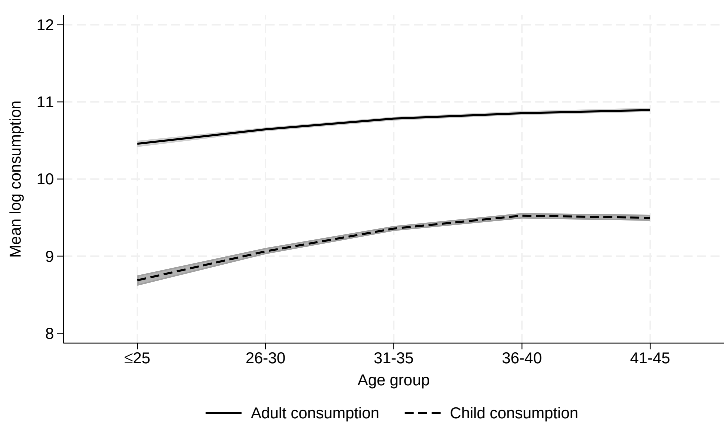
in which the first adult counts as one equivalence unit, each additional adult as 0.5, and each child as 0.3. Children's consumption is total household consumption multiplied by the children's share of these equivalence units; adult consumption takes the remainder. Table A.1 reports descriptive statistics for the three consumption variables on the pooled PSID 2005–2013 sample (one observation per household-year). Figure A.1 shows the age profiles of adult and child consumption.

Table A.1: Consumption

Variable	Age group	N	Mean	SD	p25	p50	p75
Total household consumption	≤25	801	42,452	19,661	29,452	38,469	51,103
	26–30	2,048	54,123	27,026	36,191	49,129	65,090
	31–35	2,324	65,849	37,534	43,689	57,642	78,110
	36–40	2,051	73,089	47,731	46,485	63,200	86,116
	41–45	2,022	75,139	61,268	47,720	64,660	86,705
Adult consumption (OECD split)	≤25	801	38,589	19,046	25,299	34,581	47,199
	26–30	2,048	47,264	25,064	30,363	42,620	57,405
	31–35	2,324	54,872	32,040	35,412	47,530	66,013
	36–40	2,051	59,200	38,282	38,334	50,569	70,111
	41–45	2,022	61,916	47,760	39,974	54,040	72,567
Child consumption (OECD split)	≤25	801	3,863	4,594	0	2,875	6,374
	26–30	2,048	6,859	7,296	0	5,525	11,032
	31–35	2,324	10,977	10,483	3,797	9,402	15,351
	36–40	2,051	13,889	13,017	5,877	11,929	18,947
	41–45	2,022	13,223	16,508	3,771	10,465	18,307

Note: Model 1, United States. Total household, adult, and child consumption in 2011 USD; survey-weighted. Pooled PSID 2005–2013.

Figure A.1: Age profiles of adult and child consumption



Note: Model 1, United States. Lifecycle profile of mean log adult and child consumption; survey-weighted.

Number of children

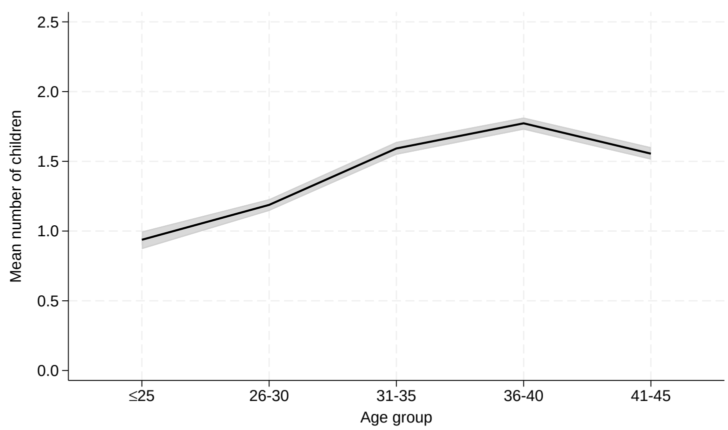
In PSID, children are household members under 18. Table A.2 reports descriptive statistics for the per-household number of children by age group. Figure A.2 shows the age profile of the per-household number of children.

Table A.2: Number of children

Variable	Age group	N	Mean	SD	p25	p50	p75
	≤25	1,454	0.940	0.990	0.000	1.000	2.000
	26–30	3,409	1.190	1.150	0.000	1.000	2.000
Number of children (under 18)	31–35	3,818	1.590	1.190	1.000	2.000	2.000
	36–40	3,743	1.770	1.190	1.000	2.000	2.000
	41–45	3,913	1.560	1.190	1.000	2.000	2.000

Note: Model 1, United States. Per-household number of children; survey-weighted.

Figure A.2: Age profiles of number of children



Note: Model 1, United States. Lifecycle profile of the number of children per household; survey-weighted.

Partner's earnings and other income sources

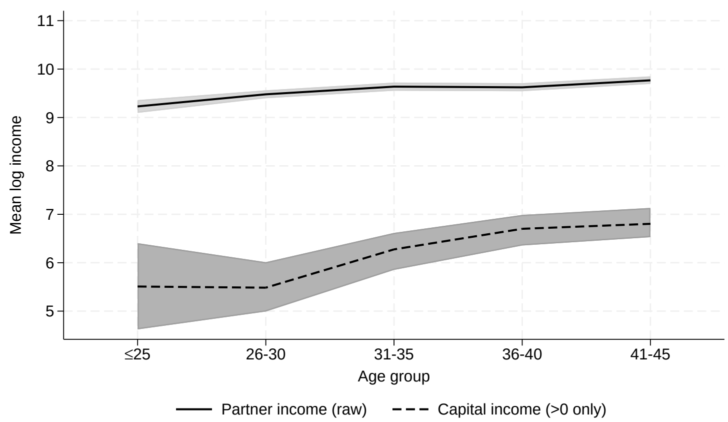
Partner's income, y_P, is defined as post-government household income minus capital income minus the head's net labor earnings; capital income is deducted because it already enters the model through the return on assets, ra_t, in the budget constraint. We smooth y_P across age via a panel fixed-effects regression on year dummies, the head's and partner's ages, education-level dummies for both partners, household size (number of adults and children), and the partner's monthly hours. Table A.3 reports descriptive statistics for the raw partner-income variable y_P by age group and for capital income y_{cap} pooled across ages 22–45. Figure A.3 shows the age profile of partner income and capital income.

Table A.3: Partner income and capital income

Variable	Age group	N	Mean	SD	p25	p50	p75
Partner income (raw)	≤25	1,389	16,399	13,433	5,062	14,300	25,021
	26–30	3,265	23,288	21,291	7,223	21,069	32,723
	31–35	3,688	29,176	29,899	7,242	23,991	39,576
	36–40	3,606	31,339	35,196	7,767	24,237	41,748
	41–45	3,744	32,905	36,511	9,746	24,477	43,081
Capital income (> 0 only)	All (22–45)	577	3,682	10,851	186	891	2,168

Note: Model 1, United States. Partner income (by age group) and capital income (> 0, pooled across ages 22–45), in 2011 USD; survey-weighted.

Figure A.3: Age profile of partner income and capital income



Note: Model 1, United States. Lifecycle profile of mean log partner income y_P (raw) and mean log capital income (conditional on $y_{\text{cap}} > 0$); survey-weighted. Windfall/inheritance income (gifts plus expected inheritances from SCF, discounted at 50%) is treated as a separate model input and does not enter this profile.

Net hourly wages and working hours

Our labor-market target variables for the household head are the annual hours of work l_t, expressed as a fraction of the maximum annual working time $24 \times 4.25 \times 5 \times 12 = 6{,}120$ hours, and the net hourly wage w_t. Annual hours and gross hourly wages are reported directly in PSID; net hourly wages are constructed as gross earnings times $(1 - \tau)$, where the household's income-tax rate τ comes from NBER TAXSIM (version 9; see Kimberlin, 2015).[56]

We estimate both lifecycle profiles via panel fixed-effects regressions: on the head's net hourly wage for w_t, and on the head's annual hours for l_t. Both regressions use the same controls: year dummies, the head's and spouse's ages, their squares, the number of adults, and

[56] The NBER's Internet TAXSIM version 9, http://users.nber.org/~taxsim/taxsim9/.

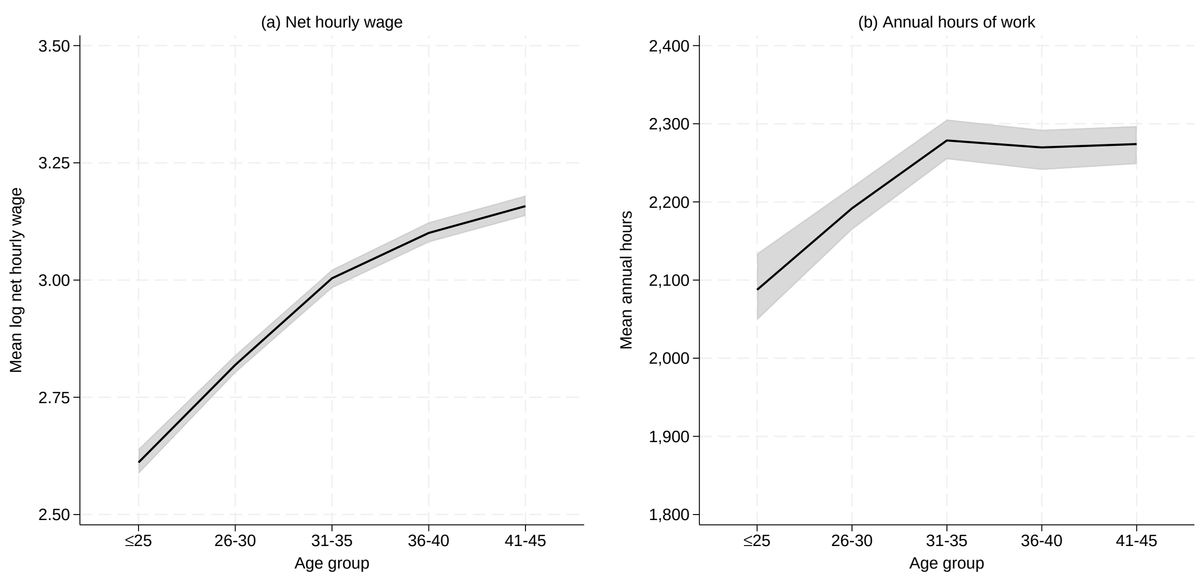
the number of children. Predicted values are collapsed by the head's age. Table A.4 reports descriptive statistics for the net hourly wage and the annual hours of work, restricted to working heads (consistent with the working sample used in the wage regression). Figure A.4 shows the age profiles of the net hourly wage and annual hours of work.

Table A.4: Net hourly wages and annual hours

Variable	Age group	N	Mean	SD	p25	p50	p75
Net hourly wages (USD/hr)	≤25	1,389	14.950	7.490	10.250	13.680	17.430
	26–30	3,265	18.850	10.070	12.270	16.620	22.400
	31–35	3,688	22.900	12.030	14.410	20.130	28.000
	36–40	3,606	25.470	13.270	15.770	22.360	32.810
	41–45	3,744	27.110	14.110	16.400	23.780	35.830
Annual hours of work	≤25	1,389	2,088	705	1,828	2,080	2,448
	26–30	3,265	2,192	672	1,960	2,147	2,500
	31–35	3,688	2,279	648	2,000	2,232	2,568
	36–40	3,606	2,270	625	1,968	2,200	2,558
	41–45	3,744	2,274	647	1,976	2,205	2,586

Note: Model 1, United States. Net hourly wages (2011 USD) and annual hours of work; working heads only; survey-weighted.

Figure A.4: Age profile of net hourly wages and hours



Note: Model 1, United States. Lifecycle profiles of (a) mean log net hourly wage and (b) mean annual hours of work; survey-weighted. Working heads only.

Net wealth

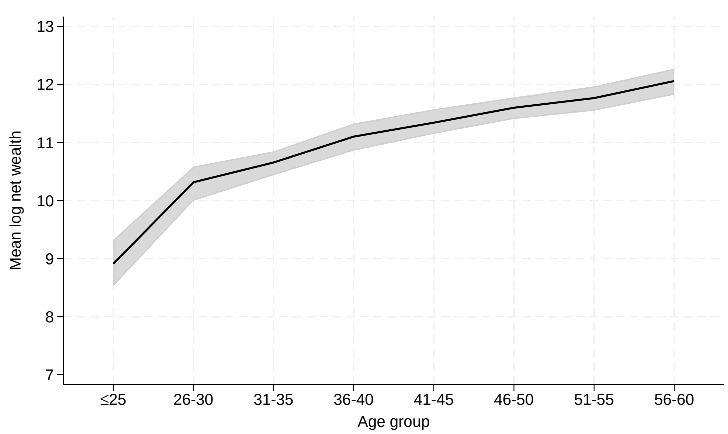
The Survey of Consumer Finances (SCF) 2007 wave records detailed household-level asset information. We construct net wealth as the sum of all financial and non-financial assets — stocks, business equity, bonds, savings/checking accounts, retirement accounts, life-insurance cash value, real estate, vehicles, and personal-use assets — net of liabilities. The moment block uses eight five-year age bins (≤ 25, 26–30, …, 56–60); the bins extend beyond the model's 22–45 horizon to anchor terminal wealth. Table A.5 reports descriptive statistics for net wealth on the SCF couples sample, restricted to the model's 22–45 horizon. Figure A.5 shows the age profile of net wealth.

Table A.5: Net wealth

Variable	Age group	N	Mean	SD	p25	p50	p75
Net wealth	≤25	162	24,671	165,172	-2,600	2,580	11,260
	26–30	249	88,853	259,851	252	13,812	81,750
	31–35	309	121,426	259,975	5,700	27,830	135,206
	36–40	378	192,509	347,782	8,860	73,690	226,204
	41–45	460	249,479	432,282	9,760	82,030	280,720

Note: Model 1, United States. Net wealth in nominal USD (SCF 2007); survey-weighted.

Figure A.5: Age profile of net wealth



Note: Model 1, United States. Lifecycle profile of mean log net wealth (conditional on NWorth > 0) across the full SCF range 22–60; survey-weighted. Defined-benefit pension wealth is excluded to keep the asset definition comparable across the United States and Germany. Net wealth is averaged over the SCF's five implicates per household and symmetrically winsorized at the 2nd and 98th percentiles.

Parental childcare time (household head)

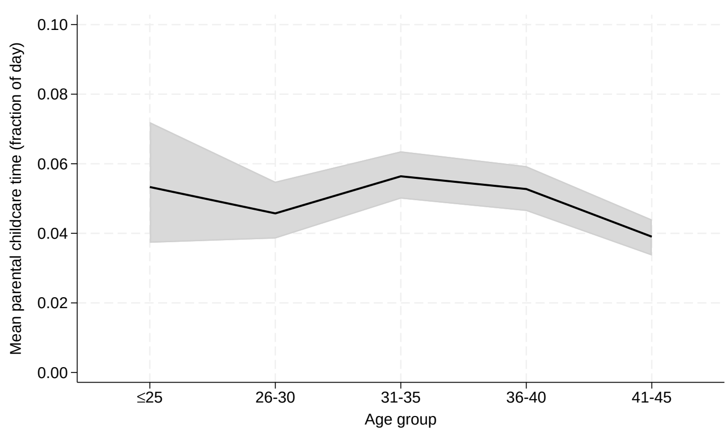
The 2014 American Time Use Survey records minute-level diary data on parental childcare activities, classified into 20 primary-childcare codes covering physical care, educational activities (reading, homework assistance), and child-health-related care. We sum across these 20 codes at the respondent level and express the daily total as a fraction of a 24-hour day. Restricting to household heads in couples, we compute the age-specific mean. Table A.6 reports descriptive statistics for the head's parental childcare time. Figure A.6 shows the age profile of parental childcare time.

Table A.6: Parental childcare time

Variable	Age group	N	Mean	SD	p25	p50	p75
Parental childcare time (fraction of day)	≤25	147	0.053	0.091	0.000	0.000	0.082
	26–30	485	0.046	0.078	0.000	0.000	0.062
	31–35	727	0.056	0.077	0.000	0.024	0.083
	36–40	743	0.053	0.080	0.000	0.021	0.080
	41–45	707	0.039	0.066	0.000	0.007	0.059

Note: Model 1, United States. Household head's parental childcare time, as a fraction of a 24-hour day; survey-weighted.

Figure A.6: Age profile of parental childcare time



Note: Model 1, United States. Lifecycle profile of mean parental childcare time of the household head (fraction of a 24-hour day); survey-weighted.

External childcare time

External childcare time is not directly recorded in PSID or other US micro datasets. We therefore infer it from the average childcare expenditure and the market price of external care.

The composition of childcare cost depends on children's age; in general, it includes the cost of after-school sitters, au pairs, childcare centers, family childcare centers, and nannies. By using the "Family Budget Calculator" provided by the Economic Policy Institute,[57] we can calculate the estimated childcare cost. The average childcare cost in the US is quite heterogeneous across different regions. For example, a household with 2 adults and 1 child will spend around \$1,472 on childcare in Washington, D.C., but only \$390 in rural Arkansas. We have set the childcare cost equal to \$550 per child per month. Using the average hourly cost of external childcare at \$8/hour, the calculated childcare time would be \$550/\$8 = 68.75 hours per month per child, and then divided by $7 \times 24 \times 4$ (monthly available time), approximately 10% of childcare time is allocated through external childcare resources. This is consistent with observed conditions in the United States.[58] Each worker typically cares for a small group of children, implying an effective labor cost of \$3–4 per child per hour, and including an overhead for facilities and administration, an effective price in the \$7–\$12 per hour range captures a realistic level faced by families when purchasing childcare. In Model 1, the structural childcare price is set at \$8 per hour for the United States, consistent with this range.

Summary of data moments

Table A.7 summarizes the nine data moments used in the US estimation. Each variable is reported by age group; the upper row of each cell is the anchor mean ($b = 0$) and the lower row in parentheses is the 95% bootstrap CI. Throughout, $b = 0$ denotes the anchor (un-resampled) draw and $b = 1, \ldots, B$ are bootstrap replications.

The combined age profiles for the seven US data moments described above are summarized in Figure A.7. Each panel plots one of the seven lifecycle profiles at the single-age resolution used in estimation, complementing the age-group summary table above.

[57] See www.epi.org/resources/budget/, and for details of the Family Budget Calculator, see Gould et al. [2015].

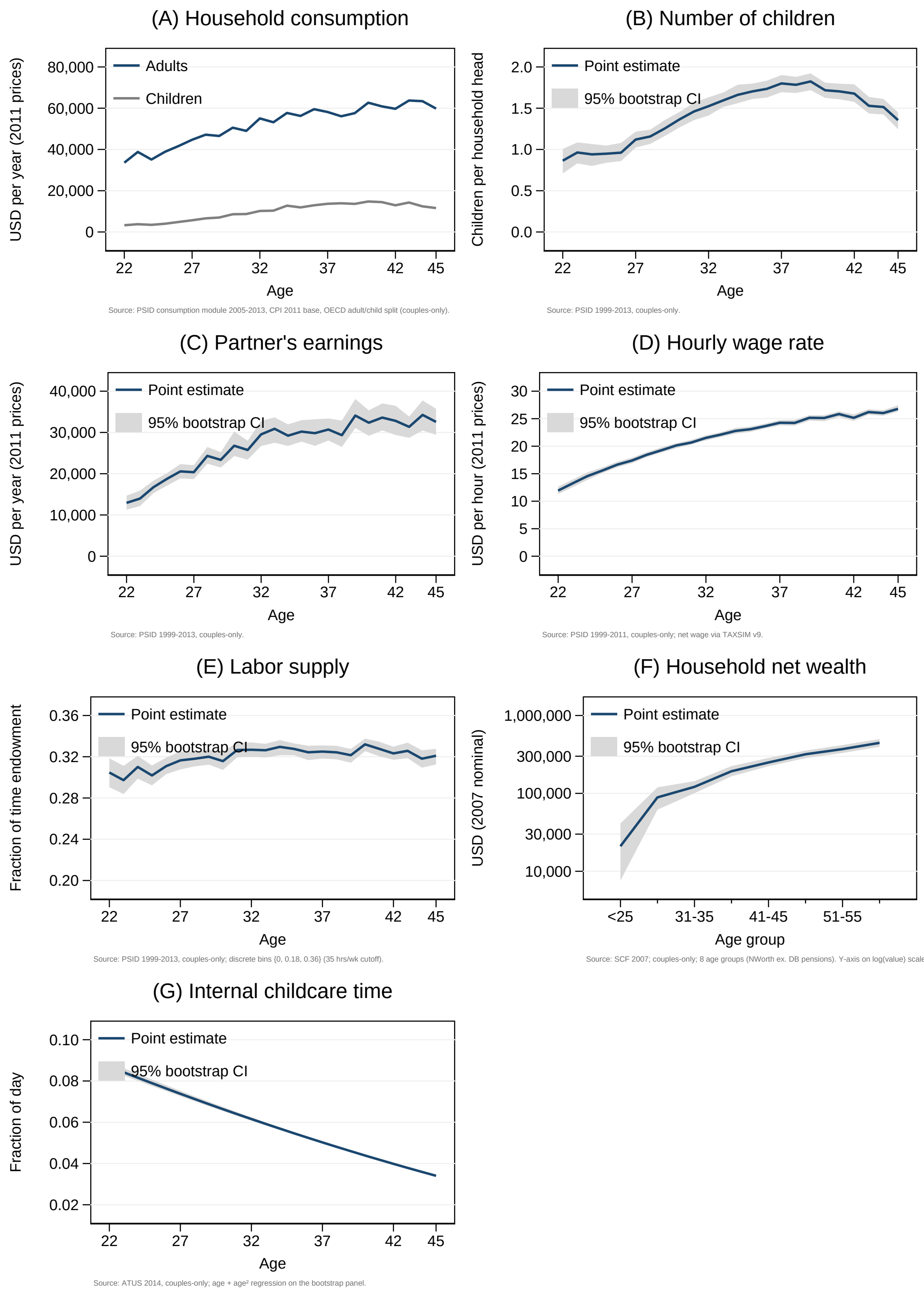
[58] Childcare workers earned a mean annual wage of approximately \$22,310 (median \$20,320) in 2015 according to the Bureau of Labor Statistics Occupational Employment Statistics survey: U.S. Bureau of Labor Statistics, *May 2015 Occupational Employment Statistics*, "39-9011 Childcare Workers," available at https://www.bls.gov/oes/2015/may/oes399011.htm.

Table A.7: Data moments used in estimation

Variable	Age group ≤ 25	26–30	31–35	36–40	41–45
Adult consumption (USD)	38,589	47,264	54,872	59,200	61,916
	(36,942, 40,145)	(45,989, 48,569)	(53,288, 56,372)	(57,325, 61,107)	(59,396, 64,561)
Child consumption (USD)	3,863	6,859	10,977	13,889	13,223
	(3,446, 4,294)	(6,470, 7,286)	(10,472, 11,530)	(13,277, 14,512)	(12,436, 14,096)
Number of children	0.940	1.190	1.590	1.770	1.560
	(0.870, 1.000)	(1.140, 1.230)	(1.550, 1.640)	(1.730, 1.810)	(1.510, 1.600)
Partner income, raw (USD)	16,399	23,288	29,176	31,339	32,905
	(15,471, 17,396)	(22,506, 24,260)	(28,055, 30,314)	(30,094, 33,029)	(31,497, 34,356)
Capital income, positive only (USD)	866	1,834	2,975	4,125	4,588
	(265, 1,489)	(664, 3,218)	(1,427, 4,811)	(2,316, 6,546)	(3,182, 6,234)
Net hourly wages (USD/hr)	14.95	18.85	22.90	25.47	27.11
	(14.50, 15.49)	(18.48, 19.29)	(22.45, 23.33)	(24.96, 25.99)	(26.57, 27.68)
Annual hours of work	2,088	2,192	2,279	2,270	2,274
	(2,049, 2,134)	(2,164, 2,219)	(2,255, 2,305)	(2,241, 2,292)	(2,248, 2,297)
Net wealth (USD)	24,671	88,853	121,426	192,509	249,479
	(7,744, 55,258)	(61,773, 122,610)	(102,069, 140,958)	(163,883, 220,458)	(212,301, 291,646)
Parental childcare time (fraction of day)	0.053	0.046	0.056	0.053	0.039
	(0.037, 0.072)	(0.038, 0.055)	(0.050, 0.064)	(0.046, 0.059)	(0.034, 0.044)

Note: Model 1, United States. Anchor mean ($b = 0$, upper row) and 95% bootstrap CI (lower row); 300 bootstrap draws; survey-weighted.

Figure A.7: Data-moment age profiles



Note: Model 1, United States. Age profiles of data moments; survey-weighted.

A.1.2 Model 1 for Germany

Our data infrastructure comprises the German Socio-Economic Panel (SOEP) for the lifecycle panel, the German Family Panel (Pairfam) for parental childcare time, and the German subset of the Eurosystem Household Finance and Consumption Survey (HFCS) for the household-wealth robustness profile. As in the United States, the Model 1 sample is restricted to couples with the household head aged 22–45.

Our primary dataset is the SOEP[59], a representative panel of the population living in Germany funded by the Leibniz Association. The data is collected annually. Every year since 1984, about 20,000 to 30,000 persons in about 15,000 households are interviewed. Most importantly for our purposes, the SOEP includes detailed information on individual- and household-level socio-demographics, i.e., family composition, labor-market characteristics, income and wealth (accumulation). For detailed information on the SOEP data, see Goebel et al. [2019]. Our data encompasses the 1984 to 2013 period.

We also use the German Family Panel (Pairfam) dataset, a multi-disciplinary household panel focusing on partnership and family dynamics, funded by the German Research Foundation. The data is collected annually, with the fieldwork and data collection conducted by TNS Infratest Social Research in Munich. The first data wave was collected in 2008/9. Pairfam surveys three birth cohorts (1971-73, 1981-83, and 1991-93). The core modules of Pairfam focus on the formation and the development of partnerships, processes of starting and expanding families, parenting and child development, and intergenerational relationships. We use the 5th wave, which was collected in 2013/2014.

The third dataset we use is the German subset of the Eurosystem Household Finance and Consumption Survey (HFCS). HFCS is a representative household survey focusing on income and wealth information. Most importantly for our purposes, it covers the balance sheets, savings, incomes, work histories, and demographic characteristics of more than 3,000 households living in Germany. The first HFCS wave was collected in 2010/11, released in 2013. For detailed information on the HFCS data, see Household Finance and Consumption Network [2013].

Consumption

SOEP household consumption is the difference between household post-government income and household savings. Adult and child consumption are then derived from the OECD equivalence scale (Section A.1.1).

[59] We use SOEP data release v30.

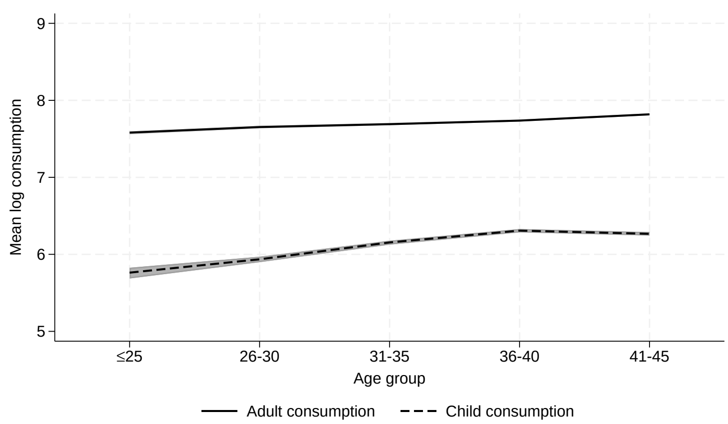
Table A.8 reports descriptive statistics on the SOEP working sample. Values are monthly EUR in 2010 prices. Figure A.8 shows the age profiles of adult and child consumption.

Table A.8: Consumption

Variable	Age group	N	Mean	SD	p25	p50	p75
Total household consumption	≤25	2,214	2,137	555	1,739	2,081	2,466
	26–30	6,755	2,398	673	1,940	2,307	2,734
	31–35	10,433	2,657	788	2,139	2,527	3,030
	36–40	12,632	2,896	909	2,293	2,749	3,278
	41–45	13,556	3,053	1,004	2,386	2,875	3,493
Adult consumption (OECD split)	≤25	2,214	2,026	542	1,633	1,953	2,378
	26–30	6,755	2,198	670	1,725	2,079	2,544
	31–35	10,433	2,294	756	1,783	2,135	2,633
	36–40	12,632	2,402	798	1,877	2,242	2,744
	41–45	13,556	2,614	896	2,020	2,444	2,988
Child consumption (OECD split)	≤25	2,214	111	188	0	0	226
	26–30	6,755	200	257	0	0	335
	31–35	10,433	363	322	0	336	571
	36–40	12,632	494	370	265	484	701
	41–45	13,556	439	392	0	380	670

Note: Model 1, Germany. Household, adult, and child consumption in 2010 EUR (monthly); survey-weighted.

Figure A.8: Age profiles of adult and child consumption



Note: Model 1, Germany. Lifecycle profile of mean log adult and child consumption; survey-weighted. SOEP does not measure consumption directly; reported savings is left-censored by survey design (the SOEP questionnaire only asks for positive monthly savings amounts), so Tobit-imputed savings are used for non-reporters. The SOEP consumption proxy is a monthly-income-based measure, unlike the annual PSID consumption expenditure in the United States.

Number of children

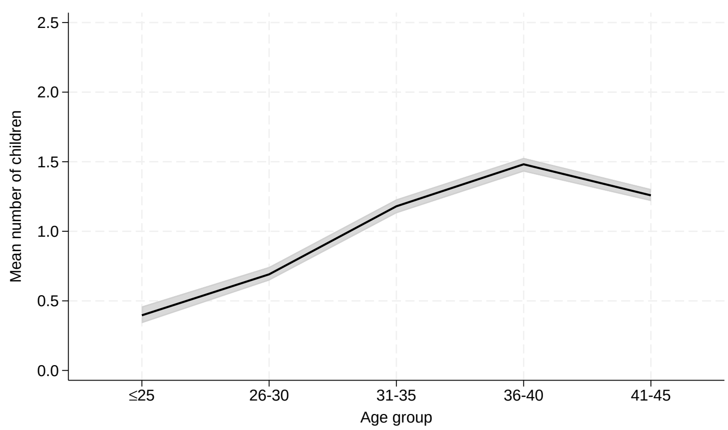
Children are SOEP household members under 18. Table A.9 reports the descriptive statistics by age group. Figure A.9 shows the age profile of the per-household number of children.

Table A.9: Number of children

Variable	Age group	N	Mean	SD	p25	p50	p75
Number of children (under 18)	≤25	2,344	0.400	0.660	0.000	0.000	1.000
	26–30	6,898	0.690	0.860	0.000	0.000	1.000
	31–35	10,519	1.180	1.000	0.000	1.000	2.000
	36–40	12,681	1.480	1.030	1.000	2.000	2.000
	41–45	13,608	1.260	1.040	0.000	1.000	2.000

Note: Model 1, Germany. Per-household number of children; survey-weighted.

Figure A.9: Age profiles of number of children



Note: Model 1, Germany. Lifecycle profile of the number of children per household; survey-weighted.

Partner's earnings and other income sources

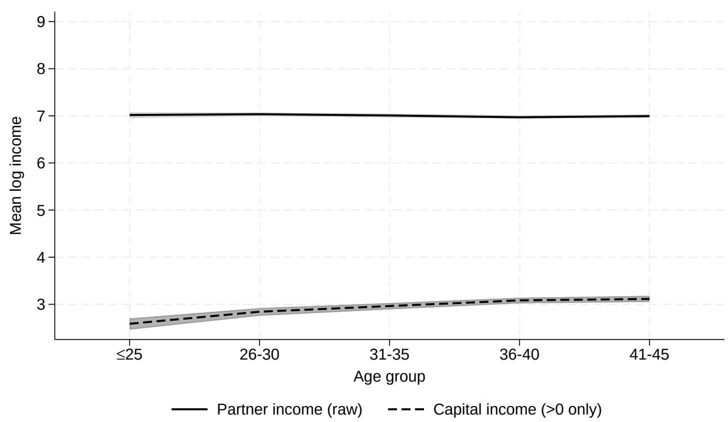
Partner's income y_P is defined exactly as in the US case (Section A.1.1): post-government household income minus capital income minus the head's net labor earnings, with capital income removed because it enters the model through the return on assets. Table A.10 reports descriptive statistics for the raw partner-income series by age group and for capital income conditional on $y_{\text{cap}} > 0$, pooled across ages 22–45. Figure A.10 shows the age profile of the SOEP-based partner income and capital income.

Table A.10: Partner income and capital income

Variable	Age group	N	Mean	SD	p25	p50	p75
	≤25	2,344	1,277	760	761	1,260	1,703
	26–30	6,898	1,359	877	702	1,306	1,890
Partner income (raw)	31–35	10,519	1,396	1,015	583	1,285	2,020
	36–40	12,681	1,372	1,097	548	1,171	1,979
	41–45	13,608	1,415	1,127	586	1,202	1,989
Capital income (> 0 only)	All (22–45)	40,588	54	124	9	19	57

Note: Model 1, Germany. Partner income (by age group) and capital income (> 0, pooled across ages 22–45), in 2010 EUR (monthly); survey-weighted.

Figure A.10: Age profile of partner income and capital income



Note: Model 1, Germany. Lifecycle profile of mean log partner income y_P (raw) and mean log capital income (conditional on $y_{\text{cap}} > 0$); survey-weighted.

Net hourly wages and working hours

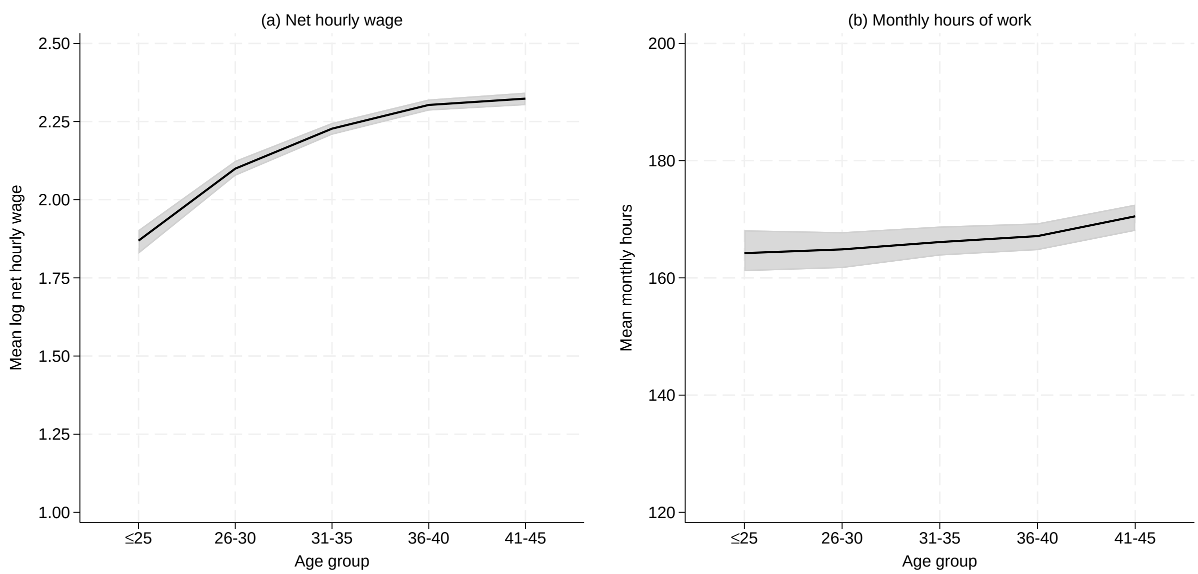
Our labor-market target variables for the household head are the monthly hours of work l_t, expressed as a fraction of the maximum monthly working time $24 \times 4.25 \times 5 = 510$ hours, and the net hourly wage w_t. Both are constructed from SOEP household heads aged 18–60 over 1995–2013. We estimate the two lifecycle profiles via panel fixed-effects regressions: on the head's net hourly wage for w_t (working observations only) and on the head's monthly hours for l_t. Both regressions use the same controls: year dummies, the head's and spouse's ages, their squares, the number of adults, and the number of children. Predicted values are collapsed by the head's age. Table A.11 reports descriptive statistics for the net hourly wage and the monthly hours of work, restricted to working heads. Figure A.11 shows the age profiles of the net hourly wage and monthly hours of work.

Table A.11: Net hourly wages and monthly hours

Variable	Age group	N	Mean	SD	p25	p50	p75
Net hourly wages (EUR/hr)	≤25	1,852	7.100	3.060	5.260	6.890	8.380
	26–30	5,662	8.830	3.450	6.680	8.400	10.430
	31–35	8,969	10.080	3.930	7.340	9.670	12.120
	36–40	11,330	10.950	4.360	7.760	10.400	13.640
	41–45	12,469	11.250	4.630	7.860	10.620	14.040
Monthly hours of work	≤25	1,852	164	50	162	170	183
	26–30	5,662	165	54	159	170	191
	31–35	8,969	166	56	157	170	191
	36–40	11,330	167	58	157	170	191
	41–45	12,469	171	56	159	170	196

Note: Model 1, Germany. Net hourly wages (2010 EUR per hour) and monthly hours of work; working heads only; survey-weighted.

Figure A.11: Age profile of net hourly wages and hours



Note: Model 1, Germany. Lifecycle profiles of (a) mean log net hourly wage and (b) mean monthly hours of work; survey-weighted. Working heads only.

Net wealth

The primary wealth source is the SOEP wealth module, implemented in 2002, 2007, and 2012. Household net wealth is deflated to 2012 prices. Because only three waves are available, we forgo fixed-effects smoothing and report pooled age-specific means.[60] The age bins follow the wealth-moment convention: eight five-year groups spanning ages 22–60. As a robustness

[60] Averages over all imputations and imputation-specific averages are very similar. This similarity is not ensured for other measures such as median or variance.

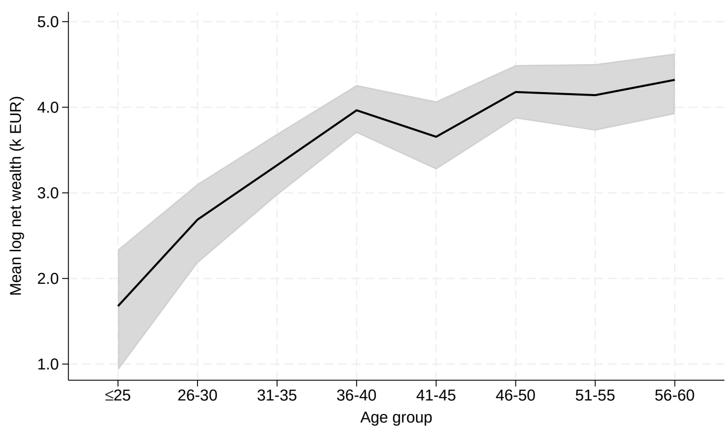
check, we re-derive the profile from the HFCS using the same procedures. Table A.12 reports descriptive statistics from the HFCS sample (thousand EUR); the SOEP-based wealth profile enters the estimation moments. Figure A.12 shows the age profile of net wealth, extending to age 60 to match the wealth-moment age-bin convention.

Table A.12: Net wealth

Variable	Age group	N	Mean	SD	p25	p50	p75
Net wealth (HFCS, k EUR)	≤25	113	28	79	0	3	16
	26–30	174	39	79	1	11	33
	31–35	182	64	110	4	30	80
	36–40	247	135	197	24	45	162
	41–45	317	149	216	5	66	212

Note: Model 1, Germany. Net wealth from the HFCS sample (thousand EUR); survey-weighted.

Figure A.12: Age profile of net wealth



Note: Model 1, Germany. Lifecycle profile of mean log net wealth (HFCS, conditional on NW > 0, in thousand EUR) across the full HFCS range 22–60; survey-weighted. SOEP net wealth is averaged across multiple imputations and symmetrically winsorized at the 2nd and 98th percentiles within each wave.

Parental childcare time (household head)

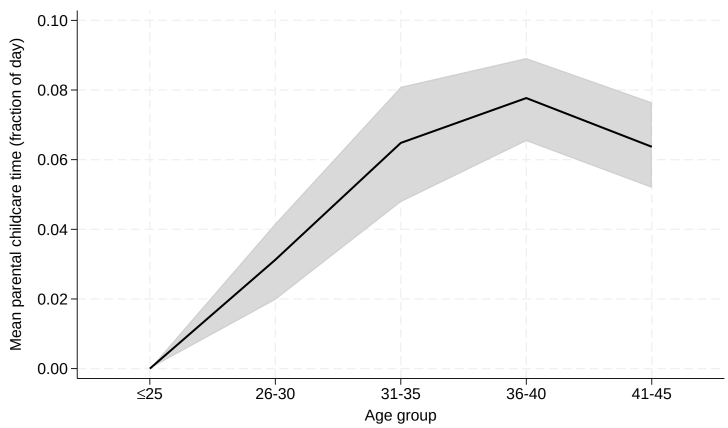
Pairfam Wave 5 records the time the household head (anchor person) spends with children in the morning and afternoon. We sum across the morning and afternoon components at the respondent level to construct m_{head}, the head's daily childcare time as a fraction of the day, and fit a quadratic in age, $m_{\text{head}} = \beta_0 + \beta_1 \,\text{age} + \beta_2 \,\text{age}^2$. The fitted curve is then evaluated on the 22–45 grid; predictions at ages without Pairfam observations are quadratic extrapolations. Table A.13 reports unweighted descriptive statistics for the raw m_{head} data underlying the quadratic fit. Figure A.13 shows the lifecycle profile of parental childcare time by age group.

Table A.13: Parental childcare time

Variable	Age group	N	Mean	SD	p25	p50	p75
Parental childcare time (fraction of day)	26–30	448	0.031	0.058	0.000	0.000	0.056
	31–35	216	0.065	0.072	0.000	0.056	0.111
	36–40	532	0.078	0.065	0.000	0.056	0.111
	41–45	300	0.064	0.056	0.000	0.056	0.111

Note: Model 1, Germany. Household head's parental childcare time, as a fraction of a day (Pairfam Wave 5 raw data); unweighted.

Figure A.13: Age profile of parental childcare time



Note: Model 1, Germany. Lifecycle profile of mean parental childcare time of the household head (m_{head}, fraction of a day); unweighted. Pairfam Wave 5 raw data; the model moment uses a quadratic fit to the same data.

External childcare time

Since 2009, SOEP contains information on how many hours per day a child attends an institution. The institutions are preschools, crèches, and nurseries (including after-school nurseries). The weighted average of total external childcare time in 2009 is 2.15 hours per day.

Comparable salaries are observed in Germany for early-childhood educators (*Erzieher*) according to the Federal Statistical Office's *Bildungsfinanzbericht*.[61] As in the United States,

[61]Statistisches Bundesamt (Federal Statistical Office of Germany), *Bildungsfinanzbericht 2015*, Wiesbaden, 2015.

the small worker-to-child ratio and facility overhead imply an effective hourly price on the same order of magnitude. In Model 1, the structural childcare price is set at €7 per hour for Germany, consistent with this range.

Summary of data moments

Table A.14 summarizes the nine data moments used in the Germany estimation. Each variable is reported by age group; the upper row of each cell is the anchor mean ($b = 0$) and the lower row in parentheses is the 95% bootstrap CI.

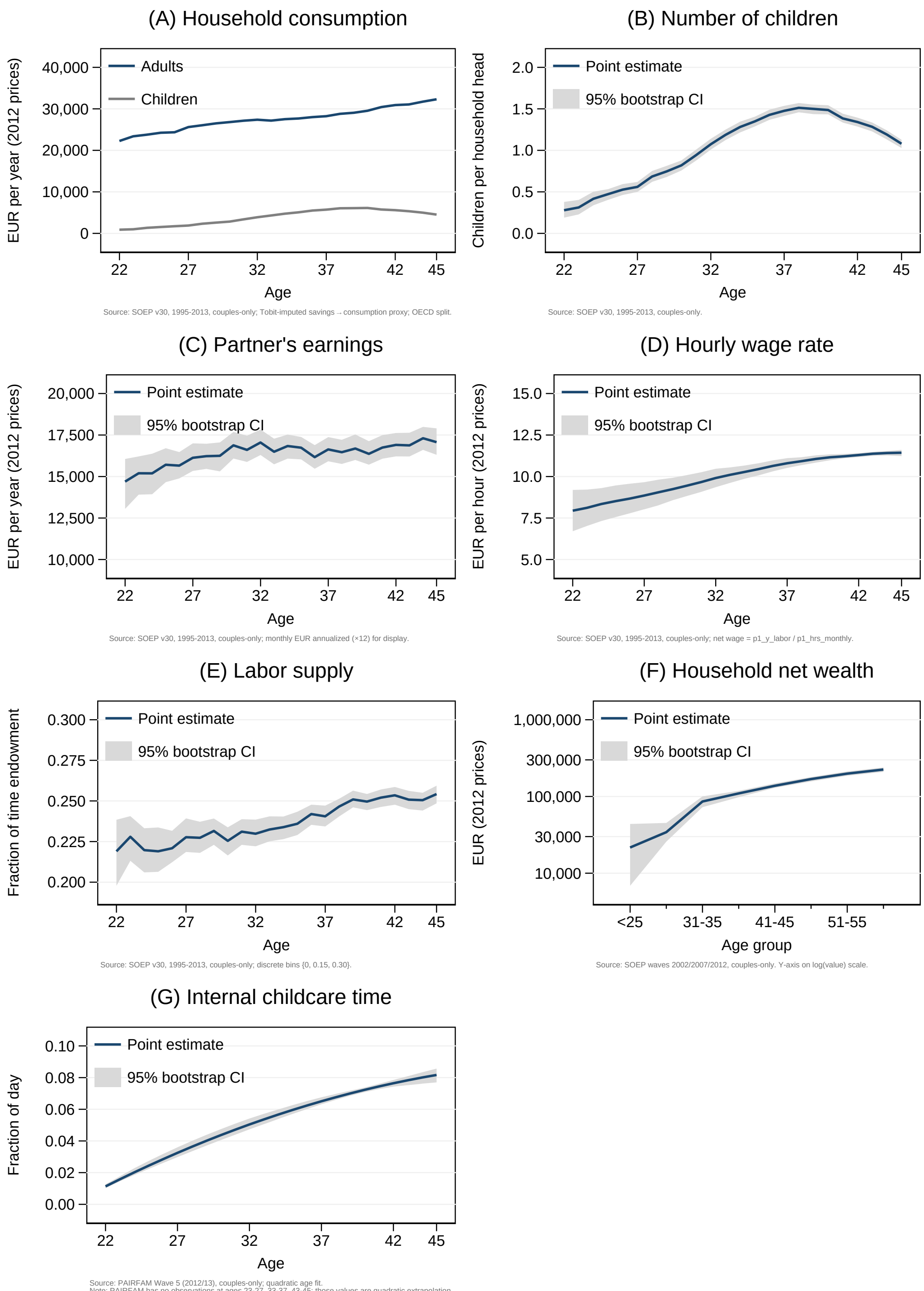
Table A.14: Data moments used in estimation

	Age group				
Variable	≤ 25	26–30	31–35	36–40	41–45
Adult consumption (EUR/month)	2,026	2,198	2,294	2,402	2,614
	(1,976, 2,066)	(2,157, 2,241)	(2,256, 2,335)	(2,369, 2,442)	(2,574, 2,655)
Child consumption (EUR/month)	111	200	363	494	439
	(93, 130)	(185, 219)	(348, 380)	(477, 509)	(423, 456)
Number of children	0.400	0.690	1.180	1.480	1.260
	(0.340, 0.460)	(0.650, 0.740)	(1.130, 1.230)	(1.430, 1.530)	(1.220, 1.300)
Partner income, raw (EUR/month)	1,277	1,359	1,396	1,372	1,415
	(1,202, 1,339)	(1,313, 1,401)	(1,353, 1,445)	(1,334, 1,422)	(1,369, 1,463)
Capital income, positive only (EUR/month)	33	43	51	58	63
	(28, 38)	(39, 47)	(46, 55)	(54, 63)	(58, 68)
Net hourly wages (EUR/hr)	7.10	8.83	10.08	10.95	11.25
	(6.82, 7.34)	(8.64, 9.04)	(9.91, 10.28)	(10.77, 11.13)	(11.05, 11.45)
Monthly hours of work	164	165	166	167	171
	(161, 168)	(162, 168)	(164, 169)	(165, 169)	(168, 173)
Net wealth (HFCS, k EUR)	28	39	64	135	149
	(14, 45)	(22, 66)	(51, 77)	(108, 162)	(124, 180)
Parental childcare time (fraction of day)	0.000	0.031	0.065	0.078	0.064
	(0.000, 0.000)	(0.020, 0.042)	(0.048, 0.081)	(0.065, 0.089)	(0.052, 0.076)

Note: Model 1, Germany. Anchor mean ($b = 0$, upper row) and 95% bootstrap CI (lower row); 300 bootstrap draws; survey-weighted.

The combined age profiles for the seven German data moments described above are summarized in Figure A.14. Each panel plots one of the seven lifecycle profiles at the single-age resolution used in estimation, complementing the age-group summary table above.

Figure A.14: Data-moment age profiles



Note: Model 1, Germany. Age profiles of data moments; survey-weighted.

A.2 Model 2

A.2.1 Model 2 for the United States

The data underlying the Model 2 estimation for the US draws on three surveys: the Current Population Survey (CPS, IPUMS extract, 2000–2025) for labor supply, fertility status, and marital structure; the American Time Use Survey (ATUS, 2014) for parental childcare time; and the Panel Study of Income Dynamics (PSID, 1999–2013) for the mean log hourly wage. Together these surveys yield 292 moments across 13 blocks (12 age-indexed blocks $\times$ 24 ages + 4 cells for the pooled "mean labor supply by number of children" block).

Sample. For each survey we restrict to women aged 22–45, pooling single women and women in couples. Labor supply is discretized to $\{0, 0.18, 0.36\}$ (non-working, part-time at 1–34 weekly hours, full-time at $\geq$ 35 weekly hours); number of children is capped at three; "married" is defined as currently married with a spouse present or absent; "birth" in the current year is proxied by the presence of a child of her own, aged zero — i.e., the woman is a mother and her youngest child has not yet reached age one. All age-indexed moments are computed as means by age, with weights given by the survey's longitudinal weight times a per-draw bootstrap weight (300 draws). Per-draw moment vectors are written out and used downstream both for the *anchor model* (the $b = 0$ draw) and for the bootstrap covariance matrix used in the estimation.

External childcare price. The Model 2 household budget constraint embeds an annualized external childcare cost parameter equal to \$43,200 per year for the United States—the full-time annual earnings of one external childcare provider.[62]

Panel (A) — Mean labor supply

The CPS sample comprises roughly 1.46 million person-year observations of women aged 22–45 (IPUMS-CPS, 2000–2025). Following the model's three labor states $\{0, \ell^p, \ell^f\}$, we

[62] Model 1 and Model 2 report the structural childcare price in different units — hourly for Model 1, annualized for Model 2 — because the two models adopt different budget-constraint conventions. (i) *Time-base normalization* — Model 1 annualizes the hourly price against the full 5,760 clock-hours-per-year basis ($12 \times 20 \times 24$) used in its budget constraint, whereas Model 2 annualizes against 1,800 working hours per year (150×12 for the United States; 160×12 for Germany), reflecting the standard full-time work schedule of an external childcare provider. (ii) *Calibration targets* — Model 1's $\tilde{p}_b$ is calibrated jointly with the structural parameters to fit the cost-share moments described in Sections A.1.1 and A.1.2, while Model 2's annualized cost parameter (\$43,200 per year in the United States; €32,640 in Germany) is calibrated to PSID/SOEP childcare-cost moments through the budget-constraint term $n_t\,(1-s)\,p_{b,t}\,m_{b,t}$, with subsidy rate $s = 0$ at the structural baseline. After undoing the time-base difference, the two implied effective prices are within the same order of magnitude.

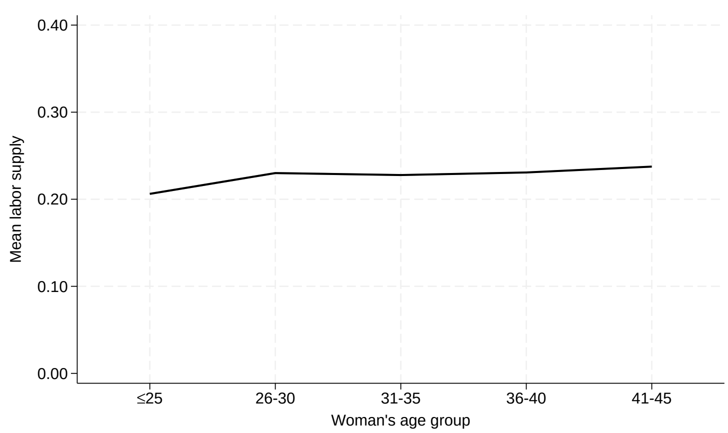
discretize observed labor supply to 0 (non-working), 0.18 (part-time, 1–34 weekly hours), and 0.36 (full-time, $\geq$ 35 weekly hours).[63] The resulting mean labor supply averages $\bar{\ell} = 0.227$ across ages, with the profile rising smoothly from 0.180 at age 22 to roughly 0.240 by age 45. A parallel calculation on the PSID under the same discretization yields $\bar{\ell} = 0.231$, a gap of only 1.8%. Table A.15 and Figure A.15 report the CPS-based mean labor supply by age; the PSID figure above is a robustness cross-check confirming the two surveys agree under the common discretization.

Table A.15: Mean labor supply

Variable	Age group	N	Mean	SD	p25	p50	p75
Mean labor supply	≤25	109,877	0.206	0.160	0.000	0.180	0.360
	26–30	152,244	0.230	0.161	0.000	0.360	0.360
	31–35	169,619	0.228	0.161	0.000	0.360	0.360
	36–40	179,939	0.231	0.160	0.000	0.360	0.360
	41–45	179,110	0.237	0.157	0.000	0.360	0.360

Note: Model 2, United States. Mean discretized labor supply; survey-weighted.

Figure A.15: Age profile of mean labor supply



Note: Model 2, United States. Mean labor supply; survey-weighted.

Panel (B) — Mean number of children

The mean number of children by age is constructed as a *synthetic 50/50 blend* of two independent sources, reconciling PSID's panel structure with the CPS's larger cross-section.[64]

[63]Concretely, conditional on the woman being employed (CPS variable `empstat` $\in \{10, 11, 12\}$), we classify her as non-working (`uhrsworkt` $= 0$), part-time ($1 \leq$ `uhrsworkt` ≤ 34), or full-time ($35 \leq$ `uhrsworkt` < 997); hours-varying observations (`uhrsworkt` $= 997$) are set to missing. The discretized variable is denoted `ls_discrete` in the data pipeline.

[64]The first leg is the PSID mean of `n_children` conditional on the woman's age in the prebuilt all-women

Both sources under-count children for women in their forties: as older children leave the household they no longer appear in the mother's reported child count. We correct for this by adding a cumulative birth-probability term at ages 40–45,

$$\hat{N}(a) \;=\; \bar{n}(a) \;+\; \sum_{s=22}^{a-18} P_{\text{birth}}(s), \qquad a \in \{40, \ldots, 45\},$$

where $P_{\text{birth}}(s)$ is the age-s birth probability used to add back the children who have already left home. The two legs use different inputs for this term: the PSID leg is corrected with the standard CPS birth-probability series, whereas the CPS-independent leg is corrected with a birth probability estimated from its own sample and recomputed within each bootstrap draw.

The PSID leg covers roughly 30,300 women observed between 1999 and 2013, while the CPS-independent leg restricts the pooled CPS sample to households satisfying the filter described above. The cumulative birth correction kicks in only at ages 40 and older, rising from 0.064 at age 40 to 0.468 at age 45. After applying the correction, the PSID and CPS-independent profiles average 1.383 and 1.381 children, respectively, with the 50/50 synthetic blend at 1.382 — a 0.1% gap that confirms the two independent sources tell the same story about completed fertility. The synthetic profile rises monotonically from 0.639 at age 22 to a peak of 1.749 at age 43. Table A.16 reports summary statistics for the per-woman number of children by woman's age group, separately for the PSID leg and the CPS-independent leg. Figure A.16 shows the age profile of the mean number of children.

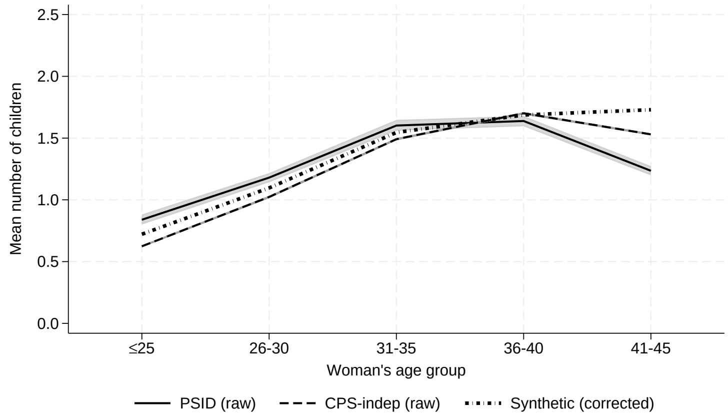
Table A.16: Mean number of children

Variable	Age group	N	Mean	SD	p25	p50	p75
Number of children (PSID)	≤25	3,635	0.840	1.050	0.000	0.000	1.000
	26–30	5,854	1.180	1.220	0.000	1.000	2.000
	31–35	5,414	1.600	1.330	0.000	2.000	2.000
	36–40	5,601	1.640	1.220	1.000	2.000	2.000
	41–45	5,884	1.240	1.150	0.000	1.000	2.000
Number of children (CPS-indep)	≤25	81,862	0.620	0.930	0.000	0.000	1.000
	26–30	139,176	1.020	1.190	0.000	1.000	2.000
	31–35	163,548	1.490	1.310	0.000	1.000	2.000
	36–40	176,096	1.700	1.310	1.000	2.000	2.000
	41–45	176,013	1.530	1.250	0.000	2.000	2.000

Note: Model 2, United States. Mean number of children (PSID and CPS-independent legs); survey-weighted.

panel. The second leg is a CPS-independent mean of `nchild` by age, restricted to the sample of women who are either married (`marst` ≤ 2), have children (`nchild` > 0), or are the household head or spouse (`pernum` ≤ 2).

Figure A.16: Age profile of mean number of children



Note: Model 2, United States. Mean number of children; survey-weighted. The figure compares the raw PSID and CPS-independent legs with the bias-corrected synthetic 50/50 blend that enters estimation; the synthetic line is shown without a band for visual clarity.

Panel (C) — Mean parental time

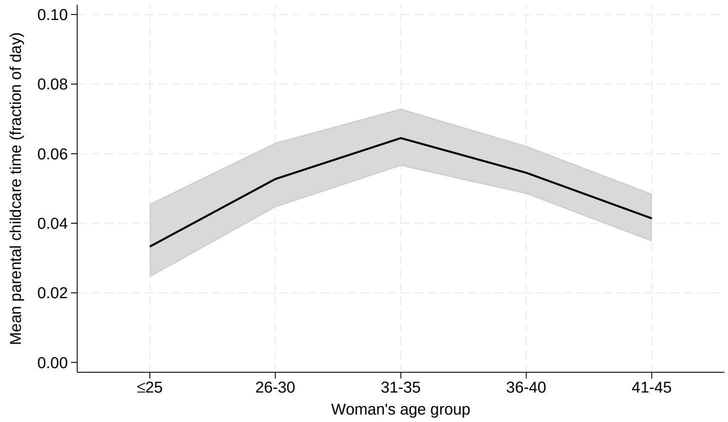
We use the 2014 ATUS to measure women's parental childcare time. Following the same activity-code dictionary used in the M1 panel of Section A.1.1, we sum twenty primary-childcare activity codes (`t030101`, `t030102`, ..., `t030399`) at the respondent level and express the daily total as a fraction of the 24-hour day, $m =$ (daily childcare minutes)/1440. Restricting to women aged 22–45, we then compute the age-specific mean. The underlying ATUS extract is identical to the Model 1 parental-time block in Section A.1.1; the only difference is that we condition on *the woman* rather than the household head. Table A.17 reports descriptive statistics for the woman's parental childcare time (as a fraction of a 24-hour day) by woman's age group. Figure A.17 shows the age profile of mean parental childcare time.

Table A.17: Mean parental childcare time

Variable	Age group	N	Mean	SD	p25	p50	p75
	≤25	248	0.033	0.073	0.000	0.000	0.023
	26–30	481	0.053	0.086	0.000	0.000	0.080
Parental childcare time (fraction of day)	31–35	568	0.064	0.084	0.000	0.028	0.102
	36–40	584	0.055	0.077	0.000	0.021	0.083
	41–45	600	0.041	0.074	0.000	0.001	0.059

Note: Model 2, United States. Woman's parental childcare time, as a fraction of a day; survey-weighted.

Figure A.17: Age profile of mean parental childcare time



Note: Model 2, United States. Mean parental childcare time of the woman; survey-weighted. Parental childcare time is expressed as a fraction of a day.

Panel (D) — Full-time rate

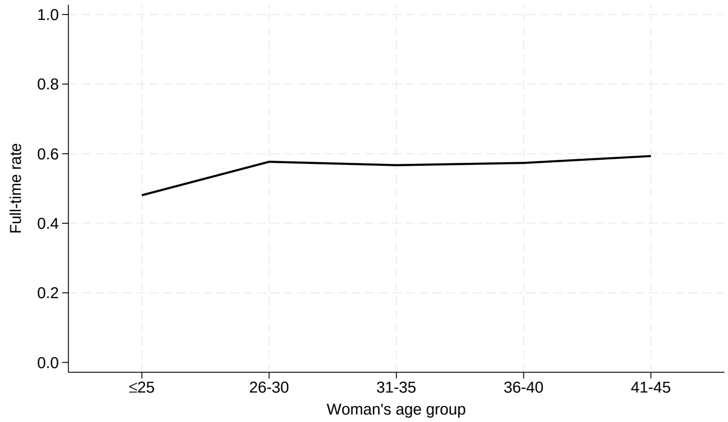
The full-time rate is the share of women whose discretized labor supply equals 0.36 (i.e. working $\geq$ 35 weekly hours), conditional on age. Together with the part-time rate in Panel (E), it disciplines the intensive-margin distribution of women's hours, which the labor-supply mean in Panel (A) alone cannot identify. Figure A.18 shows the age profile of the full-time rate.

Table A.18: Full-time and part-time rates

Variable	Age group	N	Mean	SD
Full-time rate	≤25	109,877	0.481	0.500
	26–30	152,244	0.577	0.494
	31–35	169,619	0.567	0.495
	36–40	179,939	0.574	0.495
	41–45	179,110	0.593	0.491
Part-time rate	≤25	109,877	0.184	0.388
	26–30	152,244	0.125	0.330
	31–35	169,619	0.132	0.338
	36–40	179,939	0.135	0.342
	41–45	179,110	0.133	0.340

Note: Model 2, United States. Shares working full-time and part-time; survey-weighted.

Figure A.18: Age profile of full-time rate



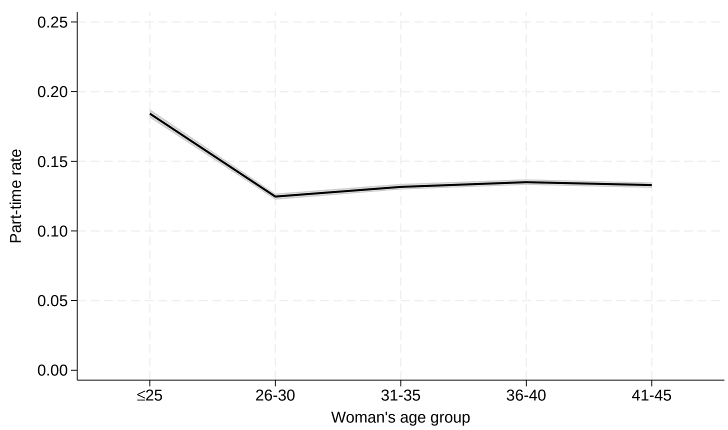
Note: Model 2, United States. Full-time rate; survey-weighted.

Panel (E) — Part-time rate

Analogously, the part-time rate is the share of women whose discretized labor supply equals 0.18 (i.e. working 1–34 weekly hours). The full-time and part-time rates are reported jointly in Table A.18 above.

Figure A.19 shows the age profile of the part-time rate.

Figure A.19: Age profile of part-time rate



Note: Model 2, United States. Part-time rate; survey-weighted.

Panel (F) — Variance of labor supply

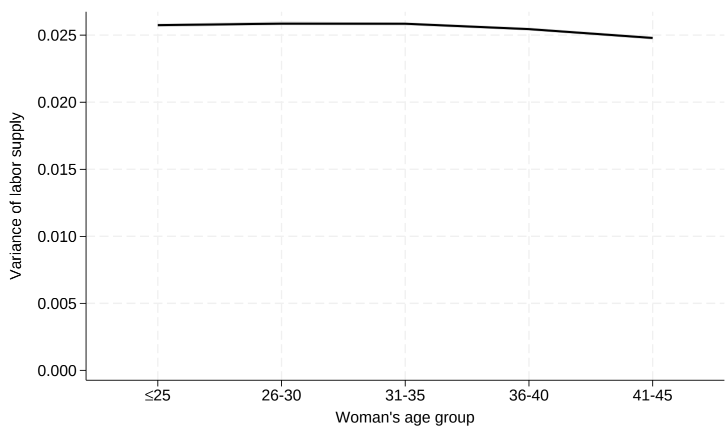
The within-age variance of discretized labor supply, $\mathrm{Var}(l \mid a)$, captures the cross-sectional dispersion of women's labor-supply choices. We include this moment because it takes only three values, so its variance is fully determined by the FT/PT shares and is therefore informative about the heterogeneity in preferences over labor supply. Table A.19 reports the within-age-group variance of labor supply, together with its standard deviation. Figure A.20 shows the age profile of the within-age variance of labor supply.

Table A.19: Variance of labor supply

Variable	Age group	N	Var	SD
Labor supply	≤25	109,877	0.026	0.161
	26–30	152,244	0.026	0.161
	31–35	169,619	0.026	0.161
	36–40	179,939	0.025	0.160
	41–45	179,110	0.025	0.157

Note: Model 2, United States. Within-age-group variance of labor supply; survey-weighted.

Figure A.20: Age profile of variance of labor supply



Note: Model 2, United States. Variance of labor supply; survey-weighted.

Panel (G) — Fraction childless

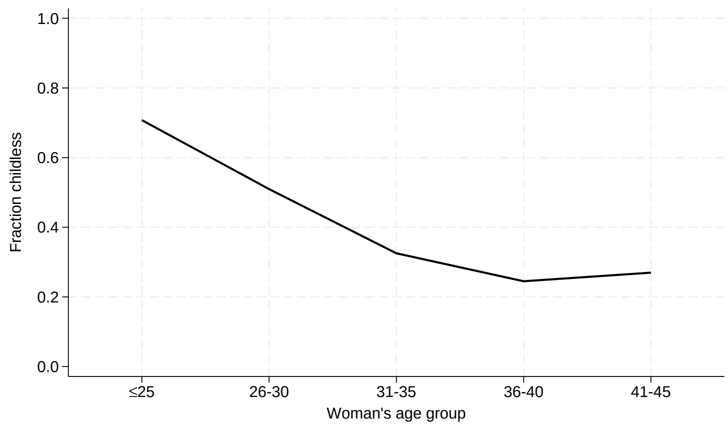
The fraction childless is the share of women with no children in the CPS sample. Combined with the mean number of children in Panel (B), it identifies the spread of the children distribution rather than only its mean. Table A.20 reports the fraction of childless women by woman's age group. Figure A.21 shows the age profile of the fraction childless.

Table A.20: Fraction childless

Variable	Age group	N	Mean	SD
Fraction childless	≤25	109,877	0.708	0.455
	26–30	152,244	0.510	0.500
	31–35	169,619	0.325	0.468
	36–40	179,939	0.245	0.430
	41–45	179,110	0.270	0.444

Note: Model 2, United States. Share with no children; survey-weighted.

Figure A.21: Age profile of fraction childless



Note: Model 2, United States. Fraction childless; survey-weighted.

Panel (H) — Mean labor supply by number of children

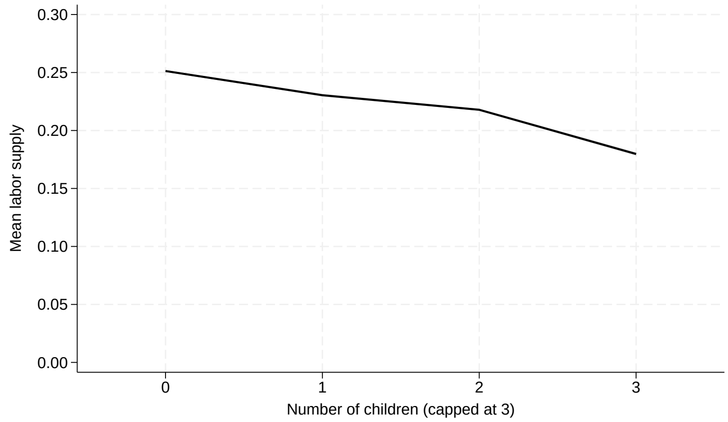
We measure the mean of discretized labor supply by capped number of children $n_{\text{cap}} \in \{0, 1, 2, 3\}$, pooled across ages 22–45. This pooled four-cell moment disciplines how strongly labor supply responds to family size, separately from the age profile of family size itself. Table A.21 reports the mean labor supply by capped number of children, pooled across ages 22–45. Figure A.22 shows the age profile of mean labor supply by number of children.

Table A.21: Mean labor supply by number of children

Variable	N children (capped)	N	Mean	SD	p25	p50	p75
Mean labor supply	0	253,251	0.251	0.153	0.180	0.360	0.360
	1	172,821	0.230	0.160	0.000	0.360	0.360
	2	218,195	0.218	0.161	0.000	0.360	0.360
	3	146,522	0.180	0.165	0.000	0.180	0.360

Note: Model 2, United States. Mean labor supply by capped number of children, pooled across ages 22–45; survey-weighted.

Figure A.22: Age profile of mean labor supply by number of children



Note: Model 2, United States. Mean labor supply by capped number of children; survey-weighted. Pooled across ages 22–45.

Panel (I) — Covariance of labor supply and number of children

We measure the within-age covariance $\mathrm{Cov}(l, n \mid a)$ between discrete labor supply and (capped) number of children, defined as $\mathrm{mean}(l \cdot n \mid a) - \mathrm{mean}(l \mid a) \cdot \mathrm{mean}(n \mid a)$ at each age. Together with the pooled Panel (H), the age-indexed covariance identifies the association between family size and women's labor supply over the lifecycle. Table A.22 reports the within-age-group covariance of labor supply and number of children, along with the corresponding mean labor supply, mean number of children, and mean of their product. Figure A.23 shows the age profile of the within-age covariance of labor supply and number of children.

Table A.22: Covariance of labor supply and number of children

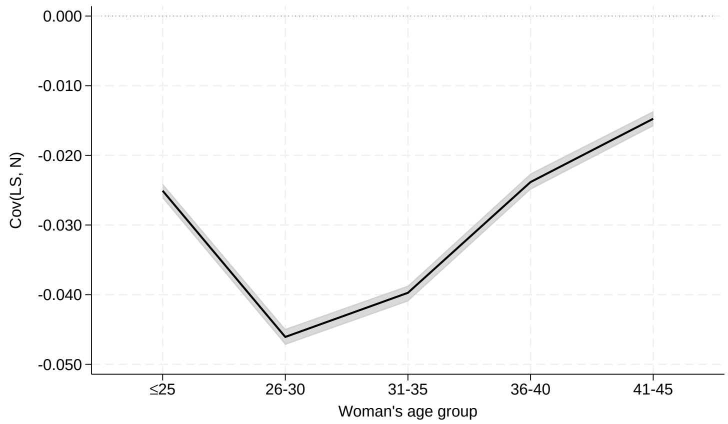
Variable	Age group	N	Cov(LS,N)	Mean LS	Mean N	Mean(LS x N)
	≤25	109,877	-0.025	0.206	0.457	0.069
	26–30	152,244	-0.046	0.230	0.886	0.158
Cov(LS, N)	31–35	169,619	-0.040	0.228	1.337	0.265
	36–40	179,939	-0.024	0.231	1.547	0.333
	41–45	179,110	-0.015	0.238	1.418	0.322

Note: Model 2, United States. Within-age-group covariance of labor supply and number of children; survey-weighted.

Panel (J) — Birth probability

The age-specific birth probability is the share of CPS women whose youngest child is age zero, indicating a birth in the current calendar year. This proxy follows the IPUMS-CPS recommendation for inferring within-year fertility events from a cross-sectional survey.

Figure A.23: Age profile of covariance of labor supply and number of children



Note: Model 2, United States. Within-age covariance of labor supply and number of children; survey-weighted.

Table A.23 reports the birth probability by woman's age group. Figure A.24 shows the age profile of birth probability.

Table A.23: Birth probability

Variable	Age group	N	Mean	SD
Birth probability	≤25	109,877	0.073	0.261
	26–30	152,244	0.091	0.288
	31–35	169,619	0.081	0.273
	36–40	179,939	0.037	0.189
	41–45	179,110	0.008	0.088

Note: Model 2, United States. Age-specific birth probability; survey-weighted.

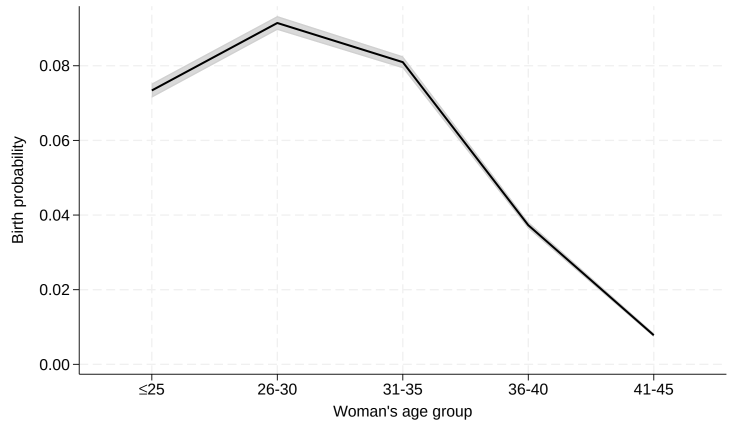
Panel (K) — Share of single women without children

We measure the share of women who are simultaneously unmarried and without children, by age. Table A.24 reports the share of women who are single and without children by woman's age group. Figure A.25 shows the age profile of the share of single women without children.

Panel (L) — Share of single women with children

We measure the share of women who are unmarried but have at least one child, by age. Together with Panel (K) these two shares pin down the joint distribution of marital status and fertility, which the Model 2 single-by-children state space requires. Table A.25 reports the share of women who are single with at least one child by woman's age group. Figure A.26 shows the age profile of the share of single women with children.

Figure A.24: Age profile of birth probability



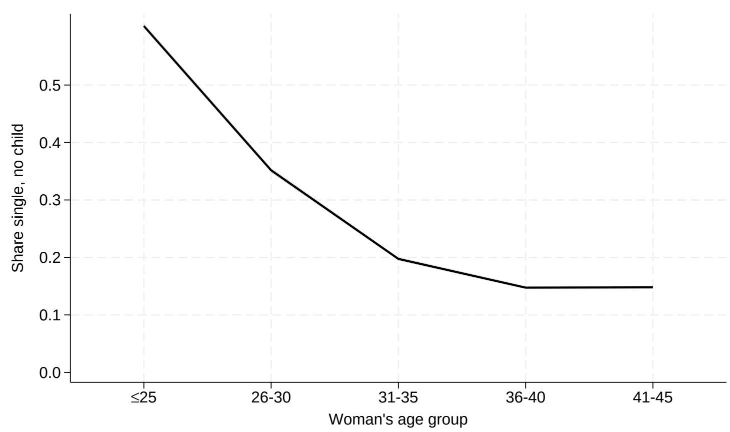
Note: Model 2, United States. Birth probability; survey-weighted.

Table A.24: Share of single women without children

Variable	Age group	N	Mean	SD
	≤25	109,877	0.603	0.489
	26–30	152,244	0.352	0.477
Share single, no child	31–35	169,619	0.197	0.398
	36–40	179,939	0.148	0.355
	41–45	179,110	0.148	0.355

Note: Model 2, United States. Share single without children; survey-weighted.

Figure A.25: Age profile of share of single women without children



Note: Model 2, United States. Share of single women without children; survey-weighted.

Panel (M) — Mean log hourly wage

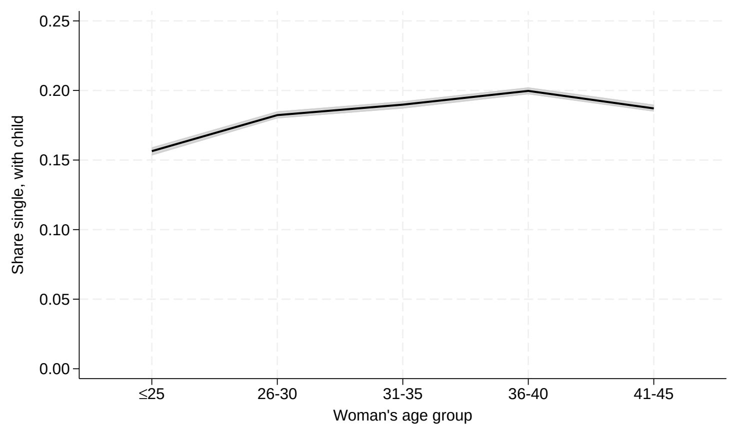
The mean log hourly wage is estimated on PSID 1999–2013 by panel fixed-effects re-

Table A.25: Share of single women with children

Variable	Age group	N	Mean	SD
Share single, with child	≤25	109,877	0.156	0.363
	26–30	152,244	0.182	0.386
	31–35	169,619	0.190	0.392
	36–40	179,939	0.200	0.400
	41–45	179,110	0.187	0.390

Note: Model 2, United States. Share single with children; survey-weighted.

Figure A.26: Age profile of share of single women with children



Note: Model 2, United States. Share of single women with children; survey-weighted.

gression at the woman level, in contrast to the Model 1 wage profile, which is estimated at the household-head level irrespective of gender.

We regress the log net hourly wage—constructed from gross wages winsorized at the year-specific p5/p95, converted to net via TAXSIM, and trimmed below \$1/hr—on year dummies, the woman's age and age-squared, and the number of children, with household fixed effects, restricted to working women ($w > 0$, year $\neq 2013$ to respect TAXSIM v9 coverage). Predicted log wages are collapsed by the woman's age. The resulting age profile is concave: wages rise steeply at younger ages and flatten after the main childbearing years, consistent with the career interruptions and wage penalties women experience around motherhood. Table A.26 reports summary statistics for the mean log hourly wage (PSID, woman-level) by woman's age group. Figure A.27 shows the age profile of the mean log hourly wage.

Summary of data moments

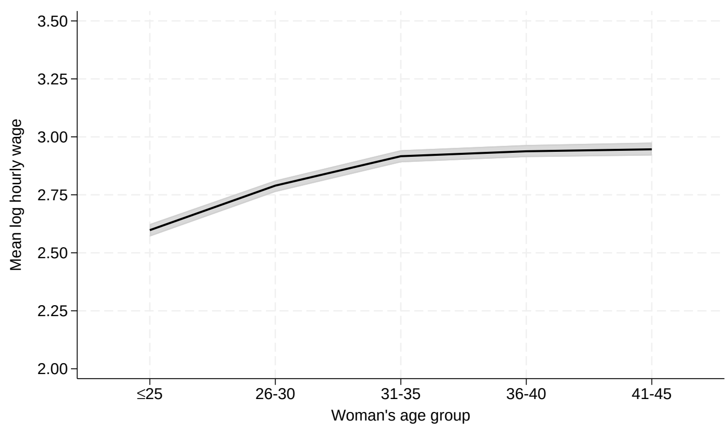
Table A.27 consolidates the ten mean-based data moments used in the US estimation. Each cell reports the anchor mean (upper row) and the 95% bootstrap CI in parentheses

Table A.26: Mean log hourly wage

Variable	Age group	N	Mean	SD	p25	p50	p75
Mean log hourly wage	≤25	3,128	2.598	0.616	2.248	2.588	2.935
	26–30	4,884	2.790	0.667	2.425	2.816	3.157
	31–35	4,428	2.917	0.684	2.519	2.940	3.336
	36–40	4,547	2.938	0.734	2.516	2.957	3.384
	41–45	4,761	2.946	0.737	2.540	2.959	3.355

Note: Model 2, United States. Mean log hourly wage (PSID, woman-level); survey-weighted.

Figure A.27: Age profile of mean log hourly wage



Note: Model 2, United States. Mean log hourly wage; survey-weighted. PSID, woman-level.

(lower row). The variance, covariance, and pooled-by-N panels are reported separately in Tables A.19, A.22, and A.21.

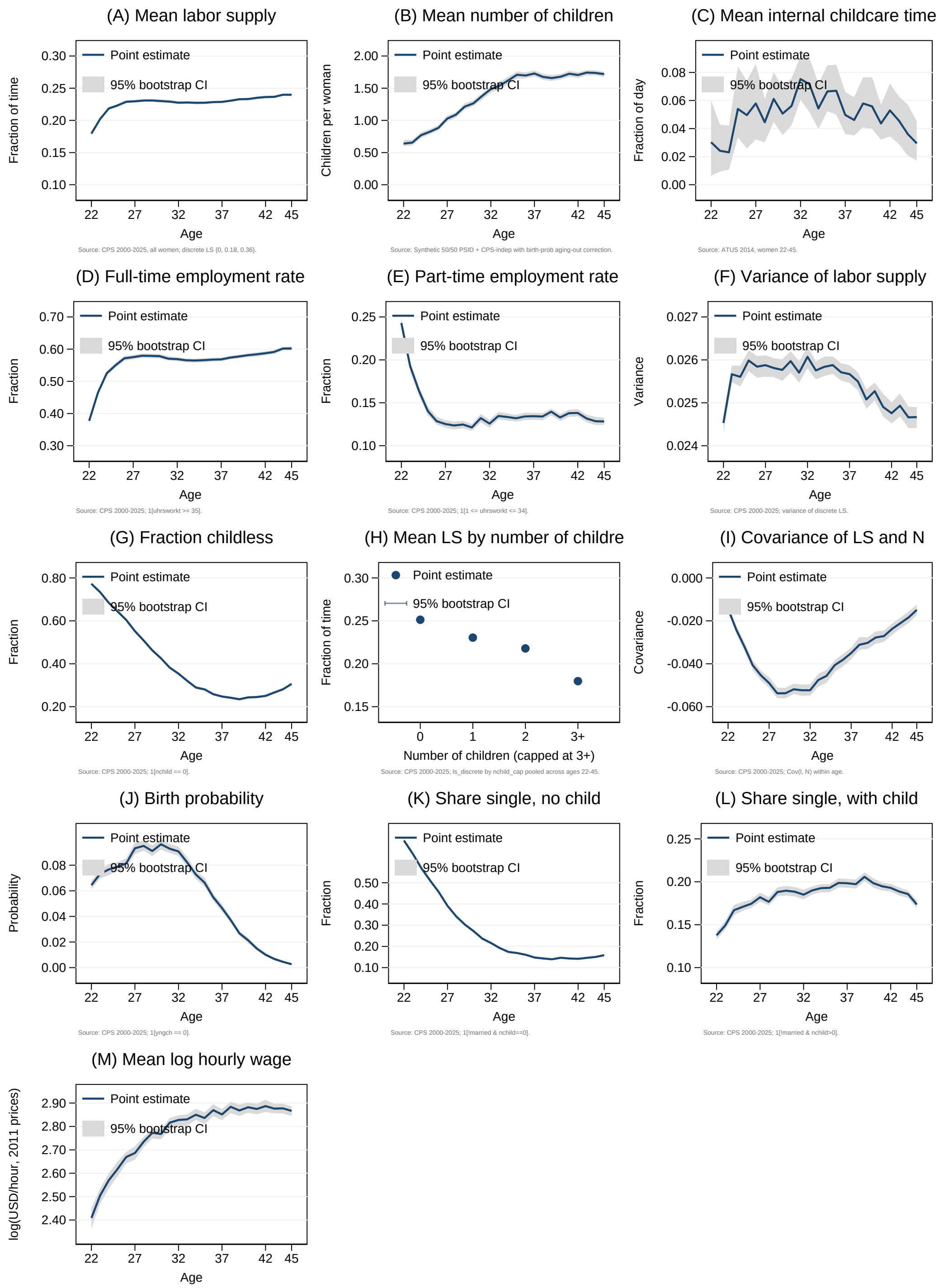
Table A.27: Data moments used in estimation (mean-based panels)

Variable	Age group ≤ 25	26–30	31–35	36–40	41–45
Mean labor supply (A)	0.206	0.230	0.228	0.231	0.237
	(0.205, 0.207)	(0.229, 0.231)	(0.227, 0.229)	(0.230, 0.232)	(0.237, 0.238)
Mean number of children (B)	0.84	1.18	1.60	1.64	1.24
	(0.80, 0.88)	(1.14, 1.22)	(1.56, 1.65)	(1.60, 1.68)	(1.20, 1.27)
Mean parental childcare time (C)	0.033	0.053	0.064	0.055	0.041
	(0.024, 0.046)	(0.045, 0.063)	(0.056, 0.073)	(0.048, 0.062)	(0.035, 0.049)
Full-time rate (D)	0.481	0.577	0.567	0.574	0.593
	(0.477, 0.484)	(0.574, 0.581)	(0.564, 0.570)	(0.570, 0.576)	(0.590, 0.596)
Part-time rate (E)	0.184	0.125	0.132	0.135	0.133
	(0.181, 0.187)	(0.122, 0.127)	(0.130, 0.134)	(0.133, 0.137)	(0.131, 0.135)
Fraction childless (G)	0.708	0.510	0.325	0.245	0.270
	(0.704, 0.711)	(0.506, 0.513)	(0.322, 0.329)	(0.242, 0.248)	(0.267, 0.273)
Birth probability (J)	0.073	0.091	0.081	0.037	0.008
	(0.071, 0.075)	(0.090, 0.093)	(0.079, 0.083)	(0.036, 0.038)	(0.007, 0.008)
Share single, no child (K)	0.603	0.352	0.197	0.148	0.148
	(0.599, 0.607)	(0.348, 0.355)	(0.195, 0.200)	(0.145, 0.150)	(0.145, 0.151)
Share single, with child (L)	0.156	0.182	0.190	0.200	0.187
	(0.153, 0.159)	(0.180, 0.185)	(0.187, 0.192)	(0.197, 0.202)	(0.185, 0.190)
Mean log hourly wage (M)	2.598	2.790	2.917	2.938	2.946
	(2.570, 2.624)	(2.762, 2.812)	(2.890, 2.943)	(2.913, 2.965)	(2.920, 2.976)

Note: Model 2, United States. Anchor mean ($b = 0$, upper row) and 95% bootstrap CI (lower row); 300 bootstrap draws; survey-weighted. Variance, covariance, and pooled-by-N panels appear in their own tables.

The combined age profiles for the thirteen US data moments described above are summarized in Figure A.28. Each panel shows one of the thirteen moment blocks by single year of age, the resolution used in the MSM estimation, with 95% bootstrap confidence bands where available; the figure is the visual counterpart of the consolidated US moment table above, which aggregates to five age groups for compactness.

Figure A.28: Data-moment age profiles



Note: Model 2, United States. Age profiles of data moments; survey-weighted.

A.2.2 Model 2 for Germany

The data underlying the Model 2 estimation for Germany draws on two surveys: the German Socio-Economic Panel (SOEP) for labor supply, fertility status, marital structure, and hourly wages; and the German Family Panel (Pairfam, Wave 5, 2013/14) for parental childcare time. Together these surveys yield 292 moments across 13 blocks, structurally aligned with the US Model 2 moment vector of Section A.2.1.

Sample. The sample is composed of single women and women in couples, constructed from SOEP via two complementary cases. In *Case A*, the woman is the survey-designated household head, and her variables enter directly; she is recorded as married if the household reports a partner. In *Case B*, the woman is the female partner of a male head, and her variables enter from the partner slot; Case B women are married by construction. After restricting to women aged 22–45, the combined panel contains roughly 61,000 person-year observations. Labor supply is discretized to $\{0, 0.15, 0.30\}$, corresponding to non-working, part-time, and full-time hours measured against a 510-hour monthly reference.[65] The number of children is capped at three, and hourly wages are winsorized at the 1st and 99th percentiles. All age-indexed moments are computed as means by age, with weights given by the SOEP household weight times a per-draw bootstrap weight, and serve both for the anchor model (the $b = 0$ draw) and for the bootstrap covariance matrix used in the estimation.

External childcare price. The Model 2 household budget constraint embeds an annualized external childcare cost parameter equal to €32,640 per year for Germany—the full-time annual earnings of one external childcare provider.[66]

Panel (A) — Mean labor supply

The SOEP woman-level panel contains roughly 61,000 person-year observations aged 22–45. Following the model's three labor states $\{0, \ell^p, \ell^f\}$ for Germany, discretized labor supply takes values 0 (non-working), 0.15 (part-time), and 0.30 (full-time). Mean labor supply averages $\bar{\ell} = 0.185$ across ages, with the profile starting at 0.211 at age 22, dipping to a minimum of 0.169 at age 35 during the peak-childbearing years, and recovering to 0.194 by age 45. The lower level relative to the United States ($\bar{\ell} = 0.227$) and the visible mid-life dip reflect the German pattern of stronger part-time-induced labor-supply reductions among

[65] Formally, $l^{\text{cont}} = \texttt{p\_hrs\_monthly}/510$, with $l^{\text{cont}} < 0.075 \rightarrow 0$ (non-working), $0.075 \leq l^{\text{cont}} < 0.225 \rightarrow 0.15$ (part-time), and $l^{\text{cont}} \geq 0.225 \rightarrow 0.30$ (full-time).

[66] See footnote 62 for the unit convention (Model 1 hourly vs. Model 2 annualized) and its reconciliation across the two countries.

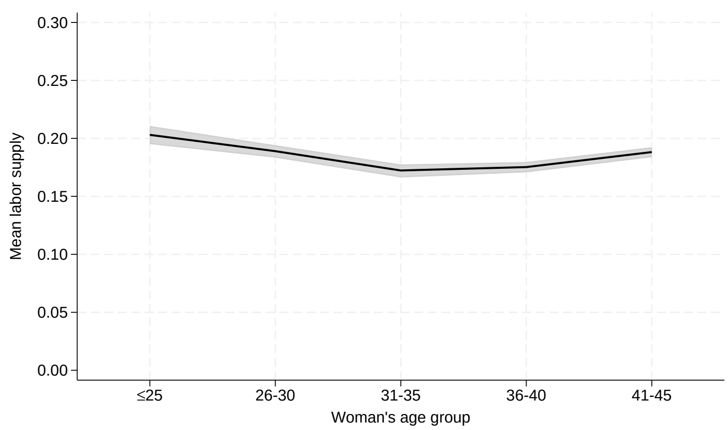
women caring for young children. The support $\{0, 0.15, 0.30\}$ differs from the US support $\{0, 0.18, 0.36\}$ because the underlying hours scale (a 510-hour monthly reference in SOEP versus 35 weekly hours in the CPS) and the FT/PT cutoffs are country-specific. Table A.28 reports summary statistics for the discrete labor-supply variable on the SOEP woman-level panel, by woman's age group. Figure A.29 shows the age profile of mean labor supply.

Table A.28: Mean labor supply

Variable	Age group	N	Mean	SD	p25	p50	p75
Mean labor supply	≤25	4,727	0.203	0.132	0.000	0.300	0.300
	26–30	10,573	0.189	0.134	0.000	0.300	0.300
	31–35	13,731	0.172	0.133	0.000	0.150	0.300
	36–40	15,551	0.175	0.126	0.000	0.150	0.300
	41–45	16,436	0.188	0.122	0.150	0.150	0.300

Note: Model 2, Germany. Mean discretized labor supply; survey-weighted.

Figure A.29: Age profile of mean labor supply



Note: Model 2, Germany. Mean labor supply; survey-weighted.

Panel (B) — Mean number of children

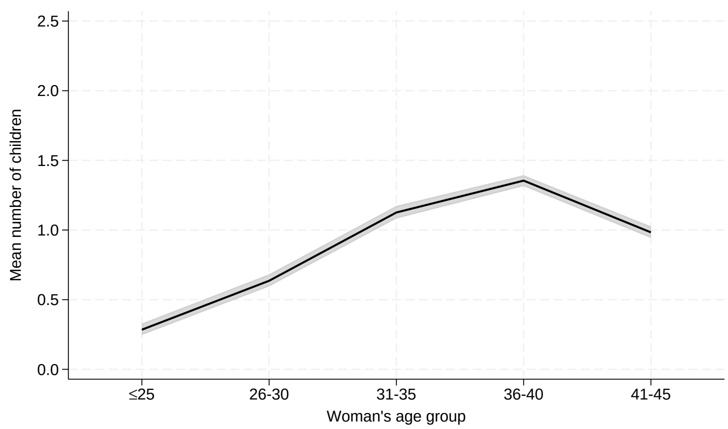
The mean number of children by age is constructed via the *biokids cumulative-births correction* rather than from the raw SOEP household-child count. The raw count understates the number of children for older women: as older children leave the parental household, they no longer appear in the mother's reported count, biasing the profile downward at older ages. The correction reconstructs each woman's cumulative biological-children count from year-on-year birth events in the SOEP biographical module and uses that count as the moment input. Table A.29 reports summary statistics for the per-woman number of children by woman's age group. Figure A.30 shows the age profile of the mean number of children.

Table A.29: Mean number of children

Variable	Age group	N	Mean	SD	p25	p50	p75
Number of children	≤25	4,727	0.280	0.590	0.000	0.000	0.000
	26–30	10,573	0.640	0.870	0.000	0.000	1.000
	31–35	13,731	1.130	1.040	0.000	1.000	2.000
	36–40	15,551	1.350	1.040	1.000	1.000	2.000
	41–45	16,436	0.980	1.000	0.000	1.000	2.000

Note: Model 2, Germany. Mean number of children; survey-weighted.

Figure A.30: Age profile of mean number of children



Note: Model 2, Germany. Mean number of children; survey-weighted.

Panel (C) — Mean parental time

SOEP does not record detailed childcare time at the respondent level. We therefore decompose mean parental time at age a into an extensive margin (the fraction of women at age a who are mothers) and an intensive margin (mean childcare time among mothers at age a):

$$\overline{m}(a) \;=\; \text{frac\_mother}(a) \;\times\; m^{\text{intensive}}(a).$$

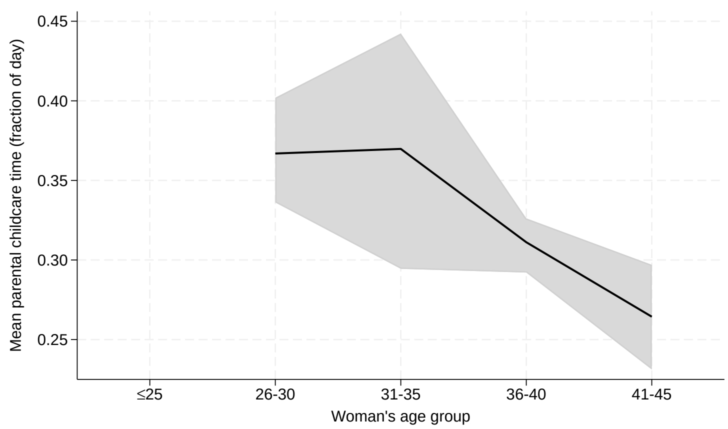
The extensive margin is computed in SOEP as the share of women with at least one child at each age. The intensive margin is fitted on Pairfam Wave 5 mothers at ages 22–42 as a quadratic in age, $m^{\text{intensive}}(a) = \beta_0 + \beta_1\, a + \beta_2\, a^2$, using the estimated coefficients $(\beta_0, \beta_1, \beta_2) = (0.211, -0.0037, 0.00002)$ and floored at zero. The product $\text{frac\_mother}(a) \times m^{\text{intensive}}(a)$ is then evaluated on the 22–45 age grid. Table A.30 reports unweighted summary statistics for the Pairfam parental childcare time variable underlying the quadratic fit, by woman's age group. Figure A.31 shows the age profile of mean parental childcare time.

Table A.30: Mean parental childcare time

Variable	Age group	N	Mean	SD	p25	p50	p75
Intensive margin (mothers only)	26–30	536	0.367	0.152	0.333	0.333	0.333
	31–35	128	0.370	0.181	0.167	0.333	0.500
	36–40	2,044	0.311	0.153	0.167	0.333	0.333
	41–45	232	0.264	0.135	0.167	0.167	0.333
Unconditional moment	≤25	—	0.029	—	—	—	—
	26–30	—	0.052	—	—	—	—
	31–35	—	0.074	—	—	—	—
	36–40	—	0.078	—	—	—	—
	41–45	—	0.057	—	—	—	—

Note: Model 2, Germany. Woman's parental childcare time (Pairfam); intensive-margin (conditional on reporting) and unconditional moments; unweighted.

Figure A.31: Age profile of mean parental childcare time



Note: Model 2, Germany. Mean parental childcare time of the woman; unweighted. Parental childcare time is expressed as a fraction of a day; Pairfam Wave 5 raw data.

Panel (D) — Full-time rate

We measure the share of women in the SOEP woman-only panel who are full-time workers (i.e., discretized labor supply equal to 0.30), conditional on age. Figure A.32(a) shows the age profile of the full-time rate.

Panel (E) — Part-time rate

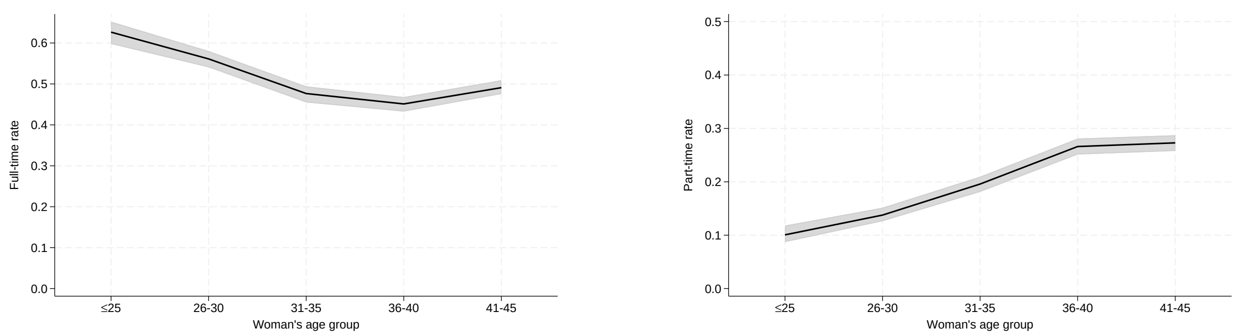
Analogously, the part-time rate is the share of women whose discretized labor supply equals 0.15, by age. The full-time and part-time rates are reported jointly in Table A.31 above. Figure A.32(b) shows the age profile of the part-time rate.

Table A.31: Full-time and part-time rates

Variable	Age group	N	Mean	SD
Full-time rate	≤25	4,727	0.627	0.484
	26–30	10,573	0.561	0.496
	31–35	13,731	0.477	0.499
	36–40	15,551	0.451	0.498
	41–45	16,436	0.491	0.500
Part-time rate	≤25	4,727	0.101	0.301
	26–30	10,573	0.138	0.345
	31–35	13,731	0.196	0.397
	36–40	15,551	0.266	0.442
	41–45	16,436	0.273	0.445

Note: Model 2, Germany. Shares working full-time and part-time; survey-weighted.

Figure A.32: Age profile of full-time and part-time rates



(a) Full-time rate (b) Part-time rate

Note: Model 2, Germany. Full-time and part-time rates; survey-weighted.

Panel (F) — Variance of labor supply

The within-age variance of discretized labor supply, $\mathrm{Var}(l \mid a)$, is computed on the SOEP woman-only panel. As in the US case, this moment is fully determined by the FT/PT/non-working shares (since labor supply has only three support points) but provides a compact summary of the cross-sectional dispersion in labor-supply choices. Table A.32 reports the within-age-group variance of labor supply, together with its standard deviation. Figure A.33 shows the age profile of the within-age variance of labor supply.

Panel (G) — Fraction childless

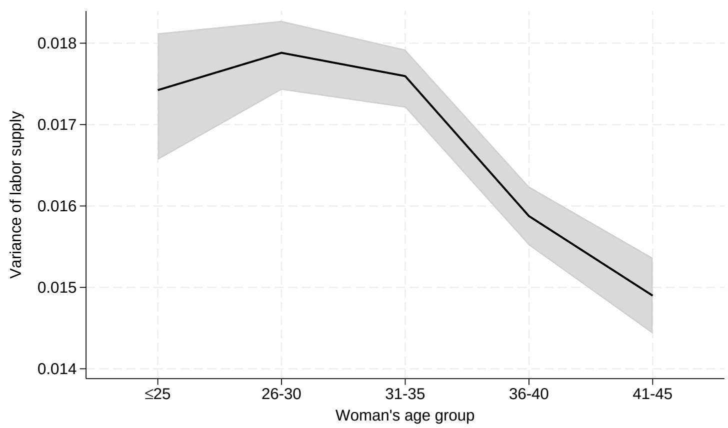
We measure the share of women with no children in the SOEP woman-only panel. Table A.33 reports the fraction of childless women by woman's age group. Figure A.34 shows the age profile of the fraction childless.

Table A.32: Variance of labor supply

Variable	Age group	N	Var	SD
Labor supply	≤25	4,727	0.017	0.132
	26–30	10,573	0.018	0.134
	31–35	13,731	0.018	0.133
	36–40	15,551	0.016	0.126
	41–45	16,436	0.015	0.122

Note: Model 2, Germany. Within-age-group variance of labor supply; survey-weighted.

Figure A.33: Age profile of variance of labor supply



Note: Model 2, Germany. Variance of labor supply; survey-weighted.

Table A.33: Fraction childless

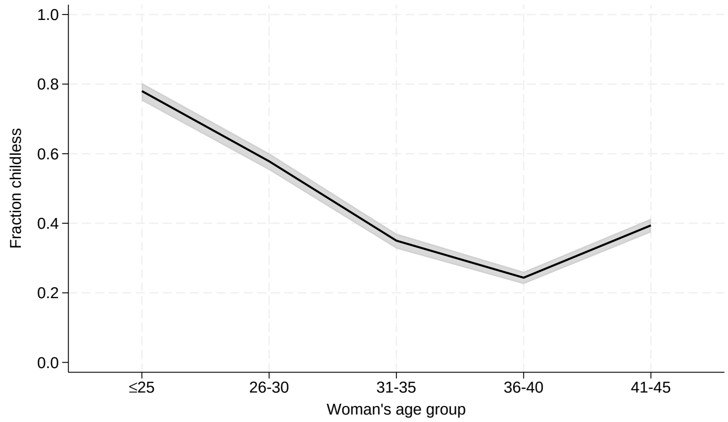
Variable	Age group	N	Mean	SD
Fraction childless	≤25	4,727	0.780	0.414
	26–30	10,573	0.578	0.494
	31–35	13,731	0.350	0.477
	36–40	15,551	0.244	0.429
	41–45	16,436	0.394	0.489

Note: Model 2, Germany. Share with no children; survey-weighted.

Panel (H) — Mean labor supply by number of children

Pooling across ages 22–45 in the SOEP woman-only panel, we measure the mean discretized labor supply by capped number of children $n_{\text{cap}} \in \{0, 1, 2, 3\}$. Table A.34 reports the mean labor supply by capped number of children, pooled across ages 22–45. Figure A.35 shows the age profile of mean labor supply by number of children.

Figure A.34: Age profile of fraction childless



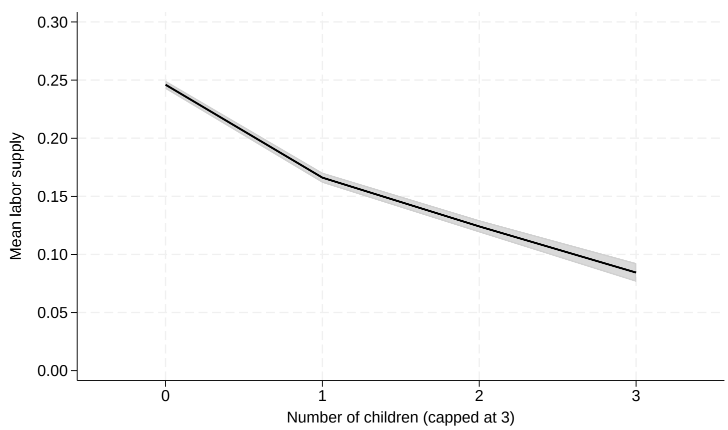
Note: Model 2, Germany. Fraction childless; survey-weighted.

Table A.34: Mean labor supply by number of children

Variable	N children (capped)	N	Mean	SD	p25	p50	p75
Mean labor supply	0	21,960	0.246	0.105	0.300	0.300	0.300
	1	17,798	0.166	0.127	0.000	0.150	0.300
	2	16,216	0.124	0.120	0.000	0.150	0.150
	3	5,044	0.084	0.111	0.000	0.000	0.150

Note: Model 2, Germany. Mean labor supply by capped number of children, pooled across ages 22–45; survey-weighted.

Figure A.35: Age profile of mean labor supply by number of children



Note: Model 2, Germany. Mean labor supply by capped number of children; survey-weighted. Pooled across ages 22–45.

Panel (I) — Covariance of labor supply and number of children

We measure the within-age covariance $\mathrm{Cov}(l, n \mid a)$, defined as $\mathrm{mean}(l \cdot n \mid a) - \mathrm{mean}(l \mid a) \cdot \mathrm{mean}(n \mid a)$ at each age in the SOEP woman-only panel. Together with the pooled

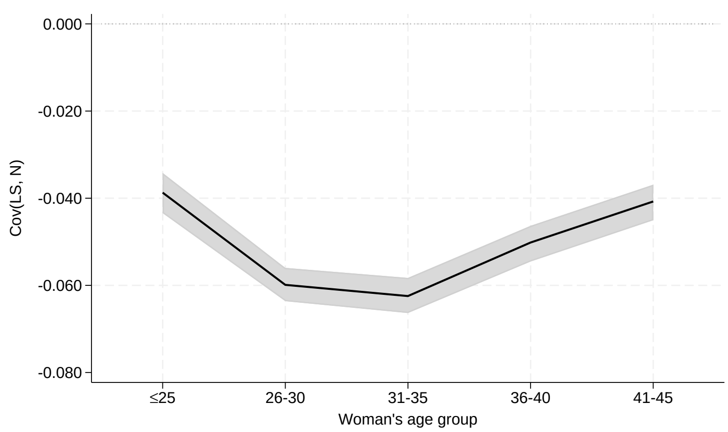
Panel (H), this age-indexed covariance identifies the association between family size and women's labor supply over the lifecycle. Table A.35 reports the within-age-group covariance of labor supply and number of children, along with the corresponding mean labor supply, mean number of children, and mean of their product. Figure A.36 shows the age profile of the within-age covariance of labor supply and number of children.

Table A.35: Covariance of labor supply and number of children

Variable	Age group	N	Cov(LS,N)	Mean LS	Mean N	Mean(LS x N)
	≤25	4,727	-0.039	0.203	0.284	0.019
	26–30	10,573	-0.060	0.189	0.628	0.059
Cov(LS, N)	31–35	13,731	-0.062	0.172	1.106	0.128
	36–40	15,551	-0.050	0.175	1.324	0.182
	41–45	16,436	-0.041	0.188	0.965	0.141

Note: Model 2, Germany. Within-age-group covariance of labor supply and number of children; survey-weighted.

Figure A.36: Age profile of covariance of labor supply and number of children



Note: Model 2, Germany. Within-age covariance of labor supply and number of children; survey-weighted.

Panel (J) — Birth probability

Unlike the US case, where we rely on the cross-sectional proxy that the youngest child is age zero, the SOEP panel allows us to identify births directly from year-on-year transitions in a woman's number of children.[67] Table A.36 reports the birth probability by woman's age group. Figure A.37 shows the age profile of birth probability.

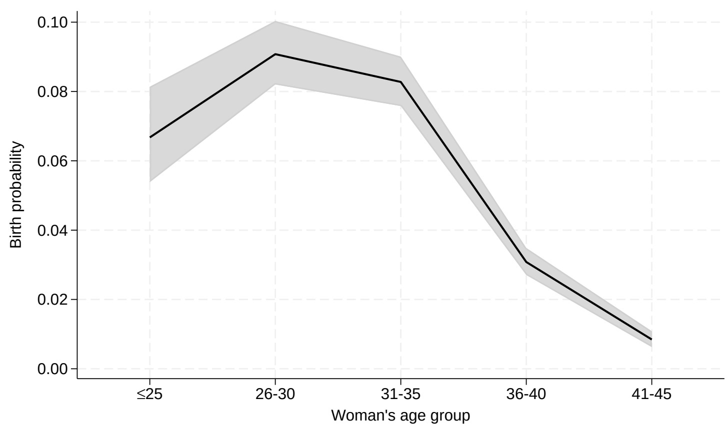
[67]At age 22, the youngest age in the sample, the year-on-year measure is undefined for first-year entrants who have no prior-year observation; we set $P_{\text{birth}}(22) = 0.040$ to continue the profile's rise into the mid-twenties, just below the values observed at ages 23 and 24.

Table A.36: Birth probability

Variable	Age group	N	Mean	SD
Birth probability	≤25	2,260	0.067	0.250
	26–30	7,353	0.091	0.287
	31–35	10,366	0.083	0.276
	36–40	12,110	0.031	0.173
	41–45	13,146	0.008	0.092

Note: Model 2, Germany. Age-specific birth probability; survey-weighted.

Figure A.37: Age profile of birth probability



Note: Model 2, Germany. Birth probability; survey-weighted.

Panel (K) — Share of single women without children

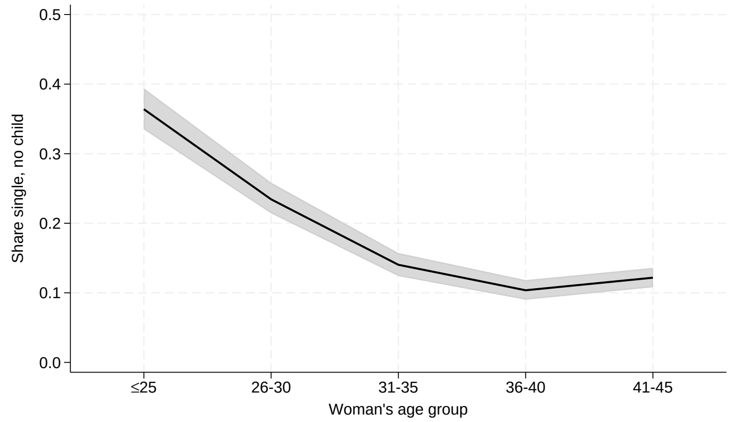
We measure the share of women who are simultaneously unmarried and without children (SOEP Case A), by age. Table A.37 reports the share of women who are single and without children by woman's age group. Figure A.38 shows the age profile of the share of single women without children.

Table A.37: Share of single women without children

Variable	Age group	N	Mean	SD
Share single, no child	≤25	4,727	0.364	0.481
	26–30	10,573	0.234	0.424
	31–35	13,731	0.140	0.347
	36–40	15,551	0.104	0.305
	41–45	16,436	0.122	0.327

Note: Model 2, Germany. Share single without children; survey-weighted.

Figure A.38: Age profile of share of single women without children



Note: Model 2, Germany. Share of single women without children; survey-weighted.

Panel (L) — Share of single women with children

We measure the share of unmarried women with at least one child, by age. Table A.38 reports the share of women who are single with at least one child by woman's age group. Figure A.39 shows the age profile of the share of single women with children.

Table A.38: Share of single women with children

Variable	Age group	N	Mean	SD
	≤25	4,727	0.016	0.126
	26–30	10,573	0.040	0.196
Share single, with child	31–35	13,731	0.068	0.252
	36–40	15,551	0.093	0.291
	41–45	16,436	0.082	0.274

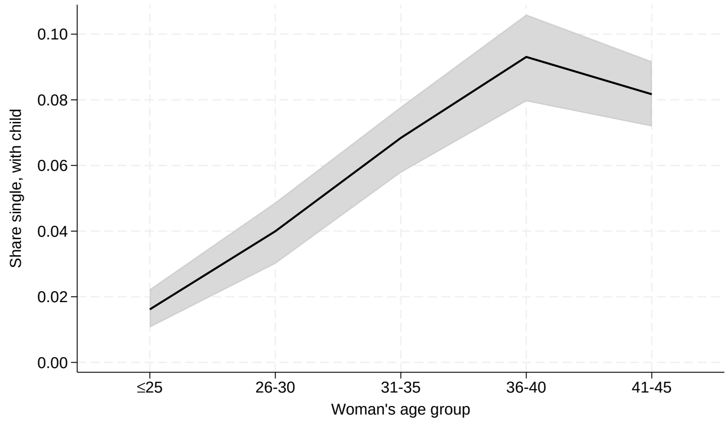
Note: Model 2, Germany. Share single with children; survey-weighted.

Panel (M) — Mean log hourly wage

The Model 2 wage moment is the cross-sectional mean of log hourly wages by woman's age.[68] Table A.39 reports summary statistics for the mean log hourly wage by woman's age group. Figure A.40 shows the age profile of the mean log hourly wage.

[68]For each Case A observation, we compute the hourly wage as labor income divided by monthly hours (when both are positive), winsorize it at the 1st and 99th percentiles, and take the log. Case B women are included only in the hours and children moments, not the wage moment, because SOEP lacks comparable wage data for women observed in the partner role.

Figure A.39: Age profile of share of single women with children



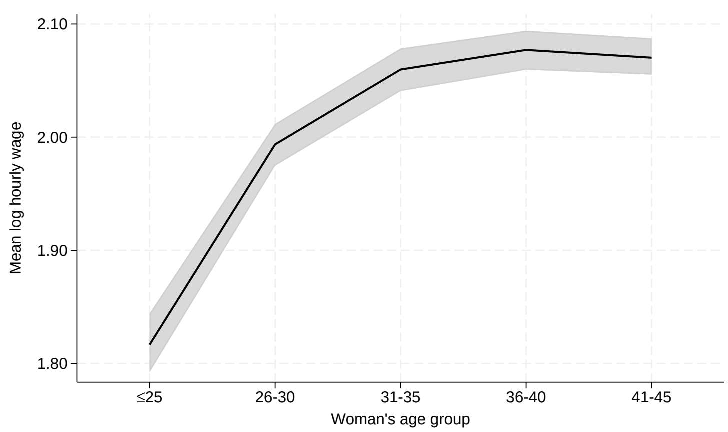
Note: Model 2, Germany. Share of single women with children; survey-weighted.

Table A.39: Mean log hourly wage

Variable	Age group	N	Mean	SD	p25	p50	p75
Mean log hourly wage	≤25	3,392	1.817	0.394	1.628	1.896	2.064
	26–30	7,267	1.994	0.374	1.812	2.045	2.230
	31–35	9,361	2.060	0.406	1.836	2.104	2.324
	36–40	11,420	2.077	0.415	1.830	2.100	2.347
	41–45	12,937	2.070	0.427	1.814	2.094	2.339

Note: Model 2, Germany. Mean log hourly wage (SOEP, woman-level); survey-weighted.

Figure A.40: Age profile of mean log hourly wage



Note: Model 2, Germany. Mean log hourly wage; survey-weighted. SOEP, woman-level.

Summary of data moments

Table A.40 consolidates the ten mean-based data moments used in the German estimation. Each cell reports the anchor mean (upper row) and the 95% bootstrap CI in parentheses (lower row). The variance, covariance, and pooled-by-N panels are reported separately in Tables A.32, A.35, and A.34.

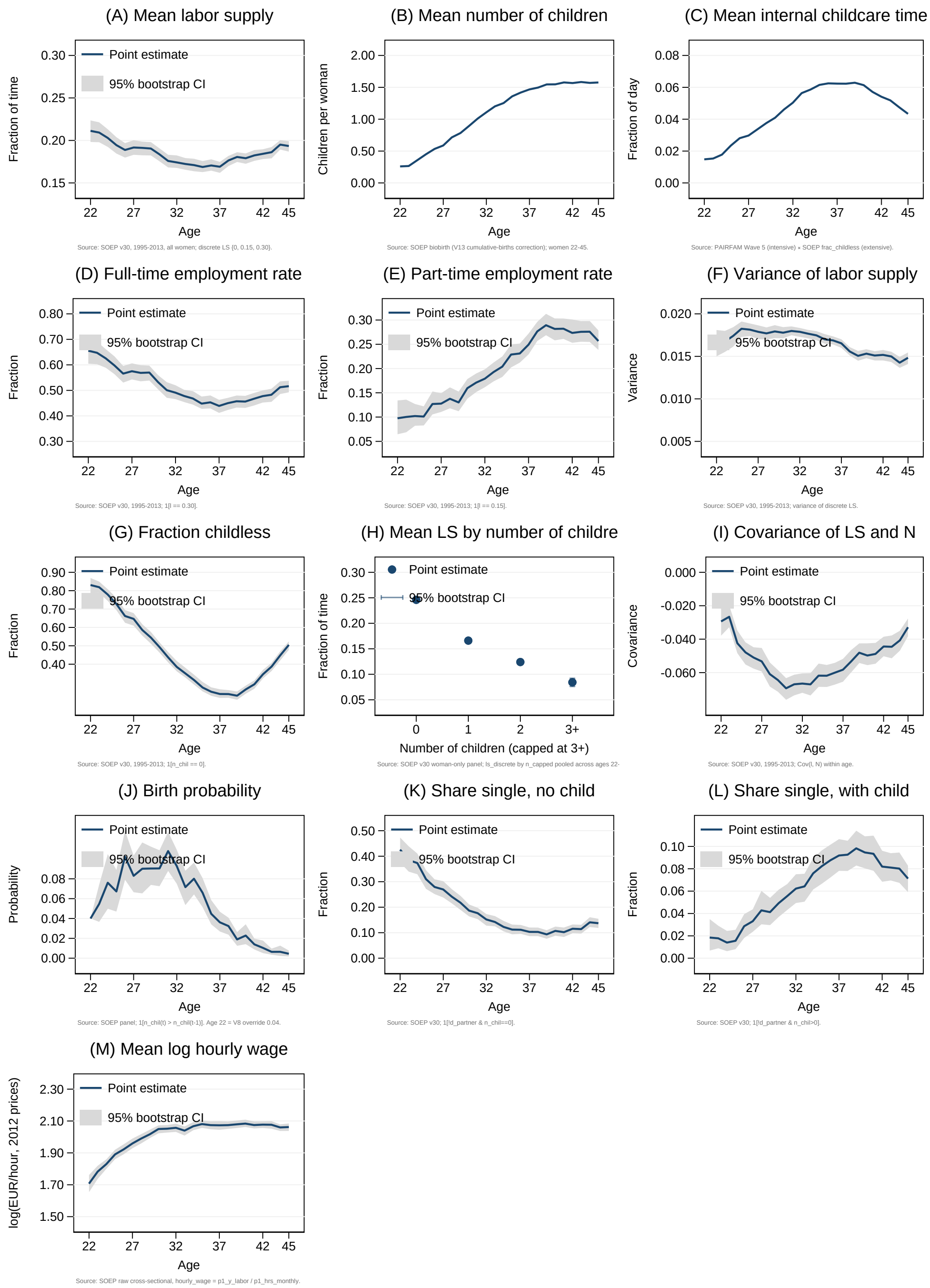
Table A.40: Data moments used in estimation (mean-based panels)

	Age group				
Variable	≤ 25	26–30	31–35	36–40	41–45
Mean labor supply (A)	0.203	0.189	0.172	0.175	0.188
	(0.195, 0.211)	(0.183, 0.194)	(0.166, 0.178)	(0.171, 0.180)	(0.184, 0.192)
Mean number of children (B)	0.28	0.64	1.13	1.35	0.98
	(0.25, 0.33)	(0.60, 0.68)	(1.08, 1.17)	(1.32, 1.39)	(0.94, 1.03)
Full-time rate (D)	0.627	0.561	0.477	0.451	0.491
	(0.598, 0.652)	(0.541, 0.580)	(0.455, 0.494)	(0.432, 0.468)	(0.475, 0.509)
Part-time rate (E)	0.101	0.138	0.196	0.266	0.273
	(0.087, 0.118)	(0.126, 0.152)	(0.181, 0.210)	(0.251, 0.281)	(0.258, 0.287)
Fraction childless (G)	0.780	0.578	0.350	0.244	0.394
	(0.753, 0.803)	(0.553, 0.601)	(0.327, 0.370)	(0.225, 0.260)	(0.374, 0.413)
Birth probability (J)	0.067	0.091	0.083	0.031	0.008
	(0.054, 0.081)	(0.082, 0.100)	(0.076, 0.090)	(0.027, 0.035)	(0.006, 0.011)
Share single, no child (K)	0.364	0.234	0.140	0.104	0.122
	(0.335, 0.394)	(0.214, 0.258)	(0.124, 0.157)	(0.090, 0.118)	(0.108, 0.136)
Share single, with child (L)	0.016	0.040	0.068	0.093	0.082
	(0.011, 0.022)	(0.030, 0.049)	(0.058, 0.078)	(0.080, 0.106)	(0.072, 0.092)
Mean log hourly wage (M)	1.817	1.994	2.060	2.077	2.070
	(1.792, 1.844)	(1.975, 2.012)	(2.041, 2.078)	(2.060, 2.094)	(2.055, 2.087)
Mean parental childcare time (C)	0.029	0.052	0.074	0.078	0.057
	(0.021, 0.037)	(0.045, 0.058)	(0.068, 0.080)	(0.074, 0.083)	(0.045, 0.069)

Note: Model 2, Germany. Anchor mean ($b = 0$, upper row) and 95% bootstrap CI (lower row); 300 bootstrap draws; survey-weighted. Variance, covariance, and pooled-by-N panels appear in their own tables.

The combined age profiles for the thirteen German data moments described above are summarized in Figure A.41. Each panel shows one of the thirteen moment blocks by single year of age, the resolution used in the MSM estimation, with 95% bootstrap confidence bands where available; the figure is the visual counterpart of the consolidated German moment table above, which aggregates to five age groups for compactness.

Figure A.41: Data-moment age profiles



Note: Model 2, Germany. Age profiles of data moments; survey-weighted.

B Vignette survey on fertility/career decisions

B.1 General overview of the module

The vignette was implemented as a satellite survey module to Pairfam in Germany and Understanding America Study in the United States. The two panels provide us with a rich set of socio-demographic background variables over respondents' biographies. This set is further extended by desired fertility in the respondent's current situation. The UAS version of the module can be downloaded from: https://uasdata.usc.edu/UAS-27.

The description of our vignette module consists of two parts:

Part A - Technical details regarding the survey module.

Part B - Description of the design of the vignette module.

B.2 Part A - Technical details

B.2.1 Target population

Our target population is respondents in the phase of family planning and career making: persons aged 18 to 50.

B.2.2 Basic structure of the survey module

Our module provides a random sequence of vignettes to each respondent. Each vignette describes features of a hypothetical household and hypothetical policies. Hereon, these features are called *environment*. An environment is characterized along five *dimensions*: availability of childcare facilities; level of child transfers; working hours of the partner; earnings of the partner; and hourly wage rate. Based on the five dimensions, we distinguish 80 hypothetical environments. A random subset of four environments is presented to each respondent. For each environment, we ask the respondent to provide a) the desired *number of working hours* and b) the desired *number of children*.

B.2.3 Definition of environments

Five dimensions characterize each environment, e, with $e = 1, ..., 80$:

(i). Availability of childcare facilities (s_e): publicly available at zero cost vs. costly.

(ii). Monthly child allowance (A_e).

(iii). Working hours of the partner ($h_{p,e}$)

(iv). Net earnings of the partner ($y_{p,e}$).

(v). Hourly net wage rate (w_e).

Table B.1 explains the categories of each dimension as implemented in the UAS and Pairfam survey.

Table B.1: Attributes of environments

Dimension	Variable	Unit	Attributes
Childcare facilities	s_e		guaranteed & costless during working hours; private responsibility and costly
Child allowance	A_e	\$/month	200; 400
		€/month	250; 500
Working hours of partner	$h_{p,e}$	hours/month	0; 80; 160
Net earnings of partner	$y_{p,e}$	\$/month	0; 800; 1,600; 2,400; 4,800
		€/month	0; 800; 1,600; 3,200
Individual wage rate	w_e	\$/month	5; 10; 20; 40
		€/month	5; 10; 30; 50

The attributes of the five dimensions distinguish 80 environments:

- From policy dimensions (s_e, A_e): $2 \times 2 = 4$
- From household dimensions $(h_{p,e}, y_{p,e}, w_e)$: $5 \times 4 = 20$ (Note that not all combinations of $h_{p,e}$ and $y_{p,e}$ are feasible.)

B.2.4 Assigning environment sequences to respondents

The following procedure explains the selection of the four environments (from the full set of 80 environments) to a respondent:

Stage 1: One environment is randomly drawn from the full set of environments. This is the first environment presented to the respondent.

Stage 2: One of the five possible attributes of the first environment is modified. Together with the four unchanged attributes of environment one, this determines the second environment presented to the respondent.

Stage 3: One of the five possible attributes of the second environment is modified. Together with the four unchanged attributes of the second environment, this determines the third environment presented to the respondent.

Stage 4: One of the five possible attributes of the third environment is modified. Together with the four unchanged attributes of the third environment, this determines the fourth environment presented to the respondent.

B.3 Part B - Design of the vignette survey

B.3.1 Introductory text

Every respondent received the following introductory text:

Introduction

People have to make many decisions every day, some of which have long-lasting implications. The present survey deals with two of these, namely planning a family and going to work. These decisions may depend on individual tastes as well as many factors, like public childcare facilities and the labor market situation. In the present survey, we want to learn more about your ideas concerning fertility and labor-supply.

Imagine you lived together with a partner, and you were in the situation of planning a family. Particularly, you had to decide how many children you wanted to have altogether, and how many hours you wanted to work. We will ask you to make this decision in several different environments. Each environment is described by:

- the availability of publicly-provided childcare
- child related tax benefits and subsidies
- the number of working hours of your partner
- the earnings of the partner
- your own net wage

B.3.2 Visual design of the vignettes

Each environment is presented along with the following instruction:

> Please imagine that the environment is described as below. In the empty fields you can type in the desired number of children. Using the slider, you can choose the desired numbers of working hours. For your convenience, a calculator at the bottom of the screen gives the level of household net earnings plus child related tax benefits and subsidies (when children are present). The number does not include social-assistance, unemployment benefits or other kinds of revenues.

Following the instructions, the first environment is provided, and the respondents are asked to fill-in the desired number of working hours (by means of a slider) and children (typing a number in the indicated empty field). Individual earnings, child related transfers, and the total household income (sum), displayed at the bottom, are immediately and automatically adapted to working hours and desired children as stated by the respondent.

External child care: private responsibility and costly **Tax benefits and subsidies per month and child: $200**	
Partner's number of working hours per month: 160 **Partner's net earnings per month: $1600**	
Your hourly net wage rate: $5 **Ideal number of own working hours per month:** 0 40 80 120 160	**Note. In case you want to have children, assume that the youngest child is two years old.**
Household composition: 2 adults **Numbers of children you would like to have:**	
Partner's net earnings per month: $1600 **Your net earnings:** ***Automatically computed*** **Child transfer:** ***Automatically computed*** **Sum:** ***Automatically computed***	

Once the respondent has filled-in the desired working hours and number of children for the first environment, the second environment is provided. It differs from the first environment with respect to a single attribute, which is highlighted. The following figure shows a prototype sequence of four environments.

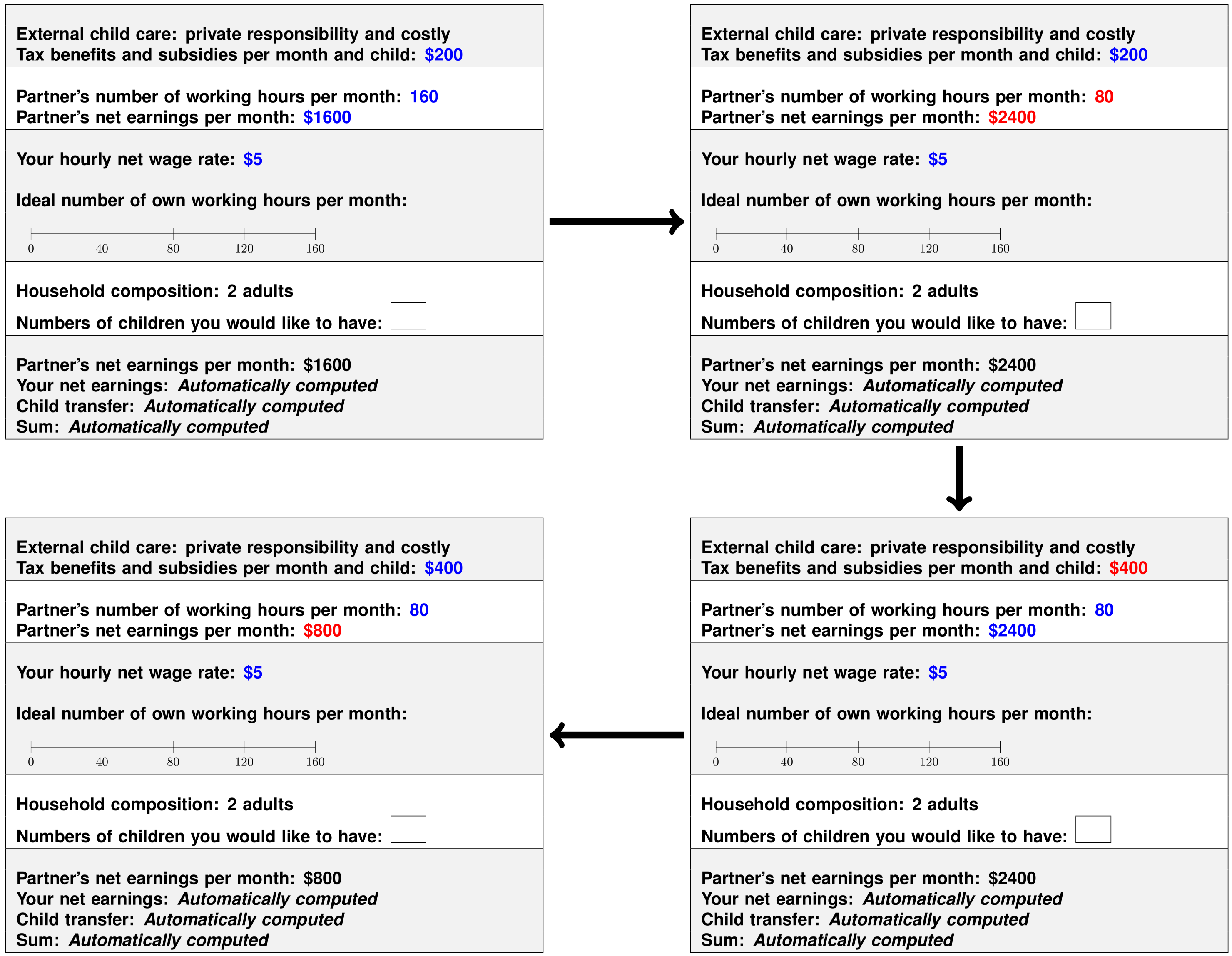


Figure B.1: Prototype sequence of four vignette environments presented to a respondent. Across consecutive environments, exactly one of the five attributes is changed (highlighted in the display); the respondent provides the desired number of working hours and children for each environment.

C Analytics and computation of Models 1 and 2

C.1 Model 1

Model 1 is a continuous-time model that can be characterized analytically to a great extent. Therefore, we present the analytics and, based on the solutions, we proceed to explaining how it is computed.

C.1.1 Analytics of Model 1

Because of the time separability of the objective function of the model, one can write two Hamiltonian functions, one for Phase 1 and one for Phase 2. The intra-temporal conditions of the problem lead to,

$$\frac{1-\theta}{\theta}\frac{C_A}{1-\underline{\ell}-\ell}=w \ . \tag{C.1}$$

Denoting utility $u^{\text{no children}}(C_A,\ell)$ by u_1,

$$u_1\equiv\left(\frac{C_A}{\sqrt{N_1}}\right)^{\theta}(1-\underline{\ell}-\ell)^{1-\theta} \ , \tag{C.2}$$

the combination of (C.1) and (C.2) gives,

$$u_1^{1-\gamma}=\kappa_4 w^{-\delta_1}C_A^{1-\gamma} \ , \tag{C.3}$$

in which κ_4 is a constant and,

$$\delta_1=(1-\theta)(1-\gamma) \ . \tag{C.4}$$

The intratemporal conditions of the problem (after some algebra involving taking derivatives with respect to time) give,

$$\frac{\dot{C}_A}{C_A}=\frac{1}{\gamma}\left(r-\rho-\delta_1\frac{\dot{w}}{w}\right) \ . \tag{C.5}$$

Solving (C.5) forward, we have,

$$C_A(s)=e^{\frac{1}{\gamma}(r-\rho)(s-t)}\left[\frac{w(s)}{w(t)}\right]^{-\delta_1}C_A(t) \ . \tag{C.6}$$

Rearranging (C.1) we get,

$$w\ell=(1-\underline{\ell})\,w-\frac{1-\theta}{\theta}C_A \ . \tag{C.7}$$

Combining (C.7) with equation (2) in the main body of the paper, and (C.6) gives,

$$\dot{a}(s) - ra(s) = \bar{w}_1(s) - \frac{1}{\theta} e^{\frac{1}{\gamma}(r-\rho)(s-t)} \left[\frac{w(s)}{w(t)}\right]^{-\delta_1} C_A(t) \ , \tag{C.8}$$

in which,

$$\bar{w}_1(t) \equiv y_p(t) + (1-\underline{\ell})\, w(t) \ . \tag{C.9}$$

Multiplying both sides of (C.8) by the integrating factor $e^{-r(s-t)}$ and integrating from t to $\bar{t}$ (for $t \leq \bar{t}$) gives,

$$e^{-r(\bar{t}-t)} a(\bar{t}) - a(t) = \bar{\omega}_1(t) - \bar{\xi}_1(t)\, C_A(t) \ , \quad t \leq \bar{t} \tag{C.10}$$

in which

$$\bar{\omega}_1(t) \equiv \int_t^{\bar{t}} e^{-r(s-t)} \bar{w}_1(s)\, ds \ , \tag{C.11}$$

and

$$\bar{\xi}_1(t) \equiv \int_t^{\bar{t}} e^{\left[\frac{1}{\gamma}(r-\rho)-r\right](s-t)} \left[\frac{w(s)}{w(t)}\right]^{-\delta_1} ds \ . \tag{C.12}$$

Equation (C.10) is crucial for deriving the consumption function, but in order to do this we need to connect consumption decisions in Phase 1 with consumption decisions in Phase 2. This connection is established below in equation (C.42), which expresses Phase-1 consumption $C_A(t)$ as a function of the boundary asset value $a(\bar{t})$ and Phase-2 anticipations. Finally, the optimal labor-supply decision for $t \leq \bar{t}$ is given after re-arranging equation (C.7) by,

$$\ell(t) = 1 - \underline{\ell} - \frac{1-\theta}{\theta} \frac{C_A(t)}{w(t)} \ , \ \ t \leq \bar{t} \ . \tag{C.13}$$

In contrast to phase (1), in phase (2) there is a larger number of decisions. Because the decision functions are complicated, we use recursive numerical solution. Yet, several analytical results facilitate our analysis, so we explicitly derive them and explain them here. The Hamiltonian in Phase 2 is,

$$\mathcal{H}_2 = e^{-\rho(t-\bar{t})} \frac{u_2^{1-\gamma}}{1-\gamma} + \lambda_2 \left[ra + y_p + w\ell + \left(A - \kappa w^{\phi} - \tilde{p}_b m_b\right) n - C_C - C_A\right]$$

in which,

$$u_2 \equiv \left\{ \left[\left(\frac{C_A}{\sqrt{N_2}}\right)^{\theta_1} \left(\frac{C_C}{\sqrt{n}}\right)^{1-\theta_1} \right]^{\theta} \left(1 - \underline{\ell} - \underline{m} - m - \ell\right)^{1-\theta} \right\}^{\alpha} \left[n^{\psi} \left(m + \zeta m_b^{\eta}\right)^{1-\psi}\right]^{1-\alpha} \ . \tag{C.14}$$

The first-order conditions are,

$$\alpha\theta\theta_1 \frac{e^{-\rho(t-\bar{t})}u_2^{1-\gamma}}{C_A} = \lambda_2 \ , \tag{C.15}$$

$$\alpha\theta\left(1-\theta_1\right)\frac{e^{-\rho(t-\bar{t})}u_2^{1-\gamma}}{C_C} = \lambda_2 \ , \tag{C.16}$$

$$\alpha\left(1-\theta\right)\frac{e^{-\rho(t-\bar{t})}u_2^{1-\gamma}}{1-\underline{\ell}-\underline{m}-m-\ell} = \lambda_2 w \ , \tag{C.17}$$

$$\frac{m+\zeta m_b^{\eta}}{1-\underline{\ell}-\underline{m}-m-\ell} = \frac{\left(1-\alpha\right)\left(1-\psi\right)}{\alpha\left(1-\theta\right)} \ , \tag{C.18}$$

$$\left[\alpha\theta\frac{1-\theta_1}{2}-\left(1-\alpha\right)\psi\right]\frac{e^{-\rho(t-\bar{t})}u_2^{1-\gamma}}{n} = \lambda_2\left(A-\kappa w^{\phi}-\tilde{p}_b m_b\right) \ , \tag{C.19}$$

$$\left(1-\alpha\right)\left(1-\psi\right)\zeta\eta\frac{e^{-\rho(t-\bar{t})}u_2^{1-\gamma}m_b^{\eta-1}}{m+\zeta m_b^{\eta}} = \lambda_2\tilde{p}_b n \ , \tag{C.20}$$

$$-\frac{\dot{\lambda}_2}{\lambda_2} = r \ , \tag{C.21}$$

together with the budget constraint given by equation (13) in the main body of the paper. A crucial analytical step is to relate all decisions to C_A and m_b. Once this step is achieved, we can insert all decisions in the budget constraint given by equation (13), and solve the differential equation for asset accumulation in the same fashion as we did with solving equation (C.8) in Phase 1, creating a recursion that iterates on solving m_b as an optimal function of time.

Combining (C.15) and (C.16) yields,

$$C_C = \frac{1-\theta_1}{\theta_1}C_A \ , \tag{C.22}$$

while combining (C.15) and (C.19) gives,

$$\left(A-\kappa w^{\phi}-\tilde{p}_b m_b\right)n = \frac{\alpha\theta\frac{1-\theta_1}{2}-\left(1-\alpha\right)\psi}{\alpha\theta\theta_1}C_A \ . \tag{C.23}$$

Combining (C.15), (C.17), and (C.18) gives,

$$m+\zeta m_b^{\eta} = \frac{\left(1-\alpha\right)\left(1-\psi\right)}{\alpha\theta\theta_1}w^{-1}C_A \ . \tag{C.24}$$

Combining (C.17), (C.20), and (C.18) leads to,

$$n = \zeta \eta w \tilde{p}_b^{-1} m_b^{\eta-1} \ , \tag{C.25}$$

and combining (C.25) with (C.23) yields a key equation that pins down $m_b(t)$ as an implicit function of $C_A(t)$, given the exogenous wage and price processes; this is the relationship we iterate on in Section 1.2 to compute the equilibrium $m_b(t)$ schedule,

$$m_b^{\eta} = \frac{A - \kappa w^{\phi}}{\tilde{p}_b} m_b^{\eta-1} - \frac{\alpha \theta^{\frac{1-\theta_1}{2}} - (1-\alpha)\psi}{\alpha \theta \theta_1 \eta \zeta} w^{-1} C_A \ , \tag{C.26}$$

which is the nonlinear equation we will be using in order to iterate on function $m_b(t)$. Equations (C.26) and (C.24) give

$$m = -\zeta \frac{A - \kappa w^{\phi}}{\tilde{p}_b} m_b^{\eta-1} + \frac{\alpha \theta^{\frac{1-\theta_1}{2}} + (1-\alpha)\left[\eta - (1+\eta)\psi\right]}{\alpha \theta \theta_1 \eta} w^{-1} C_A \ , \tag{C.27}$$

and after combining (C.27), (C.15), and (C.17), we finally obtain,

$$\ell = 1 - \underline{\ell} - \underline{m} + \zeta \frac{A - \kappa w^{\phi}}{\tilde{p}_b} m_b^{\eta-1} - \frac{\alpha \left[\eta(1-\theta) + \theta^{\frac{1-\theta_1}{2}}\right] + (1-\alpha)\left[\eta - (1+\eta)\psi\right]}{\alpha \theta \theta_1 \eta} w^{-1} C_A \ . \tag{C.28}$$

Substituting equations (C.22) through (C.28) into equation (13) in the main body of the paper, we obtain,

$$\dot{a}(t) - r a(t) = \bar{w}_2(t) - \kappa_3 C_A(t) \ , \tag{C.29}$$

in which,

$$\bar{w}_2(t) \equiv y_p(t) + \left[1 - \underline{\ell} - \underline{m} + \zeta \frac{A - \kappa w(t)^{\phi}}{\tilde{p}_b(t)} m_b(t)^{\eta-1}\right] w(t) \ , \tag{C.30}$$

and

$$\kappa_3 \equiv \frac{1}{\theta_1} + \frac{(1-\eta)\alpha \theta^{\frac{1-\theta_1}{2}} + \alpha \eta (1-\theta) + (1-\alpha)(\eta - \psi)}{\alpha \theta \theta_1 \eta} \ . \tag{C.31}$$

In order to solve equation (C.29), the next step is to deal with the intertemporal condition given by (C.21). The key is to work on (C.15), by taking logs on both sides of (C.15) and taking a derivative with respect to time. To do so, we substitute equations (C.22) through (C.28) into (C.14) in order to obtain,

$$u_2(t)^{1-\gamma} = \kappa_1 f(t) C_A(t)^{1-\gamma_2} \ , \tag{C.32}$$

in which κ_1 is a constant and,

$$f(t) = w(t)^{\delta_2}\, \tilde{p}_b(t)^{\delta_3}\, m_b(t)^{\delta_4} \quad , \tag{C.33}$$

with,

$$\begin{aligned} \delta_2 &= \left\{\alpha\left[\theta\left(1+\frac{\theta_1}{2}\right)-1\right]+(1-\alpha)(2\psi-1)\right\}(1-\gamma) \quad , \\ \delta_3 &= \left[\alpha\theta\frac{1-\theta_1}{2}-\psi(1-\alpha)\right](1-\gamma) \quad , \\ \delta_4 &= (1-\eta)\left[\alpha\theta\frac{1-\theta_1}{2}-\psi(1-\alpha)\right](1-\gamma) \quad . \end{aligned} \tag{C.34}$$

Substituting (C.32) into (C.15), taking logs on both sides, differentiating with respect to time we obtain,

$$\frac{\dot{C}_A}{C_A} = \frac{1}{\gamma}\left(r-\rho+\frac{\dot{f}}{f}\right) \quad ,$$

which has solution,

$$C_A(s) = e^{\frac{1}{\gamma}(r-\rho)(s-t)}\left[\frac{f(s)}{f(t)}\right]^{\frac{1}{\gamma}} C_A(t) \quad , \quad s \geq t \ . \tag{C.35}$$

Considering equation (C.29) at any instant $s \in [t, T]$ with $t \geq \bar{t}$, and after substituting (C.35), we obtain,

$$\dot{a}(s) - ra(s) = \bar{w}_2(s) - \kappa_3 e^{\frac{1}{\gamma}(r-\rho)(s-t)}\left[\frac{f(s)}{f(t)}\right]^{\frac{1}{\gamma}} C_A(t) \quad , \quad s \geq t \ ,$$

and we can multiply both sides of this last equation by the integrating factor $e^{-r(s-t)}$, integrating from t until T to obtain,

$$e^{-r(T-t)}a(T) - a(t) = \int_t^T e^{-r(s-t)}\bar{w}_2(s)\,ds - \kappa_3\int_t^T e^{-r(s-t)}e^{\frac{1}{\gamma}(r-\rho)(s-t)}\left[\frac{f(s)}{f(t)}\right]^{\frac{1}{\gamma}} ds \cdot C_A(t) \quad ,$$

which provides the decision rule for consumption,

$$C_A(t) = \xi_2(t)\left[a(t) - e^{-r(T-t)}a(T) + \omega_2(t)\right] \quad , \tag{C.36}$$

in which,

$$\xi_2(t) \equiv \left\{\kappa_3\int_t^T e^{\left[\frac{1}{\gamma}(r-\rho)-r\right](s-t)}\left[\frac{f(s)}{f(t)}\right]^{\frac{1}{\gamma}} ds\right\}^{-1} \quad , \tag{C.37}$$

and

$$\omega_2(t) \equiv \int_t^T e^{-r(s-t)} \bar{w}_2(s)\, ds \ . \tag{C.38}$$

To connect consumption decisions between Phase 1 and Phase 2, as equation (C.10) indicates, the optimal consumption decision for $t \le \bar{t}$ is conditional upon the level of $a(\bar{t})$. However, $a(\bar{t})$ is not one of the terminal conditions of the problem, it is endogenous. We need to connect the progression of wealth, $a(t)$, between the two life phases, and to compute consumption in Phase 1 as a function of anticipations in Phase 2 as well. In order to achieve this, consider equation (C.36) at time $\bar{t}$,

$$C_A(\bar{t}) = \xi_2(\bar{t}) \left[a(\bar{t}) - e^{-r(T-\bar{t})} a(T) + \omega_2(\bar{t}) \right] \ ,$$

which we can solve for $a(\bar{t})$ to obtain,

$$a(\bar{t}) = \frac{1}{\xi_2(\bar{t})} c(\bar{t}) - \omega_2(\bar{t}) + e^{-r(T-\bar{t})} a(T) \ . \tag{C.39}$$

Combining (C.10) with (C.39) we obtain,

$$e^{-r(\bar{t}-t)} \left[\frac{1}{\xi_2(\bar{t})} C_A(\bar{t}) - \omega_2(\bar{t}) + e^{-r(T-\bar{t})} a(T) \right] - a(t) = \bar{\omega}_1(t) - \bar{\xi}_1(t)\, c(t) \ . \tag{C.40}$$

Equation (C.6) implies,

$$C_A(\bar{t}) = e^{\frac{1}{1-\alpha_1(1-\gamma)}(r-\rho)(\bar{t}-t)} \left[\frac{w(\bar{t})}{w(t)} \right]^{-\delta_1} c(t) \ ,$$

which we can combine with (C.40) to obtain,

$$\left\{ e^{\left[\frac{1}{1-\alpha_1(1-\gamma)}(r-\rho)-r\right](\bar{t}-t)} \frac{1}{\xi_2(\bar{t})} \left[\frac{w(\bar{t})}{w(t)} \right]^{-\delta_1} + \bar{\xi}_1(t) \right\} C_A(t) =$$

$$= a(t) - e^{-r(T-t)} a(T) + \bar{\omega}_1(t) + e^{-r(\bar{t}-t)} \omega_2(\bar{t}) \ . \tag{C.41}$$

Equation (C.41) reveals the form of the consumption function for $t \le \bar{t}$, namely,

$$C_A(t) = \xi_1(t) \left[a(t) - e^{-r(T-t)} a(T) + \omega_1(t) \right] \ , \ \ t \le \bar{t} \ , \tag{C.42}$$

in which

$$\xi_1(t) \equiv \left\{ e^{\left[\frac{1}{\gamma}(r-\rho)-r\right](\bar{t}-t)} \frac{1}{\xi_2(\bar{t})} \left[\frac{w(\bar{t})}{w(t)} \right]^{-\delta_1} + \bar{\xi}_1(t) \right\}^{-1} \ , \tag{C.43}$$

and

$$\omega_1(t) = \bar{\omega}_1(t) + e^{-r(\bar{t}-t)}\omega_2(\bar{t}) \ . \tag{C.44}$$

To summarize, equations (C.42) and (C.36) comprise a decision rule which is a branch function given by,

$$C_A(t) = \begin{cases} \xi_1(t)\left[a(t) - e^{-r(T-t)}a(T) + \omega_1(t)\right] & , \quad t \leq \bar{t} \\ \xi_2(t)\left[a(t) - e^{-r(T-t)}a(T) + \omega_2(t)\right] & , \quad t \geq \bar{t} \end{cases} \ . \tag{C.45}$$

in which terms $\xi_1(t)$, $\xi_2(t)$, $\omega_1(t)$, and $\omega_2(t)$ are given by (C.43), (C.37), (C.44), and (C.38).

C.1.2 Strategy for computation of Model 1

Regarding the computation of equilibrium, there is a key facility stemming from the fact that setting $\eta = 1$ leads to an exact solution. It suffices to see equation (C.26), in which setting $\eta = 1$ allows $m_b(t)$ to have a closed form, as a function of $C_A(t)$ only (all other variables, $w(t)$ and $\tilde{p}_b(t)$ are exogenous, given by the data). At $\eta = 1$, equations (C.33) and (C.34) further imply that function $f(t)$ no longer depends on endogenous variable $m_b(t)$, but only on $w(t)$ and $\tilde{p}_b(t)$. This means that, for the special case of $\eta = 1$, the branch function (C.45) has an exact solution, allowing the model to be fitted to data conveniently using minimum-distance estimation.

For $\eta \neq 1$, our computational strategy is to first restrict the model to $\eta = 1$, in order to obtain model parameters that can serve as first guesses for minimum-distance estimation for the case of $\eta \neq 1$. The second step is to fit a polynomial function to the target data for $m_b(t)$, use the parameters of the first step and best-fit the rest of the variables of the model to the data for $\eta \neq 1$. The third step is to derive a new guess for $m_b(t)$, denoted by $m_b^{NEW}(t)$, after using equation (C.26) at the estimated parameters of the second step. Taking a convex combination between $m_b^{OLD}(t)$, the polynomial function for $m_b(t)$ from the previous iteration, and $m_b^{NEW}(t)$, we go back to the third step and continue until convergence, using the analytical features described above in this Appendix for faster computation.

C.2 Model 2

Here we describe the structure of the algorithm for our non-aggregative model, which respects cross-sectional heterogeneity. As the algorithm builds on Adda et al. [2017] and Wang [2022], which provide detailed discussions of the underlying model structure and solution methods, we focus on the novel model features.

C.2.1 Per-period utility

The per-period utility function of Model 2, given in eq. (18) of the main body, takes the form

$$u_t\big(\ell_t, \tilde{c}_t, n_t, \mathrm{mar}_t, a_t\big) \;=\; \Bigg\{$$

$$\alpha_1 \left\{ \ln\left[\underbrace{\tilde{c}_t - n_t(1-s)p_{b,t}m_{b,t} + n_t\cdot A}_{c_t} + \alpha_3(n_t)\right] - \ln\left(\underbrace{1 + 0.5\,\mathrm{mar}_t + 0.3\,n_t}_{\text{OECD equivalence scale}}\right)\right\}$$

$$+\big(1 - O_1\alpha_6\big)\left\{\sum_{k=1}^{3}\alpha_{2k}n_t\mathbb{I}_t^{\{n_t=k\}} + (1-\mathrm{mar}_t)\left[\alpha_4\ln\left(\sum_{k=1}^{3}n_t\mathbb{I}_t^{\{n_t=k\}}+1\right) + \alpha_5\mathbb{I}_t^{\{n_t=1\}}\right]\right\}$$

$$-\left\{ O_3^L\left[\mathbb{I}_t^{\{\ell_t=\ell^f\}} + \gamma_4\,\mathbb{I}_t^{\{\ell_t=\ell^p\}}\right] + O_3^C\underbrace{\sum_{j=1}^{5}\zeta_j\mathbb{I}_t^{\{a_t(j)=j\}}}_{\Gamma_1}\left[\big(1-\mathrm{mar}_t(1-\phi_b)\big)\bar{\ell}^c(a_t,n_t) - n_t m_{b,t}\right]\right\}$$

$$\times\underbrace{\left[\gamma_1 + (1-O_2\gamma_5)\left(\sum_{k=1}^{3}\gamma_{2k}n_t\mathbb{I}_t^{\{n_t=k\}}\right)\left(\sum_{j=1}^{5}\gamma_{3j}\mathbb{I}_t^{\{a_t(j)=j\}}\right)\right]}_{\Gamma}\;\Bigg\}, \tag{C.46}$$

The first line of the equation is the consumption block. Here $\tilde{c}_t$ denotes gross consumption before netting the external-childcare cost, with $c_t = \tilde{c}_t - n_t(1-s)p_{b,t}m_{b,t} + n_t \cdot A$ the net household consumption that enters the logarithm: gross consumption $\tilde{c}_t$, less the parent-paid childcare share $n_t(1-s)p_{b,t}m_{b,t}$, plus the per-child allowance $n_t \cdot A$; the function $\tilde{c}_t$ is detailed in Section C.2.5 below. The term $\alpha_3(n_t)$ is a consumption shifter that depends on the number of children, and the denominator $1 + 0.5\,\mathrm{mar}_t + 0.3\,n_t$ is the OECD equivalence scale that adjusts for household composition.

The second line is the children-utility block. The sum $\sum_{k=1}^{3}\alpha_{2k}n_t\mathbb{I}_t^{\{n_t=k\}}$ activates exactly one of the parameters $\alpha_{21}, \alpha_{22}, \alpha_{23}$ depending on the family size: when $n_t = k$, the term evaluates to $k\cdot\alpha_{2k}$, so that α_{2k} is the *per-child utility weight* at family size k. The single-parent supplementary term $(1-\mathrm{mar}_t)\big[\alpha_4\ln(n_t+1) + \alpha_5\mathbb{I}_t^{\{n_t=1\}}\big]$ captures the additional utility (or disutility) experienced by single mothers, where the inner indicator-sum $\sum_{k=1}^{3} n_t \mathbb{I}_t^{\{n_t=k\}} + 1$ simply evaluates to $n_t + 1$ and serves as the argument of the logarithm. The whole block is multiplied by the heterogeneity discount $(1 - O_1\alpha_6)$, where the binary type $O_1 \in \{0, 1\}$ indexes a woman's *fertility preference*: type $O_1 = 0$ has *high fertility preference* and retains the full children-utility, while type $O_1 = 1$ has *low fertility preference* and retains only a fraction $(1 - \alpha_6)$ of full children-utility. This high/low fertility preference structure parallels the type-specific preference convention introduced in Wang [2022]. The population shares of the two types are governed by the type-mix probability π_{O_1} (estimated within country-specific bounds; see Section 4.3 of the main body and Appendix D).

The third line is the work-disutility block. It is the sum of two pieces: a labor-time piece scaled by O_3^L, and a childcare-time piece scaled by O_3^C. The labor-time piece uses indicators only, $\mathbb{I}_t^{\{\ell_t=\ell^f\}} + \gamma_4 \mathbb{I}_t^{\{\ell_t=\ell^p\}}$, where the discrete labor choice is $\ell_t \in \{0, \ell^p, \ell^f\}$ and $\gamma_4 \in (0, 1)$ scales the part-time disutility relative to full-time. The childcare-time piece is the residual parental time $[1 - \mathrm{mar}_t(1 - \phi_b)]\bar{\ell}^c(a_t, n_t) - n_t m_{b,t}$, where $\bar{\ell}^c(a_t, n_t)$ is the calibrated parental childcare-time function (Appendix D), and the bracket $[1 - \mathrm{mar}_t(1 - \phi_b)]$ is the marital-status adjustment (married households apply factor ϕ_b; single households apply factor 1). The total external-childcare time $n_t m_{b,t}$ enters as a continuous optimization variable when $n_t > 0$, with the resulting first-order condition derived in Section 2.2.

The fourth line is the multiplicative scaling factor Γ that applies to the entire work-disutility block. It is built from a constant γ_1, plus the heterogeneity-discounted product of two sums: birth-order weights $\sum_{k=1}^{3} \gamma_{2k} n_t \mathbb{I}_t^{\{n_t=k\}}$ and youngest-child-age weights $\sum_{j=1}^{5} \gamma_{3j} \mathbb{I}_t^{\{a_t(j)=j\}}$, the product of which is scaled by the heterogeneity-discount factor $(1 - O_2\gamma_5)$. The binary type $O_2 \in \{0, 1\}$ indexes a woman's *disutility of working with children*: type $O_2 = 0$ has *high work disutility* and applies the full work-disutility scaling, while type $O_2 = 1$ has *low disutility for working with children* and retains only a fraction $(1 - \gamma_5)$ of the multiplicative work-disutility, paralleling the low-disutility-type convention in Wang [2022]. The auxiliary aggregate $\Gamma_1 = \sum_{j=1}^{5} \zeta_j \mathbb{I}_t^{\{a_t(j)=j\}}$ that appears inside the third-line childcare bracket is the corresponding ζ-weighted sum across youngest-child age bins. The branch function $a_t(k)$ partitioning youngest-child age into five bins is defined in eq. (19) of the main body. The four types $(O_1, O_2) \in \{0, 1\}^2$ together form a four-cell population, with type-mix probabilities π_{O_1}, π_{O_2} estimated within country-specific bounds (Section 4.3 of the main body; Appendix D).

C.2.2 The introduction of total external childcare time as a continuous variable

The only continuous control variable in the per-period optimization is $m_{b,t}$, the external-childcare time per child, available when $n_t > 0$. We derive its first-order condition by maximizing the per-period utility, eq. (C.46), with respect to $m_{b,t}$, treating all other variables as fixed. The resulting expression for $m_{b,t}$ then enters the residual parental-time quantity $[1-\text{mar}_t(1-\phi_b)]\bar{\ell}^c(a_t,n_t)-n_t m_{b,t}$ that appears in the work-disutility block of eq. (C.46).

Focusing on the case where $n_t>0$, the first-order conditions for maximizing with respect to $m_{b,t}$ (demand for external childcare time) are:

$$\frac{\alpha_1 n_t\,(1-s)\,p_{b,t}}{\tilde{c}_t-n_t\,(1-s)\,p_{b,t}m_{b,t}+n_t\cdot A+\alpha_3\,(n_t)}=O_3^C\cdot\Gamma\cdot\Gamma_1\cdot n_t\ ,$$

and solving for $m_{b,t}$ gives,

$$n_t m_{b,t}=\frac{\tilde{c}_t+n_t\cdot A+\alpha_3\,(n_t)-\frac{\alpha_1(1-s)p_{b,t}}{O_3^C\cdot\Gamma\cdot\Gamma_1}}{(1-s)\,p_{b,t}}$$

The expression above is the unconstrained interior-FOC solution. In addition, the optimal external-childcare time satisfies two inequality constraints from the model: (i) non-negativity, $n_t m_{b,t}\geq 0$; and (ii) an outsourcing cap that prevents the focal parent from outsourcing more than a fraction $\phi(\text{mar}_t,a_t)$ of the parental childcare time, i.e.,

$$0\ \leq\ n_t m_{b,t}\ \leq\ \phi(\text{mar}_t,a_t)\cdot\bar{\ell}^c(a_t,n_t)\,,$$

where $\phi(\text{mar}_t,a_t)$ collapses the time-varying outsourcing cap and the marriage adjustment into a single composite factor,

$$\phi(\text{mar}_t,a_t)\ \equiv\ \alpha(a_t)\cdot\big[1-\text{mar}_t(1-\phi_b)\big]\,,\qquad \alpha(a_t)\ =\ \max\big\{0,\ 0.856-0.044\,a_t\big\}\,.\tag{C.47}$$

The cap $\alpha(a_t)$ is decreasing in the age of the youngest child (older children require more non-outsourceable parental time), and the marriage factor $[1-\text{mar}_t(1-\phi_b)]$ scales the focal parent's share of total parental childcare (ϕ_b if married, 1 if single). Combining the interior FOC with these bounds gives the closed-form total external childcare time used in the estimation:

$$n_t m_{b,t}=\begin{cases}\min\Big\{\max\Big\{\dfrac{\tilde{c}_t+n_t\cdot A+\alpha_3(n_t)}{(1-s)p_{b,t}}-\dfrac{\alpha_1}{O_3^C\cdot\Gamma\cdot\Gamma_1},\ 0\Big\},\ \phi(\text{mar}_t,a_t)\,\bar{\ell}^c(a_t,n_t)\Big\} & ,\quad \text{if } n_t>0\\ 0 & ,\quad \text{if } n_t=0\end{cases}\ ,\tag{C.48}$$

with the per-child external childcare time given by $m_{b,t}=n_t m_{b,t}/n_t$ when $n_t>0$ and $m_{b,t}=0$ when $n_t=0$. This is identical to eq. (20) of the main body, where $\phi(\text{mar}_t,a_t)$ is defined inline in the same form. The optimal external-childcare time $n_t m_{b,t}$ in turn determines the residual *individual parental time* that appears in the work-disutility block of eq. (C.46). Specifically,

$$\text{individual parental time} \;=\; \big[1-\text{mar}_t(1-\phi_b)\big]\,\bar{\ell}^c(a_t,n_t)-n_t m_{b,t}\,, \tag{C.49}$$

where $n_t m_{b,t}$ is given by the clamped closed form (C.48). When $n_t=0$, both terms are zero. This is the individual parental time entering the disutility block of eq. (C.46), and matches eq. (21) of the main body.

C.2.3 Bellman equation and backward induction during the fertile and nonfertile period

This subsection collects the Bellman equations of Model 2, complementing the deterministic-max forms presented in Section 4.1 (nonfertile period, ages 46–62) and Section 4.2 (fertile period, ages 22–45) of the main body. The contribution of this section is to write out the explicit Gumbel-smoothed log-sum (Emax) forms that are actually solved in the estimation: the labor-supply choice is smoothed by an i.i.d. Type-I extreme-value preference shock of estimated scale σ_ε, and the stage-1 fertility choice in the fertile period is similarly smoothed by an analogous shock of scale σ_ε^f. The deterministic-max Bellman forms in the main body are the $\sigma_\varepsilon \to 0$ and $\sigma_\varepsilon^f \to 0$ limits. Estimation values for both smoothing scales, together with the calibration of the fecundity, marital, divorce, and job-finding logits, are documented in Appendix D.

The backward-induction algorithm runs from age $T=62$ down to $t=22$, with terminal continuation value $V_{63}\equiv 0$. At each $t\in\{46,\dots,62\}$, the value function $V_t(\Omega_t^0)$ on the reduced nonfertile state space (defined in main body Section 4.1) is computed from the Bellman recursion below, using the already-solved V_{t+1}. At each $t\in\{22,\dots,45\}$, the algorithm proceeds in two stages within the period: stage 2 (labor-supply choice) computes $V_t(\Omega_t)$ given the post-fertility state, and stage 1 (fertility choice) integrates over $f_t\in\{0,1\}$ to produce $V_t^*(\Omega_t^*)$, using the already-solved V_{t+1}^*. At the regime boundary $t=46$, we adopt the simplifying convention used in Wang [2022] that all children have left the household by age 46, so the continuation value entering the fertile-period recursion at $t=45$ evaluates V_{46} at $n_{46}=0$ and $a_{46}=\emptyset$ regardless of the fertility outcome at $t=45$.

Nonfertile period (ages 46–62). The reduced state space is

$$\Omega_t^0=(x_t,n_t{=}0,a_t{=}\emptyset,\text{mar}_t,y_{p,t},job_t,\xi_t)$$

(main body Section 4.1). Define the period-t action value

$$Q_t^0(\ell;\Omega_t^0) \;\equiv\; u_t(\ell,\tilde{c}_t,0,\mathrm{mar}_t,\emptyset) + \beta\, E\big[V_{t+1}(\Omega_{t+1}^0) \;\big|\; \ell,\Omega_t^0\big] \tag{C.50}$$

for each labor alternative $\ell \in \{0,\ell^p,\ell^f\}$. The deterministic-max Bellman recursion is given in main body Section 4.1. In the estimated specification, the labor-supply choice carries an i.i.d. Type-I extreme-value preference shock $\varepsilon_{\ell,t}$ of estimated scale σ_ε; marginalizing over the shock yields the standard log-sum (Emax) closed form

$$V_t(\Omega_t^0) \;=\; p_t^j\,\sigma_\varepsilon\,\log\left(\sum_{\ell\in\{0,\ell^p,\ell^f\}} \exp\frac{Q_t^0(\ell;\Omega_t^0)}{\sigma_\varepsilon}\right) + \big(1-p_t^j\big)\,Q_t^0(0;\Omega_t^0)\,. \tag{C.51}$$

The conditional choice probability of labor alternative ℓ given a job offer ($job_t = 1$) is the standard softmax,

$$P\big(\ell_t=\ell \;\big|\; \Omega_t^0, job_t=1\big) \;=\; \frac{\exp\big(Q_t^0(\ell;\Omega_t^0)/\sigma_\varepsilon\big)}{\sum_{\ell'\in\{0,\ell^p,\ell^f\}}\exp\big(Q_t^0(\ell';\Omega_t^0)/\sigma_\varepsilon\big)}\,.$$

With probability $1-p_t^j$, the agent receives no job offer and is constrained to $\ell_t = 0$ (involuntary unemployment); no smoothing applies in this branch.

Fertile period, stage 2 (labor; ages 22–45). Within a period $t \in \{22,\ldots,45\}$ the timing splits into stage 1 (fertility) and stage 2 (labor), solved backward within the period. The post-fertility state is $\Omega_t^* = (x_t, n_t^*, a_t^*, \mathrm{mar}_t, y_{p,t}, job_t, \xi_t)$, related to the pre-fertility state Ω_t through $n_t = n_t^* + f_t$ and $a_t = (1-f_t)\,a_t^*$, with $f_t \in \{0,1\}$ the binary fertility decision (main body Section 4.2). Define the stage-2 action value

$$Q_t(\ell;\Omega_t) \;\equiv\; u_t(\ell,\tilde{c}_t,n_t,\mathrm{mar}_t,a_t) + \beta\, E\big[V_{t+1}^*(\Omega_{t+1}^*) \;\big|\; \ell,\Omega_t\big]\,. \tag{C.52}$$

With Gumbel preference shocks of scale σ_ε, the log-sum Bellman for the labor choice is

$$V_t(\Omega_t) \;=\; p_t^j\,\sigma_\varepsilon\,\log\left(\sum_{\ell\in\{0,\ell^p,\ell^f\}} \exp\frac{Q_t(\ell;\Omega_t)}{\sigma_\varepsilon}\right) + \big(1-p_t^j\big)\,Q_t(0;\Omega_t)\,, \tag{C.53}$$

with the same softmax conditional choice probability as in (C.51), evaluated at the fertile state Ω_t. The deterministic-max form (main body Section 4.2) is the $\sigma_\varepsilon \to 0$ limit.

Fertile period, stage 1 (fertility). In stage 1 the agent chooses fertility $f_t \in \{0,1\}$ subject to the exogenous fecundity probability p_t^f (the biological chance of conceiving, age-dependent and calibrated; see Appendix D). With i.i.d. Gumbel preference shocks of estimated scale σ_ε^f on each fertility alternative, marginalizing over the shock gives the log-sum form

$$V_t^*(\Omega_t^*) \;=\; p_t^f\, \sigma_\varepsilon^f\, \log\!\left(\sum_{f_t\in\{0,1\}} \exp\frac{V_t\big(\Omega_t \;\big|\; f_t, \Omega_t^*\big)}{\sigma_\varepsilon^f}\right) + \big(1-p_t^f\big)\, V_t\big(\Omega_t \;\big|\; f_t=0, \Omega_t^*\big)\;, \quad \text{(C.54)}$$

where $V_t(\Omega_t|f_t,\Omega_t^*)$ is the stage-2 value in (C.53) evaluated at the fertile state Ω_t implied by Ω_t^* and the realized f_t. The conditional fertility-choice probability when biologically fertile (the p_t^f event) is the binary softmax,

$$P\big(f_t = 1 \;\big|\; \Omega_t^*, \text{fertile}\big) \;=\; \frac{\exp\!\big(V_t(\Omega_t|f_t=1)/\sigma_\varepsilon^f\big)}{\exp\!\big(V_t(\Omega_t|f_t=1)/\sigma_\varepsilon^f\big) + \exp\!\big(V_t(\Omega_t|f_t=0)/\sigma_\varepsilon^f\big)}\,.$$

The deterministic-max counterpart in main body Section 4.2 is the $\sigma_\varepsilon^f \to 0$ limit.

Exogenous transitions. The expectation operators in (C.50) and (C.52) integrate over the same set of exogenous transitions in both period regimes: marital status $\text{mar}_{t+1} \in \{0,1\}$ via marriage probability p_t^m and divorce probability p_t^d (age-dependent logits calibrated as in main body Section 4.1); the persistent wage shock $\xi_{t+1} \sim \mathcal{N}(0,\sigma_\xi^2)$, integrated by Gauss–Hermite quadrature; and the job-status shock job_{t+1} via the job-finding probability p_t^j (logit in age, marital status, and demographic factors). The biological fecundity probability p_t^f enters only the stage-1 fertile-period recursion (C.54). All calibrated logits and the wage-shock variance are documented in Appendix D.

C.2.4 Computational details

We solve Model 2 by backward induction on a discretized state space and simulate population-level moments by Monte Carlo. Key discretization and simulation choices are summarized below; the underlying calibration of the auxiliary processes ($p_t^j, p_t^m, p_t^d, p_t^f$, the wage-shock variance, and the type-mix probabilities) is documented in Appendix D.

State-space discretization. The discrete state variables are kept in their native finite supports: $n_t \in \{0,1,2,3\}$ (children), $\text{mar}_t \in \{0,1\}$ (marital status), $job_t \in \{0,1\}$ (job-offer status). The continuous state variables are discretized as follows. Experience x_t takes values on a finite grid spanning the working-life range. The persistent wage shock $\xi_t \sim \mathcal{N}(0,\sigma_\xi^2)$ is approximated by an S-node Gauss–Hermite quadrature with weights $\{w_s\}$ and nodes $\{\xi^{(s)}\}$. Husband income $y_{p,t}$ is discretized to a K-node country-specific grid. Calibration of the grid sizes, the bounds, and σ_ξ appears in Appendix D.

Youngest-child age grid and interpolation. The utility function (C.46) treats the age of the youngest child a_t as falling into five discrete bins via the branch function $a_t(k)$

(eq. (19) of the main body). Because backward induction must respect the natural one-period transition $a_{t+1} = a_t + 1$, we represent a_t on the natural-number grid $\{0, 1, \ldots, 13\}$. Following Adda et al. [2017], value functions are computed at the bin representative points $\{0, 3, 6, 9, 12\}$ and linearly interpolated within each bin to the full natural-number grid; values at $a_t = 13$ are obtained by linear extrapolation from the slope at the upper bin.

Heterogeneous-population solution. The four unobserved-heterogeneity types $(O_1, O_2) \in \{0,1\}^2$ are solved as four independent dynamic programs, each conditioned on the type-specific utility-discount factors $(1 - O_1\alpha_6)$ and $(1 - O_2\gamma_5)$ in eq. (C.46). Simulated moments are aggregated to a population mixture

$$\begin{aligned} M(\theta) \;=\; & \pi_{O_1}\pi_{O_2}\, M^{(1,1)}(\theta) + \pi_{O_1}(1-\pi_{O_2})\, M^{(1,0)}(\theta) \\ & + (1-\pi_{O_1})\pi_{O_2}\, M^{(0,1)}(\theta) + (1-\pi_{O_1})(1-\pi_{O_2})\, M^{(0,0)}(\theta)\,, \end{aligned} \tag{C.55}$$

where $M^{(O_1,O_2)}(\theta)$ denotes the moment vector simulated from the type-(O_1, O_2) DP. The type-mix probabilities (π_{O_1}, π_{O_2}) are estimated within country-specific bounds (Section 4.3 of the main body; calibration in Appendix D).

Simulation. For each candidate parameter vector θ, we simulate $N_{\text{sim}} = 60{,}000$ lifecycle paths from age 22 to 62, with type assignments drawn independently from (π_{O_1}, π_{O_2}). The random seed is fixed at 123 to eliminate Monte-Carlo noise across optimizer iterations and to match the seed used in Jacobian-based standard-error computation (Section 4.4 of the main body). All simulation calls in the estimation pipeline use the same seed and N_{sim}.

C.2.5 Details about function $\tilde{c}_t$

The function $\tilde{c}_t$ in the utility function given by equation (18) in the main body describes the structure of personal incomes and their relation to policies.

Specifically,

$$\begin{aligned} \tilde{c}_t &= \tilde{C}\left(n_t, \text{mar}_t, y\left(\ell_t, x_t, \xi_t\right), y_{p,t}\right) \\ &= \max\Big[\, G(y\left(\ell_t, x_t, \xi_t\right), y_{p,t}, \text{mar}_t) + (1-\text{mar}_t)\cdot f^{\text{alimony}}(n_t)\,,\; f^{\text{Soc. Assist.}}(\text{mar}_t, n_t) \Big]\;, \end{aligned} \tag{C.56}$$

the function $f^{\text{alimony}}(n_t)$ is the alimony schedule,[69] given by

$$f^{\text{alimony}}(n_t) = 200 \cdot n_t \cdot 12 \cdot \kappa_c^{\text{alim}}\;,$$

[69] The OECD equivalence-scale adjustment, formerly applied as a denominator $1 + 0.5\,\text{mar}_t + 0.3\,n_t$ to $\tilde{c}_t$, now enters the utility function (C.46) separately as the additive term $-\alpha_1 \ln(1 + 0.5\,\text{mar}_t + 0.3\,n_t)$, so $\tilde{c}_t$ itself is not scale-divided.

where the factor 12 annualizes the underlying monthly benchmark of 200 EUR per child, and κ_c^{alim} is a country-specific scaling parameter calibrated separately for the US and Germany (Appendix D).[70] The function $f^{\text{Soc. Assist.}}(\text{mar}_t, n_t)$ is the social-assistance floor, given by

$$f^{\text{Soc. Assist.}}(\text{mar}_t, n_t) = \{1 + 0.75\,\text{mar}_t + [0.65 + 0.2\,(1 - \text{mar}_t)]\, n_t\} \cdot 600 \cdot 12 \cdot \kappa_c^{\text{SA}} \ ,$$

where the factor 12 annualizes the underlying monthly benchmark of 600 EUR for a single, childless household, and κ_c^{SA} is a country-specific scaling parameter calibrated separately for the US and Germany (Appendix D). The household resources before social assistance are

$$G(y_t, y_{p,t}, \text{mar}_t) = y_t + \text{mar}_t \cdot y_{p,t} \ ,$$

where $y_t = y(\ell_t, x_t, \xi_t)$ is the focal woman's own labor income,

$$y\,(\ell_t, x_t, \xi_t) = w\,(x_t) \cdot \ell_t \cdot \exp(\xi_t) \ ,$$

and $y_{p,t}$ is spousal labor income (relevant only when married; discretized on a country-specific grid; see Appendix D).

Policy variables. The pronatalist policy parameters do not enter $\tilde{c}_t$ directly; rather, they enter the post-netting consumption argument c_t that appears inside the logarithm of the utility function (C.46):

$$c_t = \tilde{c}_t - n_t(1 - s)\, p_{b,t}\, m_{b,t} + n_t \cdot A\,,$$

where A is the per-child allowance (annual, in the same monetary unit as $\tilde{c}_t$) and $s \in [0, 1]$ is the daycare subsidy rate, both treated as exogenous policy stimuli experiments of Section 6 of the main body. The calibrated baseline values of (A, s) for the US and Germany are documented in Appendix D.

C.3 Policy-response decomposition

Section 7 reports inconsistent policy rankings for the United States: at the structural estimates, Model 1 finds the daycare subsidy more cost-effective than the allowance, while Model 2 finds the allowance more cost-effective than the daycare subsidy (Table 17). This subsection decomposes the architectural reasons behind that disagreement.

[70] In the model calibration, $\kappa_{\text{DE}}^{\text{alim}} = \kappa_{\text{DE}}^{\text{SA}} = 1.0$ (the underlying 200 and 600 EUR monthly benchmarks apply at face value), and $\kappa_{\text{US}}^{\text{alim}} = \kappa_{\text{US}}^{\text{SA}} = 0.01$. The substantial cross-country gap reflects both the different monetary-unit convention used in the US estimation and the weaker formal alimony enforcement and lower social-assistance floors in the US relative to Germany.

The contradictory policy rankings for the United States can partly be explained by technical specifics of the modeling: Model 1 represents children as an aggregative continuous count and applies the daycare-cost reduction to *all* children, while Model 2 tracks each child's age explicitly and applies the daycare subsidy only to those of daycare age. Accordingly, Model 1 amplifies the daycare-response by the inverse of the share of dependent-children years that require intensive formal care. Conversely, allowances enter Model 1's interior FOC through a single time-cost term $p_n = -A + \kappa w^{\phi} + \tilde{p}_b m_b$ whose response is muted by the offsetting movement in m_b, whereas in Model 2 they enter discrete fertility decisions directly and have a sizeable extensive-margin effect.[71]

[71] See also the architectural comment in Section 6.3: the Model 1 allowance response in the US is small not because of a calibration artefact ($\hat{m}_b$ is within 1.5% of the data target 1/12) but because of the channel-cancellation mechanism in the time-cost identity $p_n = -A + \kappa w^{\phi} + \tilde{p}_b m_b$.

D Estimation and calibration of Models 1 and 2

D.1 Model 1

We structurally estimate Model 1 via the Method of Simulated Moments with two-step LBFGS for both the United States and Germany. The estimation matches age profiles of labor supply, the number of children, parental childcare time, log adult consumption, log child consumption, and log assets, together with a single calibration target for the mean external childcare time. Inference is by Jacobian sandwich standard errors.

D.1.1 Estimation procedure

Parameter vector. Model 1 has 14 structural parameters,

$$\Theta = [\rho,\ r,\ \theta,\ \alpha,\ \alpha_1,\ \gamma,\ \psi,\ \kappa,\ \varphi,\ \zeta,\ \ell_{\text{low}},\ m_{\text{low}},\ \eta,\ \theta_1]^T .$$

Four parameters are calibrated externally to the following values: $\rho = 0.010$ (US), 0.020 (Germany); $r = 0.050$ (US), 0.032 (Germany); $\alpha_1 = 1.000$ (both countries); $\ell_{\text{low}} = 0.075$ (US), 0.097 (Germany). The asset return r is calibrated to the historical average for each country, which is higher in the US (5%) than in Germany (3.2%). The remaining 10 parameters are estimated by minimising the weighted MSM distance between model and data moments.

Two-step LBFGS. We use NLopt's `LD_LBFGS` with bound constraints in two steps. `LD_LBFGS` is the limited-memory Broyden–Fletcher–Goldfarb–Shanno quasi-Newton local optimiser exposed by Julia's `NLopt.jl` package; the `LD_` prefix denotes a local, derivative-based algorithm. We supply gradients via two-sided finite differences with relative step $h = 10^{-5}$. Step 1 minimises a priority-subset weighted distance with an upweighted labor-supply block ($k = 20$ for the US, $k = 10$ for Germany; the upweight magnitudes were calibrated separately and reflect the data block's relative information content). Step 2 is warm-started from the Step 1 solution and minimises the full seven-group weighted distance. Convergence in both steps uses absolute objective tolerance `ftol_abs` $= 10^{-3}$.

Weighting matrix. The weighting matrix W is a group-size diagonal that weights each of the seven moment groups, with the labor-supply (ℓ_{age}) and external-childcare (m_b) blocks upweighted by $k_{\ell_{\text{age}}}$ and k_{m_b} (the values above).[72] The moment covariance Σ used in the

[72] Using a fixed weighting matrix rather than the optimal (inverse-covariance) weighting avoids the small-sample bias documented by Altonji and Segal [1996].

sandwich standard errors below is obtained by bootstrap, as the covariance of the data moment vector over 300 resamples of the moment block (clustered at the household level).

Inference. Standard errors are Jacobian sandwich standard errors,

$$V = (J'WJ)^{-1}\, J'\, W\, \Sigma\, W\, J\, (J'WJ)^{-1}, \qquad \mathrm{SE}(\hat{\theta}_k) = \sqrt{V_{kk}},$$

where J is computed by 5% two-sided finite differences in θ at the MSM estimate $\hat{\theta}$, W is the diagonal weighting matrix from the previous paragraph, and Σ is the bootstrap moment covariance.

D.1.2 Auxiliary statistics

The estimation matches 124 moments organised in seven age-indexed groups,

$$G = \left[g_{l\text{-age}};\; g_{N\text{-age}};\; g_{M\text{-age}};\; g_{\log c};\; g_{\log cc};\; g_{\log a};\; g_{m_b}\right].$$

Group sizes and definitions are summarised below.

Group	Size	Ages	Definition (data side)
$g_{l\text{-age}}$	24	22–45	Mean labor supply per age
$g_{N\text{-age}}$	17	29–45	Mean number of children per age, post-$\bar{t}$
$g_{M\text{-age}}$	17	29–45	Mean parental childcare time per age, post-$\bar{t}$
$g_{\log c}$	24	22–45	Mean log adult consumption per age
$g_{\log cc}$	17	29–45	Mean log child consumption per age, post-$\bar{t}$
$g_{\log a}$	24	22–45	Mean log assets per age
g_{m_b}	1	—	Calibration target $1/12 \approx 0.083$ for mean external childcare time
total	124		

The data side of each group is computed from the couples-only sample (`d_partner==1`) at horizon 22–45. The model side is evaluated at $\hat{\theta}$ via the analytical decision rules of Section 3 and aggregated to the same age cells.

D.2 Model 2 – Remaining calibrated or estimated parameters

Model 2 is a richer model in terms of variables and parameters. Table 4 of the main body reports the full set of estimated structural parameters (37 in total) together with their Jacobian-based standard errors. The present section documents the *remaining* parameters that are not reported in Table 4: the calibrated auxiliary processes (employment, fecundity, marriage, divorce), the calibrated institutional and demographic constants, and the few additional estimated parameters that govern auxiliary processes rather than the structural utility (the employment-logit intercept π_1, the wage-process coefficients (ϕ_0, ϕ_1, ϕ_2), the experience-accumulation coefficient λ_p, and the fecundity-heterogeneity scale σ_η). Estimated parameters are reported with their Jacobian-based standard errors in parentheses; calibrated parameters are reported as plain numbers with their source noted.

D.2.1 Job-finding and human-capital (experience) parameter estimates

In order to capture how the probability of unemployment, $1 - p_t^j$, in the model, depends on age and experience, x, we use the logit

$$\ln\left(\frac{p_t^j}{1 - p_t^j}\right) = \pi_1 + \pi_2 \cdot age_t + \pi_3 \cdot age_t^2 + \pi_4 \cdot x_t + \pi_5 \cdot x_t^2 \ , \tag{D.1}$$

where p_t^j is the fraction of employed females out of the female labor force in year t in the sample. Equation (D.1) follows Wang [2026, p. 10]. Throughout this appendix, "re-estimated" refers to parameters that Wang [2022] calibrated externally; under our Model 2 specification we re-estimate them jointly with the structural parameters of the main body so that the auxiliary processes are internally consistent with the joint MSM estimation. Of the five logit coefficients, π_1 (intercept) is re-estimated jointly with the rest of Model 2, and $(\pi_2, \pi_3, \pi_4, \pi_5)$ are calibrated externally on country-specific labor-force-participation data. The wage process is

$$w_t = e^{\phi_0 + \phi_1 x_t + \phi_2 x_t^2 + \xi_t} \ , \tag{D.2}$$

where $\xi_t \sim N\left(0, \sigma_\xi^2\right)$ is an i.i.d. productivity shock. Equation (D.2) also follows Wang [2026, p. 11]. In the baseline (*headline*) specification —the estimation reported in Table 4 of the main body, as opposed to robustness variants that hold one or more wage parameters fixed at the Wang [2022] values—, the wage parameters (ϕ_0, ϕ_1, ϕ_2) and the productivity-shock standard deviation σ_ξ are re-estimated jointly with the rest of Model 2 (in contrast with Wang's calibration approach), to allow for full country-specific re-pricing of the wage schedule under the Model 2 specification. The experience-accumulation rule is

$$x_{t+1} = \left(x_t + \lambda_p \mathbb{I}_t^{\ell p} + \mathbb{I}_t^{\ell f}\right) \cdot \left[\mathbb{I}_t^{\ell p} + \mathbb{I}_t^{\ell f} + \delta \cdot \left(1 - \mathbb{I}_t^{\ell p} - \mathbb{I}_t^{\ell f}\right)\right] \ , \tag{D.3}$$

where λ_p is the part-time accumulation coefficient (estimated) and δ is the experience depreciation factor (calibrated from Wang, 2022, Table 5).

Table D.1: Job-finding, human-capital, and wage parameters under Model 2

	Germany		**United States**	
Parameter	value	source	value	source
Employment-probability logit (D.1)				
π_1 Intercept	-0.494 (0.027)	estimated	$+1.206$ (0.009)	estimated
π_2 Age	$+0.226$	calibrated	$+0.070$	calibrated
π_3 Age sq.	-0.006	calibrated	-0.003	calibrated
π_4 Experience	$+0.367$	calibrated	-0.003	calibrated
π_5 Experience sq.	-0.023	calibrated	-0.003	calibrated
Human capital				
λ_p Part-time coef.	$+0.557$ (0.052)	estimated	$+0.333$ (0.012)	estimated
δ Depreciation		0.983	calibrated	
Wage process (D.2)				
ϕ_0 Intercept	$+2.028$ (0.010)	estimated	$+2.203$ (0.003)	estimated
ϕ_1 Experience	-0.034 (0.003)	estimated	$+0.183$ (0.001)	estimated
ϕ_2 Experience sq.	$+0.003$ (0.0002)	estimated	-0.011 (0.0001)	estimated
σ_ξ Productivity-shock std.		0.146	calibrated	

Notes: Estimated parameters are reported with Jacobian-based standard errors at $5\% \cdot |\theta|$ step in parentheses; calibrated parameters are reported as plain numbers. Sources for calibrated values: Wang (2022, Table 4, p. 20) for δ; Wang (2022, Table 5, p. 20) for σ_ξ; the job-finding logit slopes $(\pi_2, \pi_3, \pi_4, \pi_5)$ are pre-calibrated from our own country-specific estimates (SOEP panel transitions for Germany; PSID panel transitions for the United States) and held fixed during our MSM estimation, while the intercept π_1 is re-estimated jointly with the rest of Model 2.

D.2.2 Fecundity by age

To address the interplay between desired and planned fertility and actual fertility, Model 2 uses an age-dependent fecundity probability p_t^f entering the stage-1 fertility decision (see eq. (16) in the main body, and Appendix C Section C.2.3). The country-specific fecundity profiles are calibrated from biology-anchored natural-fecundity estimates: the German profile follows Wang [2026, Table A.20, p. 51]; the US profile is calibrated to Tietze [1957][73] and Dunson et al. [2002]. Both profiles are applied year-by-year in the backward-induction Bellman recursion. Table D.2 reports the year-by-year values used in the Bellman equation; aggregated 5-year averages appear at the bottom of the table.

Table D.2: Fecundity probability p_t^f by age

Age group	22–25	26–30	31–35	36–40	41–45
Germany (p_t^f, average)	0.834	0.752	0.580	0.430	0.204
United States (p_t^f, average)	0.866	0.812	0.628	0.300	0.110

Notes: 5-year averages of the year-by-year calibrated p_t^f profile. The DE profile is taken from Wang [2026, Table A.20, p. 51]; the US profile is taken from Tietze [1957] and Dunson et al. [2002]. The full year-by-year profile is applied in the backward-induction Bellman recursion. The US profile features a sharper late-30s decline than the German profile.

Persistent per-woman fecundity heterogeneity σ_η. On top of the age-dependent profile p_t^f, the estimated specification adds a per-woman persistent heterogeneity term $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$, drawn once at simulation initialization and held fixed over the lifecycle, which perturbs the age-conditional fecundity logit:

$$\ln\left(\frac{p_{t,i}^f}{1 - p_{t,i}^f}\right) = \ln\left(\frac{p_t^f}{1 - p_t^f}\right) + \eta_i \,.$$

The dispersion parameter σ_η is estimated jointly with the rest of Model 2 (point estimate followed by Jacobian-based standard error in parentheses): $\sigma_{\eta,\mathrm{DE}} = 3.177\,(0.156)$, $\sigma_{\eta,\mathrm{US}} = 2.916\,(0.181)$. Setting $\sigma_\eta = 0$ recovers the homogeneous-fecundity case exactly.

D.2.3 Marriage and divorce

The marital transitions enter all Bellman equations via period-index logits,

$$\ln\left(\frac{p_t^m}{1 - p_t^m}\right) = \theta_{0m} + \theta_{1m} \cdot t + \theta_{2m} \cdot t^2 + \theta_{3m} \cdot t^3 \ , \tag{D.4}$$

[73] Tietze's estimates are the standard natural-fertility benchmark in the demographic-biology literature, free of contraceptive or behavioral distortions.

where $p_t^m = \Pr(\text{mar}_t = 1 \,|\, \text{mar}_{t-1} = 0)$ is the probability of transitioning into marriage from being single, and

$$\ln\left(\frac{p_t^d}{1 - p_t^d}\right) = \theta_{0d} + \theta_{1d} \cdot t + \theta_{2d} \cdot t^2 \ , \tag{D.5}$$

where $p_t^d = \Pr(\text{mar}_t = 0 \,|\, \text{mar}_{t-1} = 1)$ is the probability of transitioning out of marriage. Here $t = age_t - 22 + 1$ is the model period index ($t = 1$ at age 22, $t = 24$ at age 45), and the first realized transition (age $22 \to 23$) uses $t = 2$.

The country-specific logit coefficients are calibrated separately for the US and Germany. The German values follow Wang [2022] (medium-education group of her education-specific marriage and divorce calibration). The US values are calibrated by matching the model's simulated ever-married share path to its CPS counterpart and the simulated married-to-divorced transition path to the CPS panel transition rate, both via nonlinear least squares on the polynomial coefficients.

For the non-fertile period ($t > 24$, ages 46–62), we adopt the simplification $p_t^m = 1$ and $p_t^d = 0$: any single woman transitions to married with probability one, and married women never divorce.[74]

Table D.3: Marriage and divorce logit coefficients

	Germany		United States	
	Marriage	Divorce	Marriage	Divorce
θ_0 Intercept	-3.508	-3.810	-1.971	-4.039
θ_1 Age	$+0.574$	$+0.017$	$+0.426$	-0.071
θ_2 Age sq.	-0.058	-0.001	-0.019	$+0.002$
θ_3 Age cb.	$+0.002$	–	$+0.000315$	–

Notes: German values follow Wang [2022] (medium-education group of her education-specific marriage and divorce calibration). US values calibrated by matching the model's simulated ever-married share path to its CPS counterpart (marriage) and the simulated married-to-divorced transition path to the CPS panel transition rate (divorce), both via nonlinear least squares on the polynomial coefficients.

D.2.4 Childcare needs and the parental-childcare scaling parameter κ

The childcare-needs function $\bar{\ell}^c(a_t, n_t)$ is a discrete-valued branch function indexed by the age of the youngest child a_t (5 bins) and the number of children $n_t \in \{0, 1, 2, 3\}$. Under Model 2, the country-specific raw matrices below are multiplied by an estimated scaling parameter κ (`lcbar_scale`; $\theta[35]$ in the parameter vector) to allow the level of the childcare-

[74] This is consistent with the polynomial extrapolation (which already gives near-unity marriage hazard and near-zero divorce hazard outside the calibration range) and with the assumption that children have left the household by age 46 (see Appendix C Section C.2.3).

time function to adjust to country-specific data,

$$\bar{\ell}^c(a_t, n_t) \;=\; \kappa \cdot \bar{\ell}^c_{\text{raw}}(a_t, n_t)\,, \tag{D.6}$$

with $\kappa_{\text{DE}} = 0.500\,(0.017)$ and $\kappa_{\text{US}} = 0.771\,(0.029)$. The raw matrices are

$$\bar{\ell}^c_{\text{raw,DE}} = \begin{bmatrix} 0 & 16/24 & 20/24 & 24/24 \\ 0 & 15/24 & 16/24 & 20/24 \\ 0 & 12/24 & 14/24 & 18/24 \\ 0 & 8/24 & 10/24 & 12/24 \\ 0 & 6/24 & 10/24 & 12/24 \end{bmatrix}, \quad \bar{\ell}^c_{\text{raw,US}} = \begin{bmatrix} 0 & 12/24 & 13/24 & 14/24 \\ 0 & 8/24 & 10/24 & 11/24 \\ 0 & 6/24 & 8/24 & 9/24 \\ 0 & 4/24 & 5/24 & 6/24 \\ 0 & 2/24 & 3/24 & 4/24 \end{bmatrix}, \tag{D.7}$$

where rows correspond to youngest-child-age bins $\{[0,3),[3,6),[6,9),[9,12),[12,13]\}$ and columns to $n_t \in \{0,1,2,3\}$. The individual parental time entering the work-disutility block of the utility function is, as in Appendix C Section C.2.2,

$$\text{individual parental time} \;=\; [1 - \text{mar}_t(1-\phi_b)]\,\bar{\ell}^c(a_t, n_t) - n_t m_{b,t}\,, \tag{D.8}$$

where the marriage-adjustment factor $[1 - \text{mar}_t(1-\phi_b)]$ scales the focal parent's share of total parental childcare time, with bargaining parameter $\phi_{b,\text{DE}} = 0.600$ and $\phi_{b,\text{US}} = 0.500$ (both calibrated).

D.2.5 Country-specific institutional scales κ_c^{alim} and κ_c^{SA}

The alimony schedule $f^{\text{alimony}}(n_t)$ and the social-assistance floor $f^{\text{Soc. Assist.}}(\text{mar}_t, n_t)$ appearing in the household-resources function $\tilde{c}_t$ (Appendix C Section C.2.5) carry country-specific scaling factors κ_c^{alim} and κ_c^{SA} that capture institutional differences in alimony enforcement and social-assistance generosity between the two countries (effectively near-zero in the US relative to Germany). The calibrated values are

$$\kappa_{\text{DE}}^{\text{alim}} = \kappa_{\text{DE}}^{\text{SA}} = 1.000\,, \qquad \kappa_{\text{US}}^{\text{alim}} = \kappa_{\text{US}}^{\text{SA}} = 0.010\,.$$

Additionally, the subsistence-consumption shifter $\alpha_3(n_t) = -(a_{3,c} + a_{3,s} \cdot n_t) \cdot 12$ entering $\alpha_1 \ln(\tilde{c}_t + \alpha_3(n_t))$ depends on country-specific calibrated constants $(a_{3,c}, a_{3,s})_{\text{DE}} = (250, 160)$ and $(a_{3,c}, a_{3,s})_{\text{US}} = (200, 240)$, which calibrate the country-specific subsistence consumption floor (a base level plus a per-child increment). All four calibrations are taken from the country's external survey data and held fixed during estimation.

D.2.6 Labor choice set, external childcare wage, and terminal age

The discrete labor choice $\ell_t \in \{0, \ell^p, \ell^f\}$ covers a no-work option, part-time work, and full-time work, with country-specific levels (in fraction-of-day units):

$$\ell^f_{\text{DE}} = 0.30\,, \quad \ell^p_{\text{DE}} = 0.15\,; \quad \ell^f_{\text{US}} = 0.36\,, \quad \ell^p_{\text{US}} = 0.18\,.$$

Both countries are normalized against the common annual working-time-available base $H = 24 \times 4.25 \times 5 \times 12 = 6{,}120$ hrs/yr (24 hrs/day $\times$ 4.25 wks/month $\times$ 5 working days/week $\times$ 12 months), so that ℓ_t = annual hours$/H$. The same H enters the wage process $w_{\text{annual}} = w_{\text{hourly}} \times H$ (with a corresponding shift of the wage intercept ϕ_0 to preserve the implied annual wage level). Model 2 restricts labor supply to a discrete three-point choice; we therefore construct the empirical labor-supply moments using the same cutoffs, ensuring that the simulated moment vector and its data counterpart are measured consistently.

For the *United States*, $\ell^f = 0.36$ and $\ell^p = 0.18$ follow the CPS discrete cutoffs (≥ 35 hrs/week $\mapsto 0.36$; 1–34 hrs/week $\mapsto 0.18$; not employed $\mapsto 0$). Annualized: $\ell^f \cdot H = 2{,}203$ hrs/yr and $\ell^p \cdot H = 1{,}102$ hrs/yr (ratio 1:2). Cross-validating across data sources, applying these same cutoffs to PSID and CPS yields $\bar{\ell}_{\text{PSID}} = 0.231$ and $\bar{\ell}_{\text{CPS}} = 0.227$, agreeing within 1.8%, which justifies the discrete definition empirically.

For *Germany*, $\ell^f = 0.30$ and $\ell^p = 0.15$ are lower than their US counterparts, reflecting the institutional fact that German full-time workers accumulate fewer annual hours due to higher statutory vacation, shorter standard workweek, and generous parental leave. Annualized: $\ell^f \cdot H = 1{,}836$ hrs/yr and $\ell^p \cdot H = 918$ hrs/yr. These levels are aligned with Wang [2026, p. 10], who reports that "full-time work corresponds to 40 work hours per week and part-time to 20 hours, aligned with mean hours in the data," which maps to approximately 0.30/0.15 under the same H denominator. The 1:2 PT/FT hours ratio is preserved across countries, so the experience-accumulation parameter λ_p admits the symmetric prior bound $\lambda_p \in [1/3,\ 3/4]$ in both estimations.

The country-specific (ℓ^f, ℓ^p) levels therefore absorb the institutional difference in working hours (US workers at FT supply roughly 20% more annual hours than DE workers at FT) while keeping the model's discrete choice structure and time-endowment normalization invariant across countries. This preserves the comparability of cross-country FT-share, PT-share, and labor-supply-by-children moments.

The external childcare wage w^n entering the household-resources function $\tilde{c}_t$ (Appendix C Section C.2.5) is a calibrated annualized wage, taken from the country's external childcare market data:

$$w^n_{\text{DE}} = 17\,\text{EUR/hr} \times 160\,\text{hr/month} \times 12 = 32{,}640\,\text{EUR/year}\,;$$

$$w^n_{\mathrm{US}} = 24\,\mathrm{USD/hr} \times 150\,\mathrm{hr/month} \times 12 = 43{,}200\,\mathrm{USD/year}\,.$$

The terminal age is $T = 62$ for both countries, with $V_{63} \equiv 0$ as the boundary condition for the backward induction (Appendix C Section C.2.3).[75]

D.2.7 Partner's earnings $y_{p,t}$

Partner's earnings $y_{p,t}$ enter the household-resources function $\tilde{c}_t$ when the focal woman is married. The country-specific $y_{p,t}$ profile is calibrated by age from GSOEP-equivalent (Germany) and PSID (United States) age-conditional income tabulations and held fixed during estimation. The DE profile spans ages 25–45 (length 21 years) with annualized values from ∼18,378 EUR/year (age 25) to ∼24,315 EUR/year (age 45). The US profile spans ages 22–60 (length 39 years) with annualized values from ∼29,136 USD/year (age 22) to higher values at older ages. Both profiles are treated as state-conditional means; cross-sectional dispersion at a given age is not modeled.

D.2.8 Heterogeneous-population type-mix probabilities (π_{O_1}, π_{O_2})

Model 2 admits a four-type unobserved-heterogeneity structure through the binary attitudinal switches $O_1, O_2 \in \{0, 1\}$ in the per-period utility (eq. (18) of the main body, restated in Appendix C Section C.2.1):

- O_1 indexes fertility preference: type $O_1 = 0$ has high fertility preference (full children-utility), and type $O_1 = 1$ has low fertility preference (children-utility scaled by $1 - \alpha_6$).
- O_2 indexes the general disutility of effort (market work and parental childcare): type $O_2 = 0$ has high effort disutility, and type $O_2 = 1$ has low effort disutility (effort disutility scaled by $1 - \gamma_5$).

The four cells $(O_1, O_2) \in \{0, 1\}^2$ are solved as four independent dynamic programming blocks and aggregated into population moments by the type-mixture

$$M(\theta) = \sum_{O_1, O_2 \in \{0,1\}} \pi^{(O_1)}_{O_1} \pi^{(O_2)}_{O_2} \, M^{(O_1,O_2)}(\theta)\,,$$

[75] For comparison, the corresponding object in Model 1 is the external-childcare price p_b, annualized as $p_b \times 5{,}760$ over a $12 \times 20 \times 24$ clock-hour base, with $p_b = \$8$/hr (US) and €7/hr (Germany)—that is, \$46,080/year (US) and €40,320/year (Germany). These annualized prices, which are the objects that enter each model's budget constraint, are close to the Model 2 values ($w^n = \$43{,}200$ for the US and €32,640 for Germany). The wider gap in the per-hour rates reflects only the different annual-hour conventions: Model 1 spreads the price over 5,760 clock-hours, whereas Model 2 uses the 1,800 (US, 150×12) and 1,920 (Germany, 160×12) working-hour bases.

where $\pi_{O_1}^{(1)} = \pi_{O_1}$, $\pi_{O_1}^{(0)} = 1 - \pi_{O_1}$, and analogously for O_2.

The type-mix probabilities (π_{O_1}, π_{O_2}) are estimated within country-specific bounds documented in Table D.4. For Germany, the bounds are taken from Wang [2022] at $\pm 20\%$ around her central values. For the United States, the bounds are informed by attitudinal-question shares in the NLSY79 (survey items on desired family size and on the trade-off between work and family).

Table D.4: Type-mix probabilities (π_{O_1}, π_{O_2}): bounds and estimates

	Germany		**United States**	
	bounds	estimate	bounds	estimate
$\pi_{O_1} = \Pr(O_1 = 1)$	[0.600, 0.900]	0.759 (0.053)	[0.400, 0.700]	0.600 (0.008)
$\pi_{O_2} = \Pr(O_2 = 1)$	[0.272, 0.408]	0.344 (0.068)	[0.250, 0.550]	0.495 (0.041)

Notes: Estimated jointly with the rest of Model 2 within the country-specific bounds shown. Standard errors (Jacobian-based, $5\% \cdot |\theta|$ step) in parentheses. German bounds are Wang [2026] central values $\pm 20\%$. US bounds are informed by NLSY79 attitudinal-question shares.

E Dynamic bivariate probit model

E.1 Introduction

We are interested in studying two longitudinal dichotomous variables that are closely related and are likely to influence each other, y_{1it} and y_{2it}, where $i = \{1, ..., N\}$, $t = \{1, ..., T\}$. Variable y_1 is a "0-1" dummy variable, conveying information about desired working hours. Specifically,

$$y_{1i} = \begin{cases} 1 \ , & \text{if desired working hours are above individual } i\text{'s mean} \\ 0 \ , & \text{else} \end{cases} , \tag{E.1}$$

in which the individual's mean is derived by a respondent's answers to a series of vignette experiments. Similarly,

$$y_{2i} = \begin{cases} 1 \ , & \text{if desired number of children are above individual } i\text{'s mean} \\ 0 \ , & \text{else} \end{cases} . \tag{E.2}$$

In our vignette survey we vary five dimensions of an environment (see Appendix B). The dimensions of the first environment are randomly drawn for all respondents and serve as a baseline for the three subsequent environments. Because of this baseline structure, we can view the estimation of regressions analyzing the impact of exogenous dimensions on y_{1i} and y_{2i}, as the estimation of a dynamic panel with time lags. This is the reason that we add the panel dimension to variables y_{1i} and y_{2i}, using the symbols "y_{1it}" and "y_{2it}". Thus, technically, the problem we want to solve is a dynamic panel estimation of a joint decision process, i.e. *estimating a system of dynamic panel equations.* Estimating a system of dynamic panel equations is challenging because it is difficult to have an analytical expression for the likelihood function.

E.2 Maximum simulated likelihood estimation

We follow Cameron and Trivedi (2005, Ch. 12), in order to use maximum likelihood estimation (MLE) in the absence of an analytical expression the density. The key result is that simulation can lead to an estimator with the same distribution as the MLE, provided that the number of simulation draws made to compute the density for each observation goes to infinity (see Cameron and Trivedi, 2005, Section 12.4).

E.2.1 Model

For a respondent i that is observed in waves $t = 1, ..., T$, the latent-variable model for time periods $t = 2, ..., T$ is given by

$$\left.\begin{aligned} y^*_{1i,t} &= \mathbf{x}'_{1i,t}\beta_1 + y_{1i,t-1}\gamma_{11} + y_{2i,t-1}\gamma_{12} + \alpha_i + \mu_{1i,t} \\ y^*_{2i,t} &= \mathbf{x}'_{2i,t}\beta_2 + y_{1i,t-1}\gamma_{21} + y_{2i,t-1}\gamma_{22} + \alpha_i + \mu_{2i,t} \end{aligned}\right\} \tag{E.3}$$

in which $y^*_{1i,t}$ and $y^*_{2i,t}$ are the desired working hours and the desired number of children stated by respondent i in all vignette experiments. We observe $y_{1i,t} = \mathbf{1}[y^*_{1i,t} > \bar{y}^*_{1i}]$ and $y_{2i,t} = \mathbf{1}[y^*_{2i,t} > \bar{y}^*_{2i}]$, in which $\bar{y}^*_{1i}$ and $\bar{y}^*_{2i}$ are the average responses of respondent i, given all vignette experiments that were given to i. The β's and γ's are unknown parameters. The regressor vector $\mathbf{x}$ includes a constant term. Individual-specific random effects are captured by α_i, while $\mu_{1i,t}$ and $\mu_{2i,t}$ are the error terms.

We assume that the error terms $\mu_{i,t} = [\mu_{1i,t}, \mu_{2i,t}]^T$ are independent over time and jointly distributed as

$$\mu_{i,t} \sim \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \rho_\mu \\ \rho_\mu & 1 \end{bmatrix}\right) , \tag{E.4}$$

in which ρ_μ does not depend on t. The random effect, α_i, is a random variable representing unobserved individual heterogeneity and are time invariant, and it is assumed to be normally distributed as

$$\alpha_i \sim N\left(0, \sigma^2_\alpha\right) . \tag{E.5}$$

For the initial period $t = 1$, model (E.3) implies

$$\left.\begin{aligned} y^*_{1i,1} &= \mathbf{x}'_{1i,t}\beta^0_1 + y_{1i,0}\gamma_{11} + y_{2i,0}\gamma_{12} + \alpha_i + \mu_{1i,1} \\ y^*_{2i,1} &= \mathbf{x}'_{2i,t}\beta^0_2 + y_{1i,0}\gamma_{21} + y_{2i,0}\gamma_{22} + \alpha_i + \mu_{2i,1} \end{aligned}\right\} . \tag{E.6}$$

In order to estimate the determinants of initial conditions of the model, Heckman (1981b, pp. 188-189) suggests replacing the system (E.3) for $t = 1$, which is given by (E.6), with a static equation. That static equation may have different regression coefficients from (E.6), so,

$$\left.\begin{aligned} y^*_{1i,1} &= \mathbf{z}'_{1i,1}\delta_1 + \lambda_{11}\alpha_i + \varepsilon_{1i,1} \\ y^*_{2i,1} &= \mathbf{z}'_{2i,1}\delta_2 + \lambda_{21}\alpha_i + \varepsilon_{2i,1} \end{aligned}\right\}, \tag{E.7}$$

in which the error terms, $\varepsilon_{i1} = [\varepsilon_{1i,1}, \varepsilon_{2i,1}]^T$, are jointly distributed as

$$\varepsilon_{i1} \sim \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \rho_\varepsilon \\ \rho_\varepsilon & 1 \end{bmatrix}\right) . \tag{E.8}$$

In our vignette survey, for each respondent, two consecutive vignettes (t and $t+1$) differ only by one characteristic from (i)-(v) above. This feature means that the set of variables $\mathbf{x}_{1i,t}$ and $\mathbf{x}_{2i,t}$ vary very little across different t's. In practice, this limited variation causes estimation problems, since control variables tend to be collinear across t's. This collinearity feature makes the likelihood function too flat, which, in turn, makes finding minimizers quite difficult.[76]In order to cope with this problem, we make matrices $\mathbf{z}_{1i,1}$ and $\mathbf{z}_{2i,1}$ different from matrices $\mathbf{x}_{1i,1}$ and $\mathbf{x}_{2i,1}$, by omitting some variables. Specifically, matrices $\mathbf{z}_{1i,1}$ and $\mathbf{z}_{2i,1}$ contain fewer variables compared to matrices $\mathbf{x}_{1i,1}$ and $\mathbf{x}_{2i,1}$.

E.2.2 Distributional assumptions

All distributional assumptions are listed as,

$$\begin{aligned}
f(\alpha|\mathbf{x},\mathbf{z},\varepsilon,\mu) &= f(\alpha) && (A1)\\
f(\varepsilon|\mathbf{x},\mathbf{z},\alpha) &= f(\varepsilon|\alpha) && (A2)\\
f(\mu|\mathbf{x},\mathbf{z},\alpha) &= f(\mu|\alpha) && (A3)\\
\varepsilon \perp \mu \quad &| \quad \alpha && (A4)\\
f(\varepsilon_{i,t}|\varepsilon_{i,s},\alpha) &= f(\varepsilon_{i,t}|\alpha) \quad \forall \quad s \neq t && (A5)\\
f(\mu_{i,t}|\mu_{i,s},\alpha) &= f(\mu_{i,t}|\alpha) \quad \forall \quad s \neq t\ . && (A6)
\end{aligned}$$

Condition (A1) is a standard random-effects assumption. Conditions (A1)-(A3) ensure that all explanatory variables are exogenous. Condition (A4) ensures that innovations in dynamic equations and initial conditions are independent, conditional upon α. Finally, (A5) and (A6) rule out serial autocorrelation for the two pairs of Gaussian errors. For details on assumptions (A1)-(A6), see also Wooldridge (2001, Ch. 15).

E.2.3 Likelihood function

For a bivariate normal CDF, we have

$$\Pr(X_1 \leq x_1, X_2 \leq x_2) = \int_{-\infty}^{x_2}\int_{-\infty}^{x_1} \phi_2(z_1, z_{2,}\rho) dz_1 dz_2 = \Phi_2(x_1, x_{2,}\rho)$$

in which the density function is given by

$$\phi_2(x_1, x_{2,}\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{1-\rho^2}\right)\ .$$

[76]Note that for forming the likelihood function we use both systems of equations jointly, i.e., the system given by (E.6) and the system given by (E.7).

Partial derivatives which would be useful later on are,

$$\begin{aligned}
\frac{\partial \Phi_2(x_1,x_2,\rho)}{\partial x_1} &= \phi(x_1)\cdot F\left(x_2 : \rho x_1, \sqrt{1-\rho^2}\right) \\
&= \phi(x_1)\cdot \Phi\left(\frac{x_2-\rho x_1}{\sqrt{1-\rho^2}}\right) \equiv \Gamma_{x1}(x_1,x_2,\rho)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Phi_2(x_1,x_2,\rho)}{\partial x_2} &= \phi(x_2)\cdot F\left(x_1 : \rho x_2, \sqrt{1-\rho^2}\right) \\
&= \phi(x_1)\cdot \Phi\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right) \equiv \Gamma_{x_2}(x_1,x_2,\rho)
\end{aligned}$$

$$\frac{\partial \Phi_2(x_1,x_2,\rho)}{\partial \rho} = \phi_2(x_1,x_2,\rho) \equiv \Gamma_{\rho}(x_1,x_2,\rho)\ ,$$

in which $F\ (x : \mu, \sigma)$ denotes the cdf of a normal variable, x, with mean and variance μ and σ. In addition, we construct two index variables, $q_{1it} = 2y_{1i,t} - 1$ and $q_{2it} = 2y_{2i,t} - 1$, which we will conveniently use later on:

$$\begin{aligned}
q_{1it} &= 2y_{1i,t} - 1 \rightarrow \begin{cases} 1 & if \quad y^*_{1i,t} > \bar{y}^*_{1i} \\ -1 & if \quad y^*_{1i,t} \leq \bar{y}^*_{1i} \end{cases}\ , \\
q_{2it} &= 2y_{2i,t} - 1 \rightarrow \begin{cases} 1 & if \quad y^*_{2i,t} > \bar{y}^*_{2i} \\ -1 & if \quad y^*_{2i,t} \leq \bar{y}^*_{2i} \end{cases}\ .
\end{aligned}$$

For each individual i we have four different possible combinations,

$$\begin{pmatrix} \Pr(y_{1i,t}=1, y_{2i,t}=1) & \Pr(y_{1i,t}=1, y_{2i,t}=0) \\ \Pr(y_{1i,t}=0, y_{2i,t}=1) & \Pr(y_{1i,t}=0, y_{2i,t}=0) \end{pmatrix} \tag{E.9}$$

$$\begin{pmatrix} \Pr(y^*_{1i,t}>0, y^*_{2i,t}>0 & \Pr(y^*_{1i,t}>0, y^*_{2i,t}\leq 0) \\ \Pr(y^*_{1i,t}\leq 0, y^*_{2i,t}>0) & \Pr(y^*_{1i,t}\leq 0, y^*_{2i,t}\leq 0) \end{pmatrix}\ . \tag{E.10}$$

To save some space, let

$$\begin{aligned}
\mathbf{x}'_{1i,t}\beta_1 + y_{1i,t-1}\gamma_{11} + y_{2i,t-1}\gamma_{12} + \alpha_i &\equiv \tilde{\mu}_{1i,t} \\
\mathbf{x}'_{2i,t}\beta_2 + y_{1i,t-1}\gamma_{21} + y_{2i,t-1}\gamma_{22} + \alpha_i &\equiv \tilde{\mu}_{2i,t} \\
\mathbf{z}'_{1i,1}\delta_1 + \lambda_{11}\alpha_i &\equiv \tilde{\mu}_{1i,1} \\
\mathbf{z}'_{2i,1}\delta_2 + \lambda_{21}\alpha_i &\equiv \tilde{\mu}_{2i,2}\ ,
\end{aligned}$$

so, (E.10) becomes,

$$\begin{pmatrix} \Pr(\tilde{\mu}_{1i,t} + \mu_{1i,t} > 0, \tilde{\mu}_{2i,t} + \mu_{2i,t} > 0 & \Pr(\tilde{\mu}_{1i,t} + \mu_{1i,t} > 0, \tilde{\mu}_{2i,t} + \mu_{2i,t} \leq 0) \\ \Pr(\tilde{\mu}_{1i,t} + \mu_{1i,t} \leq 0, \tilde{\mu}_{2i,t} + \mu_{2i,t} > 0) & \Pr(\tilde{\mu}_{1i,t} + \mu_{1i,t} \leq 0, \tilde{\mu}_{2i,t} + \mu_{2i,t} \leq 0) \end{pmatrix},$$

which further implies,

$$\begin{pmatrix} \Pr(\mu_{1i,t} > -\tilde{\mu}_{1i,t}, \mu_{2i,t} > -\tilde{\mu}_{2i,t} & \Pr(\mu_{1i,t} > -\tilde{\mu}_{1i,t}, \mu_{2i,t} \leq -\tilde{\mu}_{2i,t}) \\ \Pr(\mu_{1i,t} \leq -\tilde{\mu}_{1i,t}, \mu_{2i,t} > -\tilde{\mu}_{2i,t}) & \Pr(\mu_{1i,t} \leq -\tilde{\mu}_{1i,t}, \mu_{2i,t} \leq -\tilde{\mu}_{2i,t}) \end{pmatrix}. \tag{E.11}$$

Given the symmetric structure of the bivariate normal distribution with zero mean and covariance ρ_μ, (E.11) implies,

$$\begin{aligned}
\Pr(\mu_{1i,t} &> -\tilde{\mu}_{1i,t}, \mu_{2i,t} > -\tilde{\mu}_{2i,t}) \\
&= \int_{-\tilde{\mu}_{2i,t}}^{+\infty} \int_{-\tilde{\mu}_{1i,t}}^{+\infty} \phi_2(z_1, z_{2,}\rho) dz_1 dz_2 \\
&= \int_{-\infty}^{\tilde{\mu}_{2i,t}} \int_{-\infty}^{\tilde{\mu}_{1i,t}} \phi_2(z_1, z_{2,}\rho) dz_1 dz_2 = \Phi_2(\tilde{\mu}_{1i,t}, \tilde{\mu}_{2i,t}, \rho) \\
\Pr(\mu_{1i,t} &> -\tilde{\mu}_{1i,t}, \mu_{2i,t} \leq -\tilde{\mu}_{2i,t}) \\
&= \int_{-\infty}^{-\tilde{\mu}_{2i,t}} \int_{-\tilde{\mu}_{1i,t}}^{+\infty} \phi_2(z_1, z_{2,}\rho) dz_1 dz_2 \\
&= \int_{-\infty}^{-\tilde{\mu}_{2i,t}} \int_{-\infty}^{\tilde{\mu}_{1i,t}} \phi_2(z_1, z_{2,}\rho) dz_1 dz_2 = \Phi_2(\tilde{\mu}_{1i,t}, -\tilde{\mu}_{2i,t}, \rho) \\
\Pr(\mu_{1i,t} &\leq -\tilde{\mu}_{1i,t}, \mu_{2i,t} > -\tilde{\mu}_{2i,t}) \\
&= \int_{-\tilde{\mu}_{2i,t}}^{+\infty} \int_{-\infty}^{-\tilde{\mu}_{1i,t}} \phi_2(z_1, z_{2,}\rho) dz_1 dz_2 \\
&= \int_{-\infty}^{\tilde{\mu}_{2i,t}} \int_{-\infty}^{-\tilde{\mu}_{1i,t}} \phi_2(z_1, z_{2,}\rho) dz_1 dz_2 = \Phi_2(-\tilde{\mu}_{1i,t}, \tilde{\mu}_{2i,t}, \rho) \\
\Pr(\mu_{1i,t} &\leq -\tilde{\mu}_{1i,t}, \mu_{2i,t} \leq -\tilde{\mu}_{2i,t}) \\
&= \int_{-\infty}^{-\tilde{\mu}_{2i,t}} \int_{-\infty}^{-\tilde{\mu}_{1i,t}} \phi_2(z_1, z_{2,}\rho) dz_1 dz_2 = \Phi_2(-\tilde{\mu}_{1i,t}, -\tilde{\mu}_{2i,t}, \rho)
\end{aligned}$$

and

$$\begin{pmatrix} \Phi_2(\tilde{\mu}_{1i,t}, \tilde{\mu}_{2i,t}, \rho)_{[1,1]} & \Phi_2(\tilde{\mu}_{1i,t}, -\tilde{\mu}_{2i,t}, \rho)_{[1,0]} \\ \Phi_2(-\tilde{\mu}_{1i,t}, \tilde{\mu}_{2i,t}, \rho)_{[0,1]} & \Phi_2(-\tilde{\mu}_{1i,t}, -\tilde{\mu}_{2i,t}, \rho)_{[0,0]} \end{pmatrix}$$

combined with

$$
\begin{aligned}
q_{1it} &= 2y_{1i,t} - 1 \rightarrow \begin{cases} 1 & if \quad y^*_{1i,t} > 0 \\ -1 & if \quad y^*_{1i,t} \leq 0 \end{cases} \\
q_{2it} &= 2y_{2i,t} - 1 \rightarrow \begin{cases} 1 & if \quad y^*_{2i,t} > 0 \\ -1 & if \quad y^*_{2i,t} \leq 0 \end{cases}
\end{aligned}
$$

allow us to directly write down the likelihood function of individual i at period t as

$$L_{i,t}\left(a_{1i,t}; a_{2i,t}\right) = \Phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho) \ . \tag{E.12}$$

Equation (E.12) implies that the contribution of the i-th individual to the likelihood is,

$$
\begin{aligned}
L_i &= \int_{-\infty}^{+\infty} \Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon) \times \\
&\quad \times \prod_{t=2}^{T} \Phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu) f_2(\alpha_i) d\alpha
\end{aligned} \tag{E.13}
$$

where $\Phi_2(\cdot, \cdot, \rho)$ is the bivariate cumulative density function of a bivariate normal distribution with means zero, unit variances, and covariances ρ. $f_2(\cdot)$ represents the normal density of the random effects α_i given the distributional assumption made above.

E.2.4 Simulated likelihood

We approximate the integral (E.13) with the simulated average

$$
\begin{aligned}
L_i^{approx} &\simeq \frac{1}{R}\sum_{r=1}^{R} \quad \Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon | \alpha_i^r) \times \\
&\quad \times \prod_{t=2}^{T} \Phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu | \alpha_i^r) \ ,
\end{aligned}
$$

where the random effects (α_i) are replaced by independent random draws (α_i^r).

If N is the number of households in the sample, RN independent draws from the standard normal distribution are taken (using a pseudo-random number generator or using Halton sequence for simulation).[77]For each household, this procedure gives R independent draws $(\widetilde{\alpha_i^r})_{[1...R]}$. We then transform $\widetilde{\alpha_i^r}$ into draws from a standard normal distribution stated

[77]Halton sequences are known for better coverage of the $[0, 1]$ interval and need less draws to achieve high precision compare with pseudo-random number generator.

in (E.5), as

$$\alpha_i^r = \sigma_\alpha \widetilde{\alpha_i^r} \ .$$

E.3 STATA ml

E.3.1 Definition of 9 equations

Here we list the first-order derivatives related to the 9 equations,

$$\begin{pmatrix} xb1t & xb2t & xb10 & xb20 & \rho_e & \rho_\mu & \sigma_\alpha & \lambda_{11} & \lambda_{21} \\ \theta 1_{it} & \theta 2_{it} & \theta 3_{it} & \theta 4_{it} & \theta 5_{it} & \theta 6_{it} & \theta 7_{it} & \theta 8_{it} & \theta 9_{it} \end{pmatrix} ,$$

(following the STATA syntax). First, let's define

$$\Omega \equiv \Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon) \prod_{t=2}^{T} \Phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu) \ .$$

For convenience in later steps we use,

BLOCK 1

$$\begin{aligned} \frac{\partial \ln L_i}{\partial \theta 1_{it}}_{(t\geq 2)} &= \frac{1}{L_i}\frac{\partial L_i}{\partial \theta 1_{it}} = \frac{1}{L_i}\frac{\partial L_i}{\partial xb1t} \\ &= \frac{1}{L_i} \cdot \int_{-\infty}^{+\infty} \left[q_{1it} \times \frac{\Gamma_{x1}(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu)}{\Phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu)} \right] \Omega \cdot f(\alpha) d\alpha \end{aligned}$$

$$\begin{aligned} \frac{\partial \ln L_i}{\partial \theta 2_{it}}_{(t\geq 2)} &= \frac{1}{L_i}\frac{\partial L_i}{\partial xb2t} \\ &= \frac{1}{L_i} \cdot \int_{-\infty}^{+\infty} \left[q_{2it} \times \frac{\Gamma_{x2}(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu)}{\Phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu)} \right] \Omega \cdot f(\alpha) d\alpha \end{aligned}$$

$$\begin{aligned} \frac{\partial \ln L_i}{\partial \theta 3_{it}}_{(t=1)} &= \frac{1}{L_i}\frac{\partial L_i}{\partial xb10} \\ &= \frac{1}{L_i} \cdot \int_{-\infty}^{+\infty} \left[q_{1i1} \times \frac{\Gamma_{x1}(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)}{\Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)} \right] \Omega \cdot f(\alpha) d\alpha \end{aligned}$$

$$\left.\frac{\partial \ln L_i}{\partial \theta 4_{it}}\right._{(t=1)} = \frac{1}{L_i}\frac{\partial L_i}{\partial xb20}$$

$$= \frac{1}{L_i}\cdot\int_{-\infty}^{+\infty}\left[q_{2i1}\times\frac{\Gamma_{x2}(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)}{\Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)}\right]\Omega\cdot f(\alpha)d\alpha$$

BLOCK 2

$$\left.\frac{\partial \ln L_i}{\partial \theta 5_{it}}\right._{(t=1)} = \frac{1}{L_i}\frac{\partial L_i}{\partial \rho_e}$$

$$= \frac{1}{L_i}\cdot\int_{-\infty}^{+\infty}\left[q_{1i1}q_{2i1}\times\frac{\phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)}{\Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)}\right]\Omega\cdot f(\alpha)d\alpha$$

$$\left.\frac{\partial \ln L_i}{\partial \theta 6_{it}}\right._{(t\geq 2)} = \frac{1}{L_i}\frac{\partial L_i}{\partial \rho_\mu}$$

$$= \frac{1}{L_i}\cdot\int_{-\infty}^{+\infty}\left[q_{1it}q_{2it}\times\frac{\phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu)}{\Phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu)}\right]\Omega\cdot f(\alpha)d\alpha$$

BLOCK3

$$\left.\frac{\partial \ln L_i}{\partial \theta 7_{it}}\right._{(t=1)} = \frac{1}{L_i}\frac{\partial L_i}{\partial \ln\sigma_{\alpha 1}} = \frac{1}{L_i}\frac{\partial L_i}{\partial \sigma_{\alpha 1}}\cdot\sigma_{\alpha 1}$$

$$= \frac{\sigma_{\alpha 1}}{L_i}\cdot\int_{-\infty}^{+\infty}\widetilde{\alpha_1^r}$$

$$[q_{1i1}\lambda_{11}\Gamma_{x1}(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon) + q_{2i1}\lambda_{21}\Gamma_{x2}(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)]$$

$$\times\frac{\Omega}{\Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)}\cdot f(\alpha)d\alpha$$

$$\left.\frac{\partial \ln L_i}{\partial \theta 7_{it}}\right._{(t\geq 2)} = \frac{1}{L_i}\frac{\partial L_i}{\partial \ln\sigma_{\alpha 1}} = \frac{1}{L_i}\frac{\partial L_i}{\partial \sigma_{\alpha 1}}\cdot\sigma_{\alpha 1}$$

$$= \frac{\sigma_{\alpha 1}}{L_i}\cdot\int_{-\infty}^{+\infty}\widetilde{\alpha_1^r}$$

$$[q_{1it}\cdot\Gamma_{x1}(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu) + q_{2it}\cdot\Gamma_{x2}(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu)]$$

$$\times\frac{\Omega}{\Phi_2(q_{1it}\tilde{\mu}_{1i,t}, q_{2it}\tilde{\mu}_{2i,t}, q_{1it}q_{2it}\rho_\mu)}\cdot f(\alpha)d\alpha$$

BLOCK 4

$$\begin{aligned}
\frac{\partial \ln L_i}{\partial \theta 8_{it}}_{(t=1)} &= \frac{1}{L_i}\frac{\partial L_i}{\partial \lambda_{11}} \\
&= \frac{1}{L_i} \cdot \int_{-\infty}^{+\infty} \left[q_{1i1}\alpha^r \times \frac{\Gamma_{x1}(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)}{\Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)} \right] \Omega \cdot f(\alpha) d\alpha
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L_i}{\partial \theta 9_{it}}_{(t=1)} &= \frac{1}{L_i}\frac{\partial L_i}{\partial \lambda_{21}} \\
&= \frac{1}{L_i} \cdot \int_{-\infty}^{+\infty} \left[q_{2i1}\alpha^r \times \frac{\Gamma_{x2}(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)}{\Phi_2(q_{1i1}\tilde{\mu}_{1i,1}, q_{2i1}\tilde{\mu}_{2i,1}, q_{1i1}q_{2i1}\rho_\varepsilon)} \right] \Omega \cdot f(\alpha) d\alpha
\end{aligned}$$

Using these blocks, a STATA program jointly estimates the system given by (E.6) and the system given by (E.7).

E.4 Estimation results

Table E.1 provides the estimation results.

Table E.1: Dynamic bivariate probit model

	Desired no. of children	Desired working hours
	United States	
hourly wage	0.0075***	0.0004
	(5.6108)	(0.2877)
partner: wage	0.0148***	-0.0067***
	(7.1570)	(-3.4567)
partner: working hrs	-0.0008*	0.0010**
	(-1.8948)	(2.4435)
Allowance (\$200 or \$400)	0.0003	-0.0000
	(1.2175)	(-0.0019)
Subsidy (subsidized vs. costly, 0/1)	-0.3027***	-0.0490
	(-6.8066)	(-1.1908)
$n^{\star}_{t-1}$	-0.1763***	0.6903***
	(-3.0734)	(14.0450)
$\ell^{\star}_{t-1}$	0.6528***	-0.1847***
	(12.5656)	(-3.2440)
N	7288	
	Germany	
hourly wage	-0.0123**	0.0029
	(-2.0448)	(0.5475)
partner: wage	0.0267***	-0.0188**
	(3.1914)	(-2.5062)
partner: working hrs	-0.0012	-0.0018*
	(-1.1317)	(-1.8862)
Allowance (€250 or €500)	0.0009**	-0.0008**
	(1.9938)	(-2.1212)
Subsidy (subsidized vs. costly, 0/1)	-0.4570***	-0.1962**
	(-4.4072)	(-2.0914)
$n^{\star}_{t-1}$	0.1154	0.5314***
	(1.1767)	(6.2320)
$\ell^{\star}_{t-1}$	0.7492***	-0.1879*
	(7.6166)	(-1.8544)
N	1,784	

Note: t statistics in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dynamic bivariate probit model with common random-effect.

F Out-of-sample validation and cross-validations of Models 1 and 2

F.1 Model 1: aggregative and deterministic

F.1.1 Construction of the simulated panel

The construction of the Model 1 simulated panel is described conceptually in Section 6.1 of the main body of the paper. Here we provide the technical implementation details: the per-respondent simulation procedure, the variable-by-variable mappings of vignette stimuli into Model 1's budget constraint, the implementation choices for cross-model consistency, and the pipeline mechanics.

The vignette sample, restricted to female respondents aged 22–42, contains $N_{\text{US}} = 3{,}056$ cells in the United States (764 respondents $\times$ 4 randomized environments) and $N_{\text{DE}} = 1{,}128$ in Germany (282 respondents $\times$ 4 environments). For each such vignette-panel cell, we solve Model 1's lifecycle problem for three modeling scenarios. This gives a total of 9,168 solutions for the US and 3,384 for the German sample. The structural parameters are held fixed at their estimated values $\hat{\theta}_{\text{M1}}$ (Section 3 of the main body of the paper) across all model solutions. Modeling scenario 1 is the country-level baseline (baseline wage and partner-income profiles, baseline allowance and daycare price). Modeling scenarios 2 and 3 substitute the vignette stimuli into the wage and partner-income paths: modeling scenario 2 (*current-period only*) does so at the respondent's age $t_0 = \text{age} - 21$ only, while modeling scenario 3 (*permanent*) does so for all $t \geq t_0$. Each solve returns model-implied age-by-age fertility $n(t)$ and labor supply $\ell(t)$ over the full lifecycle window $t \in [22, 45]$; we collect these into matrices N_{ModEnv} (24×3) and L_{ModEnv} (24×3), with rows indexed by age and columns by modeling scenario.

Two of the vignette stimuli enter Model 1 as time paths (wage $w(t)$ and partner income $y_p(t)$) and the other two as scalars (allowance A and daycare price $\tilde{p}_b$). The time paths are replaced only at the periods specified by the perception convention: modeling scenario 2 (current-period only) replaces them at t_0 alone, while modeling scenario 3 (permanent) replaces them for all $t \geq t_0$. The scalars are applied to the full lifecycle window in both perturbed modeling scenarios. Modeling scenario 1 retains the country-baseline values for all four stimuli.

The vignette module encodes childcare availability as a binary indicator $s_{\text{costly\_care}} \in \{0, 1\}$, with $s_{\text{costly\_care}} = 0$ representing subsidized (free or near-free) external daycare and $s_{\text{costly\_care}} = 1$ representing costly daycare. We implement this binary mapping in each model's native daycare-cost specification. In Model 1, the subsidized state sets the daycare price to $\tilde{p}_b = 0.05 \cdot \tilde{p}_b^{\text{anchor}}$ and the costly state retains $\tilde{p}_b = \tilde{p}_b^{\text{anchor}}$; in Model 2, the same binary maps to the subsidy rate $s = 0.95$ (highly subsidized) versus $s = 0$ (costly) on the daycare-cost term $(1-s)\, p_{b,t} m_{b,t}$ in the household budget constraint. Both parameterizations represent the vignette's intended intervention — a 95% reduction in the effective daycare cost — under each model's native specification.

We define the capped fertility $\tilde{n}(t) \equiv \min(n(t), 3)$ and use $\tilde{n}(t)$ in place of $n(t)$ for all subsequent aggregation, mirroring Model 2's discrete state ceiling $\{0, 1, 2, 3\}$. The cap is applied element-wise to the N_{ModEnv} matrix immediately after the per-modeling-scenario solve and before any post-$\bar{t}$ aggregation. The cap value 3 in both Model 1 (via the min transformation) and Model 2 (via the discrete state grid $\{0, 1, 2, 3\}$) represents *three or more children*: any fertility count of $3, 4, 5, \ldots$ is mapped to the same upper-boundary state. This convention reflects the rare-event nature of higher-parity fertility in the data and the practical need to keep each model's state space compact.

The computational pipeline runs in parallel by 9 parallel-programming workers per country; each parallel-programming worker handles a disjoint subset of the vignette-panel cells and writes a partial comma-separated-value (CSV) file, while a merge step concatenates the partial files into a single panel-level file. For each cell, the panel records the post-$\bar{t}$ mean of $\tilde{n}(t)$ under each perturbed modeling scenario,

$$\bar{n}_{\text{ModEnv}} \;\equiv\; \frac{1}{46 - \bar{t}} \sum_{t=\bar{t}}^{45} \tilde{n}(t,\, \text{ModEnv}), \qquad \text{ModEnv} \in \{2, 3\},$$

together with the vignette policy stimuli and the respondent's stated number of children. The columns $\bar{n}_2$ and $\bar{n}_3$ are the Model 1 panel's analogs of Model 2's completed fertility $\overline{n_{45}}$, and serve as the dependent variable in the FE regressions of Section 6.2 and the elasticity computations of Section 6.3.

F.1.2 Out-of-sample validation

Section 6.2 in the main body of the paper reports the cross-validation of the vignette panel and the simulated panels of Models 1 and 2 (see Table 13).

Here we provide Model-1-specific technical details: the vignette-panel cell-level sample composition, the predicted-vs-stated fertility distribution by 2×2 policy blocks (transfer × costly care), and the fixed-effects regression under three Model 1 wage-perception conventions (permanent, current period, permanent but evaluated at the respondent age).

Data and sample The vignette panels from PAIRFAM (Germany; satellite vignette module of wave 5, 2013/14) and UAS module 27 (US) are restricted to female respondents aged

22–42: 1,128 observations from 282 female respondents in Germany (four randomized environments per respondent), and 3,056 observations from 764 female respondents in the United States (of which 3,001 have a stated number of children). The structure of the Model 1 simulated panel and the vignette panel is the same: one row per vignette-panel observation.

Distributions of stated and predicted number of children Table F.1 reports the distribution of completed fertility under each of the four 2×2 policy blocks (transfer × costly care) for the vignette panel and the Model 1 simulated panel under modeling scenario 3 (permanent-wage treatment). Predicted values are evaluated at the baseline Model 1 estimate $\hat{\theta}_{\mathrm{M1}}$ in Table 3 in the main body of the paper.

Table F.1: Distribution of number of children by 2×2 policy block: vignette panel vs. Model 1 simulated panel (env-3 Permanent)

Block	Panel	0	1	2	3+	Mean	s.d.	N
US								
Low transfer / Subsidized	Vignette	18.2	20.3	33.6	27.9	1.89	1.38	699
	Sim Perm	0.0	0.0	0.0	100.0	3.00	0.01	707
High transfer / Subsidized	Vignette	16.4	17.8	40.1	25.7	1.90	1.35	751
	Sim Perm	0.0	0.0	0.0	100.0	3.00	0.01	772
Low transfer / Costly	Vignette	17.9	22.6	37.1	22.3	1.80	1.36	743
	Sim Perm	9.5	38.4	38.8	13.2	1.52	0.76	755
High transfer / Costly	Vignette	17.5	23.5	38.6	20.4	1.75	1.28	808
	Sim Perm	6.7	39.1	37.6	16.7	1.62	0.78	822
DE								
Low transfer / Subsidized	Vignette	11.4	25.0	43.2	20.3	1.81	1.10	236
	Sim Perm	0.0	0.0	0.0	100.0	3.00	0.00	236
High transfer / Subsidized	Vignette	10.2	18.4	43.0	28.5	2.01	1.13	305
	Sim Perm	0.0	0.0	0.0	100.0	3.00	0.00	305
Low transfer / Costly	Vignette	17.1	27.6	40.4	14.9	1.60	1.12	275
	Sim Perm	0.0	51.3	34.2	14.5	1.64	0.67	275
High transfer / Costly	Vignette	11.9	24.7	41.3	22.1	1.80	1.06	312
	Sim Perm	0.0	41.3	42.0	16.7	1.75	0.65	312

Note: Cell-level frequencies (%) of integer-rounded number of children; Mean and s.d. computed from continuous predictions. "Sim Perm" = modeling scenario 3 (permanent-wage treatment) at $\hat{\theta}_{\mathrm{M1}}$ with per-age $\min(n, 3)$ cap. Transfer "Low / High" corresponds to monthly child allowance \$200 / \$400 (US) or 250 / 500 Euros (DE); "Subsidized" maps to $\tilde{p}_b = 0.05 \cdot \tilde{p}_b^{\mathrm{anchor}}$, "Costly" to $\tilde{p}_b = \tilde{p}_b^{\mathrm{anchor}}$. Vignette US has $N = 3{,}001$ overall because 55 of the 764 female respondents did not state a number of children.

The 2×2 breakdown reveals policy-specific patterns. Under subsidized care (CC= 0), Model 1's Permanent-wage treatment saturates at the $n = 3$ cap in 100% of cells in both countries and across both transfer levels (Mean = 3.00, s.d. = 0.00). The vignette distribution under subsidized care, by contrast, spans the full support: in the US, vignette means are 1.89 and 1.90 under low and high transfer; in DE, 1.81 and 2.01. “Subsidized care” in the vignette corresponds to a near-free daycare condition, an informative counterfactual but an unrealistic real-world scenario. The implied substantial drop in the relative price of children raises fertility substantially, and Model 1's predictions reach the upper bound of the support across the entire sample. Under “costly care,” by contrast, daycare prices remain at empirically observed levels, and the predicted distribution moves back into a range comparable to the vignette responses. Under costly care (CC= 1), Model 1's predicted distribution matches the vignette distribution closely in both countries: low/costly US means 1.52 (sim) vs. 1.80 (vig); high/costly US means 1.62 vs. 1.75; low/costly DE means 1.64 vs. 1.60; high/costly DE means 1.75 vs. 1.80. The simulated distribution is more concentrated than the vignette (sd $\approx$ 0.7–0.8 vs. vignette $\approx$ 1.1–1.4), as expected for an aggregative deterministic model.

This dichotomy is a strength of the Model 1 design rather than a weakness. The vignette presents the daycare stimulus as a binary 0/1 contrast, “costly care” vs. “fully subsidized care”, and Model 1 reproduces the qualitative pattern that this design is intended to elicit: fertility responds sharply under full subsidization of care and moderately at the empirically grounded costly end. That the model correctly separates the two cases is itself a structural validation; it tells us that Model 1's optimization-based fertility decisions move in the right qualitative direction under large changes in the relative price of children. Having established this qualitative similarity, we proceed to the quantitative cross-validation of realistic 5–20% policy variations in Section 6.3 and the structural elasticity computation in Section 6.4 of the main body of the paper, where the smaller policy perturbations stay well within the interior of the support and the cap is not binding.

Fixed-effects regression: vignette panel vs. Model 1 simulated panel Table F.2 reports the linear fixed-effects regression with individual-level cluster-robust standard errors of the number of children on the four randomized vignette stimuli (hourly wage, partner earnings, transfer, costly care) for the vignette panel and for three Model 1 wage-perception conventions: *Permanent* (env-3 from Section F.1.1), *current-period* (env-2, the vignette wage applied at t_0 only), and *Permanent but evaluated at the respondent age* (env-3 evaluated at the respondent's vignette age a^* rather than the post-$\bar{t}$ mean).

All three Model 1 wage conventions agree with the vignette on the sign of every coefficient: hourly wage positive, partner earnings ≈ 0, transfer positive, costly care strongly negative. The magnitudes differ across conventions. The costly-care coefficient is roughly 6–8$\times$ larger in absolute value than the vignette's $-0.189/-0.250$ in both countries, because Model 1's aggregative continuous-n representation amplifies the per-cell response, the same architectural pattern flagged in Section 6.2 of the main body of the paper. The hourly-

Table F.2: Fixed-effects regression of number of children on vignette stimuli (vignette panel vs. Model 1 simulated panel, three wage-perception treatments)

	Vignette	Sim Perm	Sim Current-period	Sim Perm (at age a^*)
US				
hourly wage	+0.014***	+0.017***	+0.001***	+0.013***
	(0.002)	(0.001)	(0.0001)	(0.001)
partner earnings (per \$1,000)	+0.118***	-0.003	-0.001	-0.001
	(0.017)	(0.005)	(0.0005)	(0.004)
child allowance (per \$100)	+0.032	+0.023***	+0.012***	+0.022***
	(0.024)	(0.007)	(0.001)	(0.005)
costly care (0,1)	-0.189***	-1.321***	-1.372***	-1.309***
	(0.040)	(0.033)	(0.003)	(0.029)
R^2 (within)	0.071	0.715	0.997	0.807
N	3,001	3,056	3,056	2,468
DE				
hourly wage	+0.020***	+0.029***	+0.003***	+0.026***
	(0.005)	(0.003)	(0.0002)	(0.002)
partner earnings (per €1,000)	+0.240***	-0.019	-0.003**	-0.013
	(0.038)	(0.012)	(0.002)	(0.011)
child allowance (per €100)	+0.132***	+0.014	+0.019***	+0.037***
	(0.035)	(0.010)	(0.001)	(0.007)
costly care (0,1)	-0.250***	-1.229***	-1.661***	-1.222***
	(0.046)	(0.043)	(0.005)	(0.037)
R^2 (within)	0.154	0.758	0.998	0.821
N	1,128	1,128	1,128	1,116

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Female respondents aged 22–42. Simulated panels generated at $\hat{\theta}_{\mathrm{M1}}$ (Section 4.4 of the main body of the paper). *Sim Perm* = env-3 (vignette wage applied for all $t \geq t_0$, post-$\bar{t}$ mean). *Sim Current-period* = env-2 (vignette wage applied at t_0 only, post-$\bar{t}$ mean). *Sim Perm (at age a^*)* = env-3 (Permanent) evaluated at the respondent's vignette age a^* rather than the post-$\bar{t}$ mean. The Sim Perm (at age a^*) column has slightly smaller N because some respondents are at ages outside the panel evaluation window.

wage coefficient is closest to the vignette under the Permanent and the Permanent-but-evaluated-at-the-respondent-age treatments (0.017 and 0.013 vs. vignette 0.014 in the US; 0.029 and 0.026 vs. vignette 0.020 in DE), while the Current-period treatment produces a near-zero coefficient (0.001/0.003) because the one-period vignette wage perturbation has limited propagation through the deterministic lifecycle. The within-R^2 values rise sharply from the vignette panel (0.071/0.154) to all three simulated-panel columns (0.72–1.00), consistent with Model 1 generating systematically more deterministic responses than the survey respondents.

A powerful feature of the vignette module is its randomization design, embedded directly within the representative household panels of both countries. Each respondent is presented with the same set of randomized vignette environments, so the regression coefficients in Table F.2 identify the causal partial derivatives of fertility with respect to each stimulus from within-respondent variation, free of the selection and identification challenges that arise in observational policy variation. The Model 1 simulated panel inherits this randomization design and combines it with a separate strength: the structural framework is calibrated to established representative household surveys of both countries (PSID and CPS for the US, SOEP for Germany), through which the deep parameters $\hat{\theta}_{\mathrm{M1}}$ are calibrated to real-world age profiles of fertility, labor supply, consumption, assets, and childcare. The simulated panel therefore inherits both pieces simultaneously, experimental randomization at the within-respondent level and real-world grounding at the population level, which is what makes the cross-validation here a credible test of the structural model.

F.2 Model 2: heterogeneous and stochastic

F.2.1 Construction of the simulated panel

The construction of the Model 2 simulated panel is described conceptually in Section 6.1 of the main body of the paper. Here we provide the technical implementation details: the per-respondent dynamic-programming solve, the variable-by-variable mappings of vignette stimuli into Model 2's wage process and budget constraint, the initial-state initialization, and the forward-simulation procedure that produces the completed-fertility outcome per vignette-panel cell.

For each vignette-panel cell, we solve Model 2's full dynamic programming problem at the estimated structural parameters $\hat{\theta}_{\mathrm{M2}}$, holding $\hat{\theta}_{\mathrm{M2}}$ fixed throughout. Because Model 2 has a four-type unobserved-heterogeneity structure (defined in equation 19 in the main body of the paper) through the binary attitudinal switches $(O_1, O_2) \in \{0,1\}^2$—where O_1=1 scales child-related utility by $(1-\alpha_6)$ and O_2=1 scales childcare-time disutility by $(1-\gamma_5)$—

each cell requires four separate dynamic programming solutions (one per type), and the simulated outcome is the mixture across types with independent type probabilities π_{O_1} and π_{O_2} (estimated at the anchor). With 764 US respondents and 282 DE respondents each evaluated on four environments, the US sample yields 3,056 cells and the DE sample yields 1,128 cells. In total, we solve Model 2 12,224 times in the US and 4,512 times in Germany. Each solution is paired with $N_{\text{SIM}} = 10{,}000$ stochastic forward simulations of the lifecycle, from which the cell-level completed fertility $\overline{n_{45}}$ is obtained as the mean across simulated paths and across the four (O_1, O_2) types (π-weighted). This gives a total of 122.24 million simulated paths in the US and 45.12 million in Germany. This computationally intensive procedure contrasts sharply with Model 1's deterministic aggregative solver.

Each randomized vignette environment is characterized by five attributes: hourly net wage rate (w_e), partner working hours ($h_{p,e}$), partner monthly earnings ($y_{p,e}$), availability of childcare facilities ($s_e \in \{\text{costly, subsidized}\}$), and monthly allowance per child (A_e). We map these attributes one-for-one into Model 2's environment as follows.

- Wage. The vignette wage is fed directly into the income function by replacing Model 2's wage process. Specifically, in the specification $w_t = \exp(\phi_0 + \phi_1 x_t + \phi_2 x_t^2 + \xi_t)$, we set $\phi_1 = \phi_2 = 0$ and $\phi_0 = \log(w_e)$ for each cell, so that the deterministic component of the wage equals exactly w_e at every age (the productivity shock $\xi_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$ continues to apply, drawn independently each period from its discretized S-point grid; the vignette substitution does not alter the shock distribution).

- Partner income. The partner-income profile $y_{p,t}$ is replaced with a constant annual value $12 \cdot y_{p,e}$ at every age.

- Child allowance. The child-allowance term A in Model 2's budget constraint (entering as $n_t \cdot A$ in the cash-flow expression $\tilde{c}_t = c_t + n_t(1-s)p_{b,t}m_{b,t} - n_t \cdot A$, see Section 4 of the main body of the paper) is set to $12 \cdot A_e$, converting the vignette's per-child monthly amount to a per-child annual amount.

- Daycare cost. The daycare subsidy rate s in Model 2's budget constraint (entering as the household-paid daycare-cost term $n_t(1-s)p_{b,t}m_{b,t}$ within $\tilde{c}_t$, see Section 4 of the main body of the paper) is set to $s = 0$ when $s_e = \text{costly}$ (Model 2's baseline calibration value, representing no subsidy) and $s = 0.95$ when $s_e = \text{subsidized}$ (a near-free daycare condition, the maximum stable subsidy within Model 2's numerical range).

We initialize each respondent's lifecycle continuation at her actual life stage. The simulation starts at period $t_0 = \text{age} - 21$ (so a respondent aged 30 starts at $t_0 = 9$), with $n_0 = \min(\text{observed children}, 3)$ existing children, marital status as observed in the survey, initial experience $x_0 = \text{age} - 22$ (assuming continuous accumulation since age 22), and age of youngest child $a_0 = 3$ (a uniform default since the vignette is silent on this dimension; the value is irrelevant for respondents with no existing children). For each unique vignette environment we solve Model 2's dynamic program once and forward-simulate $N_{\text{SIM}} = 10{,}000$ stochastic draws of fertility/labor-supply trajectories from the respondent's initial state up to the end of fecundity (age 45). Predictions are π-weighted across the four heterogeneity types $(O_1, O_2) \in \{0, 1\}^2$ following Model 2's baseline specification. The structural estimated parameters $\hat{\theta}$ are taken from Table 4 of the main body of the paper and held fixed throughout.

For each vignette-panel cell we obtain two model-implied outcomes: the total number of children at age 45, $n^{\text{sim}} = \overline{n}_{45}$, and the labor supply at the respondent's actual age, $\ell^{\text{sim}} = \overline{\ell}_{t_0} \cdot 510$ (converted to monthly hours, with $510 = H/12$ where $H = 6{,}120$ is Model 2's annual time endowment), where $\overline{\cdot}$ denotes the average across the N_{SIM} stochastic draws and the four (O_1, O_2) types.

Because survey respondents are not told whether the vignette wage variation is permanent or temporary, two wage-perception conventions are natural: *permanent-wage* and *current-period-wage*. The construction described above corresponds to *permanent-wage*: a single dynamic program is solved at the modified (ϕ_0, ϕ_1, ϕ_2), and the vignette wage applies for all $t \geq t_0$. *current-period-wage*, introduced as a robustness check in Section F.2.2 below, confines the vignette wage to the initialization period via a two-DP formulation: at t_0 the agent uses the policy from DP_{Perm} (solved at the vignette ϕ_0), and from $t_0 + 1$ onward she uses the policy from DP_{N} (solved at the respondent's natural, estimated wage process). All non-wage attributes (A_e, s_e, $y_{p,e}$) are applied identically across both DPs. The two conventions span the natural range of perception assumptions: Permanent-wage applies the vignette wage to all $T - t_0$ remaining periods, while Current-period-wage applies it to a single period.

Formally, let $d^{DP}(s, t)$ denote Model 2's policy function (over labor-supply and fertility choices) at state s and age-period t under dynamic program DP, following the discrete-choice lifecycle convention of Eckstein and Wolpin [1989] and Keane and Wolpin [1997]. We solve two DPs per vignette-panel cell:

$$\begin{aligned} DP_{\text{Perm}}: &\quad \text{solve with } (\phi_0, \phi_1, \phi_2) = (\log w_e,\ 0,\ 0) \text{ and non-wage substitutions } (A,\ s,\ y_{p,t}), \\ DP_{\text{N}}: &\quad \text{solve with the estimated } (\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2) \text{ and identical non-wage substitutions } (A,\ s,\ y_{p,t}). \end{aligned}$$

Under *permanent-wage*, the agent uses $d^{DP_{\text{Perm}}}(\cdot, t)$ for every $t \geq t_0$ (single-DP execution).

Under *current-period-wage*, the agent uses the piecewise policy

$$d^{\mathrm{CP}}(s,t) = \begin{cases} d^{DP_{\mathrm{Perm}}}(s,t_0) & t = t_0, \\ d^{DP_{\mathrm{N}}}(s,t) & t > t_0. \end{cases}$$

Concretely, the realized wage in the budget equation is $w_t = w_e$ at $t = t_0$ (vignette wage) and $w_t = \exp(\hat{\phi}_0 + \hat{\phi}_1 x_t + \hat{\phi}_2 x_t^2)$ for $t > t_0$ (natural Mincer path), with $\xi_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$ drawn each period in both cases and entering as $\text{income}_t = w_t \cdot \ell_t \cdot \exp(\xi_t)$. All non-wage attributes (A, s, $y_{p,t}$) are applied identically across both DPs and across all periods.

F.2.2 Out-of-sample validation

Section 6.2 of the main body of the paper reports the joint cross-validation. Here we provide Model-2-specific technical details, following the same layout as Section F.1.2 above, with the FE regression run under two Model 2 wage-perception conventions (Permanent and Current-period).

Data and sample The vignette panel is identical to that used for Model 1 in Section F.1.2. The Model 2 simulated panel has the same row structure (one row per respondent × vignette cell) and is built per the technical construction in Section F.2.1 above.

Distribution of stated and predicted number of children Table F.3 reports the distribution of completed fertility under each of the four 2×2 policy blocks (transfer × costly care) for both the vignette panel and the Model 2 simulated panel under the Permanent-wage treatment. Predicted values are evaluated at the baseline Model 2 estimate $\hat{\theta}_{\mathrm{M2}}$ in Table 5 in the main body of the paper.

In both countries and across all four blocks, Model 2's predicted distribution stays close to the vignette. The Model 2 means range from 1.76 to 2.24 in the US (vignette: 1.75 to 1.90) and from 1.78 to 1.86 in DE (vignette: 1.60 to 2.01). Unlike Model 1, even under subsidized care, the Model 2 distribution remains spread out across the support: even at the highest $s = 0.95$ subsidy, the simulated distribution maintains substantial mass on $n = 0$, 1, and 2, with at most 40% of cells in the US (and $\leq 12\%$ in DE) hitting the $n = 3$ boundary state.

This behavior reflects Model 2's heterogeneous-agent structure: the four-type (O_1, O_2) mixture distributes the fertility response across types whose preferences for children differ by $(1 - \alpha_6)$ and $(1 - \gamma_5)$, and the $N_{\mathrm{SIM}} = 10{,}000$ stochastic forward simulations introduce additional within-type variation. The same extreme policy variation that drove Model 1's

Table F.3: Distribution of number of children by 2×2 policy block: vignette panel vs. Model 2 simulated panel (Permanent-wage)

Block	Panel	0	1	2	3+	Mean	s.d.	N
US								
Low transfer / Subsidized	Vignette	18.2	20.3	33.6	27.9	1.89	1.38	699
	Sim Perm	8.8	18.8	40.2	32.2	1.93	0.86	707
High transfer / Subsidized	Vignette	16.4	17.8	40.1	25.7	1.90	1.35	751
	Sim Perm	3.5	8.8	47.7	40.0	2.24	0.65	772
Low transfer / Costly	Vignette	17.9	22.6	37.1	22.3	1.80	1.36	743
	Sim Perm	9.5	32.2	35.1	23.2	1.76	0.90	755
High transfer / Costly	Vignette	17.5	23.5	38.6	20.4	1.75	1.28	808
	Sim Perm	6.8	8.9	50.1	34.2	2.13	0.76	822
DE								
Low transfer / Subsidized	Vignette	11.4	25.0	43.2	20.3	1.81	1.10	236
	Sim Perm	2.1	7.6	82.2	8.1	1.81	0.52	236
High transfer / Subsidized	Vignette	10.2	18.4	43.0	28.5	2.01	1.13	305
	Sim Perm	0.0	8.9	79.7	11.5	1.86	0.54	305
Low transfer / Costly	Vignette	17.1	27.6	40.4	14.9	1.60	1.12	275
	Sim Perm	3.3	16.0	68.7	12.0	1.78	0.59	275
High transfer / Costly	Vignette	11.9	24.7	41.3	22.1	1.80	1.06	312
	Sim Perm	1.6	5.1	81.4	11.9	1.83	0.56	312

Note: Cell-level frequencies (%) of integer-rounded number of children; Mean and s.d. computed from continuous predictions. "Sim Perm" = Permanent-wage treatment at $\hat{\theta}_{\mathrm{M2}}$. Transfer "Low / High" corresponds to monthly child allowance \$200 / \$400 (US) or 250 / 500 Euros (DE); "Subsidized" maps to $s = 0.95$, "Costly" to $s = 0$. Vignette US has $N = 3{,}001$ overall because 55 of the 764 female respondents did not state a number of children.

aggregative representative agent to the corner is absorbed by Model 2's cross-sectional and idiosyncratic heterogeneity. The Model 2 simulated distribution is, like Model 1's, more concentrated than the vignette (s.d. $\approx$ 0.5–0.9 vs. vignette $\approx$ 1.1–1.4), reflecting the smoothing effect of mixing over types and stochastic paths.

The contrast between Model 1 and Model 2 documented across Sections F.1.2 and F.2.2 is precisely why we report two complementary structural models in this paper. As we noted in Section 6.3 in the main body of the paper, "in structural models, estimated policy responses depend on, among others, the stipulated behavioral framework, functional forms, and other technical assumptions (see Todd and Wolpin, 2023, p. 44); potentially, many model structures could match the same estimation targets equally well, but may have different policy implications." Model 1's aggregative continuous-n framework and Model 2's heterogeneous-agent discrete-state framework are exactly two such alternative structures: both match their respective estimation targets (Section 3 for Model 1, Section 4 for Model 2), yet they generate visibly different policy implications under the same extreme vignette stimulus, Model 1 piles up at the $n = 3$ boundary, while Model 2 spreads across the support. Reporting both models side by side, and cross-validating each against the same vignette environment, makes the dependence of policy implications on modeling assumptions transparent, and provides a model-choice confidence interval on any single policy conclusion: Model 1's and Model 2's estimates jointly bound the policy implication, and the width of that interval is itself an indicator of how robust the conclusion is. A narrow gap (such as on the sign of the costly-care coefficient) signals robustness; a wide gap (such as on the magnitude of the costly-care coefficient) signals that the conclusion depends on the modeling assumption and, accordingly, should be interpreted cautiously.

The two models also differ sharply in computational cost. Model 2 is a heterogeneous-agent dynamic-programming model with stochastic forward simulation: as documented in Section F.2.1, evaluating the model at $\hat{\theta}_{\mathrm{M2}}$ across the vignette panel requires 122.24 million simulated paths in the US and 45.12 million in Germany, with each cell solved via a full backward induction over the discrete state space. Model 1, by contrast, is an aggregative deterministic lifecycle model: its inner equilibrium admits a closed-form solution at $\eta = 1$, which we use as a well-educated warm-start guess. Combined with the warm-start, the single point-estimate MSM for Model 1 completes in well under 2 minutes per country on a single machine, against Model 2's days-long run at $N_{\mathrm{SIM}} = 10{,}000$. This computational asymmetry is itself a methodological complementarity: Model 1's speed makes extensive robustness checks tractable, while Model 2's richness captures the cross-sectional heterogeneity that Model 1 cannot.

Fixed-effects regression: vignette panel vs. Model 2 simulated panel Table F.4 reports the linear fixed-effects regression with individual-level cluster-robust standard errors of the number of children on the four randomized vignette stimuli (hourly wage, partner earnings, transfer, costly care) for the vignette panel and for two Model 2 wage-perception conventions: *Permanent* (vignette wage applied for all $t \geq t_0$, single-DP solve) and *current-period* (vignette wage applied at t_0 only, two-DP construction described in Section F.2.1).

All three columns (vignette and both Model 2 wage conventions) agree on the qualitative pattern: costly care reduces fertility ($\beta < 0$ at the 1% level in every column), transfers raise it (positive in vignette and in both Model 2 columns for both countries), and the hourly-wage coefficient is positive in the US (across vignette, Permanent, and Current-period). The one departure is the DE hourly-wage coefficient, which flips sign under Model 2: the Permanent construction yields a small negative coefficient (-0.003^{***}) while the Current-period treatment renders the response statistically indistinguishable from zero (-0.001, *n.s.*). The typical ordering $|\beta^{\text{vignette}}| > |\beta^{\text{Model 2}}|$ established in the opening of this Appendix holds for the costly-care coefficient in both countries and for the DE transfer coefficient.

The closeness of the Permanent and Current-period columns shows that Model 2's responsiveness to the four policy stimuli is largely robust to the wage-perception assumption. Transfer and costly care deliver coefficients that are essentially identical across the two perception specifications in both countries (US transfer $+0.185 \to +0.182$ per \$100, US costly care $-0.108 \to -0.116$; DE transfer $+0.012 \to +0.013$ per €100, DE costly care $-0.044 \to -0.049$). Partner earnings are negligible in both columns and in both countries. The hourly-wage coefficient is the only place where the wage-perception assumption matters to first order, since the vignette wage is the stimulus being perceived; even there, the two constructions deliver the same sign in the US ($+0.002 \to +0.001$) and a quantitatively reduced but still non-positive coefficient in DE ($-0.003 \to -0.001$). The economic logic is that the wage process is the model's numéraire and scales relative prices throughout the budget constraint: changing the perception of the vignette wage shifts the marginal-response magnitudes but cannot reverse them, and the within-respondent partial derivatives that the FE regression identifies are pinned down primarily by structural preferences (taste for children, disutility of work, curvature of utility) and by the agent's natural lifecycle wage profile, both of which are common to the two constructions.

To conclude, Model 2 is substantially more complex than Model 1, with a heterogeneous-agent dynamic-programming structure, stochastic forward simulation, four unobserved-heterogeneity types, and the discrete state space $\{0, 1, 2, 3+\}$ documented in Section 4 in the main body of the paper, against Model 1's aggregative deterministic representative-agent formulation. Yet the cross-validated policy responses to the vignette policy stimuli are quantitatively

Table F.4: Fixed-effects regression of number of children on vignette stimuli (vignette panel vs. Model 2 simulated panel, permanent- and current-wage treatment)

	Vignette	Sim Permanent	Sim Current-period
US			
hourly wage	+0.014***	+0.002***	+0.001***
	(0.002)	(0.0003)	(0.0003)
partner earnings (per \$1,000)	+0.118***	+0.006	+0.011**
	(0.017)	(0.004)	(0.004)
child allowance (per \$100)	+0.032	+0.185***	+0.182***
	(0.024)	(0.016)	(0.016)
costly care (0,1)	-0.189***	-0.108***	-0.116***
	(0.040)	(0.007)	(0.008)
R^2 (within)	0.071	0.359	0.363
N	3,001	3,056	3,056
DE			
hourly wage	+0.020***	-0.003***	-0.001
	(0.005)	(0.001)	(0.0004)
partner earnings (per €1,000)	+0.240***	-0.003	-0.010***
	(0.038)	(0.002)	(0.003)
child allowance (per €100)	+0.132***	+0.012***	+0.013***
	(0.035)	(0.002)	(0.002)
costly care (0,1)	-0.250***	-0.044***	-0.049***
	(0.046)	(0.006)	(0.007)
R^2 (within)	0.154	0.244	0.216
N	1,128	1,128	1,128

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Female respondents aged 22–42. $\hat{\theta}_{\mathrm{M2}}$ from Table 5 in the main body of the paper. *Sim Permanent* = vignette wage applied for all $t \geq t_0$ (single-DP solve). *Sim Current-period* = vignette wage applied at t_0 only via two-DP construction: at t_0 the agent uses the policy from DP_{Perm} (solved under the vignette wage permanent), and from $t_0 + 1$ onward she uses the policy from DP_{N} (solved under the agent's natural wage process). Non-wage attributes (A, s, $y_{p,t}$) are applied identically across both constructions.

similar across the two models in sign, statistical significance, and broad order of magnitude (Table F.2 and Table F.4). This convergence of structurally distinct frameworks on the same vignette-based answer is itself an important finding: it shows that the combination of *vignette experiments* with *structural anchoring on representative household surveys* is robust to substantial variation in the structural model's behavioral assumptions, and that the methodological construction of pairing the two ingredients delivers stable policy implications. The vignette provides the experimental rigor; the structural model provides the counterfactual power; and the agreement we document here is the working demonstration of what Todd and Wolpin [2023] call the "best of both worlds," a methodological construction worth carrying beyond this paper.

Online Computational Appendix

for "Combining a Survey Experiment with Lifecycle Models to Evaluate Pronatalist Policies"

Joshua Goldstein [*] Christos Koulovatianos [†‡] Jian Li [§]
Carsten Schroeder [¶‖]

This draft: July 2, 2026

Abstract

This online appendix provides the detailed computational and algorithmic description of the lifecycle models, estimation procedures, and ancillary constructions used in the main paper. Section 2 documents the data pipeline; Section 3 and Section 4 describe Model 1 and Model 2; Section 5 the vignette simulation panel; Section 6 the elasticity construction; Section 7 the structure of the replication package; and Section 8 reports an end-to-end verification run of the full package from raw microdata to paper outputs.

[*] UC Berkeley.

[†] University of Luxembourg.

[‡] Center for Financial Studies (CFS), Goethe University Frankfurt.

[§] Institute for Advanced Economic Research, Dongbei University of Finance and Economics. Department of Finance, University of Luxembourg.

[¶] Department of Economics, Freie Universität Berlin.

[‖] German Socio-Economic Panel, German Institute for Economic Research, Berlin.

Contents

1 Introduction

This Online Computational Appendix accompanies the paper *Combining a Survey Experiment with Lifecycle Models to Evaluate Pronatalist Policies*. The paper presents two structural lifecycle models (Model 1, aggregative and deterministic; Model 2, heterogeneous-agent and stochastic), estimates them on United States and German microdata, validates them against a randomized-vignette survey experiment, and reports the elasticities of fertility with respect to a per-child allowance and a childcare-price subsidy. This appendix documents the *computational implementation* of every component: the data pipeline that produces the model-ready inputs (Section 2), the two structural-model implementations and their estimation procedures (Section 3, Section 4), the vignette-anchored simulated panel that links the survey responses to the structural estimates (Section 5), the three elasticity constructions that summarize the paper's policy results (Section 6), and a guide to the runnable replication package (Section 7).

Relationship to the paper and replication package. This appendix is one of three documents that together constitute the paper's full computational documentation. The *paper* gives the structural derivations, identification arguments, and substantive results. *Appendix A in the paper* gives the narrative variable-construction details for every empirical moment. The *Online Computational Appendix* (this document) gives the algorithmic and computational implementation, focusing on what the code actually computes (parameter ranges, sign conventions, discretization choices, numerical solver structure, byte-equivalence verifications). The runnable code itself appears in the companion `IER_R2_Replication/` package; the appendix and the package are designed to be read together. Section 7 provides the package's entry point.

Cross-reference conventions. Internal cross-references within the OCA follow the paper's conventions and appear as Section 3, (17), Table 1, etc. References to the paper use prose ("Section 4 in the paper", "Table 11 in the paper", "Figure 5 in the paper") because the OCA and the paper compile separately. References to the replication package use file paths (`IER_R2_Replication/docs/FILE_MAP.md`, etc.) anchored at the package root.

1.1 Glossary

The following acronyms and short codes recur throughout this appendix.

Methods.

MSM	Method of Simulated Moments
DP	Dynamic programming (Bellman recursion)
FE	Fixed effects (Table 14 vignette-vs-simulated panel regressions)
SE	Standard error
NLopt	Julia nonlinear-optimization library (M1 and M2 estimation)

Data sources.

PSID	Panel Study of Income Dynamics (US; M1, M2)
SCF	Survey of Consumer Finances (US; M1 wealth, 2007 wave)
CPS	Current Population Survey (US; M1 labor supply)
ATUS	American Time Use Survey (US; M1, M2 time use)
NLSY79	National Longitudinal Survey of Youth 1979 (US; M2 π priors)
UAS-27	Understanding America Study, wave 27 (US; vignettes)
SOEP	German Socio-Economic Panel (DE; M1, M2)
Pairfam	Panel Analysis of Intimate Relationships and Family Dynamics (DE; vignettes, 2014 satellite module)

Model and version codes.

M1	Model 1 (lifecycle dynamic-programming model; canonical estimate = v14)
M2	Model 2 (M1 + 4-type mixture; canonical estimate = V16)
Pb	Price-of-children base (policy stimulus in elasticity exercises)
A	Allowance / per-child cash transfer (policy stimulus)
Subsidy95	M1 vignette simpanel protocol (sub = 0.95 with per-age cap, Section 5.4)

Pipeline and process tags.

L1, L2, L3a, L3b, L4	Data-pipeline layered-build stages (Section 2.2)
DUA	Data Use Agreement (required for restricted-access raw surveys)
OCA	Online Computational Appendix (this document)
π_{O_1}, π_{O_2}	Latent type-share parameters in M2's 4-type mixture

2 Data pipeline

2.1 Overview

The data pipeline is the upstream layer of the replication package: it takes survey microdata (PSID, SCF, CPS, ATUS, NLSY79, and the UAS vignette survey for the United

States; SOEP and the Pairfam vignette survey for Germany) and produces the model-ready data inputs that Models 1 and 2 consume (Section 3.3, Section 4.3) and the moment vectors that the MSM estimators target (Section 3.4, Section 4.4). The implementation appears in `data_processing/` of the replication package.

Paper Appendix A documents the construction of each data variable (definitions, sample restrictions, transformations); this section focuses on the *computational* aspects of the pipeline: layered architecture (Section 2.2), random seeds and bootstrap conventions (Section 2.3), price-base and deflation conventions (Section 2.4), and byte-equivalence verification.[1] The pipeline is implemented in Stata 17 (∼25 `.do` files) with two small Julia helpers.

2.2 Architecture

The pipeline is structured as a per-country directory tree with parallel substructure for the United States (`scripts/us/`) and Germany (`scripts/de/`). Each country tree has the following layered subdirectories:

Subdirectory	Layer	Role
`variables/`	L1	Per-source raw-variable extraction from PSID, SCF, SOEP, etc.; one `.do` per source
`full_sample/`	L1→L2	Cross-source merging and panel assembly
`model_sample/`	L2	Theory-driven sample restrictions and unit-of-observation choices (M1 couples-only, M2 includes singles, etc.)
`summary_stats/`	L3a	Auxiliary summary statistics and `aux_check` verifications
`moments/`	L3b	Model-targeted moments per (model, country, bootstrap draw); 7 moment groups for M1 (byte-exact target)
`deliverables/`	L4	Final `.csv`/`.jld2` outputs that Models 1 and 2 consume directly; one aggregator script per model (`m1.do`, `m2.do`)
`figures/`	auxiliary	Country-specific Appendix A figures
`verify/`	auxiliary	Dev-time byte-equivalence checks

Layer numbers (L1, L2, L3a, L3b, L4) appear in the file headers; the L2 designation for `model_sample/` is fixed by `scripts/_config.do`'s design principle:

[1] Reproducibility is enforced by a `verify/` harness that re-derives every moment and model input and asserts agreement with the reference to a strict relative tolerance, ranging from 10^{-13} at the moment layer to 10^{-6} at the lifecycle panels. The exact per-block assertions reside in the package's `verify/` scripts.

"single seed, single bootstrap protocol; sample restrictions and unit-of-observation differences are theory-driven and applied at the model_sample layer (L2)."

Path registry. All paths and global constants are in `scripts/_config.do`. The registry exposes `PROJECT_ROOT` (the IER R2 root; settable via environment variable for relocation), `V2_ROOT`, `V2_SCRIPTS`, `V2_DATA`, `RAW_DATA_ROOT`, plus per-country output roots (`VARS_US/DE`, `FULL_SAMPLE_US/DE`, `MODEL_SAMPLE_US/DE`, `MOMENTS_US/DE`, `DELIVERABLES_US/DE`, etc.). Every per-stage `.do` file begins with `include"${V2_SCRIPTS}/_config.do"` to inherit these paths.

Shared helpers. Six `.do` files in `scripts/` provide cross-stage utilities used by both country pipelines: `_utils.do` (sample-builder primitives, weight handling, `collapse` wrappers); `_utils_moments.do` (moment-computation helpers); `_utils_stats.do`, `_utils_stats_extra.do` (summary-statistic helpers); `_utils_figures.do`, `_utils_figures_extra.do` (plotting helpers for Appendix A figures). A small Python script `scripts/_reformat_table_captions.py` post-processes the Appendix A `.tex` output to enforce a consistent caption format across the ~ 44 tables.

Aggregator scripts. Each country's `deliverables/` subdirectory contains two aggregator scripts. `m1.do` consumes the L3b per-block CSVs plus the `m1_psid_enriched` dataset and produces Model 1's $\sim 1{,}500$ output `.csv` files per country (panels A–E, the M2-anchor columns 10–11 for cross-model alignment, the M moment file, and the A-binned asset file). `m2.do` is the parallel aggregator for Model 2.

2.3 Random seeds and bootstrap conventions

The pipeline uses a single unified seed and a single bootstrap protocol across both models and both countries.

Single unified seed. `scripts/_config.do` declares

- `$NBS = 300` — the number of bootstrap draws,
- `$SEED = 12345` — the unified seed for cluster bootstrap across M1 and M2.

The same `$SEED` value is set once inside the helper function `stamp_bw` (`scripts/_utils.do`) before any `bsample` call, so every per-source bootstrap shares the same RNG state.

Bootstrap procedure: `stamp_bw`**.** For each source dataset, the helper `stamp_bw [`*`cluster_var`*`]` attaches NBS $+1=301$ bootstrap-weight variables `bw_0`, `bw_1`, ..., `bw_300` to every observation:

- `bw_0` $\equiv 1$ for every row — the *anchor* draw (no bootstrap; full sample with original weights).
- For $b=1,\ldots,300$: `set seed $SEED` (one-time); then `bsample, weight(bw_b)` (or `bsample, cluster(`*`cluster_var`*`) weight(bw_b)` when a cluster variable is supplied).

The cluster argument is dataset-specific: PSID clusters on `current_hhid` (current household identifier); CPS on `cpsidp` (IPUMS-CPS person identifier); SCF on `id`; SOEP on `hid`; UAS / Pairfam on the respondent id. When `stamp_bw` is called *without* a cluster argument (e.g., ATUS, where each observation is its own primary unit), each row is its own bootstrap unit.

Per-bootstrap moment computation. Downstream moment computations use the weight $w_b \equiv w_{\text{base}} \cdot$ `bw_b` in their `collapse [pw=...]` or `mean [pw=...]` calls. The helper `wmean_by` (`scripts/_utils.do`) wraps this loop and writes a panel of NBS $+1=301$ columns per group, one per bootstrap draw. The anchor ($b=0$) feeds the point-estimate moments; the $b\geq 1$ draws give the empirical bootstrap variances used by the standard-error machinery (Section 3.7, Section 4.7).

2.4 Price-base / deflation conventions

The pipeline uses country- and source-specific price bases. The conventions are:

Country / Model / Variable	Price base	Notes
US M1 / PSID income	2011 USD	BLS CPI deflator applied in `final_dataset.do`
US M1 / SCF wealth (2007 wave)	2007 USD	Not deflated (nominal)
US M2 / PSID-based moments	2011 USD	Same convention as US M1 PSID
DE M1 / SOEP income	2010 EUR	SOEP deflator series
DE M1 / SOEP wealth	2012 EUR	CPI Germany series; base = 2012
DE M2 / SOEP-based moments	2010 EUR	Same as DE M1 SOEP income

Wealth enters the structural model only as the time-zero asset $a(0)$, treated as an exogenously-determined initial condition.

3 Model 1 implementation

3.1 Overview

This section documents the computational implementation of Model 1. The structural specification — sample selection, identification strategy, and theoretical interpretation — is presented in Section 3 of the paper and Appendices C and D; here we focus on the implementation details: Section 3.2 explains the 12-parameter specification of Table 4 in the paper (with two additional calibrated constants bringing the implementation vector to 14 entries); Section 3.3 analyzes the two-phase dynamic programming solver; Section 3.4 lays out the two-step MSM objective with the priority-weight (group-size) weighting matrix; Section 3.5 gives details on the L-BFGS optimization pipeline; Section 3.6 elucidates identification and result verification; and Section 3.7 shows asymptotic inference.

The specification is symmetric across countries: the same 12-parameter specification (Table 4 in the paper; 14-element implementation vector with two calibrated constants) and the same MSM objective are estimated for the United States and Germany. Two configuration items differ across countries: (i) the pinned interest rate r; and (ii) the priority-weight multipliers ($k_{\ell_{\text{age}}} = 20, k_{m_b} = 10$ for the US versus $10, 100$ for Germany), chosen to bring the model's m_b above the acceptance threshold of 0.05 in both countries (calibration target $1/12 \approx 0.0833$). These multipliers affect optimization priorities but do not alter the underlying model structure.

3.2 Parameterization

The implementation uses a 14-element vector $\theta = (\theta_1, \ldots, \theta_{14})$ whose ordering and meaning follow Section 3 of the paper (Table 1). Of these 14 entries, 12 correspond to the model parameters reported in Table 4 in the paper: ten are free to be estimated within country-specific box bounds, and two (ρ, ℓ_{low}) are pinned from the warm-start. The remaining two entries are calibrated constants set outside estimation:

- The interest rate r is fixed at a country-specific value reflecting the long-run real rate: $r = 0.050$ for the United States and $r = 0.032$ for Germany (country-specific historical-average asset return).
- The pre-experience term α_1 is fixed at 1.00 for both countries (treated as a structural constant).

The pinned entries are $(\rho, \alpha_1, \ell_{\text{low}}) = (0.01, 1.00, 0.075)$ for the United States and $(0.02, 1.00, 0.097)$ for Germany.

Index	Symbol	US bounds	DE bounds	Status
Discount factor				
1	ρ	[0.010, 0.015]	[0.020, 0.025]	Pin (warm-start)
2	r	0.050	0.032	Pin (country)
Consumption preferences				
3	θ	[0.45, 0.50]	[0.40, 0.45]	Free
4	α	[0.85, 0.90]	[0.86, 0.95]	Free
6	γ	[4.0, 6.0]	[1.5, 3.0]	Free
Child preferences				
5	α_1	1.00	1.00	Constant
7	ψ	[0.475, 0.500]	[0.305, 0.500]	Free
13	η	[1.01, 1.05]	[1.01, 1.10]	Free
14	θ_1	[0.81, 0.83]	[0.84, 0.87]	Free
Cost of children				
8	κ	[3.5, 4.0]	[0.50, 1.30]	Free
9	φ	[0.30, 0.50]	[0.45, 0.55]	Free
10	ζ	[0.40, 0.95]	[0.65, 1.00]	Free
Minimum levels				
11	ℓ_{low}	[0.000, 0.075]	[0.000, 0.100]	Pin (warm-start)
12	m_{low}	[−0.10, 0.05]	[−0.20, 0.00]	Free

Table 1: Model 1 implementation vector θ (14 entries = 12 model parameters of Table 4 in the paper + 2 calibrated constants; 10 free, 2 pinned, 2 constants). For the meaning of each parameter, see Section 3 and Appendices C and D of the paper.

Weighting. The estimation objective applies country-specific upweights to a subset of the moments (Section 3.4). The two multipliers that enter the estimation are:

	United States	Germany
$k_{\ell_{\text{age}}}$ (labor-supply-by-age, upweight)	20	10
k_{m_b} (outsourced-childcare share, upweight)	10	100

These multipliers were selected by a k_{m_b} sensitivity scan (see `estimation/01_scan_k_mb.jl`) as the smallest upweights that bring the model's outsourced-childcare share m_b above the acceptance threshold of 0.05 in both countries (calibration target $1/12 \approx 0.0833$). At the estimate, $\hat{m}_b = 0.0844$ (United States) and 0.0717 (Germany).

3.3 Model structure: state space, two-phase optimization, and numerical solver

Model 1 is an aggregative, deterministic, continuous-time lifecycle model in which the household head jointly decides on consumption, labor supply, fertility, and childcare time. The structural specification is given in Section 3 of the paper; here we re-state the elements needed to describe the computational implementation and then specify the discretization and numerical solver.

State and choice variables. The household head ages from $age_0 = 22$ to $age_T = 45$. Using the model time index $t = age - age_0 + 1$, the horizon is $t \in [t_0, T]$ with $t_0 = 1$ and $T = 24.001$ (a small offset past $t_T = 24$ to avoid integration boundary issues; constants in `src/data.jl`). The interval is split at $\bar{t} = 7$ (corresponding to $agebar = 28$) into an *infertile* Phase 1 $[t_0, \bar{t}]$ and a *fertile* Phase 2 $[\bar{t}, T]$. The state variable evolved by the model is household assets $a(t)$; exogenous time-varying inputs are partner income $y_p(t)$ and hourly wage $w(t)$. The control variables are adult consumption $C_A(t)$, child consumption $C_C(t)$ (Phase 2 only), labor supply $\ell(t) \in [0, 1 - \ell_{\text{low}}]$ (continuous, with country-specific lower bound on leisure ℓ_{low}), parental childcare time $m(t)$ (Phase 2 only), outsourced childcare time per child $m_b(t)$ (Phase 2 only), and the number of children $n(t)$ (Phase 2 only; under the paper's aggregation assumption, $n(t)$ is treated as a continuous variable). Household-headcount parameters are $N_1 = 2.02$ in Phase 1 and $N_2 = 2.07$ in Phase 2 (Section 3 in the paper, Model 1).

Phase 1 optimization (infertile, $t \in [t_0, \bar{t}]$).

$$\max_{(C_A(t),\, \ell(t))_{t\in[t_0,\bar{t}]}} \int_{t_0}^{\bar{t}} e^{-\rho t}\, u^{\text{no children}}\big(C_A(t), \ell(t)\big)\, dt \quad \text{s.t.} \quad \dot{a}(t) = r\, a(t) + y_p(t) + w(t)\, \ell(t) - C_A(t), \tag{1}$$

where $u^{\text{no children}}(\cdot)$ is defined by equation (2) in the paper, with structural parameters $\theta, \gamma, \ell_{\text{low}}$.

Phase 2 optimization (fertile, $t \in [\bar{t}, T]$).

$$\begin{aligned} \max_{(C_A, C_C, \ell, m, m_b, n)_t} \int_{\bar{t}}^{T} & e^{-\rho(t-\bar{t})}\, u^{\text{with children}}\big(C_A(t), C_C(t), \ell(t), m(t), m_b(t), n(t)\big)\, dt \\ \text{s.t.} \quad & \dot{a}(t) = r\, a(t) + y_p(t) + w(t)\, \ell(t) + \big[A - \kappa\, w(t)^{\varphi} - \tilde{p}_b\, m_b(t)\big]\, n(t) \\ & \quad - C_C(t) - C_A(t), \end{aligned} \tag{2}$$

with $\tilde{p}_b = (1-s)\,p_b$ the childcare-subsidy-adjusted price (equation (15) in the paper), A the per-child allowance, and $u^{\text{with children}}(\cdot)$ the composite-good utility constructed from equations (4)–(9) in the paper, with structural parameters $\alpha, \theta, \theta_1, \psi, \zeta, \eta, \kappa, \varphi$.

Combined lifetime objective. The full optimization procedure joins Phases 1 and 2 with continuity of assets at $a(\bar{t})$:

$$V(t_0) \;=\; \max \int_{t_0}^{\bar{t}} e^{-\rho t}\, u^{\text{nc}}\, dt \;+\; e^{-\rho \bar{t}} \int_{\bar{t}}^{T} e^{-\rho(t-\bar{t})}\, u^{\text{wc}}\, dt \quad \text{s.t. (1) on } [t_0, \bar{t}] \text{ and (2) on } [\bar{t}, T]. \tag{3}$$

Numerical solver (`src/model.jl`). The continuous-time problem is solved in two stages: a closed-form analytical step seeds an initial profile, then iterative refinement converges to the full solution:

1. *Time discretization.* The continuous-time horizon is discretized with step size $\Delta t = 1/10$ (ten sub-periods per year), set by the constant `_DT` in `src/model.jl`.
2. *Phase 1 analytical step.* Under the special case $\eta = 1$ (where η is the parameter in equation (9) in the paper), the Phase 1 problem admits a closed-form solution. This is used to seed an initial profile for the outsourced-childcare time $m_b(t)$ on Phase 2.
3. *Phase 2 fixed-point iteration.* For the general case $\eta \neq 1$, the Phase 2 problem is solved by an iterative scheme: given the current $m_b(t)$ profile, the implementation fits the corresponding aggregate M_b polynomial and recovers the implied m_b profile; the new and old profiles are blended with damped weights $0.95\, m_b^{\text{old}} + 0.05\, m_b^{\text{new}}$ (constants `_DAMP_OLD`, `_DAMP_NEW`). Iteration terminates when the relative change is below `_EPS` $= 10^{-3}$ or after `_MAX_ITER` $= 80$ iterations.
4. *Integration.* Time integrals in the objective and budget constraints are computed by `QuadGK` with relative tolerance `_QUADGK_RTOL` $= 10^{-10}$.
5. *Nonlinear root-finding.* First-order conditions for the Phase 2 controls (conditional on m_b) are solved by `NLsolve`.

A single call to `matching_fun_v7` (the moment-aggregation wrapper in `src/moments_v7.jl` around the core solver `matching_fun` in `src/model.jl`) takes ≈ 0.25 s for the United States and ≈ 0.11 s for Germany on the reference machine. The L-BFGS estimation (Section 3.5) is therefore highly efficient by structural-DP standards: the full estimation takes ≈ 90 s wall for the US and ≈ 25 s for Germany.

3.4 Two-step MSM objective

Model 1 is estimated by the method of simulated moments (MSM): a two-step protocol minimizes the weighted distance between the simulated and data moments, with diagonal weights that upweight a priority subset of moments (defined below). The objective evaluator is `matching_fun_v7` in `src/moments_v7.jl`; the distance formula is `dis_v7` in `estimation/02_run_estimate.jl`.

Moment vector and groups. Let $\boldsymbol{m}(\theta) \in \mathbb{R}^{124}$ and $\boldsymbol{M}^d \in \mathbb{R}^{124}$ denote the simulated and data moment vectors. The 124 moments are organized into 7 age-by-quantity groups with the following sizes (Section 3.4 and Table 4 in the paper):

Group	n_g	Contents
ℓ_{age}	24	Mean labor supply by age (22–45)
n_{age}	17	Mean number of children by age (29–45; post-$\bar{t}$)
m_{age}	17	Mean parental childcare time by age (29–45)
$\log c$	24	Mean log consumption by age (22–45)
$\log cc$	17	Mean log child consumption by age (29–45)
$\log a$	24	Mean log assets by age (22–45)
m_b	1	Outsourced-childcare share, scalar
Total	**124**	

Weight diagonal w. The weighting is constructed by `build_weights_l_and_mb_upweight` (in `estimation/build_weights_mb.jl`). Given the country-specific upweight multipliers $k_{\ell_{\text{age}}}$ and k_{m_b} (Section 3.2) and a subset $\mathcal{G}$ of the seven groups, the per-element weight on element i in group g is

$$w_i = \frac{\text{upweight}(g)}{|\mathcal{G}| - 2 + k_{\ell_{\text{age}}} + k_{m_b}} \cdot \frac{1}{n_g}, \qquad \text{upweight}(g) = \begin{cases} k_{\ell_{\text{age}}}, & g = \ell_{\text{age}}, \\ k_{m_b}, & g = m_b, \\ 1, & \text{otherwise.} \end{cases} \tag{4}$$

By construction the weights sum to 1 across the moments in $\mathcal{G}$ (mass redistribution rather than mass addition); moments outside $\mathcal{G}$ get weight zero. The -2 correction in the denominator removes the duplicate ℓ_{age} and m_b contributions from the unit-weight tally so that the upweighted groups replace, rather than augment, their unit-weight contributions.

Distance function. For a given parameter vector θ, weighting vector $\mathbf{w}$, and step multiplier $\mu \in \{1, w_{\text{priority}}\}$ (where $w_{\text{priority}} = 10$):

$$L_{M1}^{(\mu,\mathbf{w})}(\theta) \;=\; \mu \cdot \sum_{i=1}^{124} w_i \cdot \big(m_i(\theta) - M_i^d\big)^2. \tag{5}$$

Two-step protocol. The estimation uses the following two-step sequence (executed in `estimation/02_run_estimate.jl::run_country`):

- *Step 1: priority subset.* Build $\mathbf{w}^{(1)}$ via (4) restricted to $\mathcal{G}^{(1)} = \{\ell_{\text{age}}, n_{\text{age}}, m_{\text{age}}, m_b\}$ (four priority groups). Minimise $L_{M1}^{(10,\mathbf{w}^{(1)})}$ via L-BFGS (Section 3.5).
- *Step 2: full set.* Build $\mathbf{w}^{(2)}$ via (4) restricted to $\mathcal{G}^{(2)}$ = all 7 groups (124 elements active). Minimise $L_{M1}^{(1,\mathbf{w}^{(2)})}$, warm-started at the Step-1 minimum.

The Step-1 amplification by $\mu = 10$ concentrates the optimization on the priority moments at the start; Step 2 then tunes against the full moment set with $\mu = 1$ to ensure the full-set fit.[2]

3.5 Optimization pipeline

The two-step MSM objective (5) is minimized by the L-BFGS algorithm with finite-difference gradient, as implemented in `NLopt.jl`. The entire pipeline is executed sequentially per country by `run_country` in `estimation/02_run_estimate.jl`.

Optimizer configuration.

Setting	Value
Algorithm	`NLopt.Opt(:LD_LBFGS, 14)` (L-BFGS with derivatives, 14-dim)
Bounds	`opt.lower_bounds = cfg.LB`, `opt.upper_bounds = cfg.UB`
Gradient	Central-difference finite-difference, step $h = 10^{-5}$ relative
Convergence	`ftol_abs` $= 10^{-3}$
Iteration cap	`maxeval` $= 1500$
Starting point	`cfg.PAR_initial`, strict-interior-perturbed by $\varepsilon = 10^{-8}$

The pinned parameters ($\rho, r, \alpha_1, \ell_{\text{low}}$; Section 3.2) are enforced by setting their lower and upper bounds equal in `cfg.LB` and `cfg.UB`; L-BFGS treats them as fixed throughout.

[2] The estimation objective uses the priority-weight scheme (4); the 300 bootstrap draws of $\boldsymbol{M}^d$ enter only the asymptotic-SE sandwich (Section 3.7), as the moment covariance Σ_{moments}.

Two-step run.

1. *Step 1: priority-subset solve.* Calls `solve_lbfgs` with weights $\mathbf{w}^{(1)}$ on the 4-group priority subset (Section 3.4) and multiplier $\mu = w_{\text{priority}} = 10$. Starting from the strict-interior-perturbed `PAR_initial`, L-BFGS returns the Step-1 minimum $\hat{\theta}^{(1)}$.
2. *Step 2: full-set solve.* Calls `solve_lbfgs` with weights $\mathbf{w}^{(2)}$ on the full 7-group set and multiplier $\mu = 1$. Warm-started at the strict-interior perturbation of $\hat{\theta}^{(1)}$, L-BFGS returns the final estimate $\hat{\theta} \equiv \hat{\theta}^{(2)}$.

After Step 2, the code re-evaluates `matching_fun_v7` at $\hat{\theta}$ to extract per-group sums of squared deviations and the data/model moment vectors, then writes the estimate file `locked/par_$(country).jld2` (see Section 3.1 and the README for the file schema).

Wall time. A single `matching_fun_v7` evaluation takes ≈ 0.25 s for the United States and ≈ 0.11 s for Germany on the reference machine (Section 3.3). The full two-step L-BFGS estimation takes ≈ 90 s for the US and ≈ 25 s for Germany, corresponding to ≈ 350 objective evaluations for the US and ≈ 230 for Germany across Step 1 + Step 2 combined.

3.6 Identification and result verification

Identification. The 14-element implementation vector θ has 2 calibrated constants (r, α_1), 2 parameters pinned at warm-start values (ρ, ℓ_{low}), and 10 free entries (Section 3.2), and the MSM objective (5) matches 124 moments. The free parameter vector is therefore over-identified by 114 moment conditions. As reported in the notes of Table 4 in the paper, 9/10 free parameters satisfy $|t| > 1.96$ for the United States and 5/10 for Germany at the estimate, with $\text{cond}(J'WJ) \approx 1.1 \times 10^{11}$ (US) and $\approx 1.2 \times 10^{11}$ (DE) (computed in `estimation/03_jacobian_se.jl`).

A specific identification criterion concerns the model-implied outsourced-childcare share m_b. The acceptance threshold is $m_b \geq 0.05$ (a relaxed lower tolerance), with the calibration target $1/12 \approx 0.0833$. Both countries satisfy this: $\hat{m}_b = 0.0844$ (US, just above the target) and 0.0717 (DE).

Result verification. The implementation delivers a verification script (`diagnostics/verify_v14.jl`) that performs two classes of checks against the saved estimate files in `locked/par_{us,de}.jld2`:

Structural (Part A): activates the project, includes all 7 `src/*.jl` files, loads both country configs, confirms that the external dependencies (the parameters pinned at warm-start values in `external/v11/locked/`) and the inverse-variance weighting matrices (in locked/StrEST_*.jld2) are accessible, and verifies that the saved par files have the expected schema (par length 14, BETA_D/BETA_M length 124, finite non-negative total_ssr, etc.).

Bit-exact (Part B): re-evaluates matching_fun_v7 at the saved par (no re-optimization) and compares the recomputed **BETA_M** against the saved value. On the reference replication package (2026-06-03), both countries reproduce to within floating-point precision: United States $\max_i |\text{BETA\_M}_i^{\text{recom}} - \text{BETA\_M}_i^{\text{saved}}| = 0$; Germany $= 8.88 \times 10^{-16}$ (one floating-point ULP). In the headline of Table 4 in the paper, the $\hat{m}_b$ values (0.084 US, 0.072 DE) match exactly. The bit-exactness isolates any future implementation discrepancy to the optimization step rather than the objective evaluator.

3.7 Asymptotic standard errors

Standard errors for the estimate $\hat{\theta}$ are computed once at convergence using a sandwich-formula asymptotic variance. The implementation is located in estimation/03_jacobian_se.jl.

Sandwich asymptotic variance. The asymptotic variance follows the standard MSM sandwich [Hansen, 1982, Newey and McFadden, 1994], with W from estimation and $\hat{\Sigma}_{\text{moments}}$ the bootstrap covariance of the data moment vector (300 cluster-bootstrap draws):

$$\widehat{V}(\hat{\theta}_{\text{free}}) = (\hat{J}^\top W \hat{J})^{-1} \cdot \hat{J}^\top W \, \hat{\Sigma}_{\text{moments}} \, W \hat{J} \cdot (\hat{J}^\top W \hat{J})^{-1}, \tag{6}$$

where $W = \text{diag}(\mathbf{w}^{(2)})$ is the Step-2 weight diagonal (Section 3.4); $\hat{J}$ is the 124×10 Jacobian of moment conditions with respect to the 10 free parameters at $\hat{\theta}$ (pinned columns excluded); and $\hat{\Sigma}_{\text{moments}}$ is the bootstrap covariance of the data moment vector (defined below). Standard errors are the square roots of the diagonal of $\widehat{V}$. The implementation uses the pseudo-inverse $\text{pinv}(J'WJ, \text{rtol} = \sqrt{\varepsilon_{\text{mach}}})$ in place of the direct inverse for numerical robustness.

Bootstrap moment covariance.

$$\hat{\Sigma}_{\text{moments}} = \widehat{\text{Cov}}(\boldsymbol{M}^d) = \frac{1}{B} \sum_{b=0}^{B} \left(\boldsymbol{M}^{d,(b)} - \overline{\boldsymbol{M}}^d\right)\left(\boldsymbol{M}^{d,(b)} - \overline{\boldsymbol{M}}^d\right)^\top, \tag{7}$$

where $\boldsymbol{M}^{d,(b)}$ is the 124-vector of data moments recomputed under bootstrap draw b, $\overline{\boldsymbol{M}}^d$ is the cross-draw mean, and $B = 300$ is the bootstrap count. The $\boldsymbol{M}^{d,(b)}$ are loaded via `load_data(cfg,data_dir,b;target=:anchor)` from the moment-data CSV files in `data/{us,de}/`, indexed by `bs_index = b`. The bootstrap covariance is computed inline by `bootstrap_sigma_v7` at the start of each SE run, not loaded from a cached file.

Jacobian construction. $\hat{J}$ is computed by central differencing with relative step $5\% \cdot |\hat{\theta}_k|$ (with a small absolute floor 10^{-6} for near-zero parameters), over the 10 free-parameter columns. Each column requires two full `matching_fun_v7` evaluations (Section 3.3); the Jacobian itself therefore costs only $\approx 2 \times 10 \times 0.25\,\mathrm{s} \approx 5\,\mathrm{s}$ for the US and $\approx 2\,\mathrm{s}$ for Germany. The bootstrap covariance computation dominates the SE wall time at $\approx 300\times$ (load-data time).

Reported standard errors. The standard errors reported in Table 4 in the paper (Model 1 parameter estimates with Jacobian SEs) are produced by `estimation/03_jacobian_se.jl` from (6) and (7). The saved output `locked/jacobian_se.jld2` contains, for each country, the 10-vector of SEs, the $\mathrm{cond}(J'WJ)$ scalar, the count of parameters with $|t| > 1.96$ (US: 9, DE: 5), and the `free_idx` mapping back to the 14-element parameter vector.

4 Model 2 implementation

4.1 Overview

This section documents the computational implementation of Model 2. The structural specification — sample selection, identification strategy, and theoretical interpretation — is presented in Section 4 of the paper and Appendices C and D; here we focus on the implementation details: Section 4.2 explains the 37-parameter specification with 4-type heterogeneity; Section 4.3 analyzes the dynamic programming solver and mixture aggregation; Section 4.4 lays out the augmented method of simulated moments objective; Section 4.5 gives details on the three-stage optimization pipeline; Section 4.6 elucidates identification and warm-start verification; and Section 4.7 shows asymptotic inference. The two-DP construction used to implement the vignette current-period-wage convention is documented in Section 5 (vignette simulated panel).

The specification is symmetric across countries: the same 37-element parameter vector θ is estimated under the same objective function for the United States and Germany, with prior-informed bounds on the type-share parameters π_{O_1}, π_{O_2} being the only country-specific element.

4.2 Parameterization

The implementation uses a single 37-dimensional parameter vector $\theta = (\theta_1, \ldots, \theta_{37})$ applied symmetrically to both countries (Table 2). All entries are MSM-estimated from the 292-moment vector. Their roles span the lifecycle preference and technology parameters of Section 4 in the paper, the smoothing of the discrete fertility choice, the scale of the childcare-time technology, and the shares of the two latent preference types.

Index	Parameter	Range	Role
1–33	$\theta_{1\ldots33}$	country-specific	Structural parameters (preferences, technology, demographics; see Section 4 in the paper).
34	σ_ε^f	$[0, 5]$	Scale of the fertility preference shock. At $\sigma_\varepsilon^f = 0$ the choice reduces to hard-max (deterministic).
35	$\overline{lc}$-scale	$[0.3, 1.3]$	Direct multiplier on the empirical childcare-time matrix $\bar{\ell}^{cc}$ (Table 6 in the paper, row "$\overline{lc}$-scale"; coded as `lcbar_scale`).
36	π_{O_1}	country-specific narrow	Population share of the low-fertility-preference type.
37	π_{O_2}	country-specific narrow	Population share of the low-work-disutility type.

Table 2: The 37-parameter vector θ estimated for Model 2.

Country-specific π bounds. The two type-share parameters use narrow, prior-informed bounds rather than the unit interval:

Country	π_{O_1} bounds	π_{O_2} bounds
Germany	$[0.60, 0.90]$ (Wang 0.75 $\pm$ 20%)	$[0.272, 0.408]$ (Wang 0.34 $\pm$ 20%)
United States	$[0.40, 0.70]$ (NLSY79 $\sim$0.55 $\pm$ 0.15)	$[0.25, 0.55]$ (NLSY79 $\sim$0.40 $\pm$ 0.15)

The German bounds are anchored on Wang [2026], with a $\pm$20% relative margin. The United States bounds are informed by attitudinal-question shares in the NLSY79 (see Section 4.3, footnote 38 in the paper), with a $\pm$0.15 absolute margin reflecting classification uncertainty. The narrow-bound approach combines literature-anchored bounds (which locate the feasible region) with the moment match (which refines within the region).

Mixture aggregation across the four types. With $(O_1, O_2) \in \{0,1\}^2$ there are four latent types $k \in \{1,2,3,4\}$. For each type, the dynamic program is solved under the type-specific parameter draws (Section 4.3). The aggregate moment vector at θ is the type-share-weighted mixture

$$m(\theta) \;=\; \sum_{k=1}^{4} w_k(\pi_{O_1}, \pi_{O_2})\, m^{(k)}(\theta),$$

where

- $w_1 = \pi_{O_1}\,\pi_{O_2}$ (type 1: $O_1 = O_2 = 1$);
- $w_2 = \pi_{O_1}\,(1-\pi_{O_2})$ (type 2: $O_1 = 1,\ O_2 = 0$);
- $w_3 = (1-\pi_{O_1})\,\pi_{O_2}$ (type 3: $O_1 = 0,\ O_2 = 1$);
- $w_4 = (1-\pi_{O_1})(1-\pi_{O_2})$ (type 4: $O_1 = O_2 = 0$).

The four DPs are simulated under *common random numbers*: the same random seed (`seed=123`) feeds each type's simulation, so the mixture estimator's variance is reduced and the optimization landscape is smooth in π_{O_1}, π_{O_2}.

Nesting. The specification nests two functional-form simplifications at its parameter boundaries: setting $\sigma_\varepsilon^f = 0$ recovers hardmax (deterministic) fertility choice, and concentrating π_{O_1}, π_{O_2} at unity recovers a single-type (homogeneous) economy.

4.3 Model structure: state space, choice set, and Bellman equation

This subsection re-states the dynamic-programming structure of Model 2 (originally derived in Section 4 of the paper) and specifies its smoothed implementation. The structural notation mirrors that of the paper exactly.

State space. The household's state at the beginning of period t is

$$\Omega_t = \big(\, x_t,\ n_t,\ a_t,\ \mathrm{mar}_t,\ y_{p,t},\ job_t,\ \xi_t \,\big), \tag{8}$$

where x_t is accumulated work experience, $n_t \in \{0,1,2,3\}$ is the number of children, a_t is the age of the youngest child, $\mathrm{mar}_t \in \{0,1\}$ marital status, $y_{p,t}$ partner income, $job_t \in \{0,1\}$ the job-arrival indicator, and ξ_t a productivity shock. The period index $t = age - 22 + 1$ ranges over $t = 1, \ldots, T$, with $T = 41$ ($age = 62$) the terminal model period; $V_{T+1} \equiv 0$.

The model distinguishes a fertile and a nonfertile period. For $age_t \in \{22, \ldots, 45\}$ (the fertile period), each period t is split into two stages: a stage-1 fertility decision, followed by a stage-2 labor-supply decision. For $age_t \in \{46, \ldots, 62\}$ (the nonfertile period), no further births occur and only the labor-supply decision is active. Completed fertility is therefore the count reached at age 45 ($t = 24$); because children have aged out of the household, the post-45 state carries none, as the reduced state Ω_t^0 below makes explicit:

$$\Omega_t^0 = \big(\, x_t,\ n_t = 0,\ a_t = \emptyset,\ \mathrm{mar}_t,\ y_{p,t},\ job_t,\ \xi_t \,\big) \quad \text{for } age_t \in \{46, \ldots, 62\}. \tag{9}$$

In the fertile period, the stage-2 post-decision state is

$$\Omega_t^* = \big(\, x_t,\ n_t^*,\ a_t^*,\ \mathrm{mar}_t,\ y_{p,t},\ job_t,\ \xi_t \,\big), \qquad n_t = n_t^* + f_t, \quad a_t = (1 - f_t)a_t^*, \tag{10}$$

where $f_t \in \{0, 1\}$ is the binary fertility choice ($f_t = 1$ means a newborn arrives, setting $a_t = 0$).

Experience transition. The continuous state x_t (work experience) evolves according to equation (17) in the paper:

$$x_{t+1} \;=\; \big(\, x_t \;+\; \lambda_p\, \mathbb{I}_t^{\ell p} \;+\; \mathbb{I}_t^{\ell f} \,\big) \;\cdot\; \big[\, \mathbb{I}_t^{\ell p} \;+\; \mathbb{I}_t^{\ell f} \;+\; \delta\, (1 - \mathbb{I}_t^{\ell p} - \mathbb{I}_t^{\ell f}\,) \,\big], \tag{11}$$

where $\mathbb{I}_t^{\ell p}$ and $\mathbb{I}_t^{\ell f}$ are indicators for $\ell_t = \ell^p$ and $\ell_t = \ell^f$ respectively. Equivalently in piecewise form, full-time work earns a unit of experience ($x_{t+1} = x_t + 1$); part-time work earns the fraction $\lambda_p \in (0, 1)$ ($x_{t+1} = x_t + \lambda_p$; λ_p is estimated, with a prior bound [1/3, 3/4]); and not working — whether by choice ($\ell_t = 0$) or involuntarily ($job_t = 0$) — depreciates experience multiplicatively to $x_{t+1} = \delta\, x_t$ with $\delta \in (0, 1)$. The other state variables evolve exogenously. Marriage and divorce follow age-indexed logit specifications for the transition probabilities into and out of marriage, and the job-arrival probability p_t^j follows a logit specification in age and experience. Partner income $y_{p,t}$ is a calibrated age profile that enters household resources while the woman is married. These processes and their coefficients are specified in Section 4 and Appendix D of the paper.

Deterministic Bellman equations. The deterministic version of the model (the $\sigma \to 0$ limit of the estimated specification) is governed by three optimization equations.

Nonfertile period ($age_t \in \{46, \ldots, 62\}$). The household solves a single Bellman equation in labor supply $\ell_t \in \{0, \ell^p, \ell^f\}$ (non-, part-, full-time):

$$\begin{aligned} V_t\big(\Omega_t^0\big) \;&= p_t^j \cdot \max_{\ell_t} \Big\{ u_t(\ell_t, \tilde{c}_t, n_t = 0, \mathrm{mar}_t, a_t = \emptyset) + \beta\, E\big[V_{t+1}\big(\Omega_{t+1}^0 \mid \ell_t, \Omega_t^0\big)\big] \Big\} \\ &\quad + \; (1 - p_t^j) \Big\{ u_t(\ell_t = 0, \tilde{c}_t, \ldots) + \beta\, E\big[V_{t+1}\big(\Omega_{t+1}^0 \mid \ell_t = 0, \Omega_t^0\big)\big] \Big\}, \end{aligned} \tag{12}$$

where $p_t^j \in [0,1]$ is the job-arrival probability and $\tilde{c}_t$ is the period budget (Appendix C of the paper). The $(1-p_t^j)$ branch is involuntary unemployment.

Fertile period, stage 2 (labor supply, $age_t \in \{22,\ldots,45\}$). Conditional on the realized fertility choice f_t^* from stage 1, the labor-supply Bellman is

$$\begin{aligned} V_t(\Omega_t) \;&= p_t^j \cdot \max_{\ell_t}\Big\{u_t(\ell_t, \tilde{c}_t, n_t, \mathrm{mar}_t, a_t) + \beta\, E\big[V_{t+1}^*\big(\Omega_{t+1}^* \mid \ell_t, \Omega_t\big)\big]\Big\} \\ &\quad + \,(1-p_t^j)\Big\{u_t(\ell_t = 0, \ldots) + \beta\, E\big[V_{t+1}^*\big(\Omega_{t+1}^* \mid \ell_t = 0, \Omega_t\big)\big]\Big\}. \end{aligned} \tag{13}$$

Fertile period, stage 1 (fertility). The fertility decision is not a Bellman equation but an exogenous-fecundity-modulated maximization of the stage-2 value:

$$V_t^*(\Omega_t^*) \;=\; p_t^f \cdot \max_{f_t \in \{0,1\}}\big[\, V_t(\Omega_t \mid f_t, \Omega_t^*)\big] \;+\; (1-p_t^f)\cdot V_t(\Omega_t \mid f_t = 0, \Omega_t^*)\,, \tag{14}$$

where p_t^f is the (age-decreasing) biological fecundity probability; with probability $1-p_t^f$ the agent cannot conceive, forcing $f_t = 0$.

Smoothed implementation. The estimated specification replaces the deterministic max operators above with their Type-I extreme-value (Gumbel) softmax counterparts, with scales σ_ε for the labor-supply choice and σ_ε^f for the fertility choice. Writing $Q_t^\ell(\ell;\Omega_t)$ for the bracketed expression inside the $\max_{\ell_t}$ in either (12) or (13), the labor-supply softmax is

$$\max_{\ell_t \in \{0,\ell^p,\ell^f\}} Q_t^\ell(\ell_t;\Omega_t) \;\;\rightsquigarrow\;\; \sigma_\varepsilon\, \log\!\left(\sum_{\ell \in \{0,\ell^p,\ell^f\}} \exp\!\big(Q_t^\ell(\ell;\Omega_t)/\sigma_\varepsilon\big)\right), \tag{15}$$

and, writing $Q_t^f(f;\Omega_t^*) = V_t(\Omega_t \mid f, \Omega_t^*)$, the fertility softmax (modulated by the fecundity term) is

$$V_t^*(\Omega_t^*) \;=\; p_t^f \cdot \sigma_\varepsilon^f\, \log\!\big(\exp(Q_t^f(0)/\sigma_\varepsilon^f) + \exp(Q_t^f(1)/\sigma_\varepsilon^f)\big) \;+\; (1-p_t^f)\cdot Q_t^f(0). \tag{16}$$

As $\sigma_\varepsilon, \sigma_\varepsilon^f \to 0$, (15) and (16) collapse to the deterministic forms above. The implementation estimates $\sigma_\varepsilon^f \in [0,5]$ as the parameter θ_{34} (Table 2); σ_ε is fixed at its calibration value.

Numerical solver. The model is solved by backward induction. The value functions V_t, V_t^* are tabulated on a discrete grid in the continuous state x_t (linearly interpolated for off-grid points) and exhaustively enumerated in the categorical states $(n_t, \mathrm{mar}_t, y_{p,t}, job_t)$. The age of the youngest child $a_t \in \{0,1,\ldots,13\}$ is gridded as an integer index (Appendix C of the paper). During backward induction, next-period values $V_{t+1}(x_{t+1},\cdot)$ are evaluated at the three candidate x_{t+1} values determined by (11) and at the shrunken $\delta\, x_t$ value used in the $job_{t+1} = 0$ branch.

Under the 4-type heterogeneity structure (Section 4.2), the full backward-induction routine runs once per latent type $k \in \{1, 2, 3, 4\}$, giving four type-conditional value-function tables and four sets of type-conditional simulated moments $m^{(k)}(\theta)$. The four computational tasks are parallelized via `Threads.@threads` in the implementation; common random numbers (`seed=123`) are used across types so that the mixture estimator's variance is reduced and the aggregated moment vector $m(\theta)$ is a smooth function of π_{O_1}, π_{O_2}.

4.4 Augmented MSM objective

The implementation estimates θ by the method of simulated moments (MSM), with the moment-distance function augmented by a band penalty:

$$L_{\mathrm{M2}}(\theta) \;=\; L_{\mathrm{MSM}}(\theta) \;+\; p_{\mathrm{band}}\big(\varepsilon_N(\theta)\big). \tag{17}$$

MSM core. With $m(\theta) = \sum_k w_k m^{(k)}(\theta)$ the type-mixture simulated moment vector (Section 4.2) and M^d the corresponding $N_m = 292$ data moments organized into 13 age-by-group blocks (Table 7 in the paper), the MSM core is the diagonal-weighted sum of squared deviations:

$$L_{\mathrm{MSM}}(\theta) \;=\; \sum_{i=1}^{292} w_i \cdot \big(m_i(\theta) - M_i^d\big)^2, \tag{18}$$

where the diagonal weights w_i implement priority weighting:

$$w_i \;=\; \begin{cases} 10 & \text{if } i \in \mathcal{P}, \\ 1 & \text{otherwise,} \end{cases} \qquad \big|\mathcal{P}\big| = 120, \tag{19}$$

and $\mathcal{P}$ is the set of moment indices belonging to the five priority groups, each running over the 24 ages $\{22, \ldots, 45\}$: mean labor supply, mean number of children, mean parental childcare time, birth probability, and mean log wage. The remaining 172 moments (full-time rate, part-time rate, variance of labor supply, fraction childless, mean labor supply by number of children, covariance of labor supply and children, share single without children, share single with children) carry unit weight.

Band penalty on the price elasticity of fertility. Let $\bar{N}(\theta; s)$ denote the simulated mean number of children at childcare-price subsidy rate s, averaged over ages 22–45. Recall from the paper that a subsidy s acts on the effective childcare price through $\tilde{p}_b = (1 - s)\, p_b$, so a change from $s = 0$ to $s = 0.20$ is a 20% reduction in the price faced by households. The childcare-price elasticity of fertility evaluated at $s = 0.20$ is

$$\varepsilon_N(\theta) \;=\; \frac{\left[\bar{N}(\theta; 0.20) - \bar{N}(\theta; 0)\right] / \bar{N}(\theta; 0)}{-0.20}, \tag{20}$$

where the denominator -0.20 is the percent change in childcare price ($\tilde{p}_b/p_b - 1 = -s$). Under the expected sign (subsidy raises fertility), ε_N is *negative*. The two-sided band penalty applied to $|\varepsilon_N|$ is

$$p_{\text{band}}(\varepsilon) \;=\; \lambda \cdot \begin{cases} (\varepsilon_{\text{low}} - |\varepsilon|)^2 & \text{if } |\varepsilon| < \varepsilon_{\text{low}}, \\ (|\varepsilon| - \varepsilon_{\text{high}})^2 & \text{if } |\varepsilon| > \varepsilon_{\text{high}}, \\ 0 & \text{otherwise,} \end{cases} \tag{21}$$

with $\varepsilon_{\text{low}} = 0.1$, $\varepsilon_{\text{high}} = 1.0$, and $\lambda = 100$.

Rationale. The band keeps the optimizer away from two implausible basins. When $|\varepsilon_N| < 0.1$, the estimated model exhibits essentially no fertility response to a non-trivial subsidy, contrary to the quasi-experimental evidence from Gathmann and Sass [2018] ($\hat{\varepsilon} \approx -0.6$ from a Thuringia daycare-price reform) and Bauernschuster et al. [2016] (positive fertility effects from public-childcare expansion). When $|\varepsilon_N| > 1.0$ the model is hyper-responsive — a 1% price reduction triggers more than a 1% fertility increase — again inconsistent with the literature. Outside the band, the penalty grows quadratically with distance from the band edges and discourages the optimizer from these implausible regions. Inside the band the penalty is inactive and (17) collapses to $L_{\text{M2}}(\theta) = L_{\text{MSM}}(\theta)$, so the MSM machinery determines the estimate whenever $|\varepsilon_N|$ is empirically plausible. Each objective evaluation therefore requires two type-mixture simulations of the full DP (one at $s = 0$ for the baseline moments and one at $s = 0.20$ for the elasticity); the implementation includes an optional early-drop feature (`drop_factor`) that skips the $s = 0.20$ simulation for candidates whose baseline L_{MSM} already exceeds a multiple of the running best, returning a pessimistic penalty estimate so that the BlackBoxOptim selection step demotes the candidate while preserving its rank in the population.

4.5 Optimization pipeline

The augmented objective (17) is minimized by a three-stage pipeline that combines population-based global exploration (Stage 1) with derivative-free local refinement (Stages 2 and 3). The full pipeline takes about 12 hours of wall time per country. All stages optimize the same 37-parameter vector θ (Table 2) against the same band-penalty objective (17);

what differs across stages is only the search algorithm — a population-based global search in Stage 1, then local Nelder–Mead refinement in Stages 2 and 3.

Warm-start. The pipeline is warm-started from a saved 37-vector θ_0 (`model2/external/v15c/`); verified pre-launch (see Section 4.6).

Stage 1: BlackBoxOptim exploration (4 hours). A population-based differential-evolution variant (`adaptive_de_rand_1_bin_radiuslimited` in `BlackBoxOptim.jl`)[3] with population size 70. The initial population includes the warm-start θ_0 as individual 1; the remaining 69 individuals are drawn from a Gaussian centred at θ_0 with standard deviation 0.10 times the bound width in each coordinate. Stage 1 ends after four wall-clock hours regardless of internal convergence; the best θ found is saved to `v16_{de,us}_stage1.jld2` and carried forward as Stage 2's starting point.

Stage 2: Nelder–Mead multi-restart (6.5 hours). Twelve sequential Nelder–Mead restarts on the logit-transformed parameter space (so that bounds are respected by construction). Each restart re-centers the simplex at the current best θ and runs for up to 6.5h/12 = 32.5 minutes (or until the inner NelderMead converges). Simplex shrinkage follows a hard-coded twelve-step schedule $\big((0.080, 0.50), (0.060, 0.40), (0.045, 0.30), \ldots, (0.001, 0.008)\big)$ defined in `NM_SCHEDULE_STAGE2`, with successive restarts using tighter initial simplex sizes and convergence tolerances. The end-of-Stage-2 best θ is saved to `v16_{de,us}_stage2.jld2`.

Stage 3: polish (1.5 hours). Four further Nelder–Mead restarts under the tighter four-step schedule $\big((0.004, 0.03), (0.002, 0.015), (0.001, 0.008), (0.0005, 0.004)\big)$ defined in `NM_SCHEDULE_STAGE3`. Each restart runs for up to 1.5h/4 $\approx$ 22.5 minutes. The post-polish θ is the estimate, saved to `v16_{de,us}_final.jld2`.

Warm-start benchmark fallback. At pipeline completion, the new best objective L_{MSM} (stripped of the band penalty for fair comparison) is compared against the warm-start benchmark $L^{\text{warm}}_{\text{MSM}}$ via `compare_and_accept`. If $L_{\text{MSM}} \le L^{\text{warm}}_{\text{MSM}}$, the new estimate is accepted; otherwise the estimator falls back to the warm-start, guaranteeing that the reported estimate is at least as good as the benchmark. Either way the estimated θ is preserved in `*_final.jld2`; the decision summary is written to `*_decision.jld2`; and the *accepted* θ (new estimate or

[3] "BBO" abbreviates BlackBoxOptim.jl (R. Feldt, https://github.com/robertfeldt/BlackBoxOptim.jl, v0.6.5), a Julia global-optimization package. The Stage-1 method `adaptive_de_rand_1_bin_radiuslimited` is an adaptive differential-evolution algorithm (Storn and Price, 1997, *J. Global Optim.* 11:341–359): DE/rand/1/bin with greedy selection, self-adapting control parameters, and radius-limited ("trivial-geography") donor sampling.

warm-start, whichever won) is copied to `*_accepted.jld2`, the file downstream analysis scripts load.[4]

Checkpointing. Within every stage the objective closure writes a rolling `v16_{de,us}_checkpoint.jld2` every 15 minutes containing the current best θ, so that an interrupted run can be resumed from a recent state with at most fifteen minutes of lost compute.

4.6 Identification and warm-start verification

Identification. The 37-parameter vector θ is over-identified by $N_m - 37 = 255$ moment conditions in (18). Three parameters warrant operational notes specific to the computational implementation.

Latent type-share parameters (π_{O_1}, π_{O_2}). The type shares are anchored to empirically-derived intervals (NLSY79 for the US, Wang [2026] for DE; see Section 4.2).

Policy-elasticity parameter (ε_N, an implied moment of θ). The 292 estimation moments are all on-baseline ($s = 0$), so ε_N is identified only by extrapolation from the estimated θ. The augmented band penalty (21) supplies additional prior information by ruling out the two implausible basins $|\varepsilon_N| < 0.1$ and $|\varepsilon_N| > 1.0$ as off-band.

Smoothing-scale parameter (σ_ε^f). Identified from how quickly fertility hazards transition across the age profile; the smoothing nests the hardmax limit at $\sigma_\varepsilon^f = 0$ within the estimation bounds.

Warm-start verification. The implementation delivers a pre-launch verification script (`diagnostics/verify_warmstarts.jl`) that performs three checks against the saved warm-starts in `model2/external/v15c/` (Section 4.5): (i) θ has length 37; (ii) θ lies strictly inside the estimation bounds; (iii) evaluating $L_{\text{MSM}}(\theta^{\text{warm}})$ under the implementation reproduces the saved L_{MSM} value to within 10^{-3}. On the reference replication package (2026-06-03), both countries reproduce within a strict 10^{-13} relative tolerance: Germany $L_{\text{MSM}}(\theta^{\text{warm}}_{\text{DE}}) = 6.1699 = L_{\text{MSM}}^{\text{warm, DE}}$, and United States $L_{\text{MSM}}(\theta^{\text{warm}}_{\text{US}}) = 11.7737 = L_{\text{MSM}}^{\text{warm, US}}$, both with diff $= 0.0000$ at the printed precision. The strict-tolerance agreement provides direct evidence that the implementation's DP solver, 4-type mixture aggregation, moment computation, and weighting are numerically equivalent to the reference at the warm-start.

[4] In a 30-minute compressed "smoke test" conducted in the replication package on 2026-06-03, the new estimate won the comparison in both countries (Germany: 6.1623 vs 6.1699; United States: 11.7455 vs 11.7737), so the new-estimate-wins branch of the fallback decision was exercised in the test.

4.7 Asymptotic standard errors

Standard errors for the estimate $\hat{\theta}$ are computed once at convergence using the standard MSM asymptotic variance formula.[5]

Asymptotic variance formula. Under the standard MSM regularity conditions [Hansen, 1982, Newey and McFadden, 1994], the asymptotic variance of $\hat{\theta}$ is

$$\widehat{V}(\hat{\theta}) \;=\; \big(\hat{J}^{\top}\,\hat{W}\,\hat{J}\big)^{-1}, \qquad \hat{J}_{ik} \;=\; \frac{\partial m_i(\theta)}{\partial \theta_k}\bigg|_{\theta=\hat{\theta}}, \tag{22}$$

with $\hat{J}$ the 292×37 Jacobian of moment conditions and $\hat{W}$ a diagonal weighting matrix (defined below). Standard errors are the square roots of the diagonal of $\widehat{V}$. The implementation uses the pseudo-inverse `pinv` (singular-value truncated) in place of the direct inverse for numerical robustness against ill-conditioning. The band penalty (21) is inactive at $\hat{\theta}$ in both countries, so $\hat{\theta}$ is interior to the constraint set and (22) is the standard interior-optimum MSM sandwich.

Weighting matrix. The SE computation uses the bootstrap inverse-variance diagonal:

$$\hat{W} \;=\; \operatorname{diag}\!\Big(\min\!\big(1/\hat{\nu}_i,\, 10^5\big)\Big)_{i=1}^{292}, \tag{23}$$

where $\hat{\nu}_i = \widehat{\operatorname{Var}}(M_i^d)$ is the bootstrap variance of the i-th data moment. The cap at 10^5 prevents pathologically near-zero moment variances (e.g., late-age birth probabilities, where empirical counts vanish) from dominating the SE computation. The collapsed form $(J'WJ)^{-1}$ in (22) applies the inverse-variance $\hat{W}$ throughout — both in $(J'WJ)$ and as the moment-covariance proxy. The implementation thus reports a strictly consistent estimator of the variance under the inverse-variance benchmark.[6]

Jacobian construction. $\hat{J}$ is computed by central differencing with a relative step of 5% of $|\hat{\theta}_k|$ (with a small absolute floor for parameters near zero), parallelised over the 37 parameter columns via `Threads.@threads`. Each column requires two full type-mixture simulations of the DP, each at $N_{\text{sim}} = 60{,}000$ (the production setting); the full Jacobian therefore costs $\approx 2 \times 37 = 74$ DP simulations.

[5] We thank the editor for the suggestion to use closed-form asymptotic inference for the headline standard errors, rather than re-estimation bootstrap.

[6] The full Hansen sandwich $(J'W_{\text{est}}J)^{-1}\,J'W_{\text{est}}\hat{\Sigma}W_{\text{est}}J\,(J'W_{\text{est}}J)^{-1}$ would apply if the priority weights W_{est} were substituted for $\hat{W}$ throughout. When priority weights are close to constants and the moments are nearly orthogonal under $\hat{W}$, the two formulas yield numerically close standard errors; the reported t-statistics in Table 6 in the paper satisfy $|t| > 1.96$ for all 37 parameters under either convention.

Reported standard errors. The standard errors reported in Table 6 in the paper (Model 2 parameter estimates) are produced by `diagnostics/jacobian_se.jl` from (22) and (23). All 37 parameters satisfy $|t| > 1.96$ in both countries.

5 Vignette simulated panel

5.1 Overview

The vignette simulated panel is the bridge between the randomized-vignette experiment (Section 5 in the paper) and the structural estimates of Models 1 and 2. For each respondent–vignette observation, we solve the structural model at $\hat{\theta}$ (Section 3.6, Section 4.6) with the vignette attributes inserted as the household's environment, producing a model-implied fertility outcome that matches one-to-one to the respondent's stated vignette answer. The resulting two panels — the *vignette panel* (stated) and the *simulated panel* (model-implied) — are then submitted to identical fixed-effects regressions (Section 6.2 in the paper); the resemblance of the regression coefficients is the paper's cross-validation that the structural models capture the same policy responsiveness as the survey respondents.

This section documents the implementation in detail. Section 5.2 describes the vignette inputs and their five randomized attributes; Section 5.3 the mapping from vignette stimuli to each model's budget constraint, including the two perception treatments (permanent-wage and current-period-wage); Section 5.4 the Model 1 simulation panel (`model1/sim_panel_subsidy95_cap3/`); Section 5.5 the Model 2 simulated panel (`model2/vignette_simpanel/`, including the two-DP construction used for the current-period-wage treatment); and Section 5.6 the cross-validation comparison and the interpretation of vignette-anchored versus model-anchored elasticities.

5.2 Vignette inputs

Each vignette describes a hypothetical household environment, and the respondent states a desired number of children for it. The vignette data are drawn from two country-specific surveys:

Country	Source survey	Survey eligibility	Simulation-panel cells
United States	**UAS-27** (USC Understanding America Study)	age 18–50	3,056
Germany	**Pairfam** (satellite of German Family Panel, 2014 module)	age 18–50	1,128

The simulated panel restricts to the analysis sample of *female participants aged 22–42*, matching the structural model's fertility horizon (Section 6.2 table notes in the paper). The cell counts are respondent–vignette observations: because each participant typically answered a small number of vignettes, the total number of cells is approximately the number of analysis-sample respondents multiplied by the number of vignettes each respondent answered.

Randomized attributes (per vignette). Each vignette is characterized by five randomized numerical attributes (Section 5.1 in the paper):

- w_e — the household head's hourly net wage;
- $h_{p,e}$ — the partner's working hours;
- $w_{p,e}$ — the partner's hourly net wage (so that partner labor income $y_{p,e} = h_{p,e} \cdot w_{p,e}$ is a derived quantity);
- $s_e \in \{\text{costly, subsidized}\}$ — the availability of subsidized childcare;
- A_e — the child allowance per child.

Randomization design. The five attributes are drawn independently per vignette, so that the policy stimuli (A_e, s_e) and personal-characteristics stimulus ($y_{p,e}$) are by construction orthogonal to the wage stimulus w_e. This orthogonality has an important consequence for the simulated panel: the non-wage policy responses are cleanly identified, independent of the wage uncertainty, so their sign and direction are the same under either wage-perception treatment (permanent or current-period; see Section 5.3), while their precise magnitude is somewhat treatment-sensitive.

Data access. The data are accessible from the data providers free of cost for registered users. For replication, point the package's data-build scripts at locally-obtained copies; the respective Data Usage Agreements (DUAs) and acquisition procedures are documented in Section 2 and in `IER_R2_Replication/docs/DATA_PROVENANCE.md`.

5.3 Construction of synthetic (simulated) panel

For each respondent–vignette observation, the structural model is solved again at $\hat{\theta}$, replacing selected exogenous variables with vignette stimuli while holding all structural parameters fixed at their estimated values (Model 1: Section 3.6; Model 2: Section 4.6).

Attribute-to-environment mapping. The five randomized attributes (Section 5.2) enter the structural budget constraints (for Model 1, equations (3) and (14) in the paper, the without- and with-children budgets; for Model 2, the period budget $\tilde{c}_t$ defined in Appendix C) one-to-one:

Vignette attribute	Model variable	Override formula
w_e (hourly wage)	$w(t)$ (wage path / process)	M1: $w(t) := w_e$; M2: $\phi_0 := \log w_e$, $\phi_1 = \phi_2 = 0$
$h_{p,e} \cdot w_{p,e}$ (partner)	$y_p(t)$ (partner income)	M1: $y_p(t) := h_{p,e} \cdot w_{p,e}$; M2: $y_{p,t} := 12 \cdot h_{p,e} \cdot w_{p,e}$ (annualized)
A_e (allowance)	A (per-child allowance)	M1: $A := A_e$; M2: $A := 12 \cdot A_e$ (annualized)
s_e (vignette childcare)	s (subsidy rate)	$s = 0.95$ if subsidized; $s = 0$ if costly
	$\tilde{p}_b$ (effective price)	$\tilde{p}_b = (1 - s) \cdot p_b$ in both models

For Model 2, the wage substitution sets only the deterministic component of the log-wage process $w_t = \exp(\phi_0 + \phi_1 x_t + \phi_2 x_t^2 + \xi_t)$. The idiosyncratic productivity shocks ξ_t are drawn independently from a discretized normal process at each period and are unaffected by the vignette wage assignment.

Respondent state and timing. The model is initialized at the respondent's observed state. The model period index is $t_0 = \text{age} - 21$ (so age 22 maps to $t_0 = 1$ and age 42 maps to $t_0 = 21$). Model 2 additionally fixes the initial endogenous state at the respondent's observed values: $n_0 = \min(\text{observed children}, 3)$, marital status as observed, experience $x_0 = \text{age} - 22$, and age-of-youngest-child fixed at 3 for every respondent. Model 1's aggregative formulation takes only the wage and partner-income paths for state initialization.

Two perception treatments. The vignette instrument leaves it open whether w_e is a permanent or a temporary feature of the household's environment (Section 2.2 in the paper, footnote on temporary-vs-permanent). The simulated panel reports both interpretations:

- *Permanent-wage treatment*: the wage substitution applies for all $t \geq t_0$ (the rest of the household's lifecycle).

- *Current-period-wage treatment*: the wage substitution applies only at the respondent's own period t_0; from $t_0 + 1$ onward the model's estimated wage process resumes.

The non-wage stimuli (A_e, s_e, $y_{p,e}$) are applied identically across both treatments: only the wage substitution differs.

The two treatments span the natural range of how forward-looking agents might interpret the vignette stimulus. Their model-side implementations differ between Models 1 and 2: Model 1 (deterministic) applies the substitution directly to the wage path (Section 5.4), while Model 2 (forward-looking DP) requires a *two-DP construction* for the current-period-wage case to maintain internal consistency of the agent's beliefs (Section 5.5).

5.4 Model 1 simulated panel

The Model 1 simulated panel is implemented in `model1/sim_panel_subsidy95_cap3/simulate_panel_{us,de}.jl`. The folder name encodes two convention choices that match Model 2's discrete state structure: *subsidy95* (the subsidized-childcare case uses $s = 0.95$, leaving a 5% effective price) and *cap3* (the number of children is capped per-age at $\min(n, 3)$, mirroring Model 2's discrete state ceiling $n \in \{0, 1, 2, 3+\}$).

Construction. Because Model 1 is deterministic (Section 3.3), the vignette substitutions (Section 5.3) are applied directly to the wage and partner-income paths $w(t), y_p(t)$ and to the allowance/subsidy scalars $A, \tilde{p}_b$. For each respondent–vignette observation, the script solves again the two-phase optimization ((1), (2)) at $\hat{\theta}_{M1}$ under three environments per cell (baseline, permanent-wage substitution, current-period-wage substitution) and records the resulting time-average of $n(t)$ over the fertile period.

Canonical output: $N_{\mathbf{perm}}$ at vignette age a^*. The canonical Model 1 output is the permanent-wage panel value *evaluated at the respondent's own age* a^*.[7] The reason for the "at the respondent's age" choice is that the respondent's stated answer reflects *their* life situation, so the model-predicted counterpart should be evaluated at the same temporal position.

Parallel work-sharding. The per-cell solving is fully independent across vignettes (no cross-cell dependencies in the DP), so the workload can be sharded freely. The implementation work-shards via a simple offset-and-stride scheme: nine batch scripts `run_{de,us}_`

[7] In the simpanel output schema this is `N_perm_at_age` (column 19 of the per-cell output, with `position` indexing the respondent and `N_HE[position, 3]` the env-3 = permanent-wage value at age a^*).

`w[1..9].log` each process every ninth vignette (`offset =` w`, stride = 9`) starting at offset $w \in \{1, \ldots, 9\}$, producing nine partial CSVs `outputs/{DE,US}_simpanel_w[1..9].csv` ($\approx$ 126 rows each for DE, $\approx$ 340 for US). The merge step `merge_partials.jl` concatenates these into the canonical panel `outputs/{DE,US}_simpanel.csv` (all rows merged).

Outputs. The merged panel `outputs/{DE,US}_simpanel.csv` feeds `analysis/build_regression_input.jl` which assembles the fixed-effects regression input `outputs/{DE,US}_regression_input.csv` (joining stated vignette responses with model-predicted counterparts). Downstream artifacts include `outputs/Table_H1_dist_2x2_v14.csv` (the 2×2 distribution of children counts under each treatment) and `outputs/Table_H2_fe_regression.tex` (the FE regression table in paper-ready LaTeX). Per-batch run logs are in `run_{de,us}_w*.log` for traceability.

5.5 Model 2 simulated panel

The Model 2 simulated panel is implemented in `model2/vignette_simpanel/` (see `model2/vignette_simpanel/README.md` for the package layout). It reuses the same Model 2 dynamic-programming solver (Section 4.3), evaluated at the estimated parameters $\hat{\theta}_{M2}$ (Section 4.6). The vignette stimuli are then inserted into the model before each cell is solved.

Per-cell initialization. For each respondent–vignette observation, the dynamic program is initialized at the respondent's observed state: $t_0 = \text{age} - 21$, $n_0 = \min(\text{observed children}, 3)$, marital status as observed, experience $x_0 = \text{age} - 22$, and age-of-youngest-child fixed at 3 for every respondent (a uniform choice that abstracts from the small fraction of respondents with very young children at the time of the survey).

Wage-process substitution. The vignette wage w_e enters Model 2 by replacing the wage-process intercept: $\phi_0 := \log w_e$ with $\phi_1 = \phi_2 = 0$, so that the deterministic component of $w_t = \exp(\phi_0 + \phi_1 x_t + \phi_2 x_t^2 + \xi_t)$ equals w_e at every age. The transitory productivity shock $\xi_t \sim N(0, \sigma^2)$ continues to be drawn from its estimated distribution and discretized onto the same fixed grid.

Permanent-wage treatment: single DP. Under the permanent-wage treatment, the wage-process substitution applies for all $t \geq t_0$. The implementation solves again Model 2 once at the modified wage process, computes the value functions and optimal policies for all model states Ω_t, and simulates forward from Ω_{t_0} to extract the cell's predicted mean number of children.

Current-period-wage treatment: two-DP construction. The current-period-wage treatment requires more care because Model 2 is forward-looking: at period t_0, the agent's choice of labor supply and (in stage 1) fertility depends on the wage process she believes applies for the rest of life. Naively "swapping the wage at t_0 only" would have the agent solve her t_0 problem under the estimated wage process (the only one for which her belief is internally consistent), which contradicts the vignette stimulus $w_e \neq w(\text{estimated})$ at t_0. The implementation resolves this with a *two-DP construction*:

1. Solve DP_{Perm}: the dynamic program under the permanent-wage substitution (deterministic component of $w_t = w_e$ for all t), producing policy functions $\pi_{\text{Perm}}(\Omega_t)$.
2. Solve DP_N: the dynamic program under the estimated wage process (no substitution), producing policy functions $\pi_N(\Omega_t)$.
3. Construct the agent's effective policy by switching at $t_0 + 1$:

$$\pi^{\text{cur-period}}(\Omega_t) = \begin{cases} \pi_{\text{Perm}}(\Omega_t), & t = t_0, \\ \pi_N(\Omega_t), & t \geq t_0 + 1. \end{cases}$$

At t_0 the agent acts *as if* the wage is permanent (she uses π_{Perm}, the policy of an agent who believes the wage is permanent); from $t_0 + 1$ onward she acts under the estimated wage process (she uses π_N). The agent is naïve about the policy switch at $t_0 + 1$ when making her t_0 decision — a behavioral assumption the construction makes explicit.

The non-wage stimuli (A_e, s_e, $y_{p,e}$) are applied identically across both DP_{Perm} and DP_N, so the two policy functions differ only in their wage processes.

Outputs. The Model 2 simulated panel produces a per-cell CSV with both permanent-wage and current-period-wage predicted children counts, ready for the same fixed-effects regression as the Model 1 simulated panel and the stated-vignette panel (Section 5.6).

5.6 Cross-validation comparison and interpretation

The vignette simulated panel produces two stacked panels — the *vignette panel* (stated n_i^{vig}) and the *simulated panel* (model-predicted n_i^{model}) — each indexed by respondent–vignette observation i. The cross-validation compares them via parallel fixed-effects regressions and via correlations.

Fixed-effects regression. For each panel $p \in \{\text{vig}, M1\text{-}P, M1\text{-}C, M2\text{-}P, M2\text{-}C\}$ (where P = permanent-wage, C = current-period-wage) and each country, the implementation runs the linear FE regression

$$n_i^{(p)} \;=\; \beta_w^{(p)} w_i^s + \beta_y^{(p)} y_{p,i} + \beta_C^{(p)} \mathbb{I}\{C_i = \text{costly}\} + \beta_A^{(p)} A_i + \mu_i^{(p)} + \varepsilon_i^{(p)}, \tag{24}$$

where $\mu_i^{(p)}$ is a respondent fixed effect that absorbs all time-invariant unobserved heterogeneity (and the overall intercept; no separate constant is included). The coefficients $\big(\beta_w^{(p)}, \beta_y^{(p)}, \beta_C^{(p)}, \beta_A^{(p)}\big)$ quantify each panel's policy responsiveness; their resemblance across $p \in \{\text{vig}, M1\text{-}P, M1\text{-}C, M2\text{-}P, M2\text{-}C\}$ is the paper's cross-validation that the structural models capture the same responsiveness as the stated-vignette respondents do.

The estimation runs in Stata via `model1/sim_panel_subsidy95_cap3/analysis/fe_regression.do` (M1) and the parallel M2 script; results land in `outputs/Table_H2_fe_regression.tex` ready for paper inclusion. The aggregated headline numbers appear in Table 14 in the paper, which reports the FE regressions for all five panels (vignette + M1-permanent + M1-current-period + M2-permanent + M2-current-period) side by side.

Per-respondent unit elasticities. A second cross-validation comparison (Section 6.3, Table 15, and Figure 5 in the paper) computes per-respondent unit elasticities $\varepsilon_{X,i}$ with respect to the child allowance A and the daycare price p_b, using the linear FE prediction at two counterfactual policy values $X_{\text{low}}, X_{\text{high}}$:

$$\varepsilon_{X,i} \;=\; \frac{\big(\hat{n}_{i,\text{high}} - \hat{n}_{i,\text{low}}\big)/\hat{n}_{i,\text{low}}}{\big(X_{\text{high}} - X_{\text{low}}\big)/X_{\text{low}}}, \qquad X \in \{A,\; p_b\}. \tag{25}$$

Under the linear FE specification, this unit elasticity is invariant to the magnitude of the policy change. Table 15 in the paper reports the cross-respondent mean and [min, max] of $\{\varepsilon_{X,i}\}$ for each of the five panels; Figure 5 displays the full distributions as overlaid density histograms.

Interpretation: frictionless versus real-world-anchored. The two cross-validation comparisons together support an interpretation in which the vignette panel approximates a *frictionless upper bound* on policy responsiveness (respondents report a hypothetical optimum without conditioning on their actual state — existing children, assets, employment, age), while the structural model panels provide *real-world-anchored realistic estimates* (the dynamic program accounts for state-dependent choices, fertility timing, asset accumulation, and lifetime budget constraints). The difference between the vignette response and the structural prediction can be interpreted as the magnitude of fertility responsiveness that the structural model dampens via state-dependent dynamics; symmetrically, *the structural pre-*

diction provides a benchmark for what share of the stated response is achievable once internal frictions are accounted for.

The empirical finding — that the structural panels' policy elasticities lie in the same range as the vignette panel's (paper Table 15) — is the central cross-validation result. For the specific numerical comparison (magnitudes and signs of $\beta_w^{(p)}, \beta_C^{(p)}, \beta_A^{(p)}$ across the five panels; cross-panel agreement on which policy is more cost-effective in each country; and the explicit per-respondent unit-elasticity distributions), see Tables 13, 14, 15, and Figure 5 in the paper.

6 Elasticity construction

6.1 Overview

The paper reports policy elasticities of fertility under three distinct constructions that capture different aspects of policy responsiveness. *Aggregate elasticities* (Section 6.2, Section 6.3; Table 16 in the paper) summarize the whole-economy response to a uniform policy perturbation as a single scalar per (country, model); *per-respondent unit elasticities* (Table 15 in the paper; the construction is documented in Section 5.6) give a distribution of responses across respondents, one per (respondent, panel) cell; and *vignette-FE-derived elasticities* (Section 6.4; Figure 5 in the paper) re-express the fixed-effects regression coefficients of Section 5.6 as per-respondent elasticities for visual comparison across the five panels.

This section documents the implementation of the aggregate and vignette-FE-derived elasticities, and references Section 5.6 for the per-respondent unit elasticities. Section 6.2 fixes the sign and perturbation conventions; Section 6.3 describes the per-model computation of the aggregate elasticity (Model 1 via a dedicated post-estimation script; Model 2 via a post-estimation perturbation of the policy variables); Section 6.4 the vignette-FE-derived elasticity pipeline implemented in Stata; and Section 6.5 the cross-comparison reporting in Tables 15 and 16, and Figure 5 in the paper.

6.2 Aggregate elasticity

The aggregate elasticity reports a single scalar per (country, model, policy stimulus, outcome) cell, summarizing the whole-economy mean response to a uniform policy perturbation. The aggregate elasticity uses a specific perturbation convention, documented for M1 in `model1/policy_elasticity/aggregate_n_ls_v14.jl`.

Perturbation magnitudes. Each policy stimulus is perturbed by 20% ($V = 0.20$), with the sign chosen so that the change moves in the fertility-raising direction:

- *Allowance (A)*: $A \mapsto A \cdot (1 + V) = 1.20 \cdot A$ (a 20% increase in the per-child allowance relative to the country-specific baseline).
- *Daycare price (p_b)*: $p_b \mapsto p_b \cdot (1 - V) = 0.80 \cdot p_b$ (a 20% decrease in the effective childcare price, equivalent to a 20% subsidy applied to the country's baseline p_b).

Aggregation window. Each elasticity is computed from means over a model-specific window. For Model 1, this is the Phase 2 (post-$\bar{t}$) horizon ages 28–45 (18 ages; in code, `post = tbar:data_length = 7:24`). For Model 2, the analogous aggregation is over the full simulated horizon ages 22–45 (24 ages, equivalently moment indices 25–48 of the M2 moment vector). Letting $\bar{N}$ and $\bar{L}$ denote the country-level means of the simulated fertility and labor-supply profiles over the relevant window,

$$\bar{N} \;=\; \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} N(t), \qquad \bar{L} \;=\; \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} \ell(t), \tag{26}$$

where $\mathcal{T} = \{28, \ldots, 45\}$ for Model 1 and $\mathcal{T} = \{22, \ldots, 45\}$ for Model 2.

Sign-flip convention. Let $\bar{N}^p$ denote the post-perturbation mean fertility and $\bar{N}^b$ the baseline; define the elasticity

$$\varepsilon_X^N \;=\; \sigma_X \cdot \frac{\bar{N}^p - \bar{N}^b}{\bar{N}^b} \cdot \frac{1}{v_X}, \qquad \big(v_A,\, \sigma_A\big) = \big(+V,\, +1\big), \qquad \big(v_{p_b},\, \sigma_{p_b}\big) = \big(-V,\, -1\big), \tag{27}$$

where v_X is the (signed) fractional change in policy X and σ_X is the reporting sign-flip. For the allowance policy stimulus the sign-flip is unity ($\sigma_A = +1$), so ε_A^N is the standard partial elasticity. For the childcare-price policy stimulus the sign-flip is -1, so $\varepsilon_{p_b}^N$ is the *subsidy-direction* elasticity (positive when the subsidy raises fertility). Analogous formulas with $\bar{L}$ in place of $\bar{N}$ define the labor-supply elasticities ε_X^L (`eps_LS` in the saved output).

6.3 Computation per model

The aggregate elasticity is computed by re-running the structural model at the country's estimated $\hat{\theta}$ under a perturbed environment and comparing the resulting moment means to the baseline.

Model 1 (`aggregate_n_ls_v14.jl`). For each country and each policy stimulus $X \in \{A, p_b\}$, the script:

1. Loads the estimate $\hat{\theta}$ from `model1/locked/par_$(country).jld2` and the country config with the bound widening (`cfg.LB[10] = 0.40` for the US).
2. Calls `matching_fun(par,d,W,cfg;allowance=A_b,pb_value=p_b_b)` at the baseline values (A_b, p_b^b) to obtain the baseline fertility and labor-supply profiles `out_base.model.nmodel`, `out_base.model.lmodel`.
3. Calls `matching_fun` again under the perturbed environment: $(A_b(1+V), p_b^b)$ for the allowance, or $(A_b, p_b^b(1-V))$ for the daycare price, to obtain the perturbed profiles.
4. Aggregates each profile over the Phase 2 window $\mathcal{T} = \{28, \ldots, 45\}$ to obtain $\bar{N}^b, \bar{N}^p, \bar{L}^b, \bar{L}^p$.
5. Applies the sign-flip formula (27) (and its $\bar{L}$ counterpart) to obtain ε_X^N and ε_X^L.

The results are saved to `model1/policy_elasticity/outputs/M1_aggregate_elasticity_v14.jld2` as a 2×2 nested dictionary indexed by (country $\times$ policy stimulus) $\in \{\mathrm{US}, \mathrm{DE}\} \times \{A, p_b\}$, each entry storing $(\bar{N}^b, \bar{N}^p, \bar{L}^b, \bar{L}^p, \varepsilon_X^N, \varepsilon_X^L, v)$.

Model 2. The dedicated M2 aggregate-elasticity script is `model2/analysis/policy_elasticity_table10.jl`. It defines inline `with_all(env,A_new)` and `with_sub(env,s_new)` helpers that return a copy of the country `CountryEnv` struct with the `All` or `sub` field replaced; the elasticity is then obtained by calling `solve_model_moments_v16(theta,with_all(env_base,1.2*env_base.All))` (allowance, 20% increase) or `solve_model_moments_v16(theta,with_sub(env_base,0.20))` (subsidy) and aggregating the fertility moments `mm[25:48]` over ages 22–45 to form the elasticity of mean cumulative fertility, evaluated once at $\hat{\theta}_{M2}$ from the estimate and reported in Table 16 in the paper.

Equivalent reporting. Despite the very different computational machinery (Section 3.3 continuous-time analytical seed plus fixed-point iteration; Section 4.3 discrete-time backward induction over the 4-type heterogeneity mixture), the resulting elasticities $\left(\varepsilon_A^N, \varepsilon_{p_b}^N, \varepsilon_A^L, \varepsilon_{p_b}^L\right)$ are directly comparable across the two models and across the two countries; the cross-comparison is reported in Table 16 in the paper (Section 6.5).

6.4 Vignette-FE-derived elasticities

To compare policy responsiveness across the five panels (vignette, $M1$-P, $M1$-C, $M2$-P, $M2$-C; see Section 5.6), the implementation re-applies the per-respondent unit-elasticity formula (25) from each panel's own fixed-effects regression predictions, producing a distribution of per-respondent elasticities for each (panel, country) cell. The implementation is in `elasticity/scripts/run_elasticity.do` (Stata).

Panels and inputs. The script processes 10 input `.dta` files (5 panels $\times$ 2 countries):

Panel ID	Source	Country
`vignette_{us,de}`	stated-vignette responses (UAS, Pairfam)	US, DE
`M1_{us,de}_perm`	Model 1 permanent-wage simpanel	US, DE
`M1_{us,de}_naive`	Model 1 current-period-wage simpanel	US, DE
`M2_{us,de}_perm`	Model 2 permanent-wage simpanel	US, DE
`M2_{us,de}_naive`	Model 2 current-period-wage simpanel	US, DE

Each input panel is restricted to female respondents aged 22–42. The panel structure is `xtset id sid` (respondent id, vignette sid within respondent).

Fixed-effects specification. For each panel, the script runs the 4-regressor FE regression

```
xtreg r_children s_wage s_y_partner s_c_transfer s_costly_care, fe vce(cluster id)
```

where the four stimuli are the vignette environment (Section 5.2) and clustering at the respondent level accounts for within-respondent dependence across vignettes. This is exactly the regression specification of equation (24).

Counterfactual policy values. The two policy counterfactuals take the values used in Table 15 in the paper:

- Child allowance A: \$200 vs \$400 per month (US); €250 vs €500 per month (Germany).
- Daycare price p_b (via `s_costly_care`): 0.05 (subsidized, matching the 95%-subsidy strength used in the paper's Section 6.1 simpanel construction; Section 5.4) vs 1.0 (costly).

Per-respondent elasticity formula. For each respondent i and each policy stimulus $X \in \{A, p_b\}$, the FE-predicted children counts at the two counterfactual policy values $X_{\text{low}}, X_{\text{high}}$ give

$$\begin{aligned}\hat{n}_{i,\text{high}} &= \hat{\alpha}_i + \hat{\beta}_X \cdot X_{\text{high}} + (\text{other stimuli at base}),\\ \hat{n}_{i,\text{low}} &= \hat{\alpha}_i + \hat{\beta}_X \cdot X_{\text{low}} + (\text{other stimuli at base}),\end{aligned}$$

where $\hat{\alpha}_i$ is the respondent fixed effect from the panel's FE regression. The per-respondent unit elasticity is then (25) of Section 5.6.[8]

[8] Under the linear FE specification, the per-respondent unit elasticity is invariant to the magnitude of the policy change (only the direction of $X_{\text{high}} - X_{\text{low}}$ matters). The script header verifies this empirically by

Outputs. The script writes per-panel `.csv` and `.dta` files to `elasticity/outputs/` (one file per panel $\times$ country $\times$ $\{naive, perm\}$). These ten distributions then feed `outputs/fig_elasticity_histograms.pdf` (Figure 5 in the paper), which overlays the five per-country distributions in a 2×2 grid (US transfer / US daycare / DE transfer / DE daycare).

6.5 Cross-comparison and reporting

The three elasticity constructions documented in Section 6.2, Section 6.3, and Section 6.4 feed into three paper-side reporting artifacts that together constitute the paper's cross-validation of policy responsiveness.

Table 15 in the paper (per-respondent unit elasticities). For each of the five panels and both policy stimuli $X \in \{A, p_b\}$, Table 15 reports the cross-respondent mean of the per-respondent unit elasticity $\{\varepsilon_{X,i}\}_{i=1}^{N}$ together with the $[\min, \max]$ range over the indicated sample (female participants, age 22–42).

Table 16 in the paper (aggregate elasticities). Table 16 reports a single ε_X^N per (country, model, policy stimulus) cell, computed under the sign-flip convention of (27). For Model 1 the cells come from `M1_aggregate_elasticity_v14.jld2` (Section 6.3); for Model 2 they come from analogous `model2/analysis/` scripts applied at the estimate. The aggregation window differs across models (Phase 2 ages 28–45 for M1, full horizon 22–45 for M2); Table 16 reports both sets side-by-side.

Figure 5 in the paper (overlaid histograms). Figure 5 displays the full $\{\varepsilon_{X,i}\}$ distributions from Section 6.4 as overlaid density histograms in a 2×2 grid: rows are countries (US, Germany), columns are policy stimuli (A, p_b). Each subplot superimposes the five panels' distributions on per-column x-axis ranges ($[-0.20, +0.40]$ for the allowance columns, $[-1.30, +0.20]$ for the daycare-price columns) with a dashed line at $\varepsilon = 0$; the legend identifies the five panels (*vig*, *M1_P*, *M1_C*, *M2_P*, *M2_C*).

Interpretation across modeling lenses. The three reporting artifacts share a common interpretive frame: the *vignette panel* provides a frictionless upper bound on policy responsiveness (Section 6.3 in the paper and Section 5.6); the two *structural panels* dampen this response through their respective state-dependent dynamics (M1 via the continuous-time

computing the daycare-price elasticity at $s_{\text{costly\_care}} \in \{0.5, 0.05\}$ vs 1.0: the two distributions agree to machine precision. The $s_{\text{costly\_care}} = 0.05$ version is retained for consistency with the 95%-subsidy construction of the simulated panel.

aggregative formulation, Section 3.3; M2 via the discrete-time DP with 4-type heterogeneity, Section 4.3). The empirical finding of the paper — that for Germany the three approaches are consistent and the daycare subsidy is the more cost-effective policy, while for the United States cost-effectiveness depends on which modeling lens is applied — emerges from the side-by-side comparison of these elasticities (Sections 6 and 7 in the paper).

The cost-effectiveness conclusion is anchored in a back-of-the-envelope calculation: a \$100 (€100) per-month-per-child policy yields an annual fiscal cost of 0.07%–0.24% of GDP across the two countries and two models, within a feasible budget envelope. The *within-this-envelope ranking* of the two policy stimuli is the empirical pay-off of the cross-validation work in Section 5 and Section 6.

7 Replication guide

7.1 Overview

This section orients a replicator to the `IER_R2_Replication/` package and points to the in-package documentation files that hold the operational details. The Online Computational Appendix (this document) explains the *scientific content* of each sub-pipeline (Section 2–Section 6); `IER_R2_Replication/` contains the *runnable code*, *cleaned data*, and *operational documentation*. The two are designed to be read together.

Section 7.2 sketches the package layout and the sub-package responsibilities. Section 7.3 documents the software and hardware requirements (pinned via `Project.toml`/`Manifest.toml` for the Julia components, Stata 17 for the data pipeline). Section 7.4 summarizes raw-data access (PSID, SOEP, UAS, Pairfam, etc. are restricted-access; DUAs and acquisition procedures reside in `IER_R2_Replication/docs/DATA_PROVENANCE.md`). Section 7.5 describes the build order from raw microdata to paper outputs. Section 7.6 maps each paper-side table and figure to the script that produces it (the canonical full version is located in `IER_R2_Replication/docs/FILE_MAP.md`).

7.2 Package structure

The replication package has the following top-level layout:

```
IER_R2_Replication/
|-- README.md                <- top-level orientation
|-- docs/                    <- operational documentation
|   |-- VERSION_LOCK.md      <- canonical version pointers per component
```

```
|   |-- FILE_MAP.md            <- paper Table/Figure -> script (full)
|   |-- RUNTIME.md             <- hardware, software, wall times
|   |-- DATA_PROVENANCE.md     <- raw-data sources + DUAs
|   `-- TASK.md                <- internal migration plan
|-- data_processing/          <- Stata data pipeline
|-- data/                     <- cleaned model-ready inputs (shipped per scope choice)
|   |-- us/  de/   (deliverables/, moments/, summary_stats/, ...)
|-- model1/                   <- Model 1 (estimate) - self-contained (see Section 2)
|   |-- README.md
|   |-- src/  configs/  data/  estimation/  diagnostics/  locked/
|   |-- external/v11/locked/par_*.jld2
|   `-- policy_lifecycle/  policy_elasticity/  sim_panel_subsidy95_cap3/
|       elasticity_vig_m1/
|-- model2/                   <- Model 2 (estimate) - self-contained (see Section 3)
|   |-- README.md
|   |-- src/  environments/  estimation/  diagnostics/  scripts/  results/
|   |-- analysis/  external/{v15c,v11}/
|   `-- vignette_simpanel/     <- M2 vignette simulated panel (see Section 4)
|-- elasticity/               <- Cross-model elasticity Stata pipeline
|                                 (see Section 5)
|-- figures/  tables/         <- Paper-ready outputs
`-- tests/                    <- Smoke + bit-exact tests
```

Sub-package convention. Each of `model1/`, `model2/`, `data_processing/`, and `elasticity/` is a self-contained unit with its own `README.md` (replicator-facing), an environment file (`Project.toml` for Julia sub-packages, conventions for Stata), and its own `external/` for cross-version dependencies. The sub-packages can be exercised independently: a replicator interested only in re-estimating Model 1 can ignore the others. Cross-package references (e.g., the cross-model Figure 7 generator in `model1/policy_lifecycle/` that consumes Model 2 outputs) are explicit and few.

7.3 Software and hardware requirements

The package is pinned to specific software versions to ensure reproducibility of the numerical results.

Software.

- **Julia 1.12.5** for the structural-model sub-packages (`model1/`, `model2/`, and the simulated panels). Each sub-package provides a `Project.toml` and a `Manifest.toml` that pin all transitive package versions (243 packages for `model1/`, 229 for `model2/`). Use `julia --project=. -e 'using Pkg; Pkg.instantiate()'` to materialise the locked environment.
- **Stata 17** (or compatible) for the data pipeline (`data_processing/`; $\sim$ 25 `.do` files) and for the cross-model elasticity pipeline (`elasticity/`; `run_elasticity.do`).
- **Python 3.10+** (with `matplotlib`, `numpy`, `pandas`) for three small post-processors: `model1/elasticity_vig_m1/plot_fig5_histograms.py` (Figure 5 in the paper), `literature/plot_fig_lit_forest.py` (Figure 6 in the paper), and `data_processing/scripts/_reformat_table_captions.py` (Appendix A caption-formatter).

Reference hardware. The wall-time figures cited throughout (Section 3.5, Section 4.5) are measured on a 10-core Apple Silicon machine. The package runs on any x86_64 or ARM Linux/macOS host that supports Julia 1.12 and Stata 17. Memory: $\leq$ 4 GB peak per country for the Julia components.

For the canonical reference-machine spec, per-stage wall times, and the bootstrap-precompile guidance, see `IER_R2_Replication/docs/RUNTIME.md`.

7.4 Raw-data access

The package delivers *cleaned*, model-ready data in `data/{us,de}/` but does *not* deliver the raw survey microdata. The raw surveys are restricted-access; the package's data-pipeline scripts (`data_processing/`) expect raw files in a separate `data_raw/` directory that the replicator populates after obtaining the DUAs:

Survey	Country	Used by	Access
PSID	US	M1, M2	Free; registration only
SCF (2007 wave)	US	M1 (wealth)	Public
CPS	US	M1 (labor)	Public
ATUS	US	M1, M2 (time use)	Public
NLSY79	US	M2 (π priors)	Registration only
UAS-27	US	Vignette	Registration + IRB
SOEP	DE	M1, M2	DUA + institutional
Pairfam (2014 satellite module)	DE	Vignette	DUA

DUA application typically takes weeks (SOEP and Pairfam being the slowest, $\sim$ 4–8 weeks); UAS approval is usually fast for academic affiliates. The exact filename requirements, expected raw formats, and per-source acquisition procedure are documented in `IER_R2_Replication/docs/DATA_PROVENANCE.md`. The cleaned data shipped with the package allows a replicator to verify all downstream Julia/Stata results using `make from-cleaned` (skipping the Stata data pipeline); running `make all` against the raw data confirms the cleaned data was generated correctly and exercises the entire build. A consequence is that Stage 1 itself — the construction of the cleaned data from the raw microdata — is auditable only by replicators with DUA access; replicators without DUAs verify Stages 2–5 against the shipped cleaned data and take Stage 1's correctness on trust.

7.5 Build order

A full cold build of our replication package proceeds in five stages. Stages 2a and 2b are independent and can run in parallel.

Stage	Component	Output	Reference wall
1	`data_processing/` (Stata, $\sim$ 25 `.do` files)	`data/{us,de}/deliverables/*.csv`	$\sim$ 1–3 h
2a	`model1/estimation/02_run_estimate.jl` (Julia)	`model1/locked/par_{us,de}.jld2`	$\sim$ 2 min
2b	`model2/scripts/run_v16_parallel.sh` (Julia, parallel)	`model2/results/v16_{de,us}_accepted.jld2`	$\sim$ 12 h
3	`model1/sim_panel_subsidy95_cap3/` + `model2/vignette_simpanel/` (Julia)	simpanel CSVs	$\sim$ 30 min
4	`elasticity/scripts/run_elasticity.do` (Stata)	per-panel elasticity files	$\sim$ 5 min
5	figure + table generators (Julia + Python)	`figures/Fig*.pdf`, `tables/Table*.tex`	$\sim$ 30 min

Stage 2a re-estimates M1 from the shipped warm-start (`model1/estimation/02_run_estimate.jl`); the warm-start is itself the estimate, so NLopt converges to it within a few iterations (within `FTOL_ABS` $= 10^{-3}$). Replicators who want to re-estimate M1 from a different (e.g., cold) starting point edit the warm-start specification in that script.

A skip-Stage-1 path uses the shipped cleaned data and starts at Stage 2a/2b, which is the recommended path for downstream verification. A skip-Stage-2 path uses the saved `model{1,2}` estimates and starts at Stage 3, which is the fastest end-to-end re-build of the paper outputs from estimates forward (a few hours total). Verification scripts in `model{1, 2}/diagnostics/` confirm bit-exact reproduction at the saved $\hat{\theta}$ (Section 3.6, Section 4.6).

7.6 Mapping paper outputs to scripts

The condensed mapping from paper outputs to producing scripts is:

Paper output	Producing script
Figure 2 (Model 1 fit, 2×3 with 5-yr bins)	`model1/estimation/plot_fig2_combined.jl`
Figure 3 (Model 2 fit, 2×3 with 5-yr bins)	`model2/analysis/gen_fig_paper_v16_3panel.jl`
Figure 5 (elasticity histograms, 2×2)	`model1/elasticity_vig_m1/plot_fig5_histograms.py`
Figure 6 (literature forest plot)	`literature/plot_fig_lit_forest.py`
Figure 7 (combined N profile, M1+M2)	`model1/policy_lifecycle/plot_section7_combined_N.jl`
Table 4 (M1 parameter estimates + SEs)	`model1/locked/par_{us,de}.jld2` + `jacobian_se.jld2`
Table 6 (M2 parameter estimates + SEs)	`model2/results/v16_{de,us}_accepted.jld2`
Table 14 (FE regression: vignette + M1/M2 panels)	`model1/sim_panel_subsidy95_cap3/analysis/fe_regression.do` + `model2/vignette_simpanel/src/run_fe_regression.do`
Table 15 (per-respondent unit elasticities)	`elasticity/scripts/run_elasticity.do`
Table 16 (aggregate elasticities)	`model1/policy_elasticity/aggregate_n_ls_v14.jl` + `model2/analysis/` (M2 counterpart)

The canonical full mapping (every `\includegraphics{}` and `\input{table}` in the paper, with the exact source script, inputs, and output paths) is located in `IER_R2_Replication/docs/FILE_MAP.md`. Replicators interested in regenerating a specific paper output should start from `FILE_MAP.md`.

8 End-to-end verification run

This section reports a single end-to-end `make all` run that exercises the entire pipeline documented in Section 7, from raw microdata to all paper-cited figures and tables. The run serves two purposes: (i) it confirms that the `IER_R2_Replication/` package is internally consistent and buildable in a single invocation, and (ii) it provides per-stage wall-time references for a replicator budgeting compute time.

8.1 Run conditions

Date. The run was launched on 2026-06-04 at 11:32:29 local time and completed on 2026-06-05 at 04:11:16 local time, for a total wall-clock of **16 h 39 min**. The driver was the top-level `make all` target (`IER_R2_Replication/Makefile`), with its full stdout/stderr captured to `/tmp/ier_makeall.log` and exit status 0.

Hardware. 10-core Apple Silicon machine (M-series), 32 GB RAM, macOS 14. This matches the reference specification cited throughout Section 7.3.

Software. Julia 1.12.5, Stata 17 (`StataNow/StataMP`), Python 3.10. Each Julia sub-package was used at its pinned `Manifest.toml` version.

Raw-data state. All eight raw surveys (Section 7.4) were resolved to their on-disk locations via the `data_raw/` symlink directory. The run therefore exercises the full Stage 1 Stata pipeline rather than the `make from-cleaned` skip-Stage-1 path.

Snapshot. The numerics in this section reflect the package as of 2026-06-04. Replicators verifying against a later package version should re-run `make all` and compare against the regenerated log in `IER_R2_Replication/docs/run_logs/`.

8.2 Per-stage wall times

Stage	Wall	Component and notes
1	~ 4 min	data pipeline (Stata): 11:32:29 → 11:36:44
2a	< 1 min	M1 re-estimation (Julia): converges from shipped warm-start to the estimate within FTOL_ABS $= 10^{-3}$
2b	12 h 8 min	M2 estimation (Julia, parallel): 11:36:44 → 23:44:23; DE + US in parallel, 4+4 threads on 10-core
3a-DE	~ 6 min	M1 sub=0.95 + per-age-cap simpanel: 1 128 cells in 367 s
3a-US	~ 39 min	M1 sub=0.95 + per-age-cap simpanel: 3 056 cells in 2 332 s
3b-DE	12.9 min	M2 vignette simpanel: 1 128 cells in 775 s; mean realized fertility 1.82
3b-US	38 min	M2 vignette simpanel: 3 056 cells in 2 281 s; mean realized fertility 2.02
4	~ 3 min	cross-model per-respondent elasticity (Stata)
5	~ 30 min	paper figures 2, 3, 5, 6, 7 + tables 6, 13–16
Total	**16 h 39 min**	exit code 0

Stage 2b dominates total wall-clock. Within Stage 2b, M2 DE finished at 23:44:06 and M2 US at 23:44:23, within 17 seconds of each other — consistent with the design intent of the `model2/scripts/run_v16_parallel.sh` dispatcher (Section 4.5). Stages 2a and 2b run in parallel in principle; in this run Stage 2a executed in a few seconds at the start of Stage 2b's wall, so its time is folded into Stage 2b.

8.3 Output artifacts produced

The run produced the full set of paper-cited artifacts from raw microdata:

Estimation outputs (Stage 2).

- `model1/locked/par_{us,de}.jld2` — M1 estimate, re-verified against the shipped estimate at a strict 10^{-13} relative tolerance.
- `model2/results/{DE,US}/v16_{de,us}_accepted.jld2` — the canonical M2 acceptance artifacts referenced by paper Table 6; produced freshly by Stage 2b.
- `model2/results/{DE,US}/v16_{de,us}_{final,decision}.jld2` — companion files documenting the optimizer trajectory and the accept-vs-reject decision.

Simulation panels (Stage 3).

- `model1/sim_panel_subsidy95_cap3/outputs/` — M1 vignette simulated panel under the sub $= 0.95$ + per-age $\min(n, 3)$ cap protocol of Section 5.4.
- `model2/vignette_simpanel/results/sim_panel_{de,us}.csv` — the M2 vignette simulated panel.

Figures in the paper (Stage 5). All written to `figures/`:

- `fig_model1_fit_combined_v14.pdf` — Figure 2 in the paper.
- `fig_model2_fit_combined_v16.pdf` — Figure 3 in the paper.
- `fig_elasticity_histograms.pdf` — Figure 5 in the paper.
- `fig_lit_forest.pdf` — Figure 6 in the paper.
- `fig_policy_N_combined_v14v16.pdf` — Figure 7 in the paper.

Paper Figures 1 (vignette-survey illustration, a static PDF screenshot) and 4 (out-of-sample validation diagram, built inline as TikZ) are not code-generated and are not part of the make-pipeline verification set.

Tables in the paper (Stage 5). Tables 6 (Model 2 parameter estimates), 13 (construction consistency) and 14 (FE regression), 15 (per-respondent unit elasticities), and 16 (aggregate elasticities) were regenerated via the script set documented in Section 7.6.

8.4 Sanity-check numerics

A small set of headline numerics from the run, useful as a sanity-check anchor for replicators:

M2 vignette-panel realized fertility (Stage 3b). Mean simulated completed fertility across all vignette cells (wage, partner income, allowance, and childcare) is 1.82 in DE and 2.02 in US — consistent with the realized vignette moments documented in Section 5.

Table 16 in the paper (M1 rows). Generated at Stage 5 by `model1/policy_elasticity/aggregate_n_ls_v14.jl`.

	$\varepsilon_{\bar{N}}$		$\varepsilon_{\overline{LS}}$	
Policy	Paper	Package	Paper	Package
US Allowance (+20%)	+0.013	+0.013	+0.021	+0.021
US Subsidy (price − 20%)	+1.262	+1.262	+0.016	+0.016
DE Allowance (+20%)	+0.054	+0.054	+0.101	+0.101
DE Subsidy (price − 20%)	+1.326	+1.326	+0.114	+0.114

Table 16 in the paper (M2 rows). Generated at Stage 5 by `model2/analysis/policy_elasticity_table10.jl`. The aggregation is the same elasticity-of-mean-cumulative-fertility formula $\varepsilon_{\bar{N}} = (\bar{N}^{\text{policy}} - \bar{N}^{\text{base}})/\bar{N}^{\text{base}}/0.20$ that defines the M2 internal MSM band-penalty $\varepsilon_N(\theta)$ (`model2/estimation/v16_common.jl` lines 136–137), with $\bar{N} = \frac{1}{24}\sum_{a=22}^{45} N(a)$.

Policy	Paper $\varepsilon_{\bar{N}}$	Package $\varepsilon_{\bar{N}}$
US Allowance (+20%)	+0.266	+0.2662
US Subsidy (sub $0 \rightarrow 0.20$)	+0.174	+0.1736
DE Allowance (+20%)	+0.348	+0.3479
DE Subsidy (sub $0 \rightarrow 0.20$)	+0.835	+0.8351

All eight M1 cells and all four M2 cells of Table 16 in the paper reproduce to three decimal places.

8.5 Reproducing this run

To re-run the verification end-to-end:

1. Populate `IER_R2_Replication/data_raw/` with the eight raw surveys per `docs/DATA_PROVENANCE.md`.
2. From `IER_R2_Replication/`, run `make all`.
3. On the reference hardware, expect total wall-clock of $\sim$ 16–18 h (dominated by Stage 2b's $\sim$ 12 h M2 estimation). Faster downstream-only verification is available via `make from-estimates` (skips Stages 1 and 2; runs in a few hours) or `make from-cleaned` (skips Stage 1 only; runs in $\sim$ 13–14 h).

The full log of the run described in this section is preserved at `IER_R2_Replication/docs/run_logs/make_all_2026_06_04.log` for side-by-side comparison.

References

S. Bauernschuster, T. Hener, and H. Rainer. Children of a (Policy) Revolution: The Introduction of Universal Child Care and Its Effect on Fertility. *Journal of the European Economic Association*, 14(4):975–1005, 2016.

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

L. P. Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, 1982.

W. K. Newey and D. McFadden. Large Sample Estimation and Hypothesis Testing. In R. F. Engle and D. L. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier, 1994.

H. Wang. Fertility and Family Leave Policies in Germany: Optimal Policy in a Dynamic Framework. *The Review of Economic Studies*, 2026. Accepted.