CrossMark

The gender income gap and the role of family formation revisited

A replication of Bobbitt-Zeher (2007)

Fabian Ochsenfeld[1]

Abstract This article reports the results of a replication of Bobbitt-Zeher's 2007 article "*The Gender Income Gap and the Role of Education*". Models that emulate the original specifications (by and large) reproduce the original results. However, models that adhere to Bobbitt-Zeher's theory concerning the gendered effect of family formation call into question her finding that "*values appear to matter only modestly, while family formation has virtually no effect on the income gap*".

Keywords Gender gap · Motherhood penalty · Horizontal sex segregation · Replication · Decomposition methods

JEL codes C00 · J12 · J16 · J24 · J29 · J31

Ein zweiter Blick auf die Bedeutung der Familiengründung für das Zustandekommen geschlechtsspezifischer Einkommensungleichheit
Eine Replikationsstudie zu Bobbitt-Zeher (2007)

Zusammenfassung Dieser Beitrag berichtet die Ergebnisse einer Replikation von Bobbitt-Zehers 2007 erschienenem Aufsatz „*The Gender Income Gap and the Role of Education*". Modelle, welche die ursprünglichen Spezifikationen nachbilden, replizieren (im Großen und Ganzen) die ursprünglichen Ergebnisse. Modelle, die hingegen Bobbitt-Zehers theoretischen Ausführungen bezüglich dem ge-

✉ Fabian Ochsenfeld
ochsenfeld@soz.uni-frankfurt.de

1   Institute for Sociology, Goethe University,
     Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main,
     Germany

schlechtsspezifischen Effekt der Familiengründung folgen, ziehen jedoch ihren Befund in Zweifel, wonach „*Werten nur eine bescheidene Bedeutung zukommt, während die Familiengründung praktisch keinen Effekt auf die Einkommensungleichheit hat*".

1 Introduction

The gender pay gap remains both a salient social problem and a puzzle to social scientists as it persists despite the institutionalization of egalitarian gender norms in the labor market and the reversal of the male advantage in education. To resolve the puzzle, scholars have sought to identify the relative contribution of different factors to the gender gap. Whereas research on work-family compatibility, overwork, and the motherhood penalty (Waldfogel 1997; Gangl and Ziefle 2009; Cha and Weeden 2014; Goldin 2014) suggests that the strongly gendered effect of family formation accounts for a substantial part of the gap, a highly-cited 2007 *Sociology of Education* article that scrutinized the contribution of educational factors relative to family formation among the college-educated concluded that "*family formation has virtually no effect on the income gap*" (Bobbitt-Zeher 2007, p. 13). It also found that "*values appear to matter only modestly*" (Ibid.). According to that study, gender differences in fields of study instead explain the lion's share of the income gap. In this article, I assess the robustness of these results by way of replication. This is an endeavor worth undertaking given that the original findings have important policy implications, as they suggest that improvements in work-family compatibility would do nothing to reduce economic gender inequality.

I first conduct a pure replication. Fitting statistical models that emulate those of the original study closely repro-

🕿 Springer

duce the original findings. However, I argue that these estimates are biased because Bobitt-Zeher restricts her sample to year-round full-time workers and thus conditions on labor force participation which, according to her own theoretical argument, mediates the effect of family formation on income. I show that her estimates for the importance of values are negatively biased for the same reason. Equally important, I also show that the original study misspecified and misallocated the hypothesized gendered effect of family formation in both relevant decompositions. The results from a series of models that subsequently correct for endogeneous sample selection and misspecification suggest the original study indeed severely underestimated the importance of family formation, and also that of values. According to my findings these two factors explain nearly a third of the gender income gap in the sample at hand, and thus about as much as the educational factors emphasized in the original article. My re-analysis corroborates Bobbitt-Zeher's finding that education-related factors in absolute terms explain a sizeable share of the gender income gap.

For the sake of brevity, I choose not to debate here Bobbitt-Zeher's decision to analyze incomes rather than wages or the issue of omitted variable bias in testing the theorized devaluation of college majors. These issues have been discussed elsewhere (Petersen 1989; Morgan and Arthur 2005; Gerber and Cheung 2008; Ochsenfeld 2014).

My intention is not to single out Bobbitt-Zeher's analysis for criticism. In fact, studies published in the top U. S. and German sociology journals have drawn similar conclusions based on very similar designs before and after (Marini and Fan 1997; Leuze and Strauß 2009). The purpose of this replication is thus to critically reflect on this enduring research tradition rather than to fault an individual author.

2 Overview of the original study

Bobbitt-Zeher (2007) studies the causes of income inequality between women and men in a cohort of young college graduates. She places particular emphasis on scrutinizing comprehensively the role of education and the role of family formation, arguing that the two have rarely been studied in conjunction (Ibid., p. 6). To "*understand the weight of the two sets of influences relative to one another*" (Ibid.) is thus the declared aim of her study. The author investigates this issue with data from the *National Education Longitudinal Study of 1988* (NELS:88) which she restricts to persons with 4-year college degrees who work at least 35 h per week throughout 1999 (Ibid., p. 7). She derives her main findings from two regression-based income decompositions.

The first decomposition (Table 1; Bobbitt-Zeher 2007, p. 13) starts from a baseline model with gender as the only predictor and sequentially adds variable groups that measure respondents' background, values, education, family status, and work characteristics to decompose the unconditional gender income gap into these five components. The order in which the author adds these variable groups to the set of predictors is based on their position in the life course trajectory. A reduction in the coefficient for the dummy variable *female* can thus be interpreted as a mediation and explanation of the gender income gap by the added set of covariates. Bobbitt-Zeher finds that differences in how much women and men value '*having lots of money*' explain only 7% of the gender income gap in addition to differences in background, whereas adding education-related covariates explains a third of the gap. In her analysis, the variables that measure respondents' family formation status do not further reduce the conditional income gap, but work-related factors such as hours worked, industry, and occupation account for additional 25% of the gap.

The second (Oaxaca-Blinder) decomposition (Table 2; Bobbitt-Zeher 2007, p. 14) fits income regressions separately for women and men using the same measures for background, values, education, family formation, and job characteristics as covariates. In order to calculate the *percentage explained* by gender differences in endowment with the covariates, this procedure averages over women's and men's coefficients to obtain hypothetical gender-neutral rates of return to the covariates. It then uses these hypothetical rates to estimate the degree to which gender differences in income are due to gender differences in endowment with the various covariates and to which degree the difference in outcome is instead due to gender differences in the rates of return to these endowments. Compared with the first decomposition, Bobbitt-Zeher's Oaxaca-Blinder decomposition attributes less explanatory power to education-related factors and more to work-related variables. This is because the Oaxaca-Blinder decomposition ignores the causal order of variable groups and hence ignores the fact that work-related variables mediate the effects of the education-related characteristics on income. As in the second decomposition, family formation seemingly plays no role whatsoever in the generation of women's economic disadvantage.

Based on the results from the first and second decomposition, the author concludes that besides differences in job characteristics, horizontal sex segregation in fields of study was crucial to understand income differences between college-educated women and college-educated men, whereas "*values appear to matter only modestly, while family formation has virtually no effect on the income gap for this sample of young workers.*" (Bobbitt-Zeher 2007, p. 13).

Table 1 Regression coefficients for *female* and percentage of the gender income gap explained with alternate models, including background, values, education, family formation, and work factors

Model number	Model description	Bobbitt-Zeher N = 1946			Replication 1A: pure replication N = 1924			Replication 1B: replication on unrestricted sample N = 3290			Replication 1C: replication on unrestricted sample with correction for misspecification N = 3290		
		Income gap (b_female)	% of gap explained	Δ (%)	Income gap (b_female)	% of gap explained	Δ (%)	Income gap (b_female)	% of gap explained	Δ (%)	Income gap (b_female)	% of gap explained	Δ (%)
1	Female	−6938	–	–	−6670	0	–	−5707	0	–	−5707	0	–
2	Female and background	−6643	4	4	−6373	4	4	−5496	4	4	−5496	4	4
3	Female, background, and values	−6166	11	7	−5849	12	8	−4595	19	16	−4595	19	16
4	Female, background, values, and education	−3903	44	33	−3849	42	30	−2681	53	34	−2681	53	34
5	Female, background, values, education, and family formation (not interacted with gender)	−3854	44	0	−3789	43	1	−2771	51	−2	–	–	–
5b	Female, background, values, education, and family formation (interacted with gender)	–	–	–	–	–	–	–	–	–	−1813	68	15
6	Female, background, values, education, family formation, and work	−2132	69	25	−2569	61	18	−1460	74	23	−1050	82	13

Sources: Bobbitt-Zeher (2007, p. 13); NELS:88, own (weighted) estimates.
Δ change in percentage explained relative to previous model

Table 2 Regression decompositions showing contributions of background, values, education, family formation, and work characteristics to the gender income gap

| | Bobbitt-Zeher | | Replication 2A: pure replication | | Replication 2B: replication on unrestricted sample | | Replication 2C: replication on unrestricted sample without intervening work-related variables | |
| | N = 1946 | | N = 1924 | | N = 3290 | | N = 3290 | |
Variable (endowments effects if not stated otherwise)	Diff. in income explained	% of total gap explained	Diff. in income explained	% of total gap explained	Diff. in income explained	% of total gap explained	Diff. in income explained	% of total gap explained
Background	78	1	187	3	152	3	186	3
Importance of having lots of money	383	6	401	6	592	10	911	16
Education-related								
Grades	330	5	–278	–4	–255	–4	186	3
Percentage female of college major	964	14	1008	15	883	15	911	16
Institutional selectivity	39	1	N/A	N/A	N/A	N/A	N/A	N/A
Highest degree	–93	–1	–115	–2	5	0	22	0
Family formation								
Family formation (endowments effect)	9	0	99	1	–74	–1	–157	–3
Family formation (coefficients effect)	–	–	(–158)	(–2)	276	5	918	16
Work-related								
Hours worked per week	715	10	699	10	890	16	–	–
Occupation	1431	21	1311	20	979	17	–	–
Industry	495	7	373	6	342	6	–	–
Sector	611	9	501	8	546	10	–	–
Other work factors	140	2	130	2	102	2	–	–
Total income gap	6938	–	6670	–	5707	–	5707	–

3 Pure replication

My first set of analyses (replications 1A and 2A) aims to mimic as closely as possible the original model specifications and sample restrictions in order to reproduce the original results and to provide a reference for assessing the impact of corrections on the original design later on. An exact replication in the strict sense was not feasible because I could not obtain the original code. When preparing the analysis sample I therefore emulated the original article whenever possible and resorted to qualified guesses when

instructions therein were insufficiently detailed.[1] A replication package documents these decisions in detail and is publicly and permanently available at the Harvard Dataverse (Ochsenfeld 2016a). The resulting analysis sample is slightly smaller (N = 1924) than that of the original study (N = 1946). A comparison of descriptive statistics (see Table 3 in the appendix and Bobbitt-Zeher [2007, pp. 11, 18 f.]) indicates that the independent variables are distributed very similarly but not identically. Average annual incomes are slightly lower in my sample for both

---

[1] This also meant using the same reference categories as the original study, despite the entailed problem of picking reference categories arbitrarily (Kim 2013). Diverging from Bobbitt-Zeher (2007), I coded 'other race' to 'white' and do not differentiate between married or single parents and not between number of children because the formation of these categories results in extremely sparsely populated cells that cause major problems in estimating the Oaxaca-Blinder decomposition. I also conducted a pure replication with the original categories which resulted in estimates virtually identical to those reported here.
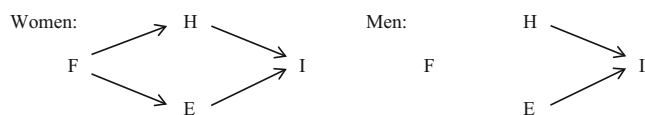
Fig. 1 Causal diagram for the effect of family formation on income for women and men. (*F* family formation, *I* income, *H* hours worked, *E* job experience. Based on Bobbitt-Zeher [2007, p. 5])

women (\$ 32,573) and men (\$ 39,243) compared to the original study (\$ 32,953 and \$ 39,891, respectively). The resulting unconditional gender income gap is thus slightly smaller in my sample (\$ 6670) than in the original study (\$ 6938).

The original analyses were conducted on the NELS:88 Restricted Use Data. Because researchers outside the U. S. cannot access these data, my replication is based on the NELS:88 Public Use Data which is identical to the Restricted Use Data except that it does not include students' SAT scores and a measure for the selectivity of the college attended. However, because these variables were shown to be entirely peripheral in the original analysis, it is nevertheless possible to very closely replicate the original results.

With regards to percentage of the income gap explained by the various variable groups in the sequential decomposition, my results almost exactly reproduce those of the original study, except for work variables where explanatory power is lower than in the original study (Table 1). I can only speculate that this may be due to potential differences in how I recoded certain industries and occupations to arrive at the categories reported in Bobbitt-Zeher (2007, p. 18 f.).

The estimates from the Oaxaca-Blinder decomposition (Table 2) are nearly identical to those reported in Bobbitt-Zeher (2007). Only for grades is this not the case, in all likelihood because this variable group encompasses 12th grade standardized test scores, 12th grade grades, and SAT scores in the original study whereas my analysis cannot include SAT scores.

If I were to draw conclusions about the importance of education-related factors relative to family formation and values based on the results of this replication exercise, these would echo Bobbitt-Zeher's. Family formation seems to explain none of the gender income gap; gender differences in the importance of '*having lots of money*' are almost irrelevant, but education-related factors – most notably segregation in college majors – explain a sizeable share (nearly a third) of women's lower incomes. In the following, I will argue that this conclusion needs to be revoked with respect to family formation and values.

4 Replication with corrections for endogenous sample restriction and misspecification

To explicate why I consider both the restriction Bobbitt-Zeher imposes on her sample and her model specifications to be at odds with her theoretical argument, I briefly summarize her argument regarding the potential effect of family formation on the gender income gap. She states that

> The effects of family formation, particularly marriage and parenthood and their impact on participation in paid labor, are implicated in gender income disparities. For example, net of other factors, such as education, women with children make 10 percent to 15 percent less than do women without children, and there is a 7 percent wage penalty for each child that a young woman has. [...] The same patterns do not hold for men; fathers experience no comparable wage penalty for their parental status. (Bobbitt-Zeher 2007, p. 4)

This argument suggests that gender *moderates* the effect of family formation on income: Whereas the effect is thought to be negative for women, it is thought to be non-existing for men. Furthermore, the author argues that labor force participation is a key mechanism that brings about the negative effect of motherhood on income:

> The impact of family formation on gender differences in earnings appears to operate through women's decreased labor force participation. Both length of job experience and part-time employment contribute to lower earnings. (Bobbitt-Zeher 2007, p. 5)

Fig. 1 illustrates these statements.

In her study, Bobbitt-Zeher chose to restrict her sample to persons working full time (≥35 h per week throughout the year). She provides no justification for this decision other than to "*avoid part-time and inconsistent workers from biasing the analysis*" (Bobbitt-Zeher 2007, p. 7). However, the reverse is likely to be true because she thereby conditions on hours worked which, according to her own theoretical argument, mediate part of the motherhood penalty (Fig. 1). In consequence, her estimates for the effect of family formation suffer from overcontrol bias (Elwert and Winship 2014).[2]

Fig. 2 suggests that Bobbitt-Zeher's decision to delete persons who work less than full-time year-round indeed disproportionately removes mothers from the sample and thus underreports the frequency of motherhood. Worse, this only retains mothers that experienced no or merely a weak effect of motherhood on hours worked: The difference in hours worked between mothers and fathers is significantly

---

[2] In addition, mothers remaining in her sample are likely selected on offered wages (Heckman 1974).

Fig. 2 Hours worked by mothers and fathers, separately for sample restricted to year-round full-time working persons (*lower row*) and sample including part-time and non-year-round working persons (*upper row*)
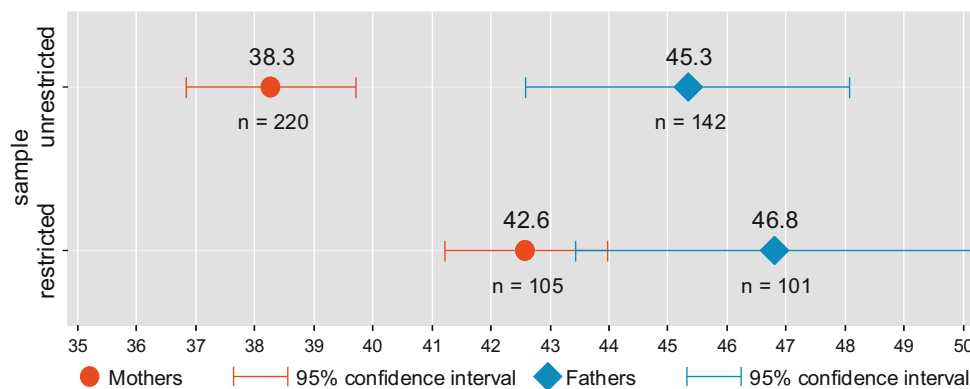
larger in the sample that includes part-time (≥10 h per week) and non-year-round workers than in the restricted sample (see Fig. 2).

The effect of values, too, is likely affected by overcontrol bias: Persons who did not find '*making lots of money*' important during high school are prone to work fewer hours and earn lower annual incomes later in life for that reason. *Hours worked* thus mediate the effect of values (i.e. importance of '*making lots of money*') on income. In this case Bobbitt-Zeher's decision to restrict the sample to persons working full-time year-round puts downward bias on her estimate for the explanatory power of this factor. Finding '*making lots of money*' unimportant in 12th grade is indeed weakly associated with not working full-time year-round 7 years later.[3]

To assess how much Bobbitt-Zeher's decision to restrict her sample to full-time year-round workers has biased the estimated explanatory power of family formation and values in her study, Table 1 (replication 1B) reports the results from a sequential decomposition that also includes persons who work less than year-round full-time and thus allows *hours worked* to bring about the effects of family formation and values. The explanatory power attributed to values in the sequential decomposition doubles from 8 to 16% – nearly half of what all education-related variables taken together contribute (34%).

The estimate for the contribution of family formation, however, continues to suggest the complete irrelevance of this factor (Table 1, replication 1B). This is, however, because Bobbitt-Zeher's models do not reflect her theoretical argument. Her statement that gender *moderates* the effect of family formation on income (see above) suggests a model that includes gender and the family formation variables (*parenthood* and *marriage*) as well as terms that interact them with gender (model 5b). However, Bobbitt-Zeher's models do not include the interaction terms (model 5).

Model 5: Income $= \beta_0 + \beta_1\text{female} + \dots + \beta_{k-1}\text{parenthood} + \beta_k\text{marriage} + \varepsilon$

Model 5b: Income $= \beta_0 + \beta_1\text{female} + \dots + \beta_{k-1}\text{parenthood} + \beta_k\text{marriage} + \beta_{k+1}\text{female}\times\text{parenthood} + \beta_{k+2}\text{female}\times\text{marriage} + \varepsilon$

Once I use model 5b instead of model 5 in the sequential decomposition on the sample that includes part-time workers, family formation adds 15 percentage points to the explanation of the gender income gap rather than nothing at all (Table 1, replication 1C). The work-related covariates in turn add much less explanatory power (13 instead of 23%) because part of their association with income stems from them mediating the motherhood penalty (Fig. 1) and therefore already gets picked up by model 5b (but not model 5).[4]

The same issue plays out somewhat differently in the Oaxaca-Blinder decomposition (Blinder 1973; Oaxaca 1973) that allows for a decomposition of the gender gap into membership, coefficients, and endowments components plus an interaction between coefficients and endowments (Fig. 3; Jones and Kelley 1984; Jann 2008).

The membership, coefficients and interaction components are often summarized into a single 'unexplained' or 'discriminatory' component. In the resulting two-fold decomposition the endowments component is conventionally referred to as the 'explained' or 'non-discriminatory' and the other components taken together as the 'unexplained' or 'discriminatory' component (Fig. 3). However, the interpretation of the sum of the membership, interaction and coefficients components as 'unexplained' is warranted if and only if indeed no theory of interest explains any part of these components (or as 'discriminatory' if indeed all components can entirely and unambiguously be interpreted in the light of discrimination theory). This condition usually

---

[3] 'Importance of having lots of money' in grade 12 is 0.1 points higher among the full-time, year-round working than among those who do not (t = 3.12).

[4] For the sake of brevity, I do not discuss here the decomposition based on model 5b conducted on the restricted sample which indicates a motherhood bonus due to endogenous selection bias.

Fig. 3 Oaxaca-Blinder decomposition

$$I_A - I_B = (\beta_{0A} - \beta_{0B}) + \sum_{j=1}^{k} \overline{X}_{jB}(\beta_{jA} - \beta_{jB}) + \sum_{j=1}^{k} (\overline{X}_{jA} - \overline{X}_{jB})(\beta_{jA} - \beta_{jB}) + \sum_{j=1}^{k} (\overline{X}_{jA} - \overline{X}_{jB})\beta_{jB}$$

$$= \text{membership} \quad + \quad \text{coefficients} \quad + \quad \text{interaction} \quad\quad + \quad \text{endowments}$$

$$= \qquad\qquad\qquad \text{'unexplained'} \qquad\qquad\qquad\qquad + \quad \text{'explained'}$$

holds in the most common application of the decomposition where human capital theory is tested against discrimination theory and wages are regressed solely on productivity-related characteristics (e. g. Braakmann 2013).

It does not hold in the study at hand, however, because Bobbitt-Zeher's statements concerning the gendered effect of family formation provide a strong theoretical ground for attributing the coefficients effects for the variables *parenthood* and *marriage* to family formation. Given Bobbitt-Zeher's above summarized argument (and given that she does not control for job experience), we do expect a direct effect of family formation on income that is more negative for women than for men. The coefficients effect regarding the *parenthood* and *marriage* variables should thus be treated as predicted by the family formation argument. In the original study, however, they were not. Instead, they remained unreported. The negligible *percentage of total gap explained* for *family formation* in the original study and in replications 2A and 2B (Table 2) merely refers to the endowment effect for family formation – the potentially higher incidence of parenthood and marriage among women compared with men. Table 2 reports both the endowments and coefficients effect for the family formation variables. For the unrestricted sample (Table 2, replication 2B), the coefficients effect of family formation explains 5% of the income gap.

This estimate, however, still does not reflect the full explanatory power of the family formation argument. To the degree that the gender difference in hours worked is an outcome of family formation, the endowment component for *hours worked* should be attributed to *family formation*.

Because virtually all other work variables (occupation, industry, sector, job training, job autonomy) can be expected to mediate the effects of education, values, and family formation, too, their inclusion in the Oaxaca-Blinder decomposition renders the decomposition results uninformative for its designed purpose, an assessment of the explanatory power of education-related variables relative to family formation. Table 2 (replication 2C) therefore shows results from a decomposition that excludes the work-related variables. These more meaningful results are in line with those from the previous sequential decomposition (Table 1, replication C) as they suggest that education, family formation, and values all independently explain sizeable shares of the income gap (Table 2, replication 2C).

Based on my preferred decomposition (Table 1, replication 1C), I conclude that values and family formation each explain approximately 15% of the gender income gap. Given that only few respondents (12%) have entered parenthood yet, I consider this to be a sizeable contribution of family formation.[5]

5 Conclusion

In a highly cited 2007 article, Bobbitt-Zeher (2007) assessed the roles of education and family formation for generating income inequality between young college-educated women and men. Her analysis produced results which led her to conclude that educational factors, in particular gender segregation in college majors, clearly dominate over family formation which she found to be entirely irrelevant for the explanation of economic gender inequality.

I argued here that Bobbitt-Zeher's conclusions concerning the effects of family formation and values must be revoked because her analysis misspecified and misallocated the hypothesized gendered effect of family formation in both relevant decompositions and, in addition, was conducted on a sample restricted in a way that negatively biases the estimated contributions of family formation and values. By way of replication I showed that when correcting for these shortcomings, the combined importance of family formation and values is comparable in magnitude to that of education-related factors. The results from another replication I conducted on comparable German data further corroborate my argument (Ochsenfeld 2016b).

Bobbitt-Zeher's original finding of a null effect for family formation can be interpreted to suggest that improvements to the work-family compatibility of workplaces or to family policies would not have the potential to reduce the gender income gap. My finding that family formation and values in sum explain almost as much as education sug-

---

[5] I conceive of this estimate as a lower bound given that inclusion in the part-time sample still requires minimal labor force participation (≥10 hrs. per week). I choose not to estimate models with a Heckman correction given that (1) this would further estrange my models from the original study and (2) the inherent endogeneity of income to hours worked renders the simultaneous estimation of income and labor force participation problematic.

gests the opposite. The analysis was conducted on a sample of young persons who only recently graduated from college and of whom only a small minority have children yet. Hence, the findings refer to a point in the life course when the importance of educational factors for gender inequality is at its maximum and the importance of family formation still close to its minimum. The more persons enter parenthood and the more years pass after graduation from college, the more will education's relative importance diminish (Braakmann 2013) and family formation's role increase. The mechanisms that bring about the motherhood penalty range from discrimination against mothers (Correll et al. 2007) to work organization (Goldin and Katz 2016), to gendered parenthood roles (Grunow et al. 2012), their interaction with family policy (Gangl and Ziefle 2009, 2015; Ochsenfeld 2012), and employment mismatch (Soerenson and Dahl 2016). These mechanisms deserve a prominent place both in our search for the causes of economic gender inequality and the design of solutions to it.

Appendix

Table 3 Descriptive statistics by gender and sample definition

Variable	Women, year-round full-time only (N = 1040)		Men, year-round full-time only (N = 884)		Women, full-time, part-time, year-round and non-year-round (N = 1833)		Men, full-time, part-time, year-round and non-year-round (N = 1457)	
	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
Income in 1999	32,573	479	39,243	659	28,420	378	34,127	603
Standardized Test Scores								
Reading (12th grade)	57.0	0.3	54.9	0.4	57.2	0.3	55.4	0.3
Math (12th grade)	57.3	0.3	59.1	0.4	57.3	0.3	59.0	0.3
Science (12th grade)	54.6	0.4	57.9	0.4	54.8	0.3	58.1	0.3
Grades								
English (12th grade)	4.55	0.07	5.46	0.11	4.60	0.07	5.48	0.08
Math (12th grade)	5.60	0.11	5.90	0.13	5.71	0.09	5.95	0.10

Table 3 Descriptive statistics by gender and sample definition (Continued)

Variable	Women, year-round full-time only (N = 1040)		Men, year-round full-time only (N = 884)		Women, full-time, part-time, year-round and non-year-round (N = 1833)		Men, full-time, part-time, year-round and non-year-round (N = 1457)	
	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
Science (12th grade)	5.32	0.09	5.52	0.12	5.35	0.08	5.53	0.09
Undergraduate GPA	3.27	0.02	3.11	0.02	3.29	0.01	3.14	0.02
College major								
Economic sciences	0.24	0.02	0.28	0.02	0.17	0.01	0.24	0.02
Math, natural science, or engineering	0.24	0.02	0.38	0.02	0.24	0.01	0.39	0.02
Social science or humanities	0.42	0.02	0.29	0.02	0.44	0.02	0.33	0.02
Education	0.10	0.01	0.04	0.01	0.15	0.01	0.04	0.01
Proportion female of major	0.60	0.01	0.48	0.01	0.61	0.01	0.49	0.01
Highest degree								
Bachelor's degree	0.92	0.01	0.95	0.01	0.88	0.01	0.91	0.01
Master's degree	0.08	0.01	0.05	0.01	0.11	0.01	0.07	0.01
Professional or doctoral degree	0.01	0.00	0.00	0.00	0.01	0.00	0.02	0.01
SES of family of origin	0.39	0.03	0.48	0.03	0.44	0.02	0.54	0.03
Race of Respondent								
White (or *other*)	0.84	0.01	0.83	0.02	0.84	0.01	0.83	0.02
Black	0.07	0.01	0.08	0.02	0.08	0.01	0.06	0.01
Latino	0.05	0.01	0.04	0.01	0.05	0.01	0.04	0.01
Asian American	0.04	0.01	0.05	0.01	0.04	0.00	0.06	0.01
Importance of having lots of money (12th grade)	2.12	0.02	2.35	0.03	2.08	0.02	2.31	0.02
Family characteristics								
Single	0.58	0.02	0.68	0.02	0.59	0.02	0.72	0.02
Married	0.40	0.02	0.31	0.02	0.39	0.02	0.27	0.02
Single Parent	0.03	0.01	0.04	0.01	0.02	0.00	0.03	0.01
Number of children	0.14	0.02	0.18	0.03	0.15	0.01	0.14	0.02
Number of hours worked, per week	44.4	0.3	47.2	0.4	41.6	0.3	43.8	0.4
Private sector	0.63	0.02	0.77	0.02	0.58	0.02	0.72	0.02
Industry								
Agriculture, forestry, fisheries, mining	0.01	0.00	0.02	0.01	0.01	0.00	0.02	0.00
Construction and allied	0.01	0.00	0.03	0.01	0.01	0.00	0.04	0.01
Manufacturing: Durable goods	0.06	0.01	0.09	0.01	0.04	0.01	0.08	0.01
Manufacturing: Nondurable goods	0.03	0.01	0.06	0.01	0.02	0.01	0.05	0.01
Utilities	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00
Wholesale distribution	0.00	0.00	0.02	0.01	0.01	0.00	0.02	0.01
Retail trades	0.10	0.01	0.08	0.02	0.09	0.01	0.09	0.01
Finance, insurance, real estate	0.12	0.01	0.17	0.02	0.10	0.01	0.14	0.01
Business, personal services	0.07	0.01	0.09	0.01	0.07	0.01	0.08	0.01
Entertainment, recreation	0.03	0.01	0.03	0.01	0.03	0.01	0.04	0.01
Professional services	0.19	0.02	0.14	0.02	0.17	0.01	0.15	0.01
Public administration, safety, military	0.05	0.02	0.06	0.02	0.03	0.01	0.06	0.01
Health care	0.16	0.02	0.04	0.01	0.18	0.01	0.06	0.01
Communications	0.06	0.01	0.07	0.01	0.06	0.01	0.08	0.01
Transportation	0.01	0.00	0.02	0.01	0.01	0.01	0.02	0.01
Hospitality	0.02	0.01	0.02	0.01	0.03	0.01	0.03	0.01
Education	0.08	0.01	0.04	0.01	0.11	0.01	0.05	0.01

Table 3 Descriptive statistics by gender and sample definition (Continued)

Variable	Women, year-round full-time only (N = 1040)		Men, year-round full-time only (N = 884)		Women, full-time, part-time, year-round and non-year-round (N = 1833)		Men, full-time, part-time, year-round and non-year-round (N = 1457)	
	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
Occupation								
Secretary, receptionist	0.03	0.01	0.01	0.00	0.03	0.01	0.01	0.00
Cashier, teller, clerk, data entry	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00
Other clerical	0.02	0.01	0.01	0.01	0.01	0.00	0.01	0.00
Personal service, cook, chef, baker	0.01	0.00	0.00	0.00	0.02	0.01	0.01	0.00
Farmer, forester, farm laborer, nonfarm laborer	0.00	0.00	0.03	0.01	0.01	0.00	0.04	0.01
Mechanic, repairer, service technician	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
Craftsman, skilled operative	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Protective service, criminal justice, military	0.03	0.01	0.06	0.01	0.02	0.00	0.06	0.01
Business and financial support services	0.07	0.01	0.05	0.01	0.06	0.01	0.04	0.01
Financial service professional	0.10	0.01	0.13	0.02	0.06	0.01	0.10	0.01
Sales, purchasing	0.10	0.01	0.09	0.01	0.09	0.01	0.09	0.01
Customer service	0.02	0.00	0.00	0.00	0.02	0.01	0.01	0.00
Legal support	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.01
Medical practice professional, services	0.03	0.01	0.01	0.00	0.03	0.01	0.01	0.00
Medical licensed professional	0.06	0.01	0.02	0.01	0.07	0.01	0.02	0.01
Educators (K-12 teachers)	0.11	0.01	0.04	0.01	0.17	0.01	0.05	0.01
Educators, instructors (non-K-12)	0.05	0.01	0.02	0.00	0.07	0.01	0.04	0.01
Human service professional	0.04	0.01	0.02	0.00	0.04	0.01	0.02	0.00
Engineer, architect, software engineer	0.03	0.01	0.11	0.01	0.02	0.00	0.09	0.01
Scientist, statistician professional	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.00
Research assistant, lab technician	0.01	0.01	0.01	0.00	0.02	0.00	0.02	0.01
Technical, professional worker	0.02	0.00	0.02	0.01	0.01	0.00	0.02	0.00
Computer systems, related professional	0.04	0.01	0.12	0.02	0.04	0.01	0.11	0.01
Computer programmer, other computer	0.02	0.01	0.02	0.00	0.01	0.00	0.02	0.00
Editor, writer, reporter	0.03	0.01	0.03	0.01	0.03	0.00	0.02	0.01
Performer, artist	0.00	0.00	0.01	0.01	0.02	0.00	0.02	0.00
Manager, executive	0.01	0.00	0.02	0.01	0.01	0.00	0.01	0.00
Manager, midlevel	0.04	0.01	0.05	0.01	0.03	0.01	0.05	0.01
Manager, supervisory, office	0.12	0.02	0.08	0.01	0.09	0.01	0.06	0.01
Health, recreational services	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
Job training	2.77	0.03	2.79	0.03	2.75	0.02	2.80	0.03
Job autonomy	0.74	0.02	0.78	0.02	0.74	0.02	0.74	0.02

Weighted estimates. Source: NELS:88

References

Blinder, A.S.: Wage discrimination. Reduced form and structural estimates. J Hum Resour **8**, 437–455 (1973)

Bobbitt-Zeher, D.: The gender income gap and the role of education. Sociol Educ **80**, 1–22 (2007)

Braakmann, N.: What determines wage inequality among young German university graduates? J Econ Stat **233**, 130–158 (2013)

Cha, Y., Weeden, K.: Overwork and the slow convergence in the gender gap in wages. Am Sociol Rev **79**, 457–484 (2014)

Correll, S., Benard, S., Paik, I.: Getting a job: Is there a motherhood penalty? Am J Sociol **112**, 1297–1339 (2007)

Elwert, F., Winship, C.: Endogenous selection bias: the problem of conditioning on a collider variable. Annu Rev Sociol **40**, 31–53 (2014)

Gangl, M., Ziefle, A.: Motherhood, Labor Force Behavior, and Women's Careers. An Empirical Assessment of the Wage Penalty for Motherhood in Britain, Germany, and the United States. Demography **46**, 341–369 (2009)

Gangl, M., Ziefle, A.: The making of a good woman: extended parental leave entitlements and mothers' work commitment in Germany. Am J Sociol **121**, 511–563 (2015)

Gerber, T.P., Cheung, S.Y.: Horizontal stratification in postsecondary education. Forms, explanations and implications. Annu Rev Sociol **34**, 299–318 (2008)

Goldin, C.: A grand gender convergence: its last chapter. Am Econ Rev **104**, 1091–1119 (2014)

Goldin, C., Katz, L.: A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation. J Labor Econ **34**, 705–746 (2016)

Grunow, D., Schulz, F., Blossfeld, H.-P.: What determines change in the division of housework over the course of marriage? Int Sociol **27**, 289–307 (2012)

Heckman, J.: Shadow prices, market wages, and labor supply. Econometrica **42**, 679–694 (1974)

Jann, B.: Making regression tables simplified. Stata J **7**, 227–244 (2007)

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

Jones, F.L., Kelley, J.: Decomposing differences between groups. A cautionary note on measuring discrimination. Sociol Methods Res **12**, 323–343 (1984)

Kim, C.: Detailed wage decompositions: revisiting the identification problem. Sociol Methodol **43**, 346–363 (2013)

Leuze, K., Strauß, S.: Lohnungleichheiten zwischen Akademikernnen und Akademikern: Der Einfluss von fachlicher Spezialisierung, frauendominierten Fächern und beruflicher Segregation. Z Soziol **38**, 262–281 (2009)

Marini, M.M., Fan, P.-L.: The gender gap in earnings at career entry. Am Sociol Rev **62**, 588–604 (1997)

Morgan, L.A., Arthur, M.M.: Methodological considerations in estimating the gender pay gap for employed professionals. Sociol Methods Res **33**, 383–403 (2005)

Oaxaca, R.: Male-female wage differentials in urban labor markets. Int Econ Rev (Philadelphia) **14**, 693–709 (1973)

Ochsenfeld, F.: Gläserne Decke oder goldener Käfig: Scheitert der Aufstieg von Frauen in erste Managementpositionen an betrieblicher Diskriminierung oder an familiären Pflichten? Kolner Z Soz Sozpsychol **64**, 507–534 (2012)

Ochsenfeld, F.: Why do women's fields of study pay less? A test of devaluation, human capital, and gender role theory. Eur Sociol Rev **30**, 536–548 (2014)

Ochsenfeld, F.: Replication data for: "The gender income gap and the role of family formation revisited: a replication of Bobbitt-Zeher (2007)". Harvard Dataverse, V1. (2016a) doi:10.7910/dvn/uemos1

Ochsenfeld, F.: The gender income gap and the roles of education and family formation: a scientific replication of Bobbitt-Zeher (2007). SocArxiv, V1. (2016b). https://osf.io/preprints/socarxiv/xf8z5

Petersen, T.: The earnings function in sociological studies of earnings inequality. Functional form and hours worked. Res Soc Stratif Mobil **8**, 221–230 (1989)

Soerenson, O., Dahl, M.: Geography, joint choices, and the reproduction of gender inequality. Am Sociol Rev **81**, 900–920 (2016)

Waldfogel, J.: The effect of children on women's wages. Am Sociol Rev **62**, 209–217 (1997)

Fabian Ochsenfeld is a sociologist at Goethe University, Frankfurt am Main.