

**The Gender Income Gap and the Roles of Education and Family Formation:
A Scientific Replication of Bobbitt-Zeher (2007)**

Fabian Ochsenfeld

Goethe University
Institute for Sociology
Theodor-W.-Adorno-Platz 6
60323 Frankfurt, Germany
Phone: +49.69.789.36631
E-mail: ochsenfeld@soz.uni-frankfurt.de

16 December 2015

Abstract: This article reports the results of a replication of Bobbitt-Zeher's 2007 *Sociology of Education* article "*The Gender Income Gap and the Role of Education*" based on comparable data from Germany. Models that emulate the original specifications successfully replicate the results. However, models that instead adhere to Bobbitt-Zeher's theory concerning the gendered effect of family formation call her key finding that "*family formation has virtually no effect on the income gap*" into question.

(72 words)

Key words: gender gap, motherhood penalty, horizontal sex segregation, replication, decomposition methods.

Word count: 3,718 words + 450 words (references), 3 tables, 1 figure. Appendix: 5 tables.

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 in the making of the gender income 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 gender gap, a 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: 13). 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 Bobbitt-Zeher’s results by way of replication. This is an endeavor worth undertaking given that (1) her findings have important policy implications, as they suggest that improvements in work-family compatibility would do nothing to reduce economic gender inequality, (2) the study ranks among the journal’s five most cited contributions published in 2007¹, and (3) other highly-cited studies have reached similar conclusions based on similar designs and modeling strategies (Marini and Fan 1997; Leuze and Strauß 2009). I first conduct a scientific (rather than pure) replication: using a different sample drawn from a different population and fitting models to these data that are almost identical to those used in the original article (Hamermesh 2007). I then report estimates from another set of models that no longer seek to emulate the statistical models of the original study but instead adhere to the theoretical reasoning regarding the gendered effect of family formation on income as explicated in the original article.

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 2005, Gerber and Cheung 2008, Ochsensfeld 2014).

¹ Rank 4 according to Web of Science, rank 3 according to Google Scholar (searched December 10, 2015).

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.: 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 U.S. data from the *National Education Longitudinal Study of 1988* which she restricts to persons with 4-year college degrees who work at least 35 hours per week throughout 1999. She derives her main findings from three regression-based income decompositions:

The purpose of the first decomposition is to assess which educational factors contribute most to an understanding of the gender income gap. To this end, the author fits seven non-nested models that contain different education-related variables plus a gender dummy and compares the coefficient for gender in these models with the coefficient for gender in two baseline models (table 1). One baseline model includes gender as the only covariate, the other baseline model adds race and parents’ socio-economic status. Bobbitt-Zeher’s main finding from this analysis is that sex segregation in fields of study can explain 39 percent of the gender income gap (model 4) and thus much more than any other education-related factor.

The second decomposition (table 2) starts from the baseline model with gender as the only predictor and sequentially adds variable groups that measure respondents’ background, values, education, family status, and work characteristics. 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 race, socio-economic background and values in total explain only 11 percent of the gender income gap, whereas adding education-related covariates explains an additional 33 percentage points of the gender 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 25 additional percentage points.

The third, Oaxaca-Blinder type decomposition (table 3) fits income regressions separately for women and men using 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 second decomposition (table 2) Bobbitt-Zeher's Oaxaca-Blinder decomposition attributes less explanatory power to education-related factors and more to work-related variables.² As in the second decomposition, family formation seemingly plays no role in the generation of women's economic disadvantage whatsoever.

Based on the results from the second and third 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 "*family formation matters not much at all.*" (Bobbitt-Zeher 2007: 15).

3. Scientific replication

In the following, I report the results from my attempt to replicate Bobbitt-Zeher's study with the German *HIS Graduate Panel Study 1997* (Fabian and Minks, 2006). This two-wave survey is suitable for a replication study, first, because it includes items which allow me to construct variables that closely match those used in the original analysis (table A1) and, second, because it provides this information for college educated persons only a few years after they entered the labor market, as is the case for the original study. The *HIS Graduate Panel Study* surveyed a nationwide sample of all individuals who received their first college degree in Germany in the academic year 1997. Graduates were sampled using a stratified cluster design and interviewed by questionnaire briefly after award of their degree in 1997/98 and for a second time in 2002/03. Most of the information I use in the analysis comes from the 2002/2003 wave. The data used in Bobbitt-Zeher's study is from 1999. At the time of the 2002/2003 wave, respondents in my sample on average graduated from high school approximately 12 years ago, whereas they are reported to have graduated from high school approximately eight years ago in the original study (Bobbitt-Zeher 2007: 7). However, students in Germany tend to study longer than in the U.S. (on average 5.4 years in my sample)

² 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.

and men in the cohort at hand were obliged to serve either in the military (12-15 months) or in civil service (15-20 months). Hence, the difference in average labor market experience between the samples analyzed here and in the original study will be significantly less than four years. To account for the stratified sampling design, I apply sampling weights as provided by HIS in all analyses. The response rate of the survey was 27 percent in the first wave. Panel attrition between the first and second wave was 35 percent, but uncorrelated with observables (Fabian and Minks, 2006: 10). Researchers can obtain the dataset and study documentation (in German) from the Leibniz Institute for Social Research under study number ZA4272. The Stata code to replicate my analysis is permanently available online (Ochsenfeld 2015).

The following, first set of analyses (replication A) is designed to mimic as closely as possible the sample restrictions and model specifications reported in the original article. I therefore restrict the sample to full-time workers (≥ 35 hours per week) with complete income data ($N=3,216$) and impute missing values by chained equations (for $N=522$, see Ochsenfeld 2015 for details). A list of all variables, including brief descriptions that allow for a comparison with the original study, and descriptive statistics can be found in the appendix (tables A1 and A2). Women's annual income in my sample is 17 percent less than men's, the gender income gap thus exactly as large as in Bobbitt-Zeher's data.

The results from the first decomposition closely replicate those of the original study (table 2): Together with background, sex segregation in college majors (i.e. percentage female) explains 43 percent of the gender income gap, compared with 39 percent in the original study. When measured instead by broad fields of study, gender differences in major selection explain somewhat less of the income gap. This also mirrors the original study. None of the other education-related characteristics mediate the unconditional gender effect from the two baseline models, so that the full model (number 9) explains 42 percent of the gender gap, somewhat more than in Bobbitt-Zeher (2007).

The second decomposition serves to scrutinize the role of educational factors in comparison with non-educational factors and thus addresses the key research question of the original study. Here, my results map almost exactly onto those of the original study (table 2, replication A). Differences in background can explain very little of the income gap in the original study. Including the degree to which respondents personally value making money reduces the gap by eight percentage points in my replication compared with the second baseline model (model number 2). In the original study, this aspect added seven percentage

points. Further adding the education-related variables allows my model to explain 41 percent of the income gap which corresponds closely with the original value (44 percent). Next, allowing for main effects of the family formation variables does not improve the model's ability to explain the gap at all, which replicates the result reported in Bobbitt-Zeher (2007). The work-related variables add slightly less to the explanation than in the original study, increasing the model's total explanatory power for the gender gap to 63 percent, compared with 69 percent.

The Oaxaca-Blinder decomposition continues to by and large emulate the results of the original study with regards to background, importance of having lots of money, institutional selectivity, highest degree, family formation, hours worked, sector and other work factors (table 3, replication A). As those from the previous, sequential decomposition, and those from the original study, the estimates suggest that horizontal segregation in college major and gender differences in work-related characteristics explain sizeable shares of the income gap, whereas all other factors contribute at best modestly to an understanding of gender differences in income. This is also true with respect to family formation: As in the original study, gender differences in the frequency of being a parent or being married are entirely irrelevant factors.

Taken together, the work-related variables explain about as much as in the original study; however, their individual contributions are distributed differently: The fact that women tend to be employed in less well-paying industries and sectors explains more in my study than in Bobbitt-Zeher's and the reverse is true with respect to occupations.

Except for this detail in attribution of explanatory power within the group of the (highly intercorrelated) work characteristics, my replication of Bobbitt-Zeher's 2007 article with similar German data was successful. In all three decompositions and even for individual variables, my estimates are close to those reported in the original study by social science standards. Against the background of recent studies that gave reason to doubt the replicability of much published work in the behavioral sciences (Open Science Collaboration 2015) this may be considered a noteworthy result.

However, I caution readers against inferring from the replicability of the original study to the validity of its main conclusion, namely that education-related factors, horizontal segregation in specific, would dominate over the impact of family formation that appears not to exist according to both the original study and its scientific replication. In the following section, I will argue that this finding is most likely an artifact of the misspecification of the effect of

family formation to then scrutinize whether the finding of a non-effect of family formation is replicable with a specification that corresponds to the original article’s theoretical argument rather than with its model specifications.

4. Replication with corrections for misspecification and endogenous sample selection

Regarding the potential effect of family formation on the gender income gap, Bobbitt-Zeher argues that

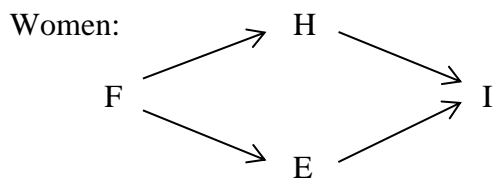
“The effects of family formation, particularly marriage and parenthood and their impact on participation in paid labor, are implicated in gender income disparities. For example, net of other factors, such as education, women with children make 10 percent to 15 percent less than do women without children, 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 2007:4)

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 2007:5)

Figure 1 illustrates this line of reasoning.

Figure 1: Directed acyclic graphs (Elwert, 2013) for the effect of family formation on income for women and men.



Notes: F: family formation; I: income; H: hours worked, E: job experience. Based on Bobbitt-Zeher (2007: 5).

Table 1: Regression coefficients for female and percentage of the gender income gap explained with alternate models, including education

Model number	Model description	Bobbitt-Zeher N=1,946		Replication A: Scientific replication N=3,216		
		Income gap (b_{female})	Percentage of gap explained	Income gap (b_{female})	s.e.	Percentage of gap explained
1	Female	-6,938	-	-7,606	(487)	-
2	Female, background	-6,643	4.2	-7,645	(490)	-0.5
3	Female, background, field of college major	-5,418	21.9	-5,269	(517)	30.7
4	Female, background, percentage female of college major	-4,244	38.8	-4,363	(580)	42.6
5	Female, background, grades at college entrance	-6,247	10.0	-7,734	(483)	-1.7
6	Female, background, college grades	-7,215	-4.0	-7,636	(487)	-0.4
7	Female, background, highest degree	-6,815	1.8	-7,547	(490)	0.8
8	Female, background, college selectivity	-6,634	4.4	-7,763	(494)	-2.1
9	Female, background, percentage female of college major, grades at college entrance, highest degree, college selectivity	-4,436	36.1	-4,396	(574)	42.2

Sources: Bobbitt-Zeher (2007:12); HIS graduate panel study, own (weighted) estimates.

Table 2: 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		Replication A: Scientific replication			Replication B: replication with correction for misspecification of family formation, full-time workers only			Replication C: replication with correction for misspecification of family formation, full-time and part-time workers		
		N=1,946		N=3,216			N=3,216			N=3,446		
		Income gap (b _{female})	% of gap explained	Income gap (b _{female})	s.e.	% of gap explained	Income gap (b _{female})	s.e.	% of gap explained	Income gap (b _{female})	s.e.	% of gap explained
1	Female	-6,938	-	-7,606	(487)	-	-7,606	(487)	-	-9,633	(491)	-
2	Female and background	-6,643	4.2	-7,645	(490)	-0.5	-7,645	(490)	-0.5	-9,619	(494)	0.1
3	Female, background, and values	-6,166	11.1	-7,005	(487)	7.9	-7,005	(487)	7.9	-8,864	(493)	8.0
4	Female, background, values, and education	-3,903	43.7	-4,127	(563)	45.7	-4,127	(563)	45.7	-5,695	(572)	40.9
5	Female, background, values, education, and family formation (not interacted with gender)	-3,854	44.4	-4,079	(568)	46.4	-	-	-	-	-	-
5b	Female, background, values, education, and family formation (interacted with gender)	-	-	-	-	-	-3,104	(632)	59.2	-3,136	(633)	67.4
6	Female, background, values, education, family formation, and work	-2,132	69.3	-2,808	(493)	63.1	-2,273	(550)	70.1	-2,178	(541)	77.4

Sources: Bobbitt-Zeher (2007:13); HIS graduate panel study, own (weighted) estimates.

Table 3: Regression decompositions showing contributions of background, values, education, family formation, and work characteristics to the gender income gap.

Variable (endowments effects if not stated otherwise)	Bobbitt-Zeher		Replication A: scientific replication			Replication C: replication with full-time and part-time workers			Replication D: replication with full-time and part-time workers		
	N=1,946		N=3,216			N=3,446			N=3,446		
	Diff. in income explained	% of total gap explained	Diff. in income explained	s.e.	% of total gap explained	Diff. in income explained	s.e.	% of total gap explained	Diff. in income explained	s.e.	% of total gap explained
Background	78	1.1	-30	(47)	-0.4	-12	(46)	-0.1	-3	(56)	0.0
Importance of having lots of money	383	5.5	308	(72)	4.0	366	(76)	3.8	763	(117)	7.9
<i>Education-related</i>											
Grades	330	4.8	-60	(57)	-0.8	-8	(49)	-0.1	-27	(59)	-0.3
Percentage female of college major	964	13.9	1,465	(296)	19.3	1,472	(295)	15.3	3,620	(284)	37.6
Institutional selectivity	39	0.6	-392	(99)	-5.2	-363	(91)	-3.8	-454	(104)	-4.7
Highest degree	-93	-1.3	26	(23)	0.3	25	(24)	0.3	68	(36)	0.7
<i>Family formation</i>											
Family formation (endowments effect)	9	0.1	-85	(140)	-1.1	-110	(73)	-1.1	-483	(102)	-5.0
Family formation (coefficients effect)	-	-	695	(377)	9.1	1,225	(365)	12.7	2,999	(409)	31.1
<i>Work-related</i>											
Hours worked per week	715	10.3	356	(89)	4.7	1,516	(174)	15.7	-	-	-
Occupation	1,431	20.6	509	(309)	6.7	445	(297)	4.6	-	-	-
Industry	495	7.1	1,950	(342)	25.6	2,170	(343)	22.5	-	-	-
Sector	611	8.8	343	(180)	4.5	273	(172)	2.8	-	-	-
Other work factors	140	2.0	155	(74)	2.0	151	(75)	1.6	-	-	-
Total income gap	6,938		7,606	(483)	-	9,633	(486)	-	9,633	(489)	-

Sources: Bobbitt-Zeher (2007:13); HIS graduate panel study, own (weighted) estimates. Male and female coefficients weighted equally (.5).

This theoretical reasoning has different implications for the sequential and the Oaxaca-Blinder decompositions. I will discuss these in turn, beginning with the sequential decomposition (table 2) that tracks the change of the slope coefficient for *female* as additional variable groups are successively added to the model. β_{female} diminishes if at least one added variable a) has a positive effect on income *and* the mean value for that variable is lower for women than for men (mediation) or b) has a negative effect on income *and* the mean value for that variable is higher for women than for men (mediation) or c) has an effect on income that is either more negative for women than for men *or* more positive for men than for women (moderation). Mediation is captured when adding the variable to the model; moderation is captured if the model additionally allows for effect heterogeneity through an interaction with gender.

Bobbitt-Zeher argues that gender *moderates* the effect of family formation on income (figure 1). This would suggest a model that includes gender and the family formation variables, parenthood and marriage, as well as terms that interact gender with parenthood and marriage (model 5b). However, Bobbitt-Zeher's models do not include the interaction terms (model 5).

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

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

The absence of change in β_{female} between models 4 and 5 in the original study and the above scientific replication (replication A) thus merely implies that college-educated women and men either do not differ in the frequency of family formation, or that family formation on average has no effect on income. Model 5b, in contrast, additionally includes interaction effects and thus corresponds to the argument made in the original article (Bobbitt-Zeher 2007: 5).

Once I conduct the sequential decomposition with model 5b instead of model 5, family formation adds 14 percentage points to the explanation of the gender gap rather than nothing at all (table 2, replication B). The work-related covariates in turn add much less explanatory power because part of their correlation with income stems from them mediating the motherhood penalty (H in figure 1), and therefore already gets picked up by model 5b (but not model 5).

Moreover, Bobbitt-Zeher chose to restrict her sample to persons working full time (≥ 35 hours per week) in order to “*avoid part-time [...] workers from biasing the analysis*” (Bobbitt-Zeher 2007: 7). However, the reverse is likely to be true because she thereby conditions on hours worked which, according to her own theoretical argument, mediates part of the motherhood penalty (figure 1). Table 2 (replication C) reports results from a sequential decomposition that also includes part-time workers and thus partly removes this source of overcontrol bias (Elwert and Winship, 2014).³ Doing so completely reverses the initial conclusion about the negligibility of family formation: Alleviated from overcontrol bias, the gendered effect of family formation now independently explains 26 percent (67 - 41) of the unconditional gender income gap which is only slightly less than the combined explanatory power of the education-related covariates (33 percent).

The same issues play 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 (Jones and Kelley 1984; Jann 2008):

$$\begin{aligned}
 I_A - I_B &= (\beta_{0A} - \beta_{0B}) + \sum_{j=1}^k \bar{X}_{jB} (\beta_{jA} - \beta_{jB}) + \sum_{j=1}^k (\bar{X}_{jA} - \bar{X}_{jB}) (\beta_{jA} - \beta_{jB}) + \sum_{j=1}^k (\bar{X}_{jA} - \bar{X}_{jB}) \beta_{jB} \\
 &= \underbrace{\text{membership} + \text{coefficients} + \text{interaction}}_{\text{'unexplained'}} + \underbrace{\text{endowments}}_{\text{'explained'}} \\
 &= \text{'unexplained'} + \text{'explained'}
 \end{aligned}$$

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. However, this summary of the membership, interaction and coefficients components is warranted if and only if there are no theoretical grounds to interpret *any* of the endowments effects as discrimination *and* the interpretation of *all* coefficients, interaction and membership effects as discrimination is indeed theoretically justified. These conditions usually hold in the most common application of the decomposition

³ The inclusion of part-time workers does not remove this bias entirely because the sample remains restricted to employed persons. Tables A4 and A5 show results for a sample that also includes the non-employed.

where human capital theory is tested against discrimination theory and wages are regressed solely on productivity-related characteristics (e.g. Braakmann 2013).

However, they do not hold in the original study and in my scientific replication because Bobbitt-Zeher provides a theoretical argument that suggests an attribution of the coefficients effects for the variables *parenthood* and *marriage* to family formation rather than to ‘unexplained’, given that job experience is not controlled for and part of the motherhood penalty must therefore be conceived of as a direct effect on income.

The negligible *percentage of total gap explained for family formation* in the original study (0.1 percent, see table 3 ‘Bobbitt-Zeher’) and in my scientific replication (-1.2 percent, see table 3, replication A) thus merely refers to the endowment effect - the potentially higher or lower incidence of parenthood and marriage among women compared with men. It does *not* refer to the gender-specific consequences of family formation emphasized by Bobbitt-Zeher’s theoretical argument. These are accounted for in the coefficients effects of family formation that are not reported in Bobbitt-Zeher’s article.

Furthermore, *proportion of the gender gap explained by family formation* suffers from overcontrol bias in the original study and my exact replication due to the partial misattribution of the endowments effect for *hours worked*. To the degree that the gender difference in hours worked is a consequence of family formation, the endowment component for *hours worked* ought to be attributed to the variable group *family formation*. Indeed, most of the gender difference in hours worked seems to be due to mothers reducing employment (Table A3). Because all other work variables (occupation, industry, sector, job training, job autonomy) can be expected to strongly mediate the effect of field of study on income, using these characteristics in the Oaxaca-Blinder decomposition renders its results generally uninformative for its designed purpose, an assessment of the explanatory power of education-related variables relative to that of family formation. A Oaxaca-Blinder composition without the intermediate work-related variables is thus much better attuned to this purpose and, based on my replication data, suggests that both education and family formation independently explain sizeable shares of the income gap (table 4, replication D), thus echoing the results from the previous sequential decomposition (table 3, replication C).

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 that was based on data from the 1988 *National Educational Longitudinal Study* 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.

In this article, I first reported results from a scientific replication of Bobbitt-Zeher's study I conducted based on the German *HIS Graduate Panel Study*. These results successfully reproduce those of the original study.

I then argued that the original study misspecified and misallocated the hypothesized gendered effect of family formation in both relevant decompositions and, in addition, was conducted with a sample restricted to full-time employed earners which negatively biases the estimated contribution of family formation. When correcting for these shortcomings in my replication, results suggest that the importance of family formation is comparable in magnitude to that of sex segregation in fields of study.

The motherhood penalty has been shown to be larger in Germany than in the U.S. (Gangl and Ziefle 2009). Therefore, I do not claim that my estimates for the importance of family formation can be directly extrapolated to the population from which the sample used in Bobbitt-Zeher (2007) was drawn. However, I do claim that the original study severely underestimates the importance of family formation for its reference population.

Acknowledgements

An earlier version of this article was presented at the Research Seminar for Empirical Social Research at University of Konstanz. I thank all participants as well as Jan Brülle for helpful comments. The HIS Graduate Panel Study data have kindly been provided by the Leibniz Institute for Social Research, Cologne.

References

- Blinder, A. S. (1973). Wage Discrimination. Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8, 437-455.
- Bobbitt-Zeher, D. (2007). The Gender Income Gap and the Role of Education. *Sociology of Education*, 80, 1-22.
- Braakmann, N. (2013). What Determines Wage Inequality Among Young German University Graduates? *Journal of Economics and Statistics*, 233, 130-158.
- Cha, Y., & Weeden, K. (2014). Overwork and the Slow Convergence in the Gender Gap in Wages. *American Sociological Review*, 79, 457-484.
- Elwert, F. (2013). Graphical Causal Models. In S. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 245-273). Dordrecht: Springer.
- Elwert, F., & Winship, C. (2014). Endogeneous Selection Bias: The Problem of Conditioning on a Collider Variable. *Annual Review of Sociology*, 40, 31-53.
- Fabian, G., & Minks, K.-H. (2006). Dokumentation des Scientific Use Files "HIS Absolventenpanel 1997". Hanover: HIS.
- Gangl, M., & Ziefle, A. (2009). 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.
- Gerber, T. P., & Cheung, S. Y. (2008). Horizontal Stratification in Postsecondary Education. Forms, Explanations and Implications. *Annual Review of Sociology*, 34, 299-318.
- Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104, 1091-1119.
- Hamermesh, D. (2007). Replication in Economics. *Canadian Journal of Economics*, 40, 715-733.
- Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 8, 453-479.
- Jones, F. L., & Kelley, J. (1984). Decomposing Differences between Groups. A Cautionary Note on Measuring Discrimination. *Sociological Methods & Research*, 12, 323-343.
- Leuze, K., & Strauß, S. (2009). Lohnungleichheiten zwischen Akademikerinnen und Akademikern: Der Einfluss von fachlicher Spezialisierung, frauendominierten Fächern und beruflicher Segregation. *Zeitschrift für Soziologie*, 38, 262-281.

- Marini, M. M., & Fan, P.-L. (1997). The Gender Gap in Earnings at Career Entry. *American Sociological Review*, 62, 588-604.
- Morgan, L. A., & Arthur, M. M. (2005). Methodological Considerations in Estimating the Gender Pay Gap for Employed Professionals. *Sociological Methods & Research*, 33, 383-403.
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14, 693-709.
- Ochsenfeld, F. (2014). Why Do Women's Fields of Study Pay Less? A Test of Devaluation, Human Capital, and Gender Role Theory. *European Sociological Review*, 30, 536-548.
- Ochsenfeld, F. (2015). Replication Data for: The Gender Income Gap and the Roles of Education and Family Formation: A Replication of Bobbitt-Zeher (2007), <http://dx.doi.org/10.7910/DVN/D80AZT>, Harvard Dataverse, V1
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349, doi: 10.1126/science.aac4716
- Petersen, T. (1989). The Earnings Function in Sociological Studies of Earnings Inequality. Functional Form and Hours Worked. *Research in Social Stratification and Mobility*, 8, 221-230.
- Waldfogel, J. (1997). The Effect of Children on Women's Wages. *American Sociological Review*, 62, 209-217.

Appendix

Table A1: Operationalization of constructs in the original and the replication study.

Construct	Measurement	
	Bobbitt-Zeher	Replication
Annual income	Annual income (\$)	12 * monthly income (gross, €)
Standardized test scores	SAT math score, SAT verbal score, ACT scores, undergraduate GPA.	Abitur grade (z-standardized), grade on college leaving certificate (z-standardized by major).
College major	4 categories: business, math/natural science/engineering, social science or humanities, education; Percentage female of major (continuous).	7 categories: economic science, math/natural science/engineering, social science or humanities, education, medical science, art, law; Percentage female of major (continuous).
Highest degree	3 categories: Bachelor's degree, Master's degree, professional or doctoral degree.	2 categories: Master's-equivalent degree (Diplom/Magister), doctoral degree.
Institutional selectivity	5 categories: Highly-selective, selective, nonselective, open-door, not ratable.	2 categories: full-university, university of applied sciences.
Marital status	3 categories: married, divorced, single.	2 categories: married, not married.
Parenthood	Number of children; 2 categories: single-parent, not a single parent.	2 categories: parent, not a parent.
Hours worked	Number of hours worked per typical week	Number of hours worked per week.
Sector	2 categories: For profit, Not for profit/government/military.	2 categories: public sector, private sector.
Industry	17 categories.	15 categories.
Occupation	31 categories.	33 categories, merged to 10 categories with $N \geq 30$ each.

Job training	2 categories: received job training in previous 12 months, received no job training in previous 12 months.	2 categories: received job training since graduation from college, received no job training since graduation from college.
Job autonomy	Level of job autonomy on the job as perceived by respondent [1;4]	Level of job autonomy on the job as perceived by respondent [0;3] Measure constructed from four items ($\alpha=0.58$): (1) own initiative is valued in my job, (2) possibility to take own financial decision in my job, (3) my work is mostly pre-defined, (4) I can organize work myself in my job.
Job values	Importance of having lots of money [1;3]	Goal for the future: making lots of money [0;4].
Race	5 categories: white, black, Latino, Asian American, other.	N/A
SES of family of origin	Composite score constructed from information on parents' education, income, and occupation.	Composite score constructed from information on mother's and father's education (college degree or not) and father's occupational status (worker or not) [-1;1].

Table A2: Descriptive statistics by gender and sample definition.

Variable	Women, full-time only (N=1,180)		Men, full-time only (N=2,036)		Women, full-time and part- time (N=1,370)		Men, full-time and part-time (N=2,076)	
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
<i>Income per year</i>	37,436	(379)	45,042	(307)	35,061	(382)	44,693	(308)
<i>Grades (z-standardized)</i>								
College entrance	.10	(.03)	.01	(.02)	.08	(.03)	.01	(.02)
College degree	.09	(.03)	.09	(.02)	.06	(.03)	.09	(.02)
<i>College major</i>								
Economic sciences	.154	(.011)	.148	(.009)	.146	(.010)	.147	(.008)
Math, natural science, or engineering	.230	(.013)	.628	(.012)	.216	(.012)	.623	(.012)
Social science or humanities	.141	(.011)	.047	(.005)	.149	(.011)	.050	(.005)
Education	.232	(.013)	.052	(.005)	.243	(.012)	.054	(.005)
Medical science	.123	(.010)	.053	(.006)	.128	(.010)	.052	(.006)
Art	.044	(.006)	.015	(.004)	.046	(.006)	.016	(.004)
Law	.076	(.008)	.058	(.006)	.072	(.007)	.058	(.006)
Proportion female of major	.566	(.006)	.305	(.005)	.574	(.006)	.308	(.005)
<i>Highest degree</i>								
Master's degree	.892	(.010)	.868	(.008)	.898	(.009)	.868	(.008)
Doctoral degree	.108	(.010)	.132	(.008)	.102	(.009)	.132	(.008)
<i>Institutional selectivity</i>								
Univ. of applied sciences	.205	(.013)	.381	(.012)	.205	(.012)	.377	(.012)
<i>SES of family of origin</i>	.65	(.03)	.43	(.02)	.66	(.03)	.43	(.02)
<i>Importance of having lots of money</i>	2.31	(.03)	2.57	(.02)	2.28	(.03)	2.56	(.02)
<i>Family characteristics</i>								
Married	.305	(.014)	.428	(.012)	.348	(.014)	.428	(.012)
Parent	.106	(.010)	.307	(.011)	.191	(.011)	.308	(.011)
<i>Number of hours worked, per week</i>	45.7	(.2)	47.1	(.2)	42.6	(.3)	46.7	(.2)

<i>Private sector</i>	.510	(.015)	.747	(.010)	.499	(.014)	.742	(.010)
<i>Industry</i>								
Agriculture, forestry, horticulture	.005	(.003)	.008	(.004)	.005	(.002)	.008	(.004)
Construction	.008	(.003)	.026	(.004)	.007	(.002)	.026	(.004)
Manufacturing	.111	(.010)	.308	(.011)	.101	(.009)	.304	(.011)
Utilities	.017	(.004)	.021	(.003)	.015	(.003)	.021	(.003)
Retail and wholesale	.022	(.004)	.020	(.003)	.021	(.004)	.019	(.003)
Finance, insurance, real estate	.054	(.007)	.053	(.005)	.051	(.006)	.053	(.005)
Business, personal services	.044	(.006)	.051	(.005)	.045	(.006)	.052	(.005)
Entertainment, recreation	.057	(.007)	.021	(.004)	.060	(.007)	.022	(.004)
Professional services	.065	(.008)	.165	(.009)	.060	(.007)	.164	(.009)
Public administration	.061	(.007)	.046	(.005)	.059	(.007)	.047	(.005)
Health care, social services	.163	(.012)	.074	(.007)	.174	(.011)	.073	(.007)
Communications	.012	(.003)	.018	(.003)	.011	(.003)	.018	(.003)
Transportation	.010	(.003)	.011	(.002)	.010	(.003)	.011	(.002)
Education	.294	(.014)	.121	(.008)	.306	(.013)	.125	(.008)
Other services	.077	(.008)	.056	(.005)	.073	(.007)	.058	(.005)
<i>Occupation</i>								
Agriculture, forestry, horticulture occupations	.010	(.003)	.011	(.004)	.010	(.003)	.011	(.004)
Engineers and technicians	.109	(.009)	.344	(.011)	.101	(.008)	.340	(.011)
Merchants	.050	(.007)	.065	(.006)	.050	(.006)	.065	(.006)
Managerial, administrative	.233	(.013)	.312	(.011)	.221	(.012)	.310	(.011)
Protective service, criminal justice, military	.061	(.007)	.048	(.005)	.057	(.007)	.049	(.005)
Writers, publishers, artists, designers	.065	(.008)	.025	(.004)	.071	(.007)	.026	(.004)
Medical services	.100	(.010)	.048	(.006)	.102	(.009)	.048	(.006)
Social service, education	.369	(.015)	.142	(.008)	.381	(.014)	.147	(.008)
Other	.003	(.002)	.004	(.001)	.006	(.003)	.004	(.001)
<i>Job training</i>	.963	(.006)	.958	(.005)	.953	(.006)	.953	(.005)
<i>Job autonomy</i>	1.79	(.02)	1.95	(.01)	1.77	(.02)	1.94	(.01)

Note: Weighted estimates. Source: HIS Graduate Panel Study.

Table A3: Hours worked per week by gender and parenthood status.

	Parents		Non-parents		Parents and non-parents	
	mean	s.e.	mean	s.e.	mean	s.e.
Full-time only						
Men	47.2	(.4)	47.1	(.2)	47.1	(.2)
Women	44.1	(.9)	45.9	(.3)	45.7	(.2)
Difference	-3.1	(1.0)	-1.2	(.3)	-1.4	(.3)
Full-time and part time						
Men	46.7	(.4)	46.8	(.2)	46.7	(.2)
Women	33.2	(.9)	44.9	(.3)	42.6	(.3)
Difference	-13.5	(.9)	-1.9	(.4)	-4.1	(.4)

Note: Weighted estimates. Source: HIS Graduate Panel Study.

Table A4: Sequential decomposition based on sample including non-employed persons.

		Replication E: replication with correction for misspecification of family formation, full-time, part-time, and non-employed persons N=3,841		
Model number	Model description	Income gap (b_{female})	s.e.	% of gap explained
1	Female	-10,610	(479)	0.0
2	Female and background	-10,545	(483)	0.6
3	Female, background, and values	-9,803	(481)	7.6
4	Female, background, values, and education	-6,616	(566)	37.6
5	Female, background, values, education, and family formation (not interacted with gender)	-6,692	(557)	36.9
5b	Female, background, values, education, and family formation (interacted with gender)	-3,073	(640)	71.0
6	Female, background, values, education, family formation, and work	-	-	-

Note: Weighted estimates. Source: HIS Graduate Panel Study.

Table A5: Oaxaca-Blinder decomposition based on sample including non-employed persons.

Replication E:			
replication with full-time, part-time, and non-employed persons			
N=3,841			
Variable (endowments effects if not stated otherwise)	Diff. in income explained	s.e.	% of total gap explained
Background	30	(56)	0.3
Importance of having lots of money	760	(112)	7.2
<i>Education-related</i>			
Grades	-13	(51)	-0.1
Percentage female of college major	3,551	(277)	33.5
Institutional selectivity	-369	(90)	-3.5
Highest degree	65	(34)	0.6
<i>Family formation</i>			
Family formation (endowments effect)	-2	(71)	0.0
Family formation (coefficients effect)	3,497	(421)	33.0
Total income gap	10,610	(479)	-

Note: Weighted estimates. Source: HIS Graduate Panel Study.