Income and Housework: A Propensity Score Matching Analysis of Gender Differences[1]

Sarah M. Ludwig-Dehm

This paper examines the effects of individuals' relative income to their partner's, on the amount of housework in a counterfactual framework. Using data from the American Time Use Survey (ATUS), a propensity score matching analysis is conducted separately for men and women, investigating whether the theoretical considerations of bargaining theory or gender performance Theory are better suited to predict the outcome. Additionally, a sensitivity analysis is used to study the bias of the estimated effects due to unobserved confounding variables, like gender ideologies. The findings show that for men as well as for women relative earnings affect their amount of housework negatively. The effect becomes weaker towards the higher shares of relative earning. This relationship is best explained by bargaining theory. Unobserved variables, like gender ideologies, do not seem to bias the results.

[1] Direct all correspondence to Sarah M. Ludwig-Dehm, 203 Oswald Tower, Department of Sociology, Penn State University, University Park, PA 16802, sml379@psu.edu

Gender equality is a much discussed topic in society. Thereby, it is especially focused on arguments about and the goal to diminish gender differences, and equal opportunities in occupations, earnings, childcare, and housework. In the traditional male-breadwinner model's view, women are the ones responsible for the work inside the house, like childcare and household chores, while men are in authority of creating income (Wang, et al., 2013). With more and more women entering the workforce, working women are a typical picture of the present (Wang, et al., 2013). This goes along with changing gender ideologies of men and women towards more egalitarian attitudes. Shifts in women's participation in the labor market as well as shifts in gender ideologies might also affect the gendered area of housework of couples. Therefore, the question arises why studies are still finding a gender gap in the amount of housework men and women are doing (see for example Bianchi, et al., 2000; Bittman, et al., 2003; Evertsson and Nermo, 2004; Schneider, 2011; South and Spitze, 1994). The gender gap is however, in line with articles suggesting that with entering the labor market women just take on additional work, because they are still taking care of most of the housework chores and childcare responsibilities (Hochschild and Machung, 1989). Hence, this study examines how the relative income of men and women who are living with a partner, affects the amount of housework they do. Specifically, it will be looked at the relationship between the propensity to have a specific share of the total couple's income and the minutes spent on housework.

There are studies that examine the effect of relative or absolute income on housework (see for example Schneider, 2011; Usdansky and Parker, 2011; Evertsson and Nermo, 2004; Greenstein, 2000). These studies, however, do not show consistent results regarding the type of relationship between these two variables. Additionally, there is still a gap in the literature on the underlying theoretical mechanisms connecting the two concepts. Especially, the possible

2

differences in the relationship mechanisms and outcomes between women and men is not clear. Another drawback of most of these studies are confounding variables that are not observed or not included in the analysis and so can lead to biased results, like gender attitudes (for a discussion see Usdansky and Parker, 2011; Greenstein, 2000).

This study goes beyond existing research in assessing the relationship between relative earnings and housework. First, unlike most articles this study uses high-quality time use data for the U.S. for a broad time frame (American Time Use Survey, 2003 to 2013) (Hofferth, Flood, and Sobek, 2013). Second, propensity score matching is used in this paper, allowing for the estimation of the effects in a counterfactual framework (Guo and Fraser, 2014). And third, a sensitivity analysis is conducted, to assess the robustness of the findings to confounding variables.

The structure of the paper is as follows: first, the two theoretical arguments most frequently used to explain the relationship between income and amount of housework, *bargaining theory* and *gender performance theory*, are presented and discussed. The section also includes a discussion of the recent literature and findings in that area. Second, the methodological problem of biased results due to selection problems is discussed and the possible solutions of Propensity Score Matching (PSM) coupled with a sensitivity analysis are explained. Third, the hypotheses are presented. Fourth, a section on data and methods, presents the American Time Use Survey (ATUS) dataset, the operationalization of the major concepts, relative income and housework, as well as of the control variables. In the same section, the methodological procedure of PSM is described. Finally, the findings and important theoretical and methodological conclusions are discussed.

THEORETICAL BACKGROUND

Bargaining Theory

An underlying assumption when explaining the share of housework among men and women with bargaining theory, is that housework is an unpleasant task that individuals are trying to avoid but that still has to be done. Therefore, couples are bargaining about it. The bargaining procedure does not have to be open and explicit, but rather is assumed to be completed in implicit negotiations. The results of these negotiations between partners should yield in contributions and gains for those involved in the household (Greenstein, 2000).

The theory assumes actors who are self-interested, meaning that if one of the partners has access to a resource that is of importance to the household, while the other has no access to it, it creates a dependency between the two individuals. The person who does not have access to the resource, therefore, holds a weaker position in the negotiations. An example of such a resource is money. Money is needed in a household, hence, wages can influence the bargaining positions of men and women (Lundberg and Pollak, 1996). The partner with less resources, meaning a lower income, has a weaker position in the negotiations (Usdansky and Parker, 2011). Taking the aspects of housework as an unpleasant task and income as a resource together, means that work in the household is provided in return for financial support (Brines, 1994). In the context of a couple household, therefore, it should be observed that as for example, the woman's earnings are rising relative to the husband's, the wife has more power to influence the distribution of housework between the two, and therefore, should do less housework (Lundberg and Pollak, 1996).

As can be seen from the assumptions made by bargaining theory, no gender differences are predicted. Some part of this is because normative and cultural factors are not included in the theoretical considerations (Brines, 1994). The relative income is expected to have the same effects for both gender. Therefore, bargaining theory predicts a linear negative relationship between relative income share and amount of housework for both women and men. These predictions could explain the gender gap in housework with a gender gap in earnings: although more women are entering the labor market, women still do more housework than men (Schneider, 2011; Evertsson and Nermo, 2004; Coltrane, 2000). This is due to lower earnings of women compared to their male partners (Greenstein, 2000; Brines, 1994). A lower relative income compared to their partner, gives women less power in negotiations about household chores. Therefore, women end up doing more work in the household than men.

Studies examining the relationship between income and amount of housework, oftentimes present findings that are in line with the predictions made by bargaining theory. Brines (1994) finds for example, a linear negative relationship between relative income share for women and individual's weekly housework hours in the United States. Similar results were also found in Sweden, shown by a study by Evertsson and Nermo (2004), who find support for bargaining theory among women. Kan (2008) found some evidence for the theory for men and women in Great Britain. Similarly, Gupta and colleagues (2009) find that higher earnings of women go along with less housework done by women. Additionally, Coltrane (2000) who reviews more than 200 articles and books, presents several studies that show that women who are employed longer and work more hours, not only earn more money but also do less housework. In almost the same manner, men who work less and have a higher educational level do more housework.

However, many studies do find gender differences in the relationship between income and work in the household. Schneider (2011) for example finds no significant effect of relative income for men. In contrast, he finds a curvilinear u-shaped relationship between the two concepts for women, meaning that as women earn a higher share of the couple's income, they do less housework. This, however, holds true only up to a specific threshold. For women, above that threshold, the relationship is reversed and the higher their share of the couple's earning, the more work they are doing in the household (Schneider, 2011). Usdansky and Parker (2011) also find a curvilinear relationship for women without a college degree, although they do not find a significant effect for other educational groups of women. While Evertsson and Nermo (2004) as well find a curvilinear effect for women in the U.S., Brines (1994) shows that men tend to do less housework the less they earn. Similarly, Greenstein (2000) predicts with his models more housework for women and less housework for men than what would be assumed by bargaining theory.

These findings show on the one hand, that there is not yet a consensus about the type of relationship between income and amount of housework. On the other hand, the findings suggest, that bargaining theory alone cannot account for the whole relationship between the two concepts, since the quadratic findings for women and the null findings for men cannot be explained. Therefore, the following section presents the theoretical considerations of gender performance theory.

Introducing the theoretical concept of doing gender (West and Zimmerman, 1987) and the assumptions of gender performance theory (Schneider, 2011), helps to account for gender differences and the quadratic relationship between income and work in the household for women. This perspective assumes that the individual behavior is influenced by expectations of others (Bittman et al., 2003). These expectations are strongly influenced by norms and values that are present in a society. Individuals want their activities and their behavior to make sense to other people Thereby, they want to be in line with prevalent social norms (Bittman et al., 2003). Some of the norms, individuals are adapting to and behaving in line with, include believes and assumptions about gender. As Schneider (2012) points out, gender is an important aspect of social interaction and used as well as reproduced and created by individuals' actions in everyday life.

An area that is highly gendered is housework (West and Zimmerman, 1987). Coltrane (2000) for example states, that housework cannot be fully understood without taking its embeddedness in gender, the household structure, and family interactions into account. Household chores are seen as functioning as fields in which gender is created and which helps to structure gender relations (Evertsson and Nermo, 2004; Coltrane, 2000). Norms about housework are strongly associated with gender roles and heavily influence housework. Women might feel, for example, that they are responsible for contributing a major part of the housework, whereas men's part appears to be optional (Bittman et al., 2003; Coltrane, 2000). Thereby, not doing household tasks is associated with masculinity for men (Schneider, 2012). The gender gap that is still existent in housework could be interpreted as showing the gendered nature of housework. Additionally, working and earning money is an area of which gender is an important

aspect. Being the main earner of income is very stable and structures expectations of men (Thébaud, 2010). Breadwinning is a central factor of reproducing masculinity. This is found across nations (Thébaud, 2010).

The norms in both areas let men and women behave in gendered ways. If however, a couple deviates from the traditional gender norms in one area, another area can be used to re-establish the norm (Brines, 1994). If, for example, a couple deviates from the gendered norms in the breadwinning area because the woman earns more than the man, the couple will try to compensate for this deviance by showing gendered behavior in a different area, e.g. household chores. This means that the woman ends up doing most of the household tasks while the man contributes only a small amount (Brines, 1994).

The curvilinear findings for women concerning income and housework can be well explained by combining bargaining theory and the concept of gender performance. Thereby, it can be argued, that bargaining behavior is shown up to a specific point, at which women earn more than men. From this point on, women will compensate for their higher income relative to their man's and show gendered work in the household, resulting in a greater amount of housework. As presented above, these patterns are found in several studies (Schneider, 2011; Usdansky and Parker, 2011; Evertsson and Nermo, 2004).

SELECTION EFFECTS

A few studies point to some methodological difficulties. Some research suggests for example, that the curvilinear effect is only driven by extreme values for persons who are earning all or nothing of a couple's income (see for example Bittman, et al., 2003). Problems can also arise

with selection effects: a confounding variable might influence the independent as well as the dependent variable. If this variable is not controlled for in the analysis, the results are likely biased. One variable that might serve as a confounder in many studies and is often unobserved (because not available in the data) is gender ideology. Men with higher or women with lower relative income could hold more traditional gender ideologies. These more traditional gender ideologies might also explain their performance of housework. Controlling for this concept might actually eliminate significant relationships between relative income and housework. A majority of the studies on the relationship between these two variables do not include gender ideology (for studies that do include the concept see Kan, 2008; Greenstein, 2000). Additionally, to my knowledge it has not been included in recent studies using U.S. data. Because gender ideologies are also not available in the dataset used for this study, I will use Propensity Score Matching (PSM) coupled with a sensitivity analysis to deal with this issue.

Propensity Score Matching allows for the comparison of individuals that are equal on all observed variables and only differ on the treatment condition (Guo and Fraser, 2014). This is accomplished by matching individuals in the treatment condition with individuals in the control condition on the probability to receive the treatment. This makes it possible to estimate the relevant effect for individuals who have the same likelihood of receiving the treatment. The objective of this procedure is to simulate an experiment, so that the treatment can be considered to be the result of random assignment (King, et al., 2007). Thereby, it is important to note that with PSM as used in this analysis, one can only estimate the average treatment effect of the treated and not the effect of the treatment for the whole population (Guo and Fraser, 2014).

PSM has some advantages over regression analysis (Harding, 2003): First, PSM does not make any assumptions about the form of the relationship between the dependent variable and the

9

independent variables. Second, estimators of PSM have smaller standard errors than estimators of regression analyses, they are more efficient. And third, collinearity is not a problem for PSM. Partly because of these advantaged, PSM is often viewed as a more appropriate method than regression analysis, and sometimes even as a method with which causal effects can be detected. While an experiment with randomized assignment uses a treatment assignment that is independent of the outcomes, neither regression analysis nor PSM can provide this. Even when using PSM, variables can bias the results, because the treatment assignment "follows a natural process that confounds group assignment with outcomes" (Guo and Fraser, 2014: 30). As already described above, an unobserved covariate, like gender ideology, might differ between the different groups of the independent variable, and might bias the results. Therefore, a sensitivity analysis is conducted to test whether the results of the PSM are robust to this kind of selection bias.

HYPOTHESES

My hypotheses are based on the recent literature on the topic of earnings and housework as well as on the theoretical considerations on combining bargaining theory and the gender performance approach. For men, I expect to see only small differences in the amount of housework between different earning levels. Therefore, the first hypothesis suggests, that:

Hypothesis 1: For men, the relative couple earning does not influence the amount of housework.

For women, I expect different results. Due to the assumption that bargaining theory and gender performance theory work together, I suspect a u-shaped curve between different earning levels and the amount of housework for women. More precisely:

Hypothesis 2: For women, the relative couple earning influences the amount of housework negatively, but only up to a specific threshold, at which the influence reverses. From this point on the influence of the relative couple earning on the amount of housework is positive.

DATA AND METHODS

The ATUS Data

For this analysis data from the American Time Use Survey (ATUS) (Hofferth, Flood, and Sobek, 2013) are used. The ATUS draws it sample from the Current Population Survey (CPS) (U.S. Census Bureau, 2014). Every month 60,000 household are selected by the CPS. Households that finish their eighth CPS interview can be selected two month after their last interview into the ATUS sample. The sampling technique of the ATUS is a three-stage process. At the first stage, states are sampled. Because the CPS oversamples states with a small population, the CPS data have to be subsampled for the ATUS to get a sample that is equally distributed across the states. At the second stage, households are sampled based on specific characteristics: race and ethnicity, presence and age of children, and the number of adults in households without any children present. In the last stage, one person, who is at least 15 years old, is selected from the household. The ATUS data are available for the years 2003 to 2013. The response rates range between 49.9

percent (in 2013) and 57.8 percent (in 2003), which is equivalent to 12,700 persons every year from 2004 to 2013[2].

The ATUS data are collected using computer assisted telephone interviewing. 50 percent of the sample are called on weekdays (10 percent on each weekday) and 50 percent on the weekend (25 percent on each day of the weekend). The core part of the interview contains a time diary. To collect this data, the respondents are asked to report their activities during a 24 hour period, starting at 4 am of the previous day and ending at 4 am at the day of the interview. For every activity reported, the respondent is asked for the duration, where the activity took place, and whether other individuals were in the same place as the respondent. This generates the specific dataset, in which each observation has its activities coded into 17 major and many more sub-categories, allowing for time use analysis. In addition, the ATUS dataset also includes information on the respondents' household composition and contains updated CPS information on the respondents' work status as well as on the respondents' partners' work status, if applicable.

Including all of these topics, compared to other datasets, the ATUS dataset has many advantages to examine the research question. The biggest advantage is the availability of time use data. The coded categories are very specific, and thus, allow for an accurate operationalization of time spent doing housework. Thereby, the analysis is not based on questions on general minutes of housework per day, but on very exact self-reported time used to do housework (see for example Shelton and John, 1996; Marini and Shelton, 1993 on the measurement of housework). Hence, not the respondents themselves, but the researcher can define which specific activities will be regarded as housework. Another advantage is the variety of variables in the dataset on the respondents' household composition, socio-demographics, as

---

[2] The sample size was higher in 2003, leading to more than 20,000 observations for this year.

well as education, occupation, and earnings. Additionally, most of these variables are also available for the respondents' partner.

<p style="text-align:center">Sample</p>

For the analysis, the datasets of the years 2003 to 2013 are used. The final sample was defined on several characteristics. First, I restrict the sample to heterosexual couples only, who are either married or cohabiting at the time of the interview. Second, I only include respondents who were between 18 and 65 years old and who had a partner who was also in that age range. Third, the sample includes only couples in which at least one partner is working and is reporting an income greater than zero. And fourth, respondents are only included if they themselves and their partner have complete data for all used variables.

<p style="text-align:center">Dependent Variable</p>

The dependent variable in this analysis is the amount of housework done by the respondent in minutes within 24 hours. The measure of minutes of work in the household follows Schneider's (2011) approach and includes several different types of housework: (a) cleaning, laundry, and sewing, (b) food preparation and clean-up, (c) interior and exterior maintenance, (d) care of lawn, garden, and houseplants, (e) care of animals and pets, (f) vehicle and appliances repair, (g) household management, and (h) shopping. The minutes reported for every single category are summed up for the overall housework variable. The total time spent on household chores is top-coded at the 99th percentile for the final sample.

Relative Earnings and Covariates

Because propensity score matching is used, the independent variable in this analysis is not a person's relative earning, but the propensity to have a high (or a low) relative earning. To measure the propensity to have a high or low relative income, first, a variable specifying the respondent's relative income of the couple income has to be created. In line with Bianchi and colleagues (2000), but in contrast to other studies (see for example Schneider, 2011; Kan, 2008; Evertsson and Nermo, 2004), the relative income variable is measured in the respondent's proportion of the total couple's income. The variable ranges from 0, contributing nothing of the couple's income, to 1, contributing all of the couple's income. From the bargaining perspective, having a zero percent share of the couple's income means being dependent on the partner and having less power in household negotiations. In contrast, having a 1, meaning earning all of the total couple's income stands for independency and a high bargaining power.

Control variables that are used in the analysis, are the respondents' and their partner's age, their races, their educational level, their employment status, the number of children in the household, an indicator of homeownership, the respondents' school enrollment and relative working hours. Race is measured dichotomously, distinguishing between white and non-white. The variable for the respondents' and their partners' education distinguishes high school or less, some college or college degree, and graduate degree or more. The respondents' employment status differentiates between employed, not employed, and not in the labor force. The partners' employment status only differentiates between employed and not employed. The number of children only includes children who are living in the household and who are under the age of 18. Homeownership is coded as a binary variable, distinguishing between owning and not owning the housing unit. School enrollment is a binary variable indicating whether or not the respondent

is enrolled at the time of the interview. The respondents' relative working hours indicate their relative hours of work per week compared to their partner. While all of these variables are used for a comparison linear regression, adjustments have to be made for the probabilities of earning a high or a low relative income.

Methodological Procedure

The treatment variable in this study is the relative earning or the share of the couple's income. Because this variable is metric, it has to be transformed into a categorical one. Due to the fact that it is not clear how the relationship between income share and housework looks like and because it is hypothesized to be a curvilinear u-shaped relationship, income share is grouped into three categories: low, middle, and high share in couple's income. The low income share group contains respondents that earn 20 percent or less of the couple's income. The middle income share group will include those respondents earning over 20 but less than 80 percent of the couple's income. And the high income share group will include respondents who earn 80 to 100 percent of the couple's total income[3]. For estimating the treatment effect, pairwise comparisons between the three groups will be conducted.

To be able to compare the results to recent studies in this field, as a first step, different linear regressions will be calculated to estimate the effect of income on the relative amount of housework for men and women. The following propensity score analysis will be conducted separately for men and women as follows: first, four logit models will be estimated. One for predicting the probabilities to be in the high income share condition, and one model for

---

[3] The main results will be described using the specified thresholds. To assess the robustness of the findings, matching and comparison results will also be reported for different thresholds.

predicting the probabilities to be in the low income share condition separately for men and women. The models will include covariates that predict whether a person will be in the low, middle, or high income condition. Second, the predicted probabilities from the logit models will be calculated. Two propensity scores will be assigned to each respondent included in the logit model. The first one is the probability to be in the high income share group and the second one is the probability to be in the low income share group. Third, treated and control subjects will be matched based on their propensity scores. Because the treatment variable in this analysis has three different conditions, three different matchings have to be done. The first one will match subjects in the low income share condition with subjects in the middle income share condition based on their probability to be in the low income share group. The second one will match subjects from the low income share group with subjects from the high income share group. And in the third one, subjects from the high income share group will be matched to individuals from the middle income share group. The second and third matchings will be based on the subject's probability to be in the high income share group. Fourth, the observed covariates will be checked on balance. Checking the balance of the variables makes sure that the propensity score model is not misspecified (Harding, 2003). As a fifth step, the actual outcomes of the treated and untreated are compared. Because in this study the treatment variable consists of three different treatment conditions, three pairwise comparisons have to be made. As a last step, a sensitivity analysis will be conducted to make sure that no biases result from a potential deviation from the conditional independence assumption, which states that the treatment assignment is not confounded given the matching on certain variables (Guo and Fraser, 2014).

RESULTS

Descriptives and Predicting Relative Income

The descriptive statistics of the variables used for the linear models are presented in Table 1. The numbers for the dependent variable housework reflect the gender gap that is still present in this field (Bittman, et al., 2003). While women in the sample report on average around 192 minutes of housework in a 24 hour period, men only report around 121 minutes. Additionally, it can be seen that men have a higher mean value on the income share variable than women. On average, men earn 66 percent of the couples' income. In contrast, women only earn 34 percent of it. This can be partly explained by the large amount of women (around 25 percent in the sample) who are not in the labor force, and so do not have an income.

Table 1 here

The respondents will not only be matched on their probabilities of being in the low or in the high income share group, but they will also be matched on their gender. Women will only be matched with women, and men only with men. Therefore, two logistic regressions have to be estimated for both men and women to predict their relative earning. Thereby, one regression predicts the probabilities to be in the low income share group. And the second regression predicts the probabilities to be in the high income share group. The probabilities from the logistic models will be used as propensity scores for the matching procedure. Two propensity scores will be assigned to each respondent in the sample. These are bound between 0 and 1. Propensity scores are estimated for three different income share thresholds presented in Table 2.

Table 2 here

Although, I use different thresholds for grouping respondents into low, medium, and high income share groups, I will mainly focus on the results using the 20 and 80 percent thresholds. Table 3 shows the distributions for these thresholds. As can be seen, more than half of the respondents are categorized as having a middle income share. This is due to the fact that the middle income share group is rather inclusive compared to the other ones. The percentages only vary minimally between women and men. However, there are large differences in the relative sizes of the low and the high income share group. Almost 37 percent of the women are grouped as having a low income share, whereas only around 9 percent of the men are grouped into this category. The reverse effect can be seen for the high income share group: over 36 percent of men fall into this group, while not even 9 percent of the women earn a high share of the couple's income. As for the descriptive statistics, this can be partly explained by the fact that more women are not in the labor force, meaning that they do not have any income. This may also show that for the average couple the man is still the breadwinner and earns more money than the woman.

Table 3 here

Linear Regressions

Two different linear regressions are estimated for comparison purposes for men and women separately: one including the relative earning variable and all covariates and one additionally including the squared relative earning term. The results of the linear regressions are reported in Table 4. The findings are similar to the ones of Schneider (2011). For men, neither the main relative earning effect nor the squared one are statistically significant. For women, however, both models show significant effects of the relative earning: The main effect of this variable is

negative, meaning that women who earn a higher share of the couples' income tend to do less housework. Introducing the squared term into the model changes this picture: the squared term is positive and statistically significant leading to a u-shaped relationship between relative income and housework, exactly what Schneider finds in his analysis (2011).

Table 4 here

Two more linear regressions are estimated that use a categorical relative earning variable instead of a metric one. That is because the PSM analysis uses group comparisons instead of a metric variable, and so these regressions are more appropriate for comparison with the PSM results. Relative earning is grouped into three groups using 20 and 80 percent as the dividing thresholds. Table 5 shows the coefficients for two of the three relative earning categories. Similar to the results of the first regressions, for men, none of the coefficients show statistical significance. For women, the results present a picture that partly deviates from the results of the first regressions. The coefficients for women show that women earning a medium and women earning a high share of the total couples' income report less hours of housework than women earning a low share. When using the medium share category as a reference (numbers not reported), the results indicate that women in the high income share group do not significantly report different hours of housework than women in the medium income share group.

Table 5 here

The findings so far suggest that the u-shaped curve between relative income and amount of housework for women found in the first regressions, might not capture the true form of the relationship. It rather seems to follow a monotonically decreasing curve that attenuates towards the high income share group.

Many different procedures can be used to match the treated with the untreated respondents based on their propensity scores. The goal is to match respondents from the treatment condition to respondents from the control condition that have the same likelihood to get the treatment. Thereby, the matching differs among others on the aspects of (a) how different are the propensity scores allowed to be for matching, (b) where to look for a matching partner, and (c) if control subjects can serve as a match for multiple treatment respondents. While defining a narrow range of difference in the propensity scores for matching may lead to better matchings, this may at the same time lead to treated units that cannot be matched, because the range is too narrow. On the other hand, if the range is defined wider, so that all treated subjects find a matching partner, some of these matches won't be very useful because of a large difference in the propensity scores. These advantages and disadvantages have to be kept in mind when making decisions about the matching procedure.

For this analysis, the algorithm of radius matching is used (see for example Becker and Ichino, 2002). This procedure makes sure that treated subjects are only matched to control subjects that are within a predefined range of the treated subject's propensity score. The trade-off between matching quality and unmatched respondents due to the radius size has to be taken into account. It is specified, that the matching procedure matches with replacement, meaning that control subjects can be matched multiple times. To check the robustness of the results obtained from this matching procedure, three different matchings are executed. The first one defines a rather large radius, or neighborhood, in which subjects are matched (r = 0.1). For the second matching, the radius is reduced by half its size to 0.05. And the third matching uses an even narrower radius size to get higher quality matchings (r = 0.01). Additionally to these three

different radius matchings, nearest neighbor matching will be used for further robustness checks. All of these matching procedures will be used for each one of the income share thresholds.

Matching Results

The quality of propensity score matching can be assessed with several different methods. One of them is to look at how many treated respondents are lying outside of the area of common support. The area of common support is defined by the propensity scores of the control respondents. If a treated respondent has a propensity score that is higher than the maximum or lower than the minimum of the propensity scores of the control respondents, it is regarded as off common support. These cases should not be used in the matching analysis, to ensure that no claims are made beyond the data. For the majority of the matchings only a small number of treated unites were off common support. If propensity scores were off common support they were dropped and not used for matching. This was especially the case for matching respondents of the high and the low income share group for men (and partly also for women). The number of propensity scores off common support increased with a decreasing radius.

Another method to check the quality of the matching procedure is to look at the covariate means for treated and control respondents before and after matching. While many covariates are likely to vary between respondents in the treatment and in the control group, after they are matched, the variables should have similar means for the two groups. Some adjustments in specifying the covariates for calculating the propensity scores had to be made to balance these variables. Some variables had to be dropped to ensure an appropriate matching quality. Graph 1 pictures the standardized percentage bias before and after matching for the used variables for

women using the 20 and 80 percent threshold with a 0.1 radius. It shows that most of the biases are reduced through matching. While the mean bias for the unmatched sample is 8.7 the mean bias for the matched sample reduces to 2.9. Only the biases for the respondent's and their partners' age are enlarged after matching. However, both change only minimally and are still in an appropriate range. This means that the matching procedure creates balance between the treatment and the control group, which is important for good quality matching.

Graph 1 here

The results of the matching are shown in Graphs 2 and 3. The graphs show the estimated average treatment effect for the treated (ATT) for men (Graph 2) and women (Graph 3) for different income share groups and matching algorithms. For each of the used thresholds the ATT for the radius of 0.1 and 0.05 are reported. Additionally, the ATT for the Nearest Neighbor matching is shown. Although three different comparisons had to be made to capture all differences between the three income share groups, the graphs only report the differences in housework compared to the medium income share group (MIS). The left part of each graph shows the differences in minutes of housework between the low and medium income share group. The right part of the graph shows the differences in minutes of housework between the high and the medium income share group. Positive numbers mean that the comparison group does on average more housework than the medium income share group. Negative numbers mean that the comparison group does less minutes of housework than the medium income share group.

First turning to the results for men, in Graph 2, it can be seen that almost all of the comparisons between the low and the medium income share group are greater 40 and statistically significant. For example, using the first dark red bar for interpretation (using the 20 and 80 percent thresholds for men and radius matching with a radius of 0.1): among men who have the

22

same propensity to be in the low income share group, men who are treated (meaning, who have an actual low share income) do, on average, 45 minutes less housework than men in the control group (meaning, who have an actual medium share income). The other bars can be interpreted accordingly. In summary, the results so far show that if men who earn only a low share of the couples' income would earn a medium share they would do around 40 minutes less housework per day.

Graph 2 here

This negative relationship however, cannot be translated to men having a high share income compared to men having a medium share income. The estimated ATTs on the right side of Graph 2 are not significantly different from zero. That means if men who earn a high share of the couples' income would earn a medium share they would not change their minutes of housework per day. For men almost all treated respondents could be matched to respondents in the control group. For the comparison between the low and the medium income share 4 respondents were off common support in each matching specification and so were not included in the matching. 2,203 respectively 2,359 (depending on the used thresholds were within common support and could be matched. For the comparison between the medium and the high income share group, the number of unmatched respondents rises to 14 and 22 (depending on the used thresholds). However, 8,533 and 7,185 respondents could be matched. The quality of the matching is therefore, not restricted.

The results for women in Graph 3, look similar to the ones for men. The comparison between the medium income share group and the low income share group yields statistically significant results for all matching procedures. The estimated ATTs for this comparison are all in the same range, showing that the effects are robust across matching procedures. Since the effects

23

are positive and the comparison is based on the probability of being in the low income share group, it can be interpreted as around 63 minutes of housework that women who are in the low income share group do less if they were in the medium income share group. That is more than one hour within a 24-hour period. Similar to men, this negative relationship does not hold true for the comparison between the medium and the high income share group. Here, none of the estimated ATTs shows statistical significance, meaning that if women who are earning a high share of the couples' income would earn a medium share they would not do a different amount of housework.

Graph 3 here

Again, in this matching the respondents that could not be matched because they were off support are very few. Comparing the low and the medium income share group, only 3 and 8 (depending on the used thresholds) could not be matched while 9,524 and 10,484 could be matched. The numbers are a little bit higher for the comparison of the medium and the high income share group (15 and 22), but the majority of respondents (2,248 and 2,044) could still be matched to respondents in the control group, enabling a good matching quality.

How do these results relate now to the proposed hypotheses? Concerning the first hypothesis, which stated that the relative earning should not influence the amount of housework men do, it can be argued that the significant results shown in the comparison between the low and the medium income share group refute the hypothesis. This means that the relative income of men does have an effect on the amount of housework they do. However, it has to be kept in mind that the comparison between the medium and the high income share group is not significant. This points to a negative relationship between relative income and housework that weakens towards the higher end of the income share distribution. The second hypothesis, which stated that for

women, a u-shaped relationship between relative income and housework will be found is only partly supported by the matching results. As hypothesized, the relative earning does affect housework negatively. However, as for men, the results point to a negative relationship between relative income and housework that extenuates towards the high income share but not to one that reverses. Therefore, the matching outcomes are in contrast to recent results that find a u-shaped curvilinear relationship between the two concepts for women as well as no existing influence of relative earning on housework for men (see for example Schneider, 2011).

Regarding the theoretical background, specifically bargaining theory and gender performance theory, the findings can be better explained by the first. Bargaining theory assumes that the higher a person's relative income the more power the person has in negotiations about e.g. household chore. Therefore, with increasing relative income less housework should be done. This should not be different for men and women. The matching results support this: for both men and women negative relationships are found. However, it is also found that the negative relationship weakens towards the higher income shares, which is not completely in line with the predictions of Bargaining Theory.

Sensitivity Analysis

Earlier it was argued that potential confounding variables can bias the estimated results between relative income and amount of housework. One of these confounding variables, which is only seldom controlled for, is gender ideology. A variable to measure gender ideology is not available in the ATUS dataset. Therefore, a sensitivity analysis is conducted to assess whether the obtained results from the PSM are changing, in case a variable influencing the outcome and the relative

income simultaneously is present. For this analysis, the results of the radius matching with the 20 and 80 percent thresholds and a radius of 0.1 are used for men and women. The procedure used in the sensitivity analysis simulates a potential confounding variable to evaluate how robust the estimated treatment effects are regarding the conditional independence assumption (Nannicini, 2007). Thereby, it is specified that the algorithm used for this test simulates a variable that influences the outcome variable, minutes of housework, and the independent variable, relative income in different directions. While housework is for example, influenced positively, the relative income would be influenced negatively. This mirrors the theoretical influence of gender ideologies on the two variables: for women, more conservative gender ideologies might go along with more minutes of housework but a lower relative income. And for men, more conservative gender ideologies might go along with less minutes of housework but a higher relative income. The differences between the baseline ATT and the simulated ATT are very low, for both men and women. The simulated ATT for men reduces by two minutes, while the simulated ATT for women reduces by 3 minutes compared to the baseline estimation. These findings suggests that the ATT estimates reported in Graph 2 and Graph 3, are rather robust to deviations from the unconfoundedness assumption. This means that gender ideology would not change the results strongly if controlled for in the analysis.

CONCLUSION AND DISCUSSION

This study examines the relationship between the relative income of men and women compared to their partner and the amount of housework they do. Particularly it was looked at the effect of the propensity to have a specific share of the couple income on the minutes of housework done within a 24-hour period. Compared to other studies examining this topic, the present study has

26

several advantages: first, the ATUS data, a high-quality time use dataset is used. Second, to be able to estimate effects in a counterfactual environment, propensity score matching is used. And third, a sensitivity analysis is conducted to assess the results' robustness in the presence of an unobserved confounding variable, like gender ideologies.

Based on the combination of the two theoretical approaches, bargaining theory and gender performance theory, as well as recent findings in the literature, it was proposed that (a) no differences between the three relative income groups should be observed for men, (b) no differences between the low and the high income shares of women should be observed, and (c) the middle income share group for women should report less housework than the low and the high income share group. The hypotheses were tested with the ATUS data from 2003 to 2013.

To calculate the propensities used in the PSM estimations, the relative income was categorized into three groups: low, middle, and high income share. The thresholds for grouping these income shares were varied. Four logistic regressions were then estimated to obtain the propensities for being in the high and the low income share group for each grouping threshold. The regressions were separately estimated for men and women. For the matching procedure radius matching with different radii and Nearest Neighbor Matching were used to test the robustness of the findings. The findings are very similar across the different matching specifications, showing their robustness.

It can be concluded that for men a negative relationship exists between their relative income and the amount of housework that they do. However, this relationship attenuates toward the higher end of the income share distribution. Also for women a negative relationship is found between the two concepts. This relationship as well weakens toward the end of the curve, meaning that the amount of housework that women with a very high income share do is not

significantly different from the amount of housework that women with a medium share do. Contrary to recent studies that find gender performance in housework among women, the PSM results show support for bargaining theory, because a low relative income is associated with more housework than a medium relative income. The difference between the OLS regression using the metric relative income and the PSM results suggest that a linear model, even if including a squared relative income term, does not appropriately capture the relationship between relative income and housework. Even when using a categorical relative income measure differences exist between the OLS regression and PSM, suggesting that individuals are selecting into income share groups.

To test the sensitivity of the ATTs to an unobserved confounding variable, a simulation estimated the ATT effect if a confounder is present. This is primarily done to test whether the results could be influenced by gender ideologies. This variable is not controlled for in most studies, not available in the ATUS dataset, but could bias the results. The findings suggest that gender ideology is a variable that does not bias the results too much when not included in the analysis.

However, this study does not go without limitations. First, one of the variables used for measuring the treatment variable, partner's earning, had to be taken from the CPS interview, which was conducted a few weeks earlier than the ATUS interview. Second, as was noted before depending on the matching procedure and the radius, some treated respondents were off common support and could not be included in the matching process. However, this can be disregarded because the numbers were very low especially compared to the treated respondents that could be matched with control subjects. Third, for this study the respondent's relative income compared to their partner's was used. However, to measure housework, the absolute amount of housework

had to be used because no measure of the partner's housework minutes was available. It would have been more appropriate to estimate a relative housework measure with a relative income measure. Additionally, while using the relative income in this sort of study is common (see for example Schneider, 2011; Usdansky and Parker, 2011; Evertsson and Nermo, 2004), Gupta (2007) does not find any significant effect for women's relative income share. In contrast, the article suggests that the absolute income has an influence on the housework, so that higher absolute earnings of women are associated with less housework (Grupta, et al., 2009).

Therefore, future studies could examine the relationship between absolute earnings and the amount of housework or include absolute earnings as a control variable to see if this turns the effect of relative income insignificant. Additionally, different datasets could be used for the same analysis to see if the findings vary. The analyses should also be done within different countries, since the context seems to be important for the strengths of the effects (Thébaud, 2010; Evertsson and Nermo, 2004).

BIBLIOGRAPHY

Becker, Sascha O. and Andrea Ichino. 2002. "Estimation of average treatment effects based on propensity scores." *The Stata Journal* 2(4):358-377.

Bianchi, Suzanne M., Melissa A. Milkie, Liana C. Sayer, and John P. Robinson. 2000. "Is Anyone Doing the Housework? Trends in the Gender Division of Household Labor." *Social Forces* 79(1):191-228.

Bittman, Michael, Paula England, Liana Sayer, Nancy Folbre, and George Matheson. 2003. "When Does Gender Trump Money? Bargaining and Time in Household Work." *American Journal of Sociology* 109(1):186-214.

Brines, Julie. 1994. "Economic Dependency, Gender, and the Division of Labor at Home." *American Journal of Sociology* 100(3):652-688.

Coltrane, Scott. 2000. "Research on Household Labor: Modeling and Measuring the Social Embeddedness of Routine Family Work." *Journal of Marriage and Family* 62(4):1208-1233.

Evertsson, Marie and Magnus Nermo. 2004. "Dependence within Families and the Division of Labor: Comparing Sweden and the United States." *Journal of Marriage and Family* 66:1272-1286.

Greenstein, Theodore N. 2000. "Economic Dependence, Gender, and the Division of Labor in the Home: A Replication and Extension." *Journal of Marriage and Family* 62:322-335.

Guo, Shenyang and Mark W. Fraser. 2014. *Propensity Score Analysis: Statistical Methods and Applications*. Second ed. Thousand Oaks, CA: SAGE Publications.

Gupta, Sanjiv, Liana Sayer, and Philip Cohen. 2009. "Earnings and the Stratification of Unpaid Time among US Women." *Social Indicators Research* 93:153-157.

Harding, David J. 2003. "Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping out and Teenage Pregnancy." *American Journal of Sociology* 109(3):676-719.

Hochschild, Arlie Russell and Anne Machung. 1989. *The Second Shift*. New York: Avon.

Hofferth, Sandra L., Sarah M. Flood, and Matthew Sobek. 2013. *American Time Use Survey Data Extract System: Version 2.4* [Machine readable database]. Technical report Maryland Population Research Center, University of Maryland, College Park, Maryland, and Minnesota Population Center, University of Minnesota, Minneapolis, Minnesota.

Kan, Man Yee. 2008. "Does gender trump money? Housework hours of husbands and wives in Britain." *Work, Employment & Society* 22(1):45-66.

King, Ryan D., Michaela Massoglia, and Ross MacMillan. 2007. "The Context of Marriage and Crime: Gender, the Propensity to Marry, and Offending in Early Adulthood." *Criminology* 45(1):33-65.

Lundberg, Shelly and Robert A. Pollak. 1996. "Bargaining and Distribution in Marriage." *The Journal of Economic Perspectives* 10(4):139-158.

Marini, M.M. and B.A. Shelton. 1993. "Measuring Household Work: Recent Experience in the United States." *Social Science Research* 22(4):361-382.

Nannicini, Tommaso. 2007. "Simulation-Based Sensitivity Analysis for Matching Estimators." *The Stata Journal* 7(3):334-350.

Schneider, Daniel. 2011. "Market Earnings and Household Work: New Tests of Gender Performance Theory." *Journal of Marriage and Family* 73:845-860.

Schneider, Daniel. 2012. "Gender Deviance and Household Work: The Role of Occupation." *American Journal of Sociology* 117(4):1029-1072.

Shelton, Beth Anne and Daphne John. 1996. "The Division of Household Labor." *Annual Review of Sociology* 22:299-322.

South, Scott J. and Glenna Spitze. 1994. "Housework in Marital and Nonmarital Households." *American Sociological Review* 59(3):327-347.

Thébaud, Sarah. 2010. "Masculinity, Bargaining, and Breadwinning: Understanding Men's Housework in the Cultural Context of Paid Work." *Gender & Society* 24(3):330-354.

U.S. Census Bureau. 2014. *American Time Use Survey User's Guide: Understanding ATUS 2003 to 2013*. U.S. Census Bureau.

Usdansky, Margaret L. and Wendy M. Parker. 2011. "How Money Matters: College, Motherhood, Earnings, and Wives' Housework." *Journal of Family Issues* 32(11):1449-1473.

Wang, Wendy, Kim Parker and Paul Taylor. 2013. *Breadwinner Moms*. Mothers Are the Sole or Primary Provider in Four-in-Ten Households with Children; Public Conflicted about the Growing Trend. Technical Report Pew Research Center.

West, Candace and Don H. Zimmerman. 1987. "Doing Gender." *Gender and Society* 1(2):125-151.

Table 1: Descriptive Statistics for Dependent, Independent, and Control Variables

	Women		Men		Total	
	Mean	SD	Mean	SD	Mean	SD
Dependent Variable						
Housework	192.43	153.09	120.76	142.46	158.17	152.37
Independent Variable						
Income Share	0.34	0.30	0.66	0.30	0.49	0.34
Controls (Respondent)						
Age	40.80	10.16	42.99	10.39	41.85	10.33
Race: white	0.73	0.45	0.73	0.44	0.73	0.44
Children present	0.70	0.46	0.69	0.46	0.70	0.46
Homeownership	0.81	0.39	0.82	0.39	0.82	0.39
Employment Status[a]						
Not Employed	3.76		2.69		3.25	
Employed	71.31		91.63		81.03	
Not in Labor Force	24.93		5.68		15.72	
School Enrollment	0.06	0.24	0.04	0.19	0.05	0.22
Education[a]						
HS or less	31.99		34.39		33.14	
Some College or College Degree	54.31		50.68		52.57	
Graduate Degree or more	13.70		14.95		14.30	
Controls (Partner)						
Age	42.85	10.34	40.91	10.27	41.93	10.35
Employed	0.90	0.30	0.68	0.47	0.80	0.40
Education[a]						
HS or less	35.59		32.27		34.00	
Some College or College Degree	50.15		54.06		52.02	
Graduate Degree or more	14.26		13.67		13.98	

a: Percentages instead of means reported
Source: ATUS, 2003-2013, own calculations

Table 2: Income Share Thresholds

Thresholds	Income Share Group		
	Low	Medium	High
20 and 80 percent	≤ 20 %	20 – 80 %	≥ 80 %
20 and 90 percent	≤ 20 %	20 – 90 %	≥ 90 %
25, 40, and 90 percent	≤ 25 %	40 – 90 %	≥ 90 %

Table 3: Income Share Group Distribution by Gender with 20 and 80 percent threshold

Income Share Group		Women	Men	Total
low	N	9,528	2,207	11,735
	%	36.94	9.34	23.75
middle	N	14,003	12,871	26,874
	%	54.29	54.48	54.38
high	N	2,263	8,547	10,810
	%	8.77	36.18	21.87
Total	N	25,794	23,625	49,419
	%	100.00	100.00	100.00

low income share group: ≤ 20 %, middle income share group: 20 – 80 %, high income share group: ≥ 80%

Source: ATUS, 2003-2013, own calculations

Table 4: OLS Regression with metric relative income variable for Men and Women

	Men		Women	
	Model 1	Model 2	Model 1	Model 2
Relative Income				
main term	2.685	-24.631	-43.458***	-152.263***
	(6.428)	(27.02)	(6.379)	(20.835)
squared term		22.126		106.788***
		(21.256)		(19.467)
N	23,597	23,597	25,762	25,762
Adj. R^2	0.0466	0.0466	0.1167	0.1177

*** $p < 0.001$, standard errors in parentheses, the models control for respondents' and partners' age, race, employment status, education, and the respondents' school enrollment, relative working hours, children in the household, and homeownership

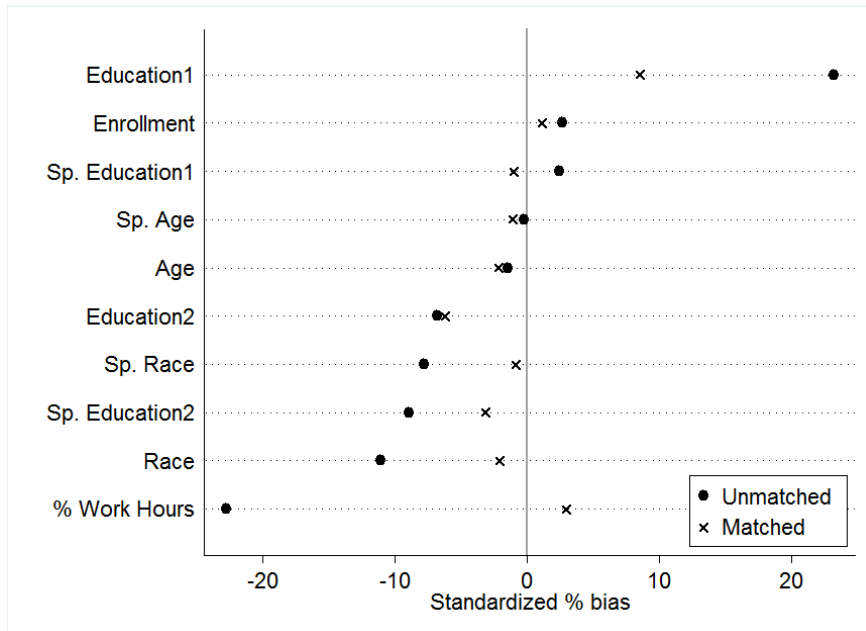Source: ATUS, 2003 – 2013, own calculations

Table 5: OLS Regression with categorical relative income variable for Men and Women

	Men	Women
Relative Income		
Ref.: low (≤ 20%)		
middle (20-80%)	-10.628	-30.533***
	(8.867)	(3.786)
high (≥ 80%)	-8.442	-36.244***
	(9.416)	(6.507)
N	23,597	25,762
Adj. R^2	0.0466	0.1174

*** $p < 0.001$, standard errors in parentheses, the models control for respondents' and partners' age, race, employment status, education, and the respondents' school enrollment, relative working hours, children in the household, and homeownership
Source: ATUS, 2003 – 2013, own calculations

Graph 1: Standardized Percent Bias across Covariates for Women

Using the 20 and 80 percent threshold and radius matching with a radius width of 0.1
Source: ATUS, 2003 – 2013, own calculations

Graph 2: Differences in Minutes of Housework between Income Groups and Matching Types for Men

Graph 3: Differences in Minutes of Housework between Income Groups and Matching Types for Women