Gender norms and income misreporting within households

Anja Roth[†] and Michaela Slotwinski[‡]

July 14, 2021

Abstract

We demonstrate that individuals' survey responses are prone to the influence of gender norms. Drawing on Swiss and Austrian data combining survey and administrative information for the same couple, we find that couples where the woman outearns her partner misreport their incomes such that they comply with the *male breadwinner norm*. This introduces a systematic bias into surveyed incomes and leads to a considerable overestimation of policy relevant measures like the gender wage gap, which is frequently based on survey data. Furthermore, surveyed income information can lead to false conclusions about individuals' labor market behavior if taken at face value.

Keywords: gender norms, female income shares, survey bias, combination survey and administrative data, income misreporting, gender wage gap

JEL classifications: D10, J01, J16

[†]University of Basel, Faculty of Business and Economics, Peter Merian-Weg 6, 4002 Basel, Switzerland. Phone: +41 (0)61 207 33 14, email: anja.roth@unibas.ch.

[‡]University of Basel, Faculty of Business and Economics, Peter Merian-Weg 6, 4002 Basel, Switzerland. ZEW – Centre for European Economic Research, L7,1, 68161 Mannheim, Germany. Phone: +41 (0)61 207 33 75, email: michaela.slotwinski@unibas.ch.

1 Introduction

Norms are understood to be important drivers of human behavior. This has been acknowledged by economists not least since the seminal work by Akerlof and Kranton (2000). In the last decade, gender norms have received ample interest by researchers as well as policy makers. While women's lives have become more emancipated and empowered, there is still evidence for strong gender inequalities. Gender norms are hypothesized to be one of the main drivers of this inequality (see, e.g., Goldin, 2006; Alesina et al., 2013; Teso, 2019; Giuliano, 2017). A growing literature in economics seeks to understand the extent to which these gender norms impact economic agents' behavior (see, e.g., Alesina et al., 2013; Fernández et al., 2004; Fernández and Fogli, 2009; Fortin, 2005; Teso, 2019).

Linking survey responses on respondents' and their partners' earnings to administrative records and studying the impact of norms on individuals' response behavior our contribution adds new evidence to this understanding. We document that the *male breadwinner norm*, which states that men are supposed to be the main earners in a couple, leads to a systematic bias in male and female income measures. Respondents misreport their own (as well as their partner's earnings) in order to adhere to traditional gender norms by placing the couple below the threshold where the woman would outearn her partner. Based on two exemplary applications, we uncover that (i) this systematic bias in surveyed incomes can lead to false conclusions about the impact of gender norms on individuals' actual labor market behavior and (ii) that it can lead to an exaggeration of official measures of gender (in)equality.

Although administrative data have become increasingly important, their availability is still quite sparse. Many studies and official figures on gender differences are based on survey data. In fact, we find that 36% of the gender wage gaps the OECD publishes for its member countries are based on survey data.[1] However, for these findings and figures to be meaningful and to reflect reality, survey data has to mirror actual behavior "well enough". This might be at risk if the surveyed item contains a systematic bias, e.g., due to social norms.

In order to detect the impact of norms on survey responses, we draw on the basic rational in the prominent contribution by Bertrand et al. (2015). They propose that the *male breadwinner norm* prompts couples to adjust their incomes such that the woman earns just less than her partner. Thus, it is the point where the women would outearn her partner (a 50 percent female income share) that marks an unwritten border distinguishing between norm compliance and norm violation. We exploit this same margin and document systematic misreporting in respondents' incomes around the point where the woman in a couple would earn more than her partner. In our data for Switzerland, individuals misreport their income in order to adhere to traditional gender norms and place the couple below the threshold where the woman would outearn her partner. The

[1] We could retrieve the original source of the GWG for 36 countries. For this assessment, employer reported earnings are counted as administrative data as they are not prone to biases due to norms.

basic idea behind the mechanism is that a violation of the male breadwinner norm entails a cost by contradicting individuals' self-perception or self-image (Akerlof and Kranton, 2000). Consistently, acknowledging that the woman earns more would conflict with the identity of an individual with traditional gender norms. While real labor market adaptations are also costly, misreporting of incomes allows survey respondents to comply with the male breadwinner norm without enduring the costs of actual income adjustments. In their original contribution, Bertrand et al. (2015) document that there is a striking discontinuity in the distribution of female income shares (female incomes as a share of couple income) in the US. We are able to replicate a discontinuity at this very threshold based on survey data. However, we do not find a discontinuity in administrative data for the very same couples in our application. The excess mass below the threshold in survey data is composed of individuals whose administrative income share lies above the threshold, who thus misreport incomes to comply with the male breadwinner norm.

While survey data would have led to the conclusion that a large share of couples just below the threshold adapted their actual labor market decisions such that they stay below the threshold, this is not what we observe. Instead, for the majority of these couples, misreporting of own and partner incomes creates the impression of adhering to the male breadwinner norm while actual behavior does not seem to respond. This insight proposes an alternative behavioral channel that might, in some cases, explain a large part of the disputed discontinuity in the distribution of income shares when measured in survey data (see, e.g., Sprengholz et al. 2020; Lippmann et al. 2020 for survey based studies and, e.g., Binder and Lam 2018; Eriksson and Stenberg 2015; Zinovyeva and Tverdostup 2021 for studies based on administrative data). Our findings suggest that the male breadwinner norm does not affect actual behavior as much as could be concluded based on survey data.

Descriptive statistics are in line with traditional gender norms being a major driver of misreporting earnings around the point where the woman outearns her partner. Focusing on couples where the woman outearns her partner based on administrative information, we find that the probability that the respondent reports a surveyed female income share below or equal to 50 percent is higher if the man is more or equally educated. Consistently, misreporting is more likely if the woman works the same or fewer hours but still outearns her partner. This is in line with the reasoning that situations which might pose a threat to the male identity lead to misreporting. Furthermore, misreporting is more prevalent among individuals from gender unequal countries, among German speaking individuals (compared to non-German speaking individuals), and among couples with a higher within-couple age difference; all measures frequently related to more traditional gender norms.

The documented misreporting has implications which go beyond the finding that in order to comply with traditional gender norms, individuals misreport their incomes in surveys. Economists and policy makers widely use survey data, e.g., to produce measures of gender (in)equality like the gender wage gap (GWG). If, as for example in our case, women's earnings are systematically underreported and those of men systematically over-reported due to gender norm considerations, this could create an upward bias in estimates

of the GWG based on survey information. Comparing GWG measures based on administrative and survey data provided by the OECD and Eurostat, we find suggestive evidence for an upward bias in survey data. Estimating the GWG in our data for Switzerland, using administrative and survey information for the very same individuals, we find that the use of surveyed incomes results in an 9.4 percent overestimation of the *true* GWG as measured in the administrative data. The bias is amplified to a 13.5 percent overestimation if the survey sample is enriched by proxy interviews, which is frequently the case to reduce survey costs (Reynolds and Wenger, 2012; Lee and Lee, 2012).[2] Furthermore, according to our results, the survey bias systematically varies with individual characteristics, like age and education. As a consequence, even a heterogeneity analysis within the same study may yield invalid findings on who is more severely affected by gender inequality. This questions the internal validity of such applications when based on survey data and underscores the importance of the use of administrative data in official statistics as well as research. It further calls for increased attention to potential biases in survey data when studying norm-sensitive items.

Finally, we present consistent evidence based on Austrian data. Four waves of the Austrian SILC also allow the comparison of surveyed and administrative incomes for the very same couples. Using this data, we are able to replicate our findings for Switzerland. Individuals misreport their incomes to place themselves below the threshold marking the male breadwinner norm, and again, we find that the survey based measure of gender (in)equality is vastly overestimated (by 21 percent). These findings corroborate that our results for Switzerland are not driven by a Swiss peculiarity nor the survey design.

Much of the literature on norm and desirability biases in surveys has been centered around under-reporting in poverty program participation and its consequences for poverty measures and the evaluation of transfer programs (see Meyer and Mittag (2019) for an overview). For example, using a matched sample of administrative data for four transfer programs and the CPS data for the US, Meyer and Mittag (2019) find that misreporting of the reception itself and of the amount of government transfers leads to biases in the evaluation of the effects of anti-poverty programs. Relatedly, Martinelli and Parker (2009) find that misreporting in surveys of self-reported program eligibility is driven not only by under-reporting due to material incentives, but also by over-reporting of goods with "status" value.

Studies considering misreporting in other survey items are rare. We are aware of three. Gil and Mora (2011) document that social norms play a role in the misreporting of individuals' body weight. Funk (2016) suggests for Switzerland that socially acceptable norms are an important driver of responses to post-vote surveys. Hariri and Lassen (2017) document that there is a social desirability bias in income reporting of high income groups in Denmark. Somewhat related literature studies the impact of survey designs on survey

[2]For instance, proxy interviews (i.e., one individual in the household reports for other household members) constitute almost 50 percent of the Current Population Survey (CPS), the most widely used data source of the US Department of Labor.

responses. Conti and Pudney (2011), for example, show that survey responses on job satisfaction are sensitive to survey context and the mode of the interview.

To the best of our knowledge, there is no clear evidence on the role of gender norms in survey responses and its consequences. However, given that there is great research interest in the impact of norms on economic behavior and that measures of gender (in)equality are at the core of many government programs, the presence of a systematic bias in survey responses translating into biased measures of gender equality seems to be of crucial importance.

The remainder of the paper is structured as follows. Section 2 describes the data for our main analysis. Section 3 presents a detailed analysis of individuals' misreporting related to the *male breadwinner norm* and documents systematic misreporting in survey-based income statements around the point where the women would earn more. Section 4 discusses and quantifies broader implications of the resulting survey bias in measures of gender (in)equality. Section 5 provides consistent evidence based on data for Austria. Section 6 offers concluding remarks.

2 Data

Our analysis primarily draws on data of the largest Swiss labor market survey (Schweizerische Arbeitskräftererhebung, SAKE) from survey years 2012 and 2015. In these years, the special questionnaire 'Social Security' including questions about partner earnings was administered in addition to the basic questionnaire eliciting respondents' earnings. The survey is based on telephone interviews where the respondent within a household is randomly chosen. This respondent then reports own and partner earnings.[3] The information from survey interviews allows us to calculate the surveyed female income share. We use the established term *female income share* throughout the paper, though incomes are in this data measures by earnings. Furthermore, we were able to retrieve the respondent's as well as his or her partner's administrative income, provided by the Swiss Central Compensation Office (ZAS) for the same years.[4,5] Observing administrative and surveyed incomes for both the survey respondent and the partner gives us the opportunity to study misreporting of ones own and the partner's incomes, as well as the difference between the administrative and the survey based female income share.

[3]More information on the exact interview process can be found in Appendix A.5.

[4]Linking survey respondents' partners to registers has only technically been feasible since 2010. Survey information on the partner's income is only surveyed as part of a special questionnaire, which, since 2010, has only been administered in 2012 and 2015. The years 2012 and 2015 are therefore the only years for which we have both surveyed and administrative income information for both the respondent and his or her partner.

[5]The respective incomes in the social security register are third party reported and there is thus no scope for misreporting.

Our sample consists of respondents with Swiss citizenship or permanent residence permit, and where both partners are in paid employment.[6] Survey income can be stated as either hourly, monthly, or yearly gross or net income.[7] The category most commonly chosen is monthly gross income, picked by 36 percent of individuals. In order to avoid biases from approximations, we focus on individuals who report both their own as well as their partner's income in the same mode (monthly net, monthly gross, or yearly gross), which is true for about 74 percent of our couples. As we study an exact threshold, the point after which the woman outearns her partner, it is important to calculate the income share earned by the woman as precisely as possible. We therefore refrain from converting monthly to yearly earnings in the case where one partner's income is stated as monthly and the other partner's income as yearly. In order to undertake such conversions, we would need to know whether a person receives, e.g., a 13. month's salary and whether this is considered when the respondent states the yearly earnings. If incomes are given in the same mode, however, they are likely comparable. The same is true for net vs. gross incomes: If one income is given as net and the other as gross we would need to know how individuals predict and perceive the difference. Any such conversions would likely introduce an error.[8] We use the survey information of those individuals who we believe are most likely to be able to report their own and their partner's income correctly. We therefore restrict the sample to individuals employed in the twelve months prior to the interview. We exclude all individuals who work shift since part of their income can vary from month to month. We concentrate on couples without institutional incentives to equalize earnings. We exclude couples where any one partner is self-employed. For self-employed, there is an incentive to distribute earnings between partners equally such that taxes are minimized. Such behavior would be reflected in a surplus of couples with an income share of exactly 50 percent visible in the administrative income share (see Zinovyeva and Tverdostup, 2021) and likely also in the surveyed income share. We further exclude all same sex couples and interviews where the randomly chosen respondent within the household was not available, as well as couples where one or both partners are above the retirement age of 65. Finally, we exclude all couples where we observe a deviation between administrative and surveyed incomes of more than 100 percent, which is true

[6]Foreigners with temporary residence permit are exposed to a special tax scheme, which among other things entails taxation at the source: In contrast to foreigners with a permanent residence permit and Swiss citizens, taxes are withheld. They might thus report their income differently, as their wage statement entails systematically different positions (see, e.g., Schmidheiny and Slotwinski, 2018).

[7]In Switzerland, the difference between gross and net income amounts to approximately 11 percent and consists of contributions to social insurances and pension payments, which are directly deducted by the employer. Except for foreigners with a temporary residence permit, taxes are paid by the individual directly. Net income therefore refers to income after social security contributions and pension payments, but before taxes. We observe a baseline underreporting of incomes of about 11 percent in our data (see, e.g., the constant terms Table A.4). The share of surveyed individuals reporting net incomes is about 30 % in the overall sample and comparable (28%) among norm-complying misreporters.

[8]We convert yearly gross into monthly gross incomes by simply dividing yearly incomes by 12.

for 44 respondents when it comes to their partner's income and 29 respondents when it comes to their own income.[9]
The main variables of interest are:

- *Survey income*: Stated income of the survey respondent and of their partner, stated by the surveyed individual.[10]
- *Surveyed female income share (FIS^{survey})*: Income share of the woman in the couple based on surveyed incomes. We define it as
  $[Survey\ inc.\ woman/(Survey\ inc.\ man\ +\ Survey\ inc.\ woman)] \times 100$.
- *Administrative income*: Survey respondent's and partner's actual earnings as recorded in social insurance registers. The variable reports total monthly gross income from employment in the month of the interview.
- *Administrative female income share (FIS^{admin})*: Income share of the woman in the couple based on administrative incomes of both partners. We define it as
  $[Admin\ inc.\ woman/(Admin\ inc.\ man\ +\ Admin\ inc.\ woman)] \times 100$.
- *Income deviation*: Deviation between survey income and administrative income for the surveyed individual and the partner. It is defined as
  $[(\ Survey\ inc. - Administrative\ inc.\ )/Administrative\ inc.] \times 100$.[11]

Other data used in additional sub-analyses are introduced at the relevant point in the paper.

3 Income misreporting and the male breadwinner norm

In this part, we study whether gender norms, and in particular the *male breadwinner norm*, lead to systematic bias in survey responses in Switzerland. We draw on work from Bertrand et al. (2015), who propose to study the distribution of female income shares to learn whether individuals labor market decisions are affected by gender norms. They document a striking discontinuity in the distribution of female income shares, measured as female incomes as a share of couple income. This discontinuity is located right at the point where a woman would outearn her partner. There is sharp bunching in female income shares just below 50 percent and missing mass above in US data. The observed statistical pattern is attributed to the *male breadwinner norm*, which states that men are supposed to be the main earners in a couple and which leads couples to sort to below

[9] We do this to exclude extreme cases where the individual reports completely unrealistic numbers in the survey.

[10] The original questions are displayed in Appendix A.5.

[11] Only survey income reported as gross can directly be compared to administrative data. We control for the fact that the income deviation between survey and administrative incomes should be larger when income is reported as net by construction by adding an indicator variable for income reported as net in all our estimations.

the threshold to comply with this norm. This study has motivated a whole literature trying to explain the discontinuity and to find the mechanisms behind it.[12] In addition to couple formation and women actively adapting their labor market outcomes in order to not outearn their partner, various drivers unrelated to gender norms have been proposed, such as the tax schedule or collective wage agreements (see, e.g., Wieber and Holst, 2015; Lippmann et al., 2020; Binder and Lam, 2018; Eriksson and Stenberg, 2015; Zinovyeva and Tverdostup, 2021).[13] We propose an additional behavioral channel (driven by norms) which has thus far not been considered. We study whether the *male breadwinner norm* leads to systematic misreporting of incomes just at the point where the women earns more than her partner. The basic idea is that a violation of the male breadwinner norm, i.e., if the women earns more, entails a cost by contradicting individuals' self-perception. Misreporting earnings such that the reported income conforms with the norm might be a cheap way to maintain ones self-concept. We expect this misreporting to occur in survey responses but also in private interactions; we obviously cannot observe the latter. In our data for Switzerland, misreporting can account for the largest part of the discontinuity in female income shares. Taking the survey data at face value would have led to false conclusions about individuals labor market behavior.[14]

We present the results on income misreporting and the male breadwinner norm in three steps: First, we compare the distribution of surveyed female income shares and administrative female income shares and test for a discontinuity at the point where the

[12] Studies using survey data, like Wieber and Holst (2015) and Sprengholz et al. (2020), find a strong discontinuity in the distribution of surveyed female income shares in Germany of roughly 60 percent, i.e., the mass just above the threshold is 60 percent lower than the mass just below. Bertrand et al. (2015) themselves use a mix of survey and administrative data for their analysis. They clearly note that there is a much larger mass of couples earning the exact same income in survey data than in administrative data and distribute the excess mass at 50 percent in surveyed female income shares to the neighbouring bins. Using the same data base as Bertrand et al. (2015), Binder and Lam (2018) show that there is a discontinuity of about 12.4 percent in US administrative data. Other studies relying on administrative data also find that the discontinuity is less distinct (Eriksson and Stenberg (2015), 22.5 percent in Sweden and Zinovyeva and Tverdostup (2021) 11.3 percent in Finland.

[13] Using administrative data Eriksson and Stenberg (2015) and Zinovyeva and Tverdostup (2021) show that the spike at the point where spouses earn the exact same income consists mainly of couples working in the same sector or for the same employer, and argue that the discontinuity at the point where the woman outearns her partner is therefore not related to traditional gender norms. Rather, the spike at the point where spouses earn the exact same income might reflect a country's institutional framework, e.g., collective agreements, minimum wages, or a progressive income tax schedule with individual taxation, which amplify incentives for couples to bunch at exactly 50 percent. This would be reflected in a large spike at this point of the distribution, which would produce a discontinuity in both survey and administrative data and which is unrelated to traditional gender norms.

[14] Descriptive evidence for the US is in line with the hypothesis that misreporting might play a role. Murray-Close and Heggeness (2018) find that the deviation between surveyed and administrative incomes is higher in couples where the woman earns more than her husband than in couples where the woman earns less than her husband. While their findings are interesting, they remain descriptive and cannot contribute to explaining the bunching below the point where the woman outearns her partner, as they compare the average of all couples below and above the threshold. As they use the full range of female income shares, it remains unclear whether the differences found are related to traditional gender norms or whether they simply reflect the different selection of individuals of couples in which the women earns more or less than her partner. Relatedly, Bursztyn et al. (2017) document that single women tend to underreport their career ambitions in situations where social norms become salient, i.e., if classmates are more likely to observe the response and larger marriage market consequences can be expected.

women earns more, i.e., at 50 percent. If there was misreporting in order to conform with the male breadwinner norm, we should observe a larger discontinuity (or a higher excess mass) in surveyed female income shares compared to administrative data. Second, to make sure that the observed differences are not driven by selection into surveys, we next draw on our main sample for which we observe surveyed and administrative incomes of the same couples. This allows us to calculate the surveyed and the administrative female income share for the same couple and assess the extent to which the observed differences can be explained by systematic misreporting of one's own and the partner's income around the threshold. These data further allow us to analyze the exact misreporting of couples where the woman outearns her partner based on administrative information, but where the respondent places the couple below the threshold or at the point where they both earn the same based on surveyed incomes. Third, we explore descriptive statistics of misreporting couples.

3.1 Distribution of female income shares

In order to compare the distributions based on surveyed and administrative incomes, we visualize them in finely binned histograms, separately to both sides of the 50 percent threshold. To test for a discontinuity just above 50 percent in the distribution of female income shares, we apply the empirical likelihood-based test by Otsu et al. (2013). This approach has several advantages over the previously proposed approach by McCrary (2008).[15] In a nutshell, it estimates the discontinuity in separate local (linear) likelihood density estimates (LLD) to both sides of the threshold. [16]

For the analysis of pure survey responses, we are able to draw on a larger sample of survey data, i.e., all SAKE waves where the question about partner incomes was asked (years 2002, 2005, 2008, 2012, and 2015).[17] However, the link with the administrative information required to analyze misreporting is only possible for the last two waves, leaving us with data from 2012 and 2015 for the detailed analysis of misreporting.

The graph on the left of Figure 1 presents the overall distribution based on surveyed incomes pooling male and female respondents. The distribution visually features a clear spike just below the margin where the woman outearns her partner and a clear discontinuity. This suggests the presence of a discontinuity in the distribution of female income shares at the point where the woman outearns her partner, just as in other countries

[15] The approach by Otsu et al. (2013) shares the good boundary properties of the local linear estimate. Additionally, the estimator is non-negative by construction, while the McCrary (2008) estimator can produce negative density estimates. See Otsu et al. (2013) for more details about the approach.

[16] As threshold we use the first observed value above 50 percent, indicated in the respective tables.

[17] In the years 2012 and 2015, the special questionnaire "Social Security" was only administered to two third of individuals questioned in the first wave of the year, which corresponds to about one third of the total sample.

(a) Surveyed incomes **(b)** Administrative incomes

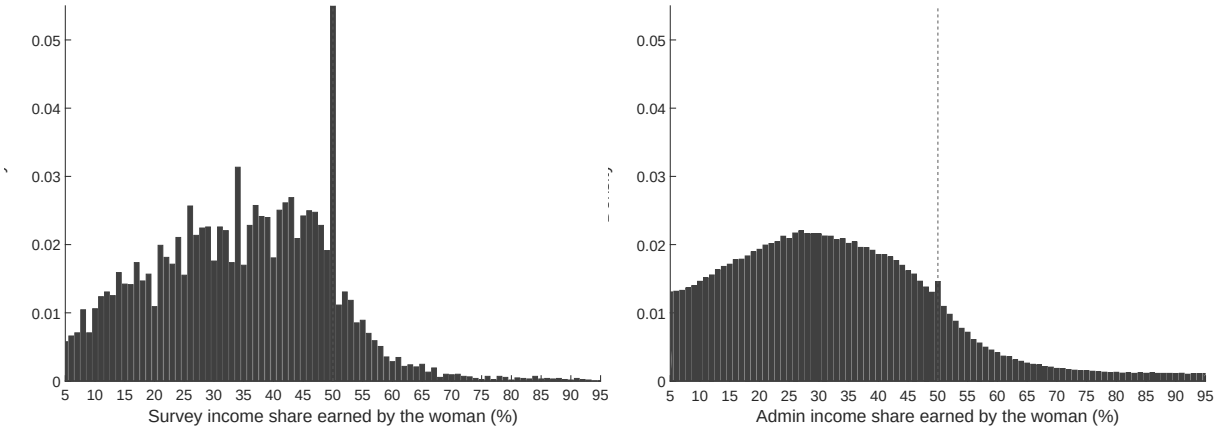


Figure 1: Overall distribution of female income shares in the couple. The shaded area represents the histogram of the underlying data in 1 percent bins. The figure on the left visualizes the distribution observed in survey data (based on SAKE survey years 2002, 2005, 2008, 2012, and 2015). The figure on the right shows the same distribution based on administrative income data for married couples (this data are described in detail in Section A.2 in the appendix). The corresponding density discontinuity estimates can be found in row (1) and row (4) of Table A.1 in Appendix A.4.

investigated in prior studies. We observe rather similar and systematic discontinuities no matter if women or men are surveyed (see Appendix Figure A.1 and Table A.1 and A.4).[18]

The overall discontinuity in the distribution based on survey data amounts to about 4 percentage points. The point estimate just below the threshold is about 4 times as high as the estimate just above the threshold; or to put it differently, the mass drops by about 75 percent at the threshold.[19] The observed drop is very similar to the one found in other studies using survey data (Sprengholz et al., 2020). As this bunching is a local phenomenon, the relative size becomes even larger, about 9 percentage points or 88 percent, if we use half the bandwidth, i.e., 3.5 percent (see Table A.2 in Appendix A.4).[20]

The graph on the right of Figure 1 presents the overall distribution based on administrative data for the whole Swiss population of married couples for 2014. These data are based on the social security register for the whole population in 2014 and are described in more detail in Section A.2 in the Appendix. The graph shows that there is no pronounced spike just below the threshold in administrative data of the whole population. The discontinuity estimates presented in row (4) of Table A.1 are much smaller but remain significant. However, it must be noted that the sample size for this calculation is huge. The drop only amounts to about 17 percent, compared to the 75 percent found in

[18]We choose the bandwidth to be 7 percent. The optimal bandwidth following McCrary (2008) would be 12.89 and that proposed by Cattaneo et al. (2018) would be 5 percent. Table A.2 in the Appendix repeats the density estimates for half the bandwidth, i.e., 3.5 percent.

[19]The conventional McCrary type approach is less precise would, however, lead to the same conclusions. See Figure A.2 in the Appendix.

[20]This is simply because the incentives to bunch are the strongest around the threshold and we see an exceptionally strong spike just below the threshold. Thus, if we reduce the bandwidth, these observations, i.e., the spike, get more weight in the estimation and the estimated discontinuity increases.

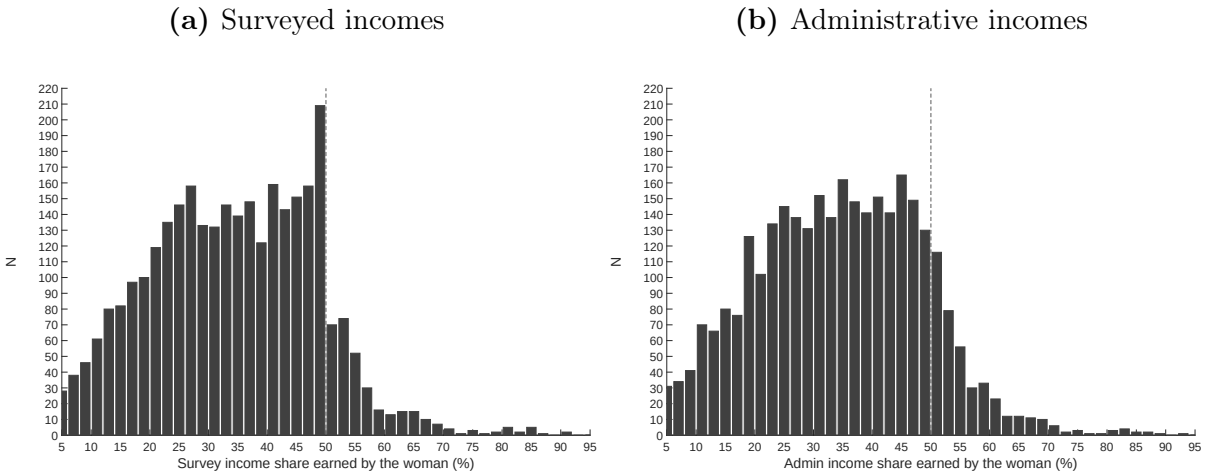


Figure 2: Pure survey and administrative distribution of female income shares for the very same couples. The distributions are presented as raw histograms in 2 percent bins. The corresponding density discontinuity estimates can be found in row (5) and row (6) of Table A.1 in Appendix A.4

survey data. This finding suggests that the spike is primarily a phenomenon of the survey data. However, it is in no way conclusive, since the population differs between the two data sources and the distribution based on administrative data can therefore not directly be compared to the distribution based on surveyed incomes. It still shows that any discontinuity that might be driven by real responses or institutional factors is magnitudes smaller than survey data would suggest. Bunching driven by institutional factors or real responses would also appear in the administrative distribution.

We use our main sample to learn whether the spike is indeed a survey artefact. While our main sample is only composed of 3,081 observations, it allows us to compare the surveyed income share earned by the woman to the administrative female income share for the very same couple. Any selection into surveys or other potential confounders can thus not explain the difference.

In Figure 2, we compare the resulting distribution of surveyed female income shares and administrative female income shares for the exact same couples. As above, we observe bunching of mass below the point where a women would outearn her partner in the survey data. When plotting the distribution for the very same couples based on administrative data, we do not observe any bunching below the 50 percent threshold. This conclusion also holds for the discontinuity estimates: While the survey distribution features a discontinuity just above 50 percent, the administrative one does not (see row (5) and row (6) of Appendix Table A.1).

The divergence between the surveyed and administrative distributions suggests that the discontinuity in survey data is a survey artifact. The finding that bunching only appears in the surveyed distribution further means that it is unrelated to any real responses to the male breadwinner norm around the threshold. In the following sections, we analyze systematic income misreporting as the driver of the divergence between the distribution of female income shares in surveyed and administrative incomes in more detail.

3.2 Misreporting of own and partner incomes

A violation of the male breadwinner norm entails a cost by contradicting individuals' self-perception. While real labor market adaptations are also costly, misreporting of actual incomes allows survey respondents to outwardly comply with the male breadwinner norm without enduring the costs a deviation from social norms would entail. In their seminal contribution, Akerlof and Kranton (2000) stress that individuals identity (or self-image) matters for their decision making. Consistently, acknowledging that the woman earns more would conflict with the identity of an individual with traditional gender norms. Misreporting therefore allows individuals to resolve the cognitive dissonance of violating traditional gender norms without bearing the loss in labor income they would incur if they adapted their labor market behavior.

3.2.1 Repositioning of couples in the distribution of female income shares

Systematic misreporting to comply with the male breadwinner norm would be reflected in women who earn more than their partner underreporting their own and/or overreporting their partner's income. A man whose partner earns just more and who wants to conform with the social norm that a woman should not earn more would, respectively, overreport his own income and/or underreport his partner's income. Such behavior would result in a strong selection of individuals around the threshold in the distribution of female income shares based on survey data. Those conforming with the norm by misreporting their own or their partner's income are placed below the threshold and those not conforming with the norm and who therefore state that the woman earns more above the 50 percent margin in the distribution of surveyed female income shares. This selection would be a consistent explanation for the divergence between the administrative and the survey based distribution of female income shares.

As we observe the position of couples in the distribution of female income shares based on administrative information, we are able to identify those individuals who cross the threshold with their survey response, i.e., couples where the woman outearns her partner based on their administrative incomes but earns less or the same as her partner according to the survey response (termed norm-complying misreporters in the following).

Figure 3 visualizes the correlation between the administrative female income share (on the x-axis) and the surveyed female income share (on the y-axis). Overall, there is a strong positive correlation between the surveyed and administrative female income share.[21] Couples where the woman outearns her partner based on administrative information but who report that the woman earns less or the same as her partner are shown in red, the opposite being true for couples in blue. We see that couples crossing the

[21] Figure A.3 in the Appendix shows how precisely individuals know their own as well as their partner's labor income. The correlation between administrative and surveyed incomes is reasonably high at 0.93 for female respondents and their own income and 0.87 for female respondents and their partner's incomes. The same numbers for male respondents are 0.92 for their own and 0.92 for their partner's income, respectively.

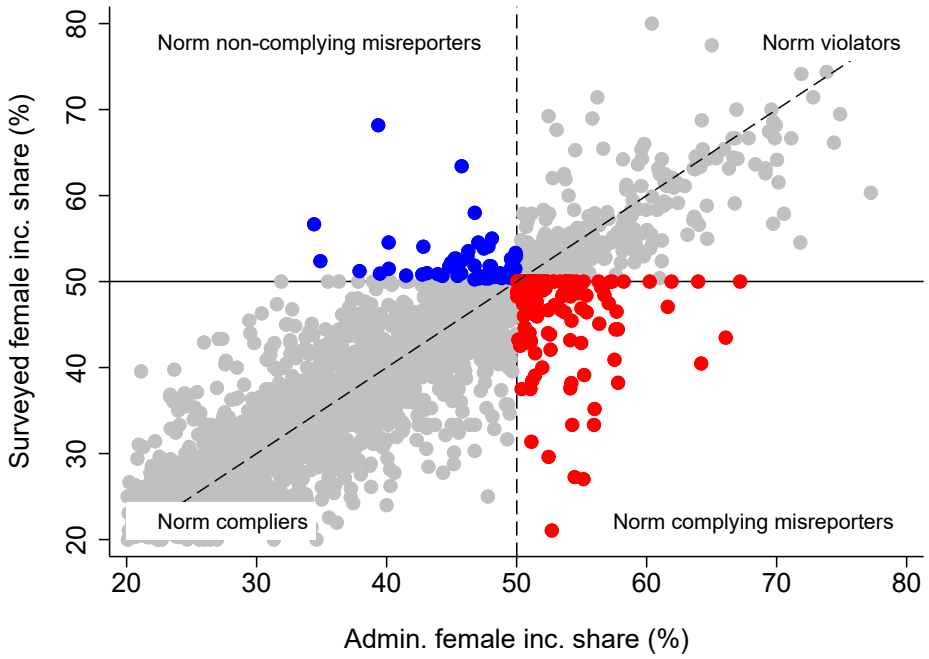


Figure 3: Surveyed vs. admin. based female income shares. This figure shows the scatter plot between the female income share resulting from survey data and from administrative data respectively, and pooling male and female respondents. The red dots mark individuals who cross the threshold from above (whose administrative female income share is above 50 percent but the surveyed one lies below or equals 50 percent) and the blue dots mark those individuals who crossed the threshold from below (whose administrative female income share is below or equal to 50 percent but the surveyed one lies above).

threshold from above (red) are placed between 51 and 70 percent in the distribution of female income shares based on administrative data and place themselves between the 20^{th} and the 50^{th} percent bin when surveyed, with the majority repositioning themselves from the range of 5 percent above 50 percent to the the five percent just below. Overall, we have 3,081 couples in the sample. In 408 of them, the woman earns more. 34 percent (141) of those couples misreport their income such that they place themselves to below the threshold. There are 2,673 couples where the woman earns less, respectively. Only 2.3 percent (61) of them misreport their income such that they place themselves to above the threshold in survey data.

The number of couples crossing the 50 percent threshold by reporting a lower female income share than what we find in the administrative data is visualized in Figure 4. The gray bars in graph (a) show the distribution of surveyed female income shares. The red bars show the number of couples in the survey distribution who cross the threshold from above and in blue the couples who cross the threshold from below. We see that there are many couples we termed norm-complying misreporters (in red) just below the threshold. In particular, about 40 percent of couples in the bar just below the threshold are placed above in the administrative distribution. This constitutes direct evidence that a considerable part of the excess mass just below the threshold is driven by couples

(a) Actual data **(b)** Randomized

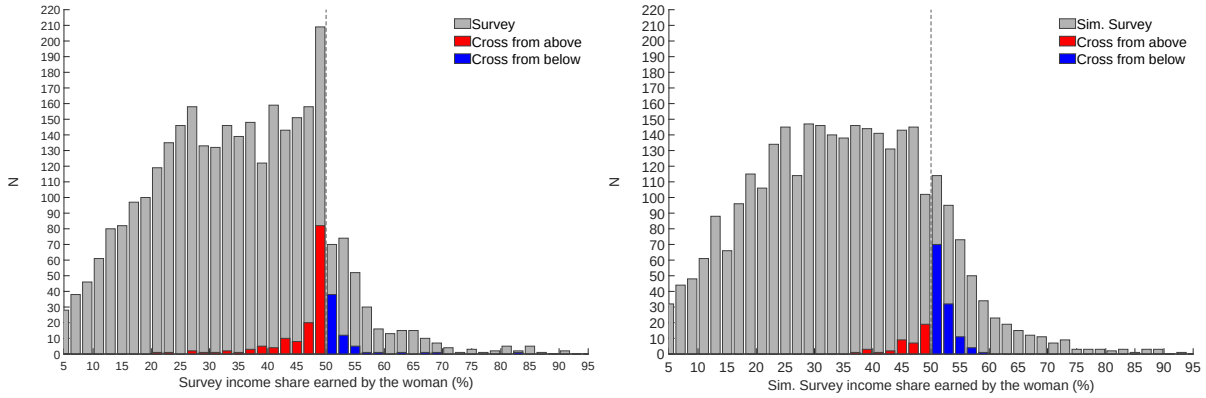


Figure 4: Histograms visualizing the number of couples finding themselves on a different side of the threshold when using survey vs. admin. data. The red bars show the number of couples where the woman outearns her partner based on administrative information but who report she earns less or the same based on survey information (norm complying misreporters). The blue bars show the number of couples where the woman earns less or the same based on administrative data and outearns the partner based on survey data (norm non-complying misreporters). The distributions are presented as raw histograms in 2 percent bins. The histogram to the left shows the distribution of surveyed female income shares and the histogram to the right visualizes the distribution of income shares and norm (non-) complying misreporters based on administrative data with random misreporting.

where the woman outearns her partner based on administrative information. There are considerably fewer individuals who place the couple above the threshold in the survey distribution and are placed below in the administrative distribution, i.e., whose deviation between survey and administrative incomes leads them to violate the norm (norm non-complying misreporters in blue).

It is important to note that the observed pattern cannot simply be explained by random deviations in income reporting of one's own or the partner's income. Theoretically, if what we observe were caused by random deviations, we would observe more norm non-complying misreporters (in blue) than norm complying misreporters (in red), as the actual mass below the threshold is considerably higher than the mass above. Figure 4 (b) visualizes a simulated distribution of surveyed female income shares. It demonstrates what a distribution of female income shares based on survey data would look like if deviations between administrative and surveyed incomes were random.[22] As expected, if misreporting were random, there would be more norm non-complying misreporters in

[22]More precisely, in order to determine the average deviation between surveyed and administrative incomes for the simulation, we regress the deviation of surveyed and administrative incomes on administrative incomes and an indicator whether incomes are reported as gross or net (since this affects the deviation). We then use each individual's administrative incomes and add a random error to simulate random reporting deviations. This error is drawn from a normal distribution with mean and standard deviation based on mean misreporting of an individual for each income, determined by the regression described before. We do this separately for men and women, as well as for their own and their partner's income.

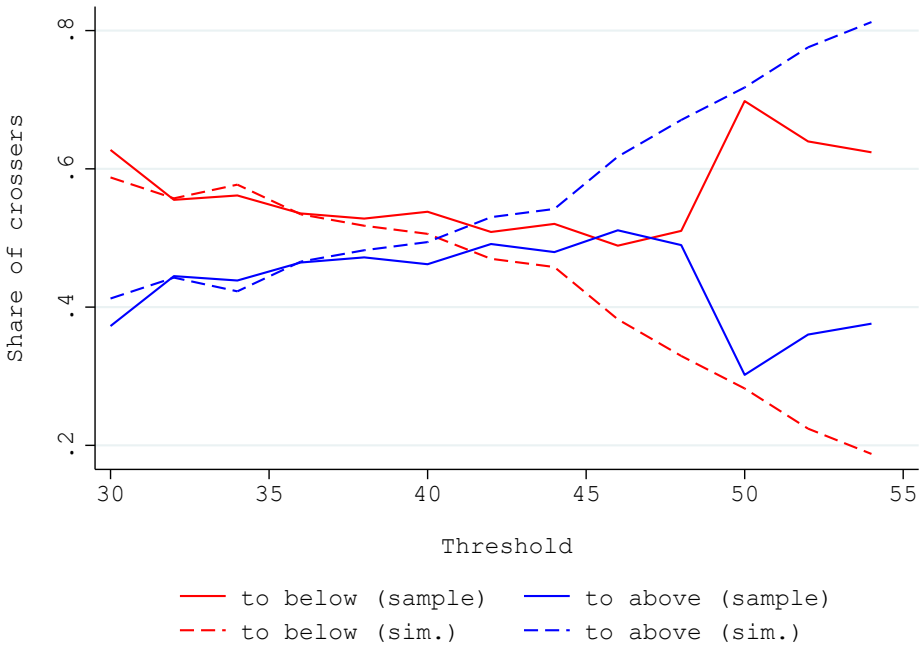


Figure 5: Share of individuals who cross a given threshold through their survey response in the original data and the simulation of random misreporting. This figure shows the share of individuals who cross a threshold through their survey response to below (red) and to above (blue), and compares this numbers for the actual data (solid line) and the simulated random misreporting (dashed line). The simulation is based on 1,000 runs of random misreporting performed as described in Footnote [22]. The threshold is set as the actual earnings share of couples based on administrative data and we evaluate the following thresholds: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54.

blue than norm complying misreporters in red.[23] This is confirmed when we repeat this random assignment of deviations in surveyed incomes 1,000 times. In our actual data, we observe that 202 couples cross the threshold, of which 61 couples (30.2%) cross from below and 141 couples (69.8%) from above. In the simulation, an average of 175.342 couples cross the threshold, of which 125.736 couples (72%) cross from below and only 49.606 couples (28%) cross from above. Consequently, in our sample, the number of individuals who place themselves below instead of above the threshold is disproportionately higher, alleviating concerns that the pattern we observe could be driven by random reporting errors. Finally, one might be concerned that misreporting of earnings in surveys is just very different than what we simulate and that our results might be an artefact of the assumptions behind the simulation exercise. Our simulation, for example, assumes that misreporting is symmetric and it would not be a valid comparison if it were skewed in reality. In order to exclude this possibility, we additionally perform a similar comparison along different thresholds, and compare the share of couples who cross a given threshold by their reporting from below or above. Figure 5 visualizes the results, i.e., the share of

[23] Figure A.4 (a) in the Appendix shows the same point, this time conditioning on the administrative female income share. Conditioning on the administrative female income share allows us to see where norm complying misreporters and norm non-complying misreporters originate.

couples crossing a given threshold from below (red) or above (blue) in the actual data (solid line) and the simulation (dashed line). The simulation and actual data move closely together up to a point close to the 50 percent female income share. They start deviating the closer we move to the 50 percent threshold marking the male breadwinner norm. The share of individuals crossing the threshold to above increases and the share crossing to below decreases if misreporting were random. This reflects the fact that as the threshold increases, the number of couples below the threshold increases and the number of couples above the threshold decreases, which increases the likelihood to cross from below to above. Up to the 45 percent threshold, actual misreporting reflects random misreporting. After that, the actual data starts to deviate from the simulation. At 50 percent, there is a sharp increase in the share of couples crossing to below and a sharp decrease in the share crossing to above. This validates that our simulation is a good comparison for random misreporting and again stresses the systematic re-positioning of couples around the threshold.

3.2.2 Estimation of misreporting of the income share earned by the woman

An alternative way to study the re-positioning is to look at systematic deviations between surveyed and administrative income shares. In order to see whether individuals whose actual income share exceeds 50 percent systematically misrepresent the actual income share, we use the deviation between the administrative and survey income share as a dependent variable and simply dummy out bins along the income distribution. Two percent bins are included for the range between 40 and 58 percent, while the observations below 40 and those above 58 percent are binned. We additionally control for an indicator for the income response mode. For the bins above 50 percent of the administrative distribution, we additionally interact the indicator for the bin with an indicator for norm complying misreporters, i.e., for couples above 50 percent in the distribution based on administrative data but with a female income share below or equal to 50 percent in the distribution based on survey data. This interaction allows us to distinguish between the behavior of norm complying misreporters and norm violators. The respective estimates show the average deviation between the survey and the administrative income share for each bin, further distinguishing between norm complying misreporters and norm violators above 50 percent.[24]

[24]More precisely we use the following specifications:

$$y_i = \beta_0 \times \mathbb{1}[FIS_i^{admin} \leq 40] + \sum_{k=1}^{9} \beta_k \times \mathbb{1}[(40 + 2 * (k-1)) < FIS_i^{admin} \leq (40 + 2 * k)] + \beta_{10} \times \mathbb{1}[58 < FIS_i^{admin}] + \rho \times mode_i + u_i, \quad (1)$$

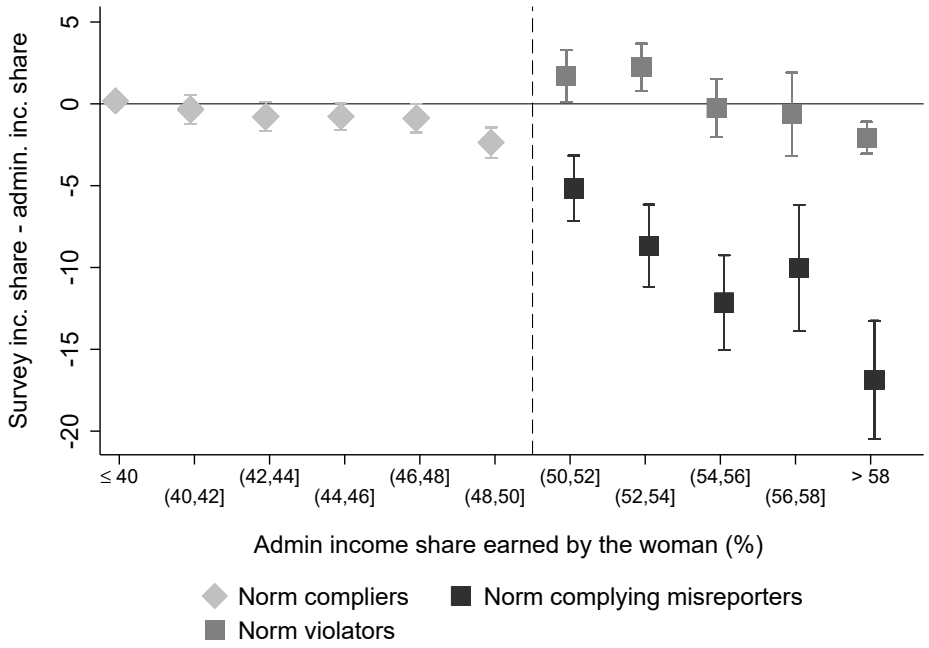


Figure 6: Deviation between the surveyed and administrative income share for bins of the distribution of administrative income shares earned by the woman. The estimates are based on regressing the deviation between the administrative and survey income share on bins of the administrative distribution and an indicator for the income response mode. For the bins above 50 percent of the administrative distribution, we additionally interact the indicator for the bin with an indicator for norm complying misreporters, i.e., for couples above 50 percent in the distribution based on administrative data but with a female income share below or equal to 50 percent in the distribution based on survey data (see, Footnote[24] for the specification). The corresponding estimates can be found in Appendix Table A.4. Estimates that alternatively condition on the survey income share can be found in Appendix Table A.5).

---

where y_i describes the outcome variable, which is either the reporting difference in the female income share (*surveyed female inc. share - admin. female inc. share*) or the income deviation (Δy =*(survey inc.- admin. inc.)/admin inc.*). The estimates of β_k therefore estimate the average in outcome y for each bin of the distribution of the administrative female income share (FIS^{admin}), controlling for the mode income is reported in (*mode*) which can either be monthly gross, monthly net, or yearly gross. The coefficients β_0 to $beta_{10}$ thus simply estimate the average dependent variable in the respective bin, or range of the administrative income share. We further interact the bins above 50 percent with an indicator taking value 1 when couples have an administrative female income share above 50 percent and a surveyed female income share below or equal to 50 percent (*norm compl. mirep.*), which allows us to distinguish between the behavior of norm complying misreporters and norm violators. We do this using the following equation:

$$\begin{aligned} y_i = \; & \beta_0 \times \mathbb{1}[RI_i^{admin} \leq 40] + \textstyle\sum_{k=1}^{9} \beta_k \times \mathbb{1}[(40 + 2 * (k-1)) < FIS_i^{admin} \leq (40 + 2 * k)] \\ & + \beta_{10} \times \mathbb{1}[58 < FIS_i^{admin}] \\ & + \Big\{ \textstyle\sum_{k=6}^{9} \gamma_k \times \mathbb{1}[(40 + 2 * (k-1)) < FIS_i^{admin} \leq (40 + 2 * k)] + \gamma_{30} \times \mathbb{1}[58 < FIS_i^{admin}] \Big\} \\ & \times \mathbb{1}[norm\ compl.\ mirep._i] + \rho \times mode_i + u_i \end{aligned} \tag{2}$$

The coefficients γ_k therefore indicate the average differential between norm complying misreporters and norm violators for the respective bin.

Figure 6 visualizes average deviations between the surveyed and the administrative female income share conditioning on the position in the administrative distribution.[25] Up to a female income share of 40 percent, there is on average no systematic deviation between the surveyed and the administrative female income share of a couple. After that, we find a visible decrease in the surveyed income share compared to the administrative income share, indicating that individuals start to underreport female incomes and/or overreport male incomes in surveys. This deviation increases for couples just below the threshold and continues to increase for couples where the woman outearns her partner based on administrative data. More specifically, norm complying misreporters, i.e., individuals who cross the threshold from above by their survey response (black squares), systematically underreport their female income share compared to other individuals in the same bin (gray squares), shown by the significantly lower estimate for norm complying misreporters compared to norm violators.

The observation that the deviation is already negative in the bins just below the 50 percent threshold indicates that couples where the woman earns less or the same as her partner based on administrative information but who are close to the 50 percent threshold already start to underreport their income share in surveys. This speaks against an equality norm argument. Such an equality norm would state that a couple's goal is to state equal earnings in surveys. Such an equality norm would be reflected in couples below 50 percent based on administrative information overreporting the surveyed income share earned by the woman, with positive estimates of β_k for norm compliers close to the 50 percent threshold.[26] The negative estimates we find for this group indicate that individuals have a preference for the men outearning his partner or that those just below the threshold preventively react to the norm and start to adapt their income reporting such that they do not violate the male breadwinner norm. For norm violators, i.e., couples where the woman earns more and who report this as such, we see two slightly positive estimates for the first two bins above the threshold. This indicates a slight overreporting of the female income share. While it is orders of magnitudes smaller than the underreporting of those who cross the threshold from above, it might still point to the fact that individuals who are willing to violate traditional gender norms are more conscious about the female income share and want to prevent the situation from being misrepresented.

3.2.3 Estimation of misreporting of own and partner income

In a next step, we shed light on how specifically respondents misreport and achieve the systematic repositioning from above to below the threshold. In a first step, we analyze whether this is achieved by a misrepresentation of male or female incomes. We analyze the

[25] The average deviation for norm compliers and norm violators for each bin of the administrative income shares is estimated by β_k, while the deviation for norm complying misreporters is estimated by the linear combination of $\beta_k + \gamma_k$.

[26] Even if we focus only on norm complying misreporters, only about 35% percent report equal earnings. Equality is thus not the main driver behind our findings.

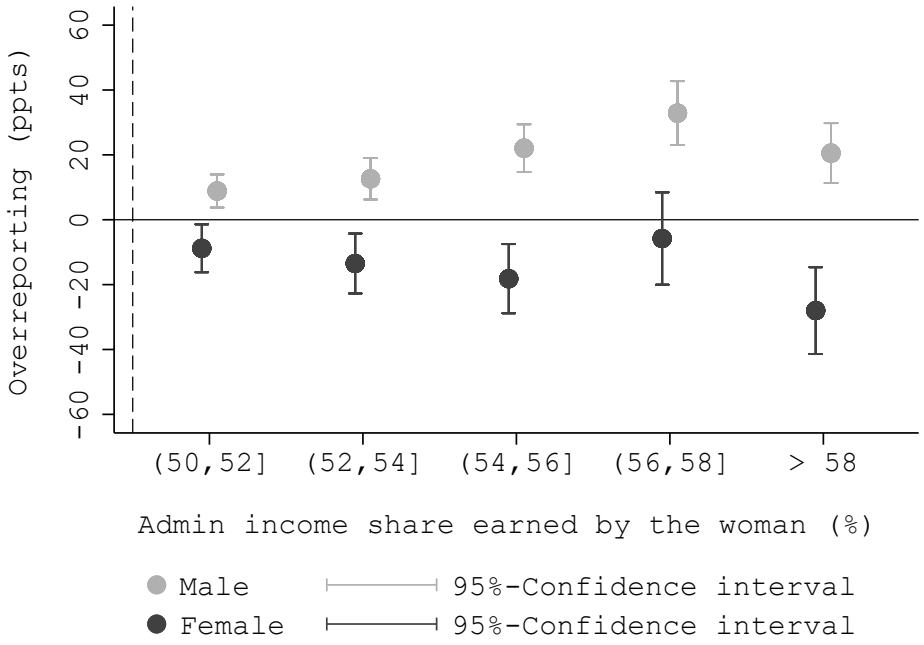


Figure 7: Excess overreporting of norm complying misreporters compared to other individuals in the same bin, conditional on the administrative distribution for women and men. It shows for each bin whether male and female incomes are misreported to fall below the 50 percent threshold.

average deviation between surveyed and administrative incomes of couples placed above 50 percent in the administrative distribution of female income shares. For bins spanning 2 percent, we calculate the average deviation between surveyed and administrative incomes of men and women controlling for the income response mode (see Equation 2 in Footnote [24] for the specification).

Figure 7 visualizes the excess deviation of norm complying misreporters (the excess income deviation of norm complying misreporters compared to norm violators in the same bin, estimated by $\hat{\gamma_k}$ in Eq. 2 in Footnote [24]). The estimates show that respondents who cross the threshold from above underreport female incomes (negative values, black dots) and simultaneously overreport male incomes (positive values, gray diamonds) compared to the average individual in the bin. The size of misreporting increases in absolute terms in the position in the distribution of administrative income shares: the farther above, the higher the required misreporting to conform with the norm. In this, the overreporting of male incomes seems to be slightly more pronounced than the underreporting of female incomes.

As one respondent per household reports the incomes of both partners in the survey we draw upon, in principle, there are three possible ways to achieve the repositioning: One can either misreport one's own or one's partner's income or both.[27]

[27]Please not that this feature of the survey procedure is unlikely to drive our results. Our results generalize to SILC Austria where each household member is surveyed individually (see Section 5).

To explore this, we run separate regressions for misreporting of own income and misreporting of partner income, and for female and male respondents. Figure 8 summarizes the results. and again visualizes the excess deviation of norm complying misreporters (the

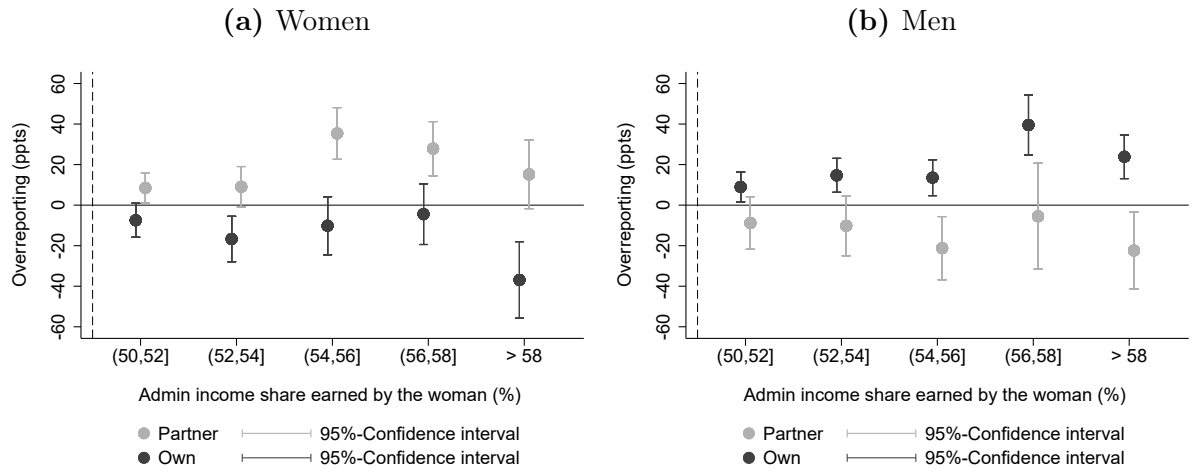


Figure 8: The figure shows excess misreporting of norm complying misreporters compared to other individuals in the same bin, conditional on the administrative distribution. We distinguish between female (left) and male (right) respondents. The figure shows for each bin whether individuals misreport their own or their partner's income to fall below the 50 percent threshold.

excess income deviation of norm complying misreporters compared to norm violators in the same bin, estimated by $\hat{\gamma_k}$ in Eq. 2 in Footnote [24]). However, this time we distinguish between female and male respondents. Panel a) shows that female respondents who cross the threshold from above underreport their own income (negative values, black dots) and simultaneously overreport the income of their partner (positive values, gray diamonds) compared to the average woman in the bin. We see a consistent picture for male respondents in panel b). Male norm complying misreporters overreport their own income and underreport the income of their partner compared to other men in the same bin. Consequently, both men and women misreport own and partner incomes in order to comply with traditional gender norms. For both male and female respondents, overreporting of male incomes seems to be clearer and more systematic.[28]

We document that in our data, the entirety of the discontinuity at the point where the woman would outearn her partner observed in survey data is explained by misreporting of incomes. Both female and male respondents in couples where the woman outearns her partner based on administrative information underreport female and overreport male income. The finding that there is no discontinuity in the distribution of female income shares based on administrative information might not necessarily be fully applicable to other countries. In Switzerland, there is, for instance, no particular tradition of collective wage agreements applying to whole sectors and there is no general minimum wage. This

[28]While Figure 8 presents the estimated differentials between norm complying misreporters and the respective reference group, Figure A.6 in the Appendix show the respective level estimates.

makes it less likely for couples to earn the exact same income, even if working in the same sector. Additionally, married couples are taxed jointly and there is no tax incentive to equalize earnings. There are thus good reasons to expect that some spike due to responses to institutional incentives would persist in other countries, as shown by Eriksson and Stenberg (2015) and Zinovyeva and Tverdostup (2021).

Nevertheless, our results stress that the use of survey data for analyses may lead to misinterpretation of real behavior, especially if the topic studied involves strong social norms. While based on survey data, we would have concluded that individuals adapt their labor market decisions to conform with the norm, real behavior seems unresponsive.

3.3 Descriptive evidence on norm-complying misreporters

In this section we descriptively present the characteristics correlated with being a norm-complying misreporter. Figure 9 shows how, on average, norm-complying misreporters differ from norm violators. We focus on couples where based on administrative data, the woman earns more than her partner. The distinction between norm complying misreporters and norm violators shows what characteristics are correlated with respondents reporting incomes in surveys which lead to a surveyed female income share below the threshold compared to couples reporting incomes such that they remain above the threshold. Panel a) presents the comparison of socio-demographics. Norm complying misreporters do not differ from norm violators when it comes to the average age of the man, the woman, nor the average household income. While there is no difference in the probability that the man in the couple has tertiary education, women in couples that place themselves below the 50 percent threshold in surveys are less likely to hold tertiary education. We find that norm complying miserporters are neither more nor less likely to have children. Except for the observation that couples where the woman is highly educated are less likely to misreport, norm complying misreporters and norm violators seem not to differ systematically.

Panel b) shows a comparison of observable characteristics known to be correlated with gender norms to proxy groups we expect to be more *traditional* and test whether these groups are indeed more likely to be norm complying misreporters. In a first step, we try to capture situations which would jeopardize the status of a man holding traditional gender norms in a relationship. This approach is based on the findings of Fisman et al. (2006) and Bursztyn et al. (2017), who document that men consider women less attractive if the woman's ambitions exceed their own.[29] The first measure we employ is the relative education within a couple. The idea is that it might be perceived as fair if the women earns more if she is also more educated than the man. A situation where the woman outearns her partner and is equally or less educated might, however, provoke discomfort and be a threat to the male identity (Akerlof and Kranton, 2000). We test for a difference in the probability that a norm complying misreporter is part of a couple where the man is

[29]It is further in line with social structure theory (Eagly and Wood, 1999).

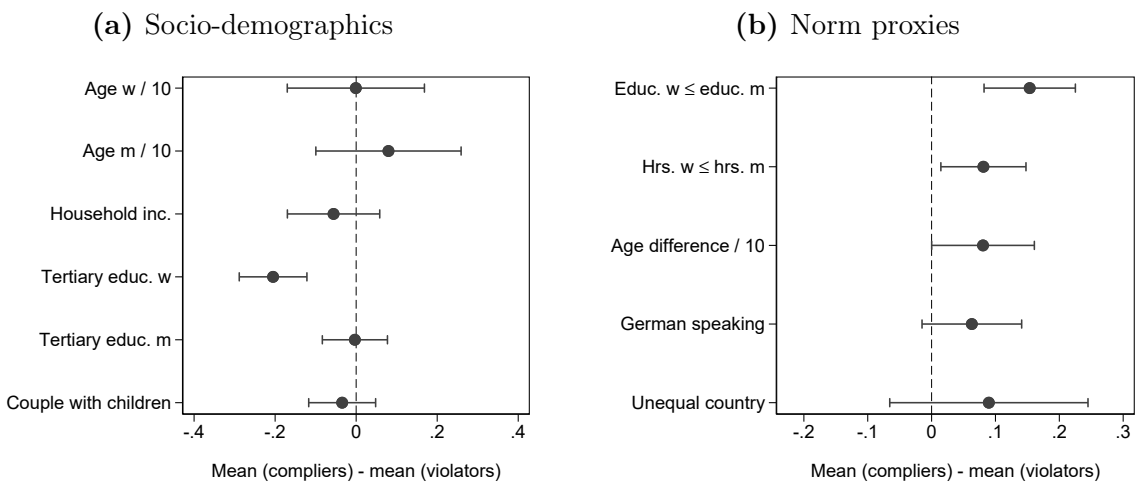


Figure 9: Average comparison between norm complying misreporters and norm violators. This Figure shows the difference in the group averages between norm violators (admin. female income share $>$ 50 percent, surveyed female income share $>$ 50 percent) and norm complying misreporters (compliers, admin. female income share $>$ 50 percent, surveyed female income share $\leq$ 50 percent) for different characteristics. The estimates only include individuals with an administrative female income share above 50 percent. Panel a) presents the comparison of sociodemographic characteristics and panel b) the comparison of observables known to be correlated with traditional gender norms. We pool the data for male and female respondents and report 90 percent confidence bounds. The corresponding estimates can be found in Tables A.6 and A.7. *Hrs.* stands for weekly work hours. *Age diff.* describes the within-couple age difference, defined as $(age_m - age_w)$. *German speaking* describes an indicator set to one for German speaking individuals (in relation to French, Italian, or Romansh speaking). *Unequal country* describes an indicator set to one for individuals with origins in a country with more traditional gender norms.

equally or more educated and is still outearned by the woman, and find that these couples are more likely to be norm complying misreporters. Or to put it differently, the share of couples where the woman is less or equally educated is 15 percentage points higher in the group of norm complying misreporters when compared to norm violators. The difference amounts to about 20 percent of the sample average (76 percent). This evidence is in line with the idea that situations which might be a threat to the male breadwinner identity provoke misreporting. Another situation, which could produce similar unease is when the woman works the same or fewer hours per week as her partner but earns more. Considering this measure, we find that couples where the woman works fewer or the same hours are systematically more likely to misreport (the difference amounts to about 10 percent of the sample average). These findings suggest that situations which jeopardize the male position within the relationship (or the male identity) seem to be a good predictor for norm-complying misreporting.

In a next step, we analyze direct proxies of traditional gender norms known in the literature. First, we exploit a proxy which we observe for all couples in the sample. Following Folke and Rickne (2020), we use the within-couple age difference as a proxy of an individual's gender norms when entering the relationship. Couples where the man is older

than the woman are assumed to be, on average, more compliant with traditional gender norms.[30] Consistently we observe that the within-couple age difference is on average higher in misreporting couples (the difference amounts to 35 percent of the sample average).

Second, we use the language the survey interview was conducted in as an approximation of gender norms. Prior studies have shown that individuals in German speaking areas of Switzerland hold more traditional gender norms than individuals in the other language regions (Italian, French, and Romansh) (see, e.g., Steinhauer, 2013). We would thus expect that norm-complying misreporters are more likely to speak German than one of the other languages. This is what we find. Norm complying misreporters seem to be on average more likely to be German speaking. While the difference is not significant at any of the conventional levels, it amounts to about 9 percent of the sample average.

As a final norm proxy, we use cultural norms in the countries of origin of migrants to approximate individuals who we expect to hold more traditional gender norms. We exploit the fact that Switzerland has a comparatively high share of immigrants and apply the epidemiological approach suggested in Fernández and Fogli (2009). We approximate an individual's norms by gender norms in the country of their ancestry. The basic idea is that individuals take part of the culture (through socialization) with them when emigrating. These norms are to some extent also transmitted intergenerationally. As these individuals live and partly grow up in the same country and institutional setting, differences in their behavior should emerge through these transmitted gender norms. Traditional gender norms are proxied by average agreement of all employed women in a country with the statement "When jobs are scarce, men should have more right to a job than women", as measured in the WVS. Splitting countries at the sample median, we define two types of origin countries: countries where average agreement is lower, which we would expect to hold less traditional gender norms, and countries where agreement is higher, which we expect to hold more traditional gender norms.[31] The difference in the likelihood that a norm complying misreporter originates from a more gender traditional country is not statistically significant at conventional levels. However, the sample is of course quite small (about 25 percent of the original sample) and the difference is considerable. Given a baseline probability of 30 percent to be from a country with traditional gender norms, norm complyiers are about 9 percentage points more likely to originate from more traditional countries, which amounts to about 30 percent of the sample mean.[32] Summing up, measures of traditional gender norms seem to be good predictors for individuals' misreporting behavior, strengthening the interpretation that individuals self-conceptions and norms are the drivers of such behavior.

[30] In our sample, men are at the median two years older than their partner.

[31] Detailed information on how we determine an individual's ancestry and the definition of equal and unequal countries can be found in Section A.3 of Appendix 2.

[32] The fact that in only 30 percent of couples where the woman outearns her partner based on administrative data, the respondent is from an unequal country serves as additional indication that a couple's female income share is related to gender norms. Based on the fact that we use the within sample median to split countries into equal and unequal, this number would be 50 percent if couples were equally distributed across female income shares.

4 Survey biases in measures of gender (in)equality

Due to a lack of high quality administrative data, social scientists have long been confined to survey data for their analyses. Survey data are sometimes also unavoidable, as they can provide information on topics such as voting behavior, division of family responsibilities, or employment in the informal sector, which cannot be elicited from administrative sources. It is all the more important for researchers to be aware of potential biases in survey information. Our evidence strikingly demonstrates that gender norms play an important role in individuals' responses to surveys and can lead to systematic misreporting.

The gender wage gap is probably the most common and policy relevant measure of gender equality. Still, a large share measures of the gender wage gap is based on survey data. For example, around 36 percent of all available OECD numbers on the GWG are based on survey data.[33] Our finding that the male breadwinner norm leads to a systematic under-reporting of female and over-reporting of male incomes around the threshold would imply that GWG numbers based on survey data might overestimate the true GWG. Eurostat, the statistical office of the European Union, on the other hand provides number on the GWG which are mainly based on (quasi-) administrative data from the structure of earnings survey (SES). The SES is an employer survey which we treat equivalently to administrative data, as there is no room for misreporting by the individual. The SES is conducted every four years.[34]

Combining the OECD and Eurostat data on the GWG allows us to explore whether there is a deviation between numbers based on survey and administrative data that would suggest an overestimation of the GWG. We are able to compare the OECD and Eurostat numbers for 31 countries. Panel a) of Figure 10 compares the average GWG numbers drawn from Eurostat and OECD reports.[35] The OECD and Eurostat use different definitions of the GWG. Even if both numbers were based on administrative data, they would thus not be identical. For example, the OECD reports the GWG at the median, while Eurostat uses the mean definition. This explains why, on average, Eurostat reports higher GWGs than the OECD. If survey based measures were overestimated, this would be reflected in a higher GWG based on this data. As the Eurostat numbers are always based on administrative data, we would expect that the differential between Eurostat and OECD measures, amounting to 20.3 percent on average, is smaller if the OECD measure is based on survey data than if it's based on administrative data. In six cases, the OECD data is

[33] According to our assessment, the GWGs of the following countries are based on survey information: Australia, Canada, Chile, Germany, Greece, Iceland, Ireland, Israel, Italy, Mexico, New Zealand, Portugal United States. This list conditions on the set of countries for which we were able identify the data source mentioned in (OECD, 2021), which additionally includes Bulgaria, Costa Rica, Croatia, Cyprus, and Malta.

[34] The data in between is based on national information and sometimes interpolations. In our analysis, we thus only draw on the years in which the SES was conducted: 2002, 2006, 2010, 2014 and 2018.

[35] We average the GWG by country and over the years 2002, 2006, 2010, 2014, and 2018 (Eurostat, 2021).

(a) Eurostat and OECD GWG **(b)** Relative deviation

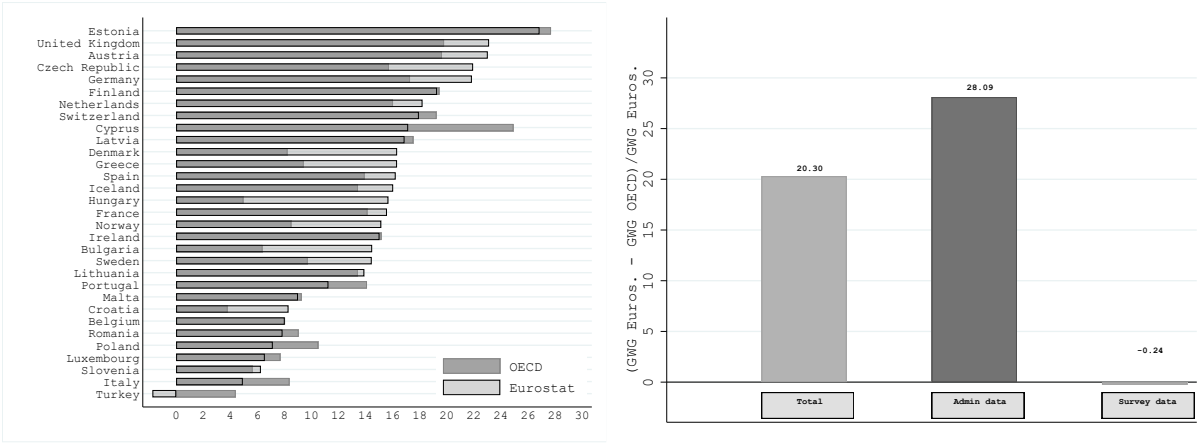


Figure 10: Comparison of the average GWG provided by Eurostat and OECD by country. The left panel shows the average GWG provided by both institutions. The data covers all countries included in both sources and the years 2002, 2006, 2010, 2014, and 2018. The right panel shows the relative deviation between the Eurostat and OECD numbers relative to the Eurostat data.

based on survey data.[36] Panel b) shows the average relative deviation between Eurostat and OECD numbers. As mentioned, the Eurostat numbers are about 20 percent higher on average than OECD numbers. In countries, where both the OECD and Eurostat numbers are based on administrative information, Eurostat numbers are about 28 percent higher than OECD numbers. This difference is likely a result of the differences in the definitions. However, if OECD numbers are based on survey information, the difference between OECD and Eurostat numbers vanishes completely. This is suggestive evidence that the OECD GWGs based on survey data are higher than they would be if they were based on administrative data, i.e., an overestimation of the GWG in survey data.

We find a consistent picture comparing the Eurostat numbers to those provided by the International Labour Organization (ILO), where some data sources are survey based as well (see Figure A.5 in the Appendix).[37]

While the comparisons presented above is suggestive evidence for a bias, it does not provide conclusive evidence of a survey bias. The GWGs are calculated differently and the numbers are probably not based on exactly the same population. Our data for Switzerland, however, offers the opportunity to assess the bias in estimates of the gender wage gap (GWG) based on survey data. We are able to assess the bias in a sample for which we observe equally many women and men and in which the selection into response by gender does not play a role as the reporting person within the household is randomly chosen.

[36]This is the case for Germany, Greece, Iceland, Ireland, Italy, Portugal.

[37]The overlap between both data sets only covers 8 countries, however. Within this group, the numbers of Austria and Portugal are based on survey information. Note that the baseline difference is much smaller, as many of the ILO numbers are based on the SES as well. This list conditions on the set of countries for which we could identify the data source from the sources mentioned in (ILOSTAT, 2021).

Table 1: Gender wage gap estimates

	GWG admin.	GWG survey	GWG admin. including partner	GWG survey with proxies
	(1)	(2)	(3)	(4)
Female	-0.149***	-0.163***	-0.171***	-0.194***
	(0.014)	(0.012)	(0.010)	(0.009)
Net	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes
N	3,078	3,078	6,071	6,071
R-squared	0.306	0.394	0.276	0.325
Overestimation		9.4 %		13.5%

Notes: OLS estimates of the gender wage gap regressing log hourly income on indicator variables for female, education, and if income is reported as net, a continuous age variable and a constant. The four specifications distinguish between the data source (survey or admin.) and whether proxy incomes are used for the estimation. We loose some observations due to missing values in education of the respondent or the partner. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Both the administrative and survey information is available for the very same individuals. Any difference can thus only be assigned to the data used.

We start by calculating the *true* GWG based on administrative information for the respondents. In particular, we regress log hourly incomes on gender, controlling for the level of education, individuals' age as well as the mode of the income response. This results in a GWG of 14.9 percent (see Table 1 column (1)). In a next step, we asses the GWG estimate based on surveyed incomes for the very same sample of individuals. Before, we demonstrated some misreporting just below the threshold and systematic and pronounced misreporting of norm complying misreporters and their own incomes above the threshold. Based on surveyed incomes, the GWG amounts to 16.3 percent (see Table 1 column (2)) and is thus overestimated by about 9.4 percent.

The bias is relevant in individuals' responses regarding their own characteristics, but might be even more pronounced when proxy responses are used. Proxy interviews are frequently used to reduce survey costs and might lead to biased information, as pointed out by Reynolds and Wenger (2012) and Lee and Lee (2012). In 2009, for example, almost half of the labor force data in the Current Population Survey (CPS) are provided by proxy respondents (Reynolds and Wenger, 2012).[38] According to our evidence the impact of norms on reported incomes is a problem that concerns both information from interviews with the target person directly and proxy interviews (individuals' responses about their partner).

[38]The CPS is one of the most frequently used US surveys in economics. Among other things, many studies on the GWG (see, e.g., Blau and Kahn, 2017, 2000, 1997; Katz and Murphy, 1992; Macpherson and Hirsch, 1995) rely on CPS data.

In a next step, we mimic the data structure of surveys with proxy information (like the CPS) and include administrative information on the respondent and the partner as independent observations. This naturally almost doubles the number of observations.[39] Based on this sample, we find a true (administrative) GWG of 17.1 percent.[40] In column (4) of Table 1, proxy income information is included, i.e., for each individual, we use both their own reported income, as well as the income they report for their partner as our data base; just like it is done in the CPS data. Now the GWG amounts to 19.4 percent (see Table 1 (4)), which overestimates the *true* GWG in the administrative data by about 13.5 percent and thus introduces a considerable bias. Both comparisons of the administrative and survey GWG differ significantly at the 10 percent level.

Many studies are not only interested in the overall level of the GWG in the population, but rather in its heterogeneity (see, e.g., Arulampalam et al., 2007; Chernozhukov et al., 2018; Blau and Kahn, 2017). Research questions include whether high or low educated women are more severely affected, or whether age is an important determinant of gender inequality. Chernozhukov et al. (2018), e.g., draw on the data from the CPS to study the heterogeneity in the GWG along marital status, education, and occupational experience. If the survey bias is more pronounced for particular groups, a comparison of the resulting GWG, even within the same study, could be invalidated as it would remain unclear whether the difference is due to real differences or due to a stronger survey bias for specific groups. Thus, even the internal validity of studies using survey data to analyze gender (in)equality might be at risk. In a next step we therefore explore whether the resulting bias varies along individuals' characteristics.

In this exercise we try to use dimensions that might also be of interest in GWG studies dealing with heterogeneity. Figure 11 visualizes the different GWG estimates for different sub-groups. In panel (a) we start by re-estimating the GWG for different age groups. We basically see that the GWG increases with age, whether we use survey or administrative data. However, the bias in the survey measure varies by groups. While the GWG for individuals aged up to 30 is underestimated by about 24 percent the GWG for the group aged between 31 and 40 is overestimated by about 18 percent and the one for those aged above 40 by about 16 percent. In panel (b), we distinguish between individuals with and without children, which is of interest in many studies asking the question whether mothers suffer a larger penalty. We observe that the GWG is somewhat larger for individuals with children. However, the difference is strikingly overestimated. While there is hardly any bias for individuals without children, the GWG for individuals with children is about 18 percent higher when based on survey data than when based on administrative data. Next, we turn to education measured by the highest educational attainment in panel (c). We see that the GWG is largest in the lowest category. However, the bias is also

[39] We loose some observations as there are missing values for the characteristics of the partner.

[40] In the survey, information on the partner's weekly work hours is only given in classes of five hours, ranging from 1 (0-5 hours) to 9 (more than 40 hours). However, we have the true weekly work hours for the respondent. Based on the this information we impute the respective average hours per class.

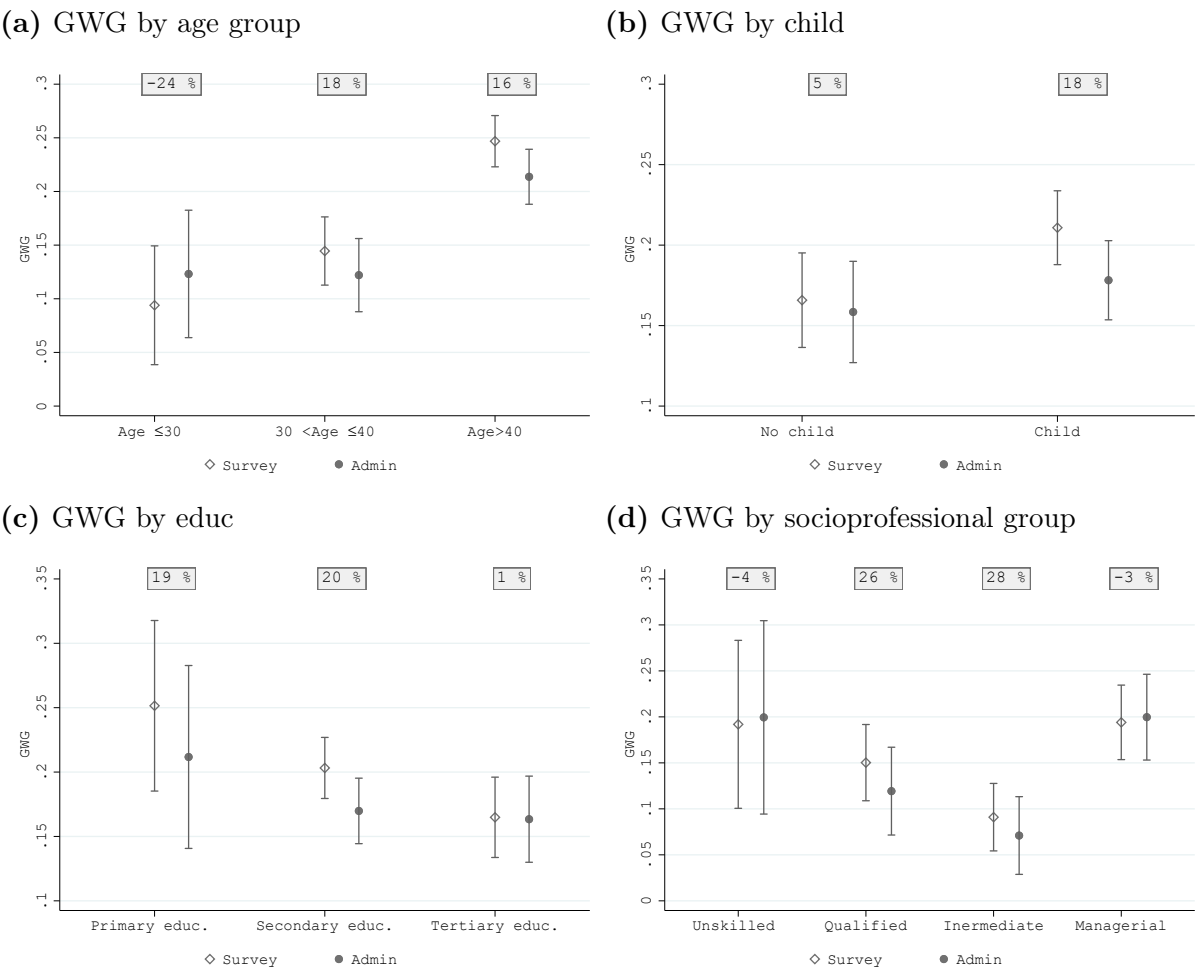


Figure 11: Comparison of GWG estimates in survey and administrative data for different groups of individuals. This figure presents the comparison of GWG estimates based on survey (hollow diamonds) and administrative (full dots) data. The resulting bias in the GWG is displayed in gray boxes. Except for panel (d), the estimates are based on the data including partner's and proxy information. Panel (d) is based on data for respondents only, as our data does not cover partner occupations.

strongest in the two categories at the bottom. The bias amounts to an overestimation of 19 percent for primary education, to 20 percent for secondary educated and is reduced to only 1 percent for tertiary educated individuals. Finally, we turn to occupations, or socioprofessional groups, in panel (d). The true GWG is slightly underestimated for unskilled occupations. It is highly overestimated for qualified (about 26%) and intermediate (about 28%) occupations, and the bias almost vanishes for managerial occupations.

Our evidence demonstrates that the survey bias varies by groups. The basic relationship seems to be valid in survey data, the survey and administrative measures mostly move in similar directions. However, the differences between groups are often biased when based on survey data. It is even unclear whether the use of survey data leads to an over or underestimation of differences. In the case of children, it clearly leads to an overestimation of the differences in the GWG between individuals with and without children. In

the case of occupations, however, it would lead to an underestimation of the difference in the GWG between, for example, intermediate and managerial occupations.

The additional analysis on the gender wage gap strikingly shows that income misreporting in surveys can lead to considerable biases in estimates of gender differences. This bias becomes more substantial if proxy information is used. Furthermore, we observe that the bias varies along individuals' characteristics, which might reflect the strength of gender norms in different subgroups. This might invalidate heterogeneity analyses even within the same study when based on survey information prone to be influenced by social norms. While we focus on the GWG, it is to be suspected that income misreporting also biases other measures of gender differences, like estimates of the *child penalty*. Unfortunately, our data is not extensive enough to test this.

5 Additional Evidence: Austria

We have shown that the male breadwinner norm affects individuals' income reporting and that there is a systematic bias in measures of the GWG based on survey information in Switzerland. However, one remaining concern is whether this finding generalizes to other countries and whether it is limited to the specific survey design where one respondent reports the incomes of both partners. While this type of data is very scare, we are able to provide additional evidence by comparing survey and administrative information for couples for another survey design and another country, namely for Austria. In this section we provide additional evidence, showing that (i) the male breadwinner norm also affects surveyed income measures in Austria and within a different survey design, and (ii) the survey bias leads to a biased GWG estimate for a more common sample in GWG studies and including typical controls.

We draw on data of the Community Statistics on Income and Living Conditions in Austria (SILC AT). The SILC is a survey that annually collects information on the living conditions of private households, mainly in member states of the European Union, therefore also in Austria. While the information on individuals' incomes was formerly elicited through a questionnaire, in 2010, Austria decided to increase the quality of the income data and to replace the survey based measures with administrative information wherever possible. In order to validate the methodology and to allow for a smoothing of the resulting structural break, the administrative data was also added to some earlier waves. This leads to the favorable situation that for the waves in 2008, 2009, 2010, and 2011, it is possible to observe the original survey based income measures as well as the merged income information from administrative sources.[41] We draw on this data to test whether our findings for Switzerland generalize for Austria.

5.1 Data, AT

In order to test whether we find evidence that the male breadwinner norm also affects survey responses in Austria and given the different survey design, we define the sample just as we did in the Swiss case.

In the SILC, all individuals aged 16 or above who are part of a selected households are interviewed personally. Thus, in contrast to the Swiss data, each person within the selected household is interviewed individually. As a consequence, each partner only reports his/her own income. We select all couples where both partners were (full or part-time) employed during the year. We thus exclude everyone with any self-employed income, as self-employed income leads to tax incentives to equalize earnings.[42] We exclude same sex couples and only include couples in which we observe surveyed and administrative incomes for both partners. We exclude couples if one partner or both partners are older than 65 or younger than 18. Our income measure is gross income from dependent employment

[41] The data from the SILC AT can be accessed by sending a data request to Statistics Austria.

[42] Our definition thus includes both married and cohabiting couples.

(Einkommen aus unselbständiger Erwerbstätigkeit, PY010).[43] We drop observations if income was imputed or stems from a proxy-interview.[44] Finally, as before, we exclude all couples where we observe a deviation between administrative and surveyed incomes of more than 100 percent. This leaves us with a sample of 1,815 couples for which we observe surveyed and administrative income for both partners.

5.2 Income misreporting, AT

The data for Austria allows us to directly compare the distributions of female income shares for the very same couples to see whether we find the same evidence for sorting around the 50 percent threshold marking the male breadwinner norm.

(a) Surveyed incomes **(b)** Administrative incomes

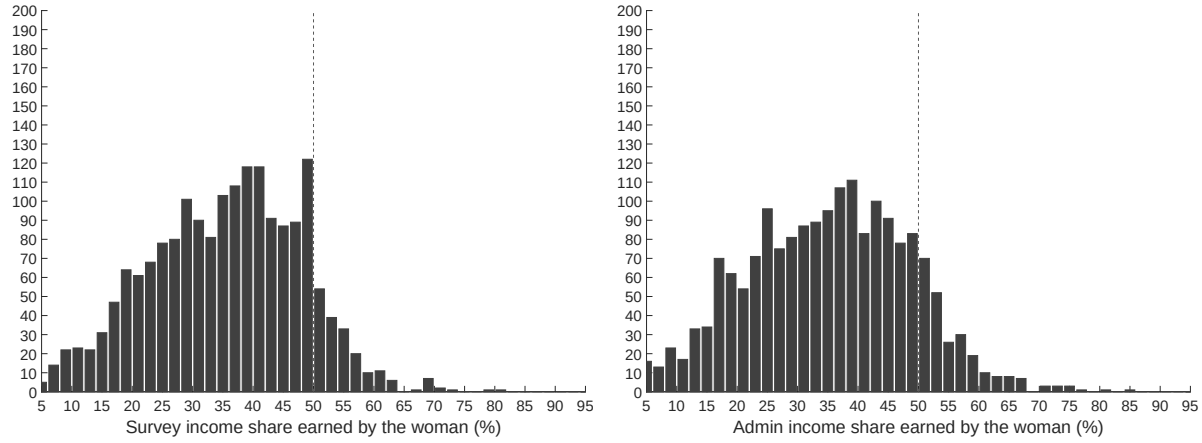


Figure 12: Pure survey and administrative distribution of female income shares for the very same couples in the SILC AT. The distributions are presented as raw histograms in 2 percent bins. The corresponding density discontinuity estimates can be found in Table A.3 in Appendix A.4

Figure 12 presents the resulting distributions of female income shares. Just as in the application for Switzerland, we observe a clear spike below the threshold in the survey data, but not in the administrative data. The corresponding discontinuity in density estimates are presented in Table A.3 in the Appendix. As before, we find a systematic discontinuity in survey data, but not in administrative data. This accentuates the finding that the spike is a survey artefact and is the result of survey misreporting in order to conform with the male breadwinner norm.

Figure 13 shows the survey distribution and the number of individuals who move themselves across the threshold through their survey response. As before, there is a considerable share of individuals whose actual earnings share is above 50 percent who report an earnings share below (red bars). Of the 1,815 couples in our sample 1,573

[43] When asked about their income, individuals are motivated to use their wage bill to provide an accurate measure.

[44] The results are virtually unaffected by this restriction.

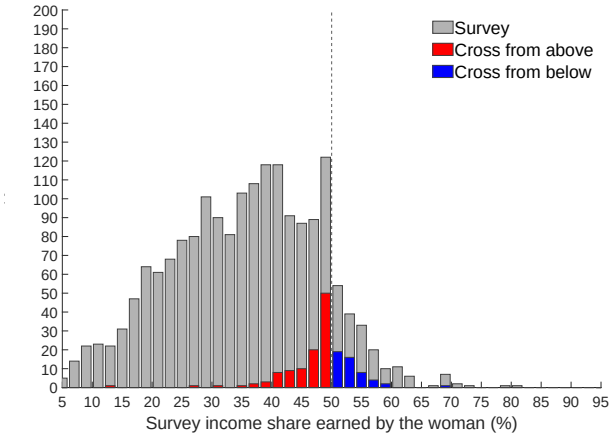


Figure 13: Histograms visualizing the number of couples finding themselves on a different side of the threshold when using survey vs. admin. data. The red bars show the number of couples where the woman outearns her partner based on administrative information but earns less or the same based on survey information (norm complying misreporters). The blue bars show the number of couples where the woman earns less or the same based on administrative data and outearns the partner based on survey data (norm non-complying misreporters). The distributions are presented as raw histograms of surveyed female income shares in 2 percent bins.

couples have a actual earnings share of below or equal to 50 percent. Only 50 (3.18 percent) of them cross the threshold though their survey response and place themselves above 50 percent in the surveyed female income share. As argued before, this is likely a result of random misreporting. Of the 242 couples who have an actual female income share of above 50 percent (where the women actually earns more), 106 (43.8 percent) report that they are positioned below the threshold. This very much mirrors the behavior we observed for Switzerland. The share is even higher in Austria compared to 34 percent in Switzerland.

Figure 14 visualizes the excess deviation of norm complying misreporters (the excess income deviation of norm complying misreporters compared to norm violators in the same bin, estimated by $\hat{\gamma}_k$ in Eq. 2 in Footnote [24]). It thus shows the estimated deviation in income misreporting between individuals in couples where the woman outearns her partner but who place themselves below the threshold compared to individuals in couples where the woman outearns her partner and who are willing to state this. Just as in the Swiss case, we observe systematic overreporting of male incomes and underreporting of female incomes. Again, the overreporting of male incomes seems to be the dominant channel.

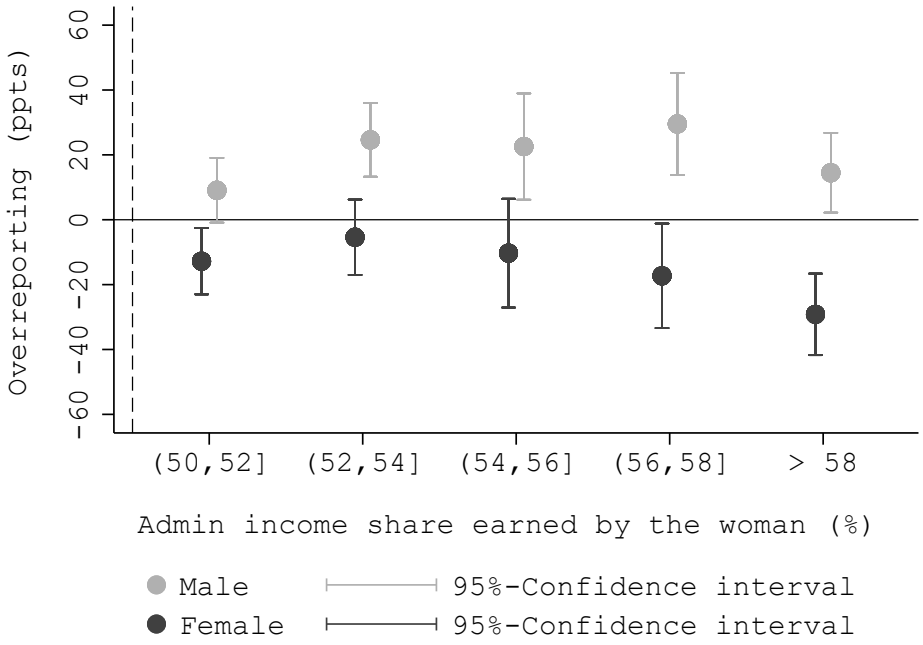


Figure 14: The figure shows excess misreporting of norm complying misreporters of male and female income compared to other individuals in the same bin, conditional on the administrative distribution. It shows for each bin whether individuals misreport their income to fall below the 50 percent threshold.

5.3 Bias in GWG estimates, AT

The Austrian data also allow us to re-evaluate the bias in the estimation of the GWG. As we observe administrative and surveyed incomes for all respondents in this data, and not only couples, we can provide GWG estimates on a more extensive sample.[45]

As above, we follow a typical GWG estimation by regressing log hourly wages on an indicator for female.[46] In the Austrian data, we observe a wide set of characteristics for all household members and can thus provide estimates controlling for the typically included characteristics. The Swiss data is limited in the information provided for the partner of the respondent. The presented estimates control for educational attainment, occupation, professional function, firm size, part-time employment, a continuous age variable and a constant.

The results are presented in Table 2. Specification (1) shows the *true* GWG based on administrative data and specification (2) the GWG based on survey information. While the true GWG, conditioning on all controls, amounts to 15.1 percent, the survey based GWG amounts to 18.3 percent. This results in an overestimation of 3.2 percentage points or 21 percent.

[45]We use similar sample restrictions as described above. The exception being that we do not restrict the sample to individuals who have a partner whose information is provided in the survey.

[46]Log hourly wages are calculated as monthly income divided by predicted monthly hours. Monthly hours are calculated based on the assumption that a month has 21 working days and using the total sum of reported weekly working hours (P030000).

Table 2: Gender wage gap estimates: SILC AT

	GWG admin.	GWG survey
	(1)	(2)
Female	-0.151***	-0.183***
	(0.011)	(0.009)
Age	Yes	Yes
Education	Yes	Yes
Occupation	Yes	Yes
Function	Yes	Yes
Firm size	Yes	Yes
Part time	Yes	Yes
Constant	Yes	Yes
N	11,894	11,894
R-squared	0.413	0.443
Overestimation		21 %

Notes: OLS estimates of the gender wage gap regressing log hourly income on indicator variables for female, education, occupation, professional function, firm size, part time employment, a continuous age variable, and a constant. The two specifications distinguish between the data source (survey or admin.) used for the estimation. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Conclusion

A growing literature in economics strives to understand whether gender norms and roles drive economic agents' behavior. Due to the lack of high quality administrative data, survey data has been the main data base available to learn about these questions for a long time. If, however, (gender) norms strongly impact individuals' response behavior, survey information might be less informative about behaviour than is commonly assumed.

In the first part of our analysis, we draw on Swiss data combining surveyed and administrative information on incomes for the same couples. Comparing surveyed and administrative incomes, we demonstrate that individuals' survey responses are strikingly prone to the influence of social norms. We exploit the fact that the *male breadwinner norm* marks an unwritten border distinguishing between norm compliance and norm violation and document that it strongly affects individuals' survey responses about their earnings. While the distribution of female income shares based on survey data features excess mass just below and a distinct discontinuity at the point where the woman would outearn her partner, we do not see a discontinuity in the distribution of female income shares based on administrative information for the very same couples. Our analysis reveals that the excess mass is driven by couples in which the women in fact outearns her partner. This result demonstrates that survey data is prone to be influenced by social norms and studies based on survey data might reach false conclusions. Based on survey data, we would have concluded that the *male breadwinner norm* influences individuals' labor market decisions, even though this is not the case in our setting.

In the second part of the analysis, we explore potential consequences of the systematic underreporting of female and overreporting of male incomes for measures of gender (in)equality. Comparing GWG measures provided by the OECD and Eurostat, we find suggestive evidence for an overestimation of gender inequality when based on survey data. This evidence is corroborated when we compare GWG estimates based on the Swiss data for which we are able to compare the GWG estimates based on administrative and survey data for the same individuals. The GWG is overestimated in survey data (by 9.4 percent) and even more so if we were to use proxy information (by 13.5 percent), a standard practice in many population surveys. Moreover, the heterogeneity analysis reveals that the bias varies along several dimensions. As a result, even the comparison of gender (in)equality within the same study may be invalid.

Finally, we present evidence supporting the interpretation that our findings for Switzerland are not driven by a Swiss peculiarity nor the survey design. Some waves of the Austrian SILC also include surveyed and administrative incomes for the same couples. Drawing on this data, we are able to replicate our findings for Switzerland. A large share of individuals in couples where the women outearns her partner misreport their incomes and place themselves below the threshold marking the male breadwinner norm. Furthermore, the Austrian data allows us to replicate the analysis on GWG estimates based on survey and administrative information for a more common sample and conditioning on the most common characteristics. Again, we find that the survey based gender (in)equality measure is vastly overestimated (by 21 percent).

Overall, our results urge for caution in using survey data to assess individuals' behavior when that behavior is associated with social norms.

Acknowledgements

Michaela Slotwinski gratefully acknowledges financial support from the *NCCR on the move* financed by the Swiss National Science Foundation. We are grateful to the FSO, and particularly to Jonas Deplazes, for supporting our research and enabling us to connect the survey and the administrative information. We are further grateful to Felipe Carozzi, Marius Faber, Olle Folke, Patricia Funk, Matthias Krapf, Armando Meier, Georg Nöldeke, Reto Odermatt, Imran Rasul, Marcus Roller, Kurt Schmidheiny, Sebastian Siegloch, Sarah Smith, Alois Stutzer, Daniel Waldenström, and Conny Wunsch and the many seminar and conference participants for helpful comments.

Data Statement

The confidential individual-level data (SESAM) were obtained under contract Nr. 180262 from the Swiss Statistical Office. The data from the Swiss Structural Survey (Strukturerhebung) were obtained under contract Nr. 170334. Access is granted for scientific research projects after review of a detailed application.[47] The data from the SILC AT (Community Statistics on Income and Living Conditions Austria) can be accessed by sending a data request to Statistics Austria.[48]

[47] See https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/erhebungen/sesam.html

[48] See https://www.statistik.at/web$_{de}$/$frageboegen$/$private_{h}aushalte$/$eu_{s}ilc$/$index.html$

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the Origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Arulampalam, W., A. L. Booth, and M. L. Bryan (2007). Is there a glass ceiling over europe? exploring the gender pay gap across the wage distribution. *ILR Review 60*(2), 163–186.

Bertrand, M., E. Kamenica, and J. Pan (2015). Gender identity and relative income within households. *Quarterly Journal of Economics 130*(2), 571–614.

Binder, A. J. and D. Lam (2018). Is there a male breadwinner norm? The hazards of inferring preferences from marriage market outcomes. NBER Working Paper 24907.

Blau, F. D. and L. M. Kahn (1997). Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of Labor Economics 15*(1, Part 1), 1–42.

Blau, F. D. and L. M. Kahn (2000). Gender differences in pay. *Journal of Economic Perspectives 14*(4), 75–99.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Bursztyn, L., T. Fujiwara, and A. Pallais (2017). 'Acting wife': Marriage market incentives and labor market investments. *American Economic Review 107*(11), 3288–3319.

Cattaneo, M. D., M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *Stata Journal 18*(1), 2034–261.

Chernozhukov, V., I. Fernández-Val, and Y. Luo (2018). The sorted effects method: discovering heterogeneous effects beyond their averages. *Econometrica 86*(6), 1911–1938.

Conti, G. and S. Pudney (2011). Survey design and the analysis of satisfaction. *Review of Economics and Statistics 93*(3), 1087–1093.

Eagly, A. H. and W. Wood (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist 54*(6), 408.

Eriksson, K. H. and A. Stenberg (2015). Gender identity and relative income within households: Evidence from Sweden. IZA DP No. 9533, IZA.

Eurostat (2021). Gender pay gap in unadjusted form (sdg_05_20). Accessed: 2021-01-28.

Fernández, R. and A. Fogli (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics 1*(1), 146–77.

Fernández, R., A. Fogli, and C. Olivetti (2004). Mothers and sons: Preference formation and female labor force dynamics. *Quarterly Journal of Economics 119*(4), 1249–1299.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics 121*(2), 673–697.

Folke, O. and J. Rickne (2020). All the single ladies: Job promotions and the durability of marriage. *American Economic Journal: Applied Economics 12*(1), 260–87.

Fortin, N. M. (2005). Gender role attitudes and the labour-market outcomes of women across OECD countries. *Oxford Review of Economic Policy 21*(3), 416–438.

Funk, P. (2016). How accurate are surveyed preferences for public policies? Evidence from a unique institutional setup. *Review of Economics and Statistics 98*(3), 442–454.

Gil, J. and T. Mora (2011). The determinants of misreporting weight and height: The role of social norms. *Economics & Human Biology 9*(1), 78 – 91.

Giuliano, P. (2017). Gender: A historical perspective. In S. L. Averett, L. M. Argys, and S. D. Hoffman (Eds.), *The Oxford Handbook of Women and the Economy*, pp. 1–32. New York, NY: Oxford University Press.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American Economic Review 96*(2), 1–21.

Hariri, J. G. and D. D. Lassen (2017). Income and outcomes: Social desirability bias distorts measurements of the relationship between income and political behavior. *Public Opinion Quarterly 81*(2), 564–576.

ILOSTAT (2021). Gender wage gap by occupation (%). Accessed: 2021-01-26.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Lee, J. and S. Lee (2012). Does it matter who responded to the survey? Trends in the us gender earnings gap revisited. *ILR Review 65*(1), 148–160.

Lippmann, Q., A. Georgieff, and C. Senik (2020). Undoing gender with institutions: Lessons from the German division and reunification. *The Economic Journal 130*(629), 1445–1470.

Macpherson, D. A. and B. T. Hirsch (1995). Wages and gender composition: Why do women's jobs pay less? *Journal of Labor Economics 13*(3), 426–471.

Martinelli, C. and S. W. Parker (2009). Deception and misreporting in a social program. *Journal of the European Economic Association 7*(4), 886–908.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Meyer, B. D. and N. Mittag (2019). Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness, and holes in the safety net. *American Economic Journal: Applied Economics 11*(2), 176–204.

Murray-Close, M. and M. L. Heggeness (2018). Manning up and womaning down: How husbands and wives report their earnings when she earns more. SESHD Working Paper No. 2018-20.

OECD (2021). Gender wage gap (indicator). doi:10.1787/7cee77aa-en. Accessed: 2021-01-12.

Otsu, T., K.-l. Xu, and Y. Matsushita (2013). Estimation and inference of discontinuity in density. *Journal of Business & Economic Statistics 31*(4), 507–524.

Reynolds, J. and J. B. Wenger (2012). He said, she said: The gender wage gap according to self and proxy reports in the current population survey. *Social Science Research 41*(2), 392–411.

Schmidheiny, K. and M. Slotwinski (2018). Tax-induced mobility: Evidence from a foreigners' tax scheme in switzerland. *Journal of Public Economics*.

Sprengholz, M., A. Wieber, and E. Holst (2020). Gender identity and wives' labor market outcomes in West and East Germany between 1984 and 2016. *Socio-Economic Review mwaa048*.

Steiner, I. and P. Wanner (2015). Towards a new data set for the analysis of migration and integration in Switzerland. nccr – on the move, Working Paper No. 1, nccr – on the move.

Steinhauer, A. (2013). Identity, working moms, and childlessness: Evidence from Switzerland. Mimeo, University of Zurich.

Teso, E. (2019). The long-lerm effect of demographic shocks on the evolution of gender roles: Evidence from the transatlantic slave trade. *Journal of the European Economic Association 17*(2), 497–534.

Wieber, A. and E. Holst (2015). Gender identity and women's supply of labor and non-market work: Panel data evidence from Germany. IZA DP No. 9471.

Zinovyeva, N. and M. Tverdostup (2021). Gender identity, co-working spouses and relative income within households. forthcoming American Economic Journal: Applied Economics.

A Appendix

A.1 Data

A.2 Administrative data on the total population

In one additional sub-analysis, we assess whether the overall administrative distribution of female income shares also exhibits a discontinuity at the 50 percent margin as well as a spike at exactly 50 percent for the universe of Swiss couples. To this end, we draw on administrative information for all households in Switzerland in 2014.[49] The data do not include an identifier for the spouse. However, the fact that we observe a household identifier, the marital status, and the date the marital status last changed allows us to get a close approximation of couples. We define individuals to be a married couple if they live in the same household (same address) and share the same date of change of marital status.[50] This should render a good approximation as it is unlikely that two married individuals live in the same household and share the same wedding date, but are not married to each other. The resulting within-couple age difference is similar in the survey and the administrative data. Men are on average about two years older than their wife. This supports our approach of matching spouses. We restrict the sample to Swiss citizens and individuals holding a permanent residence permit and who are aged between 18 and below 65 in order to match the sample used for the main analysis. We further exclude individuals with any self-employed income. The income measure we use is the raw sum of gross incomes of an individual across all employments in the given year. This allows us to compute the administrative income share of married woman for the whole population.

A.3 Definition of an individual's ancestry

In order to gain a comprehensive picture of individuals' migration background, we exploit information on the respondent's nationality and their parents' country of birth. Information on the individuals' nationality or their second nationality if they hold dual citizenship (Swiss and any foreign country) have been part of the regular SAKE survey since 2003, which means this information is available for all years we use for the analysis. We have a sample of 664 individuals with information on migration background. The information on an individual's background is supplemented with data on norms persisting in their country of ancestry, as reflected in the World Value Surveys.[51] We use average approval rates of full-time or part-time employed women with the statement "When jobs are scarce, men should have more right to a job than women" as a measure of gender norms. Norms

[49] These data combine the population register (Statpop) with income data from the social security register (ZAS by the Swiss Central Compensation Office). It was composed by Steiner and Wanner (2015).

[50] The date of marital status in our data source is available for all individuals who have changed their marital status within the last 40 years. We might thus lose observations of older cohorts. However, this information was not centrally registered before. Furthermore, we restrict the data to households with no more than 20 individuals to reduce the possible false positive rate.

[51] Data link: http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp.

for the countries Bosnia and Herzegovina, Kosovo, Serbia and Montenegro and Central Serbia are proxied by values for Serbia.[52] Based on this information, we generate a binary variable indicating whether an individual's country of origin shows average approval rates above or equal to the sample median (gender equal), or below the sample median (gender unequal). This results in two types of origin countries:

- **Equal countries:** Australia, Brazil, Canada, Finland, Germany, Italy, Montenegro, Netherlands, Peru, Slovenia, Spain, Sweden, United States
- **Unequal countries:** Albania, Algeria, Argentina, Belarus, Bosnia and Herzegovina, China, Colombia, Croatia, Ecuador, France, Hungary, India, Japan, Kosovo, Kyrgyzstan, Latvia, Lithuania, Macedonia, Pakistan, Poland, Romania, Russia, Serbia, Slovakia, South Korea, Thailand, Turkey, Ukraine, Venezuela, Vietnam

The within sample median is set by Germany (represents 28 percent of the sample) where 12 percent of surveyed women who work either full- or part-time agree with the statement that "When jobs are scarce, men should have more right to a job than women." The largest fraction of individuals with a background in a gender *equal* country is from Italy (31 percent) and the largest fraction of individuals from a gender *unequal* country is from France (11 percent).

[52] Country names defined by the BFS:
www.bfs.admin.ch/bfs/de/home/grundlagen/stgb.assetdetail.6166613.html.

A.4 Tables & Figures

Discontinuity in density estimates

In the following two tables with density discontinuity estimates following the empirical likelihood-based test by Otsu et al. (2013), c refers to the threshold used (the lowest value of female income shares exceeding 50 percent in order to only fit the density between realized values), h refers to the bandwidth, $\hat{f}_l$ reports the fit of the density coming from the left and $\hat{f}_r$ coming from the right respectively, $\hat{\theta}$ is the estimate of the discontinuity, $\hat{lr}$ is the value of the local likelihood ratio statistic under the null[53], and $\hat{f}_l/\hat{f}_r$ measures the relative size of the discontinuity.

Table A.1: Density discontinuity estimates

	c	h	$\hat{f}_l$	$\hat{f}_r$	$\hat{\theta}$	$\hat{lr}$	$\hat{f}_l/\hat{f}_r$	p-value	N
Large sample:									
Survey overall									
(1)	50.0905	7	0.055 5	0.013 8	−0.041 7	332.475 6	4.033 5	0.000 0	13,068
Survey female									
(2)	50.0905	7	0.063 9	0.016 2	−0.047 7	177.635 9	3.945 1	0.000 0	6,037
Survey male									
(3)	50.0952	7	0.048 2	0.011 7	−0.036 5	153.526 5	4.116 9	0.000 0	7,031
Administrative overall									
(4)	50.0003	7	0.013 7	0.011 5	−0.002 2	163.873 0	1.193 6	0.000 0	647,664
Small sample:									
Survey									
(5)	50.2203	7	0.048 1	0.015 6	−0.032 5	50.856 1	3.074 0	0.000 0	3,081
Administrative									
(6)	50.0225	7	0.021 4	0.023 7	0.002 2	0.437 3	0.905 3	0.508 4	3,081

Notes: Rows (1) to (3) present local likelihood ratio results for the discontinuity in the distribution of females' surveyed income shares (based on SAKE survey years 2002, 2005, 2008, 2012, and 2015). Row (4) presents the discontinuity estimate in the distribution of administrative female income shares in the full population of married couples which is described in Section A.2 of the appendix. Row (5) presents the local likelihood ratio results for the survey information and row (6) for the administrative information in our main sample including administrative and survey information for the very same couples (2012 and 2015). N stands for the number of observations with regard to the observations available for estimating the whole density in the respective sample.

[53]The null hypothesis $H_0 : \theta_0 = \theta$ for some θ can be tested by $lr(\theta)$ using $\chi^2(1)$ critical values. We test against $H_0 : \theta_0 = 0$.

Table A.2: Density discontinuity estimates, half of bandwidth

	c	h	$\hat{f}_l$	$\hat{f}_r$	$\hat{\theta}$	$\hat{l}r$	$\hat{f}_l/\hat{f}_r$	p-value	N
Large sample:									
Survey overall									
(1)	50.0905	3.5	0.116 1	0.012 3	−0.103 8	493.275 2	9.448 3	0.000 0	13,068
Survey female									
(2)	50.0905	3.5	0.123 4	0.014 5	−0.108 9	244.763 8	8.526 5	0.000 0	6,037
Survey male									
(3)	50.0952	3.5	0.111 3	0.010 5	−0.100 8	247.149 5	10.593 8	0.000 0	7,031
Administrative overall									
(4)	50.0003	3.5	0.015 9	0.011 5	−0.004 4	282.564 3	1.385 7	0.000 0	647,664
Small sample:									
Survey									
(5)	50.2203	3.5	0.095 3	0.014 8	−0.080 5	78.057 2	6.421 8	0.000 0	3,081
Administrative									
(6)	50.0225	3.5	0.020 7	0.023 9	0.003 2	0.485 5	0.865 7	0.485 9	3,081

Notes: Rows (1) to (3) present local likelihood ratio results for the discontinuity in the distribution of females' surveyed income shares (based on SAKE survey years 2002, 2005, 2008, 2012, and 2015). Row (4) presents the discontinuity estimate in the distribution of administrative female income shares in the full population of married couples which is described in Section A.2 of the appendix. Row (5) presents the local likelihood ratio results for the survey information and row (6) for the administrative information in our main sample including administrative and survey information for the very same couples (2012 and 2015). N stands for the number of observations with regard to the observations available for estimating the whole density in the respective sample.

Table A.3: Density discontinuity estimates: SILC AT

	c	h	$\hat{f}_l$	$\hat{f}_r$	$\hat{\theta}$	$\hat{l}r$	$\hat{f}_l/\hat{f}_r$	p-value	N
Survey									
(1)	50.0176	7.0000	0.041 1	0.017 3	−0.023 8	18.260 7	2.374 6	0.000 0	1,815
Administrative									
(6)	50.0010	7.0000	0.022 2	0.024 0	0.001 9	0.166 3	0.922 6	0.683 4	1,815

Notes: This table presents local likelihood ratio results for the discontinuity in the distribution of females' surveyed income shares (based on the SILC Austria). N stands for the number of observations with regard to the observations available for estimating the whole density in the respective sample.

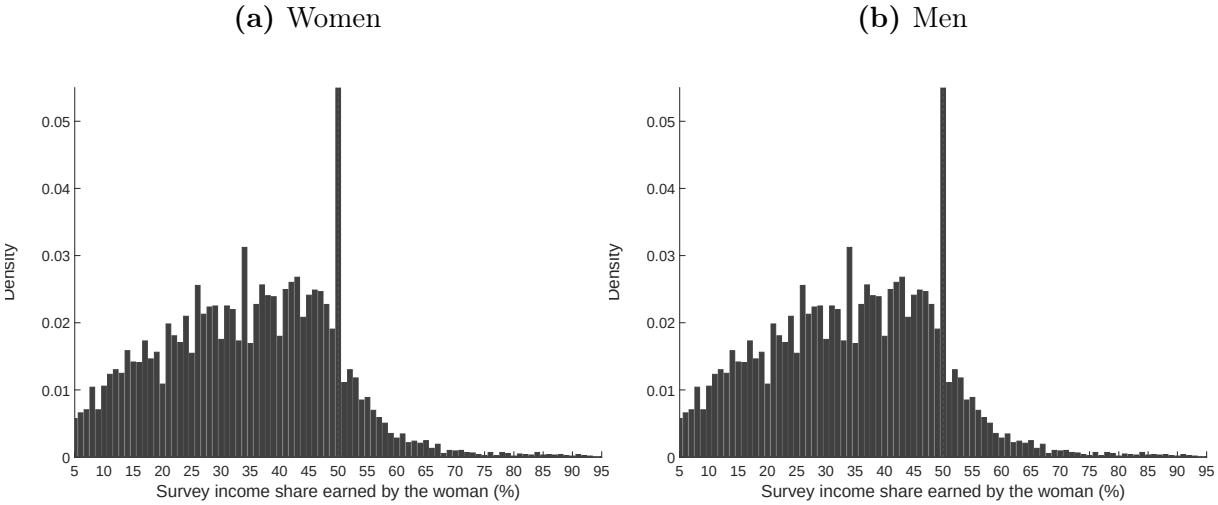


Figure A.1: Overall income share earned by the woman in a couple separately for female and male respondents. The shaded area represents the histogram of the underlying data in 1 percent bins. The corresponding density discontinuity estimates can be found in rows (2) and (3) of Table A.1.

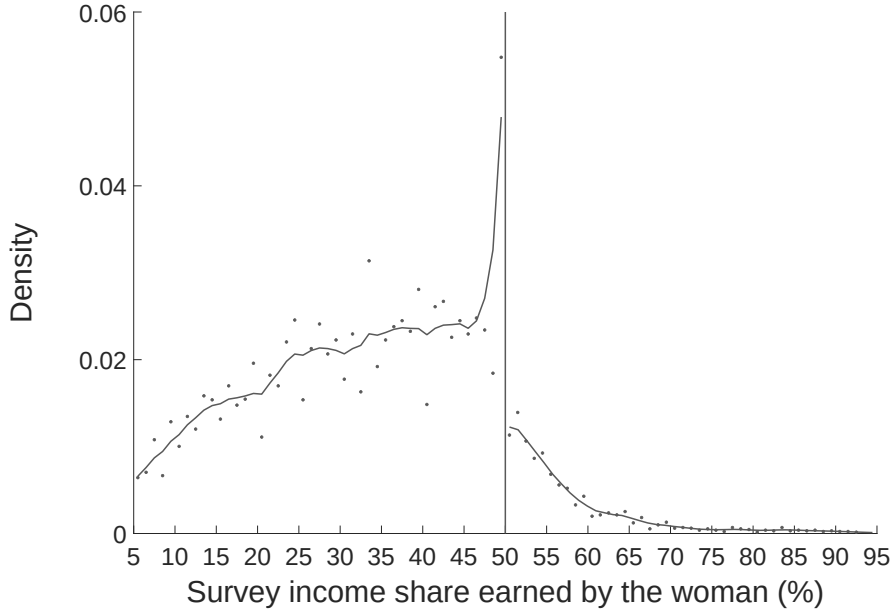


Figure A.2: Overall distribution of surveyed female income shares estimated using smoothed binned counts as proposed in McCrary (2008). The smooth is a local linear smooth using a triangular kernel and a bandwidth of 3.5%. The bin size is 1%. The automated procedure in McCrary (2008) proposes a bin size of 0.262 and a bandwidth of 13.11. It renders a point estimate of log difference in height of 0.556 with a p-value of 0.049. The manipulation test provided in the rddensity package in STATA (Cattaneo et al., 2018) proposes an optimal bandwidth of about 4 percent and also indicates a clear discontinuity with p-values<0.01.

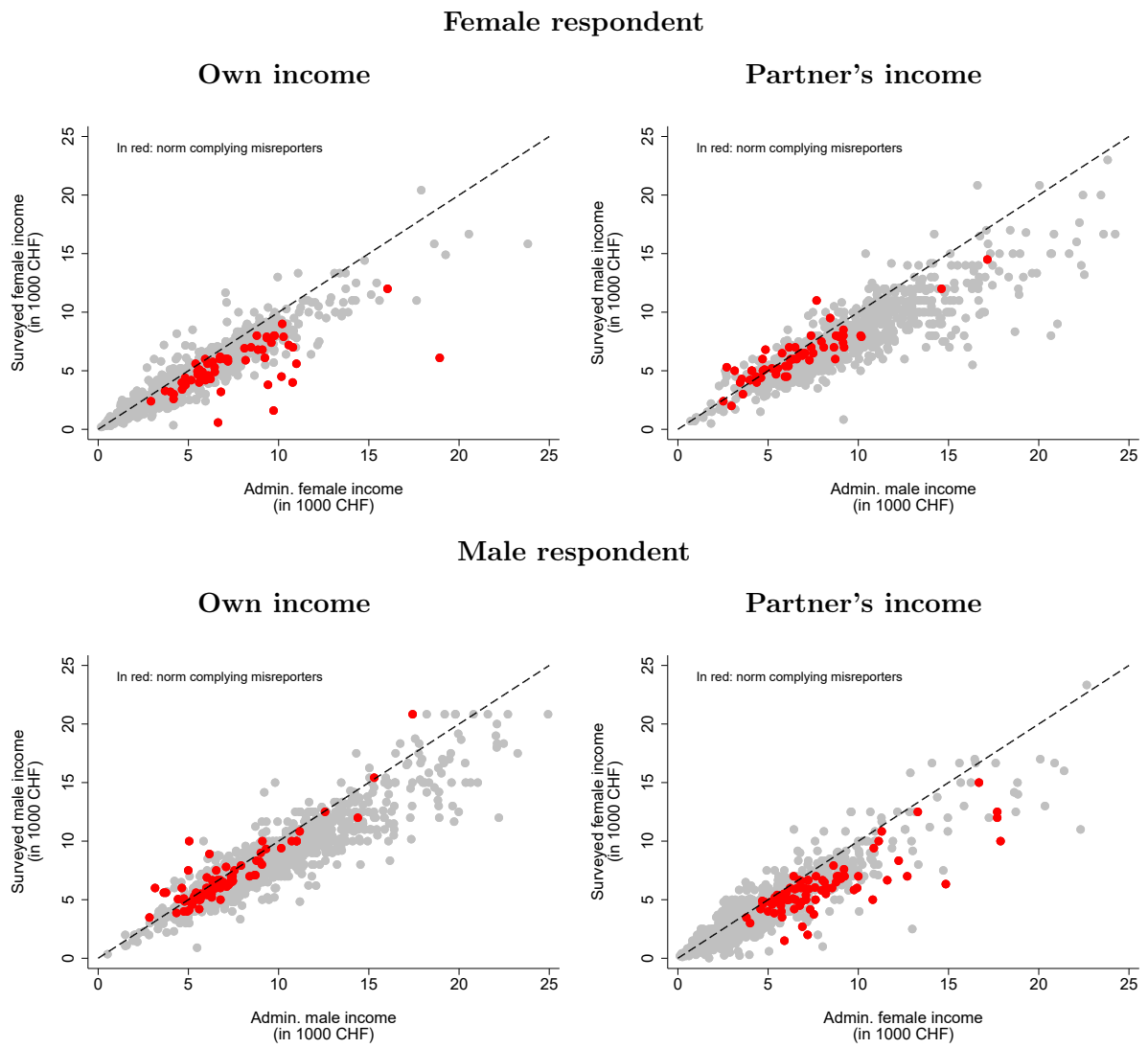


Figure A.3: Correlation between surveyed and administrative incomes for female and male respondents and for their own and their partner's income in the main sample. The dashed line shows the 45° line where surveyed and administrative incomes coincide. The red dots mark individuals who cross the threshold of just above 50 percent from above to conform with the norm that the women should not earn more than her partner (norm complying misreporters). For this correlation, we exclude 7 observations with monthly incomes of more than CHF 25,000 as these would distort this picture.

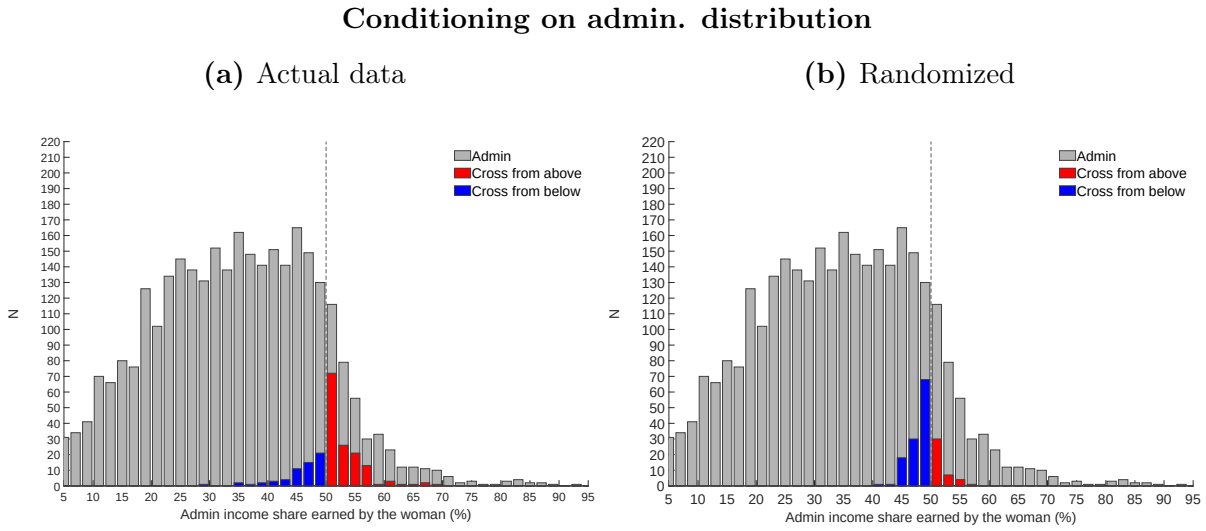


Figure A.4: Histograms visualizing the number of couples finding themselves on a different side of the threshold when using survey vs. admin. data. The red bars show the number of couples where the woman outearns her partner based on administrative information but earns less or the same based on survey information (norm complying misreporters). The blue bars show the number of couples where the woman earns less or the same based on administrative data and outearns the partner based on survey data (norm non-complying misreporters). The distributions are presented as raw histograms in 2 percent bins. The histogram to the left shows the distribution of administrative female income shares and the histogram to the right visualizes the distribution of income shares based on administrative data with random misreporting (in gray).

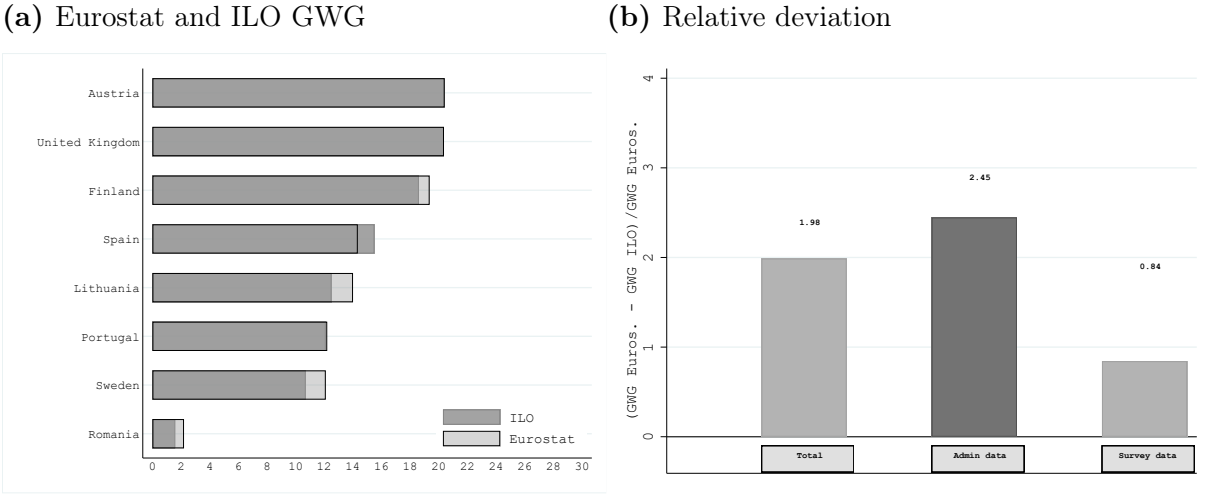


Figure A.5: Comparison of the average GWG provided by Eurostat and ILO by country. The left panel shows the average gwg provided by both institutions. The data covers all countries covered in both sources and the years 2002, 2006, 2010, 2014, and 2018. The right panel shows the relative deviation between Eurostat and ILO relative to the Eurostat number.

Misreporting estimates

Table A.4: Misreporting estimates, conditional on admin. distribution

Dependent variable	Diff. female inc. share	Overreporting (*Survey inc. – Admin. inc.*)			
		Female respondent		Male respondent	
		Own inc.	Partner inc.	Own inc.	Partner inc.
	(1)	(2)	(3)	(4)	(5)
P(0,40]	0.960**	1.816	-2.594	-3.821***	1.819
	(0.433)	(1.861)	(1.662)	(1.455)	(2.563)
P(40,42]	0.422	-1.485	-3.040	-2.752	-1.928
	(0.603)	(2.619)	(2.339)	(1.996)	(3.515)
P(42,44]	-0.008	1.903	0.581	-0.765	-1.070
	(0.613)	(2.672)	(2.386)	(2.023)	(3.562)
P(44,46]	Ref.	Ref.	Ref.	Ref.	Ref.
P(46,48]	-0.106	-2.502	1.088	-1.593	0.196
	(0.604)	(2.620)	(2.340)	(2.002)	(3.526)
P(48,50]	-1.597**	-2.496	5.603**	-2.263	-4.514
	(0.627)	(2.611)	(2.332)	(2.173)	(3.827)
P(50,52]	2.480***	2.723	-0.977	-6.616*	5.494
	(0.907)	(3.523)	(3.146)	(3.523)	(6.203)
P(52,54]	3.005***	1.874	-2.759	-7.735**	5.182
	(0.843)	(3.358)	(2.999)	(3.132)	(5.514)
P(54,56]	0.511	-1.145	-0.846	-0.221	1.059
	(0.994)	(4.251)	(3.796)	(3.344)	(5.887)
P(56,58]	0.131	-8.507	2.372	-13.944***	1.130
	(1.362)	(5.267)	(4.704)	(5.295)	(9.324)
P(58,60]	-1.297**	-4.523*	6.415***	-0.309	-2.443
	(0.644)	(2.556)	(2.283)	(2.449)	(4.311)
P(50,52] × norm complying misrep.	-5.169***	-7.455*	8.514**	8.986**	-8.796
	(1.022)	(4.274)	(3.817)	(3.763)	(6.626)
P(52,54] × norm complying misrep.	-8.669***	-16.711***	9.047*	14.682***	-10.216
	(1.279)	(5.714)	(5.103)	(4.271)	(7.521)
P(54,56] × norm complying misrep.	-12.145***	-10.173	35.350***	13.565***	-21.210***
	(1.475)	(7.303)	(6.523)	(4.516)	(7.952)
P(56,58] × norm complying misrep.	-10.027***	-4.443	27.851***	39.509***	-5.461
	(1.970)	(7.639)	(6.822)	(7.577)	(13.342)
P(58,60] × norm complying misrep.	-16.874***	-36.888***	15.183*	23.868***	-22.407**
	(1.848)	(9.644)	(8.614)	(5.492)	(9.670)
Constant	-0.770*	-10.289***	-12.735***	-7.556***	-14.538***
	(0.425)	(1.814)	(1.621)	(1.430)	(2.519)
Inc. mode	Yes	Yes	Yes	Yes	Yes
N	3,081	1,415	1,415	1,666	1,666
R-squared	0.115	0.150	0.144	0.182	0.055

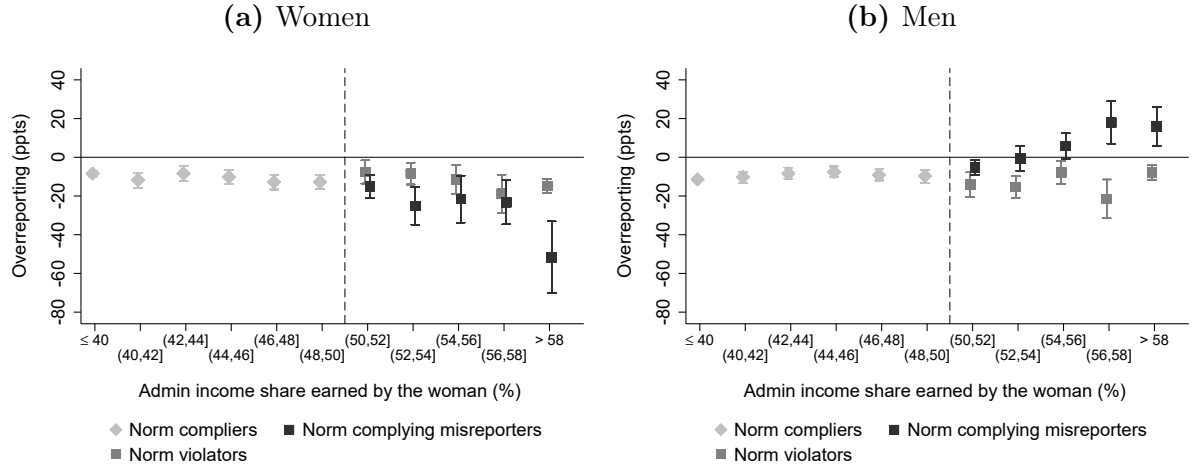
Notes: Misreporting of female income shares and incomes by bin of the admin. female income share. We distinguish between misreporting of norm compliers (admin female inc. share $\leq$ 50 percent), norm violators (admin female inc. share $>$ 50 percent, surveyed female inc. share $>$ 50 percent), and norm complying misreporters (admin female inc. share $>$ 50 percent, surveyed female inc. share $\leq$ 50 percent). The interaction term between the bins above 50 percent and the indicator for norm complying misreporters shows the additional deviation for this group. In column (1), the dependent variable is the deviation in the female income share between survey and admin. data (*surveyed female inc. share - admin female inc. share*). For columns (2) to (5), the dependent variable is misreporting, as defined above. We use the specification described in Eq. 2 and standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Misreporting estimates, conditional on survey distribution

Dependent variable	Diff. female inc. share	Overreporting (*Survey inc. − Admin. inc.*)			
		Female respondent		Male respondent	
		Own inc.	Partner inc.	Own inc.	Partner inc.
	(1)	(2)	(3)	(4)	(5)
P(0,40]	-2.499***	-1.158	0.989	1.732	-11.782***
	(0.456)	(2.027)	(1.858)	(1.553)	(2.698)
P(40,42]	-0.756	1.177	-1.320	0.941	-7.193*
	(0.635)	(2.851)	(2.613)	(2.137)	(3.713)
P(42,44]	-0.381	2.779	1.845	1.859	-0.407
	(0.625)	(2.788)	(2.556)	(2.112)	(3.669)
P(44,46]	Ref.	Ref.	Ref.	Ref.	Ref.
P(46,48]	0.193	3.037	0.729	-1.061	-4.136
	(0.621)	(2.638)	(2.419)	(2.235)	(3.882)
P(48,50]	1.511**	3.824	-2.845	-0.786	1.510
	(0.641)	(2.678)	(2.455)	(2.375)	(4.127)
P(50,52]	-0.074	2.239	0.360	-1.354	-6.602
	(0.747)	(3.178)	(2.913)	(2.670)	(4.639)
P(52,54]	-1.216	-2.083	0.667	2.433	-6.512
	(0.761)	(3.199)	(2.932)	(2.762)	(4.798)
P(54,56]	-1.652**	-2.274	6.876**	1.163	-1.472
	(0.832)	(3.401)	(3.117)	(3.165)	(5.499)
P(56,58]	-0.926	1.163	2.403	4.555	-8.792
	(1.059)	(4.052)	(3.714)	(4.653)	(8.083)
P(58,60]	1.259*	0.441	-0.997	-3.917	-6.097
	(0.679)	(2.767)	(2.537)	(2.641)	(4.589)
P(0,40] × norm complying misrep.	-19.841***	-38.398***	27.148***	12.524***	-36.730***
	(1.229)	(5.794)	(5.312)	(3.956)	(6.873)
P(40,42] × norm complying misrep.	-17.456***	-47.065***	2.148	66.176***	-13.343
	(3.030)	(16.442)	(15.073)	(8.926)	(15.507)
P(42,44] × norm complying misrep.	-11.140***	-20.157***	21.889***	13.629**	-24.245***
	(1.627)	(7.569)	(6.938)	(5.294)	(9.197)
P(44,46] × norm complying misrep.	-10.193***	-18.059**	15.467**	32.336***	-10.206
	(1.887)	(8.381)	(7.684)	(6.405)	(11.128)
P(46,48] × norm complying misrep.	-7.592***	-6.950	16.815***	16.541***	-14.160**
	(1.241)	(6.038)	(5.535)	(3.958)	(6.877)
P(48,50] × norm complying misrep.	-6.258***	-7.066**	15.026***	8.049***	-13.894***
	(0.742)	(3.405)	(3.121)	(2.567)	(4.460)
Constant	1.419***	-9.816***	-14.162***	-12.113***	-4.171
	(0.448)	(1.979)	(1.814)	(1.529)	(2.656)
Inc. mode	Yes	Yes	Yes	Yes	Yes
N	3,081	1,415	1,415	1,666	1,666
R-squared	0.165	0.160	0.108	0.191	0.090

Notes: Misreporting of female income shares and incomes by bin of the surveyed female income share. We distinguish between misreporting of norm compliers (admin female inc. share $\leq$ 50 percent), norm violators (admin female inc. share $>$ 50 percent, surveyed female inc. share $>$ 50 percent), and norm complying misreporters (admin female inc. share $>$ 50 percent, surveyed female inc. share $\leq$ 50 percent). The interaction term between the bins below 50 percent and the indicator for norm complying misreporters shows the additional deviation for this group. Please note that the bins are defined conditional on the survey distribution and that norm complying misreporters are therefore placed *below* the 50 percent threshold. In column (1), the dependent variable is the deviation in the female income share between survey and admin. data (*surveyed female inc. share - admin female inc. share*). For columns (2) to (5), the dependent variable is misreporting, as defined above. We use the specification described in Eq. 2 and standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Income deviation in own incomes, conditional on administrative income share



Income deviation in partner incomes, conditional on administrative income share

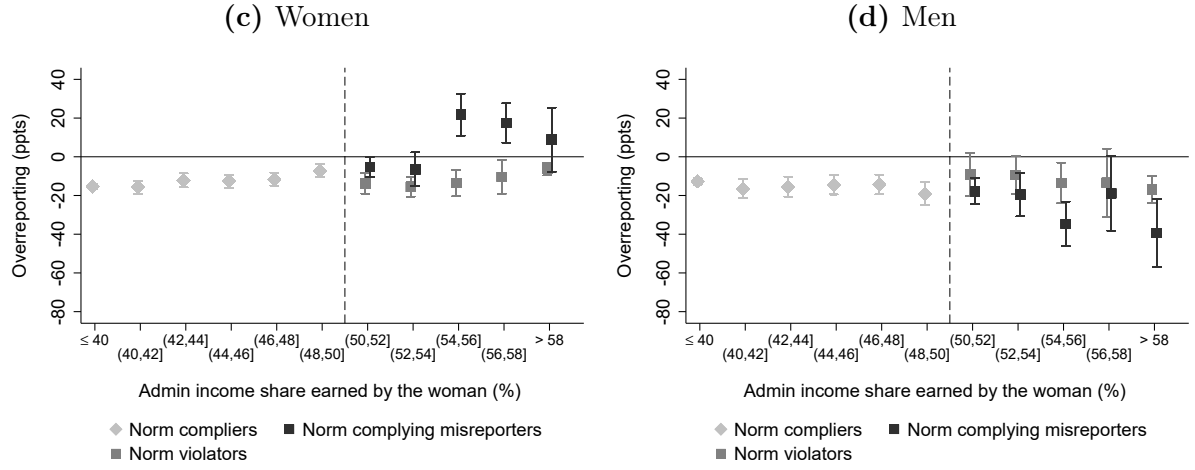


Figure A.6: These graphs show the level estimates of the deviation between surveyed and administrative incomes for norm compliers, norm violators, and norm complying misreporters conditioning on the position in the administrative distribution of female income shares. The upper panel shows the average deviations of own incomes for female (a) and male (b) respondents. The lower panel shows the average deviations of partner incomes for female (c) and male (d) respondents. 95 percent confidence bounds are displayed. We use the specification described in Eq. 2.

Table A.6: Socio-demographic characteristics of couples where the woman outearns her partner

	Mean	Mean violators	Mean compliers	Diff.	SE diff.	t-value	p-value	N
Age w / 10	3.97	3.97	3.97	-0.001	0.103	0.01	0.994	408
Age m / 10	4.21	4.18	4.26	0.080	0.109	-0.73	0.466	408
Inc. admin. hh	1.49	1.51	1.45	-0.056	0.069	0.80	0.423	408
Tert. educ. w	0.43	0.50	0.29	-0.205	0.051	4.03	0.000	405
Tert. educ. m	0.32	0.32	0.32	-0.003	0.049	0.06	0.950	405
Couple w/ children	0.37	0.38	0.35	-0.035	0.050	0.69	0.494	408

Notes: Descriptive statistics for couples where the woman outearns her partner in administrative terms, distinguishing between norm violators (admin. female income share $>$ 50 percent, surveyed female income share $>$ 50 percent) and norm complying misreporters (compliers, admin. female income share $>$ 50 percent, surveyed female income share $\leq$ 50 percent).

Table A.7: Norm proxies of couples where the woman outearns her partner

	Mean	Mean violators	Mean compliers	Diff.	SE diff.	t-value	p-value	N
Educ. w $\leq$ educ. m	0.76	0.71	0.87	0.154	0.044	-3.52	0.000	408
Hrs. w $\leq$ hrs. m	0.81	0.78	0.87	0.081	0.041	-2.00	0.046	405
Age diff. / 10	0.23	0.21	0.29	0.081	0.049	-1.64	0.101	408
German speaking	0.70	0.68	0.74	0.063	0.048	-1.33	0.186	408
Unequal country	0.30	0.27	0.36	0.090	0.095	-0.95	0.346	106

Notes: Descriptive statistics of norm proxies for couples where the woman outearns her partner in administrative terms, distinguishing between norm violators (admin. female income share $>$ 50 percent, surveyed female income share $>$ 50 percent) and norm complying misreporters (compliers, admin. female income share $>$ 50 percent, surveyed female income share $\leq$ 50 percent). *Hrs.* stand for weekly work hours. *Age diff.* describes the within-couple age difference defined as $(age_m - age_w)$. *German speaking* describes an indicator set to one for German speaking individuals (in relation to French, Italian, or Romansh speaking). *Unequal country* describes an indicator set to one for individuals with origins in a country with more traditional gender norms.

A.5 Survey procedure and questions

The income components are surveyed as follows: Close to the end of the standard SAKE questionnaire, the target person states her or his own income. Right after the standard questionnaire concludes, the special questionnaire starts. Consecutively, the respondents are asked whether they have a partner living in the same household. If yes, the follow up questions are whether the partner is employed and what is her or his income. The question about the partner's income is asked in exactly the same way and with the same options as the question about the target person's income. Thus, it is always the respondent's own income that is surveyed first before the focus changes to the partner's income. There are only a few questions in between the one on one's own and on the partner's income. It can thus be assumed that the respondent still has the response to the question about their own income present when asked about the partner's income.

```
77000 01
77000 02 Könnten Sie mir Ihren MONATSLOHN angeben ?
77000 03 Wenn's Ihnen leichter fällt, können Sie auch den Jahres- oder
77000 04 Stundenlohn angeben.
77000 05 ------------------------------------------------------------------------
77000 06
77000 07
77000 08       o LOHN ..............................<*******>
77000 09       - Arbeitet ohne Entlöhnung .........<0>
77000 10       -------------------------------------------
77000 11       - Weiss nicht .......................<X>
77000 12       - Keine Antwort .....................<Y>
77000 13
77000 14
77000 15                       *******
```

Figure A.7: Survey question on personal income. Survey question 77000 (variable IW04) is translated as: "Could you tell me your monthly salary? If it is easier for you, you may also tell me your yearly or hourly salary." There are four response options: 1. Salary (numeric) 2. Works without compensation 3. Don't know 4. No answer

.

```
77100 01
77100 02  ==> INT: Sind die angegebenen Fr. X ...
77100 03
77100 04  o BRUTTO (VOR Abzug der Sozialbeiträge):    - pro Monat .....<1>
77100 05                                              - pro Jahr ......<2>
77100 06                                              - pro Stunde ....<3>
77100 07  o NETTO (NACH Abzug der Sozialbeiträge):    - pro Monat .....<4>
77100 08                                              - pro Jahr ......<5>
77100 09                                              - pro Stunde ....<6>
77100 10
77100 11  o weiss nicht ..............................................<8>
77100 12  o keine Antwort ..............................................<9>
77100 13  o der angegebene Betrag von Fr. X ist falsch .......<0>
77100 14
77100 15
```

Figure A.8: Survey question on mode of income declaration. Survey question 77004 asks whether the reported income is net or gross; hourly, monthly, or yearly. The exact question is "Is the declared amount CHF X: 1. Gross (per month / per year / per hour), 2. Net (per month / per year / per hour)". Additional answer options are "3. "I don't know" 4. No answer 5. "The above amount of CHF X is wrong."

```
77900 01  Sie haben mir vorher gesagt, dass Ihre Ehe-/Lebenspartnerin
77900 02  erwerbstätig ist. Könnten Sie mir den MONATSLOHN von Ihrer Ehe-/
77900 03  Lebenspartnerin angeben?
77900 04  Wenn's Ihnen leichter fällt, können Sie auch den Jahres- oder
77900 05  Stundenlohn angeben.
77900 06  ------------------------------------------------------------------------
77900 07
77900 08
77900 09
77900 10        o MONATSLOHN ......................... <********>
77900 11        o Arbeitet ohne Entlöhnung .......... <0>
77900 12        ---------------------------------------------
77900 13        - Weiss nicht ....................... <X>
77900 14        - Keine Antwort ..................... <Y>
77900 15                         ********
```

Figure A.9: Survey question that defines the partner's income. Survey question 77900 (variable IW20) is translated as: "You have told me before that your spouse / partner is employed. Could you tell me the monthly salary of your spouse / partner? If it is easier for you, you may also tell me his / her yearly or hourly salary." There are four response options: 1. Salary (numeric) 2. Works without compensation 3. Don't know 4. No answer

.

77950	01	
77950	02	Sind die angegebenen#bFr. X.-#e...
77950	03	
77950	04	o BRUTTO (VOR Abzug der Sozialbeiträge): - pro Monat .....<1>
77950	05	- pro Jahr ......<2>
77950	06	- pro Stunde ....<3>
77950	07	o NETTO (NACH Abzug der Sozialbeiträge): - pro Monat .....<4>
77950	08	- pro Jahr ......<5>
77950	09	- pro Stunde ....<6>
77950	10	
77950	11	o weiss nicht .....<8>
77950	12	o keine Antwort .....<9>
77950	13	o der angegebene Betrag von Fr. X ist falsch .....<0>
77950	14	
77950	15	FORMAT !

Figure A.10: Survey question that specifies the mode of the income declaration for the partner's income. Survey question 77950 asks whether the reported partner income is net or gross; hourly, monthly, or yearly. The exact question is "Is the declared amount CHF X: 1. Gross (per month / per year / per hour), 2. Net (per month / per year / per hour)". Additional answer options are "3. "I don't know" 4. No answer 5. "The above amount of CHF X is wrong."