Revisiting the breadwinner norm: The effect of potential relative earnings on married women's labor supply*

Sarah Rosenberg†

Abstract

In light of recent research suggestive of a male breadwinner norm, I examine the relationship between female breadwinning and women's labor force participation for opposite-sex married couples in the U.S. I show that using measures based on observed income likely yields bias due to non-classical measurement error and omitted variables. Estimating the relationship with a measure for potential relative earnings unaffected by this bias gives a precise null relationship. These findings suggest aversion to female breadwinning does not impact decisions within marriage, but the results do not rule out possible effects earlier at the margin of marital sorting.

Keywords: Gender norms, female labor force participation

JEL: J16, J22

*I would like to thank seminar participants at Stockholm University, ECARES, the Institute for Fiscal Studies, the Society of Household Economics Conference, and the IZA World Labor Conference 2018, and the European Association for Labour Economists Conference for thoughtful comments and suggestions on an earlier version of this paper. I am grateful to Bram De Rock, Paula Gobbi, and three anonymous referees for insightful suggestions for improvement. I also want to thank Marianne Bertrand, Emir Kamenica, and Jessica Pan for sharing the code necessary to replicate their paper. My research is made possible by funding via a FRESH Grant from the F.N.R.S.

†Université Libre de Bruxelles, ECARES. Email: srosenbe@ulb.ac.be

1 Introduction

Following the dramatic changes in women's labor supply in the 20th century, the slowdown in women's convergence towards men's labor outcomes since the 1990s has remained a persistent question for labor economics (Blau and Kahn, 2007, 2017). Some suggest this slowdown might represent women reaching their "natural rate" of unemployment (Goldin, 2006). However, since the seminal work of Akerlof and Kranton (2000) translating the concept of social norms into a standard economics framework, economists have increasingly considered gender norms as an alternative explanation for why the gaps between men and women persist. This paper contributes towards that effort by ruling out one possible channel for gender norms: that an aversion to female breadwinning might cause women in opposite-sex marriages to decrease their labor force participation when they out-earn their husbands (or become more likely to).

When broadly defined, traditional gender norms are negatively related to female labor force participation (Fortin, 2005; Fernandez, 2007; Fortin, 2015; Goussé, Jacquemet, and Robin, 2017). However, traditional gender norms encompass a range of aspects. To understand how they affect female labor decisions, it's necessary to understand which norms matter and how they operate.

For example, norms about women's roles inside and outside the home clearly affect their labor force participation. Women working has become increasingly normalized over time, with evidence that one channel for this goes via mothers or mothers-in-law to their children (Fernández, Fogli, and Olivetti, 2004; Morrill and Morrill, 2013; Farré and Vella, 2007). Similarly, there is evidence that the "child penalty" for new mothers is also related to gender norms (Nix, Andresen, et al., 2019) and may likewise operate via inter-generational transmission of gender roles (Kleven, Landais, and Søgaard, 2019).

Norms about men may also matter for female labor force participation. The aim of this paper is to better understand the role of the "male breadwinner" norm. This norm is typically discussed in terms of two potentially salient features, that (a) husbands should earn more than their wives and/or (b) husbands should be employed in full-time paid work–or in other words, an aversion to female breadwinning and an aversion to men's lack of full-time employment. I focus on the relevance of this first point for married couples.[1]

In theory, each of the two main aspects of the male breadwinner norm can affect the labor force participation of married women at four margins: by influencing marital

[1] In recent decades, marriage rates have declined and cohabitation has risen substantially. In light of this shift, it would also be valuable to understand the norm's relevance for cohabitating couples. However, since a primary aim of this paper is to understand whether conclusions in earlier research about the norm are valid, and this earlier work is overwhelmingly focused only on married couples (Foster and Stratton (2021) are an exception), I limit the focus of the present paper to married couples.

sorting, by influencing a couple's initial decisions about work, by labor supply changes in response to potential or actual violations of the breadwinner norm within marriage, and by separation or divorce.[2] We can think about these different margins as decisions that occur at the time of marriage or decisions that occur after marriage.

Most related work in sociology and economics studies either the effects of wives' relative earnings or the effects of men's (un)employment at a particular margin, on housework, labor supply, or divorce. (Sec. 2 provides a detailed review.) Work on wives' relative earnings and men's unemployment seem at first glance to map cleanly to the concept of aversion to female breadwinning on the one hand, and aversion to men's lack of full-time employment on the other. However, studies using relative earnings may conflate these two aspects, among other problems, because observed earnings are affected by both endogenous and exogenous changes to men's employment.

To see why, let us examine what we mean when we hypothesize that an aversion to female breadwinning might lead wives to work less, do more housework, or become more likely to divorce. Consider interpretations of results from Bertrand, Kamenica, and Pan (2015) (hereafter BKP), the only paper to quantitatively study the relationship between measures of female breadwinning and wives' labor force participation: Their findings on labor have been described as indicating that "in couples in which the wife's potential income is likely to exceed her husband's (based on predicted income), the wife is less likely to be in the labor force," (Blau and Kahn, 2017) and as demonstrating that "the evidence is compelling that wives cut back on their labor force participation to avoid outearning their husbands" (Schwartz and Goñalons-Pons, 2016). BKP acknowledge explicitly that their analysis is not causal, and yet we readers are making a causal interpretation about what couples' responses would be if a wife's relative earnings exogenously increased such that she became the breadwinner.

In fact, this interpretation is sensible: it is precisely the question we need to answer to draw conclusions about the effects of an aversion to female breadwinning. The key contribution of this paper is to show that estimates based on measures of female breadwinning that are constructed with observed earnings are unlikely to be useful for answering this question.

Following a literature review and discussion of data sources, the first major section of analysis lays out the theoretical and empirical issues with using observed earnings. The starting point is that female breadwinning is not solely caused by exogenous changes to wives' and husbands' earnings. In other words, observed female breadwinning is the sum of female breadwinning that occurs due to exogenous changes in wives' and husbands' earnings (our identifying variation of interest) and other factors. If these

[2]Marital dissolution would affect the labor force participation of married women indirectly, by affecting the composition of women who remain married.

other factors were unrelated to all other variables in the model, we would simply have attenuation bias. However, if the other situations under which female breadwinning occurs are themselves related to the dependent variable, we suddenly have non-classical measurement error, and bias both towards and away from zero are possible.

Controls may not be sufficient eliminate this bias, particularly if unobserved endogenous choices by couples are important in determining female breadwinning. Couple fixed effects would solve it only if the other sources of variation in female breadwinning were time-invariant. One valid alternative solution would be to use a proxy for the likelihood of female breadwinning constructed from group-level averages. A limitation of this solution is it could also introduce a new bias, if characteristics used to define a group are correlated with unobservables that are also related to the outcome of interest, such as if education is one group characteristic, and if ability or tastes for work are related to both educational attainment and labor force participation. However, in this case, the unobservables are likely time-invariant and would thus be addressed by the inclusion of couple fixed effects.

To assess whether these conditions that would yield bias are plausible, I examine descriptive features of female breadwinning. I show that low levels of work for husbands appear to be a major factor coinciding with female breadwinning. These low levels of work owe partially to layoffs, but there also appears to be substantial evidence of endogenous choices as a factor, including husbands that quit their jobs or work in seasonal or temporary jobs, or simply working less than full-time in a given year without evidence of having been unemployed. Moreover, I show that low levels of work for husbands are related to lower levels of work and labor force participation for wives.[3] Finally, I show that the low levels of work for husbands associated with female breadwinning are indeed time-varying, with husbands' hours declining dramatically in the two to three years prior to the occurrence of female breadwinning, and returning almost to initial levels three years later.

Since these facts fulfill the theoretical requirements for bias to arise, in the next section I show that this bias matters in practice for existing results. Since BKP study labor force participation, housework, as well as divorce, their paper provides a useful frame for comparison that is also relevant for past work studying only one of these three outcomes. I re-estimate results from BKP using both cross-sectional and longitudinal data on labor force participation, housework, and divorce, with adapted measures of the likelihood of female breadwinning that are based on predicted earnings for both husbands and wives. As they do, I use the Census and American Community Survey,

[3]To be clear, this is not a claim that wives' work less in response to exogenous negative shocks to men's employment, which is the subject of the added worker effect, reviewed in Sec. 2. This correlation also reflects endogenous choices of couples about work.

the American Time Use Survey, and longitudinal data from the Panel Study on Income Dynamics. Notably, while BKP find negative estimates for the relationship between (likely) female breadwinning and wives' labor force participation, and positive estimates for divorce and wives' relative housework, using the adapted measures causes the estimates to flip signs in the cross-sectional analysis and yield precise null effects in the longitudinal analysis. I assess whether these null effects in the longitudinal analysis owe to how the predicted measure is constructed or to heterogeneity over time, but these sensitivity analyses yield similar results.

This comparative exercise supports the interpretation that estimates based on measures with observed earnings are likely biased relative to the true effect of female breadwinning owing to exogenous changes in wives and husbands earnings–otherwise we should not see such large changes in the sign or magnitude of the coefficients. However, do the null effects when using the predicted measure of female breadwinning really tell us that aversion to female breadwinning has effectively zero impact on outcomes in marriage?

To answer this question, I carry out a final analysis to show that aggregate changes in wages captured by the predicted measure rarely reverses who in the couple is the breadwinner, after accounting for average within-couple earnings differences. While it might seem like this implies the measure is underpowered to answer our question of interest, in fact, it highlights that this hypothesized mechanism–where wives decrease their labor force participation to avoid earning more than their husbands–mechanically cannot be an important contributor to gender inequality in labor outcomes. In short, the precipitating causal factor rarely occurs.

We could perhaps imagine trying to instrument our way out of the problem by searching for a large exogenous shock to relative earnings. Folk and Rickne (2019) perhaps come closest to this, by comparing politicians that narrowly win or lose elections, and thereby receive major promotions and pay increases–but in a sense, the specificity of such a shock underscores the prior point: such shocks are uncommon, yet large differences in husband-wife earnings and labor force participation are not.

In turn, these findings point to the importance of understanding the initial determination of within-couple gaps in earnings: the sorting process. While this paper contributes to the literature studying gender norms and female labor force participation by ruling out one margin for the effect of an aversion to female breadwinning on wives' labor force participation on labor decisions within marriage, it does not necessarily mean that the male breadwinner norm is unimportant. It may still be possible that such sentiments affect marital sorting. This question deserves further study.

More broadly, it contributes to the range of work using relative earnings or observed female breadwinning to study the impact of an aversion to female breadwinning on

housework and divorce, and highlights the importance of considering that estimated effects may be picking up other underlying relationships than that of interest. To be clear, it may still be interesting and useful to compare couples with female breadwinners and couples without, but we must take care in considering the underlying identifying variation in making an interpretation of such results.

The paper proceeds as follows: Sec. 2 reviews related research to motivate the hypothesis for an aversion to female breadwinning and provide context for the ways the male breadwinner norm has been tested. Sec. 3 describes the data. Sec. 4 theoretically and empirically outlines the problems with using observed earnings to construct measures of female breadwinning. Sec. 5 tests the practical relevance of these problems by re-estimating results from BKP using different measures of female breadwinning, varying whether they use observed or predicted earnings, and discusses the implications of the resulting differences. Finally, Sec. 6 concludes.

2 Literature Review

In this section, I discuss related research to provide context and motivation for the contribution of this paper. First, I review the existing work that claims to find evidence of a norm against or aversion to female breadwinning. Notably, most of these studies use regression analyses with measures of observed relative earnings or female breadwinning as the key variables of interest. I then discuss whether there is evidence beyond this type of analysis that suggests aversion to female breadwinning is likely to have an impact on labor force participation. While there is indeed some general evidence that people prefer wives not to be higher status, when considering the existing evidence on factors and characteristics related to female breadwinning, findings suggest that female breadwinning is not typically a "high-status" condition, but is instead strongly correlated to illness and unemployment for husbands. Moreover, research on the effects of husbands' unemployment suggest that it can itself affect divorce, the division of housework, and wives' labor force participation. These features of female breadwinning raise the question of how they affect the interpretation of the existing results, which this paper sets out to answer.

2.1 Evidence on aversion to female breadwinning

Researchers have attempted to infer the existence of an aversion to female breadwinning using three primary methods: First, by studying the distribution of relative income among married couples; second (and most common), by relating observed relative earnings to outcomes such as household work, divorce, or labor supply; and third, by relating

measures of potential relative earnings to these same outcomes.

The discontinuity to the right of 0.5 in the distribution of wives' shares of relative income was the headline result of BKP's study, but a series of follow-up papers have shown that it is an unreliable method of inferring an aversion to female breadwinning. This discontinuity can arise without the existence of the norm (Binder and Lam, 2020; Grow and Van Bavel, 2020) and the result depends on the point mass of couples with exactly equal income (Hederos Eriksson and Stenberg, 2015; Binder and Lam, 2020), which may simply owe to couples who own a shared business or work in the same profession (Zinovyeva and Tverdostup, 2018). Collectively, this set of studies cast doubt on the discontinuity as evidence of an aversion to female breadwinning.[4]

The second and most common approach uses variables for relative earnings or female breadwinning constructed from couples' observed earnings and regresses outcomes of interest including divorce, housework, and labor force participation on these measures of relative earnings. These measures are most commonly implemented as either a continuous variable for the wife's relative earnings or indicator variables that give a couple's position in the distribution of relative earnings. Importantly, relative earnings are generally taken as given, without the use of instrumental variables or other strategies to address the endogeneity of earnings.

Within this category in sociology, there is a substantial literature that finds evidence that husbands do more housework and wives do less among couples that have a female breadwinner, compared to those with a male breadwinner Brines (1994); Greenstein (2000), although subsequent papers have debated whether these findings represented "gender deviance neutralization" or whether they were misspecifications of the key variable and controls (Bittman, England, Sayer, Folbre, and Matheson, 2003; Evertsson and Nermo, 2004; Gupta, 2007; Gupta and Ash, 2008). England (2011) and Sullivan (2011) revisit this literature and argue that it is "much ado about nothing," since the fraction of couples whom this effect is applicable to is small and–in line with the argument of this paper–depends on female breadwinners whose husbands are often unemployed or have low income. However, this argument has not been broadly appreciated beyond this

[4]Among these, Binder and Lam are not entirely pessimistic: They propose that a more fruitful approach would be to study the joint distributions of relative earnings and household outcomes, while taking care with the point mass of couples with equal incomes. In fact, this is the second approach discussed in this section. They note that BKP indeed do this in the remainder of their paper, but given their concern for how couples with equal incomes might affect such estimation, they propose several robustness tests for future research about female breadwinning. These tests include dropping this group of couples from the sample and changing the indicator for female breadwinning to include couples with exactly equal incomes, rather than only couples where the wife has strictly greater income. Since this provides an alternative possible explanation for BKP's findings, In Appendix C I replicate and test the robustness of BKP's main results when using a binary variable for female breadwinning and show that estimates are actually unaffected by these alternative specifications. Thus, while the point mass of couples with equal income was important to explaining the discontinuity result, it does not appear to explain BKP's other results.

focused literature on housework.

A separate strand of research studies how higher earnings for wives is also generally found to be positively associated to divorce, particularly when the wife earns more (Schwartz and Goñalons-Pons, 2016; Heckert, Nowak, and Snyder, 1998; Liu, Vikat, et al., 2004; Teachman, 2010; Jalovaara, 2003; Cooke, 2006).

While much of the work discussed thus far uses cross-sectional data or treats longitudinal data as pooled cross-sections, more recent papers have used couple fixed effects with panel data to "[enhance] their purchase on causality" (England, 2011). BKP use a fixed effects approach as their method for studying the relationship between observed female breadwinning and wives' labor force participation, the gap in housework hours, and divorce. Foster and Stratton (2021) focus on relationship dissolution and take a similar approach with fixed effects.

Potential relative earnings provide the third, potentially exogenous source of variation to test aversion to female breadwinning. Compared to observed female breadwinning, deciding how to construct potential relative earnings is less straightforward, and may still lead to problems. For example, BKP also construct a contemporaneous measure of the "probability that the wife earns more" based on women's predicted earnings distributions and husbands' actual earnings. In this paper, I show that this measure actually shares similar features to using observed relative income for both husband and wife due to its use of husbands' observed income. In particular, the variation in this variable is strongly related to reductions in men's earnings, many of which are likely endogenous.

In the very same paper, BKP also use an entirely predicted measure of relative earnings based on average earnings for women and men in hypothetical marriage markets to assess its relationship with the marriage rate, which they find to be negative and large enough to be able to explain a substantial fraction of the decline in the marriage rate in recent decades. However, this analysis speaks to the potential effect of aversion to female breadwinning on initial partnering decisions, rather than within marriage.

Finally, Schwartz and Goñalons-Pons (2016) use a measure of predicted relative earnings that estimates long-run relative earnings 30 years into the future for each couple-year based on average earnings for individuals with similar demographic characteristics, to study whether there is selection into marriage such that couples with wives who have higher expected long-run earnings are more likely to divorce. This analysis is intended as a complement to the results in the same paper showing that there is a positive relationship between observed female breadwinning and divorce, but only for earlier cohorts. They find that couples with higher long-run relative earnings potential for wives actually have a lower risk of divorce among more recent cohorts.

From this review, it is clear that nearly all the research suggesting couples take costly

decisions within marriage about their division of labor or even separation to avoid or compensate for female breadwinning is based on a strategy of regressing outcomes on measures of observed relative earnings. Thus, it is important to understand whether this strategy provides an informative answer to the question of interest.

2.2 Stated and Revealed Preferences

Beyond the set of papers regressing outcomes on observed relative earnings measures, there is work using other strategies that suggests general preferences against higher-status women in couples. Several experimental papers demonstrate that men dislike for women to have higher status in heterosexual relationships (Fisman, Iyengar, Kamenica, and Simonson, 2006; Ratliff and Oishi, 2013). Despite the fact that higher-educated women are now more likely to marry than women with lower educational attainment (Bertrand, Cortés, Olivetti, and Pan, 2016; Fry, 2010; Torr, 2011; Perelli-Harris and Lyons-Amos, 2016; Bertrand et al., 2016; Pessin, 2018), even highly educated single women appear to internalize men's expectations in how they present their career ambitions (Bursztyn, Fujiwara, and Pallais, 2017).

More specific to breadwinning, in surveys couples also selectively misreport earnings and housework in order to conform with gender norms (Murray-Close and Heggeness, 2018; Roth and Slotwinski, 2019; Kan, 2008). However, survey misreporting is low cost. In contrast, lab experiments about couples' task allocation suggest gender norm considerations are still dominated by comparative advantage when couples make decisions Cochard, Couprie, and Hopfensitz (2018); Görges (2018). At the same time, people might not associate specialization on the lab tasks with the real-world status concerns previously highlighted, which could mean that these lab experiments underestimate the willingness to pay for conforming to gender norms.

In sum, evidence based on other strategies provides some evidence for preferences against higher-status women, and even against female breadwinning, but there is no existing evidence other than from regressions with relative earnings that couples actually are willing to take costly decisions about housework, labor force participation, or divorce in order to avoid or compensate for female breadwinning.

2.3 Factors related to female breadwinning and its persistence

In practice, the existing evidence compares couples with female breadwinners to couples with male breadwinners. To understand what we are identifying with this comparison, we need to understand the factors and characteristics associated with female breadwinning. Only a few papers have studied this subject.

The incidence of female breadwinning has increased over time and has gained attention in media as well as qualitative research, where the focus has often been on highly-educated women with high earnings (Tichenor, 2005). However, female breadwinning is largely a transitory phenomenon among U.S. married couples, with 60% of occurrences of female breadwinning persisting over a three-year period (Winkler, McBride, and Andrews, 2005), and less than a third persisting over five years (Winslow-Bowe, 2006). Moreover, Winslow-Bowe (2006) points out that popular depiction of higher-earning wives is often as "high-status, generally White, career women" but in fact finds evidence that "economic vulnerability and marginalization" – including couples that are in the bottom quartile of the income distribution and where the husband has low weekly hours – are strongly associated with female breadwinning. The findings of Winkler et al. (2005) support this same conclusion, showing that those couples that have a female breadwinner over three years are much more likely to have a husband in poor health compared to couples without a female breadwinner or who experience female breadwinning for a shorter period of time. In a similar spirit, Vitali and Arpino (2016) show that in cross-country regressions, men's unemployment rates are associated with female breadwinning, but more egalitarian gender norms are not. These findings suggests that female breadwinning is on average less about empowered career women than popular perception might suggest, and that men's ability to work or to find work may play an important role.

2.4 Men's unemployment and its effects

Studies of the nature of female breadwinning suggest that low levels of work and elevated levels of unemployment for husbands are one of the major associated factors. Yet existing work not related to the literature on relative earnings and female breadwinning has demonstrated that men's unemployment itself is linked to divorce, housework, and labor force participation. Is it possible that we are picking up the effects of men's unemployment in regressing these outcomes on observed measures of relative earnings and female breadwinning?

Specifically, there is robust evidence that men's lack of employment, particularly unemployment, is associated with increased risk of divorce, although women's (un)employment is not (Killewald, 2016; Sayer, England, Allison, and Kangas, 2011; Charles and Stephens, 2004; Eliason, 2012). Interestingly, this effect is larger where traditional gender norms are stronger (Gonalons-Pons and Gangl, 2021) and is not explained by the loss of income from the husband not working (Killewald, 2016; Charles and Stephens, 2004). It is possible to connect this relationship to the idea of an aversion to men's unemployment, but it could also be that the stress of unemployment causes strife.

For housework, there is mixed evidence, with some settings finding that men's unemployment actually leads to reductions in their relative share of housework (Foster and Stratton, 2018); others finding that it increases, but less than for women who unemployed (Gough and Killewald, 2011; Van der Lippe, Treas, and Norbutas, 2018; Fauser, 2019); still others find that it increases more for husbands' unemployment than for wives' unemployment (Voßemer and Heyne, 2019). These differences may reflect difficulty in defining housework in a comparable way given available data.

Finally, there is a large existing literature studying wives' labor supply responses to unanticipated negative shocks to husbands' employment, known as the "added worker" effect. Depending on the setting and circumstances, studies have generally found positive effects of job displacement or job loss for husbands on wives' labor force participation, with generally larger effects in adverse economic conditions, including recessions and under liquidity constraints (Lundberg, 1985; Fernandes and Felício, 2005; Kohara, 2010; Gong, 2011; Parker and Skoufias*, 2004; Spletzer, 1997; Mattingly and Smith, 2010; Prieto-Rodríguez and Rodríguez-Gutiérrez, 2000; Başlevent and Onaran, 2003; Ayhan, 2018; Ortigueira and Siassi, 2013).

An important distinction in considering how these three bodies of evidence inform interpretions of regressions on female breadwinning is that these literatures are overwhelmingly focused on exogenous job loss. While the previous section suggested unemployment is a factor in female breadwinning, we must keep in mind that female breadwinning can be caused by both exogenous shocks (to earnings or employment), as well as endogenous choices of couples. This point will figure importantly in understanding how regressions of outcomes on measures of relative earnings may be biased.

2.5 Synthesis

Existing evidence that aversion to female breadwinning affects married couples' labor supply, division of housework, and divorce rates depends largely on strategies that regress these outcomes on observed measures of relative earnings or indicator variables for female breadwinning. However, studies of the nature of female breadwinning itself suggest it is associated with economic marginalization, in particular low household income and low work hours for husbands. These features raise the question of whether regressions with observed measures of earnings may in fact be picking up the effects of female breadwinning for other reasons than exogenous changes in relative earnings, such as female breadwinning owing to men's unemployment. After introducing the data sources used in this paper, the subsequent section explores in greater detail what implications these other sources of variation in female breadwinning may have for this type of estimation strategy.

3 Data

Throughout the paper, I use data from four sources: the Panel Study of Income Dynamics, the U.S. Census and American Community Survey, and two datasets collected as part of the Current Population Survey, the Annual Social and Economic Supplement and the American Time Use Survey. Across all surveys, the sample of interest includes married adults where both members of the couple are between the ages of 18 and 64, where they are not retired and not enrolled in school. Summary statistics allowing for comparison of key variables across datasets are included in Appendix A.

The Panel Study of Income Dynamics is a longitudinal survey of households begun based on a representative national sample in 1968, which follows members of the original families and their descendents. The interview schedule changed from annual to biennial in 1997. I use a sample from 1979 to 2013, since 1979 is the first year consistent employment status variables are available for wives. The initial sample includes 93,263 couple-year observations.[5]

The U.S. Census and American Community Survey, which I access from IPUMS (Ruggles, Flood, Goeken, Grover, Meyer, Pacas, and Sobek, 2018), are large cross-sectional samples of the U.S. population. In this paper, I use waves 1980, 1990, and 2000 of the Census. For after 2000, when the Census Bureau began conducting the annual American Community Survey, I pool the years 2005 to 2007 to provide a comparable sample to that used in the analysis of Census data in BKP.[6] This sample includes over 4 million observations.

The Current Population Survey (likewise accessed via IPUMS) is a smaller nationally representative monthly survey. I use two modules: the Annual Social and Economic Supplement, which records more detailed information about earnings and work than the baseline interview, and the American Time Use Survey. Each module surveys a fraction of the baseline CPS population. The ATUS entails a 24-hour time diary, which provides quite accurate short-term measures of time use. The ATUS sample includes 38,638 observations between waves 2002 to 2011.

In this analysis, the PSID and the Census data are used for analyzing the relationship between female breadwinning and wives' labor force participation. In both surveys, labor force status is recorded with respect to the time of interview, while annual hours and income are recorded with respect to the prior year.[7]

The PSID and the ATUS are used for analyzing the relationship between female

[5]Not all analyses in the paper make use of the entire sample.

[6]BKP instead use 2008 to 2010, as well as 1970, but I choose to not use these years because respondents are only asked to report their exact weeks worked in the previous year through between 1980 and 2007, which allows for more precise calculation of hourly wages.

[7]The PSID asks with respect to the prior calendar year, while the Census asks with respect to the 12 months prior to the interview.

breadwinning and housework. The advantage of the PSID is that it is longitudinal and includes responses for both members of the households, with the disadvantage that it simply asks individuals to estimate how many weekly hours of housework they typically do, explicitly asks respondents to exclude childcare, and makes one person responsible for answering for both individuals. The ATUS is likely more precise, and collects information on childcare time, but it is cross-sectional and only collects time use for one individual in the household. Additionally, earnings are collected on a weekly basis in the ATUS rather than an annual as in the PSID. The CPS is used as a supplement to the PSID and ATUS to generate the same year-specific probabilities that a wife would earn more than her husband as used with Census data.

Finally, the PSID is also used to study the relationship between the likelihood that a wife earns more and the risk of marital dissolution. Since 1986, the PSID has collected marital histories for respondents, but these are not available all respondents. I follow BKP's description in their Online Appendix for coding whether a marriage ended in divorce or separation in the following year, which infers divorce or separation for individuals with incomplete marital histories.

4 The problems with observed measures of relative earnings

As discussed earlier in the literature review, there are a range of reasons that female breadwinning may occur. Broadly speaking, they can be grouped in three categories: exogenous shocks to either wives' or husbands' earnings (such as through promotions for the wife or a company-wide wage cut for the husband); negative exogenous shocks to husbands' employment (e.g. layoffs or plant closures); and endogenous decisions by the couple that result in a female breadwinner.

If we want to infer whether there is an aversion to female breadwinning, we must understand what the responses are to female breadwinning owing to the first reason only. Learning about responses to men's employment shocks is itself an interesting subject and worth studying, but while it might be related to the male breadwinner norm more broadly (in the sense of the prescription that husbands should work full-time), it cannot be interpreted as informative about female breadwinning by itself.

In this section, I outline a simple statistical framework that illustrates that these other sources of variation lead to bias if we use observed earnings to infer the effect of female breadwinning owing to exogenous changes to wives' or husbands' earnings. I identify two conditions necessary for bias to arise: first, that these sources represent a non-negligible share of the variation in female breadwinning and second, that they are

themselves related to the outcomes of interest. Then, I provide descriptive evidence on female breadwinning that suggests that these conditions are likely fulfilled. Next, I use BKP as a reference point to show how both measures entirely based on observed earnings (such as indicator variables for female breadwinning) as well those only partially based on observed earnings are likely to suffer from this bias. Finally, I discuss possible solutions to this issue.

4.1 Statistical framework

Suppose that we have perfect knowledge of the individuals in our data. We observe the three categories of factors that can contribute to female breadwinning, letting a_i represent exogenous changes to men or women's relative earnings, b_i represent negative shocks to husbands' employment, and c_i^k represents the k possible reasons a couple might endogenously change labor supply in a way that would lead to female breadwinning. We can then partition occurrences of female breadwinning by why it occurred: $F_i = F_i^a + F_i^b + F_i^c$.[8]

Each of the situations under which female breadwinning occurs as well as each of the underlying reasons it occurred may have its own relationship to female labor force participation. Supposing these relationships are linear, we can think of the true model as follows, where P represents the labor force participation of the wife (although it could also more generally represent other outcomes of interest, like divorce and housework):

$$P_i = \beta_0 + \beta_1(F_i^a) + \beta_2(F_i^b) + \beta_k(F_i^{c^k}) + \gamma_1(a_i) + \gamma_2(b_i) + \gamma_k(c_i^k) + \epsilon_i \tag{1}$$

To be clear, we imagine the underlying cause of a given type s of female breadwinning as itself potentially having its own effect γ_s on P_i, which necessitates including it as a control. Naturally, for each of the component factors F_i^s and s_i will be correlated, but we will assume all other right-hand variables are mutually uncorrelated with each other and with the error term ϵ_i.[9]

Note that the parameter of interest for our question is simply β_1: the marginal effect of female breadwinning owing to exogenous shocks to a couple's relative income on the wife's labor force participation. In fact, since the contributing factors are assumed to be uncorrelated, if we could observe F_i^a and a_i, it would be sufficient for our question to estimate the simpler model:

[8] In practice, any given instance of female breadwinning may owe partially to several of these sources, but for simplicity of exposition I characterize them as mutually exclusive. The implications for bias would be the same if we instead added weights summing to one to each of the components.

[9] This assumption may seem strong; if we did not make it we would simply have the immediate result that any direction of bias is possible.

$$P_i = \beta_0 + \beta_1(F_i^a) + \gamma_1(a_i) + \epsilon_i \tag{2}$$

Unfortunately, we do not in fact have the knowledge to be able to partition observed female breadwinning F_i into its component parts. That rules out our ability to estimate Eq. 2. Given this limitation, and still assuming that we observe a_i, b_i, and c_i^k, we would have to rewrite Eq. 1 as follows:

$$\begin{aligned} P_i &= \beta_0 + \beta_1(F_i) + \gamma_1(a_i) + \gamma_2(b_i) + \gamma_k(c_i^k) + \mu_i, \\ \mu_i &= \epsilon_i + (\beta_2 - \beta_1)b_i + (\beta_k - \beta_1)c_i^k \end{aligned} \tag{3}$$

This is a situation of non-classical measurement error, where F^b and F^c can be thought of as the "errors." The main problem that generates bias is that, in the true model, both of these components are related to P_i, whereas classical measurement error is uncorrelated with both the independent and dependent variables as well as the error term. If $\beta_2 = \beta_k = 0$, Eq. 3 would simply yield attenuation bias, since the components of F_i are mutually uncorrelated and uncorrelated with ϵ_i.

However, when $\beta_2, \beta_k \neq 0$, $\widehat{\beta_1}$ instead becomes a weighted average of $\beta_1, \beta_2, \beta_3$, where the weights in the average depend on their relative variances. As a general principle, the bias will be more severe the more variation in F_i owes to F^b and F^c relative to F^a and the greater the differences between β_1, β_2, and β_3. Moreover, the direction of the bias is determined by the signs of β_2 and β_3. In Appendix B, I prove these points in the algebraically simpler setting of two variables with no additional controls. Concretely, $\mathbb{E}[\widehat{\beta_1}] = \beta_1$ only if either $var(F_i^b) = var(F_i^c) = 0$, or if $\beta_1 = \beta_2 = \beta_3$.

One hope is that controlling for b_i and c_i^k would reduce the relative residual variation in F_i owing to F_i^b and F_i^c. There are two issues with this idea. First, we are also controlling for a_i which will likewise reduce the residual variation in F_i owing to F_i^a. To make matters worse, we said from the outset that c_i^k represents the reasons or conditions under which couples make endogenous choices about labor supply, which can then lead to female breadwinning. By definition, these reasons are almost certainly unobservable. As a result, we likely are unable to control for c_i^k at all, which leaves us instead to estimate the following model:

$$\begin{aligned} P_i &= \beta_0 + \beta_1(F_i) + \gamma_1(a_i) + \gamma_2(b_i) + \psi_i, \\ \psi_i &= \epsilon_i + (\beta_2 - \beta_1)b_i + (\beta_k - \beta_1)c_i^k + \gamma_k(c_i^k) \end{aligned} \tag{4}$$

From this, we can see we have an unusual form of non-classical measurement error,

potentially exacerbated by omitted variables bias given c_i^k is unobserved.

The discussion thus far has not included a time dimension. Another solution to consider is whether, in a longitudinal setting, couple fixed effects can alleviate this bias. Here too, the answer depends on empirical features of b_i and c_i^k. If other sources of variation in female breadwinning are largely time-invariant, then couple fixed effects would indeed be helpful, but if they are time-varying, they will not.

Finally, a straightforward solution to this bias would be to use group-level averages as proxies for F_i^a and a_i, which would be uncorrelated with b_i and c_i^k. For example, it could be possible to estimate a group-level measure of the likelihood that a wife with particular characteristics would earn more than her husband, given his characteristics, based on average earnings for similar women and men. Such proxies could then be used in order to estimate an equation along the lines of Eq. 2. Although this approach would clearly eliminate the type of bias presented here, it could potentially introduce a new bias if characteristics used to define the group are correlated with unobservables that are related to the outcome. For example, highly-educated women have higher wages than women with less educational attainment, but educational attainment itself may be related to unobserved ability and motivation, which may likewise be related to labor force participation. Fortunately, these types of unobservable characteristics are likely to be time-invariant and thus amenable to the inclusion of couple fixed effects.

In the next section, I use descriptive evidence about observed female breadwinning to highlight that the key features that yield bias are in fact present: namely, low levels of work for men owing to both exogenous and endogenous conditions represent substantial sources of variation in female breadwinning; these low-levels of work appear to be reflect time-varying changes in men's hours in the years before and after female breadwinning; and low levels of husbands' work do appear to be related to wives' labor force participation.

4.2 Features of female breadwinning

In the previous section, I showed that regressing labor force participation on observed female breadwinning will pick up an average of the effects of different situations under which female breadwinning occurs. In studying whether there is an aversion to female breadwinning on its own, however, we wish to identify solely the effect of female breadwinning owing to exogenous changes in wives' and husbands' earnings. In this section, I show that men's reductions in work, both involuntary and voluntary, are major factors in observed female breadwinning. I also provide suggestive evidence that men's reductions in work are negatively related to wives' labor force participation. Together, these two facts fulfill the two conditions needed for bias to arise, as laid out in the previous

section.

4.2.1 Men's reductions in work as a factor in female breadwinning

The first feature necessary for bias to arise in estimation using observed female breadwinning is that variation in female breadwinning owes to reasons other than shocks to couples' relative earnings. When examining factors related to female breadwinning, it's clear that men's reductions in work are a major factor.

A first way to illustrate this point is simply to graph the fraction of husbands and wives in full-time work by the wife's share of relative income, as in Fig. 1. This data is graphed by decade for the years 1979 to 2013. Couples are grouped in 0.1-width bins with respect to the wife's share of income, and full-time work is characterized as working at least 1,820 hours over the year (an average of 35 hours per week). Hours and income are both reported with reference to the full calendar year prior to the time of the interview, and markers are weighted by the number of observations per bin and time period to illustrate changes in the distributions of relative income.

If husbands' levels of work were not meaningfully related to female breadwinning, we would expect to see a flat line across Panel (a). Instead, we see that between 80-90% of husbands work full-time for all bins with relative income under 0.5. Past this threshold, the fraction working full-time declines dramatically, with a majority of men not in full-time work in couples where the wife's share of income exceeds 0.7. Wives's rates of full-time work are almost a mirror image, although lower by about 0.2 at every level of relative income. While wives' levels of work are also clearly important, the fact that they are lower at every level suggests that they are less important in determining breadwinning than husbands' work levels.

This first graph suggests that the levels of work are important, but what is the nature of husbands' lower levels of work under female breadwinning? Because hours and income are reported retrospectively for the prior calendar year (collected in time t with respect to $t-1$) while employment status is only reported with respect to the time of interview, we can glean insights by comparing husbands' employment status for the current and prior interview (in time t and $t-1$), book-ending the reference period for income.

Fig. 2 displays these before and after snapshots as the fraction of husbands that reported being unemployed, temporarily unemployed, disabled, or in the mixed category of either "housewife", student, or "other" at the time of the interview, again by the wife's binned share of earnings (with observations pooled across time). Perhaps unsurprisingly, the share of unemployment rises substantially across the distribution of relative income for both the before and after graphs. However, there is little change in the levels of

unemployment, temporary or otherwise, between time $t-1$ and time t. In contrast, it is disability and the mixed category of non-work that rise between the two time periods for couples with a female breadwinner. This difference suggests that illness or choices to study or shift to domestic work may be among reasons that contribute to low levels of work for men when female breadwinning is observed to occur.

Importantly, the "unemployed" category is not exclusive to those who were exogenously laid off, but includes all those who are currently not working and looking for work. Those currently not in work are asked about how their last job ended, which can provide additional insights about the reasons for husbands' low levels of work. Fig. 3 shows the fraction of husbands who are not in work, by why their last job ended in the prior year: layoff, plant closure, quitting, or temporary and seasonal jobs. While layoffs and plant closures account for most non-work among husbands whose wives were breadwinners, non-work owing to quitting the last job or having worked a temporary or seasonal job account for close to half. This split suggests that even among those husbands in female breadwinning couples who report "looking for work" as their employment status, their work reductions may represent endogenous choices to leave a job or to pursue a particular type of employment that is not a standard full-time job.

Finally, if these low levels of men's work associated with female breadwinning were largely time-invariant, couple fixed effects applied in a longitudinal setting could address the bias they would contribute in cross-section. However, Fig. 4 shows that men's average hours begin falling two to three years prior to the occurrence of female breadwinning, are at the lowest point at time zero–with an average reduction of four to five hundred hours per year by this point–and then return to pre-female breadwinning levels in the two to three years after the occurrence of female breadwinning. In other words, the low levels of men's work associated with female breadwinning do indeed appear to be time-varying.

To create this figure, I look at average hours of work for husbands in the four years before and after a spell of female breadwinning. For couples that experience more than one spell of female breadwinning, one instance was randomly selected as the reference period, so that each couple is represented only once.[10]

Together, these figures make clear that the first condition for bias to arise is clearly fulfilled when carrying out estimation using observed female breadwinning: Men's work reductions are an important source of variation in observed female breadwinning, and while negative employment shocks are likely an important factor in these work reductions, these work reductions appear to be related to a wide range of endogenous decisions, including husbands shifting to home work or studies, experiencing illness, quitting

[10]This approach means additional spells of female breadwinning may be included during the observed period for some couples.

their jobs, and working in temporary or seasonal positions. Moreover, the time-varying nature of these changes precludes including couple fixed effects as an easy solution.

4.2.2 Relationship with wives' labor supply

Given the facts presented above, the other sources of variation in observed female breadwinning will at minimum create attenuation bias in the estimates of the relationship between female breadwinning and labor force participation. For non-attenuating bias to arise, a second condition must be fulfilled: the other sources of variation in female breadwinning, besides exogenous shocks to relative earnings, must themselves be related to wives' labor force participation. As discussed in the literature review, prior work has studied the effects of exogenous job loss for husbands on wives' labor force participation. However, here we are interested more broadly in the relationships between husbands' and wives' levels of work and employment statuses, which may well be endogenous.

If husbands' levels of hours or employment status were unrelated to wives' labor force participation, we would expect to see roughly similar levels of labor force participation for wives with husbands of varying levels of work and employment. Fig. 5 and Fig. 6 demonstrate that this is not the case.

Fig. 5 illustrates that women with husbands who did not work full-time have lower levels of labor force participation and hours (conditional on working) across most years of the period of study. Specifically, Panel (a) of Fig. 5 shows wives' levels of labor force participation over time by whether or not husbands worked full-time in the prior year, while Panel (b) does the same for average annual hours (for women working positive hours). Splitting by full-time status highlights differences that may owe to any type of work reductions on the part of the husband. This result implies that on average, the correlation between husbands' full-time work and wives' labor force participation is positive.

This overall difference might mask variation by differing employment situations, so Fig. 6 instead displays a similar trend over time in wives' labor force participation and hours, this time by her husband's employment status: working, temporarily unemployed, unemployed (which may include people who quit their jobs as well as those laid off), disabled and those who report home work, studying or "other." Keeping in mind that the "work" category may include those who worked part-time, Panel (a) shows that wives with husbands reporting most other work statuses are less likely to be in the labor force than wives whose husbands are working, with the exception being the mixed category of home work and studying, who actually participate at higher rates than others. Conditional on working, Panel (b) suggests there are not easily discernible differences in hours across these categories, excepting again wives whose husbands are home or

studying. This graph thus supports the notion that in most situations where a husband is not working or is not working full-time, wives are less likely to work compared to their counterparts with full-time working husbands.

Finally, if the negative relationship BKP estimate between observed female breadwinning and wives' labor force participation in fact owes to the relationship between husbands' low levels of work, female breadwinning, and wives' labor force participation (and perhaps underlying unobserved factors that affect all three), then Fig. 4, illustrating the time-varying nature of husbands' reductions in work, suggests an obvious and simple test: if husbands' reductions in work that result in female breadwinning on average precede the occurrence of female breadwinning, the estimated relationship between female breadwinning and *prior* periods' labor force participation should also be negative. Such a result could of course not be squared with an interpretation that the negative relationship with subsequent labor force participation is in fact a *consequence* of female breadwinning.

In fact, in replicating BKP's baseline specification using longitudinal data on couples in the PSID but replacing the dependent variable with preceding periods for labor force participation, Fig. 7 shows just such a relationship of lower labor force participation for wives prior to the occurrence of female breadwinning, even with the inclusion of couple fixed effects. (Appendix D provides details on the implementation of this exercise.)

These facts suggest that there is a relationship between husbands' and wives' levels of work and participation in work, which would thus fulfill the second condition needed for non-attenuating bias to arise in settings where we estimate the relationship between labor force participation and observed female breadwinning.

This section and the prior section lay out in a general way how analysis using observed female breadwinning might be biased. In the next section, I lay out how the specific estimation strategies employed by BKP fit into this paradigm.

4.3 What BKP Do

BKP are ambitious in the range of datasets and outcomes they study to provide evidence for the impact of the breadwinner norm, so for clarity I begin by an overall summary of the analyses included in their paper. Subsequently, I will explain how their analytical choices relate to the issues of bias raised in the previous sections.

4.3.1 Overview

The first part of BKP studies the discontinuity in the distribution of the wife's share of relative income at 0.5 as a sign of aversion to female breadwinning. Earlier in the literature review, I summarized the set of papers that study this discontinuity analysis.

While it was a novel strategy, the later papers find that the result itself is sensitive and, from a theoretical perspective, likely not informative about the breadwinner norm.

The next part of the paper focuses on the margin of sorting, studying the relationship between potential relative earnings of randomly matched women and men in (hypothetical) marriage markets and the marriage rate. They find a negative relationship even when using an instrumental variables approach.

The remainder of BKP's work focuses on the analysis of married couples, based on the hypothesis that adjustments in labor supply and marital dissolution in relation to female breadwinning may also reflect gender identity concerns. These aspects are the focus of this paper's replication and reassessment.

Within this section, the first analysis focuses on the relationship between the predicted probability that wives would earn more than their husbands and their labor force participation and earnings, using Census and ACS data. The measure used combines predicted earnings for wives and observed earnings for husbands. They find a robust negative relationship between this measure and wives' labor force participation. To complement this analysis, they consider marital stability and marital satisfaction in relation to female breadwinning using the National Survey of Families and Households, with an indicator variable for whether the wife earned more in previous years, finding that female breadwinning is associated with lower marital satisfaction and increased probability of divorce. Next, BKP turn to studying home production using the ATUS. Here again they use an indicator variable for whether the wife earned more than the husband, though it is in terms of weekly earnings, as collected in the CPS. They interact the indicator for female breadwinning with the respondent's sex, and show that female breadwinning is positively associated with wives' housework hours but has no relationship with husbands' housework hours.

Finally, BKP reexamine the relationship between and indicator variable for female breadwinning and labor force participation, housework hours, and divorce using the PSID, with which they can include couple fixed effects. For housework, the key variable is the gap in weekly housework hours, rather than wives' absolute hours. Throughout this section, they use an indicator variable for whether the wife earned more in time $t-1$, and look at the effects on labor force participation in time t, contemporaneous housework hours in time $t-1$, and divorce probabilities in time t and $t+1$. They find strong negative and positive relationships for labor force participation and housework respectively, and a weaker positive relationship with divorce.

While the partially predicted variable of the likelihood of female breadwinning might seem to be less subject to hypothetical bias previously discussed, in the next section I show that both such a partially predicted variable as well as indicators for female breadwinning are likely affected.

4.3.2 Analyses using indicator variables for observed female breadwinning

Nearly all of BKP's analyses on married couples use indicator variables for female breadwinning. In the cross-sectional analyses of divorce and housework (using the NSFH and ATUS, respectively), as well as the longitudinal analyses of divorce, housework, and labor force participation (using the PSID), BKP's key variable of interest is an indicator for observed female breadwinning. The issue of bias has been discussed so far exactly in terms of such an indicator variable, so it's clear that the potential for bias is clearly present in these analyses.

In the longitudinal analysis BKP perform, couple fixed effects are also included–does this change whether bias is likely to arise? Given the framework laid out in Sec. 4.1 and the facts presented in Sec. 4.2, it's straightforward that couple fixed effects would only reduce or remove the bias to the extent that other sources of variation in observed female breadwinning (and their relationship to wives' labor force participation) are time-invariant. However, we have also seen that the other sources of variation owe largely to *changes* in husbands' levels of work. Thus, the issues relevant for the cross-sectional case are also relevant for the longitudinal setting.

4.3.3 Analysis using "the probability that a wife earns more"

In their cross-sectional analysis of labor force participation using the Census and the ACS, BKP construct a measure that they call "the probability that the wife earns more." While this sounds like a predicted measure of female breadwinning that might sidestep the issues applicable to indicator variables for observed female breadwinning, I show that the fact that husbands' observed earnings are used to construct it means that we still are picking up men's reductions in work as an important source of variation in this variable. In other words, the same issues apply.

BKP's goal is to generate a measure of the likelihood that a wife would earn more than her husband that can be calculated even for women who are not working. They come up with a solution to this problem by generating a predicted distribution of annual earnings for each woman and comparing it to the observed annual earnings of the husband, yielding a measure between zero and one that we interpret as the likelihood that a wife would earn more than her husband.

The measure is constructed as follows: they assign every woman (regardless of working status) a distribution of potential earnings by calculating the vigintiles of the annual earnings distribution for the working women with the same state, age group (five-year intervals), race (white, black, and other races grouped), educational attainment (five levels), and year. Denote each vigintile of earnings as $\widehat{W}^i_{w,a}$, for $i = 1, .., 19$, with the wife's demographic group given by a.

Next, they compare each of the moments of the assigned distribution for a given wife to her husband's observed annual earnings W_m, generating a value of one if a given moment for the wife exceeds the husband's earnings. Finally, they average across the 19 moments. To be concrete, this measure is computed as:

$$\text{PrWifeMore} = \frac{1}{19}\sum_{i=1}^{19} \mathbb{1}(\widehat{W}^i_{w,a} > W_m)$$

How problematic is it that husbands' observed income is used to construct this measure, given that wives' have a predicted measure of income? If husbands' reductions in work are relatively unimportant sources of variation in PrWifeMore compared to wives' predicted earnings, the potential for bias may be less. To assess, we can compare how husbands' changes in earnings relate to PrWifeMore as well as to a comparison measure where we likewise construct predicted moments of the husband's distribution in an analogous way. Specifically, we predict the moments of the husbands and wife's distributions and average across the comparisons of each corresponding moment, where a gives the demographic group of the wife and b the demographic group of the husband:

$$\text{PrWifeMore}^* = \frac{1}{19}\sum_{i=1}^{19} \mathbb{1}(\widehat{W}^i_{w,a} > \widehat{W}^i_{m,b})$$

Fig. 8 shows that reductions in husbands' earnings do appear to drive much of the variation in PrWifeMore, particularly for higher values. In contrast, PrWifeMore* does not appear to be systematically related to husbands' observed changes in earnings.

This figure is constructed using earnings data on couples from the PSID, while the predicted elements of each measure are constructed with the corresponding waves of the Current Population Survey's Annual Social and Economic Supplement, which gives me a larger sample than the PSID itself such that I can exactly follow the methodology used to estimate moments of the distributions of wages for demographic groups with the Census and the ACS.

Next, I calculate the percentage change in husbands' earnings from one year to the next. Concretely, I compute $\delta = \frac{\mu_{t-1}-\mu_{t-2}}{\mu_{t-2}}$, where μ represents husband's earnings. Since this measure is unbounded from above, I restrict the sample to those between the 1st and 99th percentiles of the distribution of either income changes.

My goal is to graphically relate the relationship between these changes in earnings and wages to PrWifeMore and PrWifeMore*. However, by nature of its construction, PrWifeMore takes on only 20 unique values, whereas PrWifeMore* has many possible values.

To facilitate comparison, I assign the couples to fixed-width bins, where the bins are centered around the 20 possible values that PrWifeMore can take on. Next, within each

bin I calculate the median and mean of the income changes for the husbands falling within that bin. To give some sense of how couples are distributed across these values, the size of the markers is weighted by the number of couples in the bin.

The key takeaway from Fig. 8 is that the median and mean income changes for husbands become increasingly negative for higher values of PrWifeMore, whereas there is no trend in median or mean income changes as PrWifeMore* increases. The relationship between PrWifeMore and reductions in men's earnings is highly suggestive that men's reductions in work for a variety of reasons present a meaningful source of variation in this measure, just as for an indicator variable for observed female breadwinning. As a result, we should have the same concerns about bias in using this variable as in using an indicator variable.

I have presented how bias may theoretically arise, the descriptive evidence that suggests it is likely to do so, and how BKP's estimation strategies fall into this scenario. Next, I will discuss solutions for estimating our intended relationship of interest and assessing the extent to which measures like those BKP (and others) use are biased.

4.4 Solutions

Can measures based on observed earnings be rescued to provide informative estimates regarding an aversion to female breadwinning? In general, individual-level data is often more informative than group-level proxies, the alternative solution previously discussed. Additionally, proxies could also introduce alternative sources of bias.

The theoretical hope for measures based on observed earnings would be that with sufficient controls, one might be able to reduce the residual variation in an observed measure of female breadwinning to the identifying variation of interest: female breadwinning owing to exogenous changes in wives' and husbands' earnings.

However, the descriptive evidence in Sec. 4.2 raises serious doubts about whether it would be possible to control for all the underlying reasons female breadwinning occurs. For a substantial fraction of couples experiencing female breadwinning, there is good reason to think that female breadwinning resulted from endogenous choices about husbands' labor supply, which may be simultaneously related to wives' labor outcomes. Although we can to some extent control for husbands' observed labor supply, we do not observe the underlying reasons which may simultaneously affect his labor supply, the occurrence of female breadwinning, and his wife's labor supply.

Compounding this problem, we do not actually know how much of the variation in observed female breadwinning owes to our variation of interest, nor how much residual variation would remain after including controls for husbands' and wives' levels of income.

Considering these issues brings us back again to the attractiveness of using group-level proxies to predict the likelihood of female breadwinning. While there is still the possibility that observable characteristics used to construct such a measure might be correlated with unobservables that are related to the likelihood to work, like ability or tastes for work, such characteristics are likely to be time-invariant, and as such can be accounted for in a longitudinal setting using couple fixed effects. Using a group-level proxy based on predicted earnings, we can have more confidence on theoretical grounds that we are in fact identifying our relationship of interest.

For this exercise, BKP provides a natural frame of comparison: we can re-estimate their results using both indicator variables for female breadwinning, their measure for the probability that wives would earn more based on husbands' observed earnings, and the analogous measure instead constructed with predicted earnings for both husbands and wives. Estimates using the latter predicted measures would thus be interpreted to represent the marginal effect of interest in this paper.

In this next section, I detail precisely the measures and specifications I will use for this analysis, and how they relate to BKP's original analyses.

5 Comparing results using different measures of relative earnings

The preceding sections show that there are clear reasons based on both the statistical framework developed and the descriptive evidence regarding female breadwinning to prefer group-level predicted measures for the probability that wives would earn more than their husbands compared to measures constructed using observed earnings from either or both members of the couple. However, it remains to be seen whether the hypothesized bias is of practical importance. To answer this question, I re-estimate cross-sectional and longitudinal results from BKP, each with five different measures of female breadwinning: an indicator for observed female breadwinning, the original "probability that the wife earns more" from BKP, and three variations on this measure. I find that when using measures based on predicted earnings, estimated coefficients either flip signs or become statistically indistinguishable from zero compared to coefficients when using measures based on observed earnings. These results suggest that this source of bias does have meaningful practical implications for this area of study.

5.1 Measures of female breadwinning

Before discussing the details of the regression specifications, I will describe the five measures of female breadwinning.

The first measure, unsurprisingly, will be an indicator variable for observed female breadwinning, WifeMore: whether the wife earns more than the husband. For the ACS/Census and the PSID, this is a comparison made based on annual wage earnings. In both cases, wages are collected with respect to the prior year, and labor force participation with respect to the time of the interview. For the ATUS, only weekly earnings are observed, and thus this is the level at which earnings are compared.

The second measure will be PrWifeMore as originally defined in BKP, where $\widehat{W}^i_{s,d}$ represents moment i of the earnings distribution for individuals of sex s, where the distribution is calculated using individuals in demographic group d that are working positive hours, and where the demographic group is defined by state, race, five-year age-group, and four levels of educational attainment, and W_m represents husbands' observed earnings:

$$\text{PrWifeMore} = \frac{1}{19}\sum_{i=1}^{19} \mathbb{1}(\widehat{W}^i_{w,a} > W_m)$$

The third measure will be the adapted version previously discussed, PrWifeMore*, which replaces the husband's observed earnings with predicted earnings calculated in an analogous way:

$$\text{PrWifeMore}^* = \frac{1}{19}\sum_{i=1}^{19} \mathbb{1}(\widehat{W}^i_{w,a} > \widehat{W}^i_{m,b})$$

An additional notable feature of the original PrWifeMore is that it is based an annual earnings rather than hourly wages. For men this may not make a major difference, since the vast majority of men that are working are working full-time, but for women the variation in women's hours is likely an important source of variation in this key measure, and arguably does not reflect variation in potential earnings.[11]

Thus, the fourth and fifth measures parallel the previous two, except they are based on hourly wages, ω:

$$\text{PrWifeMore}_{Hr} = \frac{1}{19}\sum_{i=1}^{19} \mathbb{1}(\widehat{\omega}^i_{w,a} > \omega_m)$$

[11] While there is evidence for a part-time wage penalty (e.g. Manning and Petrongolo (2008); Bardasi and Gornick (2008); McGinnity and McManus (2007)), it is on the order of 10 to 30%. A woman who works 20 hours a week in a given occupation likely has annual earnings less than half that of a woman who works 40 hours per week, but most of this difference is due to the choice of hours, not a difference in their potential earnings.

$$\text{PrWifeMore}^*_{Hr} = \frac{1}{19}\sum_{i=1}^{19} \mathbb{1}(\widehat{\omega}^i_{w,a} > \widehat{\omega}^i_{m,b})$$

For the Census/ACS, the predicted elements of these latter four measures are constructed from the Census/ACS samples themselves. For the PSID and the ATUS, given their much smaller samples, I instead use data from corresponding waves of ASEC in CPS.

5.2 Cross-sectional analyses

In this section, I re-assess two cross-sectional analyses in BKP: the analysis of labor force participation using the ACS and Census and the analysis of housework using the ATUS. I find in both cases that switching from using measures of female breadwinning based on observed income to those based on predicted income flips the sign of the original result: there is a positive relationship between the predicted likelihood of female breadwinning and labor force participation and a negative relationship with wives' housework hours.

5.2.1 Estimation

For the ACS/Census, I follow BKP in including married households with both members between the ages of 18 and 64 where the husband worked positive hours in the prior year. I will note that the sample I use is different with respect to years: BKP include decennial U.S. Census waves from 1970 to 2000, and then use a pooled sample from the 2008-2010 American Community Survey as their most recent year, while my sample does not include 1970 and has a pooled sample of the 2005-2007 waves of the American community survey as the most recent wave.[12] Table A3 provides summary statistics.

The sample of married couples is defined in nearly the same way for the ATUS as for the Census/ACS, except that couples are only excluded if both were not working as of the survey. The key variable of interest is non-market work, constructed exactly as in BKP to include child care and chores. The sample likewise includes ATUS waves 2002 to 2011.

For labor force participation, BKP's baseline estimating equation is as follows:

$$LFP_i = \beta_0 + \beta_1(\text{PrWifeMore}_{ij}) + \gamma^w_i + \beta_2(\text{lnHusbIncome}_j) + (X'_j)\beta_3 + \epsilon_i \qquad (5)$$

[12] In 2008, the American Community Survey stopped asking respondents for the exact number of weeks worked. As a result, the estimation of hourly wages (wage income divided by weeks worked multipled by usual weekly hours) is less precise than in years prior to this change.

The i subscript refers the wife and the j to the husband, so that ij identifies a couple. In all specifications, γ_i^w refers to the full set of vigintiles of the wife's estimated earnings or hourly wage distribution, with i indicating the vigintile, and additional controls include the log of husbands' annual earnings as well as year fixed effects, state fixed effects, wife and husband's race, the wife and husband's five-year age-groups, and the wife and husband's educational levels, with standard errors clustered at the level of the wife's demographic group.

For non-market work, the analogous baseline equation is:

$$\begin{aligned} TotalNMWork_i = \beta_0 + \beta_1(female_i \times WifeMore_i) + \beta_2(female_i) + \beta_3(WifeMore_i) + \\ \beta_4(lnWifeInc_i) + \beta_5(female_i \times lnWifeInc_i) + \beta_6(lnHusbInc_i) + \\ \beta_7(female_i \times lnHusbInc_i) + \beta_8(lnTotInc_i) + \beta_9(female_i \times lnTotInc_i) + \\ \beta_{10}(X_i) + \beta_{11}(female_i \times X_i) + \epsilon_i \end{aligned} \tag{6}$$

Here, i simply refers to the respondent, since housework time is only observed for one member of the household all variables are interacted with the respondent's sex. Beyond those explicitly listed, controls include include year, state, and day of the week fixed effects, dummy variables for whether only the husband or wife are working respectively, the wife and husband's race and education groups, and a quadratic in the husband and wife's ages.

To adapt these specifications to use each of the five measures of actual or probable female breadwinning, I make the following changes: when the variable of interest is whether the wife actually earned more, I include controls for the log of the husband and wife's income, the log of their total income, and dummy variables for whether only the wife or only the husband had income, excluding predicted earnings distribution measures. When I instead use his observed earnings or hourly wage in constructing PrWifeMore_k along with the predicted earnings or wages for wives, controls are included for the log of his earnings or hourly wage and the full set of predicted vigintiles for wives. Where the husband's potential earnings or wages are estimated in the same way as for women, I also control for the vigintiles of his distribution, and I exclude the controls for the log of his income or wages. In all regressions, I include state and year fixed effects and controls for the race, educational attainment,[13] and five-year age groups of both the wife and the husband.[14] Finally, for regressions with non-market work, I also

[13] I use four levels of educational attainment rather than five, grouping together those who have a college degree and those who have an additional higher degree.

[14] Age groups replace the quadratic in age used in the original non-market work analysis.

include day of the week fixed effects. When the wife's earnings are predicted, standard errors are clustered at the level of her demographic group, and when both are predicted they are clustered using the combination of their demographic groups.

5.2.2 Results

For both labor force participation and housework, I find that switching to variables that use predicted income reduces the magnitude of the original effect and, in some specifications, flips the sign.

Table 1 presents the results for labor force participation. Col. 1 in Table 1 uses an indicator for female breadwinning. The coefficient on WifeMore is negative and significant at -0.016. Col. 2 displays the result for exactly replicating BKP's cross-sectional analysis, given my sample.[15] The coefficient on PrWifeMore_1 is -0.20, which is the same sign but more negative than what they find (about -0.14), perhaps owing to that I use slightly different years. In Col. 3, where I use hourly wages but nevertheless compare wives' estimated earnings distribution to the husband's observed wage, the coefficient on PrWifeMore_2 is -0.09, still significant and negative but smaller in magnitude. In Col. 4, where the key measure is constructed using estimated distributions of annual earnings for both the husband and wife, the coefficient on PrWifeMore_3 is yet again smaller but also still negative and significant, with a point estimate of -0.052. Finally, in Col. 5, where I use estimated distributions of hourly wages for both the husband and the wife, the coefficient on PrWifeMore_4 is 0.059, which is statistically significant at the 1% level.

Results for non-market work using the ATUS, displayed in Table 2, exhibit a similar pattern of sign-switching. In this setting, all variables are interacted with a dummy for "Female" to capture the differential effects that female breadwinning may have for husbands or the wives, given that we only observe the time use of one member from the couple. Thus, for each regression, the main measure of female breadwinning and its interaction with female are reported. The total marginal effect of female breadwinning for wives implied by this approach is $\beta_1 + \beta_2(FemaleBreadwinningMeasure)$, for a given measure. Since the PrWifeMore measures are continuous, the implied marginal effects vary linearly over the distribution of PrWifeMore.

Col. 1 replicates BKP's original estimation, from Col. 3 of Table 6 in their paper. As in the original result, the coefficient on $F \times WifeMore$ is positive, though in this case it is significant only at the 10% level. The coefficient on WifeMore is close to zero and not significant. Putting these together, female breadwinning is associated with more market work for women (compared to non-breadwinning women) but not for men.

[15] This regression corresponds to Table 2, Col. 2 in their paper

Without making too much of the exact implied marginal effects, for the coefficients in Cols. 2 and 4, where the measures used include the husband's observed earnings or wage, the interactions with female are positive (and significant for Col. 4), while the coefficients on the measure without the interaction are negative. For Col. 2, using predicted earnings, neither coefficient is not significantly different from zero. Finally, for Col. 5 predicted values of the husband's and wife's wages are used, the signs flip, and the marginal effect of female breadwinning is instead negatively associated with non-market hours for women.

5.3 Longitudinal analyses

In this section, I re-assess the longitudinal analysis from BKP using the PSID to study the relationship between female breadwinning on the one hand and divorce, housework, and labor force participation on the other. Similarly to the cross-sectional analysis, I find that using measures of female breadwinning based on observed earnings results in a negative relationship with wives' labor force participation, while measures based on predicted earnings instead yield precise null coefficients. Similarly, where using an indicator variable for female breadwinning picks up a slight positive relationship with divorce, for all other measures the relationship is not statistically different from zero. For housework, none of the results yield estimates that are statistically different from zero. Although the cross-sectional results flipped signs, this may have owed to correlation between the demographic characteristics used to construct predicted measures of earnings and unobserved characteristics that also affect preferences over market work or housework. This unobserved variation is likely removed in using couple fixed effects, accounting for the null results using longitudinal data.

5.3.1 Estimation

BKP use the following estimating equation for the longitudinal analysis of labor force participation and housework:

$$y_{i,t} = \beta_1(\text{WifeMore}_{i,t-1}) + X'_{i,t-1}\beta_k + \delta_t + [\mu_{i,j}] + \epsilon_{i,t}$$

In BKP's analysis the dependent variables of interest include labor force participation, the intra-couple gap in weekly housework hours (wives' hours minus husbands' hours) and divorce. The baseline specification from BKP includes the husband and the wife's respective log earnings, the log of their total earnings, indicators for whether only the husband or the wife is working, quadratics in the husband and wife's respective ages, year and state fixed effects, and couple fixed effects.

For analyzing divorce, BKP hypothesize that since divorce takes time, there may be a longer lag before effect materialize. As such, they include an additional lag of WifeMore as well as lags for all the other income variables.

For the other four measures, I use equations along the following lines:

$$y_{i,t} = \beta_0 + \beta_1(\text{PrWifeMore}_{i,j,t}) + \gamma^w_{i,t} + \gamma^m_{j,t} + (X'_{i,j,t})\beta_3 + [\mu_{ij}] + \epsilon_{i,t} \tag{7}$$

Given that these measures are based on specific demographic characteristics, the controls for these specifications are more in line with the cross-sectional analyses in the previous section. Specifically, each regression includes year fixed effects, state fixed effects, wife and husband's race, wife and husband's five-year age-groups, and wife and husband's educational levels, with standard errors clustered at the level of the wife's demographic group, as well as a dummy variable for whether the couple has a child. If the measure of interest is based on husbands' observed earnings or wage, $\gamma^m_{i,d,t}$ represents controls for the log of his earnings or hourly wage, respectively. If earnings or wages are instead predicted for member s in the couple with demographic group d, then $\gamma^s_{i,d,t}$ represents the estimated vigintiles of the earnings or hourly wage distribution. I will estimate specifications both with and without couple fixed effects, $\mu_{i,j,t}$.

While BKP's PSID sample includes couples where at least one individual in the couple had positive income, the ACS sample includes those where the husband had strictly positive income. To provide a consistent sample across specifications, I use the latter restriction.

Since BKP use income measures from time $t-1$ and $t-2$ in the analysis of divorce, I likewise include lags of PrWifeMore and its variations, as well as and lags of the associated vigintiles of the predicted earnings distribution or husbands' log earnings or wage. I report coefficients for both time periods.

5.3.2 Results

Table 3, 4, and 5 present the results for labor force participation, housework, and divorce, respectively using the PSID. Within each table, Panel (a) shows results estimated without couple fixed effects, while Panel (b) adds these.

For labor force participation, Panel (a) of Table 3 presents is strikingly similar to the analogous cross-sectional results using the Census/ACS in Table 1. The coefficient on $WifeMore$, the indicator variable for female breadwinning, is -0.026 (compared to -0.016 in Table 1). The measure BKP use in their cross-sectional analysis, PrWifeMore, yields a coefficient of -0.142 (compare to -0.197), and changing this key measure to make it either based on predicted earnings or hourly wages makes it less negative. Again similar to the cross-sectional results, the change to use predicted rather than

observed earnings for husbands makes a bigger difference than solely the switch to hourly wages: the measure using hourly observed wages for husbands, PrWifeMore_{Hr}, still has a negative and statistically significant coefficient of -0.10. One notable difference is that both measures using predicted earnings or wages have coefficients that are not statistically different from zero, which may owe to that the PSID has less cross-sectional variation than the Census/ACS.

Panel (b) follows an identical pattern, but all coefficients have smaller magnitudes, which is understandable given that couple fixed effects absorb quite some variation. The coefficient on WifeMore in Panel (b), -0.019, is a replication of the specification in Table 5, Panel A, Col. 1 of BKP, which likewise is -0.019.

For the intra-couple gap in weekly housework hours in Table 4, coefficients on WifeMore are not significant either with or without couple fixed effects, which is consistent with the fact that this specification does not yield a significant result in BKP either (the coefficient on housework only becomes statistically different from zero when more detailed income controls are added in Cols. 2 and 3 of their Table 5, Panel C). In contrast, when using BKP's predicted measure of female breadwinning that includes husbands' observed earnings, PrWifeMore the coefficients are in both cases significant and positive: 1.8 hours without fixed effects and 1.12 hours with fixed effects. Any of the other changes, either to use predicted earnings or hourly wages for the husband, or only to switch to hourly wages for the husband, yield coefficients that are not statistically different from zero.

For divorce, with results displayed across two pages for Table 5, two lags result in twice as many coefficients, but results can be summarized succinctly: all coefficients with the exception of two are statistically indistinguishable from zero and small in magnitude. The first is the coefficient on $\text{PrWifeMore}^{*}_{t-1,Hr}$, the first lag of the measure based on predicted hourly wages for both husbands and wives. The coefficient is in fact negative, at -0.01, though only significant at the 10% level. Taking this estimate at face value, it suggests when comparing between couples, those couples where the wife has higher relative earnings potential in fact have a lower risk of divorce. This interpretation is in fact consistent with results from Schwartz and Goñalons-Pons (2016). Once couple-fixed effects are accounted for, the analogous coefficient in Panel (b) is effectively zero, suggesting that the result in Panel (a) is a consequence of selection into marriage, which is also consistent with the interpretation that Schwartz and Goñalons-Pons (2016) make.

The only other coefficient that is statistically significant is that corresponding to BKP's original specification, on WifeMore_{t-2}, the second lag of the indicator for female breadwinning. This coefficient is 0.004, which compares to BKP's estimate of 0.005 in Table 5, Panel B, Col. 1. It is perhaps not surprising that none of the other specifications yield significant coefficients, as even in BKP's original results these coefficients is no

longer significant after the inclusion of additional controls.

5.4 Robustness

The previous section found that a predicted measure of the likelihood of female breadwinning based on hourly wages for husbands and wives, $\text{PrWifeMore}^{*}_{Hr}$, yielded a precise null relationship with wives' labor force participation, the intra-couple gap in housework, and divorce in the longitudinal setting with couple fixed effects. In this section, I consider robustness exercises to assess whether there are alternative of why we find a null result with this measure.

5.4.1 Fixing the sample to couples with constant characteristics

If couples endogenously change characteristics that are used to estimate the predicted measure of the likelihood of female breadwinning, the null result may be positively biased. It could be, for example, that particularly motivated or skilled couples move to states with faster wage growth, or increase their educational attainment. Among the characteristics used for constructing PrWifeMore, race, age, and sex are generally fixed, while education and state could be endogenous.

Among couples for whom PrWifeMore is defined, about 13% do not live in the same state in the last period they are observed as a couple as in the first period they are observed. Likewise, about one-third of both husbands and wives have a higher educational degree in the last period than in the first (32% and 35% respectively). Altogether, nearly half of couples in the sample (47%) experience one of these three changes during the period they are observed in the PSID.

Panel (b) Table 6 looks at the effects of PrWifeMore on labor force participation, annual hours worked using a sample restricted to those whose observable characteristics are constant from the first period of observation as a married couple in the PSID. Given that this restriction substantially reduces the number of couples, the standard errors in these regressions are larger than the corresponding baseline regressions. Overall, however, the results are quite similar to those without the restriction, including that the coefficient for divorce without couple fixed effects is negative and significant, at -0.015. Changes in characteristics over time thus do not appear to meaningfully affect the results.

5.4.2 Variation over time

Attitudes towards gender norms have become more egalitarian over time. It's possible that the null effects from pooling all time periods are the result of heterogenous effects over time.

To address this possibility, I split the sample of couples in the PSID into quartiles by year of marriage and re-estimate the fixed effects equation for labor force participation, hours, housework, and divorce for couple-year observations with a year of marriage in each quartile.

Panels (a) to (d) of Table 8 present these results. Naturally, splitting the sample makes all standard errors larger. Even examining the point estimates, there are no evident patterns over time.

5.5 Interpreting the null results

I used PrWifeMore^*_{Hr} as the group-level proxy because it provides a close comparison with BKP's original measure that also fulfills the requirement of being uncorrelated with couples' endogenous changes in labor supply. Is this in fact a useful proxy that provides insight for our question, or would it be better to instead use a different measure? Putting it differently, can we interpret nulls associated with this variable as "precise" and informative, or are they actually uninformative? I will discuss the two aspects relevant to this question: first, how big are the implied confidence intervals in terms of the within-couple variation we observe in PrWifeMore^*_{Hr}? Second, are the aggregate changes in wages captured by PrWifeMore^*_{Hr} actually meaningfully related to whether or not wives actually earn more or have higher wages than their husbands?

When the variance in PrWifeMore^*_{Hr} is decomposed into within- and between-components, the within-component has a standard deviation of 0.15, which is only slightly less than the between-component at 0.18. We can use this summary measure to scale our null effects in terms of one-standard deviation changes in PrWifeMore^*_{Hr} to give intuition of the possible effect of a large but plausible change. For example, our estimates for labor force participation imply that the effect of a one standard deviation increase in PrWifeMore^*_{Hr} would fall within a 95% confidence interval between -0.002 and 0.004; for housework between -0.11 to 0.14 hours (corresponding to 6 fewer or 8 more minutes in the intra-couple gap per week); and for divorce between -0.0001 and 0.0001. These confidence intervals indeed seem to suggest that even for relatively large changes in PrWifeMore^*_{Hr}, the range of likely effects are quite small in magnitude.

Although these estimates may be precise, there is one additional relevant concern for considering whether the coefficients PrWifeMore^*_{Hr} are actually informative about our question of interest. In particular, do these aggregate changes in earnings actually cause changes in who in the couple earns more or has a higher wage? And if not, should we seek out a different measure with more power?

To answer this question, Table 7 considers the same specification as in the earlier longitudinal analyses in Sec. 5.3, but with the outcomes being whether the wife does

earn more or whether she has a higher wage (the latter only for those who are working positive hours).

Col. 1 gives the OLS result for whether the wife earns more income on an annual basis. The coefficient is positive and highly significant, at 0.042. With fixed effects in Col. 2, the coefficient is insignificant and close to zero (0.002). For whether the wife has a higher wage conditional on working, both coefficients are positive and significant, with 0.056 and 0.025 for the OLS and FE specifications respectively in Cols. 3 and 4. We hardly need to do the calculations to notice that if we scale these marginal effects by 0.15, a one standard-deviation increase in PrWifeMore^*_{Hr}, they will be quite small.

While at first glance, this may seem to suggest that the estimates on this variable are indeed uninformative about female breadwinning owing to exogenous changes in wives' and husbands' earnings potential, in fact we learn a very useful point: aggregate changes in wages for men and women rarely cause reversals of who in a couple has higher earnings or wages. A second implication is that exogenous reversal of breadwinners *without* an employment shock to husbands or wives is likely an exceptionally uncommon event. Folk and Rickne (2019) in fact identify such a rare type of situation, focusing on female winners of close elections that effectively receive exogenous promotions. Yet even in this situation, where the shock involves a large increase in average relative earnings for the wife, the authors largely attribute the effects they find on divorce to changes in the division of labor, since the wife's job becomes more demanding as well as higher-paid. We could likewise seek out a more specific shock to wages that has large effects for a small group of people, but by definition, such an analysis would still only be applicable to a small group in the population. Another alternative is to make a more fine-grained group-level proxy, but it's not obvious that this would yield much improvement–for example, even if we included detailed occupation codes as a characteristic, it's rare that there are differential growth patterns in occupation that would lead to a one-year increase of say, 0.1 or 0.2 in the wife's share of relative earnings.

The motivation of studying whether there is an aversion to female breadwinning goes back to understanding persistent gender inequality in labor outcomes overall. If within-couple differences in earnings are large enough to begin with that changes in aggregate wages rarely cause reversals in breadwinners, and other types of exogenous shocks rarely occur, this leads to the conclusion that an aversion to female breadwinning in itself mechanically cannot have much effect on labor force participation within marriage.

At the same time: *within-couples differences in earnings are large enough to begin with that changes in aggregate wages rarely cause reversals in breadwinners.* This fact on its own highlights that if we want to understand gender inequality in labor outcomes – for example, why women are overwhelmingly the ones to leave work or work part-time if a couple has children – we need to understand how and why such sorting patterns

arise and persist. Binder and Lam (2020) and Grow and Van Bavel (2020) show that such patterns can appear even in the absence of a breadwinner norm, but they do not actually rule out that the breadwinner norm has an impact on sorting.

6 Conclusion

In this paper, I re-examine whether an aversion to female breadwinning for married couples in the U.S. affects wives' labor force participation, the intra-couple gap in housework hours, and the risk of marital dissolution. An aversion to female breadwinning is one of the possible salient aspects of the so-called "male breadwinner norm."

To understand whether an aversion to female breadwinning affects these outcomes, we need to know how couples would respond to an exogenous increase in the wife's earnings that would make her the breadwinner. Previous work on this question has generally related observed measures of female breadwinning or relative earnings to outcomes of interest, and inferred attitudes towards female breadwinning on the basis of these estimated relationships. I outline a statistical model that demonstrates that these analyses will yield biased estimates of the true relationship of interest if observed female breadwinning occurs for other reasons that have their own relationship to the outcome. I provide descriptive evidence that shows this potential bias is likely a real practical concern: men's reductions in work, both due to layoffs (which might be considered exogenous), and due to endogenous, unobserved decisions, contribute substantially to the variation in observed female breadwinning. Moreover, men's low levels of work for a variety of reasons generally appear to be linked to lower levels of work for their wives, compared to wives of husbands normally employed or working full-time.

I use BKP as a frame of reference to show how this bias affects results, whether using measures that are wholly or partially constructed using couples' observed earnings. Whereas BKP found negative relationships between their measures and wives' labor force participation, and positive relationships with housework and the incidence of divorce, the results using measures based on predicted earnings yield either relationships of the opposite sign (for cross-sectional analyses) or precise zeros (for longitudinal analyses with couple fixed effects, my preferred specification). This supports the notion that we should be cautious of interpreting estimates on measures of observed earnings as informative about reactions to female breadwinning in itself. These findings are relevant beyond BKP's specific results for the broader literature in sociology and economics studying the effects of relative earnings. The problem of bias laid out here may be present in any studies that use relative earnings as their key variable of interest.

Finally, I show that while there are stronger grounds to draw conclusions based on estimates with measures of predicted earnings compared to observed earnings, in the

end aggregate changes in wages rarely change who in a couple is the breadwinner. This fact underscores that an aversion to female breadwinning probably has little impact on labor decisions *after marriage*, and points to the importance of understanding the sorting decisions that give rise to the initial within-couple gaps in earnings and wages. Norms may yet be part of that story.

Understanding these margins by which gender norms operate is relevant for policy intended to reduce gender inequality and encouraging women to enter the labor force. Policies like parental leave reforms, childcare subsidies, and daycare expansions have been studied as possible ways to reduce gender inequality through encouraging mothers to work and fathers to spend more time on childcare, which could reduce the "child penalty", now one of the biggest contributors to gender earnings differences in high-income countries.[16] However, this body of work suggests these types of policies have limited or mixed effects on the gender earnings and wage gaps (e.g. Nix et al. (2019); Havnes and Mogstad (2011); Bettendorf, Jongen, and Muller (2015)

By the time they have children, most individuals have already made an initial choice of partner with which to have children, whether or not they remain together. If much of the effect of norms occurs at the sorting stage, then policies incentivizing labor supply later in the life cycle may have a "ceiling" in terms of effectiveness. For example, if couples initially sort to avoid having a higher-earning wife, generous parental and childcare benefits are unlikely to lead to a substantial change in fathers rather than mothers choosing to work part-time to care for children, though they may increase women's labor force attachment. Given these possibilities, more attention should be given to studying how gender norms relate to sorting into couples and what policies might affect whether couples form more or less initially "equal" unions.

[16] To be sure, encouraging women to work and reducing gender inequality are only some of the goals of such policies, which are also often designed to support parents and children through the diverse and often unstable family structures that are increasingly common in many countries (Smock and Schwartz, 2020).

7 References

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

Ayhan, S. H. (2018). Married women's added worker effect during the 2008 economic crisis—the case of turkey. *Review of Economics of the Household 16*(3), 767–790.

Bardasi, E. and J. C. Gornick (2008). Working for less? Women's part-time wage penalties across countries. *Feminist Economics 14*(1), 37–72.

Başlevent, C. and Ö. Onaran (2003). Are married women in turkey more likely to become added or discouraged workers? *Labour 17*(3), 439–458.

Bertrand, M., P. Cortés, C. Olivetti, and J. Pan (2016). Social norms, labor market opportunities, and the marriage gap for skilled women. Technical report, National Bureau of Economic Research.

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

Bettendorf, L. J., E. L. Jongen, and P. Muller (2015). Childcare subsidies and labour supply—evidence from a large Dutch reform. *Labour Economics 36*, 112–123.

Binder, A. J. and D. Lam (2020). Is there a male breadwinner norm? The hazards of inferring preferences from marriage market outcomes. *Journal of Human Resources*, 0320–10803R1.

Bittman, M., P. England, L. Sayer, N. Folbre, and G. Matheson (2003). When does gender trump money? Bargaining and time in household work. *American Journal of Sociology 109*(1), 186–214.

Blau, F. D. and L. M. Kahn (2007). Changes in the labor supply behavior of married women: 1980–2000. *Journal of Labor Economics 25*(3), 393–438.

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

Brines, J. (1994). Economic dependency, gender, and the division of labor at home. *American Journal of Sociology 100*(3), 652–688.

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

Charles, K. K. and M. Stephens, Jr (2004). Job displacement, disability, and divorce. *Journal of Labor Economics 22*(2), 489–522.

Cochard, F., H. Couprie, and A. Hopfensitz (2018). What if women earned more than their spouses? An experimental investigation of work-division in couples. *Experimental Economics 21*(1), 50–71.

Cooke, L. P. (2006). "Doing" gender in context: Household bargaining and risk of divorce in Germany and the United States. *American Journal of Sociology 112*(2), 442–472.

Eliason, M. (2012). Lost jobs, broken marriages. *Journal of Population Economics 25*(4), 1365–1397.

England, P. (2011). Missing the big picture and making much ado about almost nothing: Recent scholarship on gender and household work. *Journal of Family Theory & Review 3*(1), 23–26.

Evertsson, M. and M. Nermo (2004). Dependence within families and the division of labor: Comparing Sweden and the United States. *Journal of Marriage and Family 66*(5), 1272–1286.

Farré, L. and F. Vella (2007). The intergenerational transmission of gender role attitudes and its implications for female labor force participation. Technical report.

Fauser, S. (2019). Time availability and housework: The effect of unemployment on couples' hours of household labor. *Social science research 83*, 102304.

Fernandes, R. and F. d. Felício (2005). The entry of the wife into the labor force in response to the husband's unemployment: a study of the added worker effect in brazilian metropolitan areas. *Economic Development and cultural change 53*(4), 887–911.

Fernandez, R. (2007). Women, work, and culture. *Journal of the European Economic Association 5*(2-3), 305–332.

Fernández, R., A. Fogli, and C. Olivetti (2004). Mothers and sons: Preference formation and female labor force dynamics. *The 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.

Folk, O. and J. Rickne (2019). All the single ladies: Job promotions and the durability of marriage. *American Economic Journal: Applied Economics Forthcoming*.

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.

Fortin, N. M. (2015). Gender role attitudes and women's labor market participation: Opting-out, aids, and the persistent appeal of housewifery. *Annals of Economics and Statistics* (117-118), 379–401.

Foster, G. and L. S. Stratton (2018). Do significant labor market events change who does the chores? paid work, housework, and power in mixed-gender australian households. *Journal of Population Economics 31*(2), 483–519.

Foster, G. and L. S. Stratton (2021). Does female breadwinning make partnerships less healthy or less stable? *Journal of Population Economics 34*(1), 63–96.

Fry, R. (2010). *The reversal of the college marriage gap*. Pew Research Center Washington, DC.

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

Gonalons-Pons, P. and M. Gangl (2021). Marriage and masculinity: Male-breadwinner culture, unemployment, and separation risk in 29 countries. *American Sociological Review 86*(3), 465–502. PMID: 34149053.

Gong, X. (2011). The added worker effect for married women in australia. *Economic Record 87*(278), 414–426.

Görges, L. (2018). Wage earners, homemakers, and gender identity–Using an artefactual experiment to understand couples' labour division choices. Technical report.

Gough, M. and A. Killewald (2011). Unemployment in families: The case of housework. *Journal of Marriage and Family 73*(5), 1085–1100.

Goussé, M., N. Jacquemet, and J.-M. Robin (2017). Marriage, labor supply, and home production. *Econometrica 85*(6), 1873–1919.

Greenstein, T. N. (2000). Economic dependence, gender, and the division of labor in the home: A replication and extension. *Journal of Marriage and Family 62*(2), 322–335.

Grow, A. and J. Van Bavel (2020). The gender cliff in the relative contribution to the household income: Insights from modelling marriage markets in 27 European countries. *European Journal of Population*, 1–23.

Gupta, S. (2007). Autonomy, dependence, or display? The relationship between married women's earnings and housework. *Journal of Marriage and Family 69*(2), 399–417.

Gupta, S. and M. Ash (2008). Whose money, whose time? A nonparametric approach to modeling time spent on housework in the United States. *Feminist Economics 14*(1), 93–120.

Havnes, T. and M. Mogstad (2011). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics 95*(11-12), 1455–1465.

Heckert, D. A., T. C. Nowak, and K. A. Snyder (1998). The impact of husbands' and wives' relative earnings on marital disruption. *Journal of Marriage and the Family*, 690–703.

Hederos Eriksson, K. and A. Stenberg (2015). Gender identity and relative income within households: Evidence from Sweden. Technical report.

Jalovaara, M. (2003). The joint effects of marriage partners' socioeconomic positions on the risk of divorce. *Demography 40*(1), 67–81.

Kan, M. Y. (2008). Measuring housework participation: The gap between "stylised" questionnaire estimates and diary-based estimates. *Social Indicators Research 86*(3), 381–400.

Killewald, A. (2016). Money, work, and marital stability: Assessing change in the gendered determinants of divorce. *American Sociological Review 81*(4), 696–719.

Kleven, H., C. Landais, and J. E. Søgaard (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics 11*(4), 181–209.

Kohara, M. (2010). The response of japanese wives' labor supply to husbands' job loss. *Journal of Population Economics 23*(4), 1133–1149.

Liu, G., A. Vikat, et al. (2004). Does divorce risk depend on spouses´ relative income? A register-based study of first marriages in sweden in 1981–1998. Technical report, Max Planck Institute for Demographic Research, Rostock, Germany.

Lundberg, S. (1985). The added worker effect. *Journal of Labor Economics 3*(1, Part 1), 11–37.

Manning, A. and B. Petrongolo (2008). The part-time pay penalty for women in Britain. *The Economic Journal 118*(526), F28–F51.

Mattingly, M. J. and K. E. Smith (2010). Changes in wives' employment when husbands stop working: A recession-prosperity comparison. *Family Relations 59*(4), 343–357.

McGinnity, F. and P. McManus (2007). Paying the price for reconciling work and family life: Comparing the wage penalty for women's part-time work in Britain, Germany and the United States. *Journal of Comparative Policy Analysis 9*(2), 115–134.

Morrill, M. S. and T. Morrill (2013). Intergenerational links in female labor force participation. *Labour Economics 20*, 38 – 47.

Murray-Close, M. and M. L. Heggeness (2018). Manning up and womaning down: How husbands and wives report their earnings when she earns more. *US Census Bureau Social, Economic, and Housing Statistics Division Working Paper* (2018-20).

Nix, E., M. E. Andresen, et al. (2019). What causes the child penalty? Evidence from same sex couples and policy reforms. Technical report.

Ortigueira, S. and N. Siassi (2013). How important is intra-household risk sharing for savings and labor supply? *Journal of Monetary Economics 60*(6), 650–666.

Parker, S. W. and E. Skoufias* (2004). The added worker effect over the business cycle: evidence from urban mexico. *Applied Economics Letters 11*(10), 625–630.

Perelli-Harris, B. and M. Lyons-Amos (2016). Partnership patterns in the United States and across Europe: the role of education and country context. *Social Forces 95*(1), 251–282.

Pessin, L. (2018). Changing gender norms and marriage dynamics in the United States. *Journal of Marriage and Family 80*(1), 25–41.

Prieto-Rodríguez, J. and C. Rodríguez-Gutiérrez (2000). The added worker effect in the spanish case. *Applied Economics 32*(15), 1917–1925.

Ratliff, K. A. and S. Oishi (2013). Gender differences in implicit self-esteem following a romantic partner's success or failure. *Journal of Personality and Social Psychology 105*(4), 688.

Roth, A. and M. Slotwinski (2019). Gender norms and income misreporting within households.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2018). Ipums usa: Version 8.0 [dataset]. *Minneapolis, MN: IPUMS*.

Sayer, L. C., P. England, P. D. Allison, and N. Kangas (2011). She left, he left: How employment and satisfaction affect women's and men's decisions to leave marriages. *American Journal of Sociology 116*(6), 1982–2018.

Schwartz, C. R. and P. Goñalons-Pons (2016). Trends in relative earnings and marital dissolution: Are wives who outearn their husbands still more likely to divorce? *RSF: The Russell Sage Foundation Journal of the Social Sciences 2*(4), 218–236.

Smock, P. J. and C. R. Schwartz (2020). The demography of families: A review of patterns and change. *Journal of Marriage and Family 82*(1), 9–34.

Spletzer, J. R. (1997). Reexamining the added worker effect. *Economic Inquiry 35*(2), 417–427.

Sullivan, O. (2011). An end to gender display through the performance of housework? A review and reassessment of the quantitative literature using insights from the qualitative literature. *Journal of Family Theory & Review 3*(1), 1–13.

Teachman, J. (2010). Wives' economic resources and risk of divorce. *Journal of Family Issues 31*(10), 1305–1323.

Tichenor, V. J. (2005). *Earning more and getting less*. Rutgers University Press.

Torr, B. M. (2011). The changing relationship between education and marriage in the United States, 1940–2000. *Journal of Family History 36*(4), 483–503.

Van der Lippe, T., J. Treas, and L. Norbutas (2018). Unemployment and the division of housework in europe. *Work, employment and society 32*(4), 650–669.

Vitali, A. and B. Arpino (2016). Who brings home the bacon? The influence of context on partners' contributions to the household income. *Demographic Research 35*, 1213–1244.

Voßemer, J. and S. Heyne (2019). Unemployment and housework in couples: Task-specific differences and dynamics over time. *Journal of Marriage and Family 81*(5), 1074–1090.

Winkler, A. E., T. D. McBride, and C. Andrews (2005). Wives who outearn their husbands: A transitory or persistent phenomenon for couples? *Demography 42*(3), 523–535.

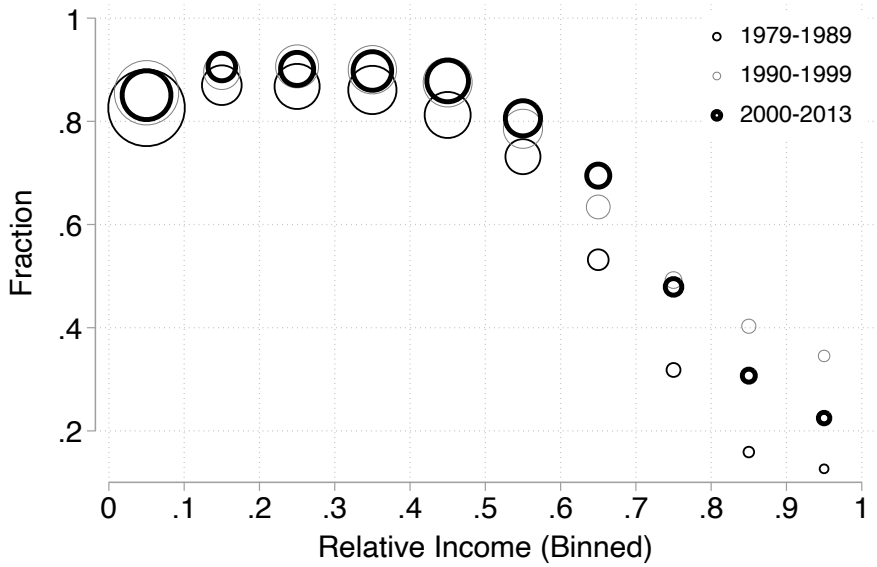
Winslow-Bowe, S. (2006). The persistence of wives' income advantage. *Journal of Marriage and Family 68*(4), 824–842.

Zinovyeva, N. and M. Tverdostup (2018). Gender identity, co-working spouses and relative income within households. Technical report, IZA.

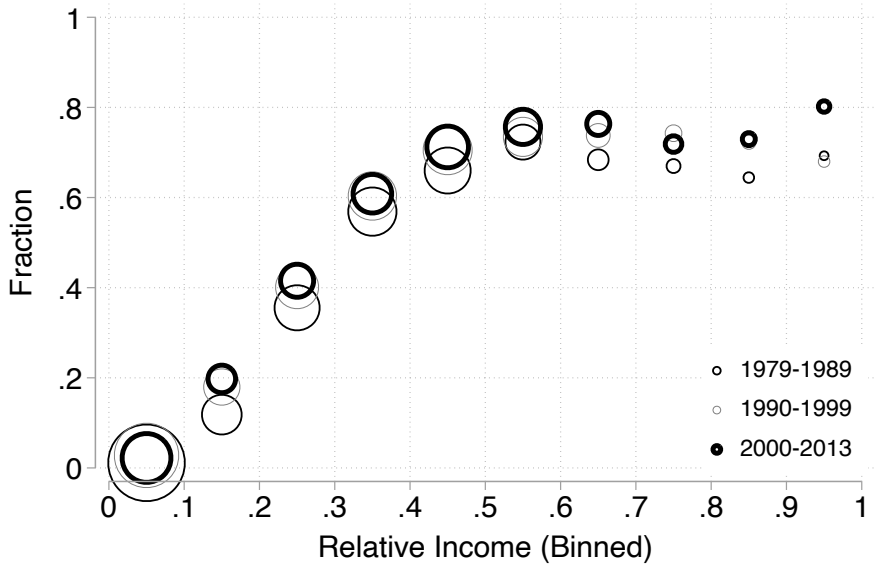
8 Figures

Figure 1: Husbands' and wives' rates of full-time work across the distribution of relative income

(a) Men's rates of full-time work



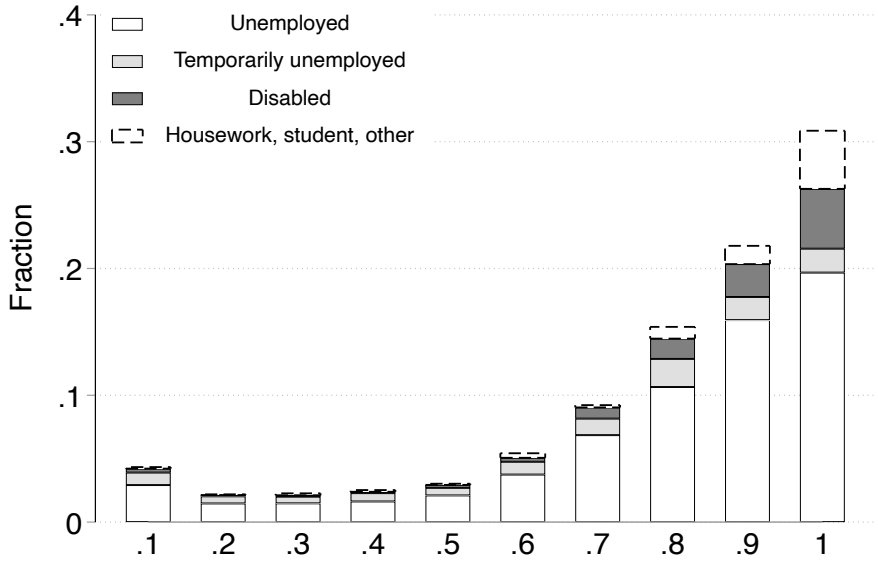
(b) Women's rates of full-time work



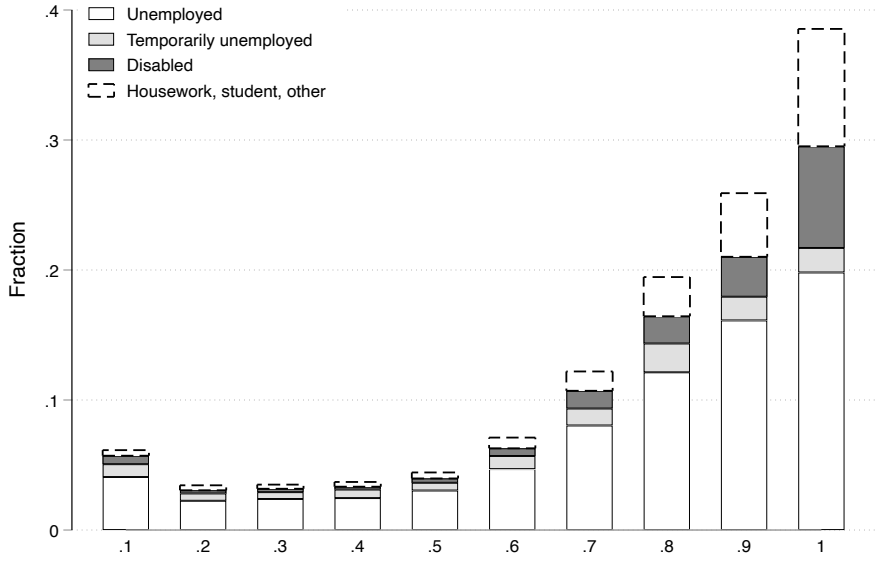
Using a sample of married couples between the ages of 18 and 64 observed between 1979 and 2013 in the Panel Study of Income Dynamics, this figure plots the rate of full-time work for wives and husbands falling within each 0.1-width bin of the distribution of relative income, calculated as the wife's share of total wage income. Relative income and hours are measured with reference to the same calendar year. Full-time work is defined as being equal to or greater than 1,820 hours in the year. The size of the markers is proportional to the number of observations within the bin.

Figure 2: Husbands' employment status in time t and $t-1$ by wife's share of relative income in the intervening year

(a) Employment status, time $t-1$

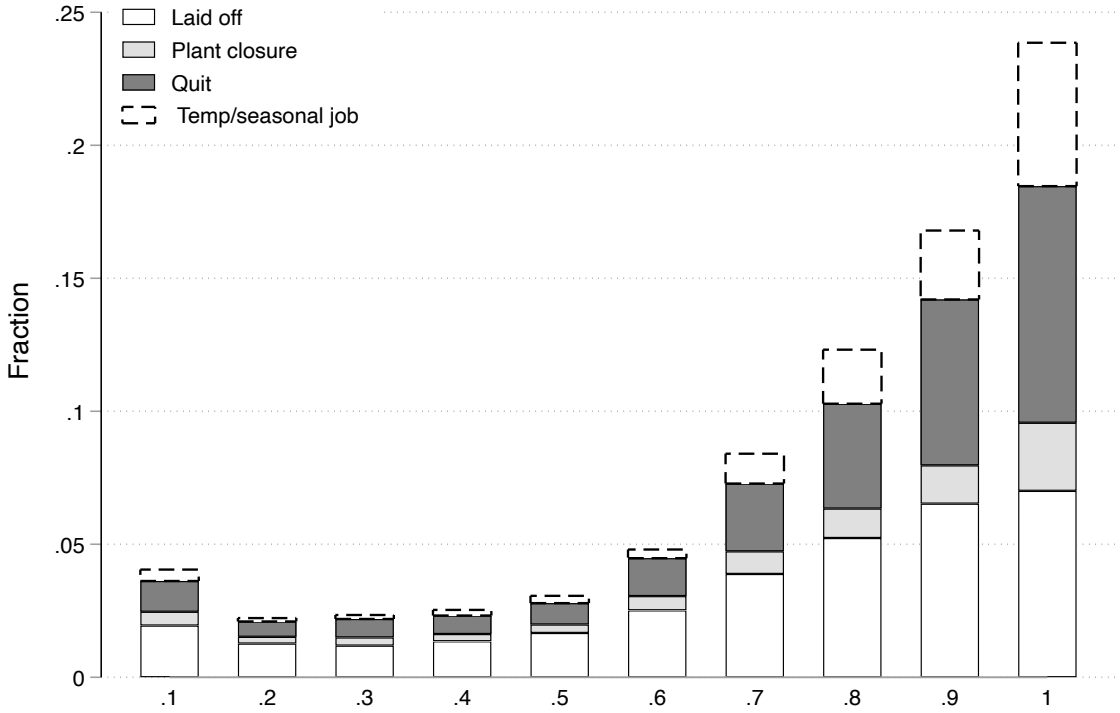


(b) Employment status, time t



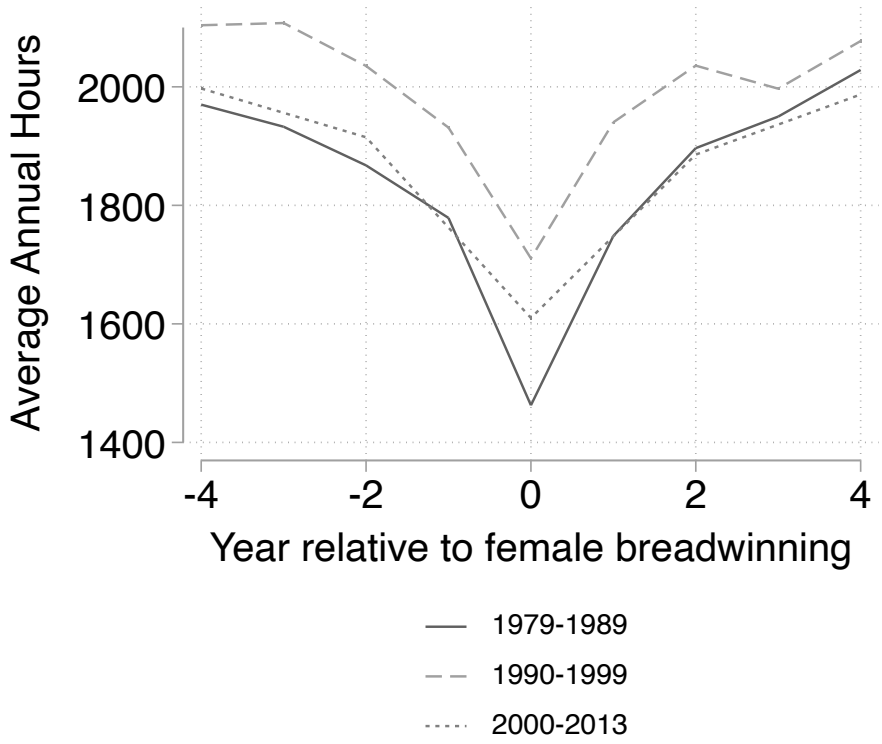
Using a sample of married couples between the ages of 18 and 64 observed between 1979 and 2013 in the Panel Study of Income Dynamics, this figure plots the fraction of husbands in each employment category within each 0.1-width bin of the distribution of relative income, calculated as the wife's share of total wage income. Relative income is for the calendar year $t-1$, while employment status is collected with respect to the time of the interviews, which are conducted between March and November.

Figure 3: Husbands' reason last job ended, if currently not in work



For individuals who are not working at the time of an interview, the Panel Study on Income Dynamics asks for the reason their last job ended in the previous calendar year. This figure displays the responses for married men between the ages of 18 to 64 married to wives in the same age range, and observed between 1979 and 2013. The figure shows the fraction of married men within each 0.1-width bins of the distribution of relative income, calculated as the wife's share of total wage income, by each possible response, treating the men who are working as having a value of zero for each possible response. If the respondent was currently unemployed and did not have a job in the prior calendar year, they are counted as long-term unemployed.

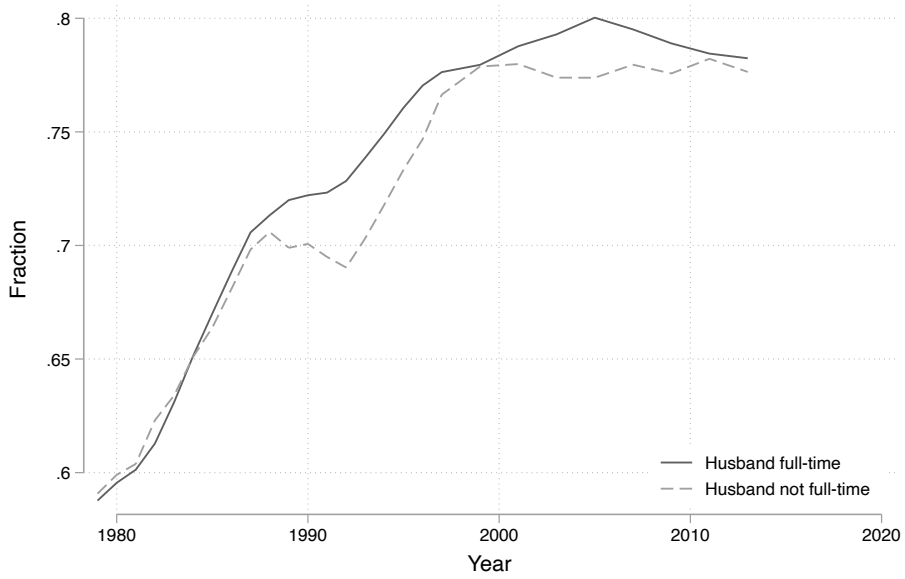
Figure 4: Changes in hours before and after female breadwinning



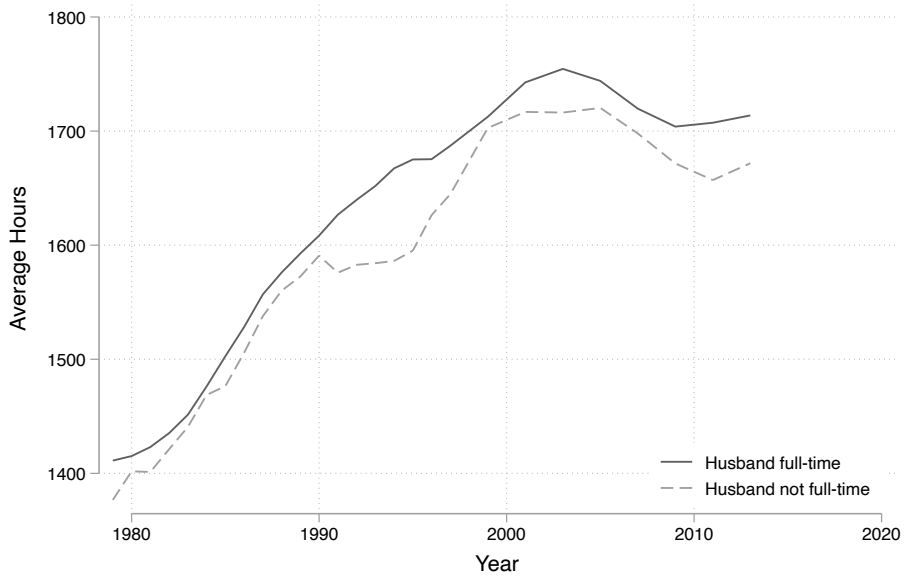
This figure shows the dynamics of husbands' annual hours of work in the four years before and after a spell of female breadwinning, using married couples between the ages of 18 and 64 observed between 1979 and 2013 in the Panel Study of Income Dynamics. For couples that were observed to experience more than one spell of female breadwinning, one instance was randomly selected as the reference period, so that each couple is represented only once.

Figure 5: Wives' average labor force participation and hours over time, by whether her husband works full-time

(a) Labor force participation



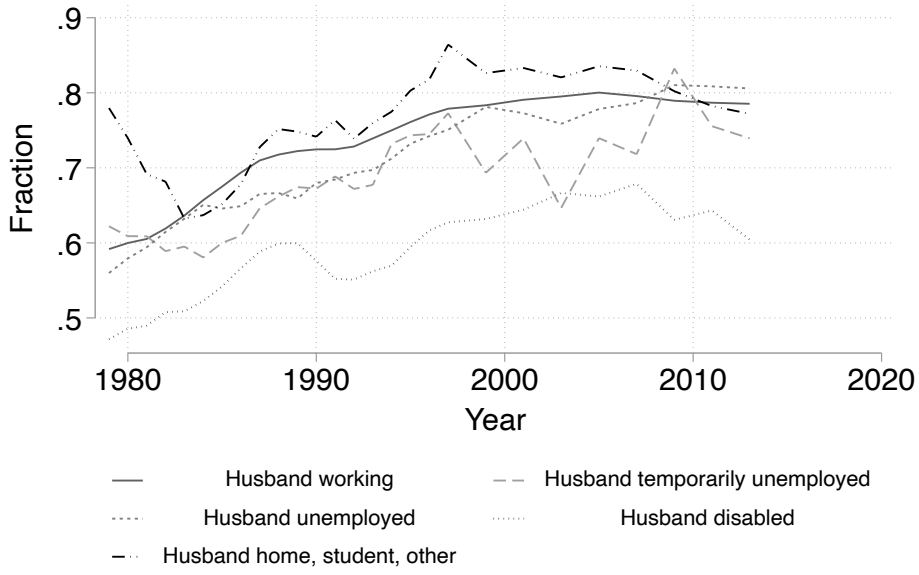
(b) Annual Hours



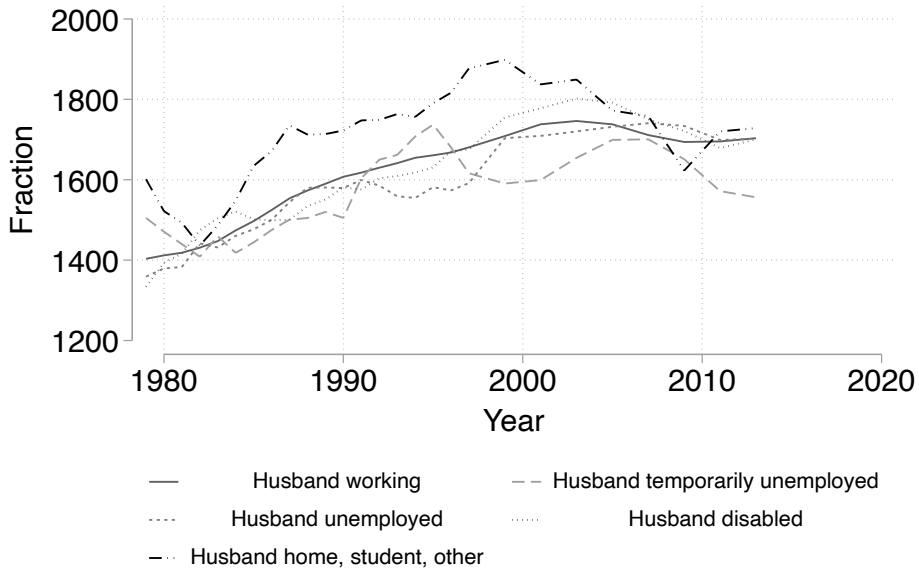
Using a sample of married couples between the ages of 18 and 64 observed between 1979 and 2013 in the Panel Study of Income Dynamics, this figure shows wives' average labor force participation rates and annual hours by whether or not her husband worked full-time. Hours and labor force participation are collected in the same interview, but hours are measured with reference to the prior calendar year, while labor force participation is with respect to the time of the interview. Raw averages by year have been smoothed using a 5-year moving average.

Figure 6: Wives' average labor force participation and hours over time, by her husband's employment status

(a) Labor force participation



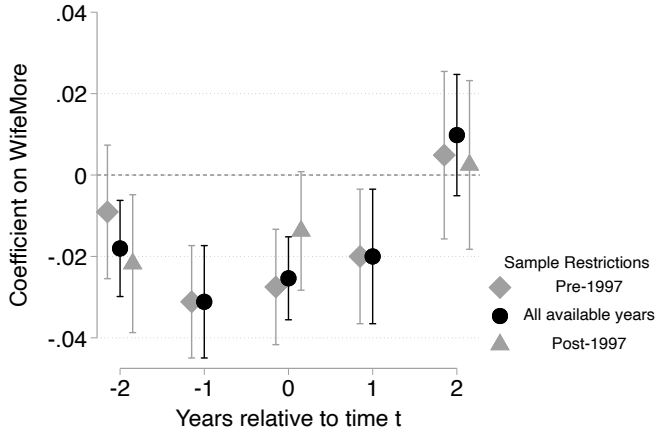
(b) Annual Hours



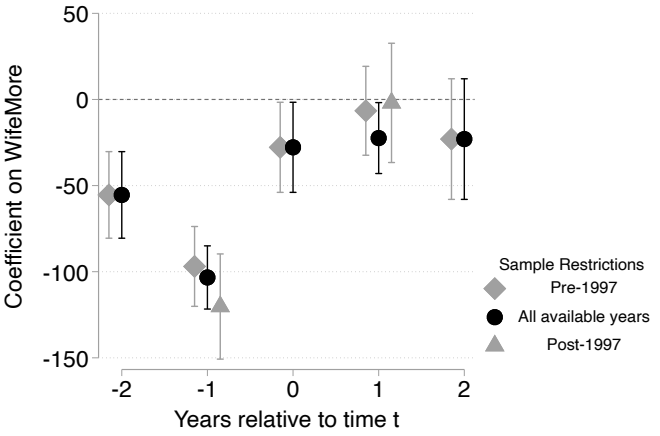
Using a sample of married couples between the ages of 18 and 64 observed between 1979 and 2013 in the Panel Study of Income Dynamics, this figure shows wives' average labor force participation rates and annual hours by her husband's employment status. Hours and labor force participation are collected in the same interview, but hours are measured with reference to the prior calendar year, while labor force participation is with respect to the time of the interview. Raw averages by year have been smoothed using a 5-year moving average.

Figure 7: Replicating BKP's longitudinal analysis before and after female breadwinning

(a) Wives' labor force participation



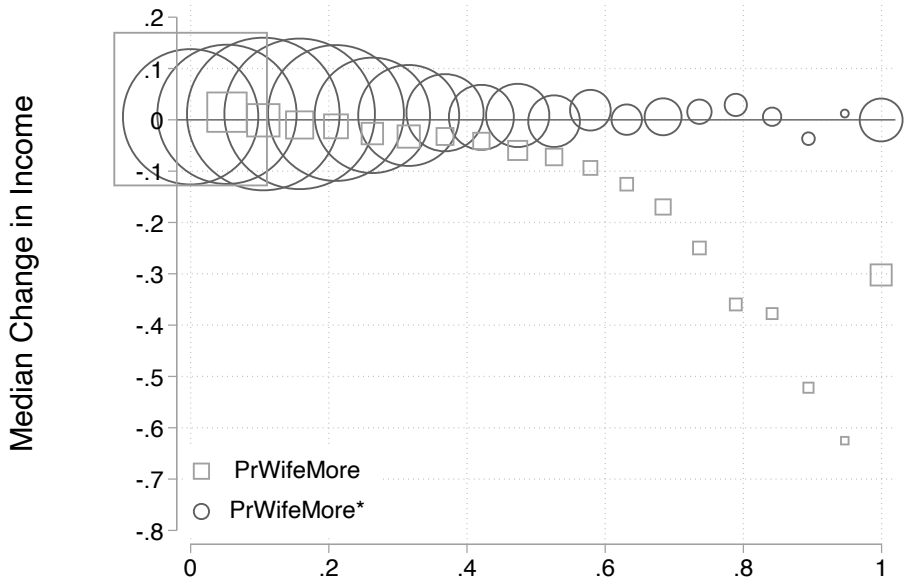
(b) Husbands' annual hours



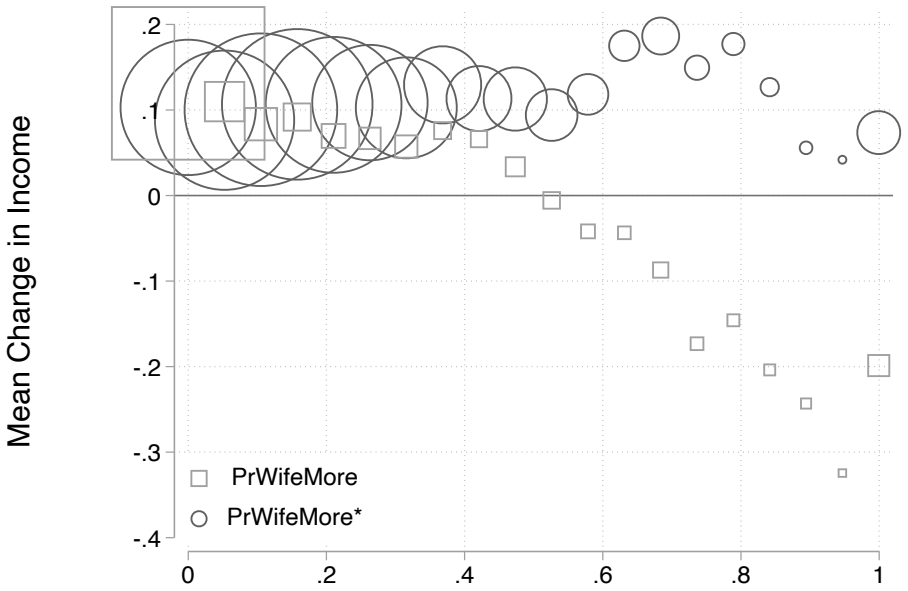
This figure is based on data from married couples between the ages of 18 and 64 from the PSID, observed between 1979-2013. It shows the coefficients on an indicator for female breadwinning in time $t-1$ from regressions on labor force participation, the intra-couple gap in weekly housework hours, and husbands' annual hours for periods $t-2$ to $t+2$. Controls are consistent with Table 5, Panel A, Col. 1 from Bertrand et al. (2015), including logs in the wife and husband's earnings and the sum of their total earnings, dummy variables for whether only the wife or only the husband had income, quadratics in the wife and husband's age, and year and state fixed effects. All controls are measured in time t, although income variables are with reference to $t-1$.

Figure 8: The relationship between changes in husbands' income and PrWifeMore

(a) Median change in husbands' income by binned values of PrWifeMore



(b) Mean change in husbands' income by binned values of PrWifeMore



This graph uses data on the income of married men between the ages of 18 and 64 from the PSID between 1979-2013, combined with information on predicted earnings based on demographic characteristics drawn from the Annual Social and Economic Supplement from the Current Population Survey to construct PrWifeMore and PrWifeMore*. Each of the markers indicate the median or mean percent change in husbands' income or wages since the last period for those couples who fall into a given bin of each PrWifeMore variable. The values between 0 and 1 are split into 20 fixed-width bins to correspond to the 20 unique values for PrWifeMore and PrWifeMore*. The size of the marker is weighted by the number of married men represented in the bin.

9 Tables

Table 1: Cross-sectional analysis of different measures of female breadwinning and labor force participation

VARIABLES	(1) LFP	(2) LFP	(3) LFP	(4) LFP	(5) LFP
WifeMore	-0.016				
	(0.001)				
PrWifeMore		-0.197			
		(0.006)			
PrWifeMore*			-0.052		
			(0.015)		
PrWifeMore_{Hr}				-0.090	
				(0.005)	
$\text{PrWifeMore}^{*}_{Hr}$					0.060
					(0.015)
Observations	4,294,326	4,294,021	4,294,021	4,294,019	4,294,019
R-squared	0.662	0.091	0.071	0.083	0.070

This table displays results of regressions using Census waves between 1980-2000 and ACS waves between 2005-2007, for women in married couples between the ages of 18 and 64 where the husband has positive income in the prior year. Wives' labor force participation is regressed on different measures of actual or predicted female breadwinning, as given in Sec. 5.1. All specifications include dummies for the year, state, educational attainment level, five-year age-group, and race of each spouse, and whether they have at least one child. For measures that are based on an estimated earnings distribution, the full set of associated vigintiles of included. For measures that are based on husband's observed earnings or wage, the log of his earnings or wage is included, while measures that are based on predicted earnings distributions for either the husband or wife also include the estimated moments of these distributions. Finally, standard errors are clustered at the level of the wife's demographic group where her earnings only are predicted, and at the level of the husband and wife's combined demographic group where earnings for both are predicted.

Table 2: Cross-sectional analysis of different measures of female breadwinning and non-market work

(a) Non-market work

VARIABLES	(1) Hours	(2) Hours	(3) Hours	(4) Hours	(5) Hours
WifeMore	0.133				
	(0.469)				
F X WifeMore	1.267				
	(0.659)				
PrWifeMore		-3.101			
		(0.968)			
FXPrWifeMore		2.453			
		(1.411)			
PrWifeMore*			0.689		
			(1.735)		
FXPrWifeMore*			0.160		
			(2.609)		
PrWifeMore$_{Hr}$				-3.825	
				(0.841)	
FXPrWifeMore$_{Hr}$				6.940	
				(1.262)	
PrWifeMore$^{*}_{Hr}$					0.743
					(1.592)
FXPrWifeMore$^{*}_{Hr}$					-1.965
					(2.072)
Observations	38,638	35,770	35,047	30,111	35,047
R-squared	0.180	0.149	0.137	0.140	0.137

This table displays results of regressions using the American Time Use Survey with the 2002 to 2011 waves, for married individuals between the ages of 18 and 65 where at least one member of the couple was working as of the time use survey. The first column replicates the specification used by Bertrand et al. (2015) in Table 6, Col. 3, with an indicator for female breadwinning. The subsequent columns consider essentially the same relationship but using slight variations on the key measure of interest, which instead proxy for the probability that a wife would earn more than her husband and vary by whether the measures are based on hourly wages or annual earnings on the one hand, and whether they used husbands' observed earnings or wages on the other. All specifications include dummies for the year, state, educational attainment level, five-year age-group, and race of each spouse. For measures that are based on an estimated earnings distribution, the full set of vigintiles of included. For measures that are based on husband's observed earnings or wage, the log of his earnings or wage is included. Finally, standard errors are clustered at the level of the wife's demographic group where her earnings only are predicted, and at the level of the husband and wife's combined demographic group where earnings for both are predicted.

Table 3: Longitudinal analysis of different measures of female breadwinning and labor force participation

(a) Without couple fixed effects

VARIABLES	(1) LFP_t	(2) LFP_t	(3) LFP_t	(4) LFP_t	(5) LFP_t
WifeMore	-0.026				
	(0.004)				
PrWifeMore		-0.142			
		(0.012)			
PrWifeMore*			-0.005		
			(0.014)		
PrWifeMore$_{Hr}$				-0.101	
				(0.011)	
PrWifeMore$^*_{Hr}$					0.014
					(0.011)
Observations	78,414	78,414	76,106	78,414	76,044

(b) With couple fixed effects

VARIABLES	(1) LFP_t	(2) LFP_t	(3) LFP_t	(4) LFP_t	(5) LFP_t
WifeMore	-0.019				
	(0.005)				
PrWifeMore		-0.054			
		(0.011)			
PrWifeMore*			-0.000		
			(0.012)		
PrWifeMore$_{Hr}$				-0.029	
				(0.010)	
PrWifeMore$^*_{Hr}$					0.005
					(0.010)
Observations	78,414	78,414	76,106	78,414	76,044
Number of FE	12,107	12,107	11,789	12,107	11,786

This table uses data on married couples between the ages of 18 and 64 in the 1979-2013 waves of the PSID to regress wives' labor force participation on measures of female breadwinning. The first column in Panel (b) replicates the specification used by Bertrand et al. (2015) in Table 5, Panel A, Col. 1, with an indicator for female breadwinning. The subsequent columns use different measures of predicted female breadwinning, as given in Sec. 5.1. Panel (b) includes couple fixed effects, while Panel (a) does not. Cols. 2-5 include dummies for the year, state, educational attainment level, five-year age-group, and race of each spouse. For measures that are based on an estimated earnings distribution, the full set of vigintiles of included. For measures that are based on husband's observed earnings or wage, the log of his earnings or wage is included. Finally, standard errors are clustered at the level of the wife's demographic group where her earnings only are predicted, and at the level of the husband and wife's combined demographic group where earnings for both are predicted.

Table 4: Longitudinal analysis of different measures of female breadwinning and labor force participation

(a) Without couple fixed effects

VARIABLES	(1) HW_{t-1}	(2) HW_{t-1}	(3) HW_{t-1}	(4) HW_{t-1}	(5) HW_{t-1}
WifeMore	-0.049				
	(0.219)				
PrWifeMore		1.833			
		(0.472)			
PrWifeMore*			-0.321		
			(0.555)		
$PrWifeMore_{Hr}$				0.256	
				(0.421)	
$PrWifeMore^{*}_{Hr}$					0.016
					(0.524)
Observations	64,819	64,819	63,161	64,819	63,116

(b) With couple fixed effects

VARIABLES	(1) HW_{t-1}	(2) HW_{t-1}	(3) HW_{t-1}	(4) HW_{t-1}	(5) HW_{t-1}
WifeMore	0.300				
	(0.231)				
PrWifeMore		1.117			
		(0.444)			
PrWifeMore*			-0.644		
			(0.512)		
$PrWifeMore_{Hr}$				0.030	
				(0.398)	
$PrWifeMore^{*}_{Hr}$					0.109
					(0.429)
Observations	64,819	64,819	63,161	64,819	63,116
Number of FE	10,034	10,034	9,863	10,034	9,861

This table uses data on married couples between the ages of 18 and 64 in the 1979-2013 waves of the PSID to regress the gap between husbands and wives in weekly housework hours on measures of female breadwinning. The first column in Panel (b) replicates the specification used by Bertrand et al. (2015) in Table 5, Panel C, Col. 1, with an indicator for female breadwinning. The subsequent columns use different measures of predicted female breadwinning, as given in Sec. 5.1. Panel (b) includes couple fixed effects, while Panel (a) does not. Cols. 2-5 include dummies for the year, state, educational attainment level, five-year age-group, and race of each spouse. For measures that are based on an estimated earnings distribution, the full set of vigintiles of included. For measures that are based on husband's observed earnings or wage, the log of his earnings or wage is included. Finally, standard errors are clustered at the level of the wife's demographic group where her earnings only are predicted, and at the level of the husband and wife's combined demographic group where earnings for both are predicted.

Table 5: Longitudinal analysis of different measures of female breadwinning and divorce

(a) Without couple fixed effects

VARIABLES	(1) Divorce$_t$	(2) Divorce$_t$	(3) Divorce$_t$	(4) Divorce$_t$	(5) Divorce$_t$
WifeMore$_{t-1}$	0.004				
	(0.003)				
WifeMore$_{t-2}$	-0.000				
	(0.002)				
PrWifeMore$_{t-1}$		0.003			
		(0.006)			
PrWifeMore$_{t-2}$		-0.002			
		(0.005)			
PrWifeMore$^*_{t-1,Hr}$			-0.003		
			(0.006)		
PrWifeMore$^*_{t-2}$			-0.002		
			(0.007)		
PrWifeMore$_{t-1,Hr}$				0.005	
				(0.005)	
PrWifeMore$_{t-2,Hr}$				0.005	
				(0.005)	
PrWifeMore$^*_{t-1,Hr}$					-0.010
					(0.006)
PrWifeMore$^*_{t-2,Hr}$					0.004
					(0.006)
Observations	57,112	54,690	53,130	63,493	53,072

(b) With couple fixed effects

VARIABLES	(1) Divorce_t	(2) Divorce_t	(3) Divorce_t	(4) Divorce_t	(5) Divorce_t
WifeMore_{t-1}	0.003				
	(0.002)				
WifeMore_{t-2}	0.004				
	(0.002)				
PrWifeMore_{t-1}		0.005			
		(0.005)			
PrWifeMore_{t-2}		0.001			
		(0.005)			
$\text{PrWifeMore}^*_{t-1,Hr}$			0.000		
			(0.006)		
$\text{PrWifeMore}^*_{t-2}$			-0.004		
			(0.006)		
$\text{PrWifeMore}_{t-1,Hr}$				0.007	
				(0.005)	
$\text{PrWifeMore}_{t-2,Hr}$				0.004	
				(0.005)	
$\text{PrWifeMore}^*_{t-1,Hr}$					-0.000
					(0.005)
$\text{PrWifeMore}^*_{t-2,Hr}$					0.003
					(0.006)
Observations	57,112	54,690	53,130	53,833	53,072
Number of FE	8,780	8,422	8,268	8,356	8,264

This table uses data on married couples between the ages of 18 and 64 in the 1979-2013 waves of the PSID to regress the gap between husbands and wives in weekly housework hours on measures of female breadwinning. The first column in Panel (b) replicates the specification used by Bertrand et al. (2015) in Table 5, Panel C, Col. 1, with an indicator for female breadwinning. The subsequent columns use different measures of predicted female breadwinning, as given in Sec. 5.1. Panel (b) includes couple fixed effects, while Panel (a) does not. Cols. 2-5 include dummies for the year, state, educational attainment level, five-year age-group, and race of each spouse. For measures that are based on an estimated earnings distribution, the full set of vigintiles of included. For measures that are based on husband's observed earnings or wage, the log of his earnings or wage is included. Finally, standard errors are clustered at the level of the wife's demographic group where her earnings only are predicted, and at the level of the husband and wife's combined demographic group where earnings for both are predicted.

Table 6: Couples with constant education and state since marriage

VARIABLES	(1) LFP_t	(2) LFP_t	(3) HW_t	(4) HW_t	(5) $Divorce_t$	(6) $Divorce_t$
$PrWifeMore^*_{Hr}$	0.014	0.011	-0.533	0.295	-0.015	0.001
	(0.014)	(0.013)	(0.555)	(0.509)	(0.007)	(0.006)
Observations	50,560	50,560	48,446	48,446	44,349	44,349
Couple FE	NO	YES	NO	YES	NO	YES
Number of FE		8,950		8,870		8,257

This table uses data on married couples between the ages of 18 and 64 in the 1979-2013 waves of the PSID. This table re-estimates the results in Col. 5 of Tables ??, ??, and ??, excluding couples where either member exhibits a change in educational attainment or state during they are observed in the sample.

Table 7: Likelihood that wife is observed to have higher earnings or wages

VARIABLES	(1) WifeMoreInc	(2) WifeMoreInc	(3) WifeMoreWage	(4) WifeMoreWage
$PrWifeMore^*_{Hr}$	0.042	0.002	0.056	0.025
	(0.010)	(0.008)	(0.013)	(0.012)
Observations	81,186	81,186	59,286	59,286
Couple FE	NO	YES	NO	YES
Number of FE		12,188		10,558

This table uses data on married couples between the ages of 18 and 64 in the 1979-2013 waves of the PSID to estimate the relationship between $PrWifeMore^*_{Hr}$ and whether the wife has higher observed earnings or wages than her husband. The first and third columns use OLS, with standard errors clustered at the wife and husband's combined demographic groups, while the second and fourth columns include couple fixed-effects. In the third and fourth columns, the sample is restricted to those couples with both members working positive hours. All specifications include the full sets of vigintiles of the wife and husband's estimated earnings distributions in a given year (calculated using corresponding CPS waves) dummy variables for year, wife and husband's age group, educational attainment, and race, as well as dummies for state, and whether the couple has a child.

Table 8: Effects of Potential Relative Income by Marriage Cohort

(a) Labor force participation

VARIABLES	(1) LFP_t	(2) LFP_t	(3) LFP_t	(4) LFP_t
$PrWifeMore^*_{Hr}$	0.015	0.018	-0.012	-0.009
	(0.021)	(0.021)	(0.020)	(0.024)
Observations	21,208	18,892	19,290	17,327
Number of FE	2,141	1,729	2,345	4,193
Year Married	<1972	1972-1981	1982-1990	1991-2013

(b) Hours

VARIABLES	(1) $Hours^w_t$	(2) $Hours^w_t$	(3) $Hours^w_t$	(4) $Hours^w_t$
$PrWifeMore^*_{Hr}$	50.699	-15.330	15.326	21.835
	(38.459)	(38.985)	(40.672)	(53.941)
Observations	19,474	17,342	17,163	13,422
Number of FE	2,025	1,599	2,131	3,397
Year Married	<1972	1972-1981	1982-1990	1991-2013

(c) Gap in weekly housework hours

VARIABLES	(1) HW_t	(2) HW_t	(3) HW_t	(4) HW_t
$PrWifeMore^*_{Hr}$	0.598	0.193	-0.064	-0.580
	(0.890)	(0.821)	(0.730)	(0.813)
Observations	19,774	17,780	19,037	17,138
Number of FE	2,138	1,719	2,333	4,182
Year Married	<1972	1972-1981	1982-1990	1991-2013

(d) Divorce

VARIABLES	(1) $Divorce_t$	(2) $Divorce_t$	(3) $Divorce_t$	(4) $Divorce_t$
$PrWifeMore^*_{Hr}$	-0.006	0.004	-0.005	-0.012
	(0.006)	(0.008)	(0.008)	(0.012)
Observations	21,208	18,892	19,290	17,327
Number of FE	2,141	1,729	2,345	4,193
Year Married	<1972	1972-1981	1982-1990	1991-2013

This table uses data on married couples between the ages of 18 and 64 in the 1979-2013 waves of the PSID. It carries out the same analysis as the fixed effects specification in Table **??** for different marriage cohorts, where marriage cohorts are defined by splitting the sample into quartiles by year of marriage.

A Summary statistics

Table A1: ACS/Census Summary Statistics

(a) Individual Characteristics

	Wives		Husbands	
	Mean	St. Dev.	Mean	St. Dev.
Demographics				
White	0.87	(0.33)	0.88	(0.33)
Black	0.06	(0.23)	0.06	(0.24)
Other Races	0.07	(0.25)	0.06	(0.24)
Age	39.12	(10.14)	41.42	(10.45)
Hispanic	0.07	(0.26)	0.07	(0.26)
Education				
<HS	0.12	(0.33)	0.14	(0.34)
HS Grad	0.43	(0.49)	0.37	(0.48)
SomeColl	0.23	(0.42)	0.22	(0.41)
College	0.22	(0.42)	0.27	(0.44)
Income and Work				
Hours (Annual)	1226	(937)	2211	(552)
Wage Income (Annual)	20149	(22403)	52445	(35457)

(b) Household Characteristics

Any Children	0.70	(0.46)
No. of Children	1.40	(1.24)
Total HH Income	79538.11	(47632.66)
Wife earns more	0.14	(0.35)
Wife higher wage	0.20	(0.40)
ProbWifeMore_1	0.17	(0.24)
ProbWifeMore_2	0.25	(0.28)
ProbWifeMore_3	0.16	(0.11)
ProbWifeMore_4	0.24	(0.13)
Observations	4293986	

The table presents summary statistics for married couples' individual and household characteristics from Census waves 1980 to 2000 and ACS waves 2005-2007. Couples are included if they are between the ages of 18 and 64. Income and wages are given in 2010 dollars.

Table A2: PSID Summary Statistics

(a) Individual Characteristics

	Wives		Husbands	
	Mean	St. Dev.	Mean	St. Dev.
Demographics				
Age	37.17	(10.07)	39.35	(10.34)
Non-White	0.20	(0.40)	0.19	(0.40)
Education				
<HS	0.11	(0.31)	0.14	(0.34)
HS Grad	0.43	(0.49)	0.39	(0.49)
SomeColl	0.24	(0.43)	0.22	(0.41)
College	0.22	(0.42)	0.26	(0.44)
Income and Work				
Wage Income (Annual)	23196	(26359)	58452	(70776)
Work Hours (Annual)	1273	(905)	2209	(661)
Housework Hours (Weekly)	22	(15)	7	(8)

(b) Household Characteristics

Any Children	0.67	(0.47)
No. of Children	1.32	(1.24)
Total HH Income	87770.70	(86752.55)
Wife earns more	0.19	(0.39)
Wife higher wage	0.22	(0.41)
ProbWifeMore_1	0.12	(0.26)
ProbWifeMore_2	0.15	(0.28)
ProbWifeMore_3	0.21	(0.19)
ProbWifeMore_4	0.28	(0.21)
Observations	93263	

The table presents summary statistics for married couples' individual and household characteristics from the PSID between 1979-2013. Couples are included if they are between the ages of 18 and 64. Income and wages are given in 2010 dollars.

Table A3: ATUS Summary Statistics

(a) Individual Characteristics

	Wives		Husbands	
	Mean	St. Dev.	Mean	St. Dev.
Demographics				
Age	40.83	(9.91)	42.98	(10.08)
Non-White	0.13	(0.34)	0.13	(0.34)
Education				
<HS	0.07	(0.25)	0.08	(0.27)
HS Grad	0.24	(0.43)	0.25	(0.43)
SomeColl	0.28	(0.45)	0.26	(0.44)
College	0.41	(0.49)	0.41	(0.49)
Income and Work				
Weekly Earnings	543	(633)	1112	(914)
Paid Work Hours	20	(29)	33	(35)
Nonmarket Hours	26	(20)	17	(19)

(b) Household Characteristics

Any Children	0.71	(0.45)
No. of Children	1.37	(1.17)
Wife earns more	0.25	(0.44)
ProbWifeMore_1	0.30	(0.34)
ProbWifeMore_2	0.35	(0.35)
ProbWifeMore_3	0.25	(0.17)
ProbWifeMore_4	0.32	(0.18)
Observations	38638	

The table presents summary statistics for married couples' individual and household characteristics from the American Time Use Survey for waves 2002 to 2011. Couples are included if they are between the ages of 18 and 65. Income and wages are given in 2010 dollars.

B Bias owing to non-classical measurement error

In Sec. 4.1, I argue that if female breadwinning occurs under different situations that each have their own effect on wives' labor force participation, regressing labor force participation on observed female breadwinning will pick up a variance-weighted average of these individual effects. Here I prove these points in the case of two variables.

Suppose the true model is as follows:

$$P = \beta_0 + \beta_1(a) + \beta_2(b) + \epsilon \tag{8}$$

Let us make the following assumptions: $cov(a, b) = cov(a, \epsilon) = cov(b, \epsilon) = 0$. To be clear, $cov(a, P), cov(b, P) \neq 0$.

Because of limitations in our ability to distinguish a from b, we observe only $F = a + b$. With this, we instead estimate:

$$P = \beta_0 + \beta_1(F) + \mu \tag{9}$$

Given what we know about the true model, the error term $\mu = \epsilon + (\beta_2 - \beta_1)(b)$.

Thus, estimating Eq. 9 will give the following:

$$\begin{aligned}
\widehat{\beta_1} &= \frac{cov(a+b, P)}{var(a+b)} \\
&= \frac{cov(a+b, \beta_0 + \beta_1(a) + \beta_2(b) + \epsilon)}{var(a+b)} \\
&= \beta_1 \left(\frac{var(a)}{var(a) + var(b)} \right) + \beta_2 \left(\frac{var(b)}{var(a) + var(b)} \right)
\end{aligned} \tag{10}$$

We can further simplify by letting $\pi = \frac{var(a)}{var(a)+var(b)}$. From this, we get the expression:

$$\widehat{\beta_1} = \beta_1(\pi) + \beta_2(1 - \pi) \tag{11}$$

From this simple expression, we can notice several interesting features. First, if $\beta_1 = \beta_2$, we have that $\widehat{\beta_1} = \beta_1$. Second, π determines whether $\widehat{\beta_1}$ is closer to β_1 or to β_2. When $\pi = 0.5$, $\widehat{\beta_1}$ is exactly the average of the two true parameters. As π approaches 1 (in other words, as the variance of b goes to zero), $\widehat{\beta_1}$ likewise approaches β_1. Conversely, as π approaches 0, $\widehat{\beta_1}$ instead goes to β_2. Finally, we cannot make a general claim about the direction of the overall bias. If β_1 and β_2 are of the same sign, the bias will be away from zero if $|\beta_2| > |\beta_1|$, and towards zero if $|\beta_2| < |\beta_1|$. If they are of the opposite sign, anything is possible, even flipping signs.

C Testing Binder and Lam's robustness checks on BKP's results

In Binder and Lam (2020), the authors demonstrate that the point mass of couples with equal income drives BKP's result when testing for a discontinuity at 0.5 in the distribution of wives' relative income among married couples. They propose that when studying gender norms, particularly one such as the male breadwinner norm, it would be preferable to examine joint distributions of the variable of interest (like relative income or female breadwinner status) with other household outcomes rather than inferring norms from univariate distributions. They acknowledge that BKP in fact do this in the sections of their paper that study the relationship between female breadwinning and within-marriage outcomes, but they also propose that such analyses should carry out additional robustness checks to test for how the point mass of couples with equal income may affect results, given its centrality in the discontinuity result. Their concern is that, conditional on extensive controls for the couples' income, the residual variation might identify a comparison between couples where the wife earns just more than half and where the wife earns exactly half of income, in which case the couples that have exactly equal income might play an important role in the results.

Given that this paper is focused on a re-assessment of precisely these within-marriage results, I carry out some of these robustness tests to evaluate whether in fact this point mass of couples is important for these results from BKP as well.

Binder and Lam propose four robustness checks for analyses where the key variable is an indicator for female breadwinning:

1. Redefining the indicator variable for female breadwinning to represent couples where the wife has greater *or equal* income to her husband, rather than only strictly greater income
2. Exclude observations where the couples have exactly equal income
3. Adding detailed occupational controls or controls for husbands and wives working in the same occupation
4. Controlling for how much of income is business income

If the point mass of couples with equal income meaningfully affects the results, then the first two robustness checks should certainly affect results. These are the robustness checks I test. I replicate the specification from Col. 1 of Table 7 from BKP for the wife's labor force participation, the intra-household gap in housework hours, and divorce using the PSID sample of married adults between the ages of 18 to 64, for years 1979 to 2013.[17]

[17] BKP have a timespan of 1969 to 2009, yet I still closely replicate their estimates.

Table A4: Testing whether couples with equal incomes affect BKP's longitudinal results

VARIABLES	LFP_t			HW_{t-1}			$Divorce_t$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SPEC.	Orig.	Redef.	Exclude	Orig.	Redef.	Exclude	Orig.	Redef.	Exclude
$WifeEarnMore_{t-1}$	-0.021	-0.020	-0.021	0.129	0.086	0.092	0.003	0.002	0.003
	(0.004)	(0.004)	(0.004)	(0.219)	(0.213)	(0.224)	(0.002)	(0.002)	(0.002)
$WifeEarnMore_{t-2}$							0.004	0.004	0.004
							(0.002)	(0.002)	(0.002)
Observations	89,148	89,148	87,902	62,404	62,404	61,528	65,052	65,058	63,382
R-squared	0.553	0.553	0.553	0.196	0.196	0.197	0.001	0.001	0.001
Number of FE	11,140	11,140	11,132	9,594	9,594	9,569	9,644	9,645	9,563

The sample includes married couples between the ages of 18 and 64 in waves 1979 to 2013 from the PSID. The columns marked "Orig." replicate regressions from Col. 1 of Table 7 in BKP for wives' labor force participation, the intra-couple gap in housework, and divorce. Each regression includes controls for a quadratic in the wife and husband's ages, the log of the wife and husband's income, the log of their total income, dummy variables for whether only the wife has income or only the husband has income, and year and state fixed effects. The main set of income variables are with respect to time $t-1$, while the divorce regressions additionally include all income-related variables with one lag for time $t-2$. Columns marked "Redef." redefine $WifeEarnMore_s$ to be equal to one if the couples have exactly equal income, while columns marked "Exclude" exclude couples with equal income from the sample.

Table A4 shows that these results are entirely unaffected by redefining the key variable for whether the wife earns more or excluding couples with equal incomes. Across specifications, the coefficient on WifeEarnMore$_{t-1}$ is for labor force participation approximately -0.02 (and highly significant; compare to BKP's estimate of -0.019), the coefficient for housework is not significantly different from zero (likewise for BKP), and for divorce coefficients for both WifeEarnMore$_{t-1}$ and WifeEarnMore$_{t-2}$ are positive at 0.002 and 0.004 respectively, although only the latter is significant (BKP have point estimates of 0.003 and 0.005 respectively, and likewise only the latter is significant). This exercise suggests that although this group of couples may have been important for the discontinuity result in BKP, they are not the driving factor in BKP's results on within-marriage outcomes.

D Before and after female breadwinning

This appendix describes how results in Fig. 7 are estimated, and present the associated tables.

BKP use the following estimating equation for the longitudinal analysis:

$$y_{i,t} = \beta_1(\text{WifeMore}_{i,t-1}) + X'_{j,t-1}\beta_k + \delta_t + \mu_i + \epsilon_{i,t}$$

The key variable of interest is WifeMore, an indicator for whether the wife earned more in time $t-1$, and in BKP's analysis the dependent variables of interest include labor force participation, the intra-couple gap in weekly housework hours (wives' hours minus husbands' hours) and divorce.[18]

I replicate the baseline specification from BKP, which includes the husband and the wife's respective log earnings, the log of their total earnings, indicators for whether only the husband or the wife is working, quadratics in the husband and wife's respective ages, year and state fixed effects, and couple fixed effects. In BKP, the coefficient on WifeMore for the gap in housework hours is not significantly different from zero in the baseline specification, so I add the controls used in the first specification for which it is highly significant, which includes the wife's share of relative income and a cubic in husbands' and wives' log earnings.

In addition to the replication, I estimate the same specifications with dependent variables from up to two calendar years before and after the reference period for female labor force participation in BKP, time t.[19] Female breadwinning is measured in time t, but with reference to time $t-1$. Finally, to highlight the role of husbands' work, I also consider husbands' annual hours as a dependent variable in the same way.

The sample for the PSID is constructed based on the same criteria as in BKP, including the restriction that only couples where at least one of the members had positive income in the year prior to the reference period, time t, are included. BKP include couples through the 2009 wave, while I include couples through 2013.

Since the PSID has been conducted biennially since 1997, we only observe variables that are collected in time $t-2$, t, and $t+2$ for this later period. In the main specification I use all years available for each relative time period, but I also estimate the results for the pre-1997 period and post-1997 period in each case as well. Finally, I restrict observations to those who are observed from time $t-2$ to time $t+2$,[20] and exclude couples who have skipped interview years within the period that they are observed in

[18] In divorce, one additional lag is included for all income-related variables with the motivation that divorce takes time.

[19] Naturally, this analysis is not feasible with divorce.

[20] I conduct an additional check removing this restriction, but the results are largely unchanged.

the survey to ensure that relative time periods are correct with respect to calendar time.

D.0.1 Results

The results are displayed in Fig. 7.[21] Although the sample size is substantially smaller than in BKP due to the restriction to observe each couple two years before and after a given observation, the main specification using all available years (in black) quite closely replicates BKP's results, with a point estimate of -0.025 for labor force participation in time t and 0.62 for housework in time $t-1$.[22]

The key feature of Panels (a) and (b) is that the coefficient for labor force participation is negative and statistically significant for periods $t-2$ to $t+1$, and it is positive (though not always significant at the 5% level) for housework over the same periods.[23] In other words, female breadwinning in time $t-1$ is associated with lower labor force participation not only after the occurrence of female breadwinning, but also during the same year that it occurs and the prior year.

Finally, Panel (c) provides additional evidence that husbands' reductions in hours are strongly associated with female breadwinning, with negative and highly significant coefficients for the periods $t-2$ to t, and the largest in magnitude being associated with year that female breadwinning is measured, $t-1$, with a coefficient of -103.

These patterns are hard to reconcile with the interpretation that women are decreasing their labor force participation due to an aversion to female breadwinning or increasing their housework hours to compensate for it. However, even if women do not decrease their labor force participation in response to female breadwinning, it does not rule out that couples might nevertheless dislike the switch to a female breadwinner, even temporarily, and it could negatively affect marital satisfaction and the risk of divorce, as several papers have found.

[21] Tabulated results may be found in Appendix **??**.

[22] These values compare to -0.019 in Col. 1 of Panel A and 1.048 in Col. 3 of Panel C, both in Table 7 of BKP. While the coefficient I estimate for housework is smaller, the 95% confidence interval includes 1.048.

[23] For both outcomes and all specifications, the coefficients are not statistically different from zero for time $t+2$.

Table A5: All couples in available years

(a) Wives' labor force participation

VARIABLES	(1) t-2	(2) t-1	(3) t	(4) t+1	(5) t+2
WifeEarnMore$_{t-1}$	-0.016	-0.029	-0.021	-0.021	0.011
	(0.006)	(0.006)	(0.004)	(0.007)	(0.006)
Observations	58,437	46,125	74,750	46,341	71,827
R-squared	0.354	0.605	0.554	0.285	0.005
Number of FE	8,779	6,740	9,721	6,770	9,713
Years	All	<'97	All	<'97	All

(b) Intra-couple gap in housework hours

VARIABLES	(1) t-2	(2) t-1	(3) t	(4) t+1	(5) t+2
WifeEarnMore$_{t-1}$	0.244	0.678	0.329	0.072	-0.115
	(0.273)	(0.341)	(0.232)	(0.339)	(0.290)
Observations	55,884	43,740	71,750	43,934	68,775
R-squared	0.069	0.201	0.191	0.096	0.042
Number of FE	8,719	6,719	9,712	6,749	9,702
Years	All	<'97	All	<'97	All

(c) Husbands' annual hours of work

VARIABLES	(1) t-2	(2) t-1	(3) t	(4) t+1	(5) t+2
WifeEarnMore$_{t-1}$	-43.580	-101.528	-22.544	-19.998	-12.056
	(11.351)	(7.840)	(11.459)	(10.001)	(14.211)
Observations	46,125	74,750	46,341	58,816	50,712
R-squared	0.104	0.255	0.060	0.005	0.001
Number of FE	6,740	9,721	6,770	8,805	6,785
Years	<'97	All	<'97	All	<'97

This table uses data on married couples between the ages of 18 and 64 from the PSID, observed between 1979-2013. Controls include logs in the wife and husband's earnings and the sum of their total earnings, dummy variables for whether only the wife or only the husband had income, quadratics in the wife and husband's age, and year and state fixed effects. All controls are measured in time t, although income variables are with reference to $t-1$. For the regressions on housework, relative income and cubics in the wife and husband's log earnings are also included.

Table A6: Restricted to couples observed between $t-2$ to $t+2$

(a) Wives' labor force participation

VARIABLES	(1) t-2	(2) t-1	(3) t	(4) t+1	(5) t+2
WifeEarnMore$_{t-1}$	-0.018	-0.031	-0.025	-0.020	0.010
	(0.006)	(0.007)	(0.005)	(0.008)	(0.008)
Observations	52,982	33,989	52,982	32,053	52,982
R-squared	0.345	0.616	0.570	0.189	0.004
Number of FE	8,105	5,071	8,105	4,732	8,105
Years	All	<'97	All	<'97	All

(b) Intra-couple gap in housework hours

VARIABLES	(1) t-2	(2) t-1	(3) t	(4) t+1	(5) t+2
WifeEarnMore$_{t-1}$	0.517	0.620	0.253	0.326	-0.179
	(0.290)	(0.419)	(0.272)	(0.422)	(0.340)
Observations	43,905	25,815	43,905	23,915	43,905
R-squared	0.066	0.161	0.199	0.059	0.022
Number of FE	7,417	4,477	7,417	4,141	7,417
Years	All	<'97	All	<'97	All

(c) Husbands' annual hours of work

VARIABLES	(1) t-2	(2) t-1	(3) t	(4) t+1	(5) t+2
WifeEarnMore$_{t-1}$	-55.427	-103.344	-27.779	-22.414	-23.001
	(12.823)	(9.362)	(13.344)	(10.506)	(17.865)
Observations	35,796	53,051	33,942	53,051	33,942
R-squared	0.097	0.222	0.073	0.004	0.001
Number of FE	5,181	8,097	4,964	8,097	4,964
Years	<'97	All	<'97	All	<'97

This table uses data on married couples between the ages of 18 and 64 from the PSID, observed between 1979-2013 (corresponding to Fig. 7). Controls include logs in the wife and husband's earnings and the sum of their total earnings, dummy variables for whether only the wife or only the husband had income, quadratics in the wife and husband's age, and year and state fixed effects. All controls are measured in time t, although income variables are with reference to $t-1$. For the regressions on housework, relative income and cubics in the wife and husband's log earnings are also included.