I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11693

Is There a Male Breadwinner Norm? The Hazards of Inferring Preferences from Marriage Market Outcomes

Ariel J. Binder
David Lam

JULY 2018

DISCUSSION PAPER SERIES

IZA DP No. 11693

Is There a Male Breadwinner Norm? The Hazards of Inferring Preferences from Marriage Market Outcomes

Ariel J. Binder
University of Michigan

David Lam
University of Michigan, NBER and IZA

JULY 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Is There a Male Breadwinner Norm? The Hazards of Inferring Preferences from Marriage Market Outcomes*

Spousal characteristics such as age, height, and earnings are often used in social science research to infer social preferences. For example, a "male taller" norm has been inferred from the fact that fewer wives are taller than their husbands than would occur with random matching. The large proportion of husbands out-earning their wives has similarly been cited as evidence for a "male breadwinner" norm. This paper argues that it is difficult and potentially misleading to infer social preferences about an attribute from observed marital sorting on that attribute. We show that positive assortative matching on an attribute is consistent with a wide variety of underlying preferences, including "female taller" or "female breadwinner" norms. Given prevailing gender gaps in height and earnings, positive sorting implies it will be rare for women to be taller than, or earn more than, their husbands – even if there is no underlying preference for shorter or lower-earning wives. In an empirical application, we show that simulations which sort couples positively on permanent earnings can largely replicate the observed distribution of spousal earnings differences in US Census data. Further, we show that an apparent sharp drop in the distribution function at the point where the wife begins to out-earn the husband results from a mass of couples earning identical incomes, a mass which we argue is not evidence of a norm for higher-earning husbands.

JEL Classification: D10, J12, J16

Keywords: economics of marriage, gender gap, economics of the family

Corresponding author:
David Lam
Institute for Social Research
University of Michigan
426 Thompson Street
Ann Arbor, MI 48104
USA
E-mail: davidl@umich.edu

* This project was generously supported by the services and facilities of the Population Studies Center at the University of Michigan (R24 HD041028). Binder gratefully acknowledges research support from an NICHD training grant to the Population Studies Center at the University of Michigan (T32 HD007339). Martha Bailey provided valuable input in the development of this paper. We also acknowledge helpful discussions and feedback from Theodore Bergstrom, Deborah Cobb-Clark, Daniel Lichter, Shelly Lundberg, Robert Pollak, Robert Willis and seminar participants at UC Santa Barbara, the 1st IZA Workshop in Gender and Family Economics, the 2017 Population Association of American Annual Meetings, the Society of the Economics of the Household (SEHO) 1st Annual Meeting, and the University of Michigan. Errors and opinions are our own.

I. Introduction

Do men prefer to be taller than their wives? Do women prefer to earn less than their husbands? Social scientists often use the attributes of spouses to infer individual preferences and social norms. For example, a number of studies seek to quantify the prevalence of a "male-taller" norm in marriage (Gillis and Avis 1980, Stulp et al. 2013) and the extent to which this norm affects inter-ethnic marriage patterns (Belot and Fidrmuc 2010). Other studies look at differences in earnings between spouses (Winkler 1998, Brennan, Barnett, and Gareis 2001, Raley, Mattingly, and Bianchi 2002), inferring from these patterns social preferences about whether husbands should earn more than their wives as well as implications of these preferences for time allocation, the division of resources, and marital stability (Schwartz and Gonalons-Pons 2016).

This paper demonstrates that the standard Beckerian marriage model generates matching patterns that suggest social norms of husbands being taller and earning more than their wives, *even when individuals prefer the reverse*. Taking the example of height, we show that a broad class of penalty functions for deviation from the social norm generates positive assortative matching on height in equilibrium, regardless of what the norm dictates about the ideal spousal height difference (including the absence of any norm). Positive assortative matching together with the prevailing gender gap in height results in an equilibrium in which few husbands are shorter than their wives—even if husbands strictly prefer to be shorter than their wives. While this result is based on features of the Beckerian marriage model, we argue that the main message of non-identifiability of preferences holds in more general conceptualizations of the marriage market.

We apply this theoretical result to the context of earnings differences between spouses, the focus of a large literature and a topic which has garnered significant attention in a prominent recent paper by Bertrand, Kamenica, and Pan (2015). With data drawn from the 2000 United States

Census, we use the Beckerian framework to match couples based on earnings. We consider two models: one in which observed earnings are taken as given, and a second in which an endogenous labor supply decision is made after marriage (to account for the fact that earned income is not an exogenous attribute). In both cases we match men and women according to the model and then simulate the resulting distribution, across couples, of the share of the couple's total earned income that was earned by the wife. Importantly, we only assume positive assortative matching—there is no explicit preference for wives to earn less than husbands. Even without imposing such a norm, our simulations succeed in reproducing the highly skewed distribution of spousal earnings differences observed in the data. That is, there are far fewer wives outearning than their husbands than vice versa, even though positive assortative matching is consistent with a wide class of preferences—including a preference for wives to outearn husbands. These simulation exercises illustrate that a literal interpretation of marital matching patterns may produce incorrect inferences about underlying preferences.

The empirical strategy pursued by Bertrand, Kamenica, and Pan (2015) (hereafter BKP) represents a compelling addition to the literature. Unlike other studies, BKP did not rely upon broad features, such as the skewness, of the distribution of attributes in marriage. Instead, they tested whether the distribution of the wife's share of total spousal earnings was continuous across the 50 percent threshold—the point at which the wife goes from earning just less to just more than her husband. They found a discontinuous dropoff in probability mass across this threshold, suggesting that couples manipulate their earnings on the margin to avoid a situation in which the wife out-earns her husband. Without assuming an explicit social norm that wives should not out-earn their husbands, it is difficult to replicate this discontinuity in our simulated matching models, implying an important role for such a norm in marital matching and earnings outcomes within marriage.

Further investigation into this discontinuity result, however, suggests that it is fragile. One issue with the spousal earnings data investigated by BKP is the presence of a mass of couples earning exactly identical incomes. This generates a mass point in the distribution of the wife's share of total earned income at 50 percent. Recognizing this feature of the data, BKP tested for a discontinuity just to the *right* of 50 percent, consistent with testing for a social norm that the wife should not strictly out-earn her husband. Using the same data source,[1] we first replicate BKP's result of a sharp dropoff in probability mass across this threshold. However, when we also test for a discontinuity just to the *left* of 50 percent, we find evidence of a sharp *gain* in probability mass. This sharp gain in mass as one moves from left of 50 percent (where the wife earns less than the husband) to 50 percent (equality) could be interpreted as evidence for a social norm that the wife should earn at least as much as her husband. Thus the data appear consistent with two nearly opposite social norms.

Even though the point mass of equal-earning couples amounts to only about one quarter of one percent of all couples, we show that its presence is responsible for these seemingly inconsistent results. The potential for the point mass to influence the results is evident from a histogram which cuts the data into very small bins (similar in size to the bins used to perform to the discontinuity test). The histogram displays a large spike in probability mass right at 50 percent, but otherwise appears fairly smooth. Accordingly, we remove the equal-earning couples from the sample and repeat the discontinuity tests. Omitting these couples eliminates the estimated discontinuities.

[1] The data are administrative earnings data from the Social Security Administration. These data are linked to a household survey (the Survey of Income and Program Participation), which permits the researcher to observe earnings of matched couples. Section IV provides further discussion.

Moreover, the resulting insignificant discontinuity estimates are similar in magnitude to those generated from our simulations based on positive assortative matching. One possible reconciliation of the evidence is a conclusion that the mass of equal-earning couples implies an *equal-earning norm*, at least for a segment of the population. However, we discuss how further evidence would be needed to endorse this conlusion.

Our investigation suggests considerable caution in inferring social norms from observed differences in spousal attributes. Given gender differences in attributes, the skewed distributions of spousal earnings and height differences can be generated by simple marriage models that result in assortative matching and make no explicit assumptions about underlying preferences. This conclusion is consistent with the recent work of Belot and Francesconi (2013), which argues that the pool of potential partners is more important than underlying preferences in the determination of who matches with whom. To be clear, our results do not imply that gender norms do not exist. Other evidence has been provided in the literature, including some additional analyses in BKP's paper. Our message is simply that the observed differences in spousal characteristics *per se* do not provide evidence regarding social norms related to those differences. Researchers should utilize other innovative methods to quantify the prevalence and consequences of such norms. Finally, while the discontinuity test performed by BKP represents such a method, its validity is limited by the realities of the data.

II. BECKER'S THEORY OF MARRIAGE AND A SIMPLE MODEL OF SORTING ON HEIGHT

Our theoretical discussion requires that we make predictions about how men and women are sorted in a marriage market. We build on Becker's (1973) economic theory of marriage, which provides well-known predictions about assortative matching on attributes. Consider a man *M* and a woman *F* who are considering marriage. We assume they marry if and if only if it makes both better off compared to alternatives. Denote the "output" of the marriage by Z_{mf}. For now, assume output can be divided $Z_{mf} = m_{mf} + f_{mf}$, where m_{ij} indicates what man i consumes when married to woman j. Although this may not be a minor assumption, since "household public goods" like children – or the earnings difference between spouses – cannot literally be divided in this way, Lam (1988) showed that the model can be applied to the case of household public goods under the assumption of transferable utility. Because output (or utility) can be divided up between husbands and wives, it is possible for men to make offers to potential wives (and women to make offers to potential husbands) of some division of output. This means that a man can in principle use "side payments" to attract a particular wife, and a woman can use side payments to attract a particular husband, making that person better off than he or she would have been with some other partner. This is a simple example of frictionless matching with transfers under transferable utility (Chiappori 2017).

Suppose we have a set of N women and N men, with marital output between woman i and man j denoted by Z_{ij}, and we consider all possible sortings of men and women. Drawing on results from other matching models in mathematics and economics, Becker showed that a competitive equilibrium in the marriage market will be the set of assignments which maximizes the sum of output across all marriages. The proof relies on a standard argument about the Pareto optimality of competitive markets. If an existing set of pairings does not maximize total output, then there must exist at least two couples who could switch partners and increase total output. Because output is transferable, it is possible to distribute the total output gains from the switch such that each individual is made better off. This will be illustrated below for a simple example of two couples sorting on height.

Becker applied this very general result to the case of sorting on some trait *A*, where we will consider woman *f* to have a trait value A_f and man *m* to have trait value A_m. We characterize marital output (which might be some measure of joint marital happiness) as a function of the values of *A* for each partner: $Z_{mf} = Z(A_m, A_f)$. Becker showed that the marriage market equilibrium will consist of positive assortative matching on *A* if

$$\frac{\partial Z(A_m, A_f)}{\partial A_m \partial A_f} > 0. \qquad (1)$$

There will be positive assortative matching if the cross-partial in (1) is positive, and negative assortative matching if the cross-partial is negative. A positive cross-partial derivative (equivalent to strict supermodularity, also known as the Spence-Mirrlees condition, as discussed in Chiappori 2017) can be interpreted as implying that the value of *A* for the husband and wife are complements, while a negative cross-partial implies they are substitutes. If, for example, having a better educated husband raises the impact of the wife's education on marital output, then we will tend to see positive assortative matching on education. We draw on this well-known result extensively below.

II.A. Illustrative Model of Sorting on Height

Some of the key theoretical points can be demonstrated with a very simple model of sorting on height in the marriage market. Denote female height by H_f and male height by H_m. Suppose there are two women: F_1 is 60" tall and F_2 is 66" tall. There are two men: M_1 is 66" tall and M_2 is 72" tall. There are two possible pairings: (F_1M_1, F_2M_2), which is positive assortative matching on height, and (F_1M_2, F_2M_1), which is negative assortative matching on height.

To find the marriage market equilibrium, we describe how the heights of couples affect marital utility. Assume that people get utility from their individual consumption and some bonus that comes from being married. The gains from marriage take the very simple form of some bonus *K* (representing, say, economies of scale in consumption or benefits of household public goods) that is offset by some penalty that depends on the height difference between spouses. *K* can be thought of in monetary or consumption units, representing in the simplest example the amount of money the couple saves by being married. The penalty associated with the height difference between couples can also be given a monetary interpretation, representing the amount of additional consumption that would be required to compensate for the disutility from a sub-optimal height difference between spouses.

Now, consider various alternative cases for the loss function associated with the height difference between spouses. For the first case, suppose that all men and women agree that the ideal marriage is one in which the husband is 6" taller than his wife. Couples in which this is not the case experience some loss of utility that increases at an increasing rate as the height difference between spouses increases. A simple example is a quadratic loss function:

$$Z(H_m, H_f) = K - (H_m - H_f - 6)^2 . \qquad (2)$$

If the husband is 6" taller than the wife then there is no loss of utility from marriage. If the husband is the same height as the wife then the loss is $(0\text{-}6)^2 = 36$. As a concrete and very literal example, this could mean that the couple would need an additional \$36 worth of consumption to make them as happy as a couple with the ideal height difference. If the husband is 12" taller than the wife then the penalty is $(12\text{-}6)^2 = 36$. With these payoff functions, we can consider the two possible pairings. If the taller man marries the taller woman and the shorter man marries the shorter woman, then each husband is 6" taller than his wife, generating a total marital utility of *2K* (zero penalty in either marriage). If partners switch, then one couple (same height) has a penalty of 36 and the other couple (taller man and shorter woman) also has a penalty of 36, for a total penalty of

72. Total marital utility is obviously highest with perfect rank-order sorting, and this is the competitive equilibrium we would expect to observe. If we started with the alternative sorting, everyone could be made better off by switching partners. If we observe the perfect rank-order sorting equilibrium and conclude that everyone prefers that husbands are taller than their wives, our inference would be correct.

Now consider a different payoff function in which the ideal couple is one in which the husband and wife have equal heights, with a penalty for height differences that is increasing in the difference:

$$Z(H_m, H_f) = K - (H_m - H_f)^2 \tag{3}$$

With perfect rank-order sorting the total penalty is now 36 + 36 = 72, since each couple is 6" from the ideal height difference. In the alternate sorting we can create one ideal couple of equal heights, generating a penalty of zero. But the other couple (the tall man and the short woman) has a height difference of 12", creating a penalty of 144. Perfect rank-order sorting produces higher total marital utility (lower total penalties). This follows from the convex penalty function, which penalizes large differences in height more than small differences.

The logic in terms of a competitive marriage market is as follows: Suppose we began with the sorting in which one couple has equal heights while the other couple has a 12" height difference. The individuals in the mismatched couple, F_1 and M_2 see that they would each be much happier if they could switch partners and have a 6" height difference instead of a 12" height difference. The question is whether F_1 would be able to induce M_1 to switch from F_2 to her. Her penalty would decline from 72 (half of 144) to 18 (half of 36) if she changed partners. The penalty for M_1 would increase from 0 to 18 (half of 36) if he switched partners. Clearly F_1 can more than compensate M_1 for changing, making him a side payment of at least 18, leaving herself better off after the switch. The exact same story can be told for M_2 inducing F_2 to switch to him. Every person will be better off after the re-sorting, so the positive assortative matching equilibrium is the one we should observe. The resulting sorting of spouses with the preferences in (3) is exactly the same as the sorting with the preferences in (2)–the sorting with positive assortative matching on height. In this second case we would be drawing an incorrect inference if we interpreted the equilibrium as resulting from a preference for men to be taller than their wives.

Taking this case even further, consider a payoff function in which the ideal couple is one in which the wife is 6” taller than her husband, with, once again, a penalty for deviations from the ideal that is increasing in the difference:

$$Z(H_m, H_f) = K - (H_f - H_m - 6)^2 \tag{4}$$

With perfect rank-order sorting the total penalty is 144 + 144 = 288, since each couple is 12” from the ideal height difference. In the alternate sorting the total penalty is 36 + 324 = 360. Once again it is positive assortative matching that produces the maximum total payoff across all marriages. If we started with negative assortative matching, a process of renegotiation analogous to the one just described should lead to a re-sorting. We therefore expect that positive sorting will be observed as the equilibrium outcome. Thus the underlying preferences (wife taller) are opposite to what is observed in equilibrium (husband taller). The reason this occurs is that the convex payoff function pushes the equilibrium toward a sorting that has small average differences between spouses. It is better to have everyone slightly off from the ideal rather than have some couples that are close to the ideal and other couples that are very far from the ideal.

The fact that three different sets of preferences produce identical equilibrium sortings is an example of a general indeterminacy of observed matches. This point is made in a recent review article by Chiappori and Salanié (2016), which considers matching models more generally, and also argues for the difficulty of inferring underlying preferences over observed attributes from marital outcomes, due to the dependence of the observed equilibrium on unobserved tastes and heterogeneity. Chiappori (2017) emphasizes that while observing that marriage is assortative can be inferred to imply that that marital surplus is supermodular in a given trait, it is impossible to determine which of the large set of supermodular functions generated the observed matches. To infer more about the underlying preferences we need additional information, such as information on transfers between partners.

II.B. A General Model of Marriage Matching on Characteristics

The conclusions reached in the examples discussed above generalize to cases with large numbers of women and men covering a large range of heights. The strong tendency for positive sorting on height to prevail in equilibrium stems from the supermodularity condition in (1). To see the intuition for this, consider a case in which the husband is shorter than the wife and there is a preference for equality. Increasing his height reduces the height gap and thus increases the total payoff from marriage. The impact of reducing the gap is larger when the initial gap is larger (from the convexity of the penalty function), so the positive impact of increasing his height is increasing in the height of the wife. Conversely, the impact of the husband's height is negative when he is taller than his wife, but this effect will be smaller when the initial gap is smaller. So, the negative impact of the husband's height becomes less negative as the wife's height increases, once again implying a positive cross-partial. This implies that there will be positive assortative matching on height.

We now more formally demonstrate how payoff functions with this convex penalty structure give rise to positive sorting on height in equilibrium.

PROPOSITION 1**.** Consider a population with N men and N women, with everyone getting married, and assume the following marital payoff function:

$$Z(H_m, H_f) = K - f(g), \tag{5}$$

where the male-female height gap $g = H_m - H_f$. If f is (strictly) convex in g, then (strict) positive sorting on height is the unique marriage market equilibrium.

Proof. See Appendix.

This result reveals that positive sorting on height in equilibrium can be consistent with both husband-taller and wife-taller norms, as well without any explicit or straightforward norms at all. The next result illustrates that if there is a substantial gender gap in the attribute distributions, as is the case for height and for income, the equilibrium implied by positive sorting is highly skewed in nature.

PROPOSITION 2**.** Consider a population with N men and N women, with everyone getting married, and assume the marital output function is again given by equation (5). If the male height distribution exhibits first order stochastic dominance (FOSD) over the female distribution,[2] then there exists a marriage market equilibrium in which no wives are taller than their husbands. Moreover, if the penalty function exhibits strict convexity, the equilibrium is unique.

Proof. See Appendix.

II.C. Extensions of the Model

Taken together, Propositions 1 and 2 indicate that a wide variety of social norms regarding spousal height differences is consistent with a skewed distribution of spousal height differences in

---

[2] That is, at any common rank in the distributions, the male attribute is larger than the female attribute. Although this may sound like a strong assumption, it is quite realistic in the cases of both height and income. For example, FOSD holds for the income distributions of husbands and wives in the 2000 US Census, the data used in our empirical investigation of income differences between spouses (see section III).

marriage. These results predict an equilibrium in which *no* wives are taller than their husbands (or, analogously, no wives earn more than their husbands). This stark result is clearly not consistent with reality. Several factors are presumably at work in actual marriage markets—couples do not match on a single trait, there are search frictions, information about payoffs is imperfect, there is not perfectly transferable utility, etc. We consider some of these issues below. The main message, which is robust to these issues, is that the link between underlying preferences about an attribute and equilibrium sorting on that attribute is not straightforward. This makes it difficult to infer preferences from the observed equilibrium sorting.

Sorting on Multiple Attributes

It is important to consider matching on multiple attributes in our current context: if the economic gains to marriage depend on attributes other than height, then the distribution of height gaps in marriage will clearly depend on how these attributes are correlated with height in the population.

As a simple example, suppose there is an additional attribute *X* which enters the marital output function, such that the economic gains from marriage unrelated to the height gap are no longer constant:

$$Z_{mf} = Z\left(X_m, H_m, X_f, H_f\right) = K\left(X_m, X_f\right) - f(H_m - H_f). \quad (7)$$

We make the following additional assumptions: $K_1 > 0, K_2 > 0, K_{12} > 0$, and $corr(X, H) > 0$. Thus *K* satisfies Becker's sufficient condition (1) for positive assortative matching on *X* in equilibrium, while *f* leads to positive assortative matching on *H* by Proposition 1. It is impossible to know without further assumptions whether the prevailing equilibrium will consist of positive sorting on *X*, on *H*, or on some function of *X* and *H*. However, given that *X* and *H* are positively correlated in the population, some degree of positive sorting on *H* must exist in equilibrium.[3] Therefore, given a significant gender gap in *H*, this model still predicts that an equilibrium in which few wives are taller than their husbands is consistent with a variety of social preferences over the spousal height gap. There could also be no social preferences regarding height whatsoever—*f* could be constant—yet the positive correlation between *X* and *H* would still lead to an equilibrium making it look as if a male-taller norm exists.

The predictions of this model are especially relevant to the case of earnings, which we will investigate empirically in sections III and IV. Lam (1988) has shown that there will tend to be positive assortative matching on earnings whenever the economic gains from marriage result from household public goods, such as children. Thus both economic incentives and social norms favor an equilibrium with positive assortative matching on earnings.

Non-Transferability of the Marital Surplus

The Becker (1973) model assumes that the gains from marriage are fully transferable between spouses via monetary payments. In this setup, both the allocation of marriages and the transfers are determined in equilibrium as prospective partners make binding agreements in the marriage market.[4] If the division of the marital surplus cannot be negotiated in the marriage market—that is, bargaining over the marital surplus occurs after marriage—the market clears on the basis of what prospective partners expect to obtain from bargaining within marriage.[5] Pollak (forthcoming) notes that such a setup is consistent with using the Gale-Shapley framework (Gale and Shapley 1962) rather than a Beckerian framework to analyze the marital equilibrium.

---

[3] If the correlation between *X* and *H* is perfect, then the equilibrium will consist of positive assortative matching on both *X* and *H*.

[4] In a finite market, the equilibrium vector of transfers is not unique. As the number of individuals in the market increases to infinity, uniqueness is achieved.

[5] For a survey of the implications of household bargaining models for distribution of resources within marriage, see Lundberg and Pollak (1996).

When the marital surplus is transferable, a system of transfers exists to push the marital equilibrium toward positive assortative matching on height. The social norm may, for example, dictate equality of heights, but transfer payments lead to an equilibrium in which all couples violate the norm by small amounts. In the case where the marital surplus is non-transferable, this result may fall apart: a husband in a perfect marriage (where he is the same height as his wife) cannot be enticed away by a wife in a miserable marriage (where she is far taller than her husband).

When the marital surplus is fully non-transferable, equilibrium marital outcomes have the potential to offer some identifying information about the underlying social norm. For example, if the ideal is for husbands to be two inches taller than wives, we might expect to see a point mass at two inches in the height gap distribution. However, the validity of this inference still hinges on how preferences interact with the prevailing height distributions. For example, in a marriage market with 10 women and 10 men, we show that two very different preference structures[6] are both consistent with an equilibrium in which the 2 shortest men are shorter than their wives, and the 8 tallest men are taller than their wives. Moreover, if attributes other than height enter the marriage calculus, it is not clear that equilibrium height differences have any relation to underlying height preferences.

The Possibility of Remaining Single

Becker's original model assumes that everyone in the marriage market gets married, though marriage rates in the United States have declined considerably since its inception (Lundberg, Pollak, and Stearns 2016). An alternative setup specifies a value of being single and requires that all marriages which form in equilibrium provide each spouse with some surplus relative to remaining single.

---

[6] See Appendix.

A marital output function given by (5) suggests that the gains from marriage, and hence marriage rates, would be lowest for men and women at the extremes of their respective height distributions. For example, if men are taller than women on average and a norm exists for husbands to be 4 inches taller than wives, we might expect the shortest men and tallest women to remain single. This suggests that by comparing the heights of those who remain single to those who marry, as well as carefully considering the gender gap in height, one might be to learn something about height preferences.[7] On the other hand, if one restricts the sample to married couples and analyzes only the distribution of height or income differences between spouses—as previous research has done—our concerns about valid identification of preferences remain.

II.D. Application to Empirical Analysis of Spousal Height Differences

A recent empirical analysis of height differences between spouses helps illustrate our point about the difficulty of inferring preferences from equilibrium matches. Stulp et al. (2013) analyze the distribution of height differences among couples in the United Kingdom's Millennium Cohort Study. They compare the actual distribution of height differences to hypothetical distributions based on random matching, drawing several inferences based on this comparison. Table I presents their data, divided into bins of 5 cm (2 inch) height differences. A key observation is that the actual distribution has fewer women who are taller than their husbands than would occur through random matching. The authors argue that this is consistent with a "male-taller" norm. They also interpret the data as supporting a "male-not-too-tall" norm, since there are fewer men who are more than 25 cm taller than their wives than would occur through random matching. In other words, they

[7] Once again, though, the occurrence of sorting on multiple attributes or for multiple reasons may compromise such inferences. In the case of earnings, for example, observing the lowest-earning men remaining single could indicate a norm for husbands to out-earn their wives, but it could also reflect economic incentives based on gains from marriage.

interpret the actual distribution as implying a social norm for husbands to be taller – but not too much taller – than their wives.

It is easy to see that the data are consistent with other social norms as well. These include what might be called a "wife-not-too-short" norm or a "heights-not-too-different" norm. In fact, a better way to describe the norm implied by Table I might be a norm to keep the difference in heights between husbands and wives close to the overall average difference in heights between men and women in the population. The three bins closest to the actual average height difference of 14.1 cm (5.5 inches) are the bins that occur more frequently in the actual distribution than in the random matching distribution. The bins with the height differences farthest from 14.1 cm are the bins that occur with the lowest frequency relative to random matching. Notice that this is exactly what will happen if there is a tendency for positive assortative matching on height, as this pushes the equilibrium toward an outcome in which the height gap is uniform across all marriages. By Proposition 1, positive assortative matching is consistent with a variety of preferences, including a preference for wives to be shorter than their husbands. Hence it seems possible that a variety of underlying preferences could produce the distribution analyzed by Stulp et al. (2013).

III. The Empirical Relevance of the Gender Gap in Earnings for the Distribution of Spousal Earnings Differences

We now apply the above insights to an empirical investigation of earnings differences between spouses, where, like height, a persistent gender gap also exists.[8] A tendency for positive sorting combined with this gender gap would lead to a skewed marriage market equilibrium in which most husbands out-earn their wives—even if there is no social norm dictating this outcome.

[8] Gender differences in wage earnings in the U.S. is well known and attributed to a variety of factors, including differential human capital and career investments, labor-market discrimination, and others (Bertrand 2010, Bailey and DiPrete 2016).

An important social question, recently investigated by Bertrand, Kamenica and Pan (2015), is whether the observed gender gap in labor market outcomes (e.g. occupation, employment, and earnings) is influenced by gender norms operating in the home, independent of the labor market. Our analysis indicates that this question cannot readily be answered by analyzing spousal earnings differences.

We simulate marriage market equilibria using observed earnings in U.S. Census data and simple matching processes. We start by matching men and women randomly. Next, we match women and men assuming positive assortative matching on observed earnings perturbed with noise (to approximate characteristics other than observed earnings influencing the marriage market). Finally, recognizing that earnings is not an exogenous attribute but is affected via a labor supply decision, we assume positive sorting on unobserved potential earnings and endogenize labor supply choices made after marriage. In the case of random sorting, spousal earnings differences are driven entirely by gender differences in earnings distributions. In the last two cases, results are driven by the gender gap in (potential) earnings combined with the assumption about assortative matching. Following BKP, we summarize spousal earnings differences by plotting the distribution of *the share of the couple's total earnings that was earned by the wife*. Thus 0.01 indicates that a wife earned 1 percent of the couple's total earnings, and 1.0 indicates that she earned all of it. 0.50 represents a couple in which wife and husband earned equal amounts.

III.A. Empirical Distributions of Spousal Earnings Differences

We begin with a sample of men and women drawn from the 5 percent sample of the 2000 U.S. Census (Ruggles et al. 2015). Following BKP, we restrict the sample to couples ages 18-65 and process earned income variables following the procedure outlined in the paper's main text and appendix. We keep only couples in which both spouses report positive earnings. Figure I displays two 20-bin histograms of the distribution of the share of total earnings earned by the wife: the one published in BKP and our replication. As in BKP, we apply a local linear smoother to the histogram bins, allowing for a break in the smoothed distribution at 0.50. The two distributions are almost identical, and both display a substantial reduction in probability mass to the right of 0.50.

For our simulation exercises, we further restrict the sample to relatively young couples (aged 18-40) without children. Our final sample consists of 109,569 dual-earning couples. FIGURE II plots the sample distribution of the wife's share of total earnings in the final sample. The main difference between this distribution and that in Figure I is that there is less mass below 0.25, which likely reflects the impact of specialization after childbearing.[9] Our simple simulations are not set up to handle the dynamic considerations of fertility and its effect on the wife's labor supply and earning potential. Nonetheless, imposing this sample restriction does not change the fact that most of the distribution lies to the left of 50 percent (where the wife earns less than the husband), and the probability mass drops sharply as one moves to the right of 50 percent. These are the stylized facts we will attempt to replicate in the following exercises.

III.B. Simulated Distributions

Random Matching of Couples

In our first simulation, we randomly match men and women in our sample into couples. Figure III displays a smoothed distribution of the wife's share of total earnings based on random matching, again allowing for a break at 0.50, overlaid on the observed distribution. The distribution

[9] This additional restriction is motivated by the well-known fact that women disproportionately reduce their working hours or exit the labor force to raise young children and later re-enter the workforce with lower earnings potential (Mincer and Ofek 1982, Hotchkiss and Pitts 2007, Attanasio, Low, and Sanchez-Marcos 2008, Bertrand, Goldin, and Katz 2010). We abstract from this endogenous specialization decision after childbearing. BKP's Appendix Figures A.1 and A.2 show similar effects of children and marital tenure on the observed distribution of the wife's share of total earnings.

generated by random matching is, perhaps surprisingly, not too dissimilar from the observed distribution—it contains a mode around 0.42 and a drop-off in mass to the right of that point. Moreover, significantly fewer wives slightly out-earn their husbands than vice versa; the point of equal earnings (0.50) corresponds to the 70th percentile of the distribution. This benchmark exercise demonstrates that the prevailing male and female earnings distributions exert a strong influence on spousal earnings differences.

Notice that Figure III follows a similar pattern to the distribution of height differences shown in Table I. The bins in Figure III that occur more frequently in the actual distribution than in the distribution with random matching are those closest to 0.42, the average wife's share of total earnings implied by random matching. (Although Figure III is in shares rather than differences, the pattern would look similar if plotted in absolute or proportional income differences.) A key feature is that the actual distribution is pushed toward the mean earnings difference and away from extremes, exactly as in our simple theoretical examples above. Following Stulp et al. (2013), one might interpret this as implying a "husband richer, but not too much richer" norm. But as we now show, the patterns are consistent with any model that generates positive assortative matching on earnings.

Positive Assortative Matching on Potential Earnings

To implement this exercise we take male and female earnings as observed in our sample (denoted as Y_i^m for males and Y_i^f for females). We create couples by matching individuals not according to observed earnings rank, but rather the rank of observed earnings perturbed with noise. That is, for each individual *i* of gender *g* we assign $W_i^g = Y_i^g + u_i$, where *u* is normally distributed white noise, and pair up males and females according to their ranks of *W*. This is consistent with at least two interpretations. One interpretation is that couples are perfectly sorted based on permanent earning potential and the white noise represents transitory earnings shocks realized after marriage. A second is that men and women care about other characteristics as well as earnings, or that marital matching is imperfect, for example due to the presence of search frictions. Under the latter interpretation, equilibrium sorting on observed earnings plus noise is the reduced form of a more complicated matching process.

Figure IV displays a simulated distribution of the wife's share of total earnings based on this simple model, with the standard deviation of u set to 16,000, overlaid on the actual distribution. The simulated distribution is very similar to the actual distribution: it exhibits a sharp drop in mass across the 50 percent threshold and contains few couples in which the wife out-earns her husband. Thus, given the gender gap in earnings distributions, the observed distribution of spousal earnings differences is largely consistent with positive assortative matching on earnings. As the previous section indicates, this matching is consistent with a wide variety of underlying preferences. It could be based on a desire for equality in spousal earnings, a preference for wives to earn more than their husbands, or economic gains from marriage related to household public goods (i.e. with no explicit preference at all for equal or unequal spousal earnings).

We next test whether the simulated drop-off in probability mass across the 0.50 threshold is discontinuous, via a Monte Carlo version of the McCrary (2008) test. We simulate 500 distributions independently from the data-generating process and test for a discontinuity at 0.50 percent in each distribution. The average point estimate is a 2.6 percent drop in mass, and the average t statistic is around -1. Thus we cannot reject the null hypothesis that our simple model generates a distribution that is smooth across the 0.50 threshold, despite there being much fewer wives who earn between 50 and 55 percent of total earnings than wives who earn between 45 and 50 percent of total earnings.

Positive Assortative Matching on Potential Earnings with Endogenous Labor Supply

One shortcoming of the previous exercise is that it treats the observed distributions of men's and women's earnings as fixed attributes, determined outside of the household. This is unrealistic; observed earnings is the product of the hourly wage rate and total hours worked in the market. A voluminous literature on household labor supply argues that household incentives, such as specialization incentives, influence especially the wife's labor supply decision. More importantly, BKP argue that social norms themselves may influence how many hours a wife chooses to work in the market: if she is at risk of out-earning her husband in a full-time job, she may work fewer hours. In this exercise, we endogenize the wife's earnings via a simple labor supply model and explore the model's predictions about the distribution of spousal earnings differences.

We assume that, for a given male *m* and female *f*, the match output function is given by

$$Z_{mf} = Z(Y_m, Y_f, P) = \frac{C^{1-\gamma}}{1-\gamma} - \psi P, \text{ with } C = 0.61(Y_m + Y_f P), \qquad (8)$$

where C is consumption of a composite good with price normalized to 1, Y_m and Y_f denote each spouse's permanent income, P is the wife's labor supply decision (constrained to be in the unit interval), γ is the coefficient of constant relative risk aversion, and ψ is the disutility incurred by the household if the wife works.[10] This specification of household utility has been used in recent work investigating determinants of wives' labor supply (e.g., Attanasio, Low, and Sanchez-Marcos 2008). It assumes household consumption of earned income is a public good with congestion; the

[10] This parameter could capture specialization incentives or social norms. Notice that the disutility faced by the household is continuous in the wife's labor supply decision—it does not change discontinuously if the wife supplies enough labor to out-earn her husband.

0.61 is a McClements scale calibration capturing consumption economies of scale in marriage.[11] We assume the marital surplus is non-transferable, so positive sorting on permanent earnings occurs in marriage market equilibrium so long as each member's permanent earnings positively affects match output.[12] It is trivial to show that this holds here (regardless of the wife's eventual labor supply decision). Assuming that each individual's potential earnings in a given period is the sum of his or her permanent earnings and a transitory shock, positive sorting on potential earnings plus noise will arise in equilibrium.

After marriage, the wife takes household potential earnings as given and chooses $P \in [0,1]$ to maximize the above utility function. With an interior solution, the wife will choose

$$P^* = \frac{\frac{1}{0.61}\left(\frac{\psi}{0.61Y^f}\right)^{-\frac{1}{\gamma}} - Y^m}{Y^f}. \tag{9}$$

If P^* lies outside of the unit interval, the appropriate corner solution applies.

To use the model to draw valid conclusions about the distribution of spousal earnings difference in marriage market equilibrium, we must reasonably calibrate it. Outside of the calibration we impose $\gamma = 1.5$, a standard value estimated in the macro literature. We assume log-normally distributed potential earnings and allow the work disutility parameter, ψ, to be heterogeneous in the population and negatively correlated with Y_f.[13] The model in total contains 8 parameters, which we calibrate by targeting 8 moments in our observed data: the means and standard deviations of male and female log observed income, the observed mean gender earnings ratio conditional on earning positive income ($P^*>0$), the observed mean gender earnings ratio

[11] To illustrate, suppose P=1 and $Y_m=Y_f$. Then the couple enjoys a higher level of joint consumption in marriage than either member would as single.

[12] Starting from perfectly positive sorting, it is easy to show that no two individuals can become better off by dissolving their current matches and matching with each other. The inability of individuals to make transfer payments means we no longer need the cross-partial assumption on the match output function to generate positive sorting on the given trait in marriage market equilibrium.

[13] This accords with estimates in the literature (Eckstein and Lifshitz 2011).

conditional on full-time work (defined in the data as at least 1600 hours worked in the last calendar year; defined in the model as P^*>0.95), the female employment rate (defined in the data as the share of wives working positive hours in the last calendar year), and the female full-time employment rate. Importantly, we do not explicitly target any moment related to marital matching or spousal earnings differences, as doing so would threaten the external validity of our inferences.

Table II summarizes the calibration. Overall the model does a good job of replicating the targets in the data. With the calibrated model we simulate the distribution of the wife's share of total spousal earnings (Figure V).[14] The simulated distribution again matches the actual distribution very closely. Although the match is not perfect, only slightly too many wives outearn their husbands relative to what is observed in reality. Performing the same Monte Carlo version of the McCrary test as in the previous exercise we also estimate a small and statistically insignificant drop-off in mass at 0.50.

In summary, simple models of spousal matching—random matching, positive assortative matching on potential earnings, and positive sorting on potential earnings with an endogenous labor supply decision—do well in generating the small incidence of wives out-earning their husbands. The positive sorting models also closely reproduce the large drop-off in probability mass across the equal-earnings threshold (0.50). However, these models fail to generate a discontinuity at this threshold.

[14] The simulation uses a sample size of 120,000 men and 120,000 women. Since around 90 percent of wives choose to work, an initial sample of 120,000 returns around 108,000 dual-earning couples, which closely matches the sample size observed in the 2000 Census.

IV. Alternative Evidence for Social Preferences that a Wife Should not Out-Earn Her Husband

The theoretical and empirical evidence suggests that social scientists wishing to test the importance of social norms need to find strategies beyond evidence of the skewed distributions of spousal attributes. The challenge in doing so makes the discontinuity found by Bertrand, Kamenica, and Pan (2015) at the equal-earnings threshold a compelling addition to the literature on social norms. The logic behind BKP's discontinuity test runs as follows. Suppose we observe the distribution of the share of total spousal earnings that was earned by the wife in the neighborhood of 0.50 (equal earnings). Suppose we find that this distribution exhibits a sharp change in probability mass at the equal-earnings threshold—that is, there are far fewer wives barely out-earning their husbands than husbands barely out-earning their wives. Because standard models of the marriage market, involving agents optimizing continuous utility functions, should not generate discontinuous equilibrium distributions, this empirical finding should be interpreted as evidence of a social penalty which applies if and only if the wife out-earns the husband. That is, couples are willing to sacrifice some of the wife's potential earnings to avoid a situation in which the wife out-earns her husband. This finding suggests an important role for gender norms in marital matching and female labor market outcomes within marriage.

BKP estimated a discontinuous drop-off in probability mass across the equal-earnings threshold in a variety of Census samples. However, as they discuss, inference is complicated by the fact that earnings are not precisely measured in Census survey data. Mis-measurement occurs for several reasons. First, earnings are reported, rather than measured directly. (Moreover, earnings for both spouses are typically reported by one household member.) Second, earnings are imputed for individuals who do not answer earnings questions, and the earnings of high-earning individuals are top-coded at a common value. Third, reported earnings are rounded (often to the nearest thousand) to minimize disclosure risk. These issues create a large point mass of couples with exactly identical earnings. Even after employing several procedures to adjust the data, BKP still found that around 3 percent of dual-earning Census couples have identical earnings. (We corroborate this finding.) To get beyond these limitations of public-use Census data they also assembled a sample of earnings records from the Social Security Administration (SSA). These data have been linked to a household survey (the Survey of Income and Program Participation, or SIPP) which allows couples to be identified.[15] In this administrative data sample, the point mass of equal-earnings couples still exists but is much smaller: only around one quarter of one percent of all dual-earning couples earn identical incomes. Reassuringly, BKP obtained a similar discontinuity result in this sample.

Without the point mass, the straightforward way to implement BKP's procedure would be to test for a discontinuity in the distribution exactly at 0.50, and interpret the finding of a significant drop-off in the density function as evidence for a social norm that the husband should out-earn his wife. The presence of the point mass presents a challenge, which BKP acknowledge in footnote 7 of their paper. To circumvent this problem, they tested for a discontinuity just to the right of 0.50. One might interpret this test as equivalent to testing whether there is a social norm dictating that the husband should *strictly* out-earn his wife. Their finding of a significant drop-off in the density function to the right of 0.50, combined with the presence of the point mass of equal earners, might suggest that couples manipulate their earnings so that the wife earns *the same as or less than* her husband.

---

[15] The data come from a pre-linked and cleaned Census Bureau data product called the Gold Standard File (GSF). Users work with synthetic versions of the data remotely and then have Census run final programs internally on the actual GSF, subject the output to a disclosure review, and then release the output. More information can be found in Benedetto, Stinson, and Abowd (2013) and here: http://www.census.gov/programs-surveys/sipp/guidance/sipp-synthetic-beta-data-product.html.

This treatment of the data seems sensible a priori, but the existence of the mass point violates one of the assumptions required by the discontinuity test—namely, that the distribution is continuous everywhere except possibly at the supposed breakpoint (McCrary 2008). Like a non-parametric regression discontinuity design, the test involves local linear smoothing of a finely-binned histogram on either side of the supposed breakpoint, and asymptotic inference is based on the size of the bins shrinking to zero at the correct rate as the number of observations increases to infinity. In BKP's application of the test, for a small bin size, the bin immediately before the breakpoint will (by containing the point mass) be taller than the bin immediately after the breakpoint. This could exert undue influence on the discontinuity estimate, especially if a small bin size and bandwidth is used to perform the test.

IV.A. Gauging the Robustness of BKP's Discontinuity Test Results

To investigate the sensitivity of the discontinuity test to the presence of the point mass, we replicate BKP's SIPP-SSA data sample and analysis. BKP constructed a sample of earnings data for all dual-earning couples aged 18 to 65 observed in the first year they were in the SIPP panel. They considered SIPP panels 1990 through 2004. We construct a sample according to the same conditions but include the 1984 and 2008 SIPP panels as well, which are available in the most recent version of the SIPP-SSA data product. We obtain a sample of around 83,000 couples—about 9,500 more than in BKP's sample.[16] Despite using a slightly different sample, the resultant distribution of the wife's share of total spousal earnings is virtually identical to BKP's, as illustrated in Figure VI.

[16] BKP report a sample size of 73,654, although it is unclear whether this number refers to all couples in their sample or all dual-earning couples.

In our replicated sample, 0.21 percent of all dual-earning couples earn identical incomes, compared to 0.26 percent in BKP's sample. To see the impact of this mass point on the distribution, Figure VII zooms in on the portion of the distribution between 45 and 55 percent, displaying histograms with a very small bin size of 0.001 (about the size used in the discontinuity tests). The top histogram retains the mass point, while the bottom histogram removes it. The two histograms look very different: the top one exhibits a large spike right at 0.50, while the bottom one does not. Moreover, though the data are noisy for such a small bin size, the histogram on the right does not look particularly discontinuous at 0.50. These illustrations suggest that the point mass may exert an undue influence on the discontinuity estimates.

Using our sample we perform 3 different versions of the McCrary test for a discontinuity in the distribution at 50 percent, based on three different treatments of the point mass: keeping the point mass and testing for a discontinuity at .500001, keeping the point mass and testing for a discontinuity at .499999,[17] and deleting the point mass and testing for a discontinuity exactly at 0.50. For each version we use 4 different sets of tuning parameters. McCrary's test procedure involves an algorithm which automatically chooses a bin size for the histogram and a bandwidth within which to apply the local linear smoother to the histogram. McCrary (2008) recommends using a smaller bandwidth than the automatically-selected one (around half the size) to conduct robust asymptotic inference. We consider the automatically selected bandwidth, which in this case is around .084; and then bandwidths of .045, .023, and .011. The last bandwidth may be too narrow for optimal statistical inference, but using successively smaller bandwidths allows us to gauge the sensitivity of the test to the presence of the point mass (which becomes increasingly dominant as the bandwidth shrinks).

[17] We also tested for discontinuities at .50001 and .49999, and .5000001 and .4999999. The results were very similar.

Table III reports the discontinuity estimates, which equal the estimated log increase in the height of the density function as one travels from just to the left of the supposed breakpoint to just to the right. A negative number thus indicates a sharp drop and a positive number indicates a sharp gain. Bolded estimates are statistically significant at the 5 percent level; italicized estimates are significant at the 1 percent level. Standard errors appear below estimates in parentheses.

The first version of the test replicates BKP's choice of retaining the point mass of couples and testing for a discontinuity just to the right of 50 percent (.500001). With the standard bandwidth and bin size, we estimate that the density function drops by a statistically significant 12.4 percent across the threshold. This is very similar to BKP's reported estimate of a 12.3 percent drop in their very similar sample (reported on p. 576). Observe that as the bandwidth shrinks, the estimate of the sharp drop rises in magnitude, such that with the smallest bandwidth we estimate a 57.5 percent drop—over 4 times as large as the first estimate. This suggests that the point estimates are sensitive to the existence of the point mass.

When we retain the point mass and test for a discontinuity just to the *left* of 50 percent, we find the exact opposite result: the density function jumps discontinuously *upward* at 50%. Once again, the estimate starts out reasonably small (6.4 percent) and becomes very large (45.1 percent) as the bandwidth shrinks. The finding of a sharp increase in the distribution at 50 percent suggests that couples manipulate earnings to avoid a situation in which the wife earns *strictly less* than her husband. Put another way, equal earnings are strongly preferred to having the wife earn slightly less than the husband (i.e. there is missing mass just to the left of 50 percent). This is nearly opposite to the social norm dictating that the wife *should not earn strictly more* than her husband, which is supported by the first version of the results.

The third column of results derives from deleting the point mass and testing for a discontinuity exactly at 50 percent. Two features stand out. First, while the estimates are negative, they are no longer statistically significant—moreover, the estimate based on the standard bandwidth matches closely the estimates generated by performing the test with the standard bandwidth on our simulated data (see section II). Second, the estimates do not rise appreciably in magnitude or statistical significance as the bandwidth shrinks, likely because the point mass is no longer present. Therefore, if we ignore the one quarter of one percent of couples earning identical incomes, the conclusion that the observed distribution of spousal earnings differences could be consistent with a variety of underlying social preferences (including no explicit social norm) is supported by the data. A related conclusion is that while BKP's discontinuity test is robust to the theoretical critique of the literature we levied in section II, it does not produce robust empirical results, given the point mass of couples earning identical incomes.

IV.B. A Further Inquiry into the Point Mass

Considering the above conclusions, it is worth exploring why the point mass exists in the first place, and what it means to remove it from the sample. For example, the existence of the point mass could indicate a social preference, in the population or a certain sub-population, for strict equality of spousal earnings. Further exploration of the 2000 Census data reveals the following facts about the couples who report identical earnings in comparison to the full sample.[18] [19] First,

[18] The Gold Standard File provides very little occupational information about the couples, which is why we use the Census for this exploration. It is important to keep in mind that the point mass of couples with identical earnings is over 10 times as large in the Census data, due to rounding of reported earnings as well as possible reporting biases. That is, many couples who report identical earnings in the Census data do not have identical administrative earnings records. However, it is reasonable to assume that couples who report identical earnings are (much) likelier than those who do not to have identical administrative records.

[19] All of these facts are based on the sample of couples in the 2000 Census 5 percent sample in which both husband and wife are age 18 to 65 with positive earnings.

couples who report identical earnings are almost six times more likely to both be self-employed than couples who report different earnings (13.0 percent versus 2.3 percent). Among couples in which husband and wife indicate being self-employed in the same occupation and industry (a likely indicator of running a family business), 34 percent report identical incomes. (These couples represent 0.18 percent of the full sample of couples.) Since income from a family business can be allocated in any way between husband and wife on tax returns, this suggests that one source of identical incomes is couples choosing to divide family business income equally for income tax purposes.[20]

In addition, there are couples in which the husband and wife do appear to earn identical salary incomes. Couples reporting that husband and wife both earn wages (i.e., are not self-employed) and report identical earnings, occupations, and industries (suggesting that they are likely to have identical jobs) constitute 0.34 percent of the sample. Elementary, middle school, and secondary teachers make up 18.9 percent of this group, by far the largest occupation. Taken together, the group of self-employed and salaried couples with identical incomes, occupations, and industries constitute 0.52 (=0.18+0.34) percent of all couples. Some of these are presumably "false positives," given the fact that Census data are self-reported and rounded. But this suggests that it is not difficult to account for the 0.2-0.3 percent of couples with identical earnings in the administrative data. Our interpretation of these cases (couples with family businesses reporting identical incomes and couples with identical earnings in occupations such as school teachers) is that they do not provide much information about a social norm related to husbands earning more than wives. They could constitute evidence for an equal-earning norm in a subset of the population,

[20] For couples filing jointly there will generally be no tax implications from the way family business income is allocated between husband and wife on Schedule C tax forms, though there might be implications for Social Security.

but they could also indicate frictions in the marriage market which lead a disproportionate share of equal-earning individuals to marry (for example, because they met through work). That is, there could be a small social penalty for the husband not out-earning his wife which is outweighed by the search cost of finding a more suitable partner.

Whatever the cause of the point mass of equal earners, we have shown that its presence compromises the validity and robustness of BKP's discontinuity test at the equal-earning threshold. It remains unclear whether observed distributions of spousal earnings differences offer identifying information about underlying social norms.

V. CONCLUSION

Our theoretical and empirical results demonstrate that it is potentially misleading to infer preferences about spousal attribute differences from their observed distribution in marriage market equilibrium. Marriage market outcomes are affected by preferences as well as the underlying distributions of attributes. If men are taller or higher-earning than women on average, preferences which lead to positive assortative matching will produce equilibria in which it is rare for women to be taller or higher-earning than their husbands. Even a preference for men to be shorter than their wives can lead to positive assortative matching and, consequently, an equilibrium in which men tend to be taller than their wives.

Our simulations produce distributions of spousal earnings shares which closely resemble the observed distribution using very simple models of assortative matching—without making any assumptions about preferences regarding husbands earning more than wives. The one feature we cannot reproduce with our simulations is the discontinuous drop-off in probability mass to the right of the equal-earning threshold, reported by Bertrand, Kamenica and Pan (2015). However, we show that this discontinuity is less informative than it first appears, since it is the result of a point mass of equal-earning couples. This mass causes a sharp drop to the right of 50 percent in the distribution of the wife's share of total earned income, which is consistent with a social norm that wives should not earn more than their husbands. But it also causes a sharp drop to the left of 50 percent, a result that is consistent with a social norm that husbands should not earn more than their wives. When we remove the point mass we do not see any evidence of a discontinuity at the equal-earnings threshold. In addition to satisfying the technical assumptions of the McCrary test, there is good reason to remove couples earning identical incomes, as these couples are predominantly joint business owners and couples in identical salaried occupations whose marriage and labor market outcomes may not reflect population preferences. Whether these individuals are retained or removed from the sample, their presence compromises the robustness of BKP's strategy of using a discontinuity test at the equal-earning threshold to infer the presence of a husband-breadwinner norm.

To be clear, our results *do not* imply that gender norms do not exist. The literature includes other types of analysis, with BKP providing other pieces of evidence in their paper that are not based on inferences drawn from the distribution of spousal earnings differences. These include analyses of marriage rates, divorce rates, labor force participation, work hours, and housework time as a function of the actual or predicted probability that the wife out-earns the husband. We are particularly intrigued by the new release of a study which finds that husbands tend to inflate, and wives deflate, reported earnings on surveys when the wife's "true" administrative earnings exceed her husband's (Murray-Close and Heggeness, 2018). It is outside the scope of this paper to analyze these other tests of the social norm hypothesis. Our argument is simply that observed differences in spousal attributes are not, in and of themselves, good evidence for social norms related to these attributes.

It is also interesting to consider whether social norms may themselves be driven by the underlying distributions of traits. In Stulp et al.'s (2013) analysis of height differences, there is a tendency for spouses to be pushed toward the actual mean difference in heights of 14 cm. We showed how this tendency can be explained as the result of positive assortative matching, with no need for a social norm related to height differences. But if there were a social norm for husbands to be 14 cm taller than their wives, it would seem surprising if some fundamental preferences coincidentally matched the actual difference in mean heights between men and women. If there is such a norm, it presumably was influenced by the actual differences in heights between men and women. A plausible explanation for such a norm could be that positive assortative matching produced distributions like those we observe, which in turn led individuals to perceive that there must be some normative reason for husbands to be taller than their wives.

This explanation is relevant to the case of earnings differences as well. Women's labor market opportunities in the United States have increased dramatically in the last 50 years, yet substantial gender career and earnings gaps remain, especially in marriage. It is possible that labor market change has outpaced social change, and slow-moving gender norms play a key role in generating these extant gender gaps in marriage. Inquiries into the existence and potential consequences of these norms are likely to continue to be an active area of research. We believe this research will be stronger and more convincing if researchers are sensitive to the challenges involved in drawing inferences about social norms from observed marriage market outcomes.

VI. Appendix

VI.A. Proofs of Propositions

Proof of Proposition 1. Compare strict positive sorting to an alternative allocation in which two couples switch partners. Specifically, consider the *i*th ranked man and woman and the *j*th ranked man and woman, $i<j$, with heights $H_{mi} > H_{mj}$ and $H_{fi}>H_{fj}$. There are two possible pairings of these two men and two women, with the following total payoffs from the two marriages:

$$\text{Payoff from pairing } A\text{: } Z(H_{mi}, H_{fi}) + Z(H_{mj}, H_{fj})$$

$$\text{Payoff from pairing } B\text{: } Z(H_{mi}, H_{fj}) + Z(H_{mj}, H_{fi}).$$

Pairing *A* represents strict positive assortative matching, while Pairing *B* represents the deviation.

Simple algebra shows that the total payoff from pairing *A* minus the total payoff from pairing *B* is:

$$\left(f(g_{ij}) + f(g_{ji})\right) - \left(f(g_{ii}) + f(g_{jj})\right) \tag{6}$$

where height gap $g_{ab} = H_{ma} - H_{fb}$ for all a and b. Recognize that the sum of the height gaps must be the same in both pairings, as the same 4 individuals are involved in each pairing. What we will now show is that of the 4 gaps, pairing *B* always contains the largest and the smallest.

By definition, man *i* is taller than man *j*, which yields $g_{ij} > g_{jj}$. Similarly, woman *i* is taller than woman *j* by definition, which yields $g_{ij} > g_{ii}$. Thus, by equality of sums, g_{ij} is always the largest gap while g_{ji} is the smallest. Strict convexity of *f* then implies that expression (6) is strictly positive, which is to say that joint marital output is strictly higher under Pairing *A* than under Pairing *B*. Therefore, whenever two men and two women are not positively sorted, a Paretoimproving system of transfers exists to restore perfect positive sorting, and thus perfect positive sorting is the unique equilibrium.

If f is merely convex, expression (6) is positive, but not necessarily strictly so. This implies that starting from perfect positive sorting, no profitable exchanges of partners can be made, and so positive sorting is an equilibrium. ■

Proof of Proposition 2. By Proposition 1, a marriage market equilibrium characterized by strict positive sorting on height exists, and is unique if the penalty function is strictly convex in the height gap. In a strict positive sorting equilibrium, the spouses of each couple have heights of identical rank in their respective distributions. Therefore, by the FOSD assumption, the husband is taller than the wife in each couple. ■

VI.B. Simple Marriage Market Example with Non-Transferable Utility.

This example illustrates the point, introduced on page 11, that even in the case of non-transferable utility, multiple preference structures can be consistent with the same equilibrium sorting of couples. Consider a marriage market with 10 men and 10 women. Male heights are distributed uniformly at 1-inch intervals from 66 inches to 75 inches. Female heights are distributed uniformly from 60 to 69 inches. Assume the same payoff structure as in equation (5): the gains to marriage are some constant, with a penalty for deviating from the ideal height gap that rises convexly in the deviation.

First, suppose that the social norm is for men to be 8 inches taller than their wives. It is easy to show that the prevailing marriage market equilibrium is one in which the 8 tallest men match with the 8 shortest women, as 8 "perfect" matches can be formed with this pairing, leaving the remaining 2 shortest men to pair with the two tallest women. Now, suppose that the social norrm is for women to be 2 inches taller than their husbands. In this case, 2 perfect matches can be created by matching the 2 shortest men with the 2 tallest women (66 inch man with the 68 inch woman; 67 inch man with the 69 inch woman), leaving the remaining 8 tallest men to match with the 8 shortest women. Thus, in both cases, the prevailing equilibrium will be one in which 80 percent of husbands are taller than their wives, and 20 percent are shorter.

VII. References

Attanasio, Orazio, Hamish Low, and Virginia Sanchez-Marcos. 2008. "Explaining Changes in Female Labor Supply in a Life-Cycle Model." *American Economic Review* 98 (4):1517-1522.

Bailey, Martha J., and Tom A. DiPrete. 2016. "Five Decades of Remarkable, but Slowing, Change in U.S. Women's Economic and Social Status and Political Participation." In *A Half Century of Change in the Lives of American Women*, edited by Martha J. Bailey and Tom A. DiPrete. New York: The Russell Sage Foundation Journal of the Social Sciences.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4):813-846.

Belot, Michele, and Jan Fidrmuc. 2010. "Anthropometry of Love: Height and Gender asymmetries in Interethnic Marriages." *Economics and Human Biology* 8 (3):361-372.

Belot, Michele, and Marco Francesconi. 2013. "Dating Preferences and Meeting Opportunities in Mate Choice Decisions." *Journal of Human Resources* 48 (2):474-508.

Benedetto, Gary, Martha Stinson, and John M. Abowd. 2013. "The Creation and Use of the SIPP Synthetic Beta." Washington, DC: United States Census Bureau. Retrieved from: https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/SSBdescribe_nontechnical.pdf.

Bertrand, Marianne. 2010. "New Perspectives on Gender." In *Handbook of Labor Economics, Volume 4B*, edited by Orley Ashenfelter and David Card, 1545-1592. Philadelphia, PA: Elsevier.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2 (3):228-55.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130 (3):571-614.

Brennan, Robert T., Rosalind Chair Barnett, and Karen C. Gareis. 2001. "When She Earns More than He Does: A Longitudinal Study of Dual-Earner Couples." *Journal of Marriage and Family* 63 (1):168-182.

Chiappori, Pierre-Andre. 2017. Matching with Transfers: The Economics of Love and Marriage. Princeton, NJ: Princeton University Press.

Chiappori, Pierre-André, and Bernard Salanié. 2016. "The Econometrics of Matching Models." *Journal of Economic Literature* 54 (3):832-61.

Eckstein, Zvi, and Osnat Lifshitz. 2011. "Dynamic Female Labor Supply." *Econometrica* 79 (6):1675-1726.

Gale, D., and L. S. Shapley. 1962. "College Admissions and the Stability of Marriage." *The American Mathematical Monthly* 69 (1):9-15.

Gillis, John Stuart, and Walter E. Avis. 1980. "The Male-Taller Norm in Mate Selection." *Personality and Social Psychology Bulletin* 6 (3):396-401.

Hotchkiss, Julie L., and M. Melinda Pitts. 2007. "The Role of Labor Market Intermittency in Explaining Gender Wage Differentials." *American Economic Review* 97 (2):417-421.

Lam, David. 1988. "Marriage Markets and Assortative Mating with Household Public Goods: Theoretical Results and Empirical Implications." *Journal of Human Resources* 23 (4):462-487.

Lundberg, Shelly, Robert A. Pollak, and Jenna Stearns. 2016. "Family Inequality: Diverging Patterns in Marriage, Cohabitation, and Childbearing." *Journal of Economic Perspectives* 30 (2):79-102.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142:698-714.

Mincer, Jacob, and Haim Ofek. 1982. "Interrupted Work Careers: Depreciation and Restoration of Human Capital." *Journal of Human Resources* 17 (1):3-24.

Murray-Close, Marta, and Misty L. Heggeness. 2018. "Manning Up and Womaning Down: How Husbands and Wives Report Their Earnings when She Earns More." U.S. Census Bureau Social, Economic, and Housing Statistics Division Working Paper No. 2018-20.

Pollak, Robert A. Forthcoming. "How Bargaining in Marriage Drives Marriage Market Equilibrium." *Journal of Labor Economics*. DOI: 10.1086/698901.

Raley, Sara B., Marybeth J. Mattingly, and Suzanne M. Bianchi. 2002. "How Dual are Dual-Income Couples? Documenting Change from 1970 to 2001." *Journal of Marriage and Family* 68 (1):11-28.

Ruggles, Steven, et al. 2015. *Integrated Public Use Microdata Series (Version 6.0) [Machine-Readable database]*.

Schwartz, Christine R., and Pilar Gonalons-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.

Stulp, Gert, Abraham P. Buunk, Thomas V. Pollet, Daniel Nettle, and Simon Verhulst. 2013. "Are Human Mating Preferences with Respect to Height Reflected in Actual Pairings?" *PLoS ONE* 8 (1).

Winkler, Anne W. 1998. "Earnings of Husbands and Wives in Dual-Earner Families." *Monthly Labor Review* 121 (4):42-48.

TABLE I

HEIGHT DIFFERENCES BETWEEN HUSBANDS AND WIVES, UK MILLENNIUM COHORT STUDY

Husband height minus wife height (cm)	Proportion in actual distribution	Proportion in distribution with random matching	Ratio of actual to random
<-10	0.6%	1.3%	0.47
-10 to -5	1.5%	2.6%	0.58
-5 to 0	1.9%	2.5%	0.77
0 to 5	8.5%	8.7%	0.97
5 to 10	16.3%	14.5%	1.12
10 to 15	21.3%	19.2%	1.11
15 to 20	20.7%	19.7%	1.05
20 to 25	15.3%	15.8%	0.97
25 to 30	8.8%	9.4%	0.94
30 to 35	3.7%	4.2%	0.87
>35	1.4%	2.1%	0.66

Note: Data taken from Table I in Stulp et al. (2013)

TABLE II
MODEL CALIBRATION

Parameter	Symbol	Calibrated Value
Mean male log earnings	μ^m	10.35
Standard deviation of male log earnings	σ^m	0.75
Mean female log potential earnings	μ^f	10.16
Standard deviation female log potential earnings	σ^f	0.70
Mean disutility of work	ψ	.0019
Standard deviation of disutility of work	σ^ψ	$\psi/2$
Correlation, disutility of work and female log earnings	ρ	-0.4
Standard deviation of transitory income shock	σ^u	13,000
Targets in the data	**Data**	**Model**
Mean male log observed income	10.35	10.35
Standard devation male log observed income	0.75	0.75
Mean female log observed income	10.00	9.98
Standard deviation female log observed income	0.87	0.87
Mean gender earnings ratio, all	0.74	0.71
Mean gender earnings ratio, full-timers only	0.80	0.79
Female labor-force participation rate	0.88	0.91
Female full-time labor-force participation rate	0.67	0.67

Notes: Calibration of marital sorting and female labor supply model discussed in section III.

TABLE III
DISCONTINUITY ESTIMATES IN THE GOLD STANDARD FILE

Bandwidth	Bin size	Treatment of point mass of couples at 0.5		
		Right of 0.5	Left of 0.5	Kick out 0.5 spike, test for break right at 0.5
.084	.0016	***-.124***	**.064**	-.034
		(.031)	(.031)	(.032)
.045	.0016	***-.184***	***.129***	-.031
		(.040)	(.040)	(.043)
.023	.0016	***-.310***	***.240***	-.040
		(.055)	(.055)	(.061)
.011	.0005	***-.575***	***.451***	-.078
		(.078)	(.081)	(.091)

Notes: The first reported bandwidth and bin size correspond to those automatically selected by the McCrary (2008) test algorithm. McCrary (2008) recommends using a smaller bandwidth than the automatically selected one, as is done in the second through fourth rows. Point estimates report the log difference in the height of the density function as one crosses from just left of the supposed breakpoint to just right of it. Bold estimates are statistically significant at the 5 percent level; italicized estimates achieve significance at the 1 percent level. Standard errors appear below point estimates in parentheses.

A. BKP Figure III

B. Replication of BKP Figure III

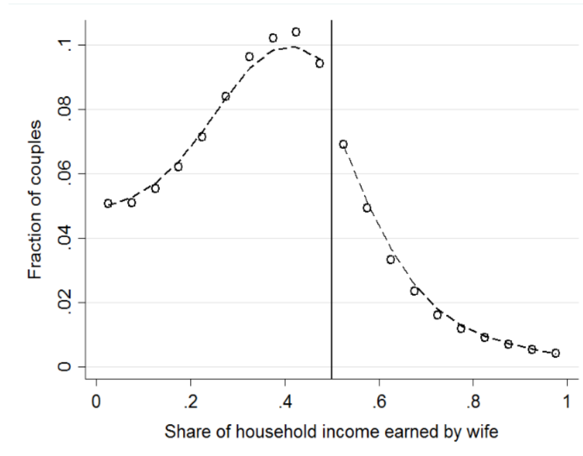


FIGURE I
Distributions of Relative Income, 2000 Census

Graph A is a screenshot of part of Figure III of BKP. Graph B is our replication. Each graph is based on a sample drawn from the 2000 Census consisting of dual-earning couples, in which both the husband and the wife are between 18 and 65 years old. Each graph plots a 20-bin histogram of the distribution of wife's share of a couple's joint income The dashed lines represent the lowess smoother applied to each histogram on either side of 0.5.

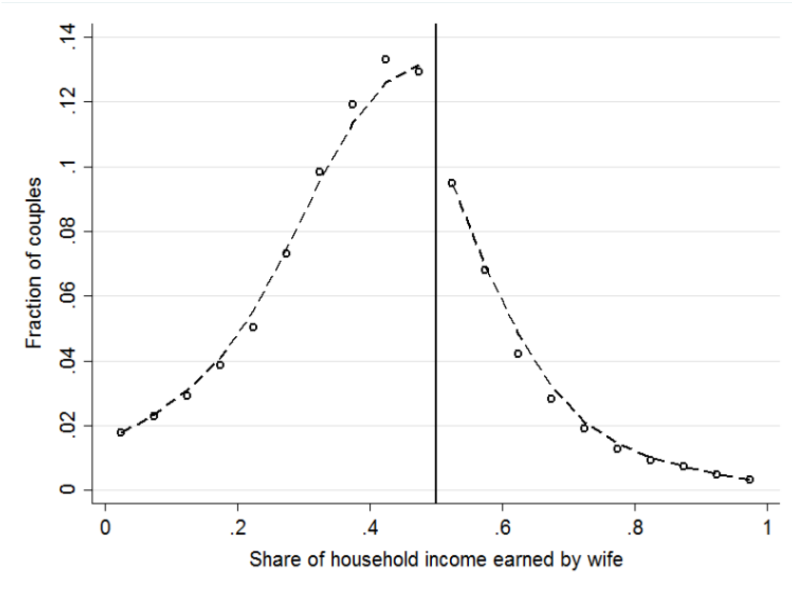


FIGURE II
Distribution of Relative Income, 2000 Census
Couples aged 18-40 without Children

The sample includes dual-earning married couples who do not have children and where both the husband and wife are between 18 and 40 years of age. The figure plots a 20-bin histogram of the observed distribution of the wife's share of total spousal earnings. The dashed lines represent the lowess smoother applied to the histogram on either side of 0.5.

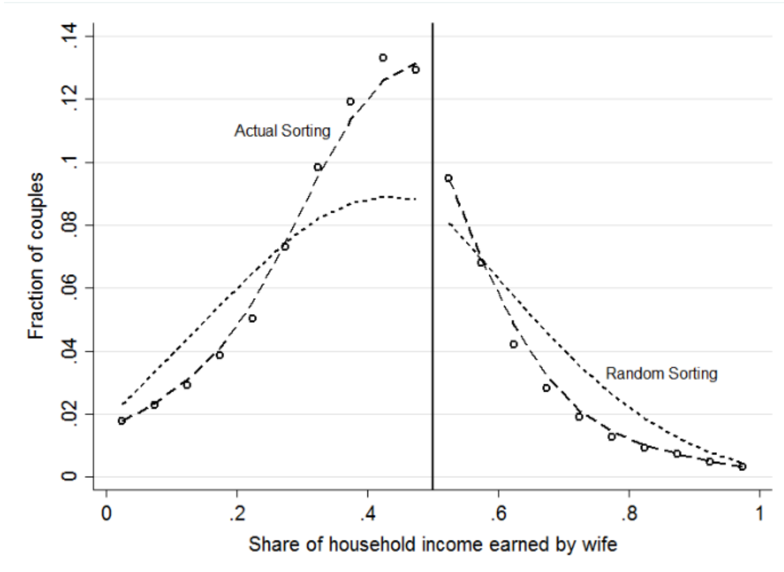


FIGURE III
Relative Income Distributions, 2000 Census: Actual and Random Sorting

The sample is the same as in Figure II. The figure plots 20-bin histograms of the observed distribution of the wife's share of total spousal earnings ("Actual Sorting") and of a simulated distribution based on random sorting of couples in the sample ("Random Sorting"). The dashed lines represent the lowess smoother applied to the histogram on either side of 0.5.

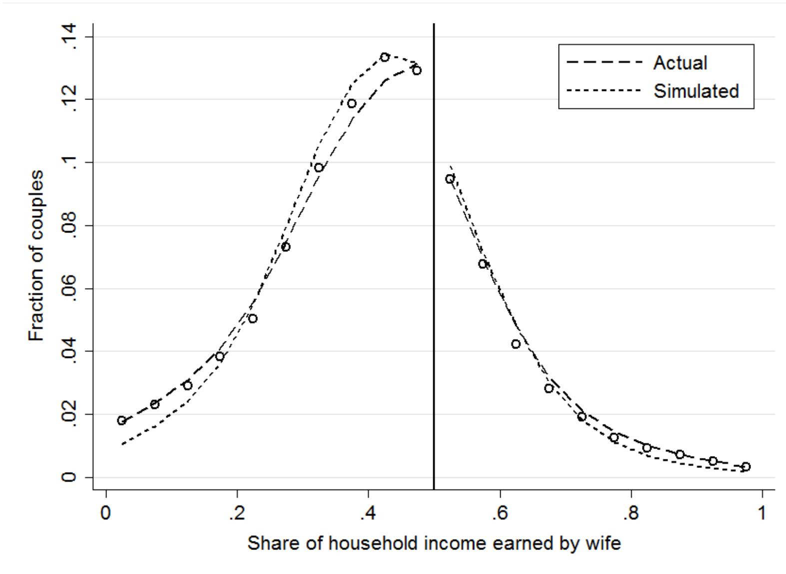


FIGURE IV
Relative Income Distributions, 2000 Census: Actual and Simulated Sorting with Exogenous Earnings

The sample is the same as in Figure II. The figure plots 20-bin histograms of the observed distribution of the wife's share of total spousal earnings ("Actual Sorting") and of a simulated distribution based on positive sorting of couples on observed earnings plus noise ("Simulated Sorting"). See section III for further detail on the simulation. The dashed lines represent the lowess smoother applied to the histogram on either side of 0.5.

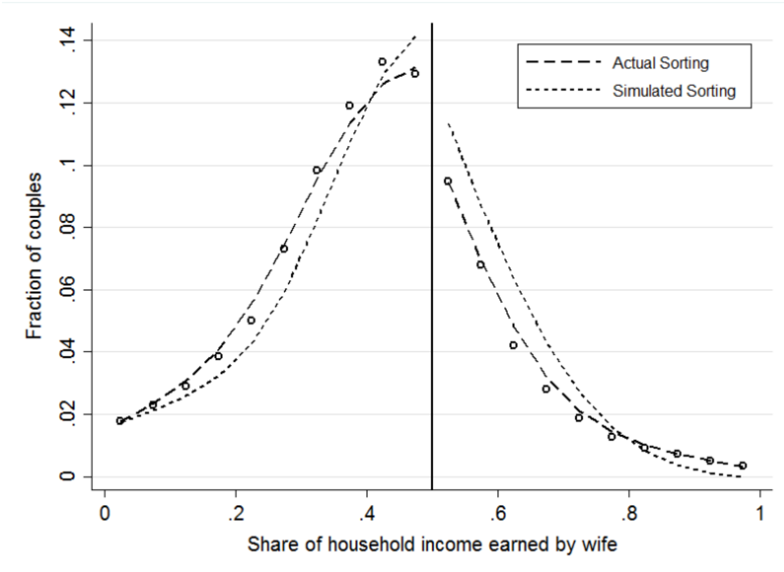
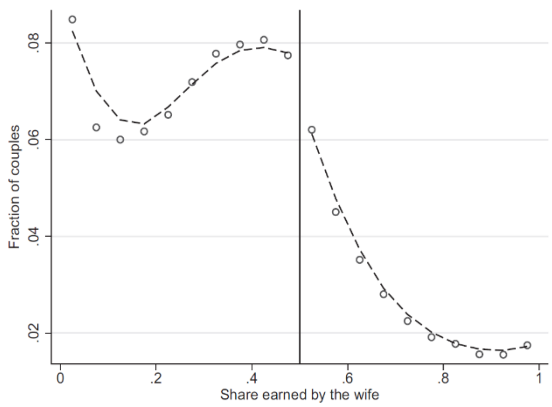


FIGURE V.
Relative Income Distributions, 2000 Census: Actual and Simulated Sorting with Endogenous Earnings

The sample is the same as in Figure II. The figure plots 20-bin histograms of the observed distribution of the wife's share of total spousal earnings ("Actual Sorting") and of a simulated distribution based on positive sorting of couples on potential earnings plus noise ("Simulated Sorting")—and in which the wife's observed earnings are endogenized via a labor supply decision. See section III for further detail on the simulation. The dashed lines represent the lowess smoother applied to the histogram on either side of 0.5.

A. BKP Figure I



B. Replication of BKP Figure I

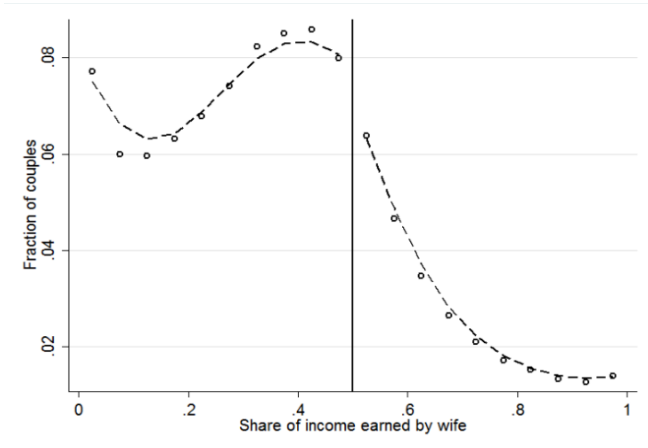


FIGURE VI
Relative Income Distributions in Administrative Data

Graph A is a screenshot of Figure I of BKP. The data underlying this graph are administrative income data from the SIPP/SSA Gold Standard File covering the 1990 to 2004 SIPP panels. Graph B is our replication of Figure I of BKP. We use the latest version of the Gold Standard File, which includes the 1984 and 2008 SIPP panels as well. For both graphs the sample includes all dual-earning couples aged 18 to 65, with income information taken from the first year the couple was observed in the SIPP panel. Both graphs plot 20-bin histograms of the observed distribution of the wife’s share of total spousal earnings. The dashed lines represent the lowess smoother applied to each histogram on either side of 0.5.

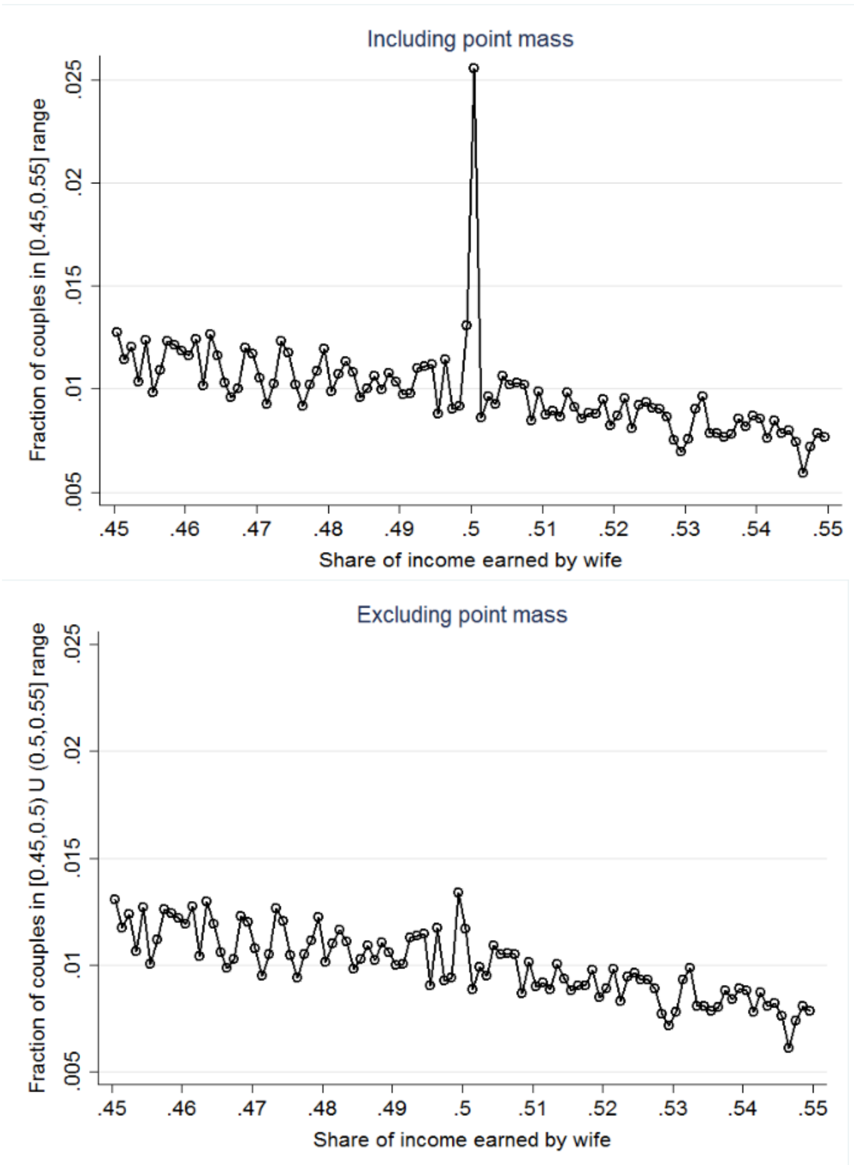


FIGURE VII
Relative Income Distributions in Administrative Data in Neighborhood of 50 Percent

The data underlying this graph are administrative income data from the SIPP/SSA Gold Standard File covering the 1984 and 1990 thru 2008 SIPP panels. For both graphs the sample includes all dual-earning couples aged 18 to 65, with earnings information taken from the first year the couple was observed in the SIPP panel. Both graphs plot histograms of the observed distribution of wife's share of total spousal earnings, restricting the sample to couples in which the wife earns between 45 and 55 percent. The graph in the top panel retains the point mass of couples earning identical incomes; the graph in the bottom panel excludes it. The bin size used in both graphs is .001; each graph contains 100 bins.