Working Papers

FEDERAL RESERVE BANK OF BOSTON

No. 16–20

The Influence of Gender and Income on the Household Division of Financial Responsibility

Marcin Hitczenko

Abstract

This paper studies how gender and income dynamics influence the division of responsibility in two-adult households for various activities, including those tasks directly related to financial decisionmaking. The data, from the 2012 Survey of Consumer Payment Choice, consist of the respondents' categorical self-assessments of their individual levels of responsibility for various tasks. A data construct, in which some households have both adults participate in the survey, is exploited to develop a penalized latent variable model that accounts for systemic response errors. The data reveal that that women, even when they are the primary earner, are much more likely than men to have the major responsibility for household shopping and bill paying. With regard to financial decisionmaking, however, there is a greater propensity to share responsibility equally, and income ranking is more important than gender in defining household roles, with higher earners more likely to have a larger share of responsibility.

Keywords: gender roles, household finance, probit models, penalized maximum likelihood

JEL Classifications: A14, C51, D13, J16

Marcin Hitczenko is a survey methodologist and a member of the Consumer Payments Research Center in the research department of the Federal Reserve Bank of Boston. His email address is marcin.hitczenko@bos.frb.org.

This version: October 17, 2016

1 Introduction

Every household must decide how to allocate responsibilities and work among its members. Cultural influences, most notably traditional gender roles (Cunningham 2001; Sayer, Cohen, and Casper 2004), and more rational approaches founded in economic theory represent at least two methods that households employ to assign tasks. For example, the "economic exchange hypothesis" suggests that partners with greater resources essentially "buy out" of doing unenviable tasks (Huston and Burgess 1979; Mannino and Deutsch 2007), while Becker (1991) posits that households allocate tasks to maximize overall utility. Both approaches suggest basing household roles on the members' relative economic standing within the household rather than on gender.

Most research on household dynamics has focused on how married couples divide "housework," a broad term that generally includes routine chores such as cleaning, child care, or house repair. Though there is a fair amount of debate about the mechanisms that determine the division of housework, it is clear that gender plays an important role in these assignments. Tasks are still viewed as inherently masculine or feminine, with responsibility for each chore more likely to be controlled by their respective gender affiliation, regardless of economic variables (Bianchi et al. 2000; Lam, McHale, and Crouter 2012; South and Spitze 1994). There is less agreement among researchers on how the income dynamics among household members affects the amount of time each member devotes to housework. Some studies find evidence of economic exchange and others argue that housework is used to neutralize gender discrepancies (Bittman et al. 2003; Gupta 2007; Parkman 2004; Schneider 2012).

By contrast, there has been considerably less research on how adult members divide household responsibility for the making of important decisions, especially those related to household finances. Since such activities relate to a household's long-term goals and well-being, assigning responsibility for major financial decisions is likely to be governed by different principles than those governing who performs routine household tasks. For example, the importance of certain financial decisions might make joint consideration more common. In addition, if one individual is to take on primary responsibility for financial planning, it seems probable that it will be the person who earns the most income and not necessarily the one with more available time. In the lone relevant study, Dobbelsteen and Kooreman (1997) find that higher wages generally correspond to a greater share of financial responsibility for each gender. Understanding how household dynamics function with respect to household decisions relating to financial planning thus offers a different perspective on household economics and gender relations in a society.

Perhaps the lack of reliable data is a major reason for the dearth of research on how households make major financial decisions. The study of housework often relies on diary-based records of the amount of time spent on various activities, readily available in surveys such as the Panel Study of Income Dynamics and the American Time

Use Survey. Comparable measures do not exist for more nebulous concepts. Because the degree of a person's involvement in decisionmaking is difficult to measure, the relevant data is effectively limited to respondents' self-assessments of household dynamics. Well-known, systematic response error can make such data difficult to analyze and lead to poor inferences (see Fowler (1995) for summary of these issues). The study by Dobbelsteen and Kooreman (1997) had to discard up to 33 percent of the sample data due to the fact that members of the same household provided inconsistent characterizations of household dynamics on a three-point Likert scale.

This paper studies the responses to four questions related to financial responsibility among two-adult, mixed-gender households taken from the 2012 Survey of Consumer Payment Choice (SCPC). The four activities in question fall at various points on the spectrum spanning routine financial activities that are commonly included in housework to the making of important financial decisions. While one question relates to household shopping, two others relate to "making decisions about savings and investments" and "making all other financial decisions," defined as deciding where to bank, choosing what payment methods to use, setting up online bill payments, and filing taxes, respectively (see Figure 2). A fourth question about the task of paying monthly bills is perhaps more ambiguous, falling somewhere between the category of a routine household chore and financial planning. The paper's analysis focuses on how, among the two adults within a household, their respective gender and relative income ranking affects the distribution of financial responsibility.

The paper is structured in the following way. Section 2 defines the analytical framework, while Section 3 introduces the survey questions and the sample used in the analysis. Section 4 studies various properties of the observed data, simultaneously arguing for the value of the data but highlighting residual effects of response error that must be accounted for to make inferences. Section 5 presents a latent variable model that relies on penalized maximum likelihood to incorporate prior notions of trend smoothness into population estimates. The results are introduced in Section 6. Finally, Section 7 provides a brief discussion of the findings and potential future work.

2 Model Framework

This paper's primary goal is to gain insight into how two-adult, mixed-gender households distribute financial responsibility among both adult members by analyzing survey questions in which the respondents declare whether their share of responsibility for a task is minor, major, or split equally with the other adult. I begin by establishing a conceptual and statistical framework for the quantities that the survey questions hope to measure as well as those that are actually observed. The desired set of responses, which can be imagined as characterizations of household dynamics in which both adult members collaborate to reach a consensus, are dubbed "household-consistent." Note

that these characterizations, even those agreed upon by household members, are influenced by biases and social pressures, and they may not be objective reflections of reality.

The mathematical model for identifying true and reported financial responsibility is as follows. For a given household, h, with two adult members, $i = 1, 2$, let $T_{hi} = 1, 2, 3$ represent household-consistent scores corresponding to a minor, equal, and a major share of responsibility, respectively. Household consistency requires that $T_{h1} + T_{h2} = 4$. Instead of T_{hi}, survey data provide $R_{hi} = 1, 2, 3$, which are interpreted the same way, but no longer have the additive restriction.

The interest is in how gender and the household members' relative economic status interact to affect household dynamics. While economic influence might best be identified by the household members' individual incomes, the data limitations only permit categorical representations of income dynamics. Thus, household h will fall into one of three types: one in which the male is the primary earner ($t[h] = 1$), one in which both adults' incomes are roughly equal ($t[h] = 2$), and one in which the female is the primary earner ($t[h] = 3$). For consistency, the male in each household is indexed with $i = 1$ and the female with $i = 2$. The paper's fundamental statistical challenge is to determine the distribution of T_{hi} as a function of $(t[h], i)$ based on observing R_{hi}.

3 Data

The data used in this work come from the 2012 Survey of Consumer Payment Choice (Schuh and Stavins 2014), an annual online survey conducted by the Consumer Payments Research Center at the Federal Reserve Bank of Boston. The SCPC samples adults, defined as those 18 years of age and older, from the RAND Corporation's American Life Panel (ALP), and gathers data about respondents' preferences and behaviors with regard to payments and certain aspects of household economics.

3.1 The 2012 SCPC Sample

The 2012 SCPC sample features 3,176 total respondents, of which 1,499 come from two-adult, mixed-gender households. Since its inception in 2003, the ALP has employed a variety of sampling strategies, some probability-based, to amass roughly $5{,}000$ individuals at the time of sample selection in September 2012. One notable sampling strategy encouraged existing panelists to invite fellow household members to join the ALP. Members recruited in this way made up around 20 percent of the ALP in 2012. Among the subset of two-adult, mixed gender households in the SCPC, there are 185 households with both adults in the sample. For notational purposes, let $s[h] = 1, 2$ define the number of adult members sampled from household h. The presence of the dual-sampled households will prove

invaluable in analyzing household trends.

Table 1 shows the number of respondents by household and member type, as well as by the number of members sampled from the household. The male is the primary wage earner in the majority of households, and there are slightly fewer households in which the female is the primary earner than those in which incomes are roughly equal. The relative proportions of household types in the 2012 SCPC are similar to the corresponding proportions among married couples in the 2008–2011 American Community Survey, in which each member has positive income and with equal incomes defined by respective shares between 40 and 60 percent of total household income (Bertrand, Kamenica, and Pan 2015). Interestingly, among single-sampled households, higher earners and females, and especially both, are more likely to be included in the sample. Section 5 discusses the potential implications of this fact on estimation.

The selection of ALP panelists into the SCPC sample is guided by the goal of creating a cohort that is representative of U.S. adult consumers. Therefore, there is reasonable concern about how well the sample of two-adult, mixed-gender households represents the U.S. population. However, as seen in Figure 1, with regard to two measurable demographic variables known to affect economic behavior, age and income (Schuh and Stavins 2010), the sample distributions in the 2012 SCPC match reasonably well to those in the 2012 Current Population Survey. Based on these findings, I feel comfortable generalizing results from the sample without incorporating any post-sampling adjustments.

3.2 Survey Data

The response data come from a set of "financial responsibility" questions in the SCPC. Each respondent is asked to assess his or her level of responsibility for "paying bills," "household shopping," "making decisions about saving and investments," and "making decisions about other household financial matters." A screenshot of the four survey question is shown in Figure 2. The response rates are very high, as only 0.4 percent of these survey questions are unanswered. Original responses correspond to a 5-point Likert scale, but, for the purposes of this paper, these responses are reduced to a 3-point scale by combining "None or almost none" with "Some," and "Most" with "All or almost all." Doing so eliminates some of the household inconsistency that derives from natural variation in delineation of the categories. For example, the difference between "most" and "almost all" or between "some" and "almost none" may be defined differently by different individuals.

Information about the age and gender of the respondents' household members comes from the "My Household Questionnaire," a brief, demographic survey that is updated for all ALP panelists every three months. Income

ranking within households is deduced from a question in the SCPC that asks respondents to specify whether their personal income is the "highest in the household," "about equal to the highest," "second highest in the household," or "third highest or lower." In the case of two-adult households, responses to this answer provide a precise relative ranking of incomes of both members. Of those households in which both adults respond, roughly 15 percent have contradictory responses (for example, both members claim to have the highest income). This could be a result of reporting error or a lack of familiarity with the other member's income. To deal with this discrepancy, all such households are classified as ones in which both members make roughly the same income. Combining information about age, gender, and income ranking allows classification of each respondent according to household and member type.

4 Characteristics of the Financial Responsibility Responses

In this section, I discuss some properties of the responses to the financial responsibility questions. I begin by highlighting aspects of the data that suggest the self-assessments, although subject to response error, offer useful insights into household dynamics. I then explore the nature of the response errors and their implications on modeling and parameter estimation.

4.1 Household-Consistency

As defined in Section 2, a necessary condition for household-consistent responses is mutually compatible ratings within a household or, using this paper's notation, that $R_{h1} + R_{h2} = 4$. Table 2 shows the distribution of the sums of household ratings, $R_{h1} + R_{h2}$, for all dual-sampled households. A majority of these households provide ratings that are household-consistent, though the rate of consistency varies substantially, from roughly 55 percent for both activities relating to financial decisionmaking to 76 percent for paying bills. The variation of consistency rates across these four activities is perhaps due to greater ambiguity in assessing the contribution levels for certain activities, particularly those more likely to be shared. This source of the inconsistency might also explain why there are more aggregate scores of 3 and 5 than there are of 2 and 6. Overall, the general trend tends toward consistent responses, which suggests that for many households, members are independently characterizing household dynamics in the same way.

In fact, with regard to paying bills, there is evidence that the self-assessments might be more reliable measures of responsibility level than recall-based estimates of the number of monthly bills paid, a seemingly natural measure of an individual's contribution. For households featured in the 2011 and the 2012 SCPC, Figure 3 plots one house-

hold member's share of financial responsibility scores $\left(\frac{R_{h1}}{R_{h1}+R_{h2}}, \text{but using the original 5-point scale}\right)$ in both years as well as that same member's share of the reported number of monthly bill payments made in both years. Despite prior expectations that responsibility levels likely reflect the share of bills paid, the two statistics are not compatible. Perhaps more meaningful is the fact that there is much more consistency in the shares of financial responsibility scores from 2011 to 2012 than in the shares of bill payments. Under the assumption that household roles are generally stable across a given year, a plausible explanation is that recall-based estimates, which are known to be affected by nontrivial response error (Marini and Shelton 1993; Press and Townsley 1998), are too noisy to reliably reflect the truth. As the SCPC data on purchases, unlike the financial responsibility question, does not distinguish household payments from all other payments, a similar comparison for household shopping cannot be made.

4.2 Response Error

Despite evidence that the SCPC responses are rooted in reality, it is undeniable that the observed data are influenced by systematic response bias. As mentioned in the previous section, it is more likely that the aggregate scores will be close to consistency than farther from consistency. In addition, Table 2 indicates that there are two to four times as many scores over 4 than under 4, suggesting a systematic tendency to overstate one's individual contribution, though the extent of this phenomenon depends on the activity. Figure 4 shows the sample distributions of responses for each of the four financial responsibility questions, but distinguishes the respondents by the number of household members featured in the sample. The tendency to shift toward higher ratings manifests itself in all respondents, not just those in dual-sampled households. This general phenomenon is consistent with "overconfidence," in which people show a favorable disconnect between their subjective self-assessment and objective reality (Moore and Healy 2008). Similar patterns of over-valuing one's own contributions to household labor when comparing self-assessments to diary data are found by Bianchi et al. (2000) and Mizan (1994). Men, in particular, have been known to inflate their role in carrying out household chores (Kamo 2000; Lundeber, Fox, and Punccohar 1994). In general, the variability of these patterns across activities suggests a complex underlying response process.

Response patterns also seem to differ between dual- and single-sampled households. The greatest such discrepancy is found in the response counts for the question on saving and investment decisions for households in which women are the higher earners (corresponding to the third plot in the third row of Figure 4). Among individuals from dual-sampled households, there is a fair amount of overall marginal consistency: a large proportion of 2 ratings, with women ($i = 2$) showing a slightly higher probability of claiming most of the responsibility compared to men ($i = 1$). However, lone household representatives, especially females, show a much higher tendency to report

higher shares of financial responsibility. Indeed, 66 percent of females (125/189) claim a majority of household financial responsibility, but only 20 percent of males (7/35) profess to have minor responsibility for these tasks. A likelihood ratio test for which the null hypothesis is that the response distribution depends only on member-type and not on the number of members sampled yields a p-value of 4.2×10^{-4}. Performing similar tests for all 12 cases yields four p-values below 0.01 and another four below 0.01, suggesting that the phenomenon of different response patterns between single- and dual- sampled households is relatively common.

From a survey methodological viewpoint, one must wonder whether the different marginal distributions of response rates are due to a selection effect or to response error. The former theory suggests that certain types of households, in this case those with one dominant financial decisionmaker, are less likely to volunteer a second adult to participate in the survey. The response-error theory might posit that individuals who know (or likely know) that both members are taking the survey may be more guarded and careful with their own self-assessment and be less likely to exaggerate their individual contribution. While the observed differences are likely due to a combination of factors, I assume that a majority of the discrepancy is a result of reporting error. Since ALP panelists were encouraged to volunteer household members far in advance of the SCPC and without the context of answering questions about the household's finances, considerations of an individual's household contribution seem unlikely to influence a household member's decision to participate. As a result, the model introduced in Section 5 assumes the same underlying distribution of household-consistent responses, but differing response error distributions for single- and dual-sampled households.

5 Methodology: Model and Estimation

In order to estimate the distributions of household-consistent responses, it is necessary to develop a methodology that adjusts for response errors. The following analysis relies on the optimization of a penalized log-likelihood function, which offers two methods to account for response error. First, a latent variable model for the response-generating process attempts to capture the relationship between household consistent and reported responses. In addition, the penalty imposes prior notions about how much the dynamics in one type of household are likely to differ from those in a different household type. All aspects of the model are described in further detail below.

5.1 Household-Consistent Model

A latent variable model for household-consistent responses, T_{hi}, takes the form

$$T_{hi} = \begin{cases} 1, & X_{hi} < -c_h \\ 2, & -c_h \leq X_{hi} \leq c_h \\ 3, & X_{hi} > c_h \end{cases},$$

for some underlying continuous variable X_{hi}. The necessary condition of $T_{h1} + T_{h2} = 4$ is imposed by $X_{h1} = -X_{h2}$. Under a probit model (McCullagh and Nelder 1989), $X_{h1} \sim \text{Normal}(\mu_h, 1)$ and $X_{h2} \sim \text{Normal}(-\mu_h, 1)$.

The simplest mean function distinguishes household members by gender and income rankings, taking the form $\mu_h = \mu(t[h])$. In addition, information about the adults' respective ages is incorporated through an interaction with gender and income ranking. Defining δ_h as the difference in age between the male and the female, an alternate model is $\mu_h = \mu(t[h]) + \alpha(t[h])\delta_h$. The distribution of δ_h for the SCPC sample, plotted in Figure 5, shows that, although men tend to be slightly older than the women with whom they are coupled, δ_h is highly concentrated around 0. This is not surprising, as life partners tend to be of similar age.

5.2 Response Error Model

To accommodate the inconsistency of responses, assume that, rather than observing T_{hi}, the observed responses are

$$R_{hi} = \begin{cases} 1, & Y_{hi} < -c_h \\ 2, & -c_h \leq Y_{hi} \leq c_h \\ 3, & Y_{hi} > c_h \end{cases}, \tag{1}$$

where $Y_{hi} = X_{hi} + \epsilon_{hi}$. In this model, ϵ_{hi} are independent across h and i, implying that one individual's response error has no effect on anyone else's. If $\epsilon_{hi} \sim \text{Normal}(\tau_h, \sigma^2_{hi})$, then

$$\begin{bmatrix} Y_{h1} \\ Y_{h2} \end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix} \mu_h + \tau_h \\ -\mu_h + \tau_h \end{bmatrix}, \begin{bmatrix} 1 + \sigma^2_{h1} & -1 \\ -1 & 1 + \sigma^2_{h2} \end{bmatrix} \right). \tag{2}$$

The decision to set both error means to τ_h, rather than have them depend on i, is driven by an identifiability issue. Namely, knowledge of $\mu_h + \tau_{h1}$ and $-\mu_h + \tau_{h2}$ does not offer a unique solution for μ_h. This restriction corresponds to minimizing the total expected shifts: $|\tau_{h1}| + |\tau_{h2}|$.

Letting θ_t define the parameters for household type t, the likelihood of the observed data for household type t takes the general form:

$$\text{L}_t(\theta_t) \quad = \prod_{\{(h,i)|t[h]=t,s[h]=1\}} \text{Prob}\left(R_{hi} \mid \theta_t\right) \times \prod_{\{h|t[h]=t,s[h]=2\}} \text{Prob}\left(R_{h1}, R_{h2} \mid \theta_t\right), \tag{3}$$

with the probabilities defined by (1) and (2).

In the case of single-sampled, two-adult households, integrating over the unsampled household member's response, the observed data likelihood will be defined by one of the two univariate Normals implied in (2): $\text{Normal}(\mu_h + \tau_h, 1+\sigma_{h1}^2)$ or $\text{Normal}(-\mu_h+\tau_h, 1+\sigma_{h2}^2)$. By the properties of a Normal distribution, the likelihood under parameters (μ, τ, σ, c) will be identical to that under $(\kappa\mu, \kappa\tau, \sigma', \kappa\, c)$, where $\kappa = \sqrt{\frac{1+\sigma^2}{1+(\sigma')^2}}$. Choosing σ instead of σ' means that household-consistent distributions are defined by a $\text{Normal}(\mu, 1)$ distribution with rating thresholds $(-c, c)$ rather than by a $\text{Normal}(\kappa\mu, 1)$ distribution with rating thresholds $(-\kappa\, c, \kappa\, c)$. The implications for household-consistent trends under the two parameterizations are different. In other words, with only single-sampled households available, it is impossible to determine how much variation in the observed responses is due to variation in household-consistent responses, defined by μ_h and c_h, and how much is introduced by response error, defined by σ_{hi}.

If both household members are sampled, this form of unidentifiability does not exist, as the added information about household consistency defines the extent of the response error. Considering equation (2), high levels of inconsistency correspond to less correlation between the latent variables Y_{h1}, Y_{h2}, thus driving up the values of σ_{h1} and σ_{h2}. For a given set of marginal distributions of the responses, larger values of σ_{hi} necessitate a larger value of c_h, which defines the distribution of T_{hi}. As a result, high rates of inconsistency correspond to estimates in which household consistent scores have a high percentage of 2s.

In light of these results, the value of single-sampled households in parameter estimation is not obvious. To avoid the aforementioned identifiability issues, the response error process of single-sampled households must be linked to that of dual-sampled households. However, the results in Section 4 show that the response error distributions cannot be the same if one assumes that selection effects do not account for the observed differences. The natural compromise is to assume that certain aspects (parameters) of the response errors are the same. Both of the models in Table 3 are considered in estimation, one which allows different response error means for single- and dual-sampled two-adult households and one which allows different response error variances.

In the case of the SCPC, there is a potential drawback to using the single-sampled households, due to the fact that household member types are not equally likely to represent their households, as seen in Table 1. While an ideal sampling design would select household members at random from within each household, it is clear that the 2012 SCPC selection process, either directly or indirectly, favored certain genders and income rankings. In order for the selection process not to bias the results, one must assume that the sample members are representative of their household type. Because such a condition is unverifiable, I consider estimates based on the inclusion of single-sampled households and those based only on dual-sampled households.

5.3 Penalty Term

A second tool that helps generate household-consistent estimates is based on the idea that if the likelihood of a particular household dynamic is a sufficiently smooth function of the income differential between the household's two members, then averaging within the three household types restricts the degree of differences between household-type aggregates. Specifically, if $p_t = [p_{t1}\ p_{t2}\ p_{t3}]$, with $p_{tj} = \text{Prob}(T_{h1} = j \mid t[h] = t)$, then one expects that p_1 (when men are the highest earners) is not too different from p_2 (when earnings are the same across both genders), which is not too different from p_3 (when women are the highest earners). Below, a simple simulation is developed to motivate the penalty term and provide insight into likely levels of disparity between $p_1, p_2,$ and, p_3.

Let inc_h be the difference in income between the male and female of a two-adult, mixed-gender household h. A smoothed density for inc_h, based on household-level data on individual incomes among relevant households in the 2012 Current Population Survey, is shown in Figure 6. Again, men tend to earn more than women.

Next, consider that the household distribution of financial responsibility is defined by a pair of linear functions of inc_h. To be valid, the set of all possible models is defined by slopes (m_1, m_2) and intercepts (b_1, b_2) such that the lines,

$$\text{Prob}(T_{h1} = 1 \mid \text{inc}_h) = m_1\text{inc}_h + b_1 \quad \text{and} \quad \text{Prob}(T_{h1} = 2 \mid \text{inc}_h) = (m_2 - m_1)\text{inc}_h + (b_2 - b_1), \tag{4}$$

do not intersect on the observed range of inc_h and all probabilities are in $[0, 1]$ (practically, I assume a linear model on the range $[-\$1e5, \$2e5]$, and values of inc_h outside of this interval are shifted to the closest point within the interval). Factors other than income differences that affect role distributions within the individual households will naturally dampen the slopes, especially if these factors are weakly correlated with income differences. Examples of 16 models are provided in Figure 6.

For each model, the penalty term, p_t, is determined by integrating with respect to the estimated distribution of income differences, which is done using Monte Carlo methods. This is accomplished by defining households with equal incomes as the 25 percent of households closest to $\text{inc}_h = 0$, roughly matching the fraction of households that declared this status in the 2012 SCPC. A useful way to quantify the difference in any two probability distributions is through the symmetrized Kullback-Leibler divergence:

$$\text{SKL}(p_t, p_{t'}) \quad = \quad \sum_{j=1}^{3} \left(p_{tj} \log\left(\frac{p_{tj}}{p_{t'j}}\right) + p_{t'j} \log\left(\frac{p_{t'j}}{p_{tj}}\right) \right), \tag{5}$$

which serves as a distance measure between distributions (Kullback and Leibler 1951). Extending this concept to

the context of this work, the quantity

$$\text{pen}(p_1, p_2, p_3) \quad = \quad \text{SKL}(p_1, p_2) + \text{SKL}(p_2, p_3) \tag{6}$$

represents a measure of change between successive household types. Larger penalty values correspond to larger absolute slopes, m_1 and m_2, meaning income differences have more drastic effects on household dynamics. Figure 7 shows the distribution of penalties if one assumes a uniform distribution on all valid models. In fact, this distribution, especially the range of observed penalty values, is fairly robust to variations of the assumed model. For example, relaxing the monotonic condition, perhaps by allowing a changing slope at $\text{inc}_h = 0$, does not substantially affect the distribution.

The distribution in Figure 7 serves as a useful guideline for evaluating the plausibility of the estimated p_t. A penalty that seems unlikely relative to the simulated distribution is evidence that the implied relationship between income differences and the responsibility distribution is too strong to reflect reality. The following subsection describes how the penalty term is used to generate estimates that have transitions consistent with the above simulation.

5.4 Fitting the Model

5.4.1 Optimization

For a set of parameters, $\theta = \{\theta_1, \theta_2, \theta_3\}$, the objective function is

$$\text{obj}(\theta) \quad = \quad -\sum_{t=1}^{3} \log \text{L}_t(\theta_t) + \lambda \times \text{pen}(p_1(\theta_1), p_2(\theta_2), p_3(\theta_3)), \tag{7}$$

where $p_t(\theta_t)$ is the implied distribution of household-consistent ratings under θ_t for household t. I use the *nlm* function in *R* to optimize equation (7) for $\lambda = 0, 10, \ldots, 80$ and let $\hat{\theta}_t(\lambda)$ represent the maximum likelihood estimates. The general model given in equation (7) is fit using both specifications given in Table 3. For all four activities, Model A greatly outperforms Model B in terms of likelihood fit, despite having one fewer parameter. As the response pattern for single-sampled households, relative to that for dual-sampled households, suggests a greater tendency to report higher ratings, the relative efficacy of a model that allows for a shift in response distributions (Model A) over one that corresponds to a diffusion of response ratings (Model B) is not surprising. The following discussion concentrates on the results from Model A.

In the adopted paradigm, θ_t naturally depends on the choice of λ. In choosing λ, the quality of fit must be balanced with the penalty term. The top panel of Figure 8 shows the relative quality of fit relative to the unpenalized

versions, as measured by the difference in log-likelihoods:

$$\Delta(\lambda) \quad = \quad \sum_{t=1}^{3} \log \mathrm{L}_t(\hat{\theta}_t(0)) - \sum_{t=1}^{3} \log \mathrm{L}_t(\hat{\theta}_t(\lambda)).$$

The second panel in Figure 8 shows the penalty values under different λ for each of the four financial activities. As expected, larger values of λ dampen the changes across household types, but in doing so, worsen the fit. For three of the four activities, including the penalty term has relatively modest effects, and the parameter estimates themselves do not vary much. For decisions about savings, there is a notable effect. More detail on the the effect of using the penalty term on parameter estimates is provided in Appendix A.

A worse fit, in terms of likelihood, corresponds to a greater divergence between the fitted model's output and the observed data; hence the model's predictive ability worsens. However, as it is doubtful that the simple model is able to fully capture the cognitive process by which response errors occur, I focus more on preserving smoothness. I thus choose the smallest value of λ for which all four penalties are less than 0.15 (based on the approximate 95^{th} percentile of the maximum of four independent draws from the distribution in Figure 8), which in the case of the 2012 SCPC data is $\lambda = 70$. Restricting the analysis only to the dual-sampled households provides similar results, as discussed in Appendix B.

6 Results

6.1 Age Effects

For all but one of the 12 activity and household-type combinations, the inclusion of a linear age term significantly improved likelihood fits, as measured by the likelihood ratio test for nested models (Wilks 1938). In all but one case, the estimated age effect was positive, suggesting that the older adults in the household are more likely to have a greater share of the responsibility for handling financial matters. Interestingly, models that did not feature age in the mean function led to different response error-related parameter estimates, but had virtually no effect on estimates of the distribution of household-consistent ratings when the age difference between the two adults is zero.

However, I am hesitant to read more into these results. A thorough model might allow for minor effects when age differences are small and tapering beyond some age-difference threshold, thus making a nonlinear term realistic. This type of relationship is difficult to discern using these data due to the scarcity of observations distant from zero, nor is such inquiry the focus of this paper.

6.2 Trends in Consistent Responses

This section studies the distribution of consistent responses, $p_t = [\text{Prob}(T_{h1} = j \mid t[h] = t)]_{j=1}^{3}$, for all t. I focus on those cases where the household members are identical in age. The estimated means and standard errors are used to effectively define the posterior distributions of $(\mu_{t[h]}, c_{t[h]})$ and, in turn, the distributions of each household member's role, p_t. Standard errors for the estimates come from the observed Fisher information matrix (Efron and Hinkley 1978).

Figure 9 shows the estimated frequency of rating distributions for each financial activity and household type. Uncertainty in these estimates is provided in Appendix C. These results suggest a clear dichotomy in the four activities. Financial decisions are much more likely to be shared equally, with over 50 percent of households estimated as doing so no matter the member's gender and relative income. The share of households equally sharing responsibility for bill payments or household shopping is much lower. Instead, women have the major responsibility for both tasks in close to 50 percent of households, regardless of their income ranking. This difference between activity types reflects both the higher proportion of 2 ratings in the data for financial decisions (see Figure 4) as well as a lower rate of consistency among dual-sampled households (see Figure 2). The dichotomy is not surprising, as financial decisions are often important enough to discuss among both adult members, whereas the other two activities are more routine tasks that are more efficiently undertaken by one individual.

Second, for all activities except household shopping, the likelihood of having more responsibility increases as an adult's relative income ranking increases. Because these three activities — paying bills, making decisions about saving and investment choices, and other financial decisions — relate to the household's finances, this result is not surprising and is consistent with the findings of Dobbelsteen and Kooreman (1997). On the other hand, with regard to shopping, there is slight evidence that bringing in a greater share of the household income increases a member's chance of doing a major portion of the shopping for the household. Perhaps this result is a manifestation of gender neutralization discussed in Bertrand, Kamenica, and Pan (2015), in which greater financial contribution by women is balanced by women having greater responsibility for housework.

These results can also be used to determine the relative impact that gender and income ranking have in determining how mixed-gender households divide responsibility for the four financial activities. This is accomplished via an ANOVA-like (Yates 1934) decomposition of variance of expected ratings based on random selection among the six member types. Consider m_{gr} to be the expected rating value of a household member with gender $g = 1, 2$

(male, female) and income ranking $r = 1, 2, 3$ (first, equal, second). Then, define

$$\begin{aligned} m_{gr} &= \mathrm{E}\left[T_{hi} \mid \text{individual } (h,i) \text{ has gender } g \text{ and income ranking } r\right] \\ m_{g\cdot} &= \mathrm{E}\left[T_{hi} \mid \text{individual } (h,i) \text{ has gender } g\right] \\ m_{\cdot r} &= \mathrm{E}\left[T_{hi} \mid \text{individual } (h,i) \text{ has income ranking } r\right]. \end{aligned}$$

Under the assumption of a uniform draw among member types, the variance of expected rankings decomposes into:

$$\begin{aligned} \mathrm{Var}\left[m_{gr}\right] = \sum_{g=1}^{2}\sum_{r=1}^{3}\left(m_{gr}-2\right)^2 &= \frac{1}{2}\sum_{r=1}^{3}\sum_{g=1}^{2}\left(m_{gr}-m_{\cdot r}\right)^2 + \frac{3}{2}\sum_{g=1}^{2}\left(m_{g\cdot}-2\right)^2 \\ &\quad + \frac{1}{2}\sum_{g=1}^{2}\sum_{r=1}^{3}\left(m_{gr}-m_{g\cdot}\right)^2 + \sum_{r=1}^{3}\left(m_{\cdot r}-2\right)^2. \end{aligned} \tag{8}$$

The first row on the right-hand side of equation (8) represents variation explained by gender: the first term reflecting variation across genders for fixed rankings and the second term reflecting variation across genders. If these terms are both zero, income ranking alone determines the distribution of ratings, and men and women have identical distributions of ratings given the same income ranking. The second row in equation (8) does the same for income ranking. Figure 10 shows the relative proportions of the gender-related contribution to the overall variance depicted on the left-hand side of equation (8) for 200 posterior draws of p_1, p_2 and p_3 for all four financial activities.

The most notable result shown in Figure 10 suggests that gender almost fully determines the likely divisions of responsibility for household shopping. In other words, the expected distribution of household-consistent ratings barely changes across income rankings for each gender. Overall, this is consistent with previous research (Bianchi et al. 2000; Lam, McHale, and Crouter 2012; South and Spitze 1994) suggesting that women are much more likely to control household shopping, as this household task is traditionally associated with women. Decisions about other financial matters, on the other hand, is predominantly linked to income ranking. Indeed, looking at Figure 9, the distribution of ratings is similar for higher-earning males and higher-earning females, and in households with equal incomes, each adult is equally likely to have the major responsibility for decisionmaking about other financial matters. Gender and income ranking have roughly equal influence on responsibility for paying bills and making decisions about saving and investments.

7 Conclusion

The financial responsibility data analyzed in this paper are self-reported, categorical assessments of respondents' contributions to a household's financial activities. This paper develops a methodology that relies on a data construct, in which a portion of two-adult households have both members take the 2012 SCPC, in order to account for response error. Prior notions about the extent of differences between the distribution of household roles among the three mixed-gender household types are incorporated through the use of penalized maximum likelihood. The ability to extract meaningful information from such data is a worthwhile goal as it allows insight into household dynamics for a wide range of abstract concepts that cannot be measured with objective measures.

From the viewpoint of understanding household dynamics, a clear downside to the SCPC data is the lack of more detailed information about the household members. The influence of factors such as education levels, the nature of the relationship between members, the number of dependents, and the time spent at work also are of interest. Such variables can be included in the models, much like age was included. The analysis is based on income ranking alone, with no ability to better classify the income dynamics of individual households. Distinguishing between a household in which one member has a slightly lower income than the other and a household in which one member is unemployed is important, but a distinction that the SCPC data does not permit.

An increased understanding of household dynamics can also come from replication of similar studies under more precise experimental conditions. Random selection of household members might give better insight into the reasons for the observed differences in response distributions among single- and dual- sampled households. Simultaneous attempts at objective measurements, easy in the case of routine tasks, along with varied structures of the self-assessment questions, might lead to a better understanding of how to effectively frame the question for more accurate results. A lesser reliance on modeling would add significant value to the use of such questions in a variety of surveys. All these observations are suggestions for further extensions and study.

References

Becker, Gary. 1991. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income Within Households." *Quarterly Journal of Economics* 130(2): 571–614.

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

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

Cunningham, Mick. 2001. "Parental Influences on the Gendered Division of Housework." *American Sociological Review* 66 (2): 184–203.

Dobbelsteen, Simone, and Peter Kooreman. 1997. "Financial Management, Bargaining, and Efficiency Within the Household: An Empirical Analysis." *De Economist* 145 (3): 345–366.

Efron, Bradley, and David V. Hinkley. 1978. "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information." *Biometrika* 65 (3): 457–487.

Fowler, Floyd J. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.

Gupta, Sanjiv. 2007. "Autonomy, Dependence, or Display? The Relationship Between Married Women's Earnings and Housework." *Journal of Marriage and Family* 69 (2): 399–417.

Huston, Ted L., and Robert L. Burgess. 1979. "Social Exchange in Developing Relationships." In *Social Exchange in Developing Relationships*, eds. Ted L. Huston and Robert L. Burgess, 3–28. New York: Academic Press.

Kamo, Yoshinori. 2000. "He Said, She Said: Assessing Discrepancies in Husbands' and Wives' Reports on the Division of Household Labor." *Social Science Research* 29 (4): 459–476.

Kullback, Solomon, and Richard Leibler. 1951. "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22 (1): 60–62.

Lam, Chun Bun, Susan M. McHale, and Ann C. Crouter. 2012. "The Division of Household Labor: Longitudinal Changes and Within-Couple Variation." *Journal of Marriage and Family* 74(5): 944–952.

Lundeber, Mary A, Paul W Fox, and Judith Punccohar. 1994. "Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments." *Journal of Educational Psychology* 86(1): 114–121.

Mannino, Ciela Anna, and Francine M. Deutsch. 2007. "Changing the Division of Household Labor: A Negotiated Process Between Partners." *Sex Roles* 56 (5): 309–324.

Marini, Margaret Mooney, and Beth Ann Shelton. 1993. "Measuring Household Work: Recent Experience in the United States." *Social Science Research* 22 (4): 361–382.

McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. Boca Raton, FL: Chapman and Hall.

Mizan, Ainon N. 1994. "Family Power Studies: Some Major Methodological Issues." *International Journal of Sociology of the Family* 24 (2): 841–854.

Moore, Don A, and Paul J Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115(2): 502–517.

Parkman, Allen M. 2004. "Bargaining Over Housework." *American Journal of Economics and Sociology* 63(4): 765–794.

Press, Julie, and Eleanor Townsley. 1998. "Wives' and Husbands' Housework Reporting: Gender, Class, and Social Desirability." *Gender and Society* 12 (2): 188–218.

Sayer, Liana C., Philip N. Cohen, and Lynne M. Casper. 2004. "Women, Men, and Work." In *The American People*, eds. R. Farley and J. Haaga, 76–106. New York: Sage.

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

Schuh, Scott, and Joanna Stavins. 2010. "Why Are (Some) Consumers (Finally) Writing Fewer Checks? The Role of Payment Characteristics." *Journal of Banking and Finance* 34(8): 1745–1758.

Schuh, Scott, and Joanna Stavins. 2014. "The 2011 and 2012 Surveys of Consumer Payment Choice." Research Data Report 14–1. Boston: Federal Reserve Bank of Boston.

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

Wilks, Samuel S. 1938. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses." *The Annals of Mathematical Statistics* 22(1): 79–86.

Yates, Frank. 1934. "The Analysis of Multiple Classification with Unequal Numbers in the Different Classes." *The Journal of the American Statistical Association* 29 (185): 51–66.

Household Type $t[h]$	**Individual Type** i	Gender	Income	**# from Single-Sampled Households**	**# from Dual-Sampled Households**	**Total**
1	1	Male	1st	304	87	760
	2	Female	2nd	282	87	
2	1	Male	Tied	107	63	394
	2	Female	Tied	161	63	
3	1	Male	2nd	76	35	335
	2	Female	1st	189	35	

Table 1: 2012 SCPC Sample Sizes by Household and Member Type

Note: "1st" means highest earner, "2nd" means second highest earner, and "Tied" means both adults earn roughly the same amount.

	Sum of Ratings: $R_{h1} + R_{h2}$				
Sum	2	3	4	5	6
Paying Bills	0.0	4.9	76.2	15.1	3.8
Shopping	0.5	7.6	65.9	20.0	5.9
Saving and Investments	1.1	14.6	54.0	26.5	3.8
Other Decisions	1.1	14.6	55.7	24.9	3.8

Table 2: Distribution of Aggregate Ratings in Dual-Sample Households

Note: Numbers in table represent the percentage of households.

	Household-Consistent Parameters	Response-Error Parameters	Total # of Parameters
Model A	$\mu_h = \mu(t[h]), c_h = c(t[h]), \alpha(t[h])$	$\tau_h = \tau(t[h], s[h]), \sigma_{hi} = \sigma(t[h], i)$	7
Model B		$\tau_h = \tau(t[h]), \sigma_{hi} = \sigma(t[h], s[h], i)$	8

Table 3: Two Model Specifications

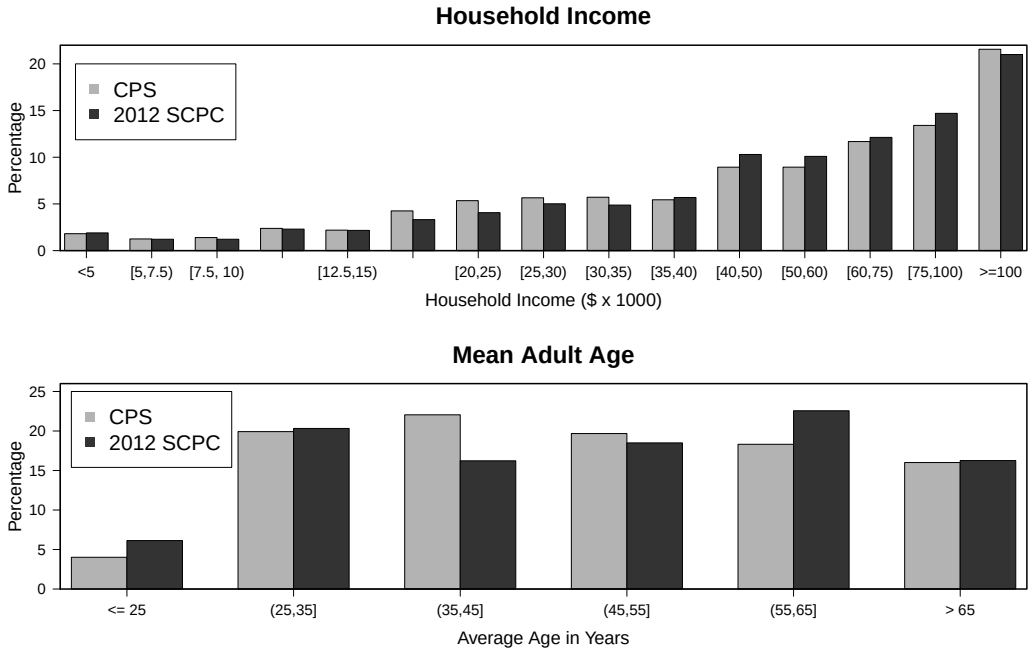


Figure 1: Distribution of Household Income and Age of Adults in Two-Adult Households in the 2012 SCPC and the 2012 Current Population Survey

Source: Author's calculations.

First, help us to understand **your** role in the financial activity of your household.

In your household, how much responsibility do **you** have for these tasks?

- *Check one per row only.*

	None or almost none	Some	Shared equally with other household members	Most	All or almost all
Paying **monthly bills** (rent or mortgage, utilities, cell phone, etc)	○	○	○	○	○
Doing regular **shopping** for the household (groceries, household supplies, pharmacy, etc)	○	○	○	○	○
Making decisions about **saving and investments** (whether to save, how much to save, where to invest, how much to borrow)	○	○	○	○	○
Making decisions about **other household financial matters** (where to bank, what payment methods to use, setting up online bill payments, filing taxes)	○	○	○	○	○

<<Back Next>>

Figure 2: Screenshot of the Financial Responsibility Questions from the 2012 SCPC

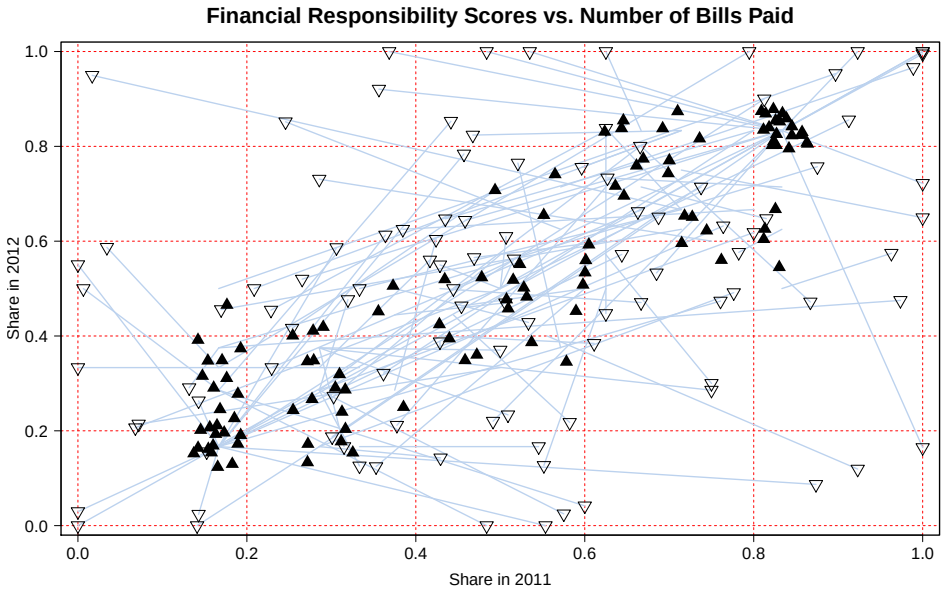


Figure 3: Two Measures of Share of Responsibility for Paying Bills Among Dual-Sampled Households who Participated in the 2011 and 2012 SCPC

Note: Shaded triangles represent an estimate based on responses to the financial responsibility question, and unshaded triangles represent an estimate based on the reported number of monthly bills paid. Lines connect responses from the same household.

Source: Author's calculations.

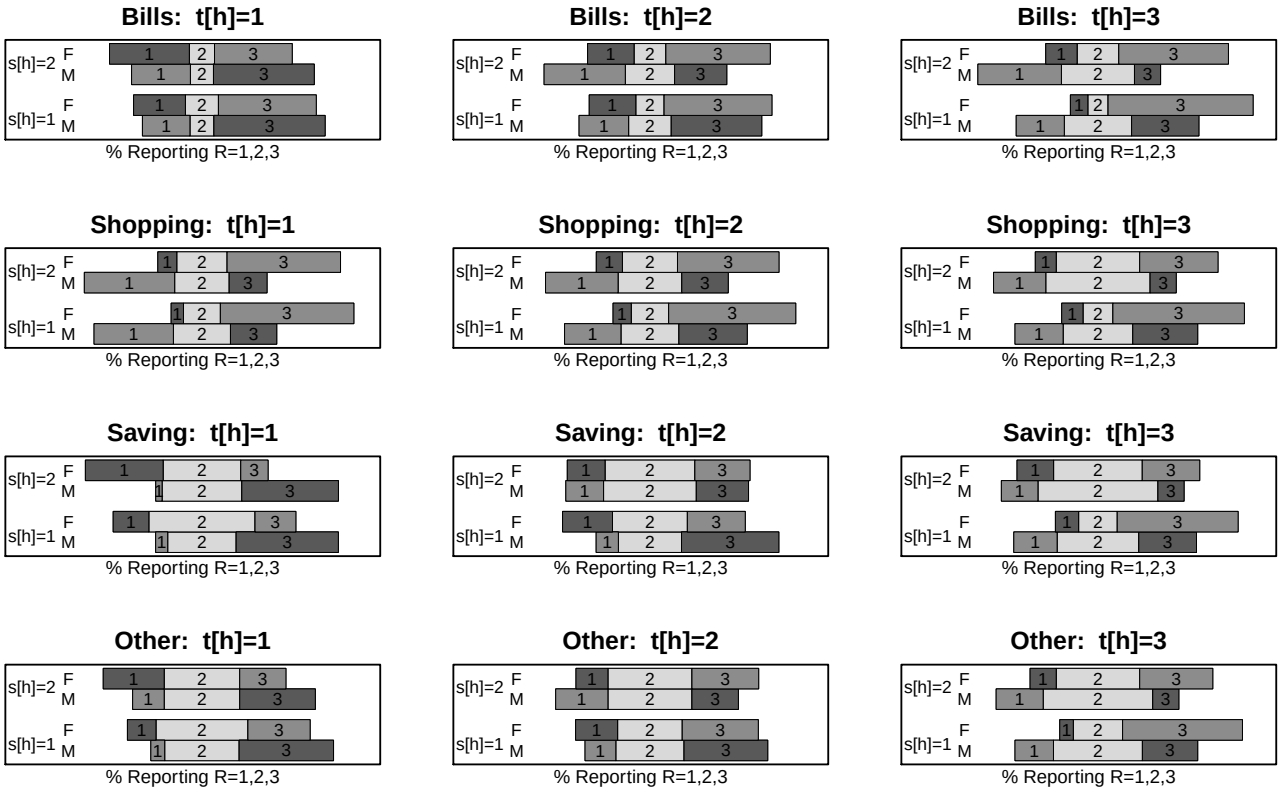


Figure 4: Distribution of Ratings in the 2012 SCPC by Household and Member Type among Single-Sampled and Dual-Sampled Households

Source: Author's calculations.

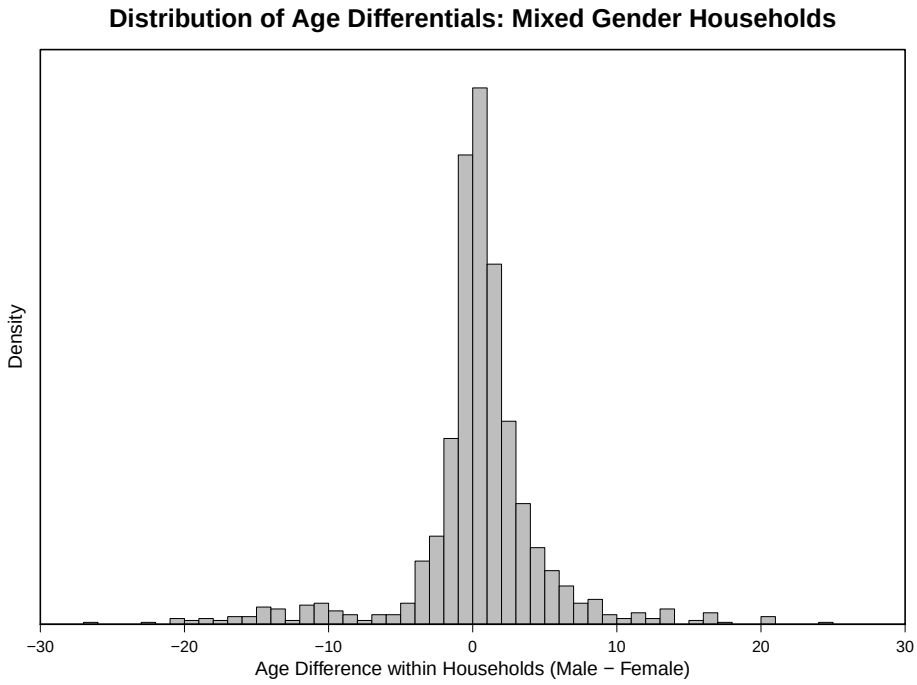


Figure 5: Distribution of Age Differences for Mixed-Gender, Two-Adult Households in the 2012 SCPC
Source: Author's calculations.

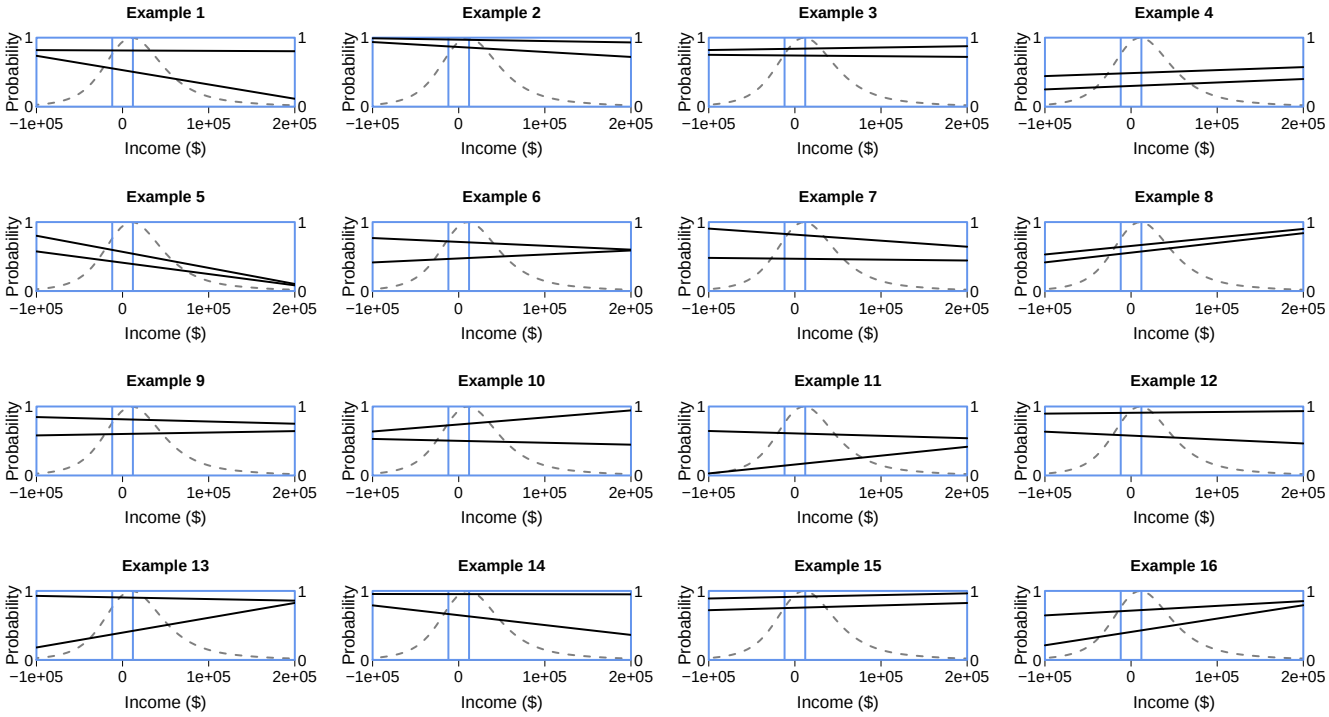


Figure 6: Sixteen Examples of Household Income Models

Note: The dashed gray lines represents the distribution of household incomes among two-adult, mixed-gender households. The vertical blue lines in each plot divide the population of households into three groups: those in which the male has a lower income (left), those in which incomes are roughly equal (middle), and those in which the female has a higher income (right). The probabilities of household consistent ratings as a function of income difference, as in (4), are depicted by black lines.

Source: Author's calculations.

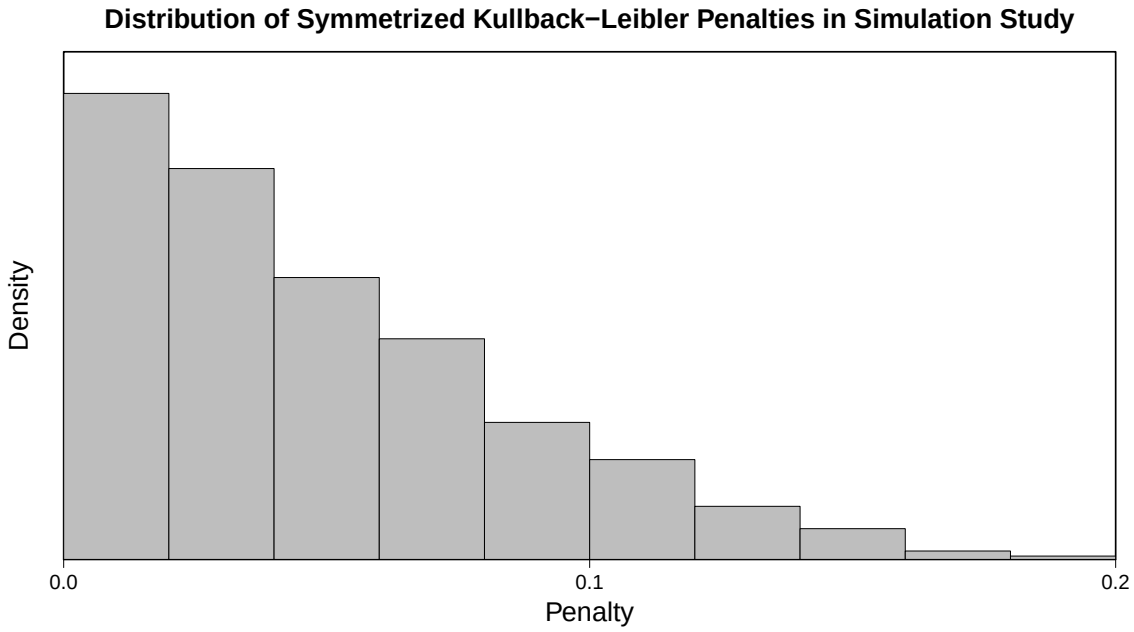


Figure 7: Distribution of Symmetrized Kullback-Leibler Penalties for Simulation Studies

Source: Author's calculations.

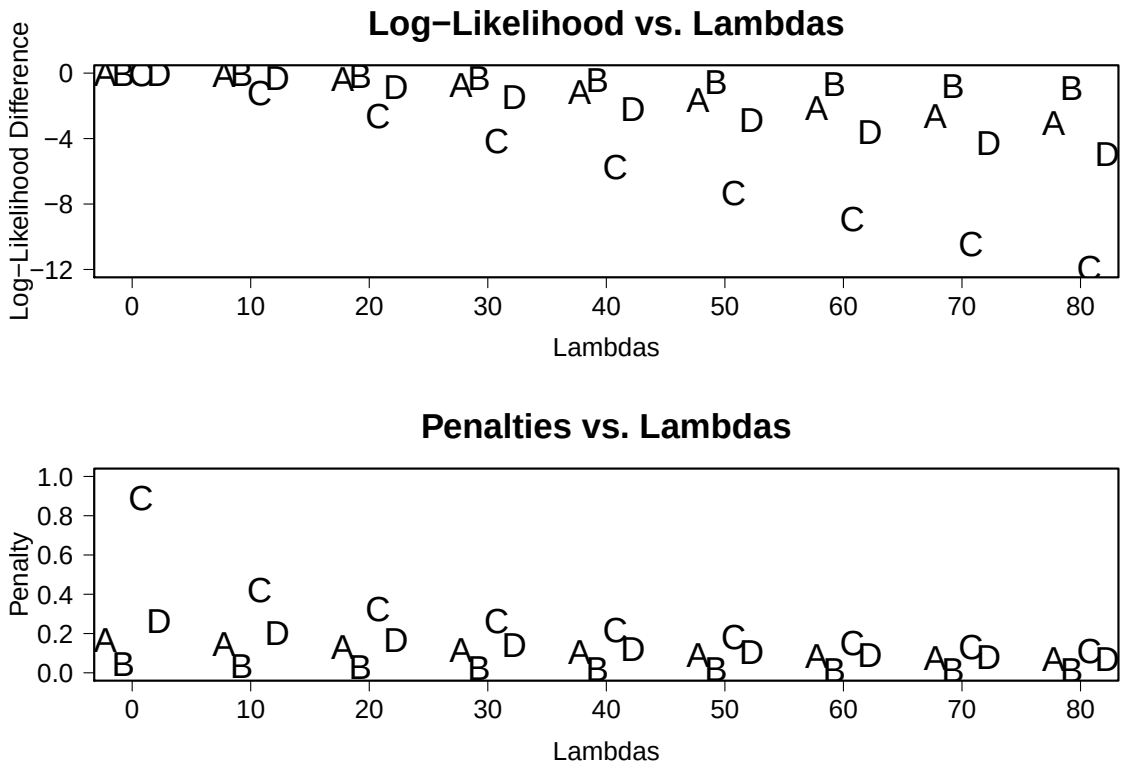


Figure 8: Effects of λ on Log-Likelihoods and Penalty Terms

Note: The activities are coded as: "A"= paying bills, "B"=household shopping, "C"=decisions about saving and investments, "D"= other financial decisions. Reported log-likelihoods are subtracted from the log-likelihood calculated at $\lambda = 0$.

Source: Author's calculations.

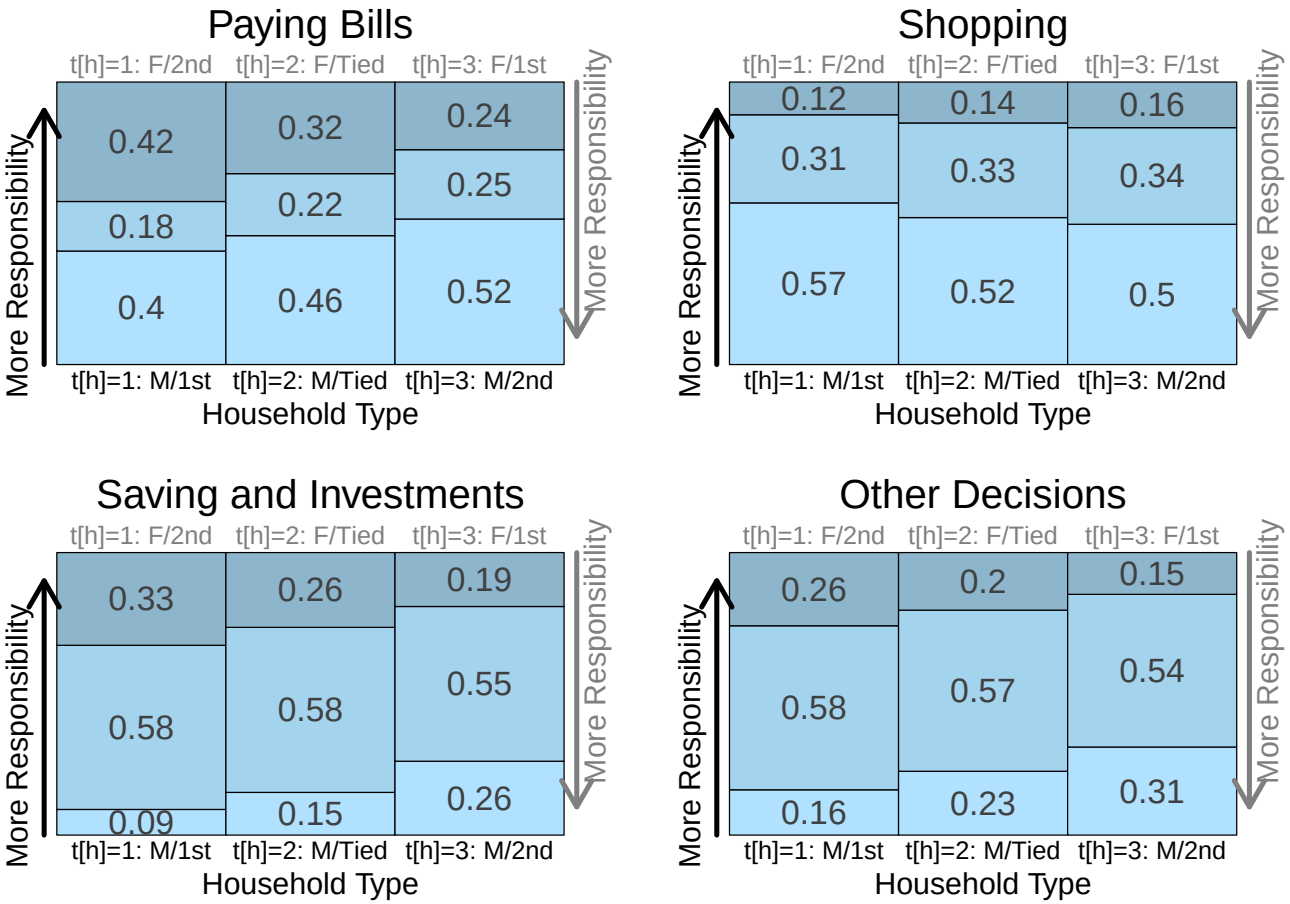


Figure 9: Expected Distribution of Household-Consistent Ratings for Each Household Type

Note: Moving from top to bottom corresponds to increasing levels of responsibility for the household member types listed along the bottom axis (and thus correspond to decreasing levels of responsibility for the opposite household member shown on the upper axis). For example, in mixed-gender households in which the male has the higher income, the probabilities of the male having a minority of, an equal share of, and a majority of responsibility for paying bills are 0.40, 0.18, and 0.42, respectively.

Source: Author's calculations.

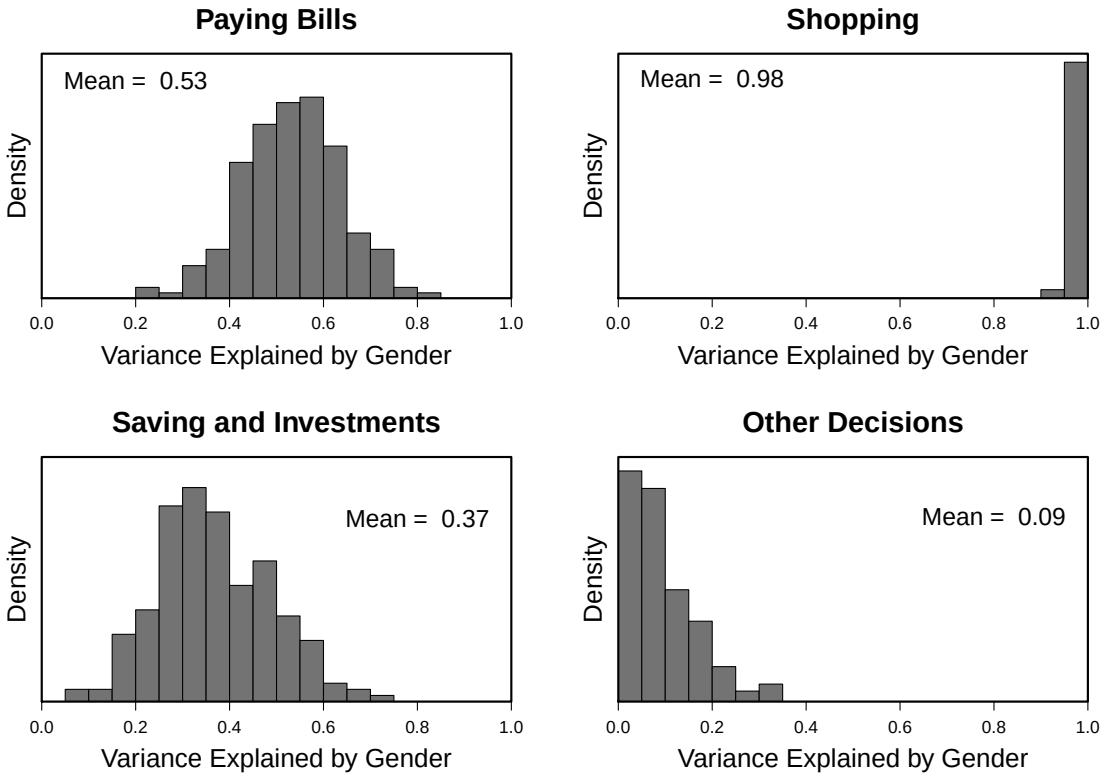


Figure 10: Proportion of Variance Explained by Gender and Income Ranking
Source: Author's calculations.

A Appendix A

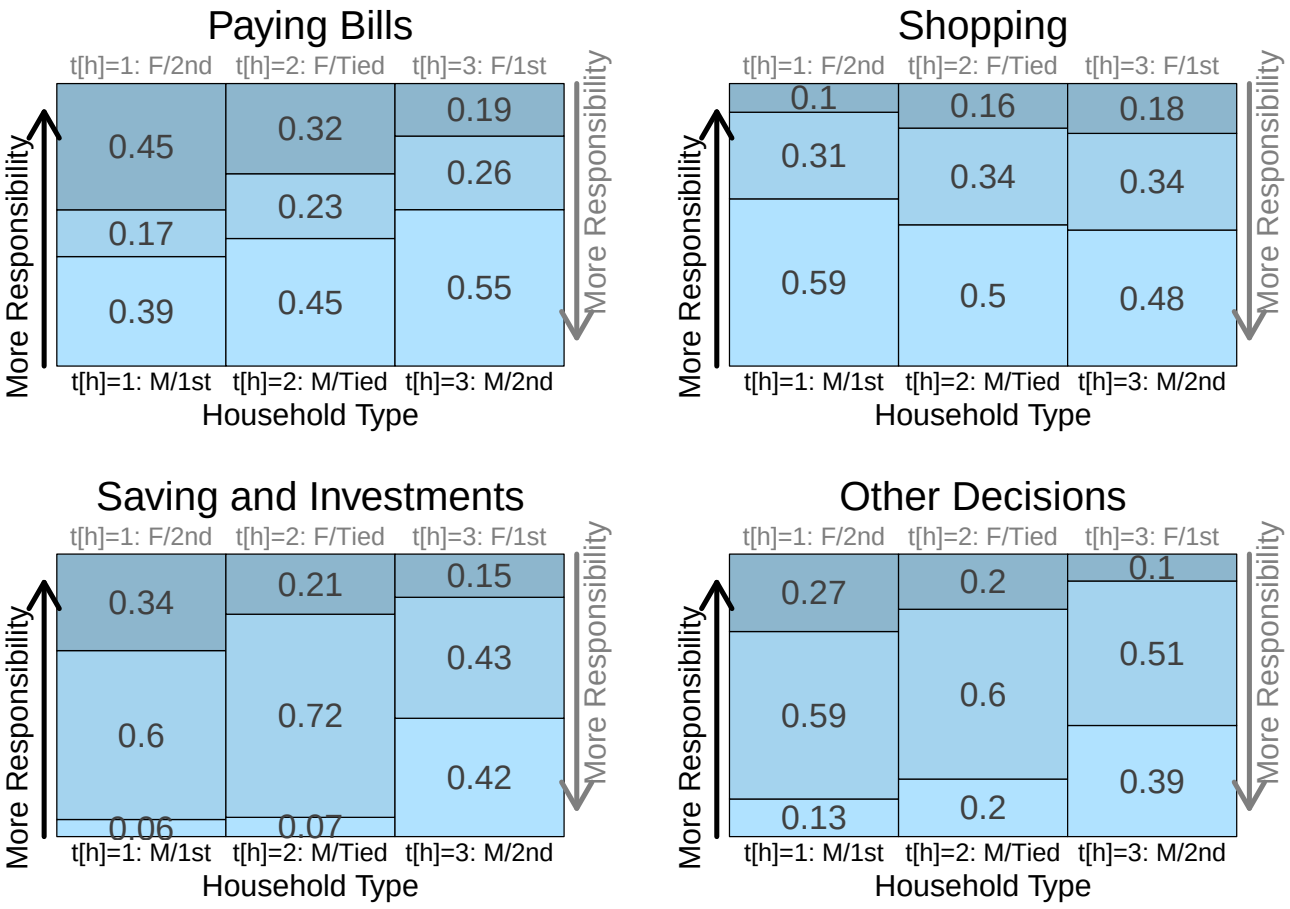
The estimated distributions of responsibility in Figure 9 were fitted using a penalty term of $\lambda = 70$. Comparable estimates with no penalty term ($\lambda = 0$) are shown in Appendix Figure 1. Except in the case of saving and investment decisions, the differences between the estimates are not very big. The latent variable model alone, with no penalty, will yield consistent estimators when the data-creation process is specified correctly. If the data-creation process is sufficiently different from what the model allows, the estimated responsibility distributions may be erroneous. As discussed in the paper, the parametric model is unlikely to be sufficient in adequately capturing the response process and household consistent ratings are heavily influenced by the consistency of dual-sampled households, for which sample sizes are not that large. I believe these are more likely explanations for the drastic differences observed for saving and investment decisions in Figure 1 and, thus, emphasize smooth transitions and enforce a nonzero penalty term.

B Appendix B

Appendix Figure 2 shows results comparable to those in Figure 9, but based on parameter estimates that use only dual-sampled households. The penalty term is similarly fixed at $\lambda = 70$. The differences between the two estimates are not substantial, which is not surprising seeing as the dually-sampled households have a big influence on the model fit.

C Appendix C

Appendix Figure 3 shows uncertainty associated with the point estimates of p_{tj} shown in Figure 9 for each of the four activities and three household types.

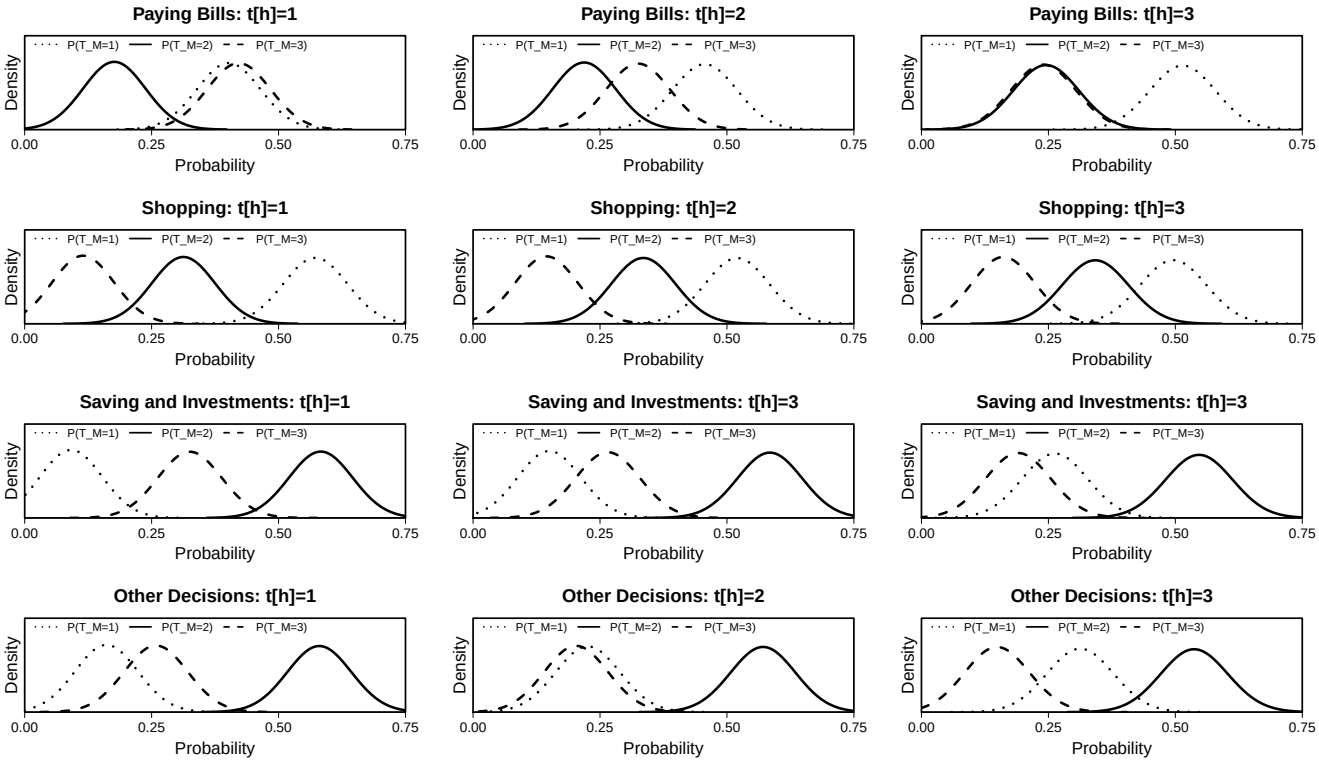


Appendix Figure 1: Expected Distribution of Household-Consistent Ratings for Each Household Type Estimated with No Penalty ($\lambda = 0$)

Source: Author's calculations.

Appendix Figure 2: Expected Distribution of Household-Consistent Ratings for Each Household Type Estimated Using Dual-Sampled Households Only

Source: Author's calculations.



Appendix Figure 3: Distribution of Rating Shares p_t for Each Activity and Household Type
Note: Distributions were generated by smoothing 2,000 draws from the posterior distributions $(\mu_{t[h]}, c_{t[h]})$.
Source: Author's calculations.