Wife vs. Husband: How Can Differences in Social Class Identity Identify Poor Quality Data?

Adan Silverio Murillo

School of Public Affairs, American University

May 17, 2018

Outline

Motivation

- Asset information and household characteristics are frequently used to conduct empirical research and to guide public policy.
- When using these variables, practitioners assume that they are less susceptible to misreporting.

Research Question

Is the data collected from surveys regarding assets and home services free of misreporting?

Data

I use data for poor households participating in Mexico's PROGRESA program.

- Questions regarding the possession of eighteen goods and eight household services were asked to both the wife and the husband, separately, in a random sample of 900 households.

Results

(1) Discrepancies in the information reported between the spouses.

- When asked about the possession of a washing machine, the information reported did not coincide in 24% of the households.

Results

(2) This result has consequences for the targeting of social programs that use proxy means test.

- 10.5% of the households would be classified as non-poor if asked to the husband, but as poor if asked to the wife.

Results

(3) The difference in the information reported between the spouses is partially explained by differences in their self-identification of social class.

- When one of the spouses self-identify with a higher social class, with respect to the social class reported by the other spouse, then he or she reports more goods and services than those reported by the other spouse.

- To address the problem of omitted variable bias, a bounding technique is implemented: Oster (2016).

How accurate is the information collected through surveys?

- Philipson and Malani (1999) point out that economists pay much more attention to the consumption of data than to the production of data.

		Does the respondent think she knows the truth?	
		Yes : $p = 0$ or 1	*No* : $p \in (0, 1)$
Does the respondent want to tell the truth?	*Yes* : $u(i \mid i) \geq u(j \mid i)$	No Problem	Knowledge Problem
	No : $u(j \mid i) \geq u(i \mid i)$	Incentive Problem	Both Problems

Fig. 1. Sources of erroneous reporting.

There have been some efforts to identify data abnormalities:

- Judge and Schechter (2009) proposed that Benford's law can be used.
- The idea behind Benfords law is that, in large data sets, numbers with a first digit of 1 are observed more often than those starting with 2, and so on.
- They find that data from developing countries are of poor quality, and data from USA are of better quality.

How accurate is the information collected through surveys?

- Meyer and Mitag (2015) find that survey data (Current Population Survey) understate the income of poor households.
- Courtemanche et al. (2017), using data from the National Longitudinal Survey of Youth- 1979, find evidence of misreporting and overreporting of participation in the Supplemental Nutrition Assistance Program (SNAP).
- Problem: Benford's law cannot be applied to dummy or categorical variables.

Why do individuals not report the truth?

- Individuals can consider the potential benefits (such as access to a social program) of cheating (underreporting information) and the cost of cheating (probability of being discovered and the potential penalties).

- Mazar and Ariely (2006) propose that in addition to the external reward mechanisms (cost-benefit analysis), there are internal reward mechanisms that affect the decisions of individuals regarding cheating.

Why do individuals not report the truth?

- Self-identity
- Self-control
- Self-esteem
- Religiosity
- Anxiety

What other aspects do individuals consider in addition to material payoffs?

Self-identity

- I hypothesize that individuals that participate in the social program PROGRESA and self-identify as middle class or above will tend to overreport the possession of assets and household services in order to behave according to their self-identity.

Data

- Poor households participating in the Program PROGRESA.
- A random sample of 900 households was selected.
- Questions regarding the possession of eighteen goods and eight household services were asked to the wife and the husband in separate interviews.
- The survey collected information on non-cognitive skills and socio-economic information.

Data

Regarding how the data were collected:

- Both interviews (of the wife and of the husband) were conducted in the household.
- The interviews were conducted in the name of a private University.

What is the quality of the PROGRESA data regarding the possession of assets?

Table: **Descriptive Statistics (goods)**

	Husband's report of possession of (%):	Wife's report of possession of (%):
Music device	59.7	57.5
Bicycle	42.7	36.5
Farm animals	29.6	31.0
Washing Machine	42.7	43.1
Gas stove	20.0	21.8
Refrigerator	63.5	65.1
Living room	23.7	23.0
Automobile	19.4	16.2
Landline	15.7	16.6
Photographic camera	8.4	6.6
Other land (apart from home)	8.3	6.5
Television	90.9	90.6
Machinery or work equipment	5.5	3.0
House, apartment or room to rent	4.3	3.7
Motorcycle	5.2	5.5
Savings	1.9	3.6
Local business	2.7	2.8
Canoe or boat	2.0	1.7

What is the quality of the PROGRESA data regarding the possession of assets?

Table: **Descriptive Statistics (goods)**

	Husband's report of possession of (%):	Wife's report of possession of (%):	Percentage that do not match (%)	Husband: Yes Wife: No	Husband: No Wife: Yes
Music device	59.7	57.5	32.9	17.5	15.4
Bicycle	42.7	36.5	30.2	18.3	11.9
Farm animals	29.6	31.0	25.6	12.1	13.5
Washing Machine	42.7	43.1	24.0	11.8	12.2
Gas stove	20.0	21.8	22.9	10.7	12.2
Refrigerator	63.5	65.1	21.3	9.8	11.5
Living room	23.7	23.0	18.7	9.7	9.0
Automobile	19.4	16.2	13.7	8.5	5.2
Landline	15.7	16.6	13.5	6.3	7.2
Photographic camera	8.4	6.6	10.9	6.4	4.5
Other land (apart from home)	8.3	6.5	10.7	6.3	4.4
Television	90.9	90.6	8.4	4.4	4.0
Machinery or work equipment	5.5	3.0	7.0	4.8	2.2
House, apartment or room to rent	4.3	3.7	7.0	3.8	3.2
Motorcycle	5.2	5.5	5.6	2.6	3.0
Savings	1.9	3.6	4.9	1.5	3.4
Local business	2.7	2.8	3.6	1.7	1.9
Canoe or boat	2.0	1.7	2.2	1.2	1.0

What is the quality of the PROGRESA data regarding the possession of household services?

Table: **Descriptive Statistics (services)**

	Husband's report of possession of (%):	Wife's report of possession of (%):	Percentage that do not match (%)	Husband: Yes Wife: No	Husband: No Wife: Yes
Toilet	74.4	75.3	24.5	11.7	12.8
Kitchen	76.9	78.4	23.8	11.2	12.6
Drainage	53.6	49.2	16.8	10.6	6.2
Gas	81.0	82.3	15.9	7.3	8.6
Hot water	16.0	14.6	15.6	8.6	7.0
Bath shower	23.2	19.0	15.4	9.9	5.5
Piped water	80.5	82.3	15.2	6.7	8.5
Electric light	95.5	96.9	4.9	1.7	3.2

Figure 1 presents the differences in goods and household services reported by the husbands and the wives.

- It is observed that only in 20% of the households the information reported about the total number of goods and household services coincide.

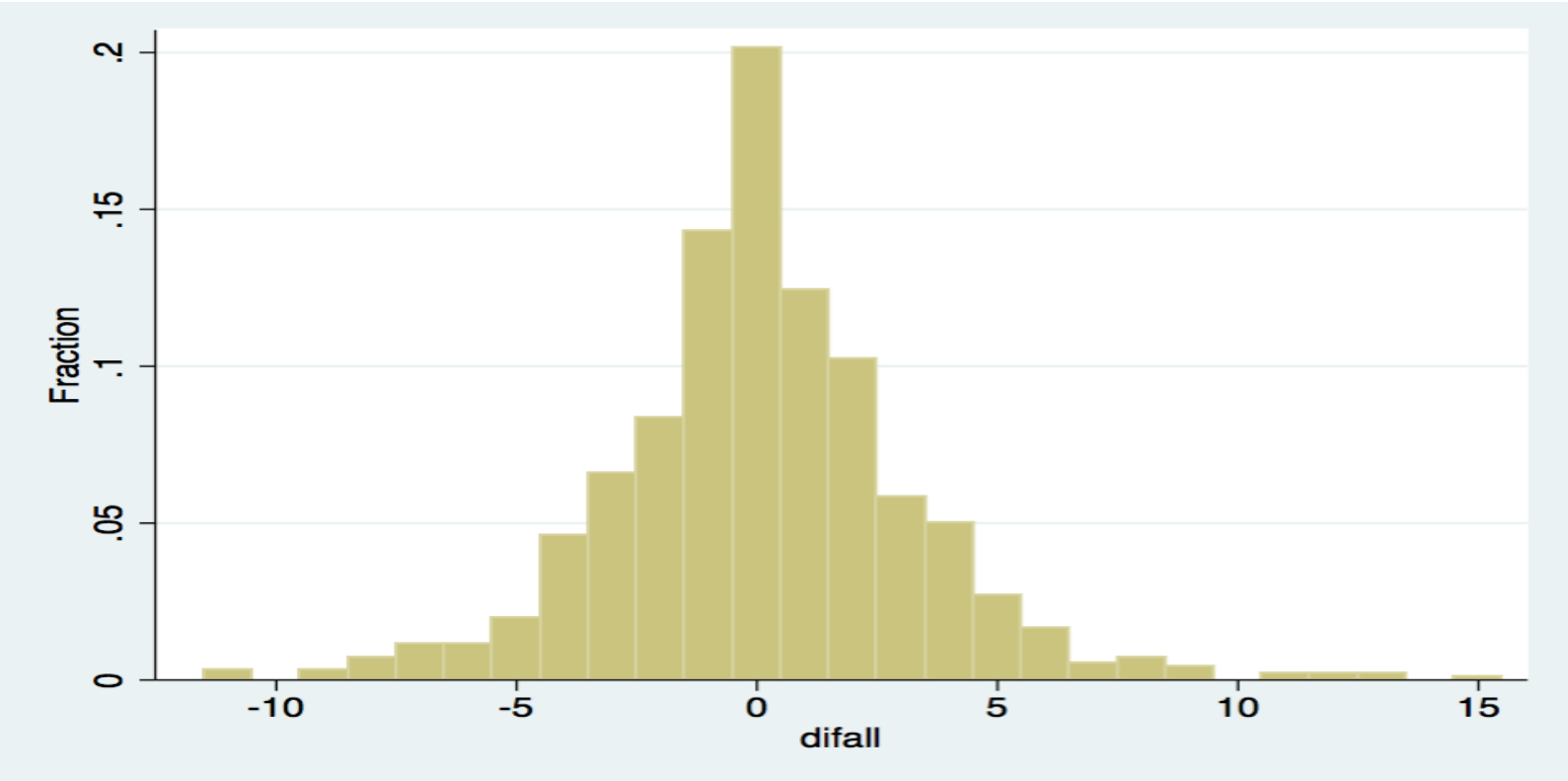

Table: **Social Class Self-identification**

		Wife		
		Poor	Medium class	Rich
Husband	Poor	253	198	8
	Medium class	141	259	16
	Rich	4	9	0

Identification Strategy

The model to estimate is given by:

$$Y_j^h - A_j^r = \beta_{1h} T_{hj} + \gamma_h X_{hj} + e_{hj}$$

$$Y_j^w - A_j^r = \beta_{1w} T_{wj} + \gamma_w X_{wj} + e_{wj}$$

Where:

- Y_j^i is an index adding goods and services reported by individual i in house j.
- A_j^r is and index adding the real number of goods and services within the household.
- T_{ij} is the self-identity of the individual i in house j.
- X is a vector of observed control variables.

Identification Strategy

I cannot observe A_j^r, thus I use the following specification:

$$Y_j^h - A_j^r - Y_j^w + A_j^r = \beta_{1h} T_{hj} - \beta_{1w} T_{wj} + \gamma_h X_{hj} - \gamma_w X_{wj} + e_{hj} - e_{wj}$$

$$Y_j^h - Y_j^w = \beta_{1h} T_{hj} - \beta_{1w} T_{wj} + \gamma_h X_{hj} - \gamma_w X_{wj} + e_{hj} - e_{wj}$$

Endogeneity Challenges

- Measurement error. The measures of self-identity is a proxy variables, so there is a potential problem of measurement error.
- Omitted variables. It is possible that self-identity is correlated with other psychological variables that affect the outcome of interest but are not present in the data.

Estimation Challenges

To address the challenges of endogeneity:

- First, I add control variables on the right hand side in order to see if such additions do not affect the coefficient of interest.
- Second, I use a recently developed bounding methodology: Oster (2016).

Table: **OLS Estimates: Effects of Social Class Self-identification on the Index of Differences in Goods and Household Services**

Dep Var: Index of differences in Assets and Services	(1)	(2)	(3)
Social Class Self-Identification (Husband)	0.177***	0.169***	0.144***
	(0.044)	(0.041)	(0.040)
Social Class Self-Identification (Wife)	0.132***	0.110***	0.097***
	(0.044)	(0.034)	(0.033)
State Fixed Effects	No	Yes	No
Municipality Fixed Effects	No	No	Yes
R^2	0.06	0.12	0.22
Observations	847	847	847

*** p <0.01, ** p <0.05, *p <0.1

Clustered standard errors displayed in parenthesis at the municipality level.

The control variables are: age, years of school,impulsiveness, self-esteem, anxiety, religiosity, mental illness, error in the number of dependents for both the wife and the husband.

Table: **Bounding Methodology: Effects of Social Class Self-identification on the Index of Differences in Goods and Household Services**

	(1) Oster (2016)	(2) Gonzalez and Miguel (2015)	(3) ($R_{max} = 1$)
	Panel A :	$0 \leq \delta \leq 1$	
Social Class Self-Identification (Husband)	[0.144, 0.146]	[0.144, 0.150]	[0.144, 0.168]
Social Class Self-Identification (Wife)	[0.097, 0.097]	[0.097, 0.098]	[0.097, 0.101]
	Panel B :	$-1 \leq \delta \leq 0$	
Social Class Self-Identification (Husband)	[0.142, 0.144]	[0.138, 0.144]	[0.120, 0.144]
Social Class Self-Identification (Wife)	[0.096, 0.097]	[0.095, 0.097]	[0.092, 0.097]

The control variables are: age, years of school,impulsiveness, self-esteem, anxiety, religiosity, mental illness, error in the number of dependents for both the wife and the husband.

Robustness Checks

- Problems in the ownership of the assets.
- Assortative mating
- Problems of information (husband did not know household participates in PROGRESA)
- Cohabiting vs. married couples

Concluding remarks

- I find important discrepancies in the information reported between the spouses.
- These discrepancies are partially explained by differences in their perception of social class.
- Researchers and policy makers use variables such as the possession of assets to proxy the real income of the households. It is assumed that these variables are less susceptible to misreporting. Yet, this paper presents evidence that contradicts this assumption.