kmr: A Command to Correct Survey Weights for Unit Nonresponse using Group's Response Rates

Ercio Munoz
CUNY Graduate Center and Stone Center on Socio-Economic Inequality

Stata Conference Chicago 2019

July 12, 2019

Motivation

Bias in inequality measures due to unit nonresponse

- There is evidence that income systematically affects survey response. For example, Bollinger et al. (2019, JPE) links internal CPS data to Social Security admin. data to show that nonresponse across the earnings distribution is U-shaped.

- Korinek et al. (2007, J. Econometrics) proposed a method to correct for unit nonresponse bias using response rates by region. Advantages: 1) It does not assume ignorability within the smallest unit and 2) relies solely on data from the survey.

- The method has been recently used with data from Egypt and EU (see Hlasny & Verne 2018a, 2018b).

Bias in inequality measures due to unit nonresponse

- There is evidence that income systematically affects survey response. For example, Bollinger et al. (2019, JPE) links internal CPS data to Social Security admin. data to show that nonresponse across the earnings distribution is U-shaped.

- Korinek et al. (2007, J. Econometrics) proposed a method to correct for unit nonresponse bias using response rates by region. Advantages: 1) It does not assume ignorability within the smallest unit and 2) relies solely on data from the survey.

- The method has been recently used with data from Egypt and EU (see Hlasny & Verne 2018a, 2018b).

Bias in inequality measures due to unit nonresponse

- There is evidence that income systematically affects survey response. For example, Bollinger et al. (2019, JPE) links internal CPS data to Social Security admin. data to show that nonresponse across the earnings distribution is U-shaped.

- Korinek et al. (2007, J. Econometrics) proposed a method to correct for unit nonresponse bias using response rates by region. Advantages: 1) It does not assume ignorability within the smallest unit and 2) relies solely on data from the survey.

- The method has been recently used with data from Egypt and EU (see Hlasny & Verne 2018a, 2018b).

This presentation

- Briefly describe the econometric method to correct for unit non-response bias suggested by Korinek et al. (2007), which estimates a micro compliance function that can be used to re-weight the survey.

- Introduce a Stata command (kmr) to implement this method (Morelli and Munoz, 2019a).

- Show the command in use with an empirical example: Inequality, total income, and poverty rate in the US estimated with the CPS correcting for unit non-response (Morelli and Munoz, 2019b).

This presentation

- Briefly describe the econometric method to correct for unit non-response bias suggested by Korinek et al. (2007), which estimates a micro compliance function that can be used to re-weight the survey.

- Introduce a Stata command (kmr) to implement this method (Morelli and Munoz, 2019a).

- Show the command in use with an empirical example: Inequality, total income, and poverty rate in the US estimated with the CPS correcting for unit non-response (Morelli and Munoz, 2019b).

This presentation

- Briefly describe the econometric method to correct for unit non-response bias suggested by Korinek et al. (2007), which estimates a micro compliance function that can be used to re-weight the survey.

- Introduce a Stata command (kmr) to implement this method (Morelli and Munoz, 2019a).

- Show the command in use with an empirical example: Inequality, total income, and poverty rate in the US estimated with the CPS correcting for unit non-response (Morelli and Munoz, 2019b).

Methodology

Intuition: 3x3 model of selective compliance

Assumption: Response does not change across regions and depends on income.
By household income, the number of answers should equal the total number of households sampled multiplied by the probability of response:

Region	Income	Sampled	Answers	Probability
1	20K	30	30	1
1	30K	30	15	1/2
2	20K	30	30	1
2	30K	30	15	1/2
2	100K	30	3	1/10
3	20K	30	30	1

However, we do not know the total number of households sampled and the probability of response by income.

Intuition

What we do know is the total number of household sampled by region, and we can use it to solve for P_i:

Region	Income	Answers	Sampled by region	Probability
1	20K	30	60	P_{20K}
1	30K	15	60	P_{30K}
2	20K	30	90	P_{20K}
2	30K	15	90	P_{30K}
2	100K	3	90	P_{100K}
3	20K	30	30	P_{20K}

Region	Answers	Sampled by region
1	$30/P_{20K} + 15/P_{30K}$	60
2	$30/P_{20K} + 15/P_{30K} + 3/P_{100K}$	90
3	$30/P_{20K}$	30

Generalization for I income groups and J geographic areas (I>J)

For each sampled household ϵ, there is a Bernoulli variable $D_{ij\epsilon}$ that equals 1 if the household response and 0 otherwise, and that the probability of response has a logistic form:

$$P(D_{ij\epsilon} = 1|X_i, \theta) = \frac{e^{X_i\theta}}{1 + e^{X_i\theta}} \tag{1}$$

Denote the mass of respondents as:

$$m_{ij}^1 = \int_0^{m_{ij}} D_{ij\epsilon} d\epsilon \tag{2}$$

with expected value:

$$E[m_{ij}^1] = m_{ij}P_i \qquad E[\frac{m_{ij}^1}{P_i}] = m_{ij} \tag{3}$$

Generalization for I income groups and J geographic areas (I>J)

Then the sum of all the ratios for a given region j:

$$\psi_{\mathrm{j}}(\theta) = \sum_{\mathrm{i}} \{ \frac{\mathrm{m}^{1}_{\mathrm{ij}}}{\mathrm{P_i}} - \mathrm{E}[\frac{\mathrm{m}^{1}_{\mathrm{ij}}}{\mathrm{P_i}}] \} = \sum_{\mathrm{i}} \frac{\mathrm{m}^{1}_{\mathrm{ij}}}{\mathrm{P_i}} - \mathrm{m_j} \tag{4}$$

Given that $\mathrm{E}[\psi_{\mathrm{j}}(\theta)] = 0$, we can stack J moment conditions $\psi_{\mathrm{j}}(\theta)$ into $\Psi(\theta)$, so:

$$\hat{\theta} = \mathrm{argmin}_{\theta} \Psi(\theta)' \mathrm{W}^{-1} \Psi(\theta) \tag{5}$$

Where W is a positive definite weighting matrix.

Syntax of the command

Syntax of the command

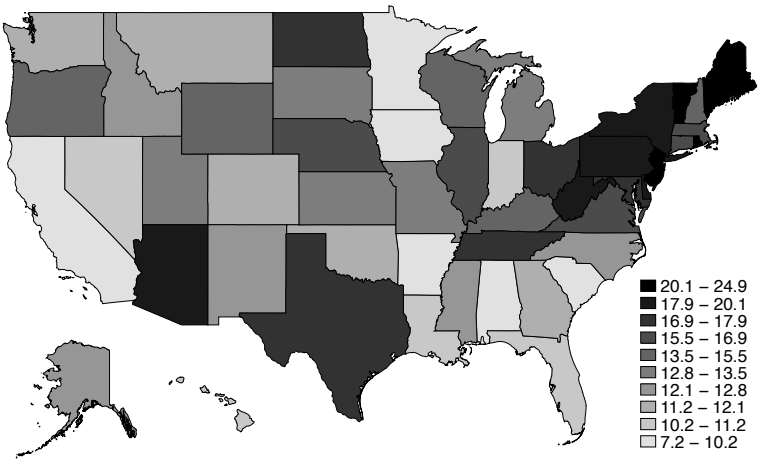
kmr [varlist] [if] [in], groups(varname) interview(varname) nonresponse(varname) options

where the options are:

- noconstant
- generate(newvarname)
- graph(varname)
- technique(string)
- start(num)
- maxiter(num)

Empirical example using the CPS

State-level variation in non-response rate in 2018



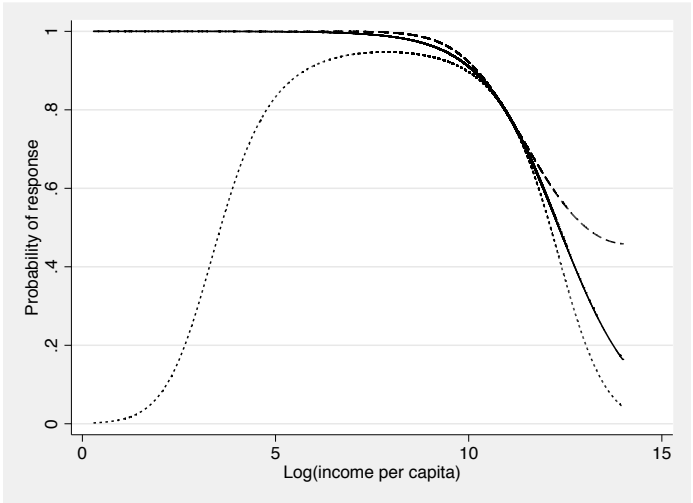
Estimates for 2018 - Gini goes from 46.5 to 50.5

Compliance function — Number of obs = 66899
AIC = 59.44
Number of groups = 51 — Schwarz = 56.8224

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ly	-.9866137	.2845151	-3.47	0.001	-1.544253	-.4289743
_cons	12.16839	3.116912	3.90	0.000	6.059354	18.27743

Generated by running:
kmr ly, groups(statefip) i(interview) n(typea)

Compliance function in 2018



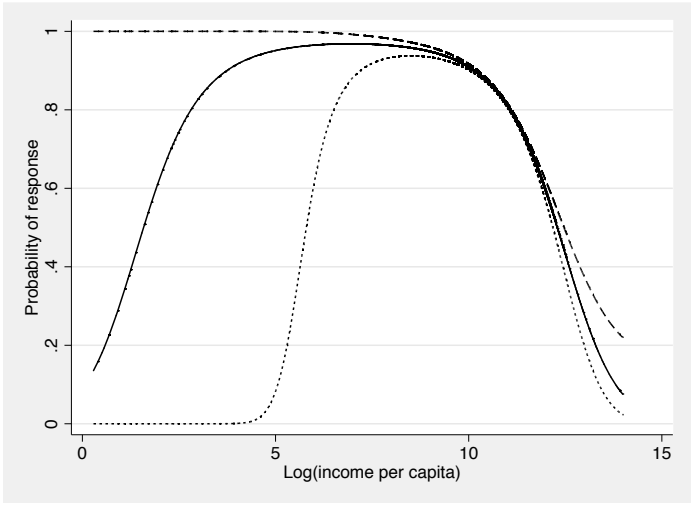
Estimates for 2018 - Gini goes from 46.5 to 53

Compliance function — Number of obs = 66899
AIC = 60.76
Number of groups = 51 — Schwarz = 58.0582

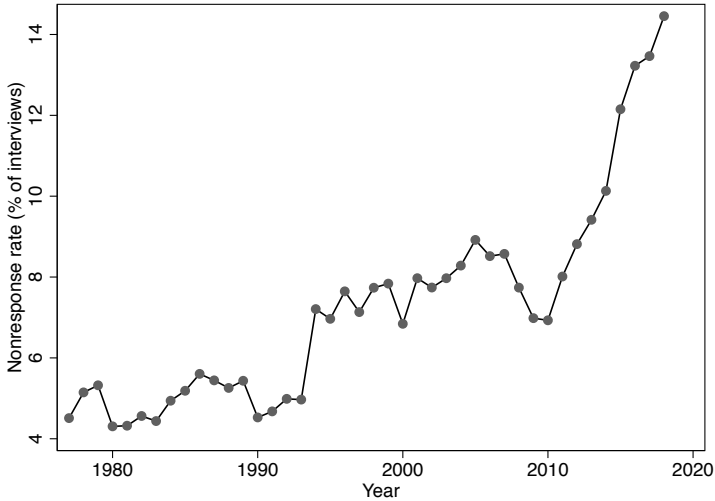
	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ly	1.653716	.9866511	1.68	0.094	-.2800843	3.587517
ly2	-.1191805	.046792	-2.55	0.011	-.2108912	-.0274699
_cons	-2.320002	5.306531	-0.44	0.662	-12.72061	8.080607

Generated by running:
kmr ly ly2, groups(statefip) i(interview) n(typea)

Compliance function in 2018 - Adding squared log of income

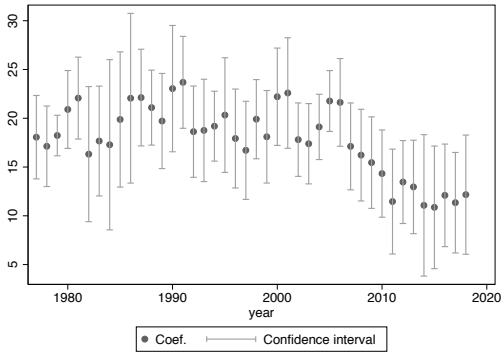


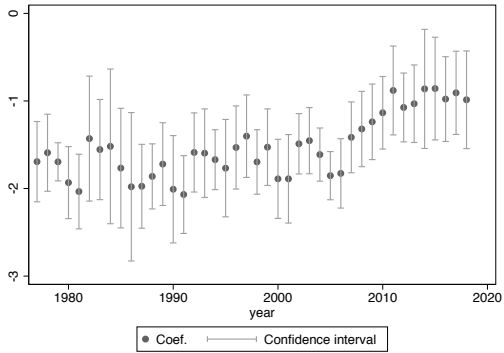
Aggregate non-response rate in the CPS on the rise



Source: Own elaboration using NBER CPS supplements.

Estimates over time $\{X_i\theta = \theta_1 + \theta_2 \log(y_i)\}$





Average correction across years 1977-2018

Table: Correction with respect to uncorrected grossed-up weights by state

Model	Gini	Top 1% income share	Total income	Poverty rate
Best model (7-year windows)	8.47%	40%	8.07%	-8.07%
$X_i\theta = \theta_1 + \theta_2 \log(y_i) + \theta_3 \log(y_i)^2$ - y/y	8.09%	36.71%	8.66%	-8.66%
$X_i\theta = \theta_1 + \theta_2 \log(y_i) + \theta_3 \log(y_i)^2$ - pooled	11.64%	52.54%	11.40%	-4.88%
$X_i\theta = \theta_1 + \theta_2 \log(y_i)$ - pooled	11.81%	46.8%	16.60%	-13.60%

References

- Bollinger, C., B. Hirsch, C. Hokayem, and J. Ziliak (2019), "Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch". Forthcoming at Journal of Political Economy.
- Hlasny and Verme (2018a), "Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods using the EU SILC Data." Econometrics 6(30)
- Hlasny and Verme (2018b), "Top Incomes and the Measurement of Inequality in Egypt." The World Bank Economic Review 32(2)
- Korinek, A., J. Mistiaen, and M. Ravallion (2007), "An Econometric Method of Correcting for Unit Nonresponse Bias in Surveys." Journal of Econometrics 136:213-235
- Morelli and Munoz (2019a), "kmr: A Command to Correct Survey Weights for Unit Nonresponse using Group's Response Rates." mimeo
- Morelli and Munoz (2019b), "Unit Nonresponse Bias in the Current Population Survey." mimeo