DK responses in surveys on inflation expectations

Natsuki Arai (Gettysburg College)
Biing-Shen Kuo (National Chengchi University)
Yasutomo Murasawa (Konan University)

January 6, 2025

Missing responses in regression analysis

Q: How do you apply regression analysis when the dependent variable has missing responses?

1. Ignore missing responses and apply OLS
2. Use a weighting method
3. Use a sample selection model
4. Others
5. Don't know

Contribution

Give an example of a regression model with DK responses, where

- OLS may suffer from sample selection bias
- ML and Heckit estimates do not coincide
- hence robust Heckit estimator is useful

Plan

Motivation and Contribution

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023, EER)

Conclusion

Plan

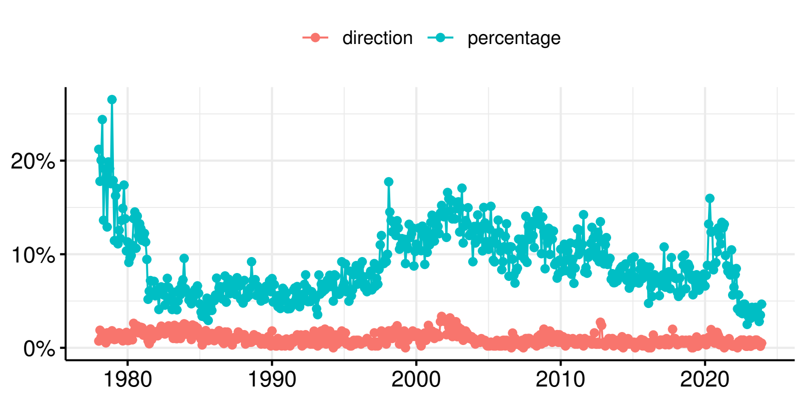
DK responses in surveys

Types of missing responses:

1. nonresponse
    a) unit nonresponse
    b) item nonresponse
2. DK response
    - Common in surveys on inflation expectations (e.g., MSC) especially for quantitative questions
    - Empirical work often discards DK responses $\Longrightarrow$ sample selection bias?

Missing response rates in the MSC

Proportion of DK responses + item nonresponses



Why discard DK responses?

Possible excuse for discarding DK responses:

1. They are ignorable $\implies$ need justification
2. Heckman-type bias correction requires strong assumptions
    - normality
    - exclusion restriction

$\implies$ use a robust estimator

Aims of this paper

1. Use a robust Heckit estimator to handle DK responses
    - developed by Zhelonkin, Genton, and Ronchetti (2016)
    - available as an R package `ssmrob`
2. Reexamine an analysis in Sheen and Wang (2023, EER)
    - Study the influence of monetary condition news on household inflation expectations
    - Use data from the MSC during 2008M12–2015M12 ('zero lower bound' period)
    - Compare OLS, ML, Heckit, and robust Heckit estimates

Findings

1. For both SR and LR expectations, OLS and ML estimates are almost identical (no sample selection bias?)
2. ML and Heckit estimates somewhat differ. For LR expectations, the bias correction term is significant $\Longrightarrow$ DK responses are not ignorable
3. Classical and robust Heckit estimates somewhat differ
4. Monetary condition news remain insignificant, supporting the conclusion of Sheen and Wang (2023)

As a robust statistical method, a robust Heckit estimator is a useful tool for ‘robustness check’ (in the true sense) when estimating a model with DK responses

Plan

Motivation and Contribution

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023, EER)

Conclusion

Sample selection model

Let

- y^* be the latent numerical response
- d be the numerical response dummy

Sample selection model

$$
\begin{aligned}
y &= \begin{cases} y^* & \text{if } d = 1 \\ \text{NA} & \text{if } d = 0 \end{cases} \\
d &= [U > 0] \\
U &= \mathbf{x}'\boldsymbol{\alpha} + z \\
y^* &= \mathbf{x}'\boldsymbol{\beta} + u \\
\begin{pmatrix} z \\ u \end{pmatrix} |\mathbf{x} &\sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \sigma_{zu} \\ \sigma_{uz} & \sigma_u^2 \end{bmatrix}\right)
\end{aligned}
$$

Sample selection bias

Outcome equation for the selected sample

$$\mathrm{E}(y|d=1,x) = x'\boldsymbol{\beta} + \mathrm{E}(u|z > -x'\boldsymbol{\alpha}, x)$$

Consider estimation of $\boldsymbol{\beta}$

- OLS estimator is inconsistent
- ML and Heckit estimators are consistent, but not widely used in the context of DK responses

Plan

Heckit estimator

Moment restrictions:

- Selection equation (probit):

$$\mathrm{E}(s\mathbf{x}h(s\mathbf{x}'\boldsymbol{\alpha})) = \mathbf{0}$$

  where $s := 2d - 1$ gives the sign and $h(.) := \phi(.)/\Phi(.)$ gives the inverse Mill's ratio

- Outcome equation for the selected sample:

$$\mathrm{E}(\mathbf{x}(y - \mathbf{x}'\boldsymbol{\beta} - \sigma_{uz}h(\mathbf{x}'\boldsymbol{\alpha}))d) = \mathbf{0}$$
$$\mathrm{E}(h(\mathbf{x}'\boldsymbol{\alpha})(y - \mathbf{x}'\boldsymbol{\beta} - \sigma_{uz}h(\mathbf{x}'\boldsymbol{\alpha}))d) = 0$$

M-estimator

Let

$$\boldsymbol{\psi}_1(\mathbf{z};\boldsymbol{\theta}) := s\mathbf{x}h(s\mathbf{x}'\boldsymbol{\alpha})$$

$$\boldsymbol{\psi}_2(\mathbf{z};\boldsymbol{\theta}) := \begin{pmatrix} \mathbf{x} \\ h(\mathbf{x}'\boldsymbol{\alpha}) \end{pmatrix} (y - \mathbf{x}'\boldsymbol{\beta} - \sigma_{uz} h(\mathbf{x}'\boldsymbol{\alpha}))d$$

where $\mathbf{z} := (d, s, y, \mathbf{x}')'$ and $\boldsymbol{\theta} := (\boldsymbol{\alpha}', \boldsymbol{\beta}', \sigma_{uz})'$

Let

$$\boldsymbol{\psi}(\mathbf{z};\boldsymbol{\theta}) := \begin{pmatrix} \boldsymbol{\psi}_1(\mathbf{z};\boldsymbol{\theta}) \\ \boldsymbol{\psi}_2(\mathbf{z};\boldsymbol{\theta}) \end{pmatrix}$$

M-estimator of $\boldsymbol{\theta}$ solves

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\psi}\left(\mathbf{z}_i;\hat{\boldsymbol{\theta}}\right) = \mathbf{0}$$

Robustness

- An estimator is robust if its influence function is bounded
- Influence function of an M-estimator:

$$\mathrm{IF}(z) \propto \boldsymbol{\psi}(z; \boldsymbol{\theta})$$

- For the Heckit estimator, $\mathrm{IF}(.)$ is unbounded

Bounded-influence estimator

- Bound $\boldsymbol{\psi}(.;\boldsymbol{\theta})$ to obtain a robust estimator
- Huber function:

$$\Psi(z) := \begin{cases} z & \text{for } |z| \leq K \\ \mathrm{sgn}(z)K & \text{for } |z| > K \end{cases}$$

- Apply a Huber function to the standardized prediction error
- Bound covariates if necessary
- Implementation is easy using `ssmrob` package for R

Bounding $\psi_1(.;\boldsymbol{\theta})$

Write

$$\psi_1(z;\boldsymbol{\theta}) = x\sqrt{h(x'\boldsymbol{\alpha})h(-x'\boldsymbol{\alpha})}r_1$$

where

$$r_1 := \frac{d - \Phi(x'\boldsymbol{\alpha})}{\sqrt{\Phi(x'\boldsymbol{\alpha})\Phi(-x'\boldsymbol{\alpha})}}$$

Let

$$\psi_1^*(z;\boldsymbol{\theta}) := w_1(x)x\sqrt{h(x'\boldsymbol{\alpha})h(-x'\boldsymbol{\alpha})}(\Psi(r_1) - \mathrm{E}(\Psi(r_1)|x))$$

where $w_1(.)$ is a weight function

Bounding $\psi_2(.;\boldsymbol{\theta})$

Write

$$\psi_2(z;\boldsymbol{\theta}) = \begin{pmatrix} x \\ h(x'\boldsymbol{\alpha}) \end{pmatrix} \sigma_w r_2 d$$

where

$$r_2 := \frac{y - x'\boldsymbol{\beta} - \sigma_{uz} h(x'\boldsymbol{\alpha})}{\sigma_w}$$

Let

$$\psi_2^*(z;\boldsymbol{\theta}) := w_2\left(\begin{pmatrix} x \\ h(x'\boldsymbol{\alpha}) \end{pmatrix}\right) \begin{pmatrix} x \\ h(x'\boldsymbol{\alpha}) \end{pmatrix} \Psi(r_2) d$$

where $w_2(.)$ is a weight function

Plan

Sheen and Wang (2023, EER)

- Study how monetary condition news affected households' readiness to spend on durables via their interest rate and inflation expectations during the 'zero lower bound' period
- We focus on one analysis, studying the influence of monetary condition news on SR and LR household inflation expectations
- Use data from the MSC during 2008M12–2015M12
- Estimate a regression equation for the percentage of inflation by OLS, ignoring nonresponses

Inflation expectations in the MSC

Direction

- **px1q1** prices up/down next year
- **px5q1** prices up/down next 5 years

Size

- **px1q2** prices % up/down next year
- **px5q2** prices % up/down next 5 years

Percentage

- **px1** price expectations 1yr recoded
- **px5** price expectations 5yr recoded

Sheen and Wang (2023) mistakenly use `px1q2`/`px5q2` instead of `px1`/`px5`

Covariates

Micro

`MPN`	news: monetary condition
`IN`	news: inflation
`ytl`	income quartiles
`age`	age of respondent
`female`	female dummy
`hsize`	household size
`edu`	education of respondent

Macro

`IP`	industrial production (growth rate at $t-1$)
`UR`	unemployment rate (at $t-1$)
`CPI`	consumer price index (growth rate at $t-1$)

Sample selection

We follow Sheen and Wang (2023):

- Use only wave 2 data to include lagged `px1`/`px5`
- Exclude respondents with missing news or demographic variables

Summary statistics

Variable	N	Mean	SD	Min	Max	NA
px1	14386	3.45	4.07	−25	25	1151
px5	14231	3.17	2.91	−15	25	1306
MPN	15537	0.000 71	0.19	−1	1	0
IN	15537	0.0077	0.23	−1	1	0
age	15537	56.70	16.15	18	97	0
hsize	15537	2.40	1.31	1	10	0
female	15537					
... No	7503	0.48				
... Yes	8034	0.52				
		⋮				

Missing responses for the percentage of inflation

		wave 1	
horizon	wave 2	observed	missing
1 year	observed	13426	960
	missing	734	417
5 year	observed	13234	997
	missing	789	517

Exclusion restriction

- Precise estimation requires a variable that affects selection but not outcome directly
- Higher inflation uncertainty may increase the likelihood of DK responses, but not the level of inflation expectations
- Include the absolute difference of the CPI inflation rate in the previous month in the selection equation
- Correct sign, but insignificant
- Still better to include

Classical estimation

Check for sample selection bias:

- Compare OLS, ML, and Heckit estimates
- Use `sampleSelection` package for R

Findings:

1. For both SR and LR expectations, ML estimates are almost identical to OLS estimates
   $\Longrightarrow$ no sample selection bias?
2. ML and Heckit estimates somewhat differ
   $\Longrightarrow$ model misspecification?
3. For LR expectations, the bias correction term is significant
   $\Longrightarrow$ DK responses are NOT ignorable?

Classical estimation

Outcome equation for `px1`

	OLS	ML	Heckit
MPN	0.17 (0.20)	0.17 (0.20)	0.22 (0.21)
IN	0.65 (0.18)***	0.65 (0.18)***	0.64 (0.19)***
Lpx1	0.24 (0.01)***	0.24 (0.01)***	0.25 (0.01)***
MPN:Lpx1	0.04 (0.04)	0.04 (0.04)	0.04 (0.04)
IN:Lpx1	0.08 (0.03)*	0.08 (0.03)*	0.09 (0.03)**
	$\vdots$		
rho		-0.01 (0.05)	-0.72
invMillsRatio			-2.77 (2.00)
Num. obs.	13426	14160	14160
Censored		734	734

Classical estimation

Outcome equation for px5

	OLS	ML	Heckit
MPN	-0.13 (0.19)	-0.13 (0.19)	-0.03 (0.22)
IN	0.53 (0.15)***	0.53 (0.15)***	0.58 (0.18)**
Lpx5	0.29 (0.01)***	0.29 (0.01)***	0.32 (0.01)***
MPN:Lpx5	0.06 (0.05)	0.06 (0.05)	0.05 (0.05)
IN:Lpx5	-0.07 (0.03)	-0.07 (0.03)	-0.06 (0.04)
	⋮		
rho		-0.01 (0.05)	-1.30
invMillsRatio			-4.13 (1.42)**
Num. obs.	13234	14023	14023
Censored		789	789

Robust estimation

Robustness check:

- Compare classical and robust Heckit estimates
- Use `ssmrob` package for R
- Set $K = 100$ (classical) or $K = 1.345$ (robust)
- Set $w_1(.) := 1$ and $w_2(\mathbf{x}_i) := \sqrt{1 - \mathbf{x}_i'(X'X)^{-1}\mathbf{x}_i}$

Findings:

- For both SR and LR expectations, the bias correction term is insignificant
- Monetary condition news remain insignificant
- Micro covariates become insignificant
- Macro covariates remain significant

Robust estimation

Outcome equation for px1		
	classical ($K = 100$)	robust ($K = 1.345$)
MPN	0.22 (0.25)	0.12 (0.19)
IN	0.64 (0.19)***	0.60 (0.14)***
Lpx1	0.25 (0.01)***	0.24 (0.02)***
MPN:Lpx1	0.04 (0.06)	0.04 (0.06)
IN:Lpx1	0.09 (0.05)	0.04 (0.05)
	⋮	
IMR1	−2.78 (2.49)	0.61 (6.23)
Num. obs.	14160	14160
Censored	734	734

Robust estimation

Outcome equation for px5		
	classical ($K = 100$)	robust ($K = 1.345$)
MPN	-0.03 (0.30)	0.15 (0.22)
IN	0.58 (0.21)**	0.43 (0.19)*
Lpx5	0.32 (0.02)***	0.31 (0.02)***
MPN:Lpx5	0.05 (0.10)	-0.01 (0.06)
IN:Lpx5	-0.06 (0.06)	-0.04 (0.06)
	⋮	
IMR1	-4.13 (1.92)*	-3.90 (3.54)
Num. obs.	14023	14023
Censored	789	789

Plan

Summary

- One cannot assume *a priori* that DK responses are ignorable. Try to estimate a sample selection model.
- ML and Heckit estimates may substantially differ, perhaps because of model misspecification and nonrobustness of classical estimators.
- Robust Heckit estimator is useful for robustness checks.

Remaining issues

1. Global misspecification
    - Semi/non-parametric estimators are not necessarily robust
    - Need a robust semi/non-parametric estimator
2. DK responses in explanatory variables
    - No sample selection bias
    - Efficiency improves by including them using DK dummies
3. Unit nonresponses
    - Use weights if ignorable
    - Some covariates (e.g., region) may be available by the sampling design

Sheen, J., & Wang, B. Z. (2023). Do monetary condition news at the zero lower bound influence households' expectations and readiness to spend? *European Economic Review*, *152*(104345).

Zhelonkin, M., Genton, M. G., & Ronchetti, E. (2016). Robust inference in sample selection models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *78*, 805–827.