DK responses in surveys on inflation expectations

Natsuki Arai[1] Biing-Shen Kuo[2] Yasutomo Murasawa[3]

SNDE 2025

[1]Gettysburg College
[2]National Chengchi University
[3]Konan University

Plan

Motivation

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023)

Results

Conclusion

Plan

Motivation

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023)

Results

Conclusion

Missing responses in surveys

Survey questions with many missing responses:

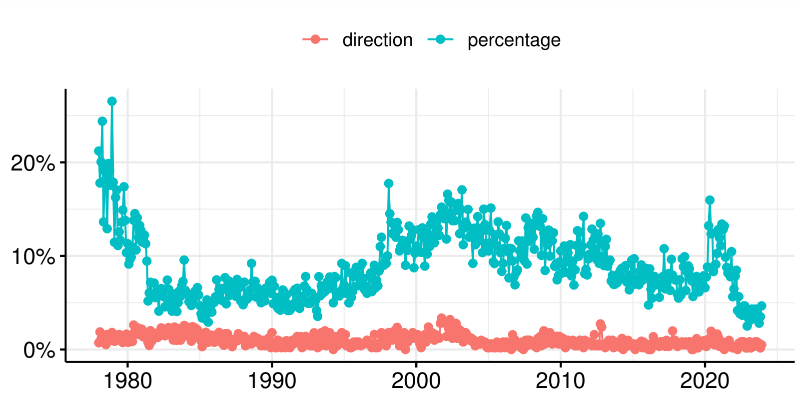
- wage
- voting behavior
- quantitative inflation expectation

Types of missing responses:

1. nonresponse
   a) unit nonresponse
   b) item nonresponse
2. DK response

Missing response rates for inflation expectations (Michigan Survey of Consumers)

Proportion of DK responses + item nonresponses



Dealing with DK responses

Recent works (on inflation expectations) discard DK responses in regression analysis:

- Sheen and Wang (2023, Eur. Econ. Rev.)
- Tsiaplias (2021, J. Appl. Econom.)
- Tsiaplias (2020, J. Econ. Dyn. Control)
- Wang, Sheen, Trück, Chao, and Härdle (2020, Macroecon. Dyn.)
- Ehrmann, Pfajfar, and Santoro (2017, Int. J. Cent. Bank.)

$\Longrightarrow$ sample selection bias?

Why discard DK responses?

Possible excuses:

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

    $\implies$ Use a robust estimator

Aim of this work

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)
    - Study the influence of monetary condition news on household inflation expectations
    - Use data from the MSC, 2008M12–2015M12 ('zero lower bound' period)
    - Compare OLS, ML, Heckit, and robust Heckit estimates

Plan

Motivation

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023)

Results

Conclusion

Sample selection model

Let

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

Sample selection model

$$y = \begin{cases} y^* & \text{if } d = 1 \\ \text{NA} & \text{if } d = 0 \end{cases}$$

$$d = [\boldsymbol{x}'\boldsymbol{\alpha} + z > 0]$$

$$y^* = \boldsymbol{x}'\boldsymbol{\beta} + u$$

$$\begin{pmatrix} z \\ u \end{pmatrix} | \boldsymbol{x} \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \sigma_{zu} \\ \sigma_{uz} & \sigma_u^2 \end{bmatrix}\right)$$

Sample selection bias

Outcome equation for the selected sample

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

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 in surveys on inflation expectations*"

Plan

Motivation

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023)

Results

Conclusion

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

Estimating functions:

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

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

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

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

(=Heckit estimator of $\boldsymbol{\beta}$)

Robustness

- An estimator is robust to outliers 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, $\boldsymbol{\psi}(.; \boldsymbol{\theta})$ is unbounded; hence NOT robust

Bounded-influence estimator

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

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

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

Plan

Motivation

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023)

Results

Conclusion

Sheen and Wang (2023, EER)

- Study the influence of monetary condition news on SR and LR household inflation expectations
- Use data from the MSC, 2008M12–2015M12 ('zero lower bound' period)
- Estimate a regression equation for the percentage of inflation by OLS, ignoring nonresponses
- Find that monetary condition news was insignificant

Inflation expectations in the MSC

Q1: Direction

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

Q2: Size (only if up/down to Q1)

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

Regressors

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 inflation expectation for the dependent variable to include lagged (wave 1) inflation expectation as a regressor
- Exclude respondents with missing news/demographic variables

Sample size

		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

- Higher inflation uncertainty may increase the likelihood of DK responses, but not the level of inflation expectations
- Include the absolute change of the CPI inflation rate in the previous month in the selection equation
- Correct sign, but insignificant
- Still better to include

Plan

Motivation

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023)

Results

Conclusion

Classical estimation

Classical estimation:

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

Parameters of interest:

1. Coefficient on MPN
2. Coefficient on the bias correction term (IMR)

Classical estimation (SR)

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)**
	⋮		
rho		−0.01 (0.05)	−0.72
invMillsRatio			−2.77 (2.00)
Num. obs.	13426	14160	14160
Censored		734	734

Classical estimation (LR)

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

Why are the ML and Heckit estimates different?
$\Longrightarrow$ Model misspecification
Two possibilities:

1. Only Heckit is consistent
2. Both ML and Heckit are inconsistent

Robustness check:

- Compare classical and robust Heckit estimates
- Use `ssmrob` package for R
- Set $K = 100$ (classical) or $K = 1.345$ (robust)

Robust estimation (SR)

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 (LR)

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

Findings

1. For both SR and LR inflation expectations, OLS and ML estimates are almost identical
   $\Longrightarrow$ No sample selection bias (?)
2. ML and Heckit estimates somewhat differ. For LR expectations, the bias correction term is significant
   $\Longrightarrow$ Sample selection bias
3. Classical and robust Heckit estimates somewhat differ
   $\Longrightarrow$ Robust estimate is more reliable
4. Monetary condition news remains insignificant
   $\Longrightarrow$ Support the conclusion of Sheen and Wang (2023)

Plan

Motivation

Regression model with DK responses

Robust Heckit estimator

Reexamination of Sheen and Wang (2023)

Results

Conclusion

Summary

- One should not simply ignore "*DK responses in surveys on inflation expectations*." Use a sample selection model.
- ML and Heckit estimates may differ, perhaps because of model misspecification.
- Use a robust Heckit estimator for a robustness check (in the true sense).

Remaining issues

1. Global misspecification
   - Our model may not be even approximately correct
   - Need a (robust) semi/non-parametric estimator
2. DK responses in the regressors
   - Can include them using DK dummies
     $\Longrightarrow$ conditional heteroskedasticity
   - Need a (robust) generalized Heckit estimator
3. Unit nonresponses
   - Need additional information, e.g., regional nonresponse rates
4. Qualitative information in DK responses
   - Can combine data on the direction and percentage of inflation to improve inference

Ehrmann, M., Pfajfar, D., & Santoro, E. (2017). Consumers' attitudes and their inflation expectations. *International Journal of Central Banking*, *47*, 225–259.

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).

Tsiaplias, S. (2020). Time-varying consumer disagreement and future inflation. *Journal of Economic Dynamics and Control*, *116*(103903).

Tsiaplias, S. (2021). Consumer inflation expectations, income changes and economic downturns. *Journal of Applied Econometrics*, *36*, 784–807.

Wang, B. Z., Sheen, J., Trück, S., Chao, S.-K., & Härdle, W. K. (2020). A note on the impact of news on US household inflation expectations. *Macroeconomic Dynamics*, *24*, 995–1015.

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.