Summary of A Natural Robustification of the Ordinary Instrumental Variables Estimator

Link to the paper.

1 Problem

- Ordinary Least Squares (OLS) estimates are biased and inconsistent when one of the covariates is correlated with the error terms (*endogenous*). This happens when (1) covariate is associated with unobserved predictors, (2) is measured with error, and/or (3) affected by the response variable.
- Solution to this problem is to use Ordinary Instrumental Variables (OIV) to consistently estimate the parameters of the model. However, when used in real datasets, OIV's are often distorted due to outliers in the response variable.
- There are existing robust methods to obtain IV estimators as listed below
    1. Minimize robust loss function (Amemiya, 1982)
    2. Robustify the estimating equations which define the OIV (Krasker and Welsch, 1985)
    3. Orthogonality conditions which give rise to the OIV in the Generalized Method of Moments
    4. Robustify at each of the stages of the OIV since OIV's can be thought of as two-stage least squares estimator (2SLS).
    5. Robust estimators for simultaneous equation models

2 Objective

The authors propose a robust instrumental variable (RIV) estimator based on the robustification of the OIV's closed form formula. The authors replace the sample moments in an OIV with a robust multivariate location and scatter S-estimators. Under certain regularity conditions, the advantage of using S-estimators is that they are consistent, asymptotically normal, affine equivariant, positive definite, have bounded influence function and can achieve maximal breakdown point regardless of dimension.

3 Methods

3.1 Estimator

The OIV estimator is given by

$$\hat{\alpha}_{OIV} = \hat{\mu}_y - \hat{\mu}_x' \hat{\beta}_{OIV}$$

$$\hat{\beta}_{OIV} = [\hat{\Sigma}_{XZ}\hat{\Sigma}_{ZZ}^{-1}\hat{\Sigma}_{ZX}]^{-1}[\hat{\Sigma}_{XZ}\hat{\Sigma}_{ZZ}^{-1}\hat{\Sigma}_{ZX}]$$

where $(\hat{\mu}, \hat{\Sigma})$ is the sample mean and sample covariance matrix based on a sample $\{(x_i, z_i, y_i)\}_{i=1}^n$ where $(x_i, z_i, y_i) \in R^p \times R^q \times R$

(For Jana: x_i's are exposures, z_i's are instruments, and y_i's are outcomes)

The authors robustify the OIV estimator by replacing the sample estimators $(\hat{\mu}, \hat{\Sigma})$ with robust multivariate location and scatter S-estimator (m, S).

$$\hat{\alpha}_{RIV} = m_Y - m_x' \hat{\beta}_{RIV}$$

$$\hat{\beta}_{RIV} = [S_{xz} S_{zz}^{-1} S_{zx}]^{-1} [S_{xz} S_{zz}^{-1} S_{zx}]$$

Task: I need to understand what are robust multivariate location and scatter S-estimators.

3.2 Properties of RIV

(1) Equivariant and Consistent

- **Task:** I need to understand what is *regression* and *carrier* equivariant.

(2) Bounded and Asymptotically Normal Influence Function

- **Task:** I need to understand what are influence functions

(3) High Breakdown Point

- **Task:** I need to learn what are breakdown points.

4 Results

Simulation

The authors compare performance of RIV with that of OIV and, Two-Stage Generalized M-estimators (2SGM) (Wagenvoort and Waldmann, 2002), Weighted Instrumental variables (WIV) (Krasker and Welsh, 1985), and Multivariate Tau estimator (Tau) (Maronna and Yohia, 1997). They considered the following regression model

$$y_i = \alpha + x_{1i}\beta_1 + x_{2i}\beta_2 + \epsilon_i \text{ for } i = 1, \ldots, n$$

The authors generated outliers of four types: (1) in the response, y, (2) in the endogenous covariate, x_1, (3) in the exogenous covariate, x_2, and (4) in the IV, z. They use MedSE which shows estimator's worst perforamnce obtained across all types of contamination. Overall, the RIV had the lowest MedSE especially when contamination percent was 20% or higher.

Heart Study

In addition, the performance of the RIV was compared to other estimators when applied to the Framingham Heart Study. The authors focus specifically on the effect of systolic blood pressure (SBP) on left atrial dimension (LAD). The other estimators, OIV, 2SGM, WIV, and Tau did not attain nearly the same level of statistical significance compared to the RIV. Particularly, the OIV estimand for the SBP is signficantly difference from the other estimators is because the OIV gives equal weights to the all observations and so is highly sensitive to outliers.

5 Conclusion

The advantage of uing RIV in datasets with outliers is that it downweights observations which signficantly depart from the bulk of the data and helps uncover the true relationships among the variables.

6 Questions

- **Task** Understand the other robust methods to obtain IV estimators and list down their pros/cons.
- **Question** What are common methods to check for validity of IV? Are there computational methods already in place?