Summary of the Simulation Work

April 6, 2016

This project is intended to compare the performance of post-Lasso (Belloni, Chen, Chernozhukov, & Hansen, 2012) and RJIVE (Hansen & Kozbur, 2014) with REL and BC-REL (Shi, 2015) in a linear IV model. We summarize the data generating process of the simulation, implementation of post-Lasso and RJIVE on generated data and the simulation result in this short summary.

Data Generating Process

The data generating process we adopt in this simulation work is from the B.4 section of Shi (2015). The data is generate as follows.

The structural equation is

$$e_i^{(0)} = y_i - (x_{i1}, x_{i2})\beta \tag{1}$$

where $e_i^{(0)}$ is the structural error and (x_{i1}, x_{i2}) and two endogenous variables. The true reduced-form equations for the endogenous variables are

$$x_{i1} = 0.5z_{i1} + 0.5z_{i2} + e_i^{(1)} \tag{2}$$

$$x_{i2} = 0.5z_{i3} + 0.5z_{i4} + e_i^{(2)} \tag{3}$$

where (e_i^1, e_i^2) are the reduced-form errors and each endogenous variables is supported by two relevant IVs from a large number of IVs $(z_{ij})_{j=1}^m$ orthogonal to the structural error. We generate $(z_{ij})_{j=1}^m \sim i.i.d\, N(0,1)$ and

$$\begin{bmatrix} e_i^{(0)} \\ e_i^{(1)} \\ e_i^{(2)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, 0.25 \begin{bmatrix} 1 & rho & rho \\ rho & 1 & 1 \\ rho & 0 & 1 \end{bmatrix} \right) \tag{4}$$

where rho stands for the magnitude of endogeneity.

In the simulation we try combinations of dimensionality $n = 120\, or\, 240$ and $m = 80\, or\, 160$ and set the $rho = 0.6$ and $\beta_0 = (1, 1)'$. The DGP script is **dgpLinearIV.m**, whose outputs are the dependent variable y $(n \times 1)$, endogenous variable x $(n \times 2)$ and instruments Z $(n \times m)$.

Post-Lasso

The script **post_lasso.m** implements the post-lasso estimation on the generated data and output $\hat{\beta}$. This function is a modified version of the original codes provided by authors of the paper(Belloni et al., 2012). Tuning parameters we used in the function are same as the original codes. We firstly select instruments for each endogenous variables by Lasso using the function **LassoShooting2.m** and then run two-stage-least-square regression including the selected instruments by **tsls.m**. These two functions are provided by the authors(Belloni et al., 2012).
To drive **LassoShooting2.m**, we need two supportive scripts **prepareArgs.m** and **process_options.m**.

Regularized JIVE

The script **RJIVE.m** implements the regularized JIVE estimator. According to Hansen and Kozbur (2014), the ridge-regularized JIVE estimator is defined as

$$\tilde{\beta} = (\sum_{i=1}^{n} \hat{\Pi}^{\wedge\prime}_{-i} Z_i X_i')^{-1} (\sum_{i=1}^{n} \hat{\Pi}^{\wedge\prime}_{-i} Z_i y_i) \tag{5}$$

where

$$\hat{\Pi}^{\wedge}_{-i} = (Z'Z - Z_i Z_i' + \wedge' \wedge)^{-1} (Z'X - Z_i X_i') \quad and \quad \wedge' \wedge = C^2 m I_n \tag{6}$$

Following the footnote on page 295 of the paper(Hansen & Kozbur, 2014), the tuning parameter C is the standard deviation of the residuals obtained by regressing X on a column of ones, that is

$$\varepsilon = (I_n - X(X'X)^{-1}X') \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}' \tag{7}$$

and

$$C = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^2} \tag{8}$$

Regularized LIML

The script Rliml.m implements the regularized LIML estimator proposed by Carrasco and Tchuente (2015). The code is mainly a modification of relevant pieces of the original code adopting Tikhonov regularization written by the authors(Carrasco & Tchuente, 2015).
Regularized LIML is defined as

$$\hat{\beta} = (X'(P^{\alpha} - \nu I_n)X)^{-1}X'(P^{\alpha} - \nu I_n)y \tag{9}$$

where

$$\nu = \nu_{\alpha} = \min_{\beta} \frac{(y - X\beta)'P^{\alpha}(y - X\beta)}{(y - X\beta)'(y - X\beta)} \tag{10}$$

We construct the matrix P^{α} following the procedure in authors' original code. In the optimization step, we make use of 'fmincon' function, set the initial value as the estimation of β from LIML, $\hat{\beta}_{LIML}$, algorithm as interior point algorithm. To avoid that too many estimation results are outliers, we set an upper bound $(6, 6)'$ and a lower bound $(-4, -4)'$ for the optimization result in the options of 'fmincon' function.

Simulation Results

The master file is **master_IV.m**. For each combination of n and m, we fix the seed of random number generator and simulate both post-Lasso estimator and Regularized JIVE for 500 replications. We report only the estimation of the first parameter β_1.
After obtain the estimation result, we use the function **output_bias_rmse.m** to compute the bias and RMSE. In the function, we firstly check if there are any outliers to ensure

1. the estimated value $\hat{\beta}_1$ is not infinity or NaN.

2. the estimated value is not so far away from the true value to avoid ruining the overall performance of the estimator, i.e. $|\hat{\beta}_1 - \beta_{1,0}| < 15$.

Based on the criteria, we dropped 2 outliers when $(n, m) = (120, 80)$ using post-Lasso and 5 outliers when $(n, m) = (120, 160)$ using post-Lasso. Under other circumstances, no outliers are detected.

Then we compute the bias and RMSE by

$$bias = \frac{1}{R}\sum_{r=1}^{R} \hat{\beta}_1^{(r)} - \beta_{1,0}^{(r)} \tag{11}$$

$$RMSE = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\beta}_1^{(r)} - \beta_{1,0}^{(r)})^2} \tag{12}$$

The simulation result is summarized in the following table:

	(n,m)	(120,80)	(120,160)	(240,80)	(240,160)
post-Lasso	bias RMSE	0.0301 0.5963	0.0122 0.7255	0.0003 0.0417	0.0020 0.0416
RJIVE	bias RMSE	-0.0136 0.1006	-0.0321 0.1980	-0.0062 0.0482	-0.0052 0.0625
RLIML	bias RMSE	-0.0019 0.0694	0.2529 0.9574	-0.0026 0.0434	-0.0015 0.0470

References

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, *80*(6), 2369 - 2429.

Carrasco, M., & Tchuente, G. (2015). Regularized liml for many instruments. *Journal of Econometrics*, 427 - 442.

Hansen, C., & Kozbur, D. (2014). Instrumental variables estimation with many instruments using regularized jive. *Journal of Econometrics*, 290 - 308.

Shi, Z. (2015). Econometric estimation with high-dimensional moment equalities. Retrieved from `http://ssrn.com/abstract=2491102`