UNIVERSTÄT ST. GALLEN

HOCHSCHULE FÜR WIRTSCHAFTS-, RECHTS- UND SOZIALWISSENSCHAFTEN

Data Analytics II: Causal Econometrics

Simulation Study

Lauritz Storch

[redacted]

lauritz.storch@student.unisg.ch

[redacted]

May 1 2022

[redacted]

Table of Content

1. Setting

1.1 Parameter of Interest

The parameter of interest is the so-called local average treatment effect (LATE). It represents the treatment effect of a specific sub-sample which were assigned to the treatment and was first discussed by Imbens and Angrist (1994). Under a setting with a Bernoulli treatment effect, the LATE can be denoted as:

$$LATE = \frac{IIT}{IIT_D} = \frac{E[Y_i(z=1)] - E[Y_i(z=0)]}{E[d_i(z=1)] - E[d_i(z=0)]}$$

Where IIT is the average effect of experimental assignment on the outcomes without considering that the group is split into treated and untreated sub-samples. The IIT_D measures the treatment effect under consideration of the subjects which are assigned to the treatment. Under full compliance, it can be shown that $IIT = ATE.$

1.2 Identification Strategy

While the IV method fails to capture either the ATE or $ATET$, because the method measures effects on subgroups, it is a common strategy to identify the $LATE$ for the same reason. Treated subjects are compliers in the IV set which means that these subjects' treatment status can be manipulated by an IV.

In a typical IV setting, one can observe that we have an indirect effect on our instrument through one or several explanatory variables on our dependent variable. Therefore, the explanatory variable is also dependent on the instrument. Regressing our dependent variable on, both, the explanatory and instrument variables would lead to a bias in our model. Hence, a first stage is performed by regressing the dependent explanatory variable on the instrument. The fitted values are then used to replace the dependent explanatory variable to obtain the reduced form regression.

First stage: $X_{endog} = \delta_0 + \theta_1 * Z_1 + \delta_2 * X_{exog} + \hat{U}$

Second stage: $Y = \beta_0 + \beta_1 * X_{endog} + \beta_2 * X_{exog} + U$

Reduced form: $Y = \alpha_0 + \alpha_1 * Z_1 + \alpha_2 * X_{exog} + V$

Where $\alpha_0 = \beta_0 + \beta_{endog} * \delta_0$, $\alpha_1 = \beta_{endog}\theta_1$ and $\alpha_2 = \beta_2 + \beta_{endog} * \delta_2$.

The identifying assumptions of the IV methods are:

1. The instrument Z has a causal effect on X. (relevance assumption)

2. Z affects the outcome Y only through X. (exclusion restriction)

3. Z does not share common causes with the outcome Y. In other words, Z has no confounding effect on Y. (exchangeability assumption)

1.3 Estimators under Consideration

The following estimators, 2SLS and LIML, are both k-class estimators of the form:

$$\beta_{k-class} = (\boldsymbol{X}'(\boldsymbol{I} - \kappa\boldsymbol{M})\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \kappa\boldsymbol{M})\boldsymbol{Y} \qquad \text{where:} \quad \beta_{k-class} = [\beta_{exog}, \ \beta_{endog}]$$

$$\boldsymbol{X} = [\boldsymbol{X}_{exog}, \boldsymbol{X}_{endog}]$$

If we specify κ as λ, the smallest eigenvalue of an underlying calculation, the k-class estimator becomes the LIML estimator. If κ is equal to one, then the k-class estimator will become the 2SLS estimator. Therefore, both estimators are closely related.

1.3.1 2SLS approach

The Two-stage Least Square (2SLS) is the most common approach for the IV method with multiple instrument variables. The approach is the same as described previously. Notation wise, the estimator for the 2SLS approach is denoted as:

$$\beta_{2SLS} = \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1}\widehat{\boldsymbol{X}}\boldsymbol{Y} \qquad \text{where:} \ \widehat{\boldsymbol{X}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$$

$$\boldsymbol{Z} \equiv (\boldsymbol{1}, \boldsymbol{X_1}, \dots, \boldsymbol{X_{K-1}}, \boldsymbol{Z_1}, \dots, \boldsymbol{Z_M})$$

$$\boldsymbol{X} \equiv (\boldsymbol{1}, \boldsymbol{X_1}, \dots, \boldsymbol{X_K})$$

Under some specific conditions, one can show that the 2SLS estimator is a consistent and asymptotically normal distributed.

1.3.2 LIML approach

The limited information maximum likelihood (LIML) estimator is similar to the 2SLS. If the underlying equations (see first and second stage) are exactly identified, then both estimators are equal. However, the LIML is a good estimator if weak instruments are found in the first stage. The LIML estimator is denoted as:

$$\beta_{LIML} = (\boldsymbol{X}'(\boldsymbol{I} - \lambda\boldsymbol{M})\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \lambda\boldsymbol{M})\boldsymbol{Y}$$

The LIML approach is also asymptotically normal distributed the under right specification.

1.3.2 Performance measures

For the performances measure I consider the bias, variance, mean squared error, and standard errors. With increasing sample size, the performance measures of 2SLS and LIML estimators should vanish due to the asymptotic normal distribution. This must be considered when defining the data generating processes (*DGP*) in the following step.

2. Simulation Design

2.1 Data Generating Process 1[1]

The first DGP consists of the following linear regression:

First stage: $X_2 = d_0 + d_1 * Z_1 + d_2 * X_1 + v$ $u, v \sim N(0,1)$

Second stage: $Y = b_0 + b_1 * X_2 + b_2 * X_1 + u$ $X, Z_1 \sim uniform(0,1)$

Reduced form: $Y = \beta_0 + \beta_1 * Z_1 + \beta_2 * X_1 + \hat{u}$ $b_0 = d_0 = 0,\ b_1 = 0.3,$ $b_2 = 0.8, d_1 = 1, d_2 = 0.4$

In this data generating process I will reduce the number of observations to 5. The number of simulations is set to 20. The reduction of observations and simulations might lead the result that the LIML performs worse than the 2SLS. Generally, the LIML should almost always lead to similar results under a correctly specified model.

2.2 Data Generating Process 2

The first DGP consists of the following linear regression:

First stage: $X_2 = d_0 + d_1 * Z_1 + \sum_{i=2}^{20} d_i * z_i + d_2 * X_1 + v$ $u, v \sim N(0,1)$

Second stage: $Y = b_0 + b_1 * X_2 + b_2 * X_1 + u$ $X, Z_1 \sim uniform(0,1)$

Reduced form: $Y = \beta_0 + \beta_1 * Z_1 + \beta_2 * X_1 + \sum_{i=2}^{20} \beta_{i+1} * z_i + \hat{u}$ $b_0 = d_0 = 0,$

$b_1 = 0.3, b_2 = 0.8, d_1 = 1,$ $d_2 = 0.4, d_i = 0.0001$

The number of observations is increased to 1000 and the number of simulations is set to 100. In the first stage are now 19 more instrument variables added. All have the same weak coefficient. Since 2SLS performs better in a weak IV setting, the LIML should perform better in this DGP.

2.1 Data Generating Process 3

The first DGP consists of the following linear regression:

First stage: $X_2 = d_0 + d_1 * Z_1 + d_2 * X_1 + v$ $u, v \sim N(0,1)$

Second stage: $Y = b_0 + b_1 * X_2 * Z_1 + b_2 * X_1 + u$ $X, Z_1 \sim uniform(0,1)$

Reduced form: $Y = \beta_0 + \beta_1 * Z'_1 * Z_1 + \beta_2 * X_1 + \hat{u}$ $b_0 = d_0 = 0,\ b_1 = 0.3,$ $b_2 = 0.8, d_1 = 1, d_2 = 0.4$

The number of observations and the number of simulations is equal to DGP 2. The instrument variable will also enter the second stage and, hence, hurt the exchangeability assumption. Both estimators should perform very bad.

---

[1] **Disclaimer:** Change of notation. I fitted the notation above to the lecture slides to make comparison easier. The following notation will follow my own notation from the code. The underlying principles stay the same.

3. Results

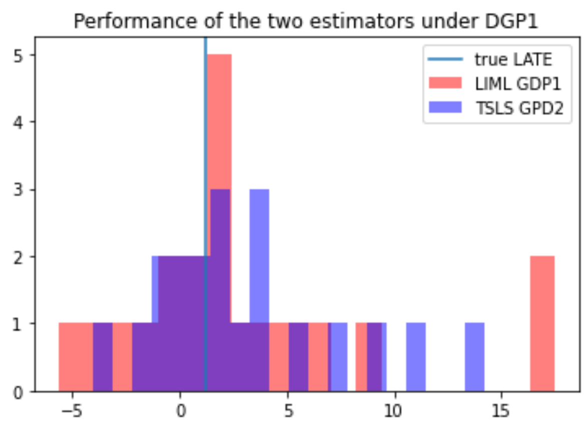
3.1 Summary of the Results DGP1

Before starting the comparison on my data, I want to emphasize that the comparison will almost only include the assessment of beta 1 since it focuses only on the LATE comparison of both estimators. Other betas are also presented but not explained due to completeness.

As we see in table 1, the inference statistics show that the 2SLS estimator performs better than the LIML estimator in DGP1. Variance, mean squared error and standard error are all smaller for the 2SLS estimator. Only the bias is almost identical. As the number of observations is reduced, this effect is due to randomness. The graph shows that this is only due to the reduction of observations and simulations. This is on purpose since the LIML is, according to literature, commonly performing better than 2SLS. The graph also shows that I have several outliers which will influence the result of the performance measures tremendously.

Data Generating Process (DGP) 1 Estimation Results:

	bias LIML	bias TSLS	var LIML	var TSLS	mse LIML	mse TSLS	se LIML	se TSLS
beta 0	-0.47605	-0.51233	8.55238	7.46065	8.77900	7.72314	2.93918	2.74518
beta 1	1.80471	1.81508	33.81145	19.35678	37.06844	22.65130	5.84406	4.42180
beta 2	-0.63788	-0.29712	30.31974	12.34456	30.72664	12.43284	5.53408	3.53118

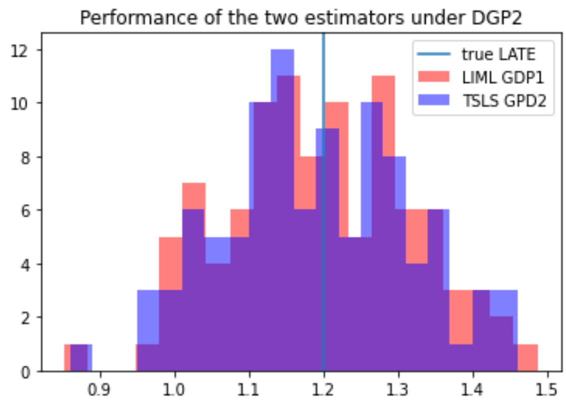


3.2 Summary of the Results DGP2

The LIML estimator should perform better in a setting with many weak instruments. This is not fully correct here. One can only observe that the bias is less for the LIML estimator. In all other cases, the LIML performs almost identically. The first conjecture for these results might be the number of observations since both are asymptotically normal distributed which should lead to a decrease in our performance measures. A reduction to 500 and 100 observations did not confirm this conjecture. Therefore, the assumption that the LIML performs better than the TSLS does not hold in this setting.

Data Generating Process (DGP) 1 Estimation Results:

	bias LIML	bias TSLS	var LIML	var TSLS	mse LIML	mse TSLS	se LIML	se TSLS
beta 0	0.01774	0.01970	0.07479	0.01885	0.07511	0.01924	0.27486	0.13798
beta 1	−0.00655	−0.00901	0.01528	0.01538	0.01533	0.01546	0.12425	0.12465
beta 2	0.02073	0.00630	0.01743	0.00398	0.01786	0.00402	0.13268	0.06339
beta 3	−0.31525	−0.32171	0.01379	0.00271	0.11317	0.10620	0.11801	0.05230
beta 4	−0.34715	−0.32144	0.01593	0.00366	0.13644	0.10698	0.12687	0.06079
beta 5	−0.33562	−0.31985	0.01391	0.00265	0.12655	0.10496	0.11854	0.05175
beta 6	−0.34386	−0.31969	0.01550	0.00279	0.13374	0.10499	0.12514	0.05311
beta 7	−0.31867	−0.31562	0.01133	0.00268	0.11288	0.10230	0.10698	0.05206
beta 8	−0.33456	−0.32754	0.01436	0.00300	0.12628	0.11028	0.12042	0.05503
beta 9	−0.30544	−0.32636	0.01471	0.00335	0.10801	0.10987	0.12192	0.05821
beta 10	−0.31615	−0.32609	0.01470	0.00254	0.11464	0.10887	0.12184	0.05061
beta 11	−0.30048	−0.31672	0.01427	0.00296	0.10456	0.10328	0.12005	0.05469
beta 12	−0.32239	−0.31656	0.01604	0.00360	0.11998	0.10381	0.12729	0.06029
beta 13	−0.31194	−0.31285	0.01643	0.00287	0.11374	0.10074	0.12882	0.05380
beta 14	−0.31517	−0.32515	0.01352	0.00258	0.11285	0.10831	0.11687	0.05107
beta 15	−0.32420	−0.33129	0.01545	0.00274	0.12056	0.11249	0.12494	0.05264
beta 16	−0.32242	−0.32381	0.01602	0.00246	0.11997	0.10731	0.12719	0.04981
beta 17	−0.32939	−0.32915	0.01472	0.00312	0.12322	0.11146	0.12194	0.05613
beta 18	−0.32695	−0.31560	0.01317	0.00301	0.12007	0.10261	0.11535	0.05514
beta 19	−0.32046	−0.32015	0.01692	0.00297	0.11961	0.10547	0.13072	0.05476
beta 20	−0.31000	−0.31363	0.01285	0.00287	0.10895	0.10123	0.11394	0.05382
beta 21	−0.32715	−0.33098	0.01247	0.00317	0.11950	0.11272	0.11222	0.05662

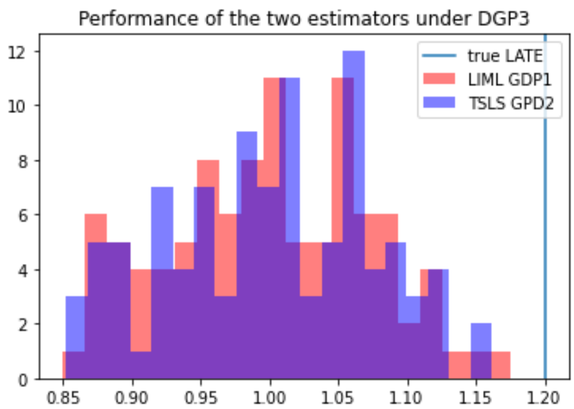


3.3 Summary of the Results DGP3

As previously explained, the multiplication of the instrument variable with the covariates hurts the exchangeability assumption. As we see, all inferences are influenced negatively. With a focus on beta, it can also be said that the LIML and 2SLS perform almost equally badly. Only the standard error is significantly higher for the 2SLS. The graph also indicates that all estimators are dense around a wrong parameter value.

Data Generating Process (DGP) 1 Estimation Results:

	bias LIML	bias TSLS	var LIML	var TSLS	mse LIML	mse TSLS	se LIML	se TSLS
beta 0	−0.14678	0.02115	0.00173	0.00164	0.02328	0.00209	0.04182	2.74518
beta 1	−0.20041	−0.19948	0.00543	0.00550	0.04559	0.04529	0.07403	4.42180
beta 2	0.19374	−0.14305	0.00550	0.00115	0.04304	0.02162	0.07456	3.53118



References

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American statistical Association, 90(430), 431-442.