Explorations in NISE Estimation

Blankmeyer, Eric

April 2021

Explorations in NISE Estimation

Eric Blankmeyer

Email eb01@txstate.edu

April 2021

Abstract. Ordinary least squares, two-stage least squares and the NISE estimator are applied to three data sets involving equations from microeconomics and macroeconomics. The focus is on simultaneity bias in linear least squares and on the ability of the other estimators to mitigate the bias.

Explorations in NISE Estimation

1. Introduction

The detection and correction of simultaneous-equation bias in ordinary least squares (OLS) remains a challenging issue in the estimation of linear models. If exclusion restrictions are the basis for the identification of parameters, researchers typically rely on instrumental variables (IV), for example two-stage least squares (TSLS) or limited information maximum likelihood (LIML). “The various methods that have been developed for simultaneous-equations models are all IV estimators” (Greene 2003, 398). This option fails, however, if the instrument set is weakly correlated with the endogenous variables or is in effect an omitted variable from the equation of interest. “Those who use instrumental variables would do well to anticipate the inevitable barrage of questions about the appropriateness of their instruments” (Leamer 2010, 35).

An alternative to IV is the Non-Instrumental Simultaneous-Equation (NISE) estimator, which is applied in this paper to the consistent estimation of equations from microeconomics and macroeconomics. A researcher may select NISE in several situations: (i) observations on the instruments are unavailable or incomplete; (ii) the instruments are found to be weak; (iii) they fail Sargan’s J test for exogeneity; or (iv) the researcher simply wants a second opinion about her IV estimates. The papers by Blankmeyer (2017a, 2017b,2018,2020) provide analytical details, simulations and additional applications while Chow (1964, 533-537, 542-543) shows how the estimator that I call NISE is related to canonical correlation, TSLS and LIML. The next section provides a concise description of NISE. Section 3 revisits the wage equation from Klein’s Model I; the supply of business loans is modeled in section 4; and the derived demand for nursing services is the subject of section 5. The final section offers several conclusions and caveats.

2. The NISE estimator

In the simultaneous linear equation

$$\mathbf{Y}\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad (1)$$

observations on G endogenous variables are collected in a matrix **Y** while **X** contains H exogenous variables. Also **γ** and **β** are vectors of parameters to be estimated, and the vector **u** has spherical gaussian disturbances with $E(\mathbf{u}) = E(\mathbf{X}\mathbf{u}) = \mathbf{0}$. There are L exogenous variables that appear in other linear equations; and because $L \geq G$, exclusion restrictions are sufficient to identify equation (1). A researcher wants to estimate equation (1) only and may have no usable data on the instruments. Since the Jacobian term does not appear in the log likelihood (Davidson and MacKinnon 1993, 644), the NISE estimator simply minimizes

$$F = (\mathbf{Y}\boldsymbol{\gamma} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}\boldsymbol{\gamma} - \mathbf{X}\boldsymbol{\beta}) - \lambda[\boldsymbol{\gamma}^T(\mathbf{Y}^T\mathbf{Y})\boldsymbol{\gamma} - 1] \ . \qquad (2)$$

With standard software that computes the largest squared canonical correlation between **Y** and **X, γ** is estimated by **c,** the canonical coefficients of **Y;** and **β** is estimated by the OLS regression of **Yc** on **X**. A researcher may then choose to renormalize the equation, dividing both sides by an element of **c**. Finally, a pairs bootstrap will approximate the sampling errors of these NISE coefficients.

3. The wage equation

Klein's Model I is a small dynamic macroeconomic model of the U. S. economy. Estimated using annual data from 1921 to 1941, the model has been a standard feature of econometrics textbooks (e.g. Greene 2003, 411-420). In one equation aggregate private-sector wages are a function of GNP, GNP lagged one year, and a trend. Although lagged GNP and the trend are exogenous, the wages and GNP are jointly endogenous so the OLS estimate is potentially biased and inconsistent. Several aggregate variables are available as instruments: government expenditures, taxes, and the one-year lags of profits and the capital stock.

Table 1 displays three estimates of the wage equation. The coefficients for OLS and TSLS are virtually identical. However the instrument set is significantly correlated with the TSLS residuals according to Sargan's J test (p-stat = 0.018) so the TSLS coefficients –and presumably the OLS coefficients-- are problematic.

NISE is not affected by this issue so I use a Hausman test to compare the coefficients of GNP in the OLS and NISE regressions (Greene 2003, 81). The test strongly rejects the null hypothesis that the two coefficients are equal. Of course this conclusion is subject to the caveat that the effective sample size is hardly of the asymptotic order.

4. The supply of business loans

I estimate the U. S. banking industry's supply function for business loans based on monthly data from January 1983 through December 2006 (cp. Maddala 1988, pp. 313-317). The time series, not seasonally adjusted, are from the FRED archive at the Federal Reserve Bank of St. Louis. The log of the total value of loans outstanding is regressed on the prime rate (the "price" variable) and on three included exogenous variables: the 3-month Treasury bill rate, its one-month lag, and the log of total bank deposits. (The banks can of course invest their deposits in Treasury bills as an alternative to business loans.) The demand side of the business-loan market provides two instrumental variables: the corporate bond rate and the log of the industrial production index.

Table 2 shows that the supply-price elasticities for NISE and TSLS do not differ significantly, but they are significantly larger than the OLS supply elasticity – again a likely instance of the latter estimator's simultaneity bias. All three included exogenous variables have the expected signs, and all are statistically significant except the lagged Treasury bill rate in the OLS regression.

The instruments for TSLS are adequate. In the first-stage regression the corporate bond rate's t-statistic is -2.53, and the t-statistic for log industrial production is 7.76. The instruments are also valid: the significance level of Sargan's J test is 0.25. For this data set, where TSLS performs acceptably, its coefficients are very similar to the NISE coefficients.

I did not explore issues of non-stationarity in the data set since it seems unlikely that unit roots can be detected reliably in time series of only 24 years duration (cp. Pindyck 1999, p. 7).

5. The demand for nurses

Drawing on a data base of the Texas Health and Human Services Commission (2002), I estimate the demand curve for nursing services in Texas long-term care facilities. The sample is comprised of 743 for-profit nursing homes licensed by the state in 2002. According to the textbook model of a competitive market, the price of a resource depends on the amount of the resource used in combination with other inputs, and it also depends on the price of the good or service produced–in this case a nursing facility's average revenue per resident day. In conjunction with the supply curve for the resource, this resource-demand function determines the wage rate.

I focus on the demand function for the services of licensed vocational nurses (LVN), also called licensed practical nurses, who have typically completed one or two years of formal training and who work under the supervision of registered nurses (RN) and physicians. In the log-linear model the jointly endogenous variables are the total LVN hours worked during 2002 and the average hourly LVN wage rate. The included exogenous variables are the total hours worked by RN, by nurse's aides (AIDE), and by laundry and housekeeping personnel (L+H) together with the number of beds in the facility and the revenue per resident day. The excluded exogenous variables would presumably be the determinants of the LVN supply curve, e. g., each LVN's age, the number of young children in the family, a spouse's income, and the local cost of living. However, these potential instruments are absent from the data set so I compare OLS and NISE.

In Table 3 the coefficients are statistically significant except for RN hours in the NISE regression. Both regressions show that the demand for LVN hours is inelastic with respect to the hourly wage; but the NISE coefficient for the LVN wage is significantly larger in magnitude than its OLS counterpart, probably a consequence of OLS simultaneity bias.

6. Conclusions and caveats

When a linear model may be subject to simultaneity bias, NISE is proposed as an alternative (or a complement) to IV estimators. This paper has explored simultaneity bias in three equations from microeconomics and macroeconomics.

Obviously the key to successful IV estimation is one or several strong instruments, and *the key to successful NISE estimation is one or several strong included exogenous variables (***X** *in equation 1).*

In this paper I have attributed to simultaneity bias the significant differences between certain pairs of OLS and NISE coefficients. Of course that conclusion cannot be categorical since other specification problems or data issues may also skew the estimates. However I note that NISE is specifically designed to deal with simultaneity bias and is ineffective against bias in other situations where IV is often applied, e. g. a regressor contaminated by measurement error or an omitted regressor. If these issues were predominant in the three data sets, the relevant OLS and NISE coefficients would probably not differ significantly.

For many linear models a pairs bootstrap can produce NISE standard errors, but the bootstrap should be applied to a robust estimator of dispersion like the median absolute deviation or the Qn statistic (Rousseeuw and Croux 1993; Maronna et al. 2006, chapter 2). Besides limiting the distortions due to outlying observations, a robust version of the standard error is required since the NISE coefficients are not guaranteed to have finite second moments, as Anderson (2010) explains in the context of LIML.

References

Anderson, T., 2010. The LIML estimator has finite moments ! *Journal of Econometrics* 157: 359-361.

Blankmeyer, E., 2017a. Structural-equation estimation without instrumental variables. Available as arXiv paper 1709.09512 at https://arxiv.org.

Blankmeyer, E., 2017b. Flying without instruments. Social Science Research Network paper 2674667, available at http://www.ssrn.com/en/.

Blankmeyer, E., 2018. Simultaneous-equation estimation with no instrumental variables. *Applied Economics Letters* 25: 1097-1100. DOI: 10.1080/13504851.2017.1397847.

Blankmeyer, E., 2020. NISE estimation of an economic model of crime. Available as arXiv paper 2003.07860 at https://arxiv.org.

Chow, G., 1964. A comparison of alternative estimators for simultaneous equations. *Econometrica* 32: 532-553.

Davidson, R., J. MacKinnon, 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Greene, W., 2003. *Econometric Analysis*, fifth edition. Upper Saddle River NJ: Prentice Hall.

Leamer, E., 2010. Tanatalus on the road to Asymptopia. *Journal of Economic Literature* 24: 31-46.

Maddala, G., 1988. *Introduction to Econometrics*. London: MacMillan.

Maronna, R., R. Martin, V. Yohai, 2006. *Robust Statistics: Theory and Methods*. Chichester, England: John Wiley & Sons.

Pindyck, R., 1999. The long-run evolution of energy prices. *The Energy Journal* 20: 1-27.

Rousseeuw, P., C. Croux, 1993. Alternatives to the median absolute deviation. *Journal of the American Statistical Association* 88: 1273-1283.

Texas Health and Human Services Commission, 2002. *2002 Cost Report – Texas Nursing Facility*. Austin, Texas.

Table 1. The wage equation in Klein's Model I
private wages in constant dollars, n = 21
(standard errors under coefficients*)

	OLS	**TSLS**	**NISE**
GNP	0.439	0.439	0.346
	0.032	0.040	0.042
GNP lagged 1 year	0.146	0.147	0.235
	0.037	0.043	0.026
linear trend	0.130	0.130	0.153
	0.032	0.032	0.037
Durbin-Watson	1.958	1.965	2.152

* Standard errors for OLS and TSLS are the conventional i.i.d. estimates.
The NISE standard errors are computed as in Blankmeyer (2017a, p. 12).

Table 2. The business-loan supply function

ln loans, n = 287

(standard errors under coefficients *)

	OLS	NISE	TSLS
prime rate	0.173	0.395	0.349
	0.061	0.079	0.055
treasury bill rate	-0.123	-0.187	-0.174
	0.062	0.056	0.041
treasury bill rate	-0.027	-0.189	-0.155
lagged one month	0.039	0.054	0.054
ln total bank deposits	0.678	0.495	0.533
	0.120	0.095	0.095

* For OLS and TSLS, the standard errors are heteroskedasticity- and autocorrelation-consistent (HAC), Newey-West version. A stationary block bootstrap estimates the standard errors for the NISE regression.

	NISE - OLS	TSLS-NISE	TSLS-OLS
	bootstrap standard errors under differences in coefficients)		
prime rate	0.222	-0.046	0.176
	0.068	0.074	0.063

Table 3. The LVN demand function

ln LVN hours, n = 743

standard errors are shown under coefficients*

	OLS	NISE	NISE -OLS
ln LVN hourly	-0.231	-0.710	-0.479
wage	0.086	0.111	0.065
ln number of beds	0.167	0.220	
	0.040	0.040	
ln RN hours	-0.050	-0.041	
	0.023	0.024	
ln aide hours	0.629	0.619	
	0.039	0.041	
ln L+H hours	0.229	0.237	
	0.042	0.044	
ln revenue per	0.406	0.486	
resident-day	0.068	0.067	
R-squared	0.784		
largest squared canonical correlation		0.797	

*For NISE and the NISE-OLS difference, the standard error is the Qn statistic computed from a pairs bootstrap.
For OLS the usual i.i.d standard errors are reported.