Explorations in NISE Estimation

Blankmeyer, Eric

April 2021

Explorations in NISE Estimation

Eric Blankmeyer

Email eb01@txstate.edu

September 2022

Abstract. Ordinary least squares, two-stage least squares and the NISE estimator are applied to three data sets involving equations from microeconomics. 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).

"It is remarkable that in many IV studies, the discussion and justification of conditional IV independence does not pay much attention to the time period in which the control variables are measured, i.e. whether this happens prior to instrument assignment or at a later point. In particular, there seems to exist a wide spread consensus that it is reasonable to use IV methods in cross-sectional data, where outcomes and controls are measured in the same period. [However] …credible controls need to be measured prior to treatment assignment. Otherwise, they might be affected by the treatment so that conditioning on them likely introduces selection bias" (Deuchert and Huber 2007, 411-412.)

An alternative to IV is the Non-Instrumental Simultaneous-Equation (NISE) estimator, which is applied in this paper to the consistent estimation of three equations from microeconomics. 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, and the appendix contains a more detailed derivation. In section 3 a supply function is estimated for business loans; the derived demand for nursing services is the subject of section 4; and in section 5 a spatial demand function for houses is developed. 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 $E(\mathbf{u}) = E(\mathbf{Xu}) = \mathbf{0}$. 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 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 1 shows that the supply-price elasticities for NISE and TSLS do not differ significantly, but they are significantly larger than the OLS supply elasticity – 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).

"Simultaneity is a concern in much of empirical economic analysis. One approach that has been employed to avoid the problems associated with simultaneity is to replace the suspect explanatory variable with its lagged value" when instrumental variables are unavailable or perform poorly (Reed 2015, 897). The author mentions a dozen examples of this strategy and remarks that it "is common across a wide variety of disciplines in economics and finance. Many appear in top journals including the *American Economic Review*, the *Journal of Finance*, the *Economic Journal*, and the *Journal of Banking & Finance*, and are highly cited" (ibid).

Using analytical methods and simulations, Reed shows that the replacement strategy does not in fact solve the simultaneity problem: the estimated parameters are still biased and inconsistent. For example, if the supply function for business loans is estimated by OLS when the prime rate has been replaced by its one-month lag, the coefficient is 0.156, very similar to the biased estimate in Table 2 -- 0.173. NISE may often be a better option in the absence of good instrumental variables.

4. 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 824 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 2 the coefficients are statistically significant except for RN hours. 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.

5. A spatial demand function for houses

I estimate a conventional partial-equilibrium demand function for houses. The price is specified in terms of dollars per square foot of living space and the quantity in terms of total square feet of living space, both variables in logarithms. The spatial data set "house" (Bivand et al., 2020) provides information on 25,357 single-family homes sold in Lucas County, Ohio between 1993 and 1998. The demand for a house probably depends not only on its price and features but also on the typical prices and features of other homes in the neighborhood. Given the geographic coordinates of a particular house, spatial software identifies the k nearest houses and reports their average price and attributes. Specifically the algorithm produces a vector Wprice of 25,357 observations whose i-th element is the average log price of the k houses nearest house i. Likewise the i-th row of a vector Wlotsize is just the average log lot size of the same k neighbors of house i. LeSage and Pace (2009, chapter 1) have additional discussion of the W operator. Choosing k –the size of the neighborhood— is an exercise in model selection with non-nested hypotheses. LeSage and Pace (2014) explain why the model-selection criteria may not be very sensitive to the choice of k.

The quantity variable is to be regressed on the house's own price, its age when sold, the year of sale, the number of bedrooms, the total number of rooms, and the log of the property's lot size; additional regressors are Wprice and Wlotsize with k = 200 neighbors. In a textbook model of supply and demand the OLS estimate of the demand elasticity is expected to be biased and inconsistent because it ignores the simultaneous determination of price and quantity (e.g., Greene 2003, 378-379). Potential instruments for the housing model include supply-side variables like the costs of land, labor and materials; but these are absent from the data set so I compare OLS and NISE. The results are displayed in Table 3, where each regressor has the expected sign and is statistically significant. But OLS indicates that housing demand is quite inelastic (-0.090) while NISE estimates that the demand is moderately inelastic (-0.659). The implausible OLS estimate may well reflect simultaneity bias.

The coefficient of W(ln price) is a kind of substitution elasticity: in this partial-equilibrium context, it measures how the demand for a house with specific features responds, *ceteris paribus*, to variations in the average price of neighboring houses. Again the NISE estimate is notably larger than the OLS estimate.

This microeconomic demand function is not be confused with the *hedonic pricing model* often used in spatial econometrics (e.g., LeSage and Pace 2009, 42-43.) In particular the hedonic model lacks a continuous-valued quantity variable and therefore does not produce conventional estimates of price elasticity and substitution elasticity.

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. I have explored simultaneity bias in three equations from microeconomics. The key to successful IV estimation is obviously 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.

Bivand, R., J. Nowosad, R. Lovelace, 2020. *Datasets for Spatial Analysis* (R package "spData") available at https://cran.r-project.org/.

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.

Deuchert, E., M. Huber, 2017. A cautionary tale about control variables in IV Estimation. *Oxford Bulletin of Economics and Statistics* 79: 411-425.

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.

LeSage, J.,R. Pace, 2009. *Introduction to Spatial Econometrics*. London: CRC Press.

LeSage, J.,R. Pace, 2014. The biggest myth in spatial econometrics. *Econometrics* 2:217-249. https://doi.org/10.3390/econometrics2040217

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.

Reed, R. 2015. On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics and Statistics* 77: 1273-1283.

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

Theil, H., 1971. *Principles of Econometrics*. New York: John Wiley & Sons.

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

Appendix: a derivation of the NISE maximum likelihood estimator (mle)

In equation (1), the covariance matrix of **Y**, denoted **Θ**, has rank G and is consistently estimated by its usual sample counterpart, denoted **V**. Likewise the covariance matrix of **X**, denoted **ψ**, has rank H. When the NISE log-likelihood is concentrated by "partialing out" **X** from each column of **Y, t**he residuals have a covariance matrix **∑**, of rank G, which is consistently estimated by its sample counterpart, denoted **S**. Then the variance of the jointly endogenous variables can be decomposed into the residual variance and the "regression" variance, say

$$\boldsymbol{\gamma}^T\boldsymbol{\Theta\gamma} = \boldsymbol{\gamma}^T\boldsymbol{\Sigma\gamma} + \boldsymbol{\delta}^T\boldsymbol{\psi\delta} \;. \qquad \text{(A1)}$$

Now NISE bears a formal resemblance to LIML, whose derivation is provided by Davidson and MacKinnon (1993, 644-649) and by Theil (1971, 502-504 and 679-686). LIML and NISE maximize similar constrained Gaussian log-likelihoods and have similar normalizations. For NISE, the normalization constraint is $\boldsymbol{\gamma}^T\boldsymbol{\Theta\gamma} = 1$ (Chow 1964, 533-537 and 542-543; cp. Theil 1971, 684 and Anderson 2010, 359-361). However, NISE does not use the other LIML constraints, which involve the excluded exogenous variables **W**, i. e., the instruments. These latter constraints increase the asymptotic efficiency of LIML relative to NISE, but the premise of NISE is that valid observations on the instruments are not available.

In short, a derivation of NISE as mle follows the derivation of LIML if one omits every term involving variables excluded *a priori* from the equation of interest. Given a random sample of n observations **X** and **Y**, maximization of the NISE log likelihood is equivalent to the minimization of

$$F^* = 0.5n\log|\boldsymbol{\Sigma}| + 0.5\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - 0.5\rho(\boldsymbol{\gamma}^T\boldsymbol{\Theta\gamma} - 1) \,, \qquad \text{(A2)}$$

where ρ is a Lagrange multiplier (cp. Theil 1971, 679). **S** is a consistent estimator of **∑**, and this substitution means that the second term in (A2) is a constant, not dependent on **∑.** When (A1) is applied to the third term of (A2) and $\partial F^*/\partial\boldsymbol{\Sigma}$ is set equal to zero, it follows that

$$n\boldsymbol{\Sigma}^{-1} - \rho\boldsymbol{\gamma\gamma}^T = \mathbf{0} \qquad \text{(A3)}$$

(e.g., Theil 1971, 31-32 or Greene 2003, 839-840).

Premultiplying (A3) by $\boldsymbol{\Sigma}$ and postmultiplying by $\boldsymbol{\Theta\gamma}$,

$$\boldsymbol{\Theta\gamma} - (\rho/n)\boldsymbol{\Sigma\gamma\gamma}^T\boldsymbol{\Theta\gamma} = \mathbf{0} \tag{A4}$$

or $$[\boldsymbol{\Theta} - (\rho/n)\,\boldsymbol{\Sigma}]\boldsymbol{\gamma} = \mathbf{0} \tag{A5}$$

or $$[\boldsymbol{\Sigma} - \lambda\boldsymbol{\Theta}]\boldsymbol{\gamma} = \mathbf{0} \;, \tag{A6}$$

where $\lambda = n/\rho$. The corresponding determinantal equation is

$$|\boldsymbol{\Sigma} - \lambda\boldsymbol{\Theta}| = |\,\boldsymbol{\Sigma\Theta}^{-1} - \lambda\mathbf{I}\,| = 0 \quad ; \tag{A7}$$

and minimization requires the smallest characteristic value,

$$\lambda_{min} = \boldsymbol{\gamma}^T\boldsymbol{\Sigma\gamma} \,/\, \boldsymbol{\gamma}^T\boldsymbol{\Theta\gamma} \;, \tag{A8}$$

which is a real positive fraction.

To operationalize the NISE estimation procedure, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Theta}$ are replaced by their consistent estimators, **S** and **V** respectively. Then $\boldsymbol{\gamma}$ is estimated by **c,** the characteristic vector corresponding to λ_{min}. Finally $\boldsymbol{\beta}$ is estimated by the OLS regression of **Yc** on **X.** However, as mentioned in section 2 of this paper, it will usually be more convenient to process **X** and **Y** using standard software for canonical correlation.

Table 1. 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.

bootstrap standard errors under differences in coefficients)

	NISE - OLS	NISE-TSLS	TSLS-OLS
prime rate	0.222	0.046	0.176
	0.068	0.074	0.063

Table 2. Estimates of the LVN demand model
(the dependent variable is log LVN hours)
standard errors are shown under coefficients *
n = 824

	OLS	**NISE**	**NISE-OLS**
log LVN hourly wage	-0.396	-0.649	-0.253
	0.104	0.120	0.059
log number of beds	0.158	0.192	
	0.040	0.036	
log RN hours	0.045	0.045	
	0.033	0.033	
log aide hours	0.669	0.683	
	0.072	0.070	
log L+H hours	0.138	0.143	
	0.058	0.055	
log revenue per	0.350	0.384	
resident-day	0.075	0.074	
R-squared	0.827		
largest squared			
canonical correlation		0.832	

* HAC standard errors with Newey-West /Bartlett window are reported for all regression coefficients except the the NISE coefficient for log LVN hourly wage and its difference from the corresponding OLS coefficient, which are computed from a pairs bootstrap using the robust Qn estimate of scale.

Table 3. House demand function
(the dependent variable is ln square foot of living space)
standard errors are shown under coefficients*
n = 25,357

	OLS	**NISE**
ln price / square foot of	-0.090	-0.659
living space	0.005	0.006
W(ln price)	0.081	0.618
	0.007	0.009
number of bedrooms	0.047	0.428
	0.004	0.005
ln lotsize	0.091	0.151
	0.004	0.006
number of rooms (total)	0.175	0.155
	0.002	0.003
age	-0.281	-0.478
	0.011	0.017
year of sale	0.003	0.023
	0.001	0.001
W(ln lotsize)	-0.008	-0.092
	0.005	0.008
R-squared	0.715	0.791

*HAC standard errors with Newey-West/Barlett window for all coefficients except the NISE price elasticity, which is the Qn statistic computed from a pairs bootstrap.