E C O N O M I C S   B U L L E T I N

A note on the comparative performance of the Zheng and Elisson−Elisson tests for omitted variables in regression models

Lawrence Dacuycuy
Graduate School of Economics, Kyoto University

Abstract

In this study, we compare the performance of the Zheng (1996) and Elisson−Elisson (2000) tests for omitted variable problems in parametric regression models. The study finds that the Elisson and Elisson test has better finite sample performance relative to the Zheng test. The results also confirm the suitability of the Elisson and Elisson test for testing models in multi−−dimensional settings.

Financial and administrative support from the Japan International Cooperation Agency (JICA) and Japan International Cooperation Center (JICE) are gratefully acknowledged. I would also like to thank my supervisor, Prof. Kimio Morimune and Prof. Yatchew for allowing me to use and modify his Splus scripts for conducting the simulation exercises.
Citation: Dacuycuy, Lawrence, (2005) "A note on the comparative performance of the Zheng and Elisson−Elisson tests for omitted variables in regression models." *Economics Bulletin,* Vol. 3, No. 21 pp. 1−6
Submitted: March 1, 2005. **Accepted:** May 2, 2005.
URL: http://www.economicsbulletin.com/2005/volume3/EB−05C10002A.pdf

1. Introduction

Developments in nonparametric consistent tests have provided a stream of econometric tools to ascertain the validity of economic models. Recent applications, however, have only focused on establishing the consistency of a variety of procedures using Monte Carlo simulation methods. Miles and Mora (2003) used various smoothing and non–smoothing based test procedures to validate Mincerian wage equations. They also conducted simulation studies to determine the sensitivity of various procedures to heteroskedasticity which is common in cross–sectional studies. However, aside from consistency studies, it is also of interest to know the performance of the said tests relative to other specification errors. Elisson and Ellison (2000), henceforth referred to as EE, noted that the tests may be applied to verify specific forms of specification errors.

To our knowledge, no study has been done comparing the performance of the Zheng (1996) and the EE test for omitted variables. The Zheng test has an unsatisfactory property of ranking last in simulation studies by Miles and Mora despite its purported good performance in Zheng. Dacuycuy (in press) evaluated the finite sampling properties of the Zheng test for specific specification errors like omitted and irrelevant variable problems. Based on the said study, the Zheng test performs well in testing for omitted variables but performs poorly in testing for irrelevant variables. Using the same simulation design as in Dacuycuy (in press), this study provides comparative empirical evidence regarding the performance of the the Zheng and EE tests in testing for omitted variable problems.

The paper is organized as follows: Section 2 briefly reviews the features of the Zheng and EE tests. Section 3 details the design of the tests. Section 4 analyzes the comparative performance of the two tests and the last section concludes.

2. Consistent tests for omitted variables

Consider a regression model, $E[y_i|x_i] = h(x_i;\theta)$. Following Fan and Li (1996), the true model that justifies omitting other regressors occurs when the expectation of y conditional on $Z \subset X$ is just the expectation of y when the conditioning variable is X. This means that whether a group of insignificant variables (W), where $W \cup Z = X$, is included or not, the conditional expectation may not be affected.

Two independent but closely related tests have been developed. The Zheng test evaluates the conditional moment $E[\epsilon E[\epsilon|X]f(x)]$, where ϵ corresponds to the residuals of a root – n consistent regression using kernel based smoothing techniques. The sample analogue of the said population moment is a U–statistic defined in Zheng as

$$V_n = \frac{1}{n(n-1)} \sum_{i=1} \sum_{j \neq i} \frac{1}{h^k} K\left(\frac{x_i - x_j}{k}\right) e_i e_j \tag{1}$$

where e, h, K and k denote the residual, smoothing or bandwidth parameter, non–negative kernel function and the dimension of X, respectively. The consistent variance estimator for the test is given by

$$\hat{\Sigma}_n = \frac{2}{n(n-1)} \sum_{i=1} \sum_{j \neq i} \frac{1}{h^k} K^2\left(\frac{x_i - x_j}{k}\right) e_i^2 e_j^2 \tag{2}$$

Given the norming factor of $nh^{d/2}$, $T_n = V_n/(\hat{\Sigma}_n)^{1/2} \sim N(0,1)$. Being a consistent test, if the null is invalid, then $T_n \to \infty$. The test statistic is asymptotically normally distributed.

The EE approach exploits information provided by residual based quadratic forms. As stated in their work, the idea is that misspecification manifests itself in spatial correlation among residuals.

The test statistic is written as

$$T_n^{EE} = \frac{e'W_{ij}e}{\sqrt{2}\mathbf{s}(e'W_{ij}e)} + \frac{\sum_{k=0}^{d} \beta_k}{\sqrt{2}\mathbf{s}(W_{ij})} \tag{3}$$

The first term represents the quadratic form based on the model's residuals. $\mathbf{s}(e'W_{ij}e)$ refers to the spectral radius of the quadratic form which acts as the estimator for the variance. The weight matrix, W, is computed similar to Zheng except that it is normalized. The second term is the finite sample correction proposed by EE when the specification error of interest is omitted variable problem. As in EE, β_k is the coefficient pertaining to the k^{th} regressor under the null in the regression of $WX_{.k}$ on X, where the first column consists of 1s. β_0 is the estimate for the intercept in the regression of $WX_{.1}$ on X. Similar to the Zheng test statistic, the EE test statistic converges to the normal distribution and consistent in that $T_n^{EE} \to \infty$ if the null is invalid.

Note that the EE test introduces a finite sample correction term to address a specific misspecification problem while the Zheng test can be readily applied without major modifications. Both tests, however, are easy to implement.

3. Investigating finite sample performance: Omitted variables

The investigation applies the similar design used in Dacuycuy (in press). We define the following variables and their respective distributions. Let $X_1 \sim N(0,1)$, $X_2 = \frac{\zeta_1 \times \zeta_2}{2}$ and $\varepsilon \sim N(0,1)$, where $\zeta_i \sim N(0,1)$ for $i = 1,2$. The omitted variable Z has a standard normal distribution. The number of replications will be maintained at 1000 and following Zheng, the set of sample sizes includes 100 to 700.

Simple bandwidth selection rules from the work Miles and Mora will be employed. Miles and Mora used a bandwidth which is represented by $\lambda_{mm} = k\hat{\sigma}_x n^{-1/7}$ for the models with 3 continuous variables, where $\hat{\sigma}_x$ is the sample standard deviation of a vector of observations, x. To facilitate the test, the test statistic would be computed following a set of different values for the constant k.

For both tests, the multivariate representation of the kernel function will be the multiplicative or product kernel which is expressed as $K(u) = K(u_1) \times K(u_2) \times \cdots \times K(u_k)$, where $u_k = \frac{(x_{ik} - x_{jk})}{h_k}$ and $K(u_k) = \frac{15}{16}(1-u^2)^2$ for $u_k \in [-1,1]$. The Zheng test uses the raw kernel weights for the residuals. For the EE test, the row sum of elements in the weight matrix sums to 1.

The statistical hypotheses for examining the test performance are

$$\begin{aligned}
H_0 : y_i &= \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2^2 \\
DGP0 : y_i &= 1 + X_1 + X_2 + X_2^2 \\
DGP1 : y_i &= 1 + X_1 + X_2 + X_2^2 + Z \\
DGP2 : y_i &= 1 + X_1 + X_1^2 + X_2 + X_2^2 + Z + Z^2 + Z^3 + Z^4
\end{aligned}$$

In DGP1, y is linear in the omitted variable Z. On the other hand, DGP2 renders the relationship nonlinear.

4. Simulation results

Comparative results in table 1 indicate that the EE test has superior estimated powers relative to that of the Zheng test when it comes to detecting an omitted variable problem. In fact, the EE test attains maximum power in testing the null against DGP1 and registers high estimated powers in testing the null against DGP2. Consistent with the empirical literature, power depends on the bandwidth parameter except in alternative DGP1 under the EE test wherein the rates of rejection are uniform across possible bandwidth choices. In the Zheng test, a higher bandwidth results in greater power.

However, the estimated sizes still underestimate their respective nominal sizes (Li, 1999). Relative to the EE test, the Zheng test registers favorable nominal sizes. The results may be affected by the increase in the dimension of the regressor vector. It is also observed that the nominal size in the Zheng test generally respond negatively to increasing k, which affect the bandwidth λ. The same is true for the EE test.

5. Concluding remarks

The study compares the finite sample performance of the Zheng and EE tests when the specification error concerns the omission of relevant variables in parametric regression. Consistent with studies on finite sample performance of the said tests, the results point to the superiority of the EE test in terms of estimated power. It also confirms the suitability of the EE test for testing models in multi–dimensional settings as evidence in the study by Miles and Mora. However, the estimated sizes are still undersized.

A possible line of investigation is to determine a suitable finite sample correction factor for examining irrelevant variable problems. Moreover, the roles of alternative bandwidth selection schemes as well as distribution approximation methods on the respective tests' power performance may also be examined. These empirical matters are left for future research.

Acknowledgements

Financial and administrative support from the Japan International Cooperation Agency (JICA) and Japan International Cooperation Center (JICE) are gratefully acknowledged. I would also like to thank my supervisor, Prof. Kimio Morimune and Prof. Yatchew for allowing me to use and modify his Splus scripts for conducting the simulation exercises.

References

Bierens, H. (1990) A consistent conditional moment test of functional form, Econometrica, 58, 1443 – 1458.

Dacuycuy, L. (in press) On the finite sampling properties of the Zheng test for omitted and irrelevant variable problems. Forthcoming in Applied Economics Letters.

Ellison, G. and Ellison, S.F. (2000) A simple framework for nonparametric specification testing, Journal of Econometrics, 96, 1 – 23.

Fan, Y., Li, Q. (1996) Consistent model specification tests: omitted variables, parametric and semiparametric functional forms, Econometrica, 64, 865–890.

Li, Q. (1999) Consistent model specification tests for time series econometric models, Journal of Econometrics, 92, 101–147.

Li, Q. and S.Wang (1998) A simple consistent bootstrap test for a parametric regression function, Journal of Econometrics, 87, 145 - 165.

Miles, D. , Mora, J. (2003) On the performance of nonparametric specification tests in regression models, Computational Statistics and Data Analysis, 42, 477 –490.

Yatchew, A. (2003) Semiparametric Regression for the Applied Econometrician, Cambridge University Press.

Zheng, J.X. (1996) A consistent test of functional form via nonparametric estimation technique, Journal of Econometrics, 75, 263–289.

Table 1: Comparative performance of tests for omitted variable problems

		Zheng test						Elisson–Elisson test					
		Null		*DGP*1		*DGP*2		*Null*		*DGP*1		*DGP*2	
Constant (k)	Sample Size	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
2.0	100	.007	.040	.207	.311	.391	.475	.006	.014	.999	1.00	.893	.951
	200	.014	.049	.313	.428	.609	.661	.008	.017	1.00	1.00	.906	.946
	500	.014	.047	.627	.715	.830	.856	.007	.016	1.00	1.00	.958	.986
2.5	100	.012	.035	.304	.409	.507	.577	.004	.006	1.00	1.00	.921	.972
	200	.014	.051	.412	.510	.645	.690	.008	.017	1.00	1.00	.934	.970
	500	.011	.042	.645	.727	.827	.850	.010	.024	1.00	1.00	.968	.987
3.0	100	.016	.034	.486	.592	.691	.749	.003	.006	1.00	1.00	.970	.989
	200	.014	.043	.559	.645	.742	.779	.005	.008	1.00	1.00	.969	.987
	500	.010	.036	.712	.778	.847	.867	.005	.016	1.00	1.00	.975	.992
3.5	100	.014	.025	.753	.815	.874	.899	.002	.004	1.00	1.00	.975	.985
	200	.017	.039	.760	.805	.868	.889	.006	.007	1.00	1.00	.988	.997
	500	.018	.036	.815	.850	.895	.905	.003	.009	1.00	1.00	.985	.994