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

Robust exogeneity tests in the presence of outliers

Sunil Sapra
California State University, Los Angeles

Abstract

Exogeneity testing is studied in the presence of outliers in response variables. Robust tests based on least absolute deviations (LAD) and M estimators are proposed and illustrated with an application to Mroz (1987) data. Our simulation results show that the proposed robust tests outperform the traditional Hausman test for exogeneity in terms of empirical power in the presence of outliers in response variables. Nevertheless, unlike the conventional Hausman test, which is undersized, the empirical size of the LAD-based exogeneity test exceeds its nominal size.

I am grateful to the associate editor, Professor Timothy Conley and an anonymous referee for several helpful comments.
Citation: Sapra, Sunil, (2006) "Robust exogeneity tests in the presence of outliers." *Economics Bulletin*, Vol. 3, No. 29 pp. 1-6
Submitted: July 18, 2006. **Accepted:** November 13, 2006.
URL: http://economicsbulletin.vanderbilt.edu/2006/volume3/EB-06C30036A.pdf

1. Introduction

Applications of exogeneity testing are common in social sciences. The tests help researchers decide whether to use instruments for any of the regressors to ensure the consistency of estimates. Testing exogeneity also helps ensure that the more efficient estimator between the ordinary least squares (OLS) and instrumental variables (IV) estimator is used given that OLS is more efficient than IV if the regressors are exogenous. Hausman's test for exogeneity (Hausman (1978)) is based on comparing the OLS and two stage least squares (2SLS) estimators of regression coefficients and determining if the differences are statistically significant. Although these tests perform well in the absence of outliers, they may not have good statistical properties in the presence of outliers. Alternative robust tests based on robust estimation methods may provide desirable alternatives to tests based on non-robust estimators. Koenker and Bassett (1982) developed robust hypothesis tests for hypotheses based on least absolute deviations (LAD) estimators.

The purpose of this paper is to present two robust tests for exogeneity hypotheses in simultaneous equation models and to compare their power and size properties to those of the conventional exogeneity tests in the presence of outliers in response variables. Our tests are modifications of the Hausman test for exogeneity and are based on LAD (Koenker and Bassett (1978)) and M-estimators (Huber (1981)) instead of least squares estimators. Peracchi (1991) showed that a general class of multidimensional tests based on M-estimators inherit the efficiency and robustness properties of the estimators on which they are based. Our robust exogeneity tests based on LAD or M estimators are motivated by this important finding and can similarly lead to more reliable inference in the presence of outliers in the data on response variables. Section 2 presents the proposed robust exogeneity hypothesis tests and section 3 presents the LAD-F test and the M-F test on which they are based. Section 4 provides results of some simulation experiments. Section 5 illustrates the tests with an application to an economic data set and compares a robust exogeneity test to its non-robust counterpart and section 6 concludes.

2. Robust Exogeneity Tests

In what follows, we present exogeneity tests based on LAD and M-estimators as simple modifications of a regression-type Hausman exogeneity test presented in Hausman (1978, 1983) and Wooldridge (2002).

Consider the following structural equation in a simultaneous equation system.

$$y = x_1'\beta_1 + \beta_2 x_2 + u, \tag{1}$$

where x_1 is a $(k_1 x1)$ vector of exogenous regressors, x_2 is a regressor, which is potentially endogenous, u is the error term and is iid with a common c.d.f. F. Assume that the regularity conditions for the validity of F_{LAD} and F_M tests to be discussed in section 3 hold (see Koenker and Bassett (1978) and Huber (1981)). Let x be a $(kx1)$ the vector of all exogenous variables in the equation system. Assume also that

$$E(x'u) = 0. \tag{2}$$

The proposed robust tests for exogeneity of x_2 are based on an artificial compound regression model and are conducted as follows.

$Step 1$: Estimate the regression (reduced form for x_2)

$$x_2 = x\pi + v \tag{3}$$

by LAD or M procedures and compute the residuals $\hat{v}$.

$Step 2$: Estimate the following augmented regression by the same robust estimation procedure as used in step 1.

$$y = x_1'\beta_1 + \beta_2 x_2 + \gamma\hat{v} + u^*. \tag{4}$$

Then a test of exogeneity of x_2 is a test of $H_0: \gamma = 0$ against $H_1: \gamma \neq 0$. The test is conducted using robust versions of the F - tests based on LAD first proposed by Koenker and Basset (1982) or those based on M estimators presented in Birkes and Dodge (1993). These tests are discussed next.

3. The F-tests based on LAD and M Estimators

Consider the linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \varepsilon. \tag{5}$$

The test statistic for testing $H: \beta_{q+1} = ... = \beta_p = 0$ based on LAD proposed by Koenker and Bassett (1982) and presented in Birkes and Dodge (1993), which is used in step 2 in section 2 above is

$$F_{LAD} = (SAR_{restricted} - SAR_{unrestricted})/(p-q)(\hat{\tau}/2), \tag{6}$$

where $(p-q)$ denotes the number of restrictions, $SAR = \sum|\hat{\varepsilon}_i|$ and

$$\hat{\tau} = \sqrt{m}\,[\hat{\varepsilon}_{(k_2)} - \hat{\varepsilon}_{(k_1)}]/4, \tag{7}$$

k_1 and k_2 are the closest integers to $(m+1)/2 - \sqrt{m}$ and $(m+1)/2 + \sqrt{m}$ respectively, $m = n - (p+1)$ and $\hat{\varepsilon}_{(k_1)}$ and $\hat{\varepsilon}_{(k_2)}$ are k_1 - *th* and k_2 - *th* LAD residuals when these residuals are arranged in increasing order. Furthermore, the F - test in section 2 above, based on Huber's M estimator in Birkes and Dodge (1993) under conditions therein is

$$F_M = (STR_{restricted} - STR_{unrestricted})/(p-q)(\hat{\lambda}), \tag{8}$$

where $STR = \sum \rho(\hat{\varepsilon}_i)$, $\hat{\lambda} = (n/m)\sum \hat{\varepsilon}_i{}^*/(n-p-1)$, ρ is the Huber function defined as

$$\begin{aligned} \rho(\hat{\varepsilon}) &= \hat{\varepsilon}^2 \text{ if } |\mathrm{k}| \leq \hat{\varepsilon}, \\ &= 2k|\hat{\varepsilon}| - k^2 \text{ if } \hat{\varepsilon} < -k \text{ or } \mathrm{k} < \hat{\varepsilon}, \end{aligned} \tag{9}$$

$\hat{\varepsilon}^* = \hat{\varepsilon}$ if $|\hat{\varepsilon}| \leq 1.5\hat{\sigma}$, $\hat{\varepsilon}^* = -1.5\hat{\sigma}$ if $\hat{\varepsilon} < -1.5\hat{\sigma}$, $\hat{\varepsilon}^* = 1.5\hat{\sigma}$ if $\hat{\varepsilon} > 1.5\hat{\sigma}$ and $\hat{\sigma} = 1.483\text{MAD}$, where MAD is the median of the absolute deviations $|\hat{\varepsilon}_i|$ and m is the number of M residuals between $-1.5\hat{\sigma}$ and $1.5\hat{\sigma}$. Both F_{LAD} and F_M follow $F(p-q, n-p-1)$ distribution in small samples and $\chi^2(p-q)$ in large samples under H (see Birkes and Dodge (1993)).

4. Monte Carlo Results

In this section, we present results on the empirical power and size properties of robust and non-robust tests focusing on LAD-F test exclusively.

4.1 Empirical Power Properties of Robust and Non-robust Exogeneity Tests

We adapt the design of an experiment in Rousseeuw and Leroy (1986) (pp. 67-68). Samples of various sizes were generated according to the following model.

$$y_1 = 2y_2 + u_1, \tag{10}$$

$$y_2 = 3y_1 + 4x_1 + 5x_2 + u_2. \tag{11}$$

In what follows, our objective is to compare the power properties of a robust exogeneity test for the exogeneity of y_2 based on the LAD-F test and the conventional Hausman test.

<u>Experiments with low $\rho = \text{Corr}(y_2,u_1)$ (Small departures from H_0)</u>

One hundred samples of sizes 50, 100, 200 and 500 were generated. We generated 60% of the observations according to the model in (10) and (11), where $x_1 \sim U[1,3]$, $x_2 \sim U[4,6]$ and $u = (u_1,u_2)'$ is bivariate normally distributed as $N(0, \Sigma)$ with $\sigma_{11} = 4$, $\sigma_{22} = 8$, and $\sigma_{12} = \sigma_{21} = 8$. Then 40% new observations were added for which $x_1 \sim U[1,3]$, $x_2 \sim U[4,6]$ as for the first set of observations, but $u = (u_1,u_2)'$ is bivariate normally distributed as $N(0, \Sigma)$ with $\sigma_{11} = 10$, $\sigma_{22} = 100$, and $\sigma_{12} = \sigma_{21} = -10$. This led to 40% outliers in observations on y_1 and y_2 in the pooled sample and generated values of $\rho = \mathrm{Corr}(y_2,u_1)$ close to 0. In the first step, an LAD regression of y_2 on the exogenous variables, x_1 and x_2, and the instrument x_1^2 was estimated and the LAD residuals were computed. In the second step, an LAD regression of y_1 on y_2 and the residuals from the first step was estimated. Exogeneity of y_2 was tested by testing whether the coefficient of the residuals is zero. The x_1 and x_2 samples were held constant across replications of each experiment. Results on powers of the robustified Hausman test based on the LAD-F test for exogeneity and the conventional Hausman test for exogeneity based on the conventional t-test for small departures from the null hypothesis of exogeneity of y_2 reflected in small values of ρ are displayed in table 1. As the sample size increases, the powers of both the LAD-F test and non-robust Hausman test tend to increase. The LAD-F test consistently outperforms the non-robust Hausman test for all sample sizes. The power of each test is low for samples of sizes 50, 100 and 200 since the instruments used are weak as reflected in the low correlation between y_2 and x_1^2.

Table 1

Empirical Powers of the tests with low values of ρ, 40% outliers and $\alpha = .05$

Sample Size	LAD-F Test	Non-robust Hausman t-test
n=50	.49	.46
n=100	.64	.58
n=200	.76	.74
n=500	.93	.89

Experiments with high $\rho = \mathrm{Corr}(y_2,u_1)$ (Large departures from H_0)

One hundred samples of sizes 50, 100, 200 and 500 were generated so that ρ was extremely high (close to .9) as follows. As in experiments with low ρ values above, $x_1 \sim U[1,3]$ and $x_2 \sim U[4,6]$. However, 60% of the observations in each sample were generated according to the model in (10) and (11) with $u_1 \sim N(100,9)$, $u_2 \sim U[10,16]$ and uncorrelated to which 40% new observations were added. For the new observations, $u_1 \sim U(1,5)$, $u_2 \sim N(500,20)$ and the two were uncorrelated. Nevertheless, y_2 and u_1 were highly correlated in all samples with ρ exceeding .9 in all samples. Exogeneity of y_2 was tested as discussed earlier. Table 2 displays the results on power properties of the two tests for extremely high values of ρ and with 40% outliers. Not surprisingly, both tests perform poorly in this situation. Table 3 displays the results on power properties of the two tests for extremely high values of ρ and with 10% outliers. Surprisingly, the percentage of outliers in the sample appears to have a marginal effect on the performance of both tests if ρ is extremely high. However, the gap between the powers of the two tests appears to increase with the percentage of outliers in observations on y_1 and y_2.

Table 2

Empirical Powers of the tests with high values of ρ, 40% outliers and $\alpha = .05$

Sample Size	LAD-F Test	Non-robust Hausman t-test
n=50	.13	.07
n=100	.16	.12
n=200	.30	.24
n=500	.45	.41

Table 3

Empirical Powers of the tests with high values of ρ, 10% outliers and $\alpha = .05$

Sample Size	LAD-F Test	Non-robust Hausman t-test
n=50	.20	.09
n=100	.26	.12
n=200	.36	.14
n=500	.40	.17

4.2 Empirical Sizes of Robust and Non-robust Exogeneity Tests

One hundred samples of sizes 50, 100, 200 and 500 were generated according to the following model derived from model (10)-(11) by dropping y_1 from the right-hand side of equation (11) and choosing u_1 and u_2 to be uncorrelated.

$$y_1 = 2y_2 + u_1, \tag{12}$$

$$y_2 = 4x_1 + 5x_2 + u_2. \tag{13}$$

For each sample, we generated 60% of the observations according to the model in (12) and (13), where $x_1 \sim U[1,3]$, $x_2 \sim U[4,6]$ and $u = (u_1,u_2)' \sim N(0, \Sigma)$ with $\sigma_{11} = 4$, $\sigma_{22} = 8$, and $\sigma_{12} = \sigma_{21} = 0$. Then 40% new observations were added for which $x_1 \sim U(1,3)$, $x_2 \sim U(4,6)$ as for the first set of observations, but $u = (u_1,u_2)' \sim N(0, \Sigma)$ with $\sigma_{11} = 10$, $\sigma_{22} = 100$, and $\sigma_{12} = \sigma_{21} = 0$. Consequently, the structural equation system (12)-(13) is recursive and y_2 is exogenous. Exogeneity of y_2 was tested as discussed in section 4.1. The results on empirical sizes for nominal significance levels of 5% and 10% are displayed in Tables 4 and 5 respectively.

Table 4

Empirical Sizes of the tests with 40% outliers and nominal $\alpha = .05$

Sample Size	LAD-F Test	Non-robust Hausman t-test
n=50	.10	.02
n=100	.08	.04
n=200	.09	.04
n=500	.08	.06

Table 5

Empirical Sizes of the tests with 40% outliers and nominal $\alpha = .10$.

Sample Size	LAD-F Test	Non-robust Hausman t-Test
n=50	.15	.04
n=100	.16	.07
n=200	.14	.08
n=500	.12	.09

It is clear that the LAD-F test tends to overreject H_0 at both nominal significance levels while the non-robust Hausman test tends to underreject H_0 at both nominal significance levels for all of the sample sizes considered. Furthermore, the empirical size of each test approaches its nominal size as the sample size increases.

5. An Empirical Application: Application to Mroz data

The data used here is from Mroz (1987) and consists of 428 observations. The data on the two possibly endogenous variables, log(wage) and educ, contain 23 and 18 outliers (observations with z-scores of more than 2 in absolute value) respectively. Consider the following example on testing for exogeneity of education in a wage equation for working women discussed in Wooldridge (2002) (pp. 120-122).

$$\log(\text{wage}) = \beta_0 + \beta_1\text{exper} + \beta_2\text{exper}^2 + \beta_3\text{educ} + u \tag{14}$$

where educ is possibly endogenous. Wooldridge (2002) chooses parents' education and husband's education as instruments for educ. Unlike Wooldridge (2002), as discussed in section 2.1 above, we estimate the following regression by LAD instead of OLS in step 1.

$$\text{educ} = \delta_0 + \delta_1\text{exper} + \delta_2\text{exper}^2 + \delta_3\text{motheduc} + \delta_4\text{fatheduc} + \delta_5\text{huseduc} \tag{15}$$

and compute the residuals $\hat{v}$. In step 2, the following augmented regression is estimated by LAD.

$$\log(\text{wage}) = \beta_0 + \beta_1\text{exper} + \beta_2\text{exper}^2 + \beta_3\text{educ} + \gamma\hat{v} + u. \tag{16}$$

Then a test of exogeneity of educ is a test of $H_0 : \gamma = 0$ versus $H_1 : \gamma \neq 0$. The results are in Table 6.

Table 6
Test Statistics for Tests for Exogeneity of education in a wage equation

Null Hypothesis	LAD-F Test	M-F Test	Non-robust Hausman t-Test
H_0: γ=0	0.084912866	.003737	1.65

The extremely low values of the LAD-F and the M-F test statistics (based on Huber's M estimator) under H_0 support the exogeneity hypothesis. However, the conventional non-robust Hausman test based on Hausman regression-type t-test yields a value, which is statistically significant at 10% significance level. Thus, the robust LAD-F and the M-F tests confirm exogeneity of education while the conventional Hausman test rejects it. It is also worth noting that the p-values for the three tests are significantly different and could lead to different inferences.

6. Conclusions

This paper has presented robust tests of exogeneity based on LAD and M estimators and compared their empirical powers and sizes with those of the conventional regression-based Hausman test in the presence of outliers in response variables. The proposed tests appear to have better power properties than their non-robust counterpart in the presence of outliers in response variables. Nevertheless, unlike the conventional Hausman test, which is undersized, the robust test tends to overreject the null hypothesis of exogeneity. In the presence of outliers in response variables as well as regressors, tests based on high-breakdown point estimators such as LTS, LMS and S estimators (Rousseeuw and Leroy (2003)) may lead to improved performance in terms of power and size. These tests can also be extended to testing exogeneity in nonlinear models. Finally, our tests are based on the assumption of spherical errors and will need to be modified if the errors are nonspherical. Robust versions of the heteroscedasticity-robust Hausman tests discussed in Wooldridge (2002) or block-bootstrap corrections of these tests discussed in Li (2006) can be developed. These topics will be pursued in future research.

References

Birkes, D. and Y. Dodge (1993) *Alternative Methods of Regression*. Wiley, New York.
Hausman, J. (1978) Specification Tests in Econometrics, *Econometrica* 46, 1251-1271.
------------- (1983) Specification and Estimation of Simultaneous Equation Models, in *Handbook of Econometrics*, Volume1, ed. Z. Griliches and M. D. Intrilligator. Amsterdam, North Holland, 391-448.
Huber, P. (1981) *Robust Statistics*, John Wiley & Sons, New York.
Koenker, R. and G. Bassett (1978) Regression quantiles, *Econometrica*, **46**, 33-50.
-----------------------(1982) Tests of linear hypotheses and l1 estimation, *Econometrica*, **50**, 1577-1583.
Li, J. (2006) The block bootstrap test of Hausman's exogeneity in the presence of serial correlation, *Economics Letters* **91**, 76-82.
Mroz, T. A. (1987) The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions, *Econometrica* **55**, 765-799.
Peracchi, F. (1991) Robust M-tests, *Econometric Theory*, **7**, 69-84.
Rousseeuw, P. J. and A. Leroy (2003) *Robust Regression and OutlierDetection*, John Wiley & Sons, New York.
Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA.