Validity and Power of the Hausman Test Under Weak Instruments

Anya Demko
Department of Economics, Reed College, Portland, Oregon, 97202, USA
(Dated: May 5, 2012)

This paper examines the validity and power of the Hausman Test under weak instruments. Monte-Carlo simulations are used to generate an instrumented variable of varying strength and varying degrees of correlation with the error of the second stage regression. The validity and power of the Hausman Test is checked under these different circumstances. The Hausman Test is found to be invalid under weak instruments and its power varies depending on instrument strength.

I. INTRODUCTION

Endogeneity of a regressor in Ordinary Least Squares regressions leads to biased estimators. Correlation between a regressor and the error term of a regression is a violation of the Ordinary Least Squares assumption $cov(x, \epsilon) = 0$. A natural question to ask if whether we can test if a suspect variable is in fact endogenous. Jerry Hausman developed a test the relies on the technique of instrumental variables. Since their introduction, instrumental variables have become a popular way to deal with endogenous regressors. Proper instruments are variables that are correlated with an endogenous regressor but are not themselves correlated with the error, ϵ. The general idea of the Hausman test is as follows: If the variable is not endogenous, then asymptotically there should be no difference between the OLS and the IV estimates; they should converge to the same value. The Hausman test operates under the following assumptions:

$$H_0 \; : \; cov(x, \epsilon) = 0$$
$$H_a \; : \; cov(x, \epsilon) \neq 0$$

meaning that rejection of the null hypothesis indicates an endogenous regressor, while failure to reject is support for an exogenous regressor. OLS estimates are more efficient than IV estimates, meaning that if a regressor is exogenous, then OLS is the preferred estimation technique.

This paper asks the question of whether the Hausman Test is valid under weak instruments. To test instrument strength, we regress the endogenous variable against the exogenous variables of the regression and the instruments. If the joint F-statistic of the instruments (or T-statistic if there is only one) is greater than 10, then the instruments are considered strong. This value was determined by Stock and Yogo after taking into account the biases produced by OLS and IV regression. We will use Monte-Carlo simulations to generate a suspect variable which is correlated to varying degrees with its instruments. The Hausman Test is then performed and we will check the distribution of the X^2 statistics to test for validity.

The next question we will address is the power of the Hausman Test to pick up the correlation between an instrumented variable and the error, ϵ, of the second stage regression given the varying strengths of this instrument. To answer these questions, we will generate the different circumstances described above using Monte-Carlo simulations, and test the Hausman Test given the fact that we will know the true answers. This procedure is gone into with more depth below.

II. PROCEDURE

A. Monte-Carlo Simulations

The Monte-Carlo simulations and subsequent tests based off of this generated data were performed with Stata do-files. These files are included in the Appendix. Also included in the do-files are detailed comments for each step. We use a sample size of 1000 for the original data. We know that IV estimates are asymptotically consistent, thus we want to insure that the sample size is large enough to allow for this. Were the sample size not large enough, then this would bias the future results.

The first step of this study is to find the F-statistics associated with different standard deviations used as input to generate the error term for the instrumentation process, or the first stage regression. This is done with the *ChooseSigma* do file in Appendix A.

B. Validity

Once standard deviations associated with instrument strength are determined, we can then proceed with IV regressions and the Hausman Test. This is accomplished with the *HausmanTest* do file in Appendix B. We generate our instruments with no correlation to the error term ϵ, thus we know the null hypothesis of $cov(x, e) = 0$ to be true. First we perform the uninstrumented regression and store the estimates. Then we use Stata's command IVregress which automatically performs a two stage least squares regression and store the instruments from that. Finally we preform the Hausman test to check if these two sets of estimates are the same. We note now that Stata uses the incorrect number of degrees of freedom for the X^2 results it reports. The degrees of freedom should be the same as the *number of suspect endogenous variables.*

Stata will use the number of instruments provided for the variable, which is not necessarily the same. We correct for this in our do-file my manually telling Stata the df to be used.

We simulate the *HausmanTest* do-file with 2000 repetitions. We store the X^2 statistics and their p-values as variables. From the X^2 statistic we generate the variable *fail* which indicates whenever the X^2 stat is over the critical value. We will use this *fail* variable to see if the test fails the expected amount of time.

C. Power

For power, we alter the *HausmanTest* do-file to now have the error term ϵ of the second stage regression correlated with the instrumented variable's error term, ν. This is the *HausmanTestCorrelation* do-file in Appendix D. To do this, we make ϵ a function of ν, and we first use the *Correlation* do file in Appendix C to find the right formula to use. We find the best functions to use to create .25, .5, and .75 correlation are:

$$\begin{aligned} \epsilon &= .25\nu + \text{rnormal}(0, \sigma_\nu) \\ \epsilon &= .6\nu + \text{rnormal}(0, \sigma_\nu) \\ \epsilon &= 1.2\nu + \text{rnormal}(0, \sigma_\nu) \end{aligned}$$

where σ_ν is the standard deviation used to generate the error term ν.

We then simulate the *HausmanTestCorrelation* do-file using 2000 repetitions and report the results similar to the way described in the validity section.

III. RESULTS

The results of the preliminary search for instrument using the ChooseSigma do-file is summarized in Tab **??**.

TABLE I: Strengths of instruments dependent on standard deviation of the generated error term.

	F-Statistic	
Standard Deviation	Median	Mean
5		
10		
15		
20		
25		
30		
35		
40		
45		

We choose to use standard deviations of 10, 15, 25, 35, and 45 to represent our instrument strengths. This provides a range of instruments which are undoubtably strong, strong by the Stock and Yogo standard, statistically significant, borderline statistically significant, and no longer statistically significant, respectively.

TABLE II: Strengths of instruments dependent on standard deviation of the generated error term.

	Fails		H_0: Same
Standard Deviation	Expected	Observed	p-value
10	100	107	.47
15	100	86	.16
25	100	60	0
35	100	42	0
45	100	25	0

A. Validity

Tab. **??** summarizes the result of the test as to whether the null hypothesis is rejected the expected amount of time. This is done by looking at the number of fails. We will compare the difference between the expected and observed numbers of fail. We can see that for strong instruments, the expected and observed number of fails is statistically the same, meaning that a true null hypothesis of exogeneity is rejected 5% of the time as expected. This is not true for the three weak instruments. The observed number of fails is less than the expected for all weak instruments, meaning that a true null hypothesis is rejected *less* than 5% of the time. Note that we also see this downward trend even amongst the strong instruments. The five percent critical values for the weak instruments were

$$\sigma = 25, \quad \sigma = 35, \quad \sigma = 45,$$

which shows a steady decrease, which indicates a general leftward shift of the Hausman X^2 statistics. This means that the Hausman X^2 statistics themselves no longer follow a X^2 distribution. This is further indicated by Fig. **??** and Fig. **??** which are a series of graphs generated by the Stata pchi command, which generates a chi-squared probability plot. This plot graphs a variable against it's position were it part of a chi-squared distribution. If a chi-squared distribution is followed, we expect this relation to be one to one and thus to see a straight line with slope 1. Fig. **??** shows that as instrument strength decreases, the top of the distribution falls away from the projected line, indicating that less of the distribution than expected is located in the tail.

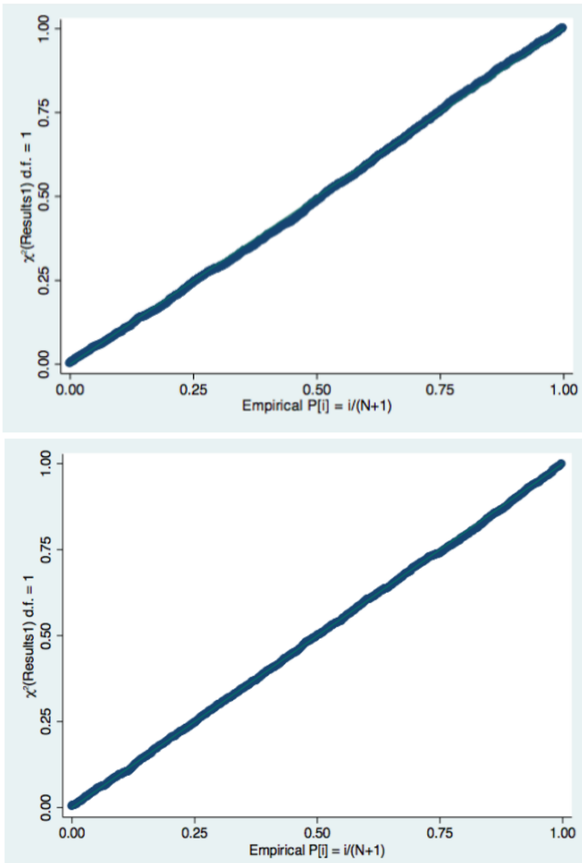

FIG. 1: X^2 Probability Plots for the X^2 statistics generated by Hausman Test for strong instruments. If a X^2 distribution is followed, we expect to a straight line of slope one. This is true for the two strong instruments graphed here.

B. Power

Tab. **??** summarizes the percentage rejections by the Hausman Test for instruments of varying strengths and varying correlation with ϵ.

We find that for strong instruments, all correlations of 50% or more are rejected, however, even for extremely strong instruments, minor correlations of around 25% are not rejected all of the time. As instrument strength decreases, power of the test decreases as well.

IV. INTERNAL AND EXTERNAL VALIDITY

Given the generated nature of the data, internal validity issues are directly controlled based on the inputs to the Monte-Carlo simulations. There are no external validity issues, as this study is generally just looking at the reliability of a statistical test, and leaves any external application of the results up to the future users of the Hausman Test.

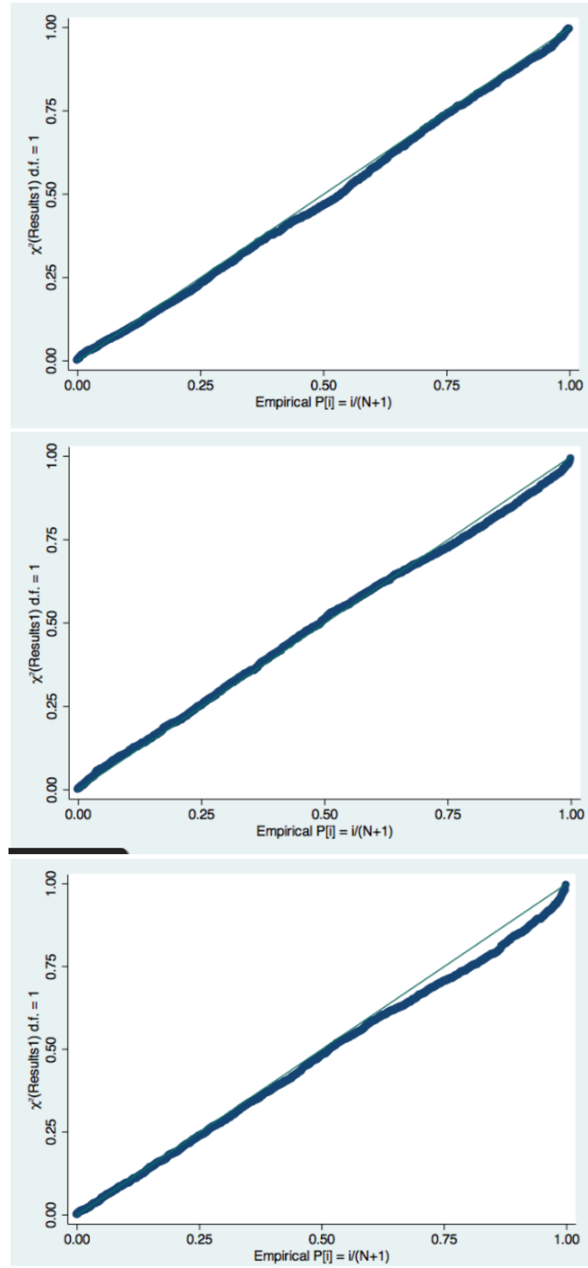

FIG. 2: X^2 Probability Plots for the X^2 statistics generated by Hausman Test for the three weak instruments. We start to see the far end of the graph drop away from the expected line.

V. CONCLUSION

A. Validity

We find that the Hausman Test is not valid for weak instruments. The X^2 statistics generated by Monte-Carlo simulations for weak instruments no longer follow the expected distribution. Further tests which could expound on this would be to find the exact instrument strength for which the failure rate drops below 5% percent. From

TABLE III: Percent failure of the Hausman Test dependent on correlation and instrument strength.

Standard Deviation	Correlation	Fail
	.25	89.5%
10	.5	100%
	.75	100%
	.25	56%
15	.5	100%
	.75	100%
	.25	18%
25	.5	81%
	.75	100%
	.25	7%
35	.5	39%
	.75	92%
	.25	3%
45	.5	16%
	.75	60%

the data we have, it would be reasonable to assume that this occurs somewhere around an F-statistic of 10, which is the generally used threshold for a strong instrument. Also, we could use a greater variety of standard deviations and track the decline of the critical values in more detail. This would give a more comprehensive view of how and when the test becomes invalid.

B. Power

The power of the test is proportional to the strength of the instruments, which is a conclusion to be expected. Interestingly enough though, we find that even for strong instruments, weak correlations between the instrument and the error term are not rejected 100% of the time. Remember: Failure to reject a null hypothesis does not directly imply that the null hypothesis is not true. This is supported by our results. This is a cautionary note to not rely completely on statistics. If theory states that a variable may be endogenous, yet statistical tests cannot detect it, *this does not mean* that endogeneity is not present! The results from this test give light on the levels of correlation needed for the Hausman Test to become reliable, and leaves it up to the econometrician to now decide what levels of potential endogeneity are acceptable.

Interestingly, we note that even for weak instruments, high correlation of at least 75% is detected a reasonable amount of time. However, as just determined in the validity section: the Hausman test becomes invalid for weak instruments, and thus the results of these trial may be called into question.

Further tests would be to make a more detailed study of what level of correlation is needed for the Hausman Test to reject 100% for each instrument strength. This could be done for more variation of strong instruments. Overall, this deeper study would give a very good picture as to how reliable the Hausman Test is. This would allow econometricians to be more informed when making decisions as to the results of the statistical tests they use.