2007

The Hausman test, and some alternatives, with heteroskedastic data

Viera Chmelarova
Louisiana State University and Agricultural and Mechanical College

THE HAUSMAN TEST, AND SOME ALTERNATIVES, WITH HETEROSKEDASTIC DATA

A Dissertation

Submitted to the Graduate School of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Economics

by
Viera Chmelarova
Magister, Comenius University, Slovakia, 1995
M.S., Louisiana State University, USA, 2002
May 2007

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Dr. R. Carter Hill for his guidance, help, support and patience through this entire journey. He has been much more than one would have expected from an advisor. I would also like to thank the remaining committee members: Dr. Eric T. Hillebrand, Dr. M. Dek Terrell, Dr. David M. Brasington and Dr. Bin Li for their valuable comments and suggestions. Special thanks go to my Baton Rouge host family and friends for their continuous support, caring and encouragement. Without them, I would never have made it this far.

TABLE OF CONTENTS

ABSTRACT

The Hausman test is used in applied economic work as a test of misspecification. It is most commonly thought of (wrongly some would say) as a test of whether one or more explanatory variables in a regression model is endogenous. There are several versions of the test available with modern software, some of them suggesting opposite conclusions about the null hypothesis. We explore the size and power of the alternative tests to find the best option. Secondly, the usual Hausman contrast test requires one estimator to be efficient under the null hypothesis. If data are heteroskedastic, the least squares estimator is no longer efficient. Options for carrying out a Hausman-like test in this case include estimating an artificial regression and using robust standard errors, or bootstrapping the covariance matrix of the two estimators used in the contrast, or stacking moment conditions leading to two estimators and estimating them as a system. We examine these options in a Monte Carlo experiment. We conclude that in both these cases the preferred test is based on an artificial regression, perhaps using a robust covariance matrix estimator if heteroskedasticity is suspected. If instruments are weak (not highly correlated with the endogenous regressors), however, no test procedure is reliable. If the test is designed to choose between the least squares estimator and a consistent alternative, the least desirable test has some positive aspects. We also investigate the impact of various types of bootstrapping. Our results suggest that in large samples, wild (correcting for heteroskedasticity) bootstrapping is a slight improvement over asymptotics in models with weak instruments. Lastly, we consider another model where heteroskedasticity is present – the count data model. Our Monte Carlo experiment shows that the test using stacked moment conditions and the second round estimator has the best performance, but which could still be improved upon by bootstrapping.

1. INTRODUCTION

Arguably the most important research tool for the empirical researcher is regression analysis. Since the mid-1970's there has been a great deal of concern for regression models in which some explanatory variables are endogenous, that is, they are correlated with the regression error term. Traditionally such problems had been considered within the context of simultaneous equations models, in which explicit assumptions about endogeneity and exogeneity were required, and multiple equation regression systems were built and estimated. At some point in the development of empirical research such models were abandoned in favor of single equation regressions that *may* have endogenous regressors. The doubt in that statement is present because structural equations for these potentially endogenous regressors were never explicitly specified. We can speculate that economists studying complex microeconometric behaviors were unable, or unwilling, to build systems models that described the data.

One outcome of this change in philosophy was the recognition of the requirement for a statistical test of whether a regression model included endogenous regressors. The usual least squares estimator is inconsistent if there are endogenous regressors. If there are endogenous regressors, and if valid instrumental variables are available, then the instrumental variables, or two-stage least squares, estimator is consistent and should be used. Instrumental variables estimators can be very imprecise relative to the least squares estimator if good instruments are not available, and finding good instruments is no easy task. Thus it is imperative that researchers have a good way of determining when they must use the instrumental variables estimator, and when they do not have to.

In a series of papers by Durbin (1954), Wu (1973, 1974) and Hausman (1978) tests were proposed that can be applied to the problem of detecting endogenous regressors. In this chapter the essentials of the problem are outlined. Chapter 2 summarizes the literature and examines the behavior of various tests in the regression model with homoskedastic errors. This dissertation studies the problem of testing for endogenous regressors in regression models when errors are heteroskedastic. In Chapter 3 asymptotically valid tests are considered in linear regression models with heteroskedastic errors. We explore using Monte Carlo experiments the behavior of alternative tests under various degrees of heteroskedasticity, and with instruments of different quality. Chapter 4 introduces the use of bootstrapping to obtain test statistic critical values. Our goal is to obtain a test with proper size in finite samples, and to examine its power. Alternative resampling schemes are considered to accommodate heteroskedastic errors. In Chapter 5 count data regression models with potentially endogenous regressors are considered. Tests for endogeneity in this model have some of the same features as the linear regression model with heteroskedastic errors, yet offer new complexities due to the model's nonlinearity. Chapter 6 concludes and summarizes.

In the remainder of this chapter we introduce the fundamental problems in the context of the simple linear regression model in a very intuitive, nontechnical discussion[1]. We also observe the perplexing default tests presented by several large computer software providers, which has introduced a measure of confusion into the application of tests for endogeneity in econometric practice. It is an objective of this work to clear the confusion and offer clear guidelines to empirical researchers who wish to test for the presence of potentially endogenous regressors.

1.1 Least Squares Estimation

In the linear regression model we make a number of assumptions about the data generating process. One of these is fundamental if the least squares estimator is to be consistent. Let us denote the simple linear regression model as

(1.1) $$y_i = \beta_1 + \beta_2 x_i + e_i, \quad i = 1, \ldots, N$$

Values of x_i and y_i are obtained by random sampling. We usually assume that the random errors e_i have certain properties, namely that they have zero mean and are uncorrelated with the regressor,

(1.2) $$E(e_i) = 0$$

(1.3) $$\operatorname{cov}(x_i, e_i) = 0$$

Under assumptions (1.1) - (1.3) the least squares estimators are consistent [if the distribution of x_i values is well behaved]; that is, they converge to the true parameter values as $T \to \infty$. If serial correlation and/or heteroskedasticity are present, then the least squares estimator is no longer "best," but it is still consistent.

---

[1] The discussion in the next few sections draws heavily on *Undergraduate Econometrics, 2nd Edition* by Hill, Griffiths and Judge (Wiley, 2001), Chapter 13.

There are some very standard situations in which correlation between x and e is anticipated to exist:

- when a relevant variable is omitted and is related to the included right-hand-side variables;
- simultaneous equations models;
- models including explanatory variables which are measured with error; and
- models in which a lagged dependent variable is included as a regressor, and serial correlation is present.

In each of these cases the usual least squares estimation procedure is no longer appropriate. If assumption (1.3) is not true, and consequently x_i and e_i are correlated, then the least squares estimators are inconsistent. They do not converge to the true parameter values even in very large samples. Furthermore, none of our usual hypothesis testing or interval estimation procedures are valid. Thus when x is random, the relationship between x and e is the crucial factor when deciding whether least squares estimation is appropriate or not. If the error term is correlated with x (any x in the multiple regression model) then the least squares estimator fails.

To demonstrate why the least squares estimator fails when $\text{cov}(x_i, e_i) \neq 0$ we will use a small simulation Let the systematic portion of the regression model be

(1.4) $$E(y_i \mid x_i) = \beta_1 + \beta_2 x = 1 + 1 \times x_i .$$

Using random number generators, we create $N = 100$ correlated pairs of x_i and e_i values with correlation .9, such that e_i has mean zero and constant variance. These values are shown in Figure 1.1. We then create y_i values by adding e_i to $E(y_i)$, given in (1.4). Applying least squares estimation to these data we obtain the least squares estimates b_1 and b_2, yielding the fitted regression line $\hat{y} = b_1 + b_2 x$. In Figure 1.2 we plot the fitted line and the true regression function $E(y_i) = 1 + x_i$. Note that the data values *are not* randomly scattered around the true regression function, because of the correlation we have created between x and e. The least squares principle works by fitting a line through the "center" of the data. When x and e are correlated the least squares idea is not going to work. The systematic overestimation of the slope, and underestimation of the intercept will not go away in larger samples, and thus the least squares estimators are not correct on average even in large samples. The least squares estimators are inconsistent. In this case

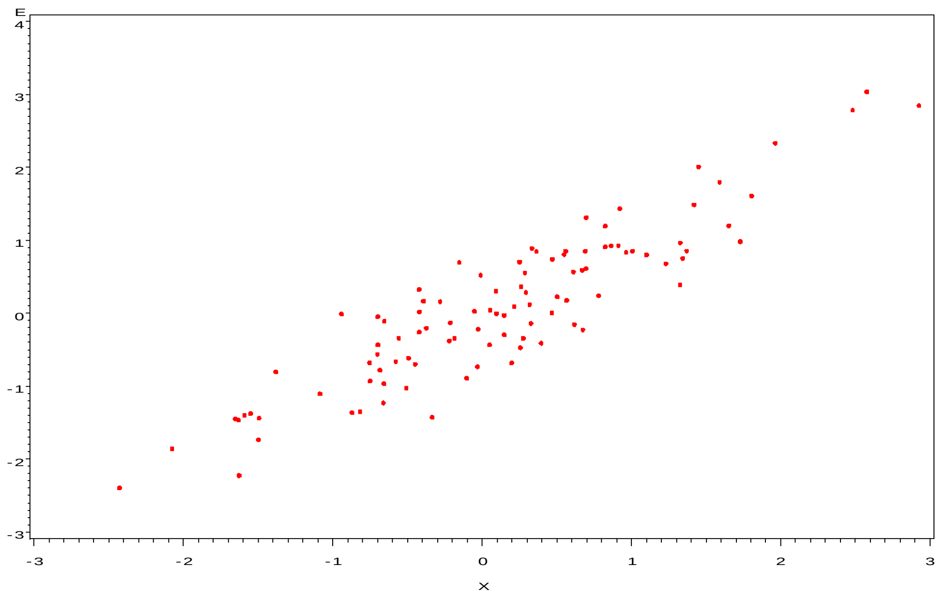


FIGURE 1.1. CORRELATED X AND E

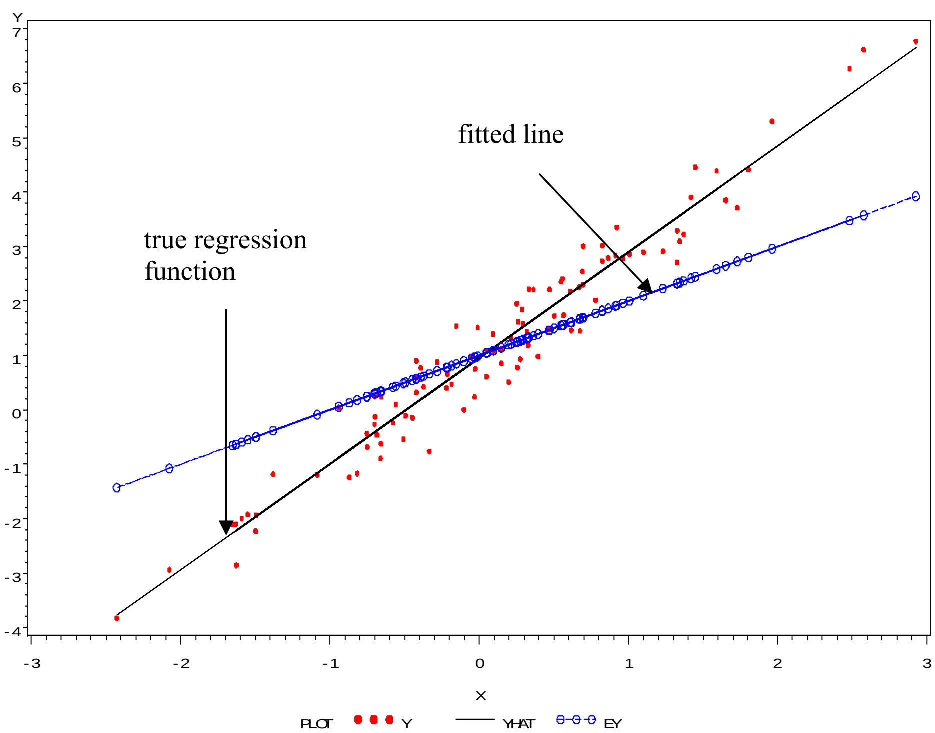


FIGURE 1.2. THE TRUE AND FITTED REGRESSION FUNCTIONS

we must consider alternative estimation procedures such as two-stage least squares, or instrumental variables estimation.

1.2 Instrumental Variables Estimation

Suppose that there is a variable, z_i, called an instrumental variable, which satisfies the moment conditions

$$E(z_i e_i) = 0 \quad \Rightarrow \quad E\left[z_i\left(y_i - \beta_1 - \beta_2 x_i\right)\right] = 0$$

$$E(e_i) = 0 \quad \Rightarrow \quad E\left(y_i - \beta_1 - \beta_2 x_i\right) = 0$$

The corresponding sample moment conditions are:

$$\frac{1}{N}\sum\left(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i\right) = 0$$
$$\frac{1}{N}\sum z_i\left(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i\right) = 0$$

Solving these equations leads us to method of moments estimators, which are usually called the instrumental variable estimators,

$$\hat{\beta}_2 = \frac{N\sum z_i y_i - \sum z_i \sum y_i}{N\sum z_i x_i - \sum z_i \sum x_i} = \frac{\sum(z_i - \bar{z})(y_i - \bar{y})}{\sum(z_i - \bar{z})(x_i - \bar{x})}$$
$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

These new estimators have the following properties:

- they are consistent
- in large samples the instrumental variable estimators have approximate normal distributions
- the variance of the instrumental variables estimator is

  $\operatorname{var}\left(\hat{\beta}_2\right) = \dfrac{\sigma^2}{\sum(x_t - \bar{x})^2 r_{zx}^2}$, where r_{zx}^2 is the squared sample correlation between the instrument z

  and the random regressor x.

1.3 When Surplus Instruments Are Available

Usually, however, we have more instrumental variables at our disposal than are necessary. For example, let w be a variable that is correlated with x but uncorrelated with e, so that we have a 3rd moment conditions

$$E(w_t e_t) = E\left[w_t\left(y_t - \beta_1 - \beta_2 x_t\right)\right] = 0$$

An alternative that uses all of the moment conditions is to choose values for $\hat{\beta}_1$ and $\hat{\beta}_2$ satisfying the 3 equations as closely as possible. The two-stage least squares estimation procedure satisfies this objective and can be implemented by (i) regressing *x* on a constant term, *z* and *w*, and obtain the predicted value $\hat{x}$; and (ii) using $\hat{x}$ as an instrumental variable for *x*. The resulting estimator is

$$\hat{\beta}_2 = \frac{\sum(\hat{x}_i - \bar{x})(y_i - \bar{y})}{\sum(\hat{x}_i - \bar{x})(x_i - \bar{x})}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

The appropriate estimator of the variance of $\hat{\beta}_2$ is

$$\widehat{\text{var}}\left(\hat{\beta}_2\right) = \frac{\hat{\sigma}^2_{IV}}{\sum(\hat{x}_i - \bar{x})^2}$$

where $\hat{\sigma}^2_{IV} = \frac{1}{N-2}\sum\left(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i\right)^2$. The estimated variance can be used as a basis for *t*-tests of significance and interval estimation of parameters.

1.4 The Hausman Test

The ordinary least squares estimator fails if there is correlation between an explanatory variable and the error term. The instrumental variables estimator can be used when the least squares estimator fails. How do we test for the presence of a correlation between an explanatory variable and the error term, so that we can use the appropriate estimation procedure? Let the null hypothesis be $H_0 : \text{cov}(x,e) = 0$. If the null hypothesis is true, both the least squares estimator and the instrumental variables estimator are consistent. Thus, in large samples the difference between them converges to zero. That is, $q = \left(b_{ols} - \hat{\beta}_{IV}\right) \to 0$. Naturally if the null hypothesis is true, use the more efficient estimator, which is the least squares estimator.

The alternative hypothesis is $H_1 : \text{cov}(x,e) \neq 0$. If the alternative hypothesis is true, the least squares estimator is not consistent, and the instrumental variables estimator is consistent, so $q = \left(b_{ols} - \hat{\beta}_{IV}\right) \to c \neq 0$. If the null hypothesis is not true, we should use the instrumental variables estimator, which is consistent.

There are several forms of the test for these null and alternative hypotheses. One form of the test directly examines the differences between the least squares and instrumental variables estimator, as we have described above. Many computer software programs implement this contrast test for the user. An alternative form of the test is very easy to implement. In the regression $y_i = \beta_1 + \beta_2 x_i + e_i$ we wish to know whether x is correlated with e. Let z_i be an instrumental variables for x_i. Then carry out the following steps:

- Estimate the model $x_i = a_0 + a_1 z_i + v_i$ by least squares, and obtain the residuals $\hat{v}_i = x_i - \hat{a}_0 - \hat{a}_1 z_i$.

  If there are more than one explanatory variables that are questionable, repeat this estimation for each one, using all available instrumental variables in each regression.

- Include the residuals computed in step 1 as an explanatory variable in the regression, $y_i = \beta_1 + \beta_2 x_i + \delta \hat{v}_i + e_i$. Estimate this "artificial regression" by least squares, and employ the usual t-test for the hypothesis of significance

  $H_0 : \delta = 0 \left(\text{no correlation between } x \text{ and } e\right)$
  $H_1 : \delta \neq 0 \left(\text{correlation between } x \text{ and } e\right)$

- If more than one variable is suspect, the test will be an F-test of joint significance of the coefficients on the included residuals.

1.5 Mroz Supply Equation

To illustrate the tests described above, we use a popular text book example. Example 9.5 in Wooldridge [*Econometric Analysis of Cross-Section and Panel Data*] is based on Mroz (1987). Based on a sample of 428 working women in 1975 we wish to estimate the labor supply function

$$hours = f\left[\log\left(wage\right), educ, const, age, kidslt6, kidsge6, nwifeinc\right]$$

We will use the software Stata, Version 9.2 and actually list the commands required for each alternative test. The single instrumental variable experience (*exper*) will be employed.

The contrast test is obtained by applying instrumental variables estimation (the IVREG2 command), followed by the least squares estimation (the REG command). After each estimation the results are stored for later recall.

```
. ivreg2 hours (lwage=exper) educ age kidslt6 kidsge6 nwifeinc

. estimates store iv

. reg hours lwage educ age kidslt6 kidsge6 nwifeinc

. estimates store ls
```

The contrast test is implemented using the HAUSMAN command.

```
. hausman iv ls
```

The test result that is reported is

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       iv           ls         Difference          S.E.
-------------+----------------------------------------------------------------
       lwage |    1772.323    -17.40781        1789.731        586.8068
        educ |    -201.187    -14.44486       -186.7422         66.9675
         age |   -11.22885    -7.729976       -3.498876        8.867692
     kidslt6 |   -191.6588    -342.5048         150.846        166.4163
     kidsge6 |   -37.73247    -115.0205        77.28804        55.07016
    nwifeinc |   -9.977746    -4.245807       -5.731939        6.104631
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from ivreg2
          B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic

                  chi2(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        9.30
                Prob>chi2 =      0.1573
```

Note that the test statistic is reported to be chi-square with 6 degrees of freedom, and a p-value of .1573. Based on these results we would (incorrectly) fail to reject the null hypothesis that log(wage) is uncorrelated with the regression error. The degrees of freedom for this test reflect that the contrast is based on a comparison of the regression coefficients other than the intercept. In carrying out the contrast test it is more conventional to include the intercept in the contrast, and this is achieved by making a simple modification to the HAUSMAN command.

```
. hausman iv ls, constant

    Test:  Ho:  difference in coefficients not systematic

                  chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        9.30
                Prob>chi2 =      0.2317
```

The chi-square test now has 7 degrees of freedom, and the reported p-value is .2317, again leading to failure to reject the null hypothesis that log(wage) is exogenous[2].

[2] This test is also the default in SAS's PROC MODEL.

Stata provides several other options with the HAUSMAN command. The option SIGMAMORE forces the contrast test to use the estimator of σ^2 based on the least squares estimates. Applying this command, along with the CONSTANT option yields

```
. hausman iv ls, constant sigmamore

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       33.56
                Prob>chi2 =      0.0000
```

The chi-square test statistic is now reported to have but one degree of freedom and the p-value is zero to four decimals. With this simple option we have obtained a totally different result.

The option SIGMALESS forces the contrast test to use the estimator of σ^2 based on the instrumental variables estimates. Applying this command, along with the CONSTANT option yields

```
. hausman iv ls, constant sigmaless

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        9.51
                Prob>chi2 =      0.0020
```

This test results in a chi-square statistic with one degree of freedom, a test statistic value that is close to the default test statistic value, yet a p-value of .002, leading us to reject the null hypothesis that log(wage) is not correlated with the regression error.

Stata provides a further option using the IVREG2 post-estimation command IVENDOG. This command yields

```
. ivendog

Tests of endogeneity of: lwage
H0: Regressor is exogenous
    Wu-Hausman F test:                 36.37992  F(1,420)    P-value = 0.00000
    Durbin-Wu-Hausman chi-sq test:     34.11764  Chi-sq(1)   P-value = 0.00000
```

Now we are presented with an *F*-test, with one numerator degree of freedom indicating that one hypothesis is being tested, and a chi-square test with one degree of freedom, but with a different numerical value than any of the previous chi-square values.

Finally, we can implement the regression based test in Stata using a few simple commands.

```
. reg lwage exper educ age kidslt6 kidsge6 nwifeinc

. predict v, residuals

. reg hours lwage educ age kidslt6 kidsge6 nwifeinc v
```

hours	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
lwage	1772.323	301.2612	5.88	0.000	1180.156	2364.491
educ	-201.187	35.44555	-5.68	0.000	-270.8598	-131.5142
age	-11.22885	5.342384	-2.10	0.036	-21.72999	-.7277101
kidslt6	-191.6589	99.25389	-1.93	0.054	-386.7551	3.4374
kidsge6	-37.73248	32.26388	-1.17	0.243	-101.1513	25.68631
nwifeinc	-9.977746	3.637582	-2.74	0.006	-17.12788	-2.827612
v	-1844.847	305.8648	-6.03	0.000	-2446.063	-1243.63
_cons	2478.435	332.2004	7.46	0.000	1825.453	3131.417

The *t*-statistic value for included residuals is −6.03, leading us to reject the null hypothesis that its coefficient is zero, and leading us to include that log(wage) is indeed endogenous. A quick calculation shows that the square of this *t*-statistic is the *F*-value reported by IVENDOG.

Why are there so many options? Why do some options imply that one hypothesis is being tested, and others imply that we testing a number of hypotheses equaling the number of model parameters, or one less than the number of model parameters? These are the questions that motivated this dissertation.

In Chapter 2 the asymptotic theory related to the tests is explored, and it is shown that the correct number of degrees of freedom for this test of endogeneity is the number of potentially endogenous variables in the regression that are being tested, in our example this is one. However other issues arise in each example related to the degree of endogeneity of the variables under scrutiny, and the number and strength of the available instrumental variables. If all available instruments are weak no test or instrumental variables estimation procedure is reliable. We explore the power of the alternative tests using Monte Carlo experiments, and compare these results to those previously reported in the literature. As to why software companies provide users with a default chi-square test with the wrong number of degrees of freedom is unclear. Our guess is that being able to compute many contrast tests, in many contexts, with the same code is powerfully attractive, yet in the context of the problems we consider fatally flawed.

Furthermore, all of the standard theory reviewed in Chapter 2 is developed under the assumption of homoskedastic errors. Most applications of tests for endogeneity are carried out with cross-sectional, or panel, data. In such data we routinely anticipate heteroskedasticity problems. Chapter 3 examines how the classical tests, and some alternatives, perform under heteroskedasticity. We are able to conclude that the preferred test is the *t*-test in the artificial regression, or the *F*-test if several potentially endogenous regressors are being tested, perhaps made robust by one form or another of a heteroskedasticity corrected covariance matrix.

Wong (2000) applied bootstrapping to obtain the critical value of the chi-square test for endogeneity. This approach is attractive as it should give tests of closer to proper size in finite samples. In recent years the explosion of literature on bootstrapping has included alternatives that may be effective when errors are heteroskedastic. We explore using bootstrapped tests in Chapter 4.

In Chapter 5 the tests for endogeneity in Chapters 3 and 4 are extended to exponential regression models for count data. A contrast test follows from recently developed instrumental and GMM estimators for count data models. In addition, Wooldridge (2002) suggests that an LM test similar to the regression based test for endogeneity can be employed. We consider these tests as well as those based upon a GMM estimator obtained by stacking moment conditions from the restricted and unrestricted models, as suggested by Creel (2004). Final remarks and questions for further consideration are in Chapter 6.

2. THE DURBIN-WU-HAUSMAN TESTS UNDER HOMOSKEDASTICITY

2.1 Literature Review

The idea that if a model is correctly specified, estimates by any two consistent methods should be close to each other, was introduced by Durbin (1954). Wu (1973, 1974) used a similar approach and suggested four tests for testing the assumption that regressors in the linear regression model are statistically independent of the disturbance term. The statistics are computationally cumbersome and thus they have not been used in empirical studies. His theoretical and Monte Carlo results (based on one endogenous regressor, two included and two excluded instruments model) indicated that the test statistics called T_2 has better properties than the other tests. He also suggested that the choice of the estimation method should be not only based on the correlation between the regressors and the error term but also on the coefficients of the first stage regression. Wu (1974) proposed the test T_2 to be the basis for a new estimator (the pre-test estimator) that is equal to the OLS estimator if the null of regressor's exogeneity is accepted and is equal to the IV estimator if the null is rejected. The computationally more convenient Hausman (1978) test is based on looking for a statistically significant difference between an efficient estimator under the null hypothesis of no misspecification and a consistent estimator under the alternative hypothesis that misspecification is present. There are more versions of the test depending on which estimators of the asymptotic covariance matrix are used. Nakamura and Nakamura (1981) proved that the Hausman test statistics are equivalent to the statistics proposed by Durbin (1954) and Wu (1973, 1974). The version of the Hausman statistic that uses the OLS estimate of the error variance is equivalent to the test statistic proposed by Durbin (1954) and separately by Wu (1973) (his T_4 statistic, which is a monotone transformation of T_2, thus has the same properties.). The version of the Hausman statistic that is formed using the IV estimate of the error variance was also first proposed by Wu (1973) (his T_3 statistic).

Asymptotically, all the different versions of the test statistic have a chi-square distribution, with degrees of freedom equal to the number of potentially endogenous regressors at most as Hausman and Taylor (1981) demonstrated. However, the rank of the covariance matrix that is usually reported is the dimension of the whole parameter vector.

In the following, we present another case where the routinely used (SAS computed) Hausman test statistics leads to partially incorrect conclusions about regressors endogeneity. Gaston and Trefler (1994) investigate the effects of international trade policy on wages in U.S. manufacturing industries in 1983. In their data, tariffs and non-tariff barriers are negatively correlated with industry wage premium meaning that wages in protected industries are lower than in unprotected industries. The correlation is economically large. One explanation of the high correlation could be that the impact of tariffs reflects the endogeneity of protection, that the level of wages influences the decision to protect an industry, in this case policy makers protect low-wage industries. However, Gaston and Trefler (1994) find no evidence of such process. In Table 4 (p. 582), they present results of the OLS and 2SLS regression of the wage premium on tariffs, non-tariff trade barriers, exports, imports, import growth, intra-industry trade and a constant. To test for endogeneity of protection (tariffs and non-tariff trade barriers), they compute the Hausman test statistic and compare it to the critical value of a chi-square distribution with 7 degrees of freedom, which is the total number of regressors. The value of the test statistic is 5.95, thus they find no evidence in favor of endogeneity of protection. However, if we use the number of degrees of freedom equal to the number of potentially endogenous regressors (two), the test rejects protection exogeneity at 10% and fails to reject it at 5% very marginally (the critical value of a chi-square distribution with 2 degrees of freedom is 5.99).

One of the reasons for the incorrect number of degrees of freedom computed is that in finite samples, the Hausman test statistic can have a distribution that is different from the one predicted by the asymptotic theory, especially if we do not make the assumptions of the classical normal model. Kariya and Hodoshima (1980) show that under the assumption of normal disturbance terms the exact conditional distribution of the Wu-Hausman statistic is the double non-central *F* distribution. They prove the non-unbiasedness (or biasedness) of the test when the critical point is greater than the ratio of the denominator and numerator degrees of freedom. The non-unbiasedness (biasedness) of the test implies that when the alternative is true, there exist some parameter points in which acceptance of the null is more likely than in the case where the null hypothesis is true thus, the type II error of the test may be relatively high. The Hausman test that uses the OLS estimate of the error variance seems to have the best properties. It has an asymptotic distribution that is a mixture of a non-central chi-square distribution with a random non-centrality parameter. Under the null it converges to a chi-square distribution. When the instruments are relevant, the stronger the instruments, the fewer the number of instruments and the stronger the endogeneity, the higher the power of the test. Under the alternative, the Hausman test with the OLS estimate of the error variance has greater asymptotic power than the tests that use the difference of the covariance matrices or the IV estimate of the error variance.

When the instruments are irrelevant, in general the two-stage least squares estimator is not consistent and has a nonstandard asymptotic distribution. Maddala and Jeong (1992) examine the behavior of the IV estimator in the one-regressor-one instrument model when the correlation between the regressor and the instrument is very low. Their results indicate that standard statistical inference in such circumstances may be very misleading. As Nelson and Starz (1990a, b) and Bound, Jaeger and Baker (1995) find, the 2SLS estimator is biased in the direction of the OLS estimator, and the 2SLS standard error is small relative to the bias. Buse (1992) shows that there is no systematic relationship between bias and the number of instruments. The estimated bias will increase with the number of excess instrumental variables only if the proportional increase in the number of instruments is faster than the rate of increase in R^2 measured relative to the fit of the endogenous variables on the exogenous variables. Blomquist and Dahlberg (1999) also study the small sample performance of the 2SLS, the LIML and four new jackknife IV estimators under weak instruments. They find that the LIML and the new jackknife estimators have a smaller bias but a larger variance than the 2SLS. In terms of root mean square error, neither LIML nor the new estimators perform uniformly better than the 2SLS. The properties of the estimators are specific to each data-generating process and sample size. To obtain reliable estimates, better instruments and/or larger samples are required. Staiger and Stock (1997) propose reporting the first stage F statistics and/or the bias measures. Shea (1997) suggests a simple partial R^2 measure of instrument relevance for multivariate models. However, Hall, Rudebusch and Wilcox (1996) show that such pretesting does not appear to work and may even make matters worse as far as test reliability is concerned. Their simulation results indicate that a use of relevance measures may actually exacerbate the poor finite-sample properties of the IV estimator as also described by Nelson and Starz (1990a, b): "The probability distribution of the IV estimator based on a "good" instrument (as identified by the use of a relevance statistic as a screening device) can be even more distorted than the one based on a random (unscreened) instrument. This error arises because those instruments that are identified as having high relevance for the regressors in the sample are also likely to have higher endogeneity in the sample." Hall, Rudebusch and Wilcox (1996) recommend that we do not rely on asymptotic theory but focus on an analysis of the distributions of the estimated parameters and hypothesis tests conditional on the realized value of the relevance statistic. In our case we should take into account the sample correlation between regressors and instruments.

Since the finite-sample distribution of IV estimators departs noticeably from the asymptotic normal distribution under low relevance of instruments, the asymptotic distributions of test statistics and variance estimators are also nonstandard. Conventional asymptotics treat the coefficients on the instruments in the first stage as nonzero and fixed which implies that the *F* statistic increases to infinity with the sample size. Thus, when the means of these *F* statistics are small, these asymptotic approximations break down as demonstrated by Wang and Zivot (1998). Staiger and Stock (1997) show that in the case of a single potentially endogenous variable the distribution of the IV esimator depends on the number of instruments and the noncentrality parameter whose limit is a noncentral Wishart random variable. The noncentrality parameter is not consistently estimable and thus asymptotically valid confidence regions cannot be constructed by directly inverting the statistics. Staiger and Stock (1997) suggest Anderson-Rubin confidence regions and confidence regions based on Bonferroni's inequality. However, as Revankar (1978) demonstrates the Anderson-Rubin test statistic is less efficient than the Wu statistics and also less efficient than the likelihood ratio test, as shown by Hwang (1980), but it is easier to compute. Conversely, Davidson and MacKinnon (2006) found that under weak instruments, none of the Student's *t* (or Wald), Kleibergen's *K* (2002), likelihood ratio and the Anderson-Rubin (1949) test statistics has any real asymptotic power against local alternatives.

In the following, we compare the size and power of the alternative versions of the Hausman test and related classical tests under homoskedasticity, investigate the impact of the strength of the instruments and show that the usual version of the test is not always the best option.

2.2 Contrast Tests

The linear regression model is $y = X\beta + u$. Initially we consider that the errors are homoskedastic with variance σ^2, thus $V(u) = \sigma^2 I$. Let the $n \times K$ matrix of explanatory variables X [$plim\left(X'X/n\right) = Q_{XX} \neq 0$] be partitioned as $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$, where

- X_1 is $n \times K_1$ and potentially endogenous, $plim\left(X_1'u/n\right) \neq 0$, and
- X_2 is $n \times K_2$ and assumed exogenous, $plim\left(X_2'u/n\right) = 0$

Assume that there is an $n \times L$ matrix $Z = \begin{bmatrix} Z_1 & X_2 \end{bmatrix}$ of instruments that are uncorrelated with the error, $plim\left(Z'u/n\right) = 0$. X_2 contains the included or "internal" instruments. The matrix Z_1 is $n \times L_1$ and comprises the excluded or "external" instruments, which do not appear in the regression equation. Valid instruments must also be correlated with the regressors so that $plim\left(Z'X/n\right) = Q_{ZX} \neq 0$.

The number of instruments $L = L_1 + K_2$ must be greater than or equal to the number of regressors K. So L_1, the number of external instruments, must be greater than or equal to K_1, the number of potentially endogenous variables. This is the necessary order condition. We also assume that Q_{ZX} is of full rank K.

The least squares estimator $\hat{\beta}_{OLS} = \left(X'X\right)^{-1} X'y$ is consistent if X and u are uncorrelated, and with homoskedastic errors $V\left(\hat{\beta}_{OLS}\right) = \sigma^2 \left(X'X\right)^{-1}$ with $\hat{\sigma}^2_{OLS} = \left(y - X\hat{\beta}_{OLS}\right)' \left(y - X\hat{\beta}_{OLS}\right) \Big/ \left(n - K\right)$. If X and u are correlated then the least squares estimator is inconsistent. Define the matrix $P_Z = Z\left(Z'Z\right)^{-1} Z'$. A consistent estimator is the instrumental variables (IV) estimator $\hat{\beta}_{IV} = \left(X'P_Z X\right)^{-1} X'P_Z y$ which has covariance matrix $V\left(\hat{\beta}_{IV}\right) = \sigma^2 \left(X'P_Z X\right)^{-1}$ and $\hat{\sigma}^2_{IV} = \left(y - X\hat{\beta}_{IV}\right)' \left(y - X\hat{\beta}_{IV}\right) \Big/ \left(n - K\right)$. If X is uncorrelated with the error then the $\hat{\beta}_{IV}$ estimator is inefficient relative to $\hat{\beta}_{OLS}$. In fact it can be quite a bit less efficient, depending on the quality of the instruments.

If we are not sure about the endogeneity of a subset of regressors then we can resort to a test belonging to the Durbin-Wu-Hausman (DWH) family. If we define $q = \left(\hat{\beta}_{IV} - \hat{\beta}_{OLS}\right)$ then the Hausman (1978) test statistic is $H = q'\left[V\left(\hat{\beta}_{IV}\right) - V\left(\hat{\beta}_{OLS}\right)\right]^{+} q$ where "+" denotes a generalized inverse. Under the null hypothesis that both estimators are consistent [or that X_1 is uncorrelated with the error term] then $H \overset{a}{\sim} \chi^2_{(K_1)}$ (see Appendix) . However, within class of contrast (comparing two estimators) tests there are a number of options

a. $H_1 = q'\left[V\left(\hat{\beta}_{IV}\right) - V\left(\hat{\beta}_{OLS}\right)\right]^{+} q$ and $H_1 \overset{a}{\sim} \chi^2_{(K_1)}$

b. $H_{1s} = q'\left[V\left(\hat{\beta}_{IV}\right) - V\left(\hat{\beta}_{OLS}\right)\right]^{+} q$ and $H_{1s} \overset{a}{\sim} \chi^2_{\left(rank\left[V\left(\hat{\beta}_{IV}\right) - V\left(\hat{\beta}_{OLS}\right)\right]\right)}$. This is the default in SAS. In STATA the default omits the contrast of constant terms. The constant is included with an extra option. The SAS computation is `H = q'*ginv(v0-v1)*q`, and the degrees of freedom as `df=round(trace(ginv(v0-v1)*(v0-v1)))`. As a reminder, the STATA default is the contrast excluding the constant term, and with degrees of freedom equaling the rank of the resulting covariance difference, which is usually $K - 1$ and not K_1 unless all regressors are potentially endogenous.

c. $H_2 = q'\left\{\hat{\sigma}^2_{IV}\left[\left(X'P_Z X\right)^{-1} - \left(X'X\right)^{-1}\right]\right\}^{+} q$ and $H_2 \overset{a}{\sim} \chi^2_{(K_1)}$ [STATA `SIGMALESS` option]

d. $H_3 = q'\left\{\hat{\sigma}^2_{OLS}\left[\left(X'P_Z X\right)^{-1} - \left(X'X\right)^{-1}\right]\right\}^{+} q$ and $H_3 \overset{a}{\sim} \chi^2_{(K_1)}$ [STATA `SIGMAMORE` option])

e. $H_{3a} = q'\left\{\hat{\sigma}^2_{ML}\left[\left(X'P_Z X\right)^{-1} - \left(X'X\right)^{-1}\right]\right\}^{+} q$ where $\hat{\sigma}^2_{ML} = \left(y - X\hat{\beta}_{OLS}\right)'\left(y - X\hat{\beta}_{OLS}\right)\Big/ n$ and $H_{3a} \overset{a}{\sim} \chi^2_{(K_1)}$ [STATA `IVENDOG` option following `IVREG2; Durbin-Wu-Hausman chi-sq test`. See Baum, Schaffer and Stillman (2003)].

2.3 Artificial Regression Version 1

In addition to the contrast tests there are some auxiliary regression equivalents. Consider the contrast vector

$$
\begin{aligned}
q &= \hat{\beta}_{IV} - \hat{\beta}_{OLS} \\
&= \left(X'P_Z X\right)^{-1} X'P_Z y - \left(X'X\right)^{-1} X'y \\
&= \left(X'P_Z X\right)^{-1}\left[X'P_Z y - X'P_Z X\left(X'X\right)^{-1} X'y\right] \qquad (2.1) \\
&= \left(X'P_Z X\right)^{-1}\left[X'P_Z\left(I - X\left(X'X\right)^{-1} X'\right)y\right] \\
&= \left(X'P_Z X\right)^{-1} X'P_Z M_X y
\end{aligned}
$$

The test of whether $q \to 0$ asymptotically is equivalent to testing whether $X'P_Z M_X y$ has zero mean asymptotically. The matrix of fitted values in the reduced form regression of X on Z

$P_Z X = Z(Z'Z)^{-1} Z'X = \hat{X} = \begin{bmatrix} \hat{X}_1 & X_2 \end{bmatrix}$ and $M_X \hat{X} = \begin{bmatrix} M_X \hat{X}_1 & 0 \end{bmatrix}$. Thus the portion of $X'P_Z M_X y$ that is relevant is $X_1'P_Z M_X y$. Defining $\tilde{P} = M_X P_Z X_1$, consider the artificial regression

(2.2) $\quad y = X\beta^* + P_Z X_1 \delta + error = X\beta^* + \hat{X}_1 \delta + error$

We use β^* to distinguish from the instrumental variables estimator since they are not numerically equal. Applying the Frisch-Waugh-Lovell (FWL) Theorem we obtain

(2.3) $\quad M_X y = M_X P_Z X_1 \delta + error \Rightarrow \tilde{y} = \tilde{P}\delta + error$

It follows that the test for $\delta = 0$ is testing for zero correlation between residuals from the regression of *y* on *X* and some transformation of the potentially endogenous regressors. The least squares residuals from (2.3) are

$$\tilde{y} - \tilde{P}\hat{\delta} = M_X y - M_X P_Z X_1 \left(X_1'P_Z M_X P_Z X_1 \right)^{-1} X_1'P_Z M_X y = M_{\tilde{P}} M_X y ,$$

where $M_{\tilde{P}} = I - M_X P_Z X_1 \left(X_1'P_Z M_X P_Z X_1 \right)^{-1} X_1'P_Z$.

The unrestricted sum of squared residuals $SSR_{un} = y'M_X M_{\tilde{P}} M_X y$ and an estimator of the error variance that is consistent under the null hypothesis that $\delta = 0$ is $\tilde{\sigma}^2 = SSR_{un} / (n - K - K_1)$. Under the null hypothesis that $\delta = 0$, the restricted model is the usual regression model with $SSR_{rest} = y'M_X y$. Taking the difference we obtain

$$SSR_{rest} - SSR_{un} = y'M_X y - y'M_X M_{\tilde{P}} M_X y = y'M_X \left(I - M_{\tilde{P}} \right) M_X y = y'\tilde{P}\left(\tilde{P}'\tilde{P}\right)^{-1} \tilde{P}'y$$

using the fact that $M_X \tilde{P} = \tilde{P}$. Therefore a test that is asymptotically equivalent to the Hausman test is an *F*-test of the null hypothesis that $\delta = 0$ in the artificial regression (2.2), with the test statistic being

(2.4) $\quad F_{DWH} = \dfrac{y'\tilde{P}\left(\tilde{P}'\tilde{P}\right)^{-1} \tilde{P}'y}{\tilde{\sigma}^2 K_1} \overset{a}{\sim} F_{K_1, n-K-K_1}$

2.4 Artificial Regression Version 2

An alternative version of the artificial regression is

(2.5) $\quad y = X\beta + M_Z X_1 \eta + error = X\beta + (\hat{X}_1 - X_1)\eta + error = X\beta + \hat{V}_1 \eta + error$

In this regression instead of augmenting the original model with the predicted values of the endogenous regressors from the reduced form, we add to the original model the residuals from reduced form equations. The *F*-test that results is identical to the earlier one because

$$M_X y = M_X M_Z X_1 \eta + error$$

Note that

$$M_X M_Z X_1 = M_X \left(I - P_Z\right) X_1 = M_X X_1 - M_X P_Z X_1 = -M_X P_Z X_1$$

since $M_X X_1 = 0$. So the regression

$$M_X y = M_X M_Z X_1 \eta + error = M_X P_Z X_1 \left(-\eta\right) + error = M_X P_Z X_1 \delta + error$$

has the same sum of squared residuals as in the earlier case (2.2), and thus the *F*-test of $\eta = 0$ in (2.5) is identical to that for $\delta = 0$ in (2.4). This test is referred to in STATA (STATA `IVENDOG` option following `IVREG2)` as the `Wu-Hausman F-test`.

This version of the artificial regression has another convenient property. The least squares estimator of β in (2.5) is numerically equal to the instrumental variables estimator $\hat{\beta}_{IV}$. To see this write

$$\begin{aligned} y &= X\beta + M_Z X_1 \eta + error \\ &= X\beta + M_Z X \begin{bmatrix} \eta \\ \eta_2 \end{bmatrix} + error \end{aligned}$$

This follows because

$$M_Z X = X - \hat{X} = \begin{bmatrix} X_1 & X_2 \end{bmatrix} - \begin{bmatrix} \hat{X}_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1 - \hat{X}_1 & 0 \end{bmatrix} = \begin{bmatrix} M_Z X_1 & 0 \end{bmatrix}$$

Let $\tilde{X} = M_Z X$ and $M_{\tilde{X}} = I - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}' = I - M_Z X\left(X'M_Z X\right)^{-1} X'M_Z$. Then we can apply the FWL theorem again to the artificial regression

$$y = X\beta + M_Z X \begin{bmatrix} \eta \\ \eta_2 \end{bmatrix} + error = X\beta + \tilde{X}\tilde{\eta} + error$$

Multiplying both sides by $M_{\tilde{X}}$ we obtain

(2.6) $$M_{\tilde{X}} y = M_{\tilde{X}} X\beta + error = \hat{X}\beta + error$$

since

$$M_{\tilde{X}} X = X - M_Z X\left(X'M_Z X\right)^{-1} X'M_Z X = X - M_Z X = P_Z X = \hat{X} .$$

The least squares estimator from (2.6) is then

$$\hat{\beta} = \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}'M_{\tilde{X}}\, y$$

Multiplying out the final term we find

$$\begin{aligned} \hat{X}'M_{\tilde{X}}\, y &= \hat{X}'\left(I - M_Z X \left(X'M_Z X\right)^{-1} X'M_Z\right) y \\ &= X'P_Z\left(I - M_Z X \left(X'M_Z X\right)^{-1} X'M_Z\right) y \\ &= X'P_Z\, y = \hat{X}'y \end{aligned}$$

since $P_Z M_Z = P_Z - P_Z = 0$. Therefore

$$\hat{\beta} = \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}'M_{\tilde{X}}\, y = \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}'y = \hat{\beta}_{IV}$$

2.5 Other Classical Tests Equivalents

There are also other classical tests equivalents. Silvey (1959), Engle (1984), Holly (1982), Andrews and Fair (1988) demonstrated that under very general conditions, the LR, LM, Wald and the Hausman test statistics are locally equivalent. Another way how to obtain the Hausman test statistics is to multiply the squared multiple correlation coefficient R^2 of the ordinary least squares regression of the constant unity on the original regressors and the reduced form errors by the number of observations n. The product nR^2 is asymptotically distributed as chi-square with degrees of freedom equal to the number of endogenous regressors (White (1987)). Ruud (1984) and Newey (1985) also show that tests asymptotically equivalent to the Hausman test can be computed as the score test. Davidson and MacKinnon (1990) show that for any test statistic that can be computed with artificial regression (for example a LM test) there is a Durbin-Wu-Hausman version that can be based on similar artificial regression. Holly and Monfort (1986) describe cases where quadratic forms based on linear combinations of the constrained and unconstrained estimators of all the parameters of a model are asymptotically equivalent to the classical test statistics. Their main result is that the rank of appropriate information matrices should be equal to the number of parameters of primary interest so that no information is lost. Newey and McFadden (1994, p. 2222) and Gourieroux and Monfort (1989, p. 73) demostrate the asymptotic equivalence of comparable classical-type and Hausman-type statistics constructed with the information matrix-type equality imposed and not imposed, respectively. Dastoor (2003) shows the equality of comparable extended families of classical type and Hausman-type statistics.

2.6 The Estimators under Homoskedasticity

Though it is not a primary objective in our study, we do report the Monte Carlo means of the alternative estimators and their root-mean-squared-errors (RMSE) for the slope parameters. The coefficients of the potentially endogenous regressors X_1 are of primary interest. The estimators whose performance we report are OLS (b_{2OLS}), instrumental variables (b_{2IV}) and optimal 2-step GMM (b_{2GMM}). In addition we report results for the pre-test estimators using the usual *t*-statistic. This estimator is

$$b_{2pt\alpha} = \begin{cases} b_{2OLS} & \text{if } t < t_c \\ b_{2IV} & \text{if } t \geq t_c \end{cases}$$

where t_c is the critical value for a *t*-distribution with $n - K - 1$ (degrees of freedom from the artificial regression) at significance level $\alpha = .05$ (b_{2pt05}) or $\alpha = .20$ (b_{2pt20}). Similarly we define a pre-test estimator defined on the default SAS statistic **`ho1s`**

$$b_{2pt\alpha s} = \begin{cases} b_{2OLS} & \text{if } H_{1s} < \chi^2_c \\ b_{2IV} & \text{if } H_{1s} \geq \chi^2_c \end{cases}$$

2.7 A Monte Carlo Experiment under Homoskedasticity

We adopt a slight variation of the Monte Carlo set up of Creel (2004). For the most part we use 40,000 simulations. Using Cameron and Trivedi's [2005, p. 252] test size calculation, this implies that the half-width of 95% interval estimates for the .01, .05, .10 and .20 test sizes will are 0.00098, 0.00214, 0.00294, and 0.00392 respectively [3]. We consider the regression

(2.7) $$y = \beta_1 + \beta_2 x + u = 0 + x + u$$

Data are generated by specifying

(2.8) $$\begin{bmatrix} x \\ v \\ z_1 \\ z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & 0 & 0 \\ \rho_2 & 0 & 1 & \rho_4 \\ \rho_3 & 0 & \rho_4 & 1 \end{bmatrix} \right)$$

The key features are that

[3] Creel used 100,000 simulations: the half-widths are 0.00062, 0.00135, 0.00186, 0.00248

- endogeneity is controlled by the parameter ρ_1 which takes the values 0, .1, .2, .3, .4 and .5
- the strength of the instruments is controlled by ρ_2 and ρ_3 which take the values .1, .3 and .5
- the correlation between the instruments is controlled by ρ_4 which takes the values 0 and .5
- samples of sizes n = 100 and n = 200 are considered.

Our choice of parameter values is limited by the necessity of positive definiteness of the data generating process covariance matrix. Previous literature mostly uses just-identified models with one endogenous regressor. We choose two instruments and report results for models with two moderate, two strong instruments and one strong and one moderate instrument since we expect the test performance to be highly dependent on the instrument strength. For future reference, the various symbols are summarized below:

- n = sample size
- ρ_1 controls endogeneity
- ρ_2 controls strength of IV #1
- ρ_3 controls strength of IV #2
- ρ_4 correlation between instruments
- α = nominal level of significance
- **ho1** = H_1 contrast with K_1 = 1 df
- **ho1s** = H_1 contrast with K = 2 df [SAS default and STATA default with `CONSTANT` option]
- **ho2** = H_2 contrast with IV variance estimator [`sigmaless`]
- **ho3** = H_3 contrast with OLS variance estimator [`sigmamore`]
- **ho3a** = H_{3a} contrast with MLE variance estimator [`Wu-Hausman chi-square`]
- **t** = t test of residual coefficient in auxiliary regression [`DWH F-test`]

2.8 Discussion of Asymptotic Results under Homoskedasticity

We report the results for just a few scenarios of instrumental variable strength: Case 1 has two strong instruments, both having correlation .5 with the endogenous regressor. Case 2 is two moderately strong instruments, both having correlation .3 with the endogenous regressor. Case 3 has one strong instrument (.5 correlation) and one weak instrument (.1 correlation). When both instruments are weak none of the tests performs well. In all reported results the instruments are uncorrelated with each other. If the instruments have a .5 correlation, power is slightly reduced.

2.8.1 The Effect of Sample Size

All the results can be found in an appendix that is available upon request. Here we point out a few findings. In Table 2.1 we consider the effect of sample size. The "`Table`" numbers within the table body refer to its location in the appendix tables that can be provided upon request. Sample sizes n = 100 and 200 are examined and we have assumed one strong and one weak instrument. The effect of having a larger sample on test power is dramatic, which in itself is troubling since we have only 1 regressor. One would think that 100 observations constitutes a "large sample" with a simple regression, but this is not so.

When the endogeneity control $\rho_1 = 0$ the tests should reject with frequency α. The contrast test **`ho1s`**, the SAS/STATA default test, fails this criterion horribly for both sample sizes shown in Table 2. For example, $\alpha = .05$ the rejection frequency for **`ho1s`** when n = 100 is .00462, and when n = 200 it is .00958. The test **`ho1`**, which is based on separate error variance estimates for OLS and IV estimation, but uses the correct degrees of freedom, $K_1 = 1$, has size that is too low. The same is true of **`ho2`**, the contrast test with the IV estimate of the variance. The other tests seem to have about the right size most of the time. However, the size of the **`t`**-test does not improve with a larger sample size. The pattern is more noticeable under weak instruments. As shown by Nelson and Startz (1990b), under weak instruments, the bias of the instrumental variable estimator may increase with larger sample size and lower degree of endogeneity. Consequently, test size distortions may be worse in larger samples.

In the second part of the table, Table 2.1a, we show the size corrected power of these tests. They are virtually identical, with **`ho3`**, **`ho3a`** and the **`t`**-test having a slight power edge at the third decimal. Thus the problems of the tests **`ho1`**, **`ho1s`** and **`ho2`** seem to be associated with incorrect size.

2.8.2 The Effect of Instrument Strength

In Table 2.2 and Figure 2.1 we report 3 cases with n = 100. The rejection frequencies in the top panel correspond to the case with two strong instruments, the second panel to two weaker instruments, and the third panel to one strong and one weak instrument. When two strong instruments are available all the tests, except **`ho1s`**, reject the exogeneity of the suspect variable in excess of 80% of the samples with a correlation between the endogenous regressor and error of $\rho_1 = .3$. Even with two strong instruments (recall that there is only one potentially endogenous regressor) the sizes of **`ho1`**, **`ho1s`** and **`ho2`** are too small. Figure 2.2 compares the effect of instrument strength on individual tests. If we have two weaker instruments, with correlation to the regressor $\rho_2 = \rho_3 = .3$, all of the tests perform substantially worse at all degrees of endogeneity. We confirm that it is better to have one strong instrument than two weak ones, though even in this case the power of the tests is reduced relative to the case with two strong instruments.

2.8.3 Estimator Bias Results

The sample means of the estimators are pictured in Figure 2.3 for the homoskedastic case. Recall that the true value of $\beta_2 = 1$. The bias in the OLS estimator increases with the degree of endogeneity of the regressor *x*. The instrumental variables estimators have virtually no bias even with $n = 100$. The pretest estimators behave predictably. For a given pretest estimator the bias is smaller when $\alpha = .20$ is chosen since the null is rejected with greater frequency and the OLS estimator abandoned in favor of IV estimation. The pretest estimator based on the SAS $\chi^2_{(2)}$ test is worse in terms of bias than the other pretest estimators because it leads to rejection of the null a smaller percentage of the time. In all cases the bias begins and ends at zero, with a positive "hump" in the middle owing to us finding ourselves in that shadowy twilight-zone of indecision between heaven (exogenous regressor not requiring any IV's) and the other place (endogenous regressor requiring us to find good IV's).

2.8.4 Estimator RMSE Comparisons

In Figure 2.4 we compare the root-mse's of the alternative estimators. The RMSE of the OLS estimator exceeds that of the IV estimators when the degree of endogeneity $\rho_1 \geq .2$. When the degree of endogeneity is low the OLS estimator is a better choice. The pretest estimator is never the best choice, but we find something interesting nonetheless. SAS's test leads to a better pre-test estimator performance when the degree of endogeneity is low ($\rho_1 \leq .2$) and performance that is at least comparable to the other pre-test estimators when $\rho_1 = .3$. Thus the SAS/STATA default test is "best" for the intrepid researcher who does

TABLE 2.1. PERCENT REJECTIONS (HOMOSKEDASTIC CASE) - EFFECT OF SAMPLE SIZE (rho2 = .5, rho3 = .1)

Table 3: Percent rejections, n=100, gamma=0, rho2=.5, rho3=.1

alpha=0.05

rho1	ho1	ho1s	ho2	ho3	ho3a	t
0.0	0.03103	0.00462	0.03492	0.04868	0.05100	0.04905
0.1	0.06015	0.01110	0.06678	0.08685	0.09035	0.08750
0.2	0.16303	0.04380	0.17625	0.21255	0.21893	0.21385
0.3	0.37157	0.14758	0.39082	0.44390	0.45280	0.44572
0.4	0.65575	0.36843	0.67393	0.72377	0.73070	0.72530
0.5	0.90073	0.70530	0.90950	0.93037	0.93302	0.93107

Table 6: Percent rejections, n=200, gamma=0, rho2=.5, rho3=.1

alpha=0.05

rho1	ho1	ho1s	ho2	ho3	ho3a	t
0.0	0.04310	0.00958	0.04482	0.05170	0.05310	0.05200
0.1	0.11480	0.03655	0.11910	0.13130	0.13325	0.13167
0.2	0.36412	0.17572	0.37222	0.39455	0.39837	0.39548
0.3	0.71972	0.50085	0.72570	0.74503	0.74838	0.74572
0.4	0.94747	0.85115	0.94952	0.95488	0.95588	0.95513
0.5	0.99825	0.98923	0.99835	0.99860	0.99862	0.99860

Table 2.1a. Size Corrected Power

Table 21: Size Corrected Power, n=100, gamma=0, rho2=.5, rho3=.1

alpha=0.05

rho1	ho1	ho2	ho3	ho3a	t
0.0	0.05003	0.05003	0.05003	0.05003	0.05003
0.1	0.08868	0.08868	0.08905	0.08905	0.08905
0.2	0.21502	0.21502	0.21663	0.21663	0.21663
0.3	0.44712	0.44712	0.44960	0.44960	0.44960
0.4	0.72540	0.72540	0.72808	0.72808	0.72808
0.5	0.93090	0.93090	0.93227	0.93227	0.93227

TABLE 2.2. PERCENT REJECTIONS (HOMOSKEDASTIC CASE) - EFFECT OF INSTRUMENT STRENGTH (n = 100)

Table 1: Percent rejections, n=100, gamma=0, rho2=.5, rho3=.5

alpha=0.05

rho1	ho1	ho1s	ho2	ho3	ho3a	t
0.0	0.04305	0.00833	0.04743	0.05185	0.05407	0.05225
0.1	0.14422	0.04793	0.15398	0.16425	0.16910	0.16512
0.2	0.48360	0.25853	0.50103	0.51818	0.52638	0.51992
0.3	0.86940	0.69960	0.87823	0.88633	0.89058	0.88740
0.4	0.99375	0.97162	0.99457	0.99503	0.99535	0.99513
0.5	0.99997	0.99980	0.99997	0.99997	0.99997	0.99997

Table 2: Percent rejections, n=100, gamma=0, rho2=.3, rho3=.3

alpha=0.05

rho1	ho1	ho1s	ho2	ho3	ho3a	t
0.0	0.02510	0.00290	0.02908	0.04963	0.05198	0.05017
0.1	0.03852	0.00500	0.04317	0.07118	0.07433	0.07175
0.2	0.09077	0.01553	0.10103	0.14815	0.15275	0.14920
0.3	0.20095	0.04805	0.21743	0.29288	0.29947	0.29433
0.4	0.38405	0.13003	0.40730	0.50550	0.51323	0.50738
0.5	0.64030	0.31525	0.66180	0.75120	0.75750	0.75257

Table 3: Percent rejections, n=100, gamma=0, rho2=.5, rho3=.1

alpha=0.05

rho1	ho1	ho1s	ho2	ho3	ho3a	t
0.0	0.03103	0.00462	0.03492	0.04868	0.05100	0.04905
0.1	0.06015	0.01110	0.06678	0.08685	0.09035	0.08750
0.2	0.16303	0.04380	0.17625	0.21255	0.21893	0.21385
0.3	0.37157	0.14758	0.39082	0.44390	0.45280	0.44572
0.4	0.65575	0.36843	0.67393	0.72377	0.73070	0.72530
0.5	0.90073	0.70530	0.90950	0.93037	0.93302	0.93107

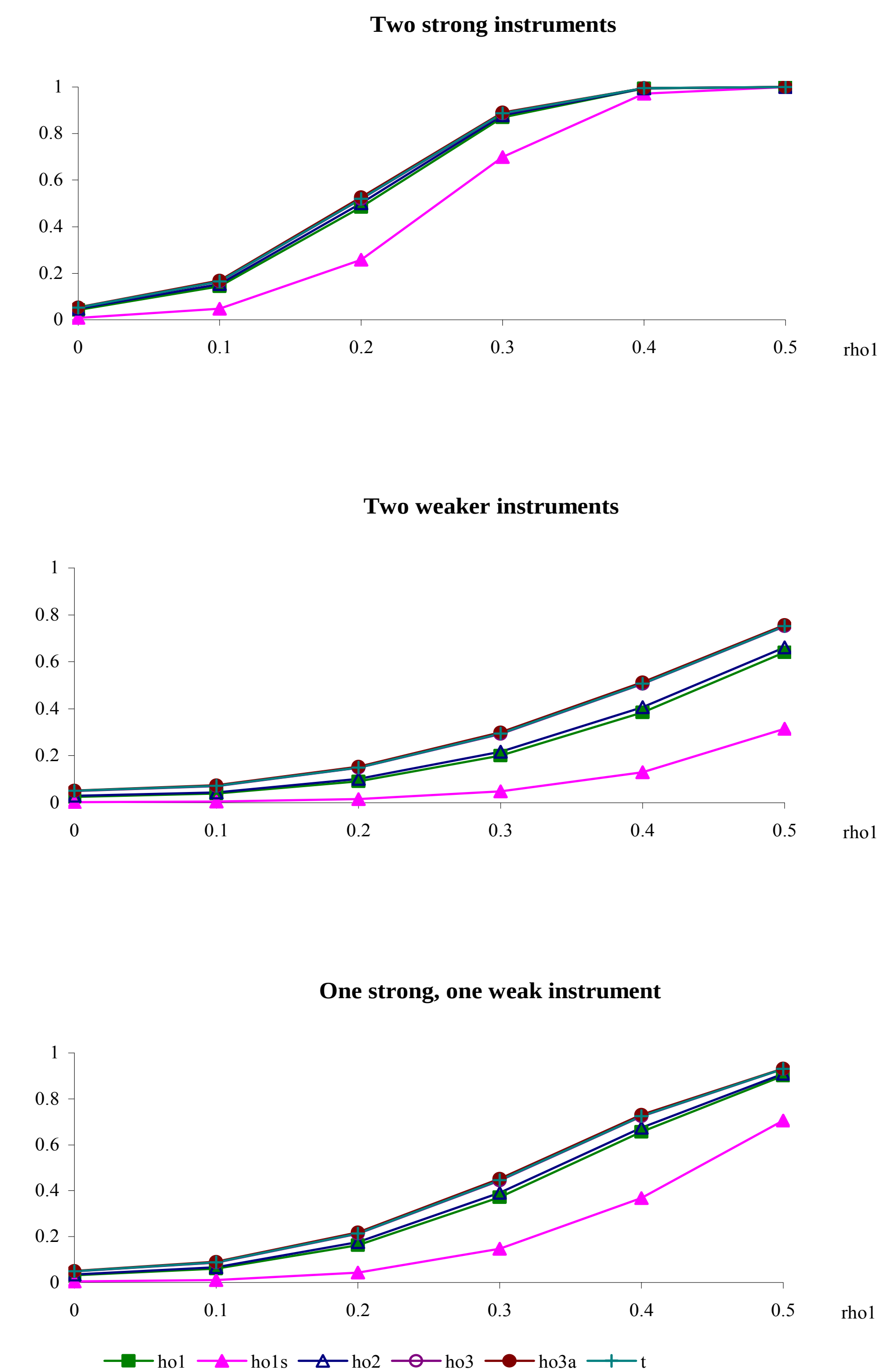


FIGURE 2.1. PERCENT REJECTIONS UNDER HOMOSKEDASTICITY (n = 100)

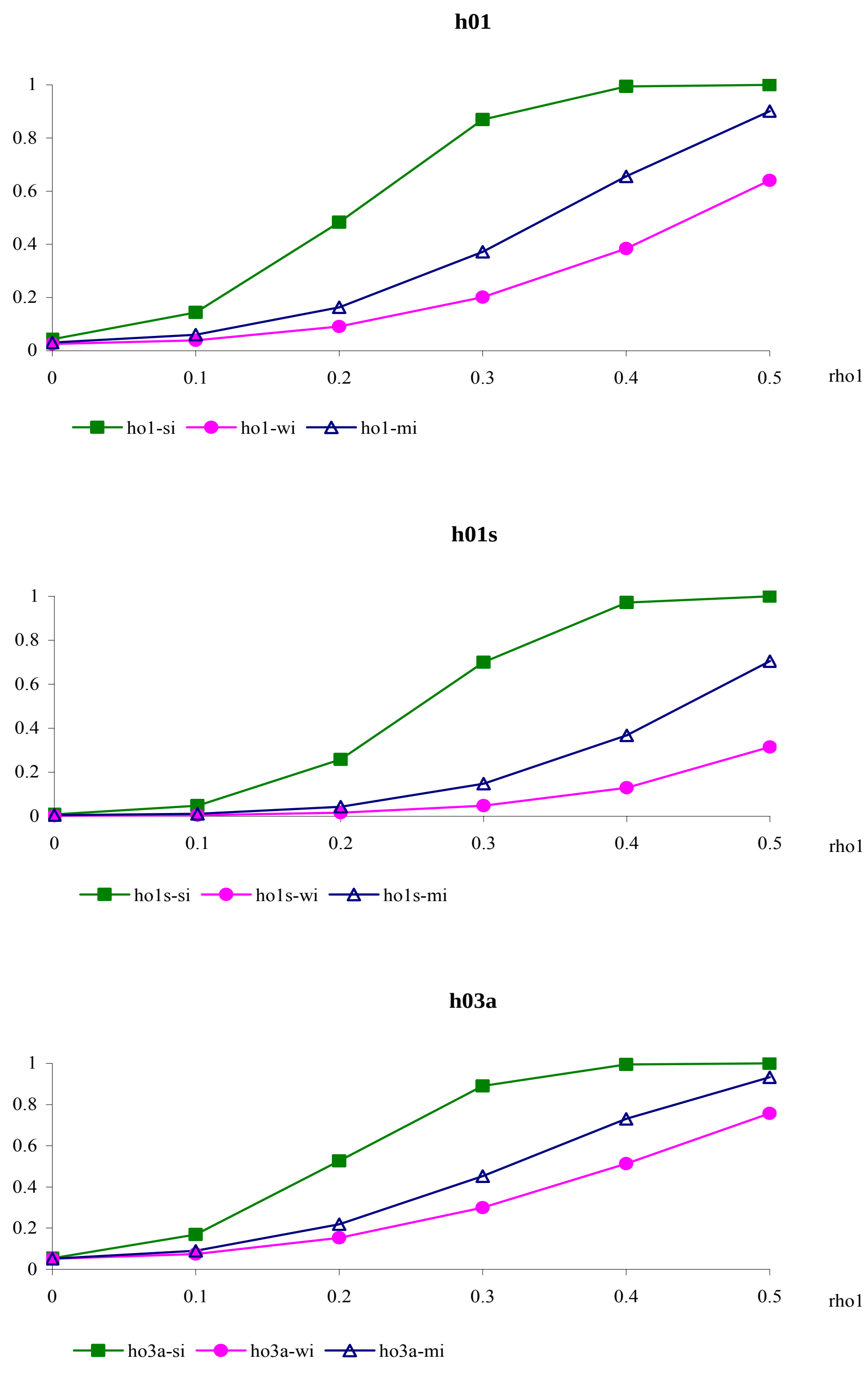


FIGURE 2.2. PERCENT REJECTIONS UNDER HOMOSKEDASTICITY - EFFECT OF INSTRUMENT STRENGTH (n = 100)

*si – two strong instruments, wi – two weaker instruments, mi - one strong, one weak instrument

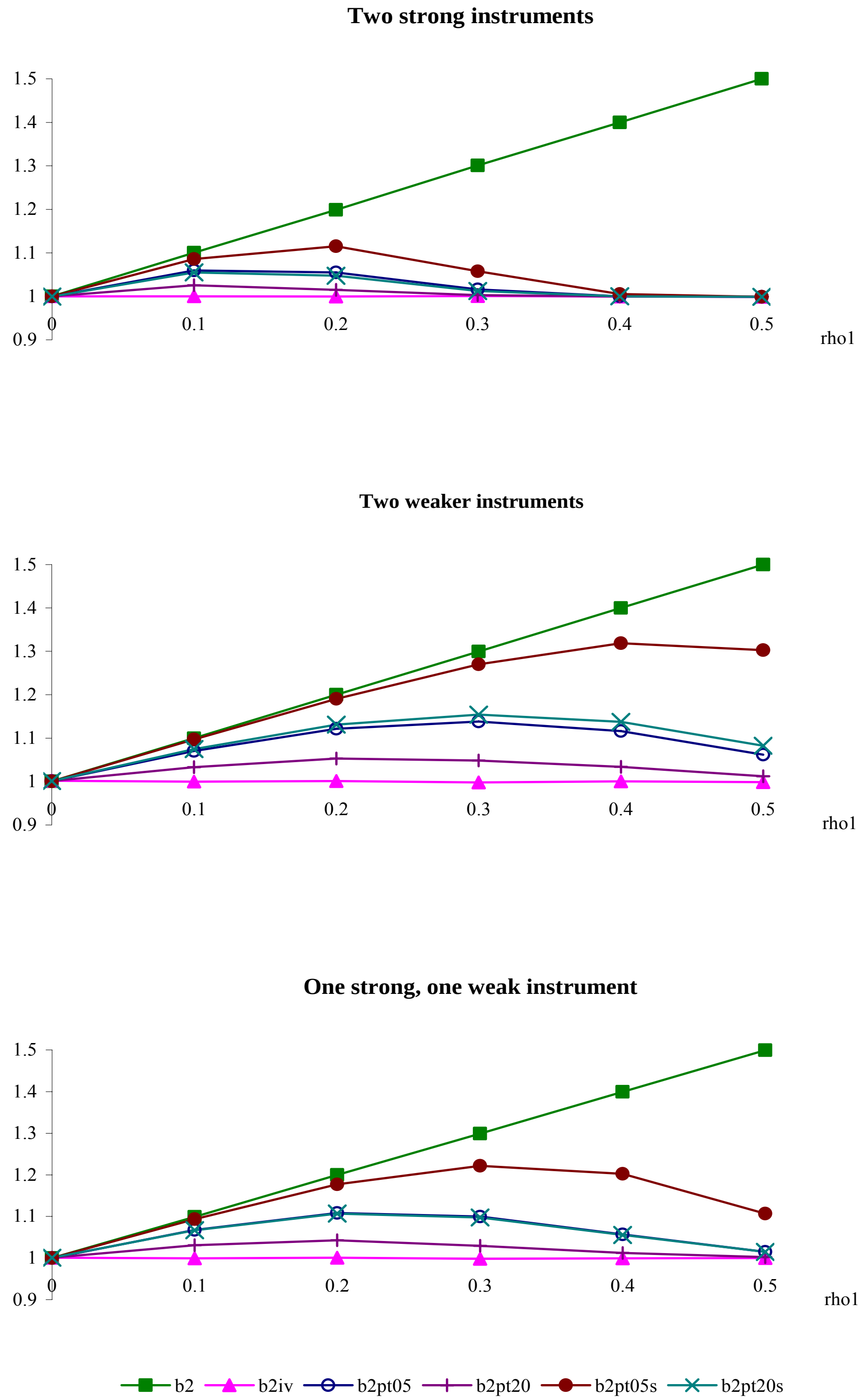


FIGURE 2.3. ESTIMATOR MEAN VALUES UNDER HOMOSKEDASTICITY (n = 100)
*pt – pre-test estimators defined in Section 2.6

not know what to do, and whose regressor may be exogenous or slightly endogenous. This seems consistent with a large company view of "corporate conservatism" and providing a product for the general, non-technical, world. Figure 2.5 compares the impact of instrument strength on root-mse's of the alternative estimators. For low degrees of endogeity there is virtually no diference among the root-mse's of the pre-test estimator based on the SAS's test given various instrument strength.

2.9 Bootstrapping under Homoskedasticity

Wong (1996) shows that bootstrapping is a good alternative to usual asymptotic theory in the homoskedastic case. Bootstrapping allows to make inferences without making strong distributional assumptions and without the need for analytic formulas for the sampling distribution parameters. It obtains the empirical distribution function from data by treating the sample as if it were the population and resampling it with replacement many times in order to generate an empirical estimate of the entire sampling distribution of a statistic. In order for the bootstrap to work we must accept that the empirical distribution function of the sample is good approximation of the population distribution function. Hall (1992) shows that bootstrap approximations converge at the rate $\sqrt{n}$. This is the same as the standard asymptotic approximations, so we should not expect bootstrap methods to improve the rate of convergence. However, Hall demonstrated that if the bootstrap is applied to asymptotically pivotal statistics, it can provide asymptotic refinements. The rate of convergence of the bootstrap is increased to n for one-sided distributions and to $\sqrt{n^3}$ for symmetrical distributions. Thus, the bootstrap provides better small sample performance than traditional asymptotic inference procedures.

We implement a bootstrapping approach to obtain superior finite sample critical values for our tests. We are interested in whether bootstrapping improves the performance of the Hausman-type tests and also the tests based on the artificial regression since the evidence especially under weak instruments is mixed. We compare the impact of the strength of the instruments and investigate whether bootstrapping is an improvement over asymptotics.

2.9.1 Bootstrapping Performance under Weak Instruments

The weak-instrument asymptotic expression for the t statistics depends non-trivially on the parameters [Staiger and Stock (1997)]. Dufour (1997) demonstrates that Wald-type statistics depending on values of a

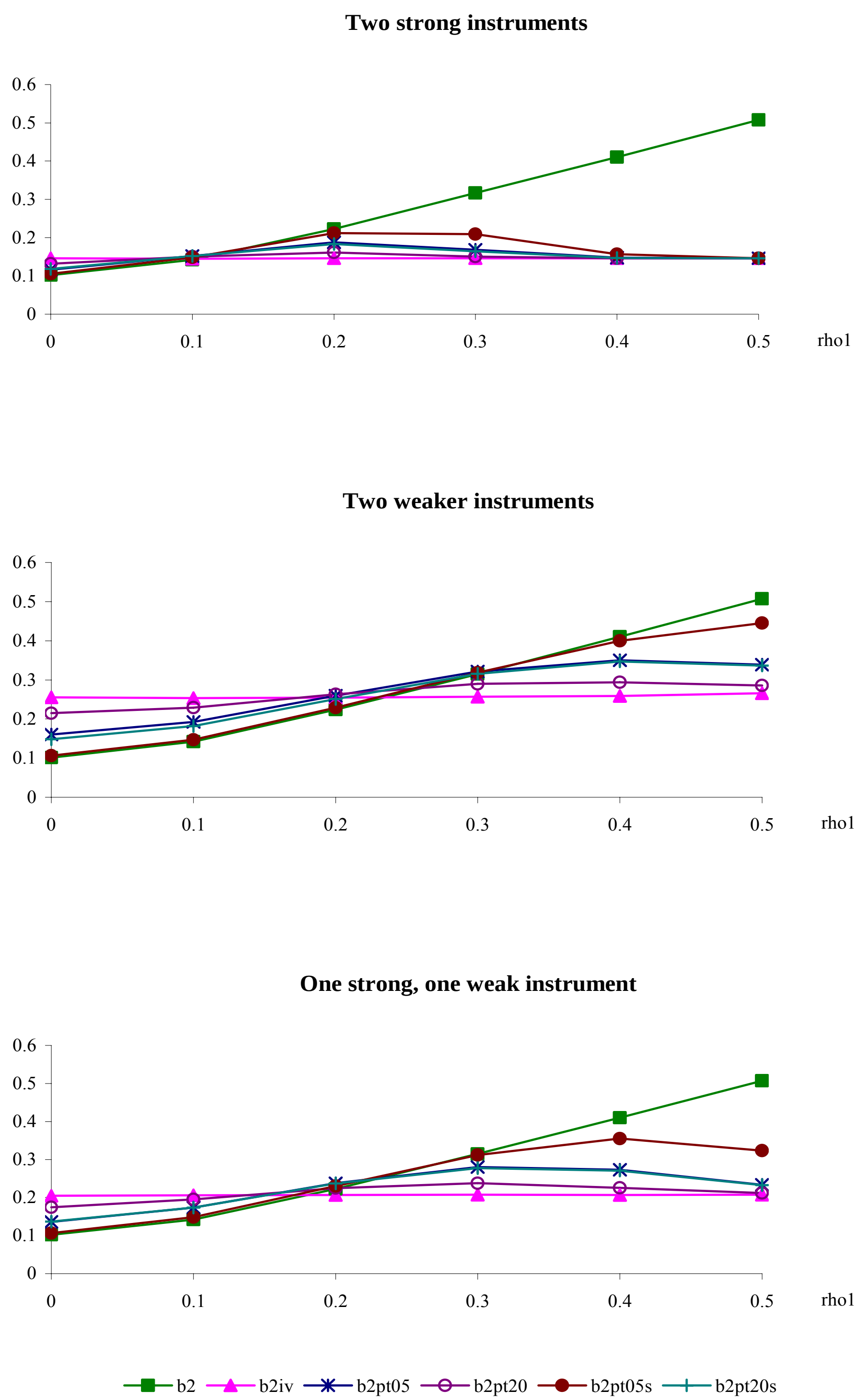
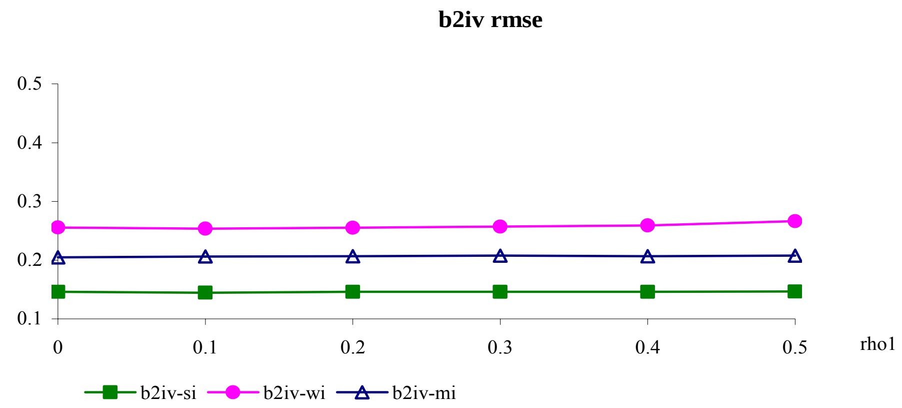
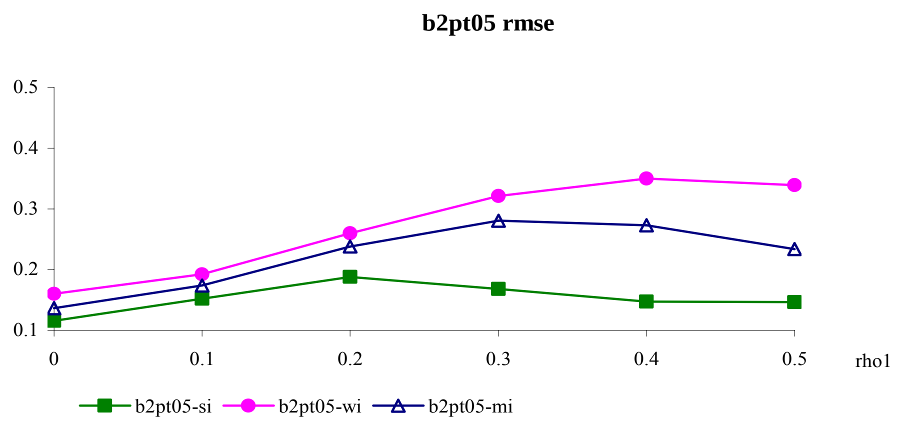


FIGURE 2.4. ESTIMATOR RMSE UNDER HOMOSKEDASTICITY (n = 100)
*pt – pre-test estimators defined in Section 2.6





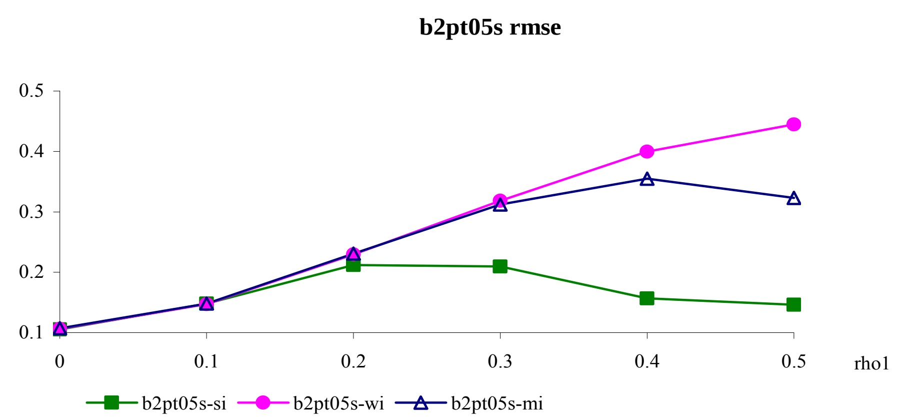


FIGURE 2.5. ESTIMATOR RMSE UNDER HOMOSKEDASTICITY - EFFECT OF INSTRUMENT STRENGTH (n = 100)
*si - two strong instruments, wi – two weaker instruments, mi - one strong, one weak instrument

locally almost unidentified parameter cannot be pivotal functions either. Edgeworth expansions of estimators and test statistics include denominator terms that are close to zero, consequently making the higher-order terms dominate the lower-order terms. Thus, the bootstrap may provide very little improvement over first-order asymptotic theory in weak-instrument cases as also shown by Hahn, Hausman, and Kuersteiner (2002), Horowitz (2001), and Rothenberg (1984). However, Wong (1996) provides contradictory evidence. He considers a just-identified model, with one potentially endogenous regressor and one instrument. The bootstrapping procedure he uses is resampling residuals obtained from the OLS regression. A basic requirement for bootstrapping to be valid is that resampling needs to be done on independently and identically distributed variables which is fullfiled if residuals are homoskedastic and the null of exogeneity of regressors is imposed. He shows that the bootstrapped Hausman test performs better than the first-order asymptotics and the improvement is significant if the instruments are weakly relevant. It should be noticed that in his tests, Wong (1996) seemed to use the incorrect number of degrees of freedom (equal to the total number of regressors) and thus the power of his asymptotic tests was too low. In his Monte Carlo experiment, there are two regressors of which one is a constant. He does not state the number of degrees of freedom used explicitly but he says that in general the number of degrees of freedom of the Hausman test is equal to K which is the total number of regressors in the model.

To improve bootstrap tests properties, Dufour (1997) suggests to use likelihood-based methods combined with projection techniques. Brown and Newey (2002) show that the empirical likelihood bootstrap provides an estimator of the *t*- and overidentification statistics that is asymptotically efficient. Their bootstrapping for GMM is based on resampling from the empirical likelihood distribution that imposes the moment restrictions rather than from the empirical distribution. Inoue (2006) and Kleibergen (2002) present experiments which also indicate that the bootstrap can lead to size improvements for the irrelevant instruments case in the GMM context.

Moreira, Porter and Suarez (2004) show that bootstrap performance depends upon regularity of the statistic examined and not upon the quality of regressors. They provide evidence that the bootstrap provides improvements for the score statistic and conditional likelihood ratio statistics up to the first order even in models with weak instruments. On the other hand, bootstrapping the Wald test offers improvements over first-order asymptotics only when instruments are good. The reason is that the Wald statistic distribution depends on the reduced form equation coefficients and thus on the instrument strength and the number of instruments.

Davidson and MacKinnon (2006) proposed new procedures for bootstrapping which use more efficient estimates of the parameters of the reduced form equation (augmented by the residuals from restricted estimation of the structural equation). Among the bootstrap Student's *t* (or Wald), Kleibergen's *K*, likelihood ratio and the Anderson-Rubin tests, the conditional (on the reduced-form covariance matrix) likelihood ratio test performed the best. The Wald test had less power than the other tests when the instruments were weak.

2.10 A Monte Carlo Experiment - Bootstrapping under Homoskedasticity

As already mentioned, the data generating process of bootstrap samples should be as close as possible to the data generating process that generated the observed data. A basic requirement for bootstrapping to be valid is that resampling needs to be done on independently and identically distributed variables.

The first step in developing bootstrap versions of a test is computing the test statistic of interest $\hat{\theta}$ in the usual way and the estimation of the model that represents the null hypothesis which is in our case OLS. We obtain the constrained (as under the null) parameter estimates $\hat{\beta}$. The bootstrap procedure is:

(1) Compute the predicted residuals $\hat{u} = y - X\hat{\beta}$

(2) Resample $\hat{u}$, obtain u^* by drawing *n* times at random with replacement from $\hat{u}$.

(3) Construct pseudo data y^* by the formula $y^* = X\hat{\beta} + u^*$.

(4) Estimate the test statistic $\hat{\theta}^*$ using X and y^*.

(5) Repeat steps (2) - (4) *B* times.

The bootstrap critical value at level α is the (1 − α) quantile of the empirical distribution function of $\hat{\theta}^*$. The bootstrap *p* value may be estimated by the proportion of bootstrap samples that yield a statistic greater than $\hat{\theta}$. This method provides asymptotic refinements.

Freedman (1981) proposed pairs bootstrap that resamples the regressand and regressors together from the original data: a pairs bootstrap sample (y^*, X^*, Z^*) is obtained by drawing rows with replacement from (y, X, Z).

The bootstrap can be used to perform also tests without asymptotic refinements. If we draw B bootstrap samples of size n with replacement yielding the bootstrap estimates $\hat{\beta}^*$ and $\tilde{\beta}^*$, the variance of the contrast of the two estimators can be estimated consistently by [Cameron and Trivedi (2005), p. 378]:

$$(2.9) \qquad \hat{V}_{Boot}\left(\hat{\beta}-\tilde{\beta}\right)=\frac{1}{B-1}\sum_{b=1}^{B}[(\hat{\beta}_b^*-\tilde{\beta}_b^*)-(\bar{\hat{\beta}}_b^*-\bar{\tilde{\beta}}_b^*)][(\hat{\beta}_b^*-\tilde{\beta}_b^*)-(\bar{\hat{\beta}}_b^*-\bar{\tilde{\beta}}_b^*)]'$$

where $\bar{\hat{\beta}}_b^* = B^{-1}\sum_{b=1}^{B}\hat{\beta}_b^*$ and $\bar{\tilde{\beta}}_b^* = B^{-1}\sum_{b=1}^{B}\tilde{\beta}_b^*$. The Hausman test statistic can then be computed as

$$(2.10) \qquad H_{Boot}=\left(\hat{\beta}-\tilde{\beta}\right)'\left(\hat{V}_{Boot}\left(\hat{\beta}-\tilde{\beta}\right)\right)^{-1}\left(\hat{\beta}-\tilde{\beta}\right)$$

Furthermore, we examine the performance of the bootstrap t test based on the auxiliary regression.

We use the same set up as in the previous sections. We perform 1,000 Monte Carlo simulations and 1,000 bootstrap replications. We consider the regression (2.7) and generate data by using (2.8). To be consistent with Wong (1996), we employ the percentile method and bootstrap the critical values.

We summarize the notation we use:

- `hb1` = bootstrap H_1 contrast with K_1 = 1 df
- `hb1s` = bootstrap H_1 contrast with K = 2 df [SAS default and STATA default with `CONSTANT` option]
- `hb2` = bootstrap H_2 contrast with IV variance estimator [`sigmaless`]
- `hb3` = bootstrap H_3 contrast with OLS variance estimator [`sigmamore`]
- `hb3a` = bootstrap H_{3a} contrast with MLE variance estimator [`Wu-Hausman chi-square`]
- `tb` = bootstrap t test of residual coefficient in auxiliary regression [`DWH F-test`]

2.11 Discussion of Bootstrapping Results under Homoskedasticity

We investigate the performance of the classical Hausman tests and the t-test under homoskedasticity. We report results of the percentile bootstrap method with asymptotic refinements. Bootstrapped samples were generated by resampling OLS residuals as in Wong (1996).

2.11.1 The Effect of Instrument Strength

In Figure 2.6 we report percent rejections of the four contrast tests and the *t*-test. With two strong instruments, the bootstrap *t*-test performs the best, the contrast tests using the difference of covariance matrices or the IV error variance estimator perform the worst. Under weak instruments, the bootstrap *t*-test performs the worst. If one strong and one weak instruments are available, the relative performance of the tests is the same as in the case with strong instruments.

In Figure 2.7.1, we compare the effect of instrument strength on individual tests. For low degrees of endogeneity, the stronger the instruments the worse the performance of the bootstrap contrast tests whereas the expected outcome (the stronger the instruments the higher the power) is true for the bootstrap *t*-test. As shown by Nelson and Startz (1990b), the finite sample distribution of the instrumental variable estimator is very different from its asymptotic counterpart. If instruments are irrelevant, the finite sample distribution collapses around a point of concentration (which is inversely related to the degree of endogeneity), at which the true density of the estimator equals zero. Further, the bias of the instrumental variable estimator may increase with low degrees of endogeneity. Thus, the performance of the contrast tests which use the instrumental variable estimator may be non-standard. Another issue could be also the small number of Monte Carlo samples.

Our results are comparable to Wong's who uses only one instrument and more Monte Carlo samples to compute the contrast test with the covariance matrix equal to the difference of covariance matrices of the respective estimators. If $\rho_1 = .5$ and $\rho_2 = .4$, the power of the bootstrap test in Wong (1996) is .741. The power of our bootstrap test is .684 if $\rho_1 = .5$, $\rho_2 = .5$ and $\rho_3 = .1$. If weak instruments is available and endogenity is weak too ($\rho_1 = .3$ and $\rho_2 = .3$), Wong's power is .162. Our experiment results in power equal to .36 if $\rho_1 = .3$, $\rho_2 = .3$ and $\rho_3 = .3$. Under low degrees of endogeneity, the power of Wong's test is not increasing the stronger the instruments.

To illustrate the importance of the type of bootstrapping used, in Figure 2.7.2 we present results for pairs bootstrapping. With pairs bootstrapping, the stronger the instruments the higher the power of the contrast test which is also valid for the *t*-test for high degrees of endogeneity. Under weak endogeneity, the *t*-test in the model that uses weak instruments reaches higher power than in the model with strong instruments. However, the differences are very small.

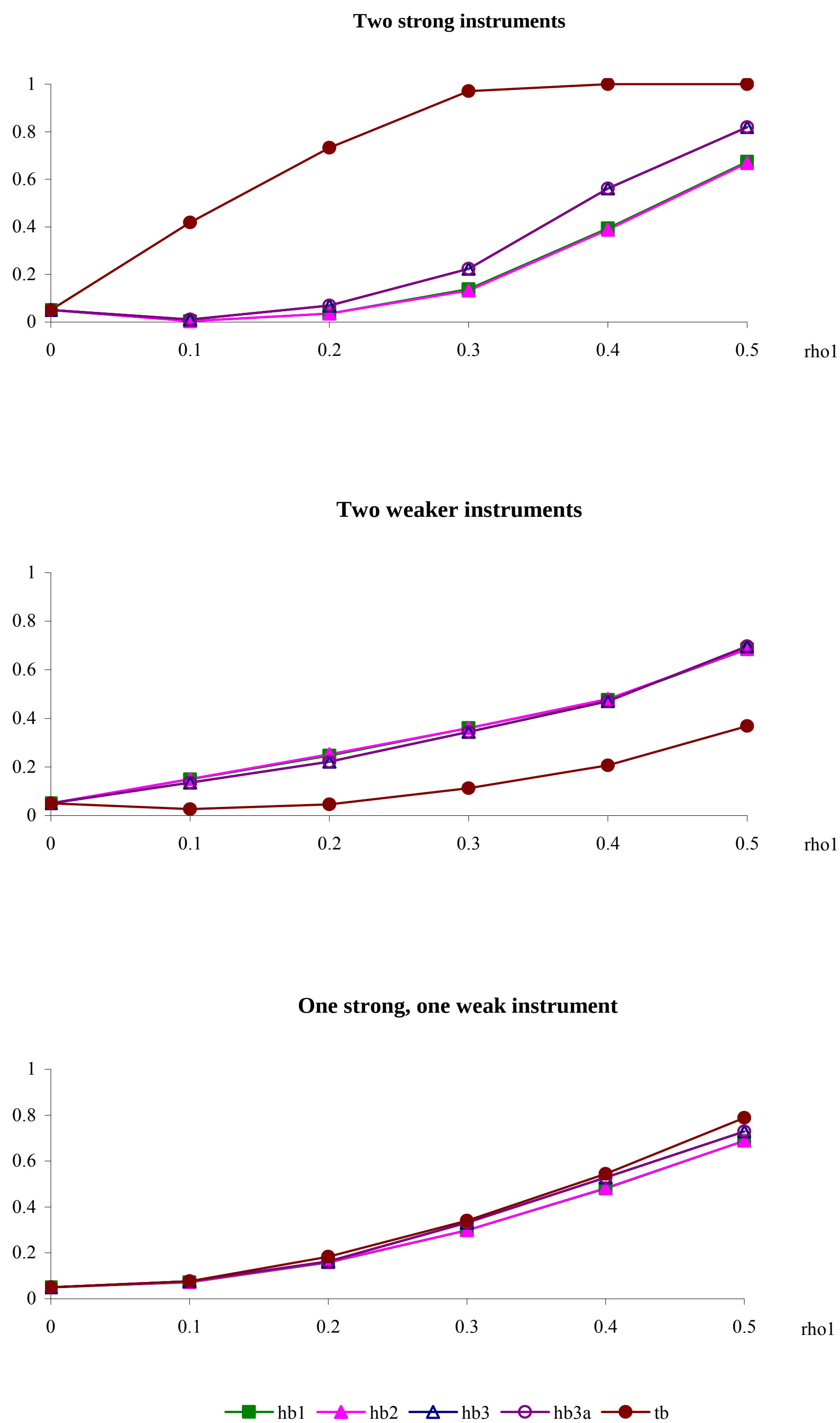
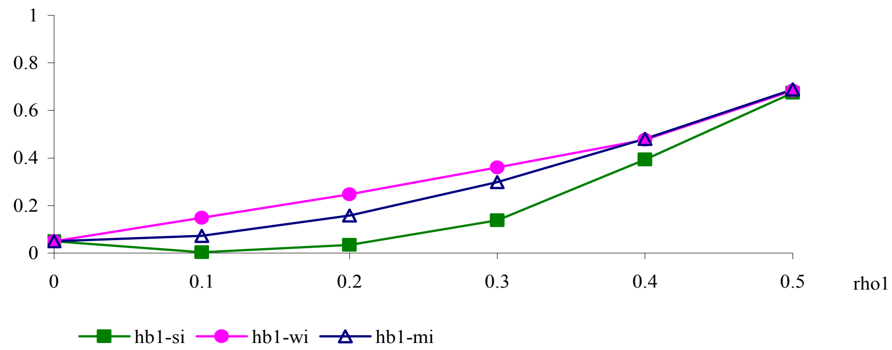
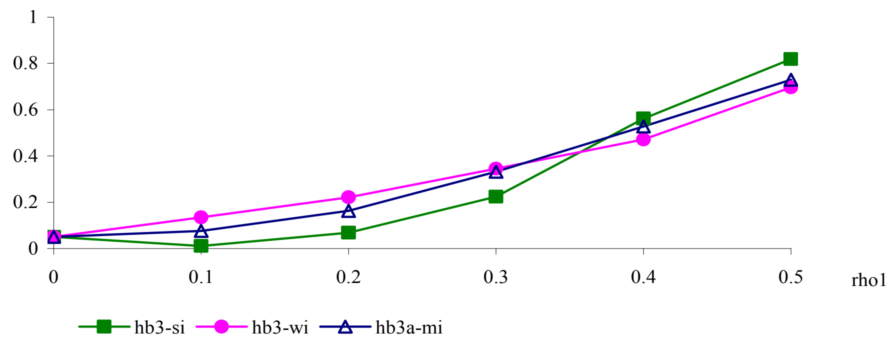


FIGURE 2.6. PERCENT REJECTIONS (BOOTSTRAPPING) UNDER HOMOSKEDASTICITY (n = 100)

hb1



hb3



tb

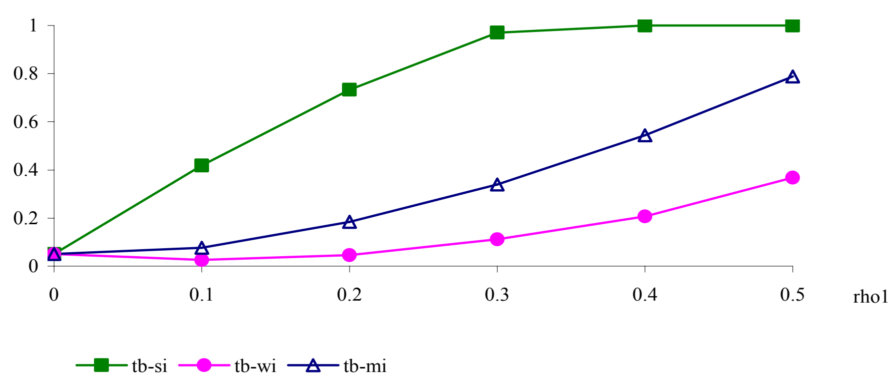


FIGURE 2.7.1. PERCENT REJECTIONS (BOOTSTRAPPING) UNDER HOMOSKEDASTICITY - EFFECT OF INSTRUMENT STRENGTH (n = 100)
Residual bootstrapping
*si – two strong instruments, wi – two weaker instruments, mi **-** one strong, one weak instrument

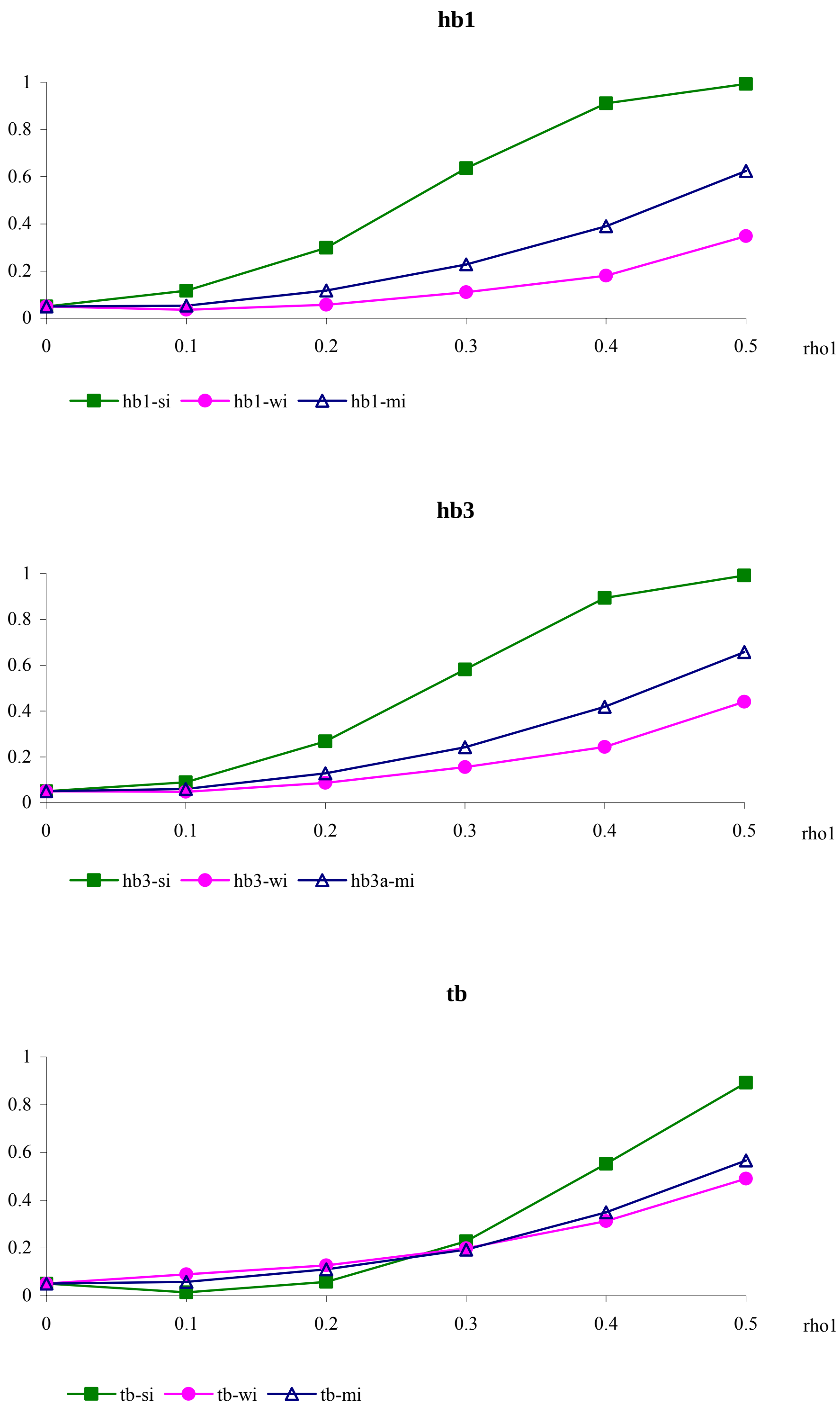
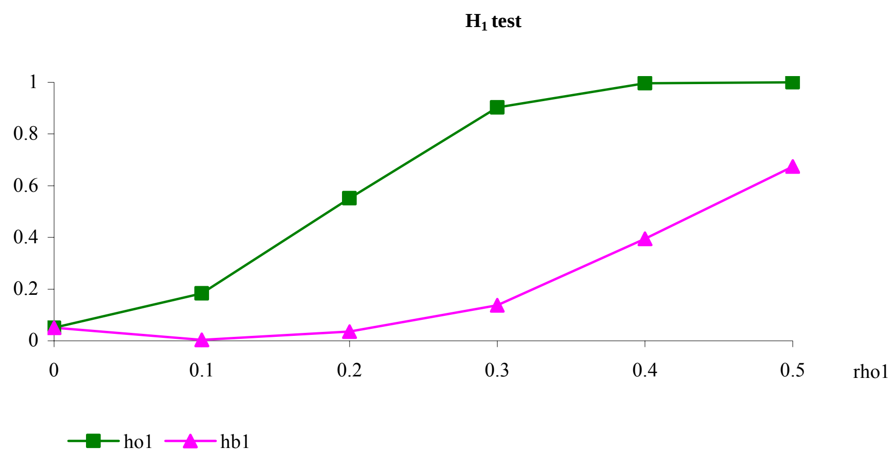
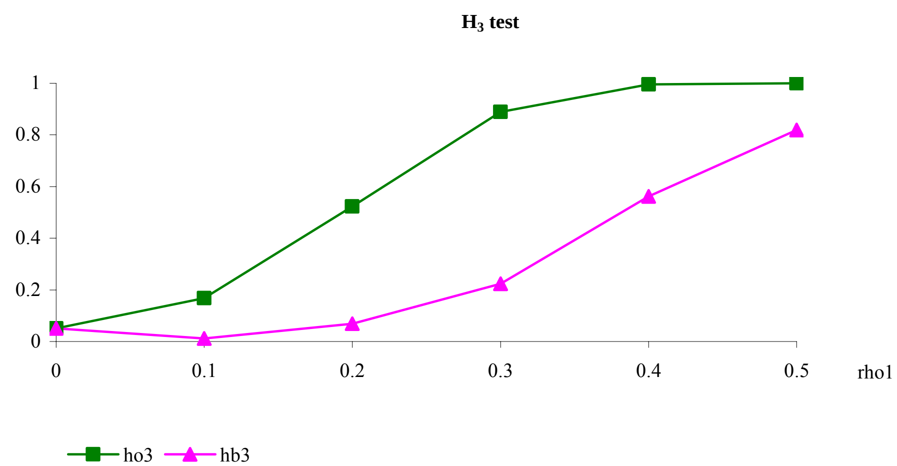


FIGURE 2.7.2. PERCENT REJECTIONS (BOOTSTRAPPING) UNDER HOMOSKEDASTICITY - EFFECT OF INSTRUMENT STRENGTH (n = 100)
Pairs bootstrapping
*si – two strong instruments, wi – two weaker instruments, mi **-** one strong, one weak instrument





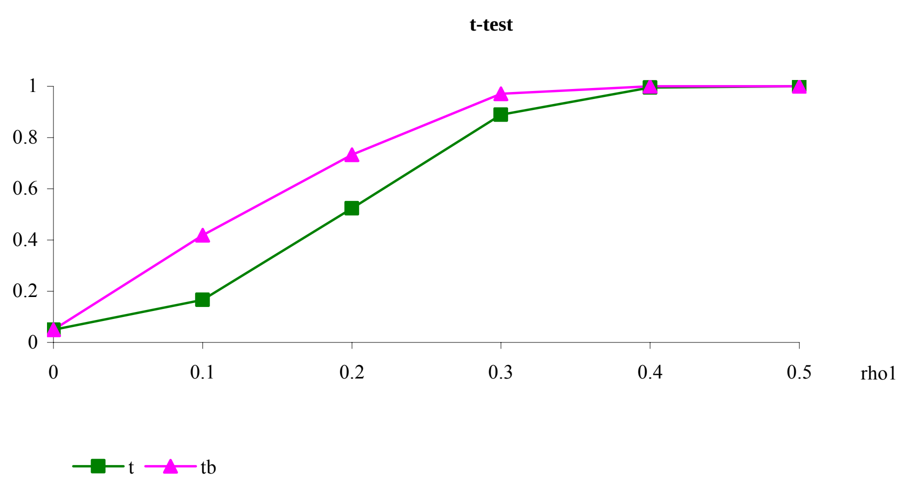


FIGURE 2.8.1. PERCENT REJECTIONS UNDER HOMOSKEDASTICITY - EFFECT OF BOOTSTRAPPING (n = 100)
Strong instruments

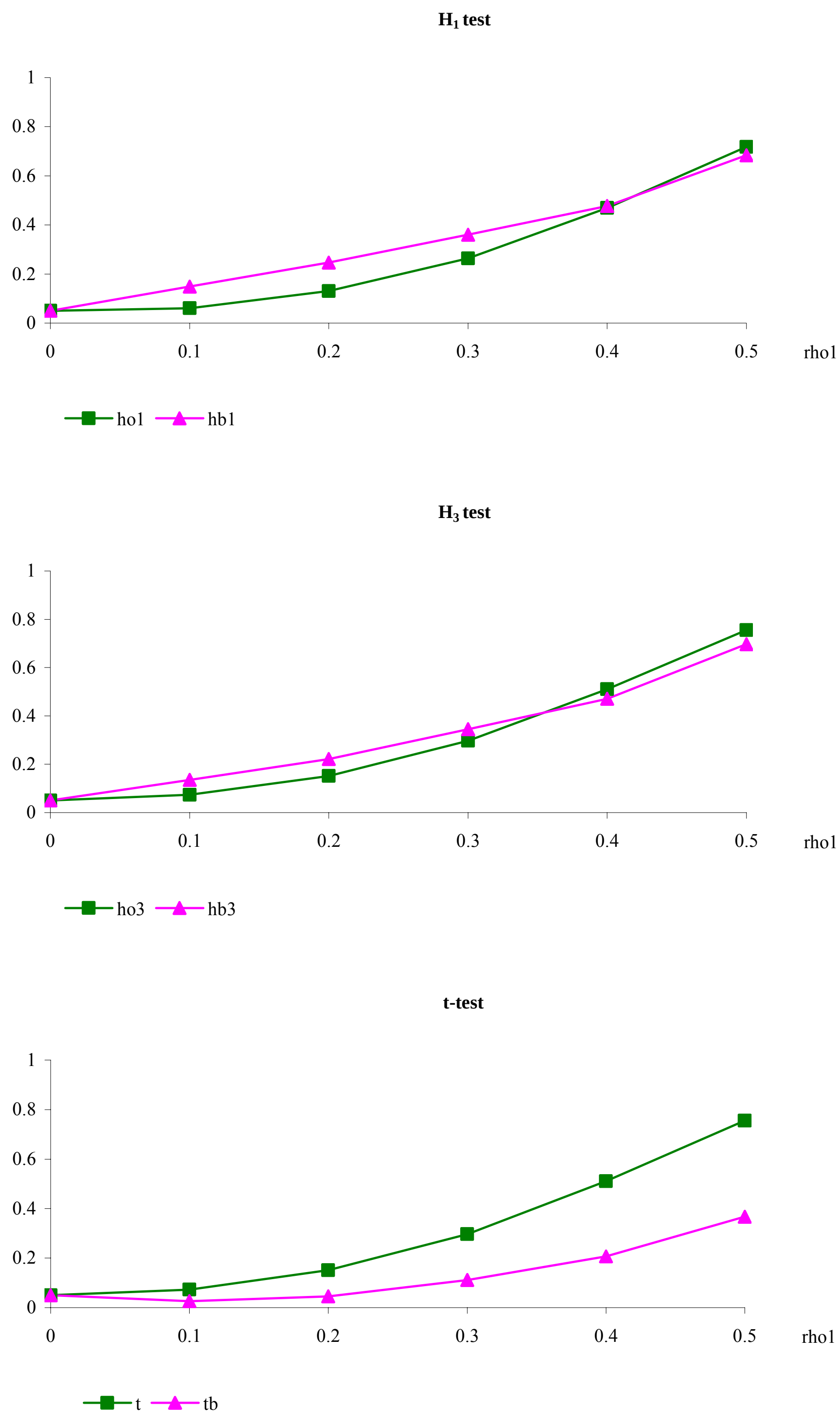


FIGURE 2.8.2. PERCENT REJECTIONS UNDER HOMOSKEDASTICITY - EFFECT OF BOOTSTRAPPING (n = 100)
Weaker instruments

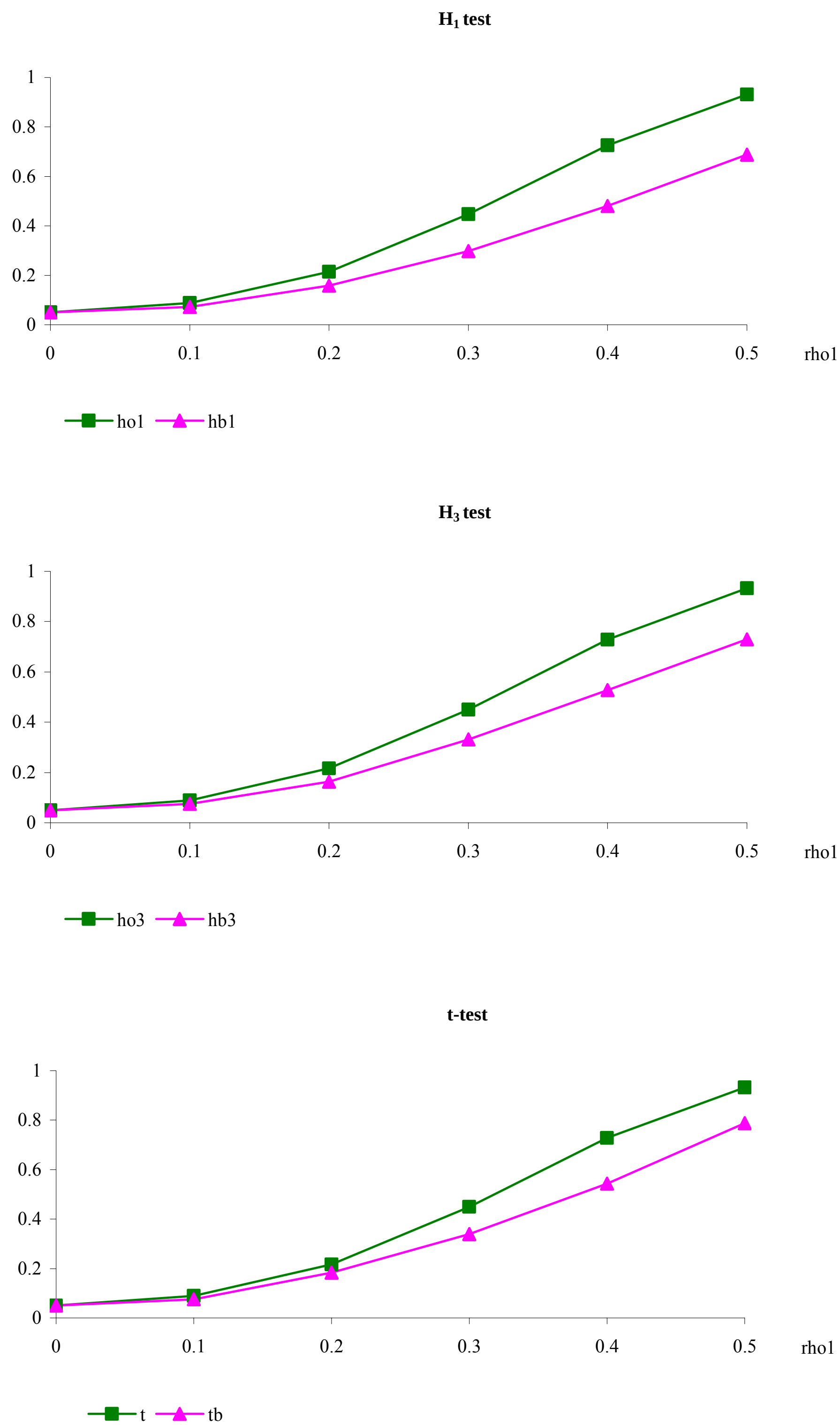


FIGURE 2.8.3. PERCENT REJECTIONS UNDER HOMOSKEDASTICITY - EFFECT OF BOOTSTRAPPING (n = 100)
One strong, one weak instrument

2.11.2 The Effect of Bootstrapping under Homoskedasticity

In Figure 2.8 we compare the effect of bootstrapping on individual tests. We again use the residual bootstrapping approach and distinguish between a strong instruments case, a weak instrument case and a one strong and one weak instrument case. It should be noticed that the number of Monte Carlo samples for the asymptotic and bootstrapped tests is different because the length of the bootstrap experiment would be enourmous if the same number was used. Bootstrapping improves the performance of the t-test under strong instruments and the performance of the contrast tests under weak instruments for low degrees of endogeneity, otherwise the asymptotic tests perform better. Wong's bootstrap tests are an improvement over asymptotic results but it seems that in his asymptotic tests, he used the number of degrees of freedom that is equal to the total number of regressors and not to the number of potentially endogenous variables. Bootstrap tests seem to be sensitive also to the number of instruments used, conclusions made for just-identified models cannot be generalized and applied to over-identified models.

2.12 Summary of Findings under Homoskedasticity

When endogeneity is absent or weak, the default tests in STATA and SAS lead to pretest estimators with lower root mean squared error than pretest estimators based on generally superior tests. In terms of test performance, changing the degrees of freedom to K_1 improves the test, but it does not produce the best test among those considered. The best (and easiest) test to use of the F or t test in an artificial regression that takes the original specification and augments it with the residual vectors from the OLS estimates of the reduced form equations that express each potentially endogenous variable as a function of all available instruments. The preferred test is a significance test of these included residual vectors. The tests are affected by instrument strength. Having more and stronger instruments is preferred, but one strong and one weak instrument, in our case of a single endogenous regressor, is better than having two "middling" instruments. Collinearity between the instruments does slightly reduce the performance of tests.

The performance of the bootstrapped tests is not an improvement over the asymptotic tests except the t-test in the model with two strong instruments and the H_1 and H_{3a} tests in the model with two weak instruments for low degrees of endogeneity. These results are very sensitive to the type of bootstrapping used.

3. HAUSMAN TESTS WITH HETEROSKEDASTICITY

The difficulty with the Hausman test in a general context , for example if heteroskedasticity is present, is that the variance of the contrast,

$$V\left(\hat{\beta}-\tilde{\beta}\right)=V\left(\hat{\beta}\right)+V\left(\tilde{\beta}\right)-2\,\mathrm{cov}\left(\hat{\beta},\tilde{\beta}\right)$$

includes the term $\mathrm{cov}\left(\hat{\beta},\tilde{\beta}\right)$ that is not routinely calculated. There are several approaches one might take. The first, using GMM and following Creel (2004) is to directly estimate the covariance term. The second, as suggested by Cameron and Trivedi (2005, p. 378) is to use bootstrapping. The third is to follow the path of Davidson and MacKinnon (1993, p. 399) and use a robust artificial regression, also suggested by Wooldridge (2002, p. 119).

In the following, we describe the approaches based on GMM which include (i) the Creel's (2004) system (ii) the naive contrast test between OLS and GMM, (iii) the auxiliary regression *F*-test that is equivalent to this contrast, and (iv) the test based on the difference in the GMM objective functions. We compare the performance of all of the GMM based tests with the standard tests, first ignoring heteroskedasticity. This includes the artificial regression tests based on the *F* or *t*-distributions [`DWH F-test`], the contrast test with the common estimate of the error variance being from the OLS [`sigmamore`] or IV [`sigmaless`] regressions, and the chi-square test using the ML estimate of the error variance [`Wu-Hausman chi-square`]. Then, we make the auxiliary regression tests robust to heteroskedasticity and investigate whether the performance improved relative to the standard tests ignoring the problem.

3.1 Direct Testing in a GMM Framework

We wish to estimate the linear regression model $y = X\beta + u$ and test for endogeneity in the presence of heteroskedastic errors . We will assume that observations are cross-sectional and that the errors u_i and u_j are uncorrelated, but that the errors are heteroskedastic with $\mathrm{var}\left(u_i\right)=\sigma_i^2$. Creel (2004) suggests that we combine the sets of moment conditions that lead to OLS and IV estimators into a single estimation problem. The moment conditions leading to OLS are

$$E\left[x_i\left(y_i - x_i'\beta\right)\right] = 0,$$

and those leading to the IV estimator are

$$E\left[z_i\left(y_i - x_i'\beta\right)\right] = 0.$$

These have the sample analogs

$$h_1 = \frac{1}{n}\sum h_{1i} = \frac{1}{n}\sum x_i\left(y_i - x_i'\beta_1\right)$$

and

$$h_2 = \frac{1}{n}\sum h_{2i} = \frac{1}{n}\sum z_i\left(y_i - x_i'\beta_2\right)$$

Then we stack these two sets of moment conditions into

$$h = \frac{1}{n}\sum h_i = \frac{1}{n}\sum \begin{bmatrix} h_{1i} \\ h_{21} \end{bmatrix}$$

In this context the GMM estimator [see Cameron and Trivedi (2005) p. 173 and following] minimizes

$$Q(\beta) = \left[\frac{1}{n}\sum h_i\right]' W_n \left[\frac{1}{n}\sum h_i\right]$$

where $\beta' = \left(\beta_1' \; \beta_2'\right)$. The optimal weight matrix W_n is the inverse of the asymptotic covariance matrix of the moment conditions. That is, if

$$S_0 = plim \frac{1}{n}\sum h_i h_i'$$

then the optimal weight matrix is $W_0 = S_0^{-1}$. If $G_0 = E\left[\partial h / \partial \beta'\right]$, then [Cameron and Trivedi (2005), p. 176]

$$\sqrt{n}\left(\hat{\beta}_{OGMM} - \beta\right) \xrightarrow{d} N\left[0, \left(G_0' S_0^{-1} G_0\right)^{-1}\right]$$

The natural estimator of for S_0 is

$$\hat{S} = \frac{1}{n}\sum h_i\left(\hat{\beta}\right) h_i\left(\hat{\beta}\right)'. \tag{3.1}$$

Then, the two-step GMM estimator minimizes

$$Q(\beta) = \left[\frac{1}{n}\sum h_i\right]' \hat{S}^{-1} \left[\frac{1}{n}\sum h_i\right] \tag{3.2}$$

The optimal estimator $\hat{\beta}_{OGMM}$ has estimated asymptotic covariance matrix

$$V\left(\hat{\beta}_{OGMM}\right)=\frac{1}{n}\left(\hat{G}'\hat{S}^{-1}\hat{G}\right)^{-1}$$

where

$$\hat{G}=\frac{1}{n}\sum\frac{\partial h_i}{\partial\beta'}\bigg|_{\hat{\beta}}$$

In our case

$$\hat{G}=\frac{1}{n}\begin{bmatrix}-X'X & 0\\ 0 & -Z'X\end{bmatrix}$$

Now let us assume that in the "first round" the weight matrix is block diagonal

$$W_n=\begin{bmatrix}W_1 & 0\\ 0 & W_2\end{bmatrix}$$

Assume that $W_1 = I_K$ and that W_2 is optimal considering only the moment conditions h_2, then the first round estimates of $\beta'=\left(\beta_1'\ \beta_2'\right)$ are $\hat{\beta}_1=\hat{\beta}_{OLS}$, the OLS estimator of the original model $y=X\beta+u$ and $\hat{\beta}_2=\hat{\beta}_{GMM}$, the optimal two-step estimator of the original model using instruments *Z*. Using these first round estimates we can obtain $\hat{S}$ and then the two-step GMM estimator

$$\text{(3.3)}\qquad \hat{\beta}_{GMM\,2}=\left\{\begin{bmatrix}X'X & 0\\ 0 & Z'X\end{bmatrix}'\hat{S}^{-1}\begin{bmatrix}X'X & 0\\ 0 & Z'X\end{bmatrix}\right\}^{-1}\begin{bmatrix}X'X & 0\\ 0 & Z'X\end{bmatrix}'\hat{S}^{-1}\begin{bmatrix}X'y\\ Z'y\end{bmatrix}$$

with asymptotic covariance matrix

$$\text{(3.4)}\qquad V\left(\hat{\beta}_{GMM\,2}\right)=n\left\{\begin{bmatrix}X'X & 0\\ 0 & Z'X\end{bmatrix}'\hat{S}^{-1}\begin{bmatrix}X'X & 0\\ 0 & Z'X\end{bmatrix}\right\}^{-1}$$

Then, a Wald test of the null hypothesis that $H_0: plim\left(\hat{\beta}_{OLS}-\hat{\beta}_{GMM}\right)=0$ can be based upon their "improved" versions in $\hat{\beta}_{GMM\,2}$ and using $V\left(\hat{\beta}_{GMM\,2}\right)$. If $R=\left[I_K\quad -I_K\right]$ then the test statistic is

$$\text{(3.5)}\qquad W_2=\left(R\hat{\beta}_{GMM\,2}\right)'\left[RV\left(\hat{\beta}_{GMM\,2}\right)R'\right]^{-1}\left(R\hat{\beta}_{GMM\,2}\right)$$

Under the null hypothesis the test statistic W_2 presumably has an asymptotic distribution with K degrees of freedom, or the rank of $RV\left(\hat{\beta}_{GMM2}\right)R'$ if that is less than K. Note that $X = \left[X_1 \quad X_2\right]$ and $Z = \left[Z_1 \quad X_2\right]$, thus asymptotically the covariance matrix of the two sets of moment conditions $\hat{S}$ (3.1) will be singular (the number of linearly independent columns will be less than $2K$). In a just identified case, the rank of $\hat{S}$ should equal K_1 since we are testing the endogeneity of K_1 regressors, consequently W_2 has K_1 degrees of freedom too.

As an alternative test we can focus on the coefficients of the potentially endogenous regressors X_1. If $R_1 = \left[I_{K_1} \quad 0 \quad -I_{K_1} \quad 0\right]$ the corresponding Wald statistic will be called W_{2a} and it will have K_1 degrees of freedom.

Creel (2004) cites the findings of Burnside and Eichenbaum (1996) that Wald tests based on GMM estimators have improved properties when restrictions implied by the model, or null hypothesis, are imposed. With that in mind we obtain the restricted GMM estimator by minimizing the objective function (3.2) subject to the null hypothesis that within $\beta' = \left(\beta_1' \; \beta_2'\right)$ the equality $\beta_1 = \beta_2 = \overline{\beta}$ holds. The resulting estimates

$$(3.6) \qquad \hat{\beta}_r = \begin{bmatrix} \hat{\overline{\beta}} \\ \hat{\overline{\beta}} \end{bmatrix}$$

are used to obtain $\hat{S}_r = \frac{1}{n}\sum h_i\left(\hat{\beta}_r\right)h_i\left(\hat{\beta}_r\right)'$ where the subscript "r" denotes this restricted estimation result. The corresponding estimated asymptotic covariance matrix is

$$(3.7) \qquad V\left(\hat{\beta}_{GMM2r}\right) = n\left\{\begin{bmatrix} X'X & 0 \\ 0 & Z'X \end{bmatrix}' \hat{S}_r^{-1} \begin{bmatrix} X'X & 0 \\ 0 & Z'X \end{bmatrix}\right\}^{-1}$$

We compute analogs of W_2 and W_{2a}, which we call W_{2r} and W_{2ar}, with degrees of freedom under the null hypothesis (that X_1 is not endogenous) of K (but maybe K_1) and K_1, respectively.

While focusing on GMM estimation, we offer one more test. Cameron and Trivedi (2005, p. 245) note that for the efficient GMM estimator the difference test statistic

$$(3.8) \qquad D = n\left[Q\left(\hat{\beta}_r\right) - Q\left(\hat{\beta}_u\right)\right]$$

has an asymptotic chi-square distribution with number of degrees of freedom equal to the number of hypotheses imposed to obtain $\hat{\beta}_r$. To implement this test we compare the objective function (3.1) at the estimates $\hat{\beta}_{GMM2}$ [from (3.2)] and $\hat{\beta}_r$ from (3.4). This test statistic should have degrees of freedom *K* under the null hypothesis.

3.2 Testing with a Robust Artificial Regression

Consider again the contrast in (2.1), $q = \hat{\beta}_{IV} - \hat{\beta}_{OLS} = \left(X'P_Z X\right)^{-1} X'P_Z M_X y$. The question, as before, is whether $X'P_Z M_X y$, or its relevant part, $X_1'P_Z M_X y$ has zero mean asymptotically. Again consider the artificial regression (2.2) and apply the FWL theorem

(3.9) $M_X y = M_X P_Z X_1 \delta + error \Rightarrow \tilde{y} = \tilde{P}\delta + error$.

Following Davidson and MacKinnon (1993, p. 401), we assume that the regression errors *u* are heteroskedastic, so that $Euu' = \Omega = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right)$, and that

$$plim \frac{1}{n} \tilde{P}'\Omega\tilde{P}$$

exists and can be consistently estimated, under the null δ = 0, by

$$\frac{1}{n} \tilde{P}'\hat{\Omega}\tilde{P}$$

where $\hat{\Omega} = diag\left(\hat{u}_1^2, \ldots, \hat{u}_n^2\right)$ and $\hat{u}_i = y_i - x_i'\hat{\beta}_{OLS}$. The least squares estimator $\hat{\delta} = \left(\tilde{P}'\tilde{P}\right)^{-1} \tilde{P}'\tilde{y}$ has an asymptotic covariance matrix that can be estimated by

(3.10) $V\left(\hat{\delta}\right) = \left(\tilde{P}'\tilde{P}\right)^{-1} \tilde{P}'\hat{\Omega}\tilde{P}\left(\tilde{P}'\tilde{P}\right)^{-1}$

Thus the Wald statistic for the null hypothesis that δ = 0 is

(3.11) $W = \hat{\delta}'\left[V\left(\hat{\delta}\right)\right]^{-1} \hat{\delta} = y'\tilde{P}\left(\tilde{P}'\hat{\Omega}\tilde{P}\right)^{-1} \tilde{P}'y \overset{a}{\sim} \chi^2_{(K_1)}$

Once again, the robust Wald or *F*-test can be based on the artificial regression (2.5) that includes the residuals of the reduced form regression for the potentially endogenous variable(s) instead of their predicted value(s).

We could have done things more directly by considering $X_1' P_Z M_X y = X_1' P_Z M_X u = \tilde{P}' u$. If we assume that $n^{-\frac{1}{2}} \tilde{P}' u \xrightarrow{d} N\left(0, plim\left[n^{-1} \tilde{P}' \Omega \tilde{P}\right]\right)$ Then

$$W = \left(n^{-\frac{1}{2}} \tilde{P}' u\right)' \left[plim\left(n^{-1} \tilde{P}' \Omega \tilde{P}\right)\right]^{-1} \left(n^{-\frac{1}{2}} \tilde{P}' u\right)$$
$$= \left(n^{-\frac{1}{2}} \tilde{P}' y\right)' \left[plim\left(n^{-1} \tilde{P}' \Omega \tilde{P}\right)\right]^{-1} \left(n^{-\frac{1}{2}} \tilde{P}' y\right)$$
$$\cong y' \tilde{P} \left(\tilde{P}' \hat{\Omega} \tilde{P}\right)^{-1} \tilde{P}' y \overset{a}{\sim} \chi^2_{(K_1)}$$

3.3 Another Robust Artificial Regression Test

Suppose we form a contrast between OLS and another linear estimator that is consistent if some regressors are endogenous. Let it take the form [MacKinnon, 1992, p. 125]

(3.12) $\hat{\beta}_A = \left(X'AX\right)^{-1} X'Ay$

where A is $n \times n$ and with rank no less than K. Then the contrast with the OLS estimator is

(3.13) $q = \hat{\beta}_A - \hat{\beta}_{OLS} = \left(X'AX\right)^{-1} X'AM_X y$

and the question is whether $X'AM_X y$ has zero mean asymptotically. Going through the same steps as in Section 3.2 we can form the artificial regression

(3.14) $y = X\beta^* + AX\delta + error$

which leads to the estimator

(3.15) $\hat{\delta} = \left(X'AM_X AX\right)^{-1} X'AM_X Ay$

and a robust Wald test statistic for the hypothesis $\delta = 0$ of

(3.16) $W_{DM} = y'M_X AX\left(X'AM_X \hat{\Omega} M_X AX\right)^{-1} X'AM_X y \overset{a}{\sim} \chi^2_{(rank[\bullet])}$

where the degrees of freedom are the rank of the matrix inversed in the quadratic form. The rank is nominally K but $M_X AX$ may not have rank K depending on the choice of A.

One logical choice for A, since we would like to form the contrast with the most efficient consistent estimator is $Z\hat{S}^{-1}Z'$, so that the estimator $\hat{\beta}_A = \hat{\beta}_{GMM}$, which is the optimal two-step GMM estimator of β in the original model $y = X\beta + u$. In this case $\hat{S} = Z'D_nZ$ where D_n is a diagonal matrix of squared residuals based either on the OLS or IV estimators.

3.4 The Estimators under Heteroskedasticity

As in the homoskedastic case, we report the Monte Carlo means of the alternative estimators and their root-mean-squared-errors (rmse) for the parameter β_2 (2.7). The slope parameter, and coefficient of the potentially endogenous regressor *x*, is of primary interest. The estimators whose performance we report are OLS (b_{2OLS}), instrumental variables (b_{2IV}) and optimal 2-step GMM (b_{2GMM}). In addition we report results for the pre-test estimators using the usual *t*-statistic. This estimator is

$$b_{2pt\alpha} = \begin{cases} b_{2OLS} & \text{if } t < t_c \\ b_{2IV} & \text{if } t \geq t_c \end{cases}$$

where t_c is the critical value for a *t*-distribution with $n - K - 1$ (degrees of freedom from the artificial regression) at significance level α = .05 (b_{2pt05}) or α = .20 (b_{2pt20}). Similarly we define a pre-test estimator defined on the default SAS statistic H_{1s}

$$b_{2pt\alpha s} = \begin{cases} b_{2OLS} & \text{if } H_{1s} < \chi_c^2 \\ b_{2IV} & \text{if } H_{1s} \geq \chi_c^2 \end{cases}$$

Finally based on the robust *t*-test **tr**, we define a pretest estimator that chooses between the OLS estimator and the 2-step optimal GMM estimator,

$$b_{2prt\alpha} = \begin{cases} b_{2OLS} & \text{if } t < t_c \\ b_{2GMM} & \text{if } t \geq t_c \end{cases}$$

3.5 A Monte Carlo Experiment under Heteroskedasticity

We use the same design as under homoskedasticity

(3.17) $y = \beta_1 + \beta_2 x + u = 0 + x + u$

Data are generated by specifying

(3.18) $$\begin{bmatrix} x \\ v \\ z_1 \\ z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & 0 & 0 \\ \rho_2 & 0 & 1 & \rho_4 \\ \rho_3 & 0 & \rho_4 & 1 \end{bmatrix} \right)$$

The key features are that

- endogeneity is controlled by the parameter ρ_1 which takes the values 0, .1, .2, .3, .4 and .5
- the strength of the instruments is controlled by ρ_2 and ρ_3 which take the values .1, .3 and .5
- the correlation between the instruments is controlled by ρ_4 which takes the values 0 and .5
- samples of sizes n = 100 and n = 200 are considered.

The error u in (3.17) is made heteroskedastic by specifying

(3.19) $u = (1 + \gamma x)v$

- The parameter γ controls the degree of heteroskedasticity and takes the values 0 (homoskedasticity), .5 and 1.

To give some idea of consequences of these choices consider the estimation results based on one sample for various design points (Table 3.1). Columns 1-3 demonstrate the effect of the heteroskedasticity control parameter γ. The Breusch-Pagan LM test shows significant heteroskedasticity with sample size n = 100.

TABLE 3.1. RESULTS FOR A SINGLE SAMPLE FOR VARIOUS MC DESIGNS

DESIGN	1	2	3	4	5	6	7	8
n (sample)	100	100	100	100	100	200	200	100
γ (hetero)	0	.5	1	1	1	1	1	1
ρ_1 (endog)	0	0	0	.5	.5	.5	.5	.5
ρ_2 (IV1)				.3	.5	.5	.3	.3
ρ_2 (IV2)				.3	.1	.5	.1	.1
Bols	.944	.892	.83	1.37	1.37	1.27	1.27	1.37
Bpagan		20.86	37.16	15.9	15.9	24.66	24.66	15.9
1st Stage F				15.36	13.47	113.83	13.17	4.05
Shea R^2				.24	.21	.54	.12	.08
Biv				.944	.75	.87	.85	.775
p(DWH)				.045	.0067	.0000	.0902	.1571
p(WH)				.044	.0067	.0000	.0886	.1519

For future reference, estimator legend and notation are summarized below:

- n = sample size
- ρ_1 controls endogeneity

- ρ_2 controls strength of IV #1
- ρ_3 controls strength of IV #2
- ρ_4 correlation between instruments
- α = nominal level of significance
- **ho1** = H_1 contrast with K_1 = 1 df
- **ho1s** = H_1 contrast with K = 2 df [SAS default and STATA default with `CONSTANT` option]
- **ho2** = H_2 contrast with IV variance estimator [`sigmaless`]
- **ho3** = H_3 contrast with OLS variance estimator [`sigmamore`]
- **ho3a** = H_{3a} contrast with MLE variance estimator [`Wu-Hausman chi-square`]
- **t** = *t* test of residual coefficient in auxiliary regression [`DWH F-test`]
- **hc** = H_1 contrast, GMM and OLS, with HC_1[4] correction, with K = 2 df
- **hc1** = **hc** test using K_1 =1 df
- **tr** = robust *t* test in auxiliary regression; HC_1 correction
- **tr2** = robust *t* test in auxiliary regression; HC_2 correction
- **tr3** = robust *t* test in auxiliary regression; HC_3 correction
- **w2** = Creel system test using 2nd round estimator K = 2 df
- **w21** = Creel test **w2** with K_1 = 1 df
- **d** = Cameron & Trivedi (2005, p. 245) *D* with K = 2 df
- **d1** = **d** test with K_1 = 1 df
- **w2r** = Creel test **w2** with covariance evaluated at estimates restricted by null, K = 2 df
- **w2r1** = **w2r** with K_1 = 1 df
- **DMF** = Davidson & MacKinnon type *F* contrast test comparing GMM to OLS, K = 2 df

[4] See Davidson & MacKinnon, 1993, p. 554, for discussion of the HC modifications.

3.6 Discussion of Asymptotic Results under Heteroskedasticity

In Figure 3.1 we present the performance of selected tests under heteroskedasticity. The relative performance of tests is independent of the instrument strength. The contrast test **ho3a** and the **t** test valid under homoskedasticity have the highest power also under heteroskedasticity.

The Davidson and MacKinnon style contrast test between the OLS estimator and the GMM estimator given in (3.16) has rejection frequencies that are even lower than the Creel (1996) type tests. In the following, we take a closer look at the effect of heteroskedasticity on the tests and the estimators.

3.6.1 The Effect of Heteroskedasticity

Table 3.2 contains some results showing the effects of heteroskedasticity on test performance. In the upper panel we give the percent rejections in the homoskedasticity case for easy reference. The parameter in control of the heteroskedasticity is γ, with the heteroskedasticity structure shown in (3.19)

$$u = (1 + \gamma x)v$$

If $\gamma = 0$ the errors are homoskedastic. We consider two degrees of heteroskedasticity, with $\gamma = .5$ and $\gamma = 1$. How strong is the heteroskedasticity? The indication from the single sample results reported in Table 3.1, is that the Breusch-Pagan statistic [$\chi^2_{(1)}$ under the null of no heteroskedasticity] is 15.9 in the former case and 24.66 in the latter. The middle and lower panels show the effects of heteroskedasticity for the cases with n = 100 and one strong instrument. When heteroskedasticity is present we continue to examine three tests that are valid for homoskedastic errors, **ho3**, **ho3a** and **t**. Our curiosity here is how heteroskedasticity affects these standard tests, which are the best in the homoskedasticity case. Going from the top to the bottom panel in Table 3.2 we see that for the case in which x is endogenous, $\rho_1 > 0$, the standard tests reject less frequently, suggesting lower power when errors are heteroskedastic. The effect of heteroskedasticity strength on these tests is also illustrated in Figure 3.2. Of these three **ho3a**, the contrast test based on the maximum likelihood estimate of the error variance, seems to fare the best, but this is a function of the too low size of **ho3** and **t** in this case, and differences disappear in the size corrected results [Table 3.3b]. The other tests in Table 3.2 are designed to cope with heteroskedasticity in one way or another. First, there are the robust t-tests **tr**, **tr2** and **tr3**. These tests are justified by the consideration

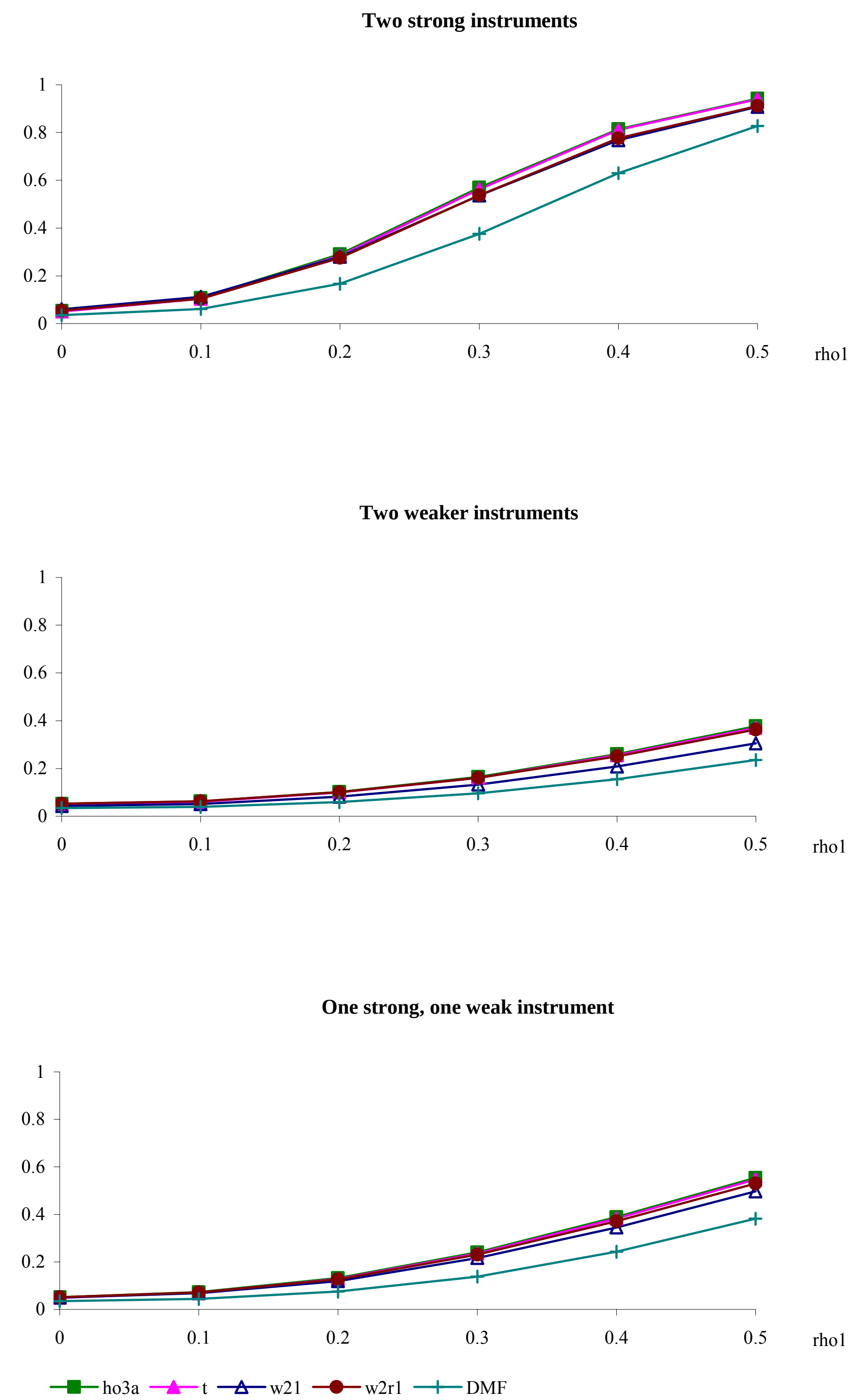


FIGURE 3.1. PERCENT REJECTIONS UNDER HETEROSKEDASTICITY (gamma = 1, n = 100)

of the artificial regression (3.9), and the robust Wald [or equivalent *F*-test, or *t* when there is a single potentially endogenous regressor] statistic (3.11). The simplest robust *t*-statistic rejects the true null three *t* tests, but this effect disappears after size correction [Table 3.3b]. It is interesting to note that the usual *t*-test has size corrected power that is even slightly higher than the three "robust" *t*-tests.

The results of two of Creel's test statistics are also considered. First, **w21** is the system test (3.5) that is based on a second round improvement of the GMM estimator (3.3) and with usual covariance matrix (3.4), but which sets the degrees of freedom of the chi-square test to one instead of the nominal number of restrictions being tested, which is two. With this modification the test rejects near 5% of the time when heteroskedasticity is stronger, only about 4.2% of the time in the medium heteroskedasticity case. The second of Creel's tests that we consider, **w2r1**, replaces the usual GMM covariance estimator with a restricted version (3.7), based the restricted GMM estimator (3.6), but again it presumes one degree of freedom rather than the nominally correct two degrees of freedom. This test represents a substantial improvement over **w21**. It has close to the correct size in both heteroskedasticity cases and greater power at all levels of regressor endogeneity. Clearly **w2r1** is preferred to **w21**. However, after size correction, the usual and robust *t*-statistic based tests perform better over all levels of endogeneity beyond the weakest [that is, for $\rho_1 \geq .2$] in both the medium and strong heteroskedasticity cases.

The test statistic **d1**, shown in (3.8), based on the difference of the two GMM objective functions is again compared to the critical value from a chi-square with one degree of freedom rather than the nominally correct two. The percentage rejections of **d1** and **w21** are identical and upon reflection one can see why the tests should be asymptotically equivalent. They are both testing the null hypothesis $H_0 : plim\left(\hat{\beta}_{OLS} - \hat{\beta}_{GMM}\right) = 0$. The test **w21** is a Wald test of the equality of the estimates obtained by jointly considering two sets of moment conditions, and the **d1** test compares the values of the GMM objective function obtained from restricted and unrestricted estimation. Neither of these tests are computationally

TABLE 3.2. PERCENT REJECTIONS - EFFECT OF HETEROSKEDASTICITY STRENGTH

Table 3: Percent rejections, n=100, gamma=0, rho2=.5, rho3=.1

alpha=0.05

rho1	ho1	ho1s	ho2	ho3	ho3a	t
0.0	0.03103	0.00462	0.03492	0.04868	0.05100	0.04905
0.1	0.06015	0.01110	0.06678	0.08685	0.09035	0.08750
0.2	0.16303	0.04380	0.17625	0.21255	0.21893	0.21385
0.3	0.37157	0.14758	0.39082	0.44390	0.45280	0.44572
0.4	0.65575	0.36843	0.67393	0.72377	0.73070	0.72530
0.5	0.90073	0.70530	0.90950	0.93037	0.93302	0.93107

Table 9: Percent rejections, n=100, gamma=.5, rho2=.5, rho3=.1

alpha=0.05

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.04743	0.05005	0.04793	0.05345	0.05115	0.04540	0.04245	0.04245	0.04990	0.03782
0.1	0.07788	0.08067	0.07848	0.08595	0.08285	0.07545	0.06755	0.06755	0.07995	0.05248
0.2	0.17815	0.18380	0.17938	0.19165	0.18638	0.17258	0.15118	0.15118	0.17865	0.11107
0.3	0.35985	0.36717	0.36157	0.37728	0.36838	0.34823	0.30815	0.30815	0.35490	0.23197
0.4	0.58835	0.59645	0.59030	0.60413	0.59467	0.57263	0.52065	0.52065	0.57622	0.42625
0.5	0.82073	0.82610	0.82215	0.82778	0.82070	0.80660	0.75053	0.75053	0.80288	0.68075

Table 15: Percent rejections, n=100, gamma=1, rho2=.5, rho3=.1

alpha=0.05

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.04830	0.05072	0.04877	0.05335	0.04970	0.04312	0.04983	0.04983	0.05100	0.03470
0.1	0.06938	0.07197	0.07007	0.07318	0.06928	0.06155	0.06772	0.06772	0.07095	0.04423
0.2	0.12780	0.13242	0.12880	0.13578	0.12970	0.11770	0.11837	0.11837	0.12610	0.07470
0.3	0.23453	0.24070	0.23585	0.24452	0.23535	0.21728	0.21600	0.21600	0.23113	0.13845
0.4	0.38100	0.38867	0.38245	0.39163	0.37967	0.35497	0.34485	0.34485	0.37078	0.24213
0.5	0.54563	0.55395	0.54755	0.55358	0.54048	0.51585	0.49715	0.49715	0.53017	0.38115

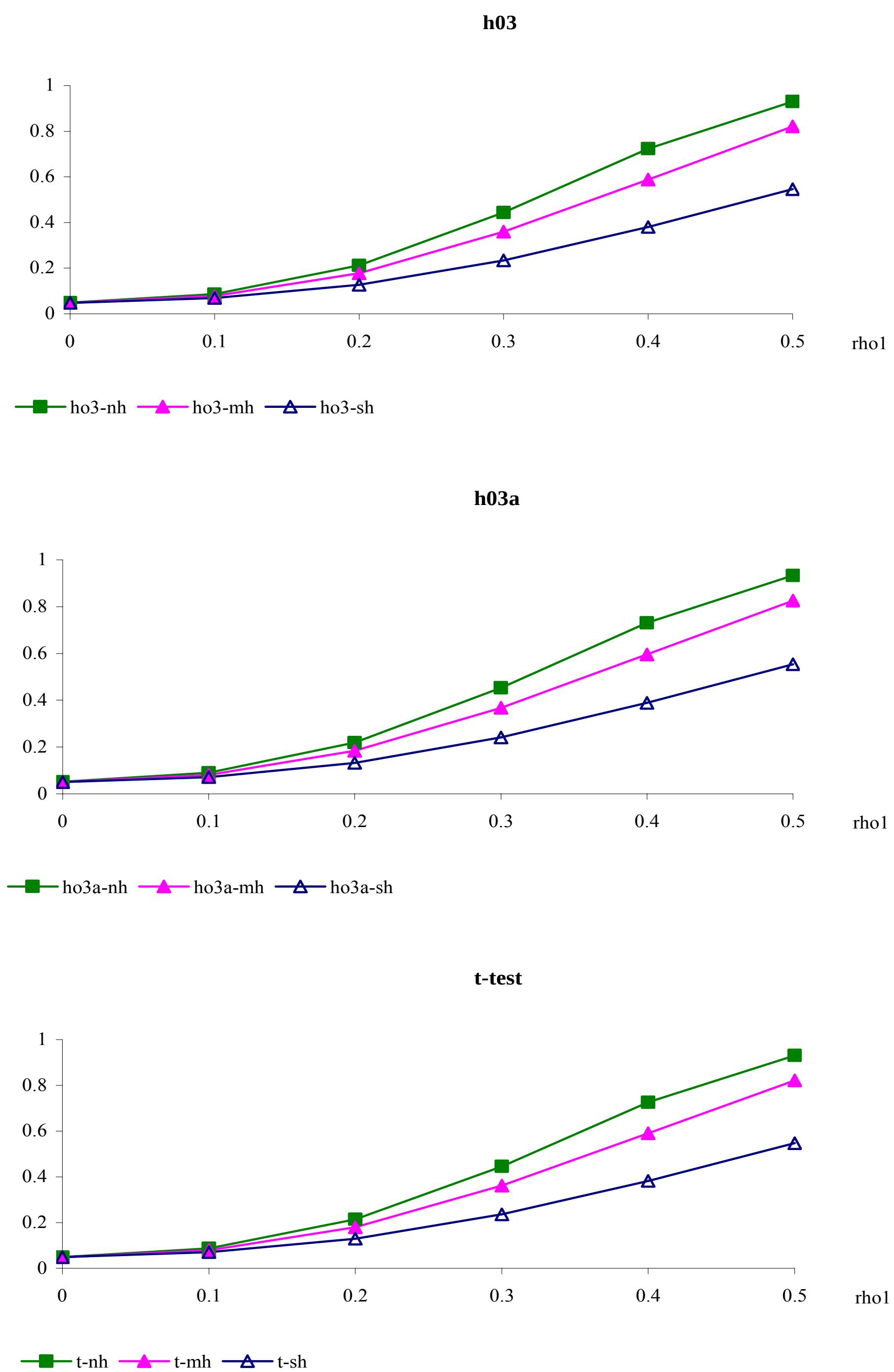


FIGURE 3.2. PERCENT REJECTIONS - EFFECT OF HETEROSKEDASTICITY STRENGTH (n = 100)

*nh – no heteroskedasticity (gamma = 0), mh – medium heteroskedasticity (gamma = .5), sh – strong heteroskedasticity (gamma = 1)

TABLE 3.3. HETEROSKEDASTIC CASE: VARIATIONS

Table 15: Percent rejections, n=100, gamma=1, rho2=.5, rho3=.1

alpha=0.05

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.04830	0.05072	0.04877	0.05335	0.04970	0.04312	0.04983	0.04983	0.05100	0.03470
0.1	0.06938	0.07197	0.07007	0.07318	0.06928	0.06155	0.06772	0.06772	0.07095	0.04423
0.2	0.12780	0.13242	0.12880	0.13578	0.12970	0.11770	0.11837	0.11837	0.12610	0.07470
0.3	0.23453	0.24070	0.23585	0.24452	0.23535	0.21728	0.21600	0.21600	0.23113	0.13845
0.4	0.38100	0.38867	0.38245	0.39163	0.37967	0.35497	0.34485	0.34485	0.37078	0.24213
0.5	0.54563	0.55395	0.54755	0.55358	0.54048	0.51585	0.49715	0.49715	0.53017	0.38115

Table 3.3a. Larger Sample Size

Table 18: Percent rejections, n=200, gamma=1, rho2=.5, rho3=.1

alpha=0.05

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.05140	0.05253	0.05162	0.05305	0.05142	0.04797	0.05248	0.05248	0.05175	0.04192
0.1	0.08910	0.09123	0.08943	0.09195	0.09002	0.08475	0.08708	0.08708	0.08940	0.06605
0.2	0.21395	0.21660	0.21448	0.21873	0.21468	0.20620	0.20650	0.20650	0.21238	0.14670
0.3	0.42543	0.42920	0.42605	0.43035	0.42480	0.41240	0.40742	0.40742	0.41920	0.30470
0.4	0.65470	0.65853	0.65535	0.65870	0.65255	0.64153	0.62840	0.62840	0.64507	0.51570
0.5	0.83890	0.84113	0.83925	0.83455	0.83013	0.82155	0.81288	0.81288	0.82475	0.72575

Table 3.3b. Size Corrected Power

Table 33: Size Corrected Power, n=100, gamma=1, rho2=.5, rho3=.1

alpha=0.05

rho1	ho3	ho3a	t	tr	tr2	tr3	w2	d	w2r	DMF
0.0	0.05003	0.05003	0.05003	0.05003	0.05003	0.05003	0.05003	0.05003	0.05003	0.05003
0.1	0.07155	0.07155	0.07155	0.06982	0.06955	0.06970	0.06797	0.06797	0.06990	0.06140
0.2	0.13160	0.13160	0.13160	0.13055	0.13012	0.13008	0.11868	0.11868	0.12445	0.10028
0.3	0.23943	0.23943	0.23943	0.23670	0.23588	0.23585	0.21638	0.21638	0.22898	0.17968
0.4	0.38685	0.38685	0.38685	0.38240	0.38033	0.37875	0.34547	0.34547	0.36773	0.29538
0.5	0.55173	0.55173	0.55173	0.54455	0.54125	0.53815	0.49800	0.49800	0.52670	0.44377

routine, to our knowledge, in econometrics software, thus their dominance by usual and robust *t*-tests makes their use unlikely without further evidence, at least in the regression context[5].

Finally we consider the Davidson and MacKinnon style contrast test between the OLS estimator and the GMM estimator given in (3.15). The Wald test statistic is given in (3.16) and we report the *F*-test version, **DMF**, which is W_{DM} divided by number of independent columns of $M_X AX$, which in this case is two. The size of this test is too low, and its size corrected power is lower than the other tests, thus unfortunately there appears little to recommend it in practice. Its performance does improve in larger samples, as shown in Table 3.3a.

3.6.2 Estimator Bias Results

Recall that the true value of $\beta_2 = 1$. In Table 3.4 and Figure 3.3 we compare alternative degrees of heteroskedasticity. Increasing heteroskedasticity slightly increases the bias of the GMM estimator, worsens the performance of the pre-test estimators, but with relative performance about the same. However, the robust *t*-test may produce a pre-test estimator with slightly less bias in the heteroskedastic cases.

3.6.3 Estimator RMSE Comparisons

In Table 3.5 and Figure 3.4 we compare the root-mse's of the alternative estimators with increasing levels of heteroskedasticity. The RMSE of the OLS estimator exceeds that of the IV and GMM estimators when the degree of endogeneity $\rho_1 \geq .2$ (when $\gamma = 0$ or $\gamma = .5$) and if $\rho_1 \geq .3$ when $\gamma = 1$. When the degree of endogeneity is low the IV estimator is a better choice than GMM. The pretest estimator is never the best choice. However, SAS's test leads to a better pre-test estimator performance when the degree of endogeneity is low ($\rho_1 \leq .2$) and performance that is at least comparable to the other pre-test estimators when $\rho_1 = .3$. Thus also under heteroskeadasticity the SAS/STATA default test is "best" for the intrepid researcher who does not know what to do, and whose regressor may be exogenous or slightly endogenous. Its advantage diminishes with an increase in sample size, as shown in Table 3.6a. In Table 3.6b and Figure 3.5 we see that weaker instruments also enhance the conservative approach, while stronger instruments reduce its appeal.

---

[5] Stata has a post-estimation command called `suest` that computes the "robust" covariance matrix of stacked estimators from potentially different models. It is currently implemented for ML based models that return a score vector. It is based on Wessie (1999), cited in Creel (2004).

TABLE 3.4. ESTIMATOR MEANS - EFFECT OF HETEROSKEDATICITY STRENGTH

Table 39: Estimator mean values, n=100, gamma=0, rho2=.5, rho3=.1

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	1.00041	1.00093	1.00077	1.00042	1.00099	1.00023	1.00051	1.00063	1.00087
0.1	1.09899	0.99930	0.99934	1.06730	1.03067	1.09370	1.06658	1.06507	1.02992
0.2	1.19934	1.00046	1.00033	1.10842	1.04266	1.17687	1.10676	1.10522	1.04149
0.3	1.29937	0.99782	0.99801	1.09958	1.02925	1.22154	1.09758	1.09666	1.02853
0.4	1.39954	0.99942	0.99950	1.05654	1.01206	1.20254	1.05561	1.05561	1.01201
0.5	1.49955	1.00002	0.99991	1.01491	1.00220	1.10693	1.01464	1.01522	1.00195

Table 45: Estimator mean values, n=100, gamma=.5, rho2=.5, rho3=.1

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.99949	0.99888	0.99917	1.00073	0.99926	0.99975	1.00044	1.00035	0.99934
0.1	1.09976	0.99992	0.99877	1.06907	1.03277	1.09465	1.06854	1.06765	1.03145
0.2	1.20004	1.00051	0.99824	1.11507	1.04744	1.18054	1.11377	1.11230	1.04517
0.3	1.29934	0.99833	0.99478	1.12064	1.04012	1.24011	1.11863	1.11644	1.03699
0.4	1.39962	1.00047	0.99579	1.09208	1.02380	1.25264	1.09017	1.08741	1.02008
0.5	1.49968	0.99738	0.99130	1.03991	1.00532	1.19023	1.03969	1.03505	0.99963

Table 51: Estimator mean values, n=100, gamma=1, rho2=.5, rho3=.1

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.99870	0.99788	0.99784	0.99957	0.99793	0.99899	0.99929	0.99919	0.99823
0.1	1.10180	1.00169	0.99901	1.07124	1.03470	1.09695	1.07081	1.07163	1.03334
0.2	1.19773	0.99863	0.99254	1.12338	1.05418	1.18254	1.12208	1.12223	1.05086
0.3	1.29824	0.99682	0.98928	1.15203	1.05719	1.25949	1.14941	1.15044	1.05261
0.4	1.39938	0.99960	0.98819	1.15214	1.05026	1.30989	1.14986	1.14909	1.04214
0.5	1.50026	0.99976	0.98630	1.12702	1.03483	1.32444	1.12478	1.12236	1.02500

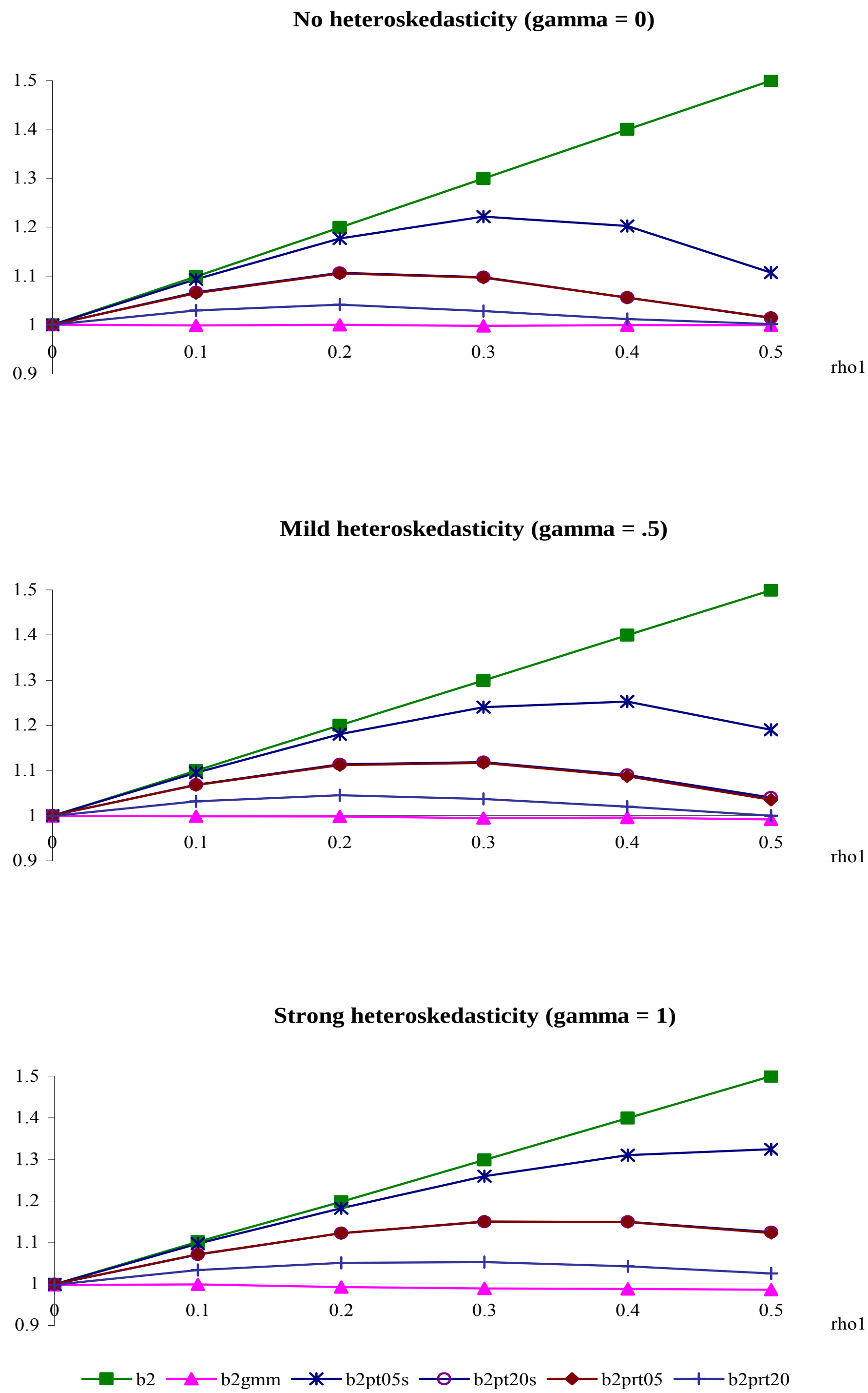


FIGURE 3.3. ESTIMATOR MEANS - EFFECT OF HETEROSKEDATICITY STRENGTH (n = 100, rho2 = .5, rho3 = .1)

*pt, prt – pre-test estimators defined in Section 3.4

TABLE 3.5. ESTIMATOR RMSE: EFFECT OF HETEROSKEDASTICITY

Table 57: Estimator rmse, n=100, gamma=0, rho2=.5, rho3=.1

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.10195	0.20466	0.20627	0.13638	0.17423	0.10693	0.13681	0.13852	0.17570
0.1	0.14170	0.20623	0.20783	0.17374	0.19500	0.14826	0.17346	0.17549	0.19575
0.2	0.22281	0.20674	0.20837	0.23770	0.22500	0.23085	0.23662	0.23750	0.22527
0.3	0.31489	0.20749	0.20940	0.28020	0.23746	0.31218	0.27763	0.27871	0.23770
0.4	0.41019	0.20669	0.20828	0.27285	0.22542	0.35492	0.27048	0.27239	0.22644
0.5	0.50717	0.20764	0.20962	0.23353	0.21210	0.32322	0.23233	0.23560	0.21373

Table 63: Estimator rmse, n=100, gamma=.5, rho2=.5, rho3=.1

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.13285	0.23995	0.24099	0.16787	0.20738	0.13759	0.16777	0.16870	0.20793
0.1	0.16705	0.23879	0.24011	0.19877	0.22289	0.17339	0.19829	0.19965	0.22330
0.2	0.24061	0.24043	0.24121	0.26238	0.25497	0.24959	0.26087	0.26110	0.25477
0.3	0.32837	0.24107	0.24163	0.31228	0.27386	0.33162	0.31018	0.31020	0.27344
0.4	0.42231	0.24495	0.24518	0.32933	0.27324	0.39614	0.32654	0.32706	0.27364
0.5	0.51874	0.24767	0.24838	0.30438	0.26063	0.40986	0.30252	0.30409	0.26143

Table 69: Estimator rmse, n=100, gamma=1, rho2=.5, rho3=.1

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.19988	0.31894	0.31961	0.23764	0.28262	0.20493	0.23757	0.23677	0.28202
0.1	0.22443	0.31964	0.32038	0.26306	0.29539	0.23176	0.26277	0.26106	0.29483
0.2	0.28438	0.32293	0.32377	0.31841	0.32610	0.29433	0.31718	0.31690	0.32593
0.3	0.36639	0.32495	0.32548	0.37828	0.35368	0.37615	0.37661	0.37669	0.35349
0.4	0.45627	0.33132	0.33025	0.42240	0.37106	0.45670	0.41999	0.42081	0.37039
0.5	0.55225	0.33814	0.33807	0.44205	0.37471	0.52095	0.43900	0.44179	0.37527

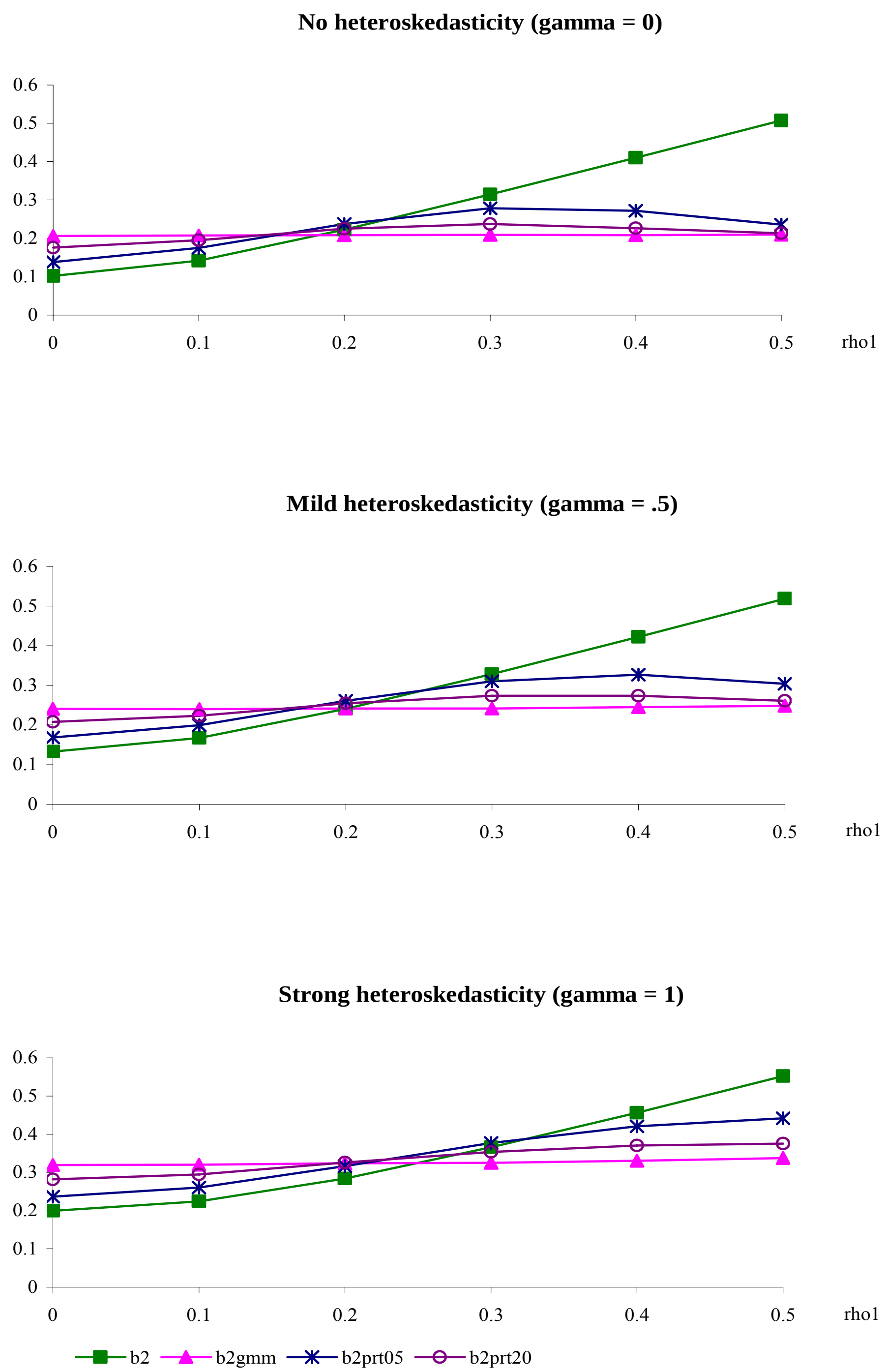


FIGURE 3.4. ESTIMATOR RMSE - EFFECT OF HETEROSKEDATICITY STRENGTH (n = 100, rho2 = .5, rho3 = .1)

*prt – pre-test estimators defined in Section 3.4

TABLE 3.6. ESTIMATOR RMSE UNDER HETEROSKEDASTICITY- VARIATIONS (gamma = 1)

Table 69: Estimator rmse, n=100, gamma=1, rho2=.5, rho3=.1

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.19988	0.31894	0.31961	0.23764	0.28262	0.20493	0.23757	0.23677	0.28202
0.1	0.22443	0.31964	0.32038	0.26306	0.29539	0.23176	0.26277	0.26106	0.29483
0.2	0.28438	0.32293	0.32377	0.31841	0.32610	0.29433	0.31718	0.31690	0.32593
0.3	0.36639	0.32495	0.32548	0.37828	0.35368	0.37615	0.37661	0.37669	0.35349
0.4	0.45627	0.33132	0.33025	0.42240	0.37106	0.45670	0.41999	0.42081	0.37039
0.5	0.55225	0.33814	0.33807	0.44205	0.37471	0.52095	0.43900	0.44179	0.37527

Table 3.6a. Sample Size

Table 72: Estimator rmse, n=200, gamma=1, rho2=.5, rho3=.1

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.14122	0.22264	0.22296	0.16823	0.19815	0.14889	0.17221	0.16759	0.19789
0.1	0.17392	0.22278	0.22297	0.19852	0.21404	0.18355	0.20115	0.19801	0.21385
0.2	0.24668	0.22380	0.22383	0.25655	0.24172	0.25672	0.25489	0.25564	0.24162
0.3	0.33520	0.22638	0.22646	0.29632	0.25565	0.32687	0.29022	0.29645	0.25598
0.4	0.42959	0.22984	0.22967	0.30428	0.25248	0.36641	0.29478	0.30446	0.25327
0.5	0.52698	0.23382	0.23313	0.28547	0.24515	0.36013	0.27621	0.28774	0.24574

Table 3.6b. Instrument strength

Table 68: Estimator rmse, n=100, gamma=1, rho2=.3, rho3=.3

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.19969	0.37888	0.37996	0.26153	0.32581	0.20380	0.25067	0.26039	0.32463
0.1	0.22381	0.37928	0.38018	0.28181	0.33828	0.22930	0.27074	0.27977	0.33777
0.2	0.28846	0.38429	0.38493	0.34191	0.37090	0.29451	0.33260	0.34059	0.36928
0.3	0.36667	0.39096	0.39045	0.40816	0.40893	0.37375	0.39834	0.40643	0.40708
0.4	0.45573	0.39530	0.39578	0.47223	0.43631	0.46185	0.46345	0.46843	0.43553
0.5	0.55341	0.40450	0.40422	0.51930	0.45511	0.55136	0.51578	0.51639	0.45353

Table 67: Estimator rmse, n=100, gamma=1, rho2=.5, rho3=.5

rho1	b2	b2iv	b2gmm	b2pt05	b2pt20	b2pt05s	b2pt20s	b2prt05	b2prt20
0.0	0.19994	0.24639	0.24691	0.21393	0.23112	0.20312	0.21617	0.21346	0.23138
0.1	0.22523	0.24642	0.24624	0.23805	0.24462	0.23134	0.23938	0.23814	0.24440
0.2	0.28697	0.24770	0.24750	0.28264	0.26457	0.29108	0.28049	0.28185	0.26407
0.3	0.36734	0.25071	0.25071	0.30345	0.26804	0.34186	0.29738	0.30334	0.26883
0.4	0.45853	0.25316	0.25348	0.29009	0.26155	0.34504	0.28394	0.29282	0.26241
0.5	0.54873	0.25563	0.25538	0.27050	0.25814	0.30869	0.26739	0.27536	0.25902

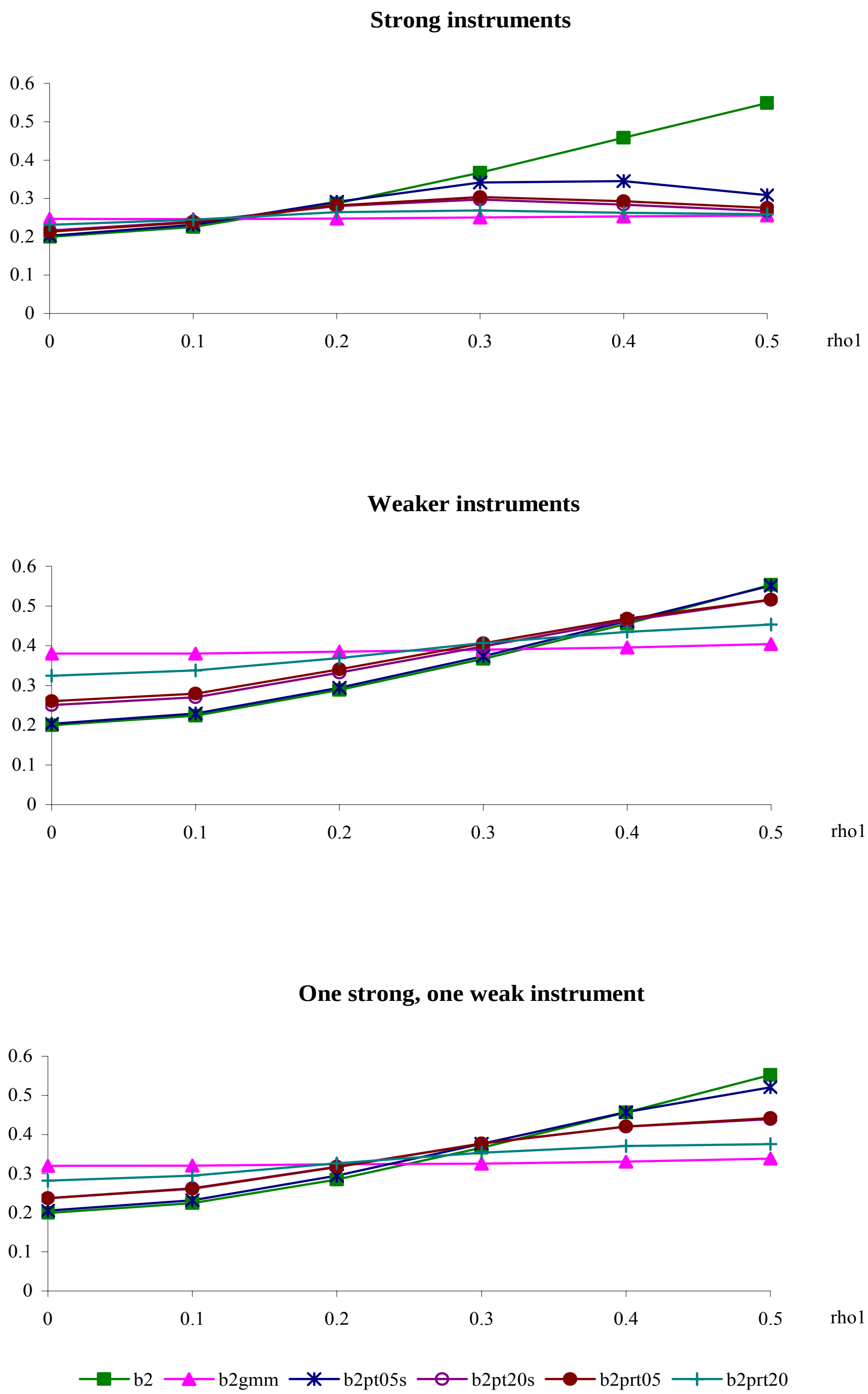


FIGURE 3.5. ESTIMATOR RMSE UNDER HETEROSKEDASTICITY (gamma = 1, n = 100)
*pt, prt – pre-test estimators defined in Section 3.4

3.7 Summary of Findings under Heteroskedasticity

In the case of heteroskedastic errors the simple *t*-test (ignoring heteroskedasticity) and the robust *t*-statistic **`tr2`** (HC_2 correction) seem to have reliable size, with the robust *t*-tests **`tr`** and **`tr3`** improving with larger samples. The statistic **`tr2`** had size closer to the correct value in the smaller sample. After size correction the differences between the alternative *t*-tests were largely eliminated. In many of the scenarios the simple *t*-test that ignores the heteroskedasticity seems to have a slight edge in size-corrected power. Based on our results there is no strong evidence that any of the other tests is superior.

Among the more specialized tests Creel's test **`w2r1`** with the covariance matrix evaluated at estimates restricted by null when degrees of freedom are specified as one rather than two performed the best. However, neither of the tests based on GMM was as powerful as the *t*-tests in our experiments.

Wong (1996) shows that bootstrapping is a good alternative to usual asymptotic theory in the homoskedastic case. In the next chapter, we extend bootstrapping to the heteroskedastic case.

4. BOOTSTRAPPING UNDER HETEROSKEDASTICITY

4.1 Bootstrap Tests under Heteroskedasticity

If we do not make the assumptions of classical normal model, the Hausman test statistic has a finite sample distribution that is different from the one predicted by the asymptotic theory. Bootstrapping allows to make inferences without making strong distributional assumptions and without the need for analytic formulas for the sampling distribution parameters. Thus, it could also help in the presence of heteroskedasticity. We are interested in whether bootstrapping improves the performance of the Hausman-type tests based on the Creel's system in GMM framework and also the Wald test based on the artificial regression since the evidence especially under weak instruments is mixed. We compare the impact of different types of bootstrapping on all the Hausman test alternatives typically used under heteroskedasticity and try to suggest the best for testing endogeneity.

As already mentioned in Chapter 2, the data generating process of bootstrap samples should be as close as possible to the data generating process that generated the observed data. With heteroskedasticity of unknown form, this might be a problem. A basic requirement for bootstrapping to be valid is that resampling needs to be done on independently and identically distributed variables. Thus for heteroskedastic data, bootstrapping the residuals is not valid. Freedman (1981) proposed pairs bootstrap that resamples the regressand and regressors together from the original data: a pairs bootstrap sample (y^*, X^*, Z^*) is obtained by drawing rows with replacement from (y, X, Z). However, randomly resampling pairs does not impose the restriction of exogeneity of regressors. Mammen (1993) and Flachaire (1999) suggested an improved version of the pairs bootstrap where the resampling scheme respects the null hypothesis.

Another technique that solves the problem of unknown heteroskedasticity is a wild bootstrap technique proposed by Liu (1988). The wild bootstrap generating process is

(4.1) $$y_t^* = X_t\hat{\beta} + a_t\hat{u}_t\varepsilon_t^*$$

where $\hat{u}_t$ is the OLS residual, ε_t^* is white noise with a zero mean and variance equal to 1. a_t is the diagonal element correction of the heteroskedasticity consistent covariance matrix estimator (we use HC_2

correction) as suggested by MacKinnon and White (1985). For the model with one regressor, Liu showed that the first three moments of the bootstrap distribution of a heteroskedasticity consistent covariance matrix estimator based statistic mapped those of the true distribution of the statistic up to order n^{-1}. Mammen (1993) provided further evidence that the wild bootstrap is asymptotically justified. He proposed that the white noise follows the two-point distribution

$$(4.2) \qquad PD_1: \ \varepsilon_t^* = \begin{cases} -(\sqrt{5}-1)/2 \doteq -0.618 & \text{with probability } p = (\sqrt{5}+1)/(2\sqrt{5}) \doteq 0.7236 \\ (\sqrt{5}+1)/2 \doteq 1.618 & \text{with probability } 1-p \doteq 0.2764 \end{cases}$$

so that $E(\varepsilon_t^*) = 0$, $E(\varepsilon_t^{*2}) = 1$, and $E(\varepsilon_t^{*3}) = 1$. Horowitz (1997, 2001) shows that this wild bootstrap performs much better than paired bootstrap in the case of heteroskedasticity and works well even if there is no heteroskedasticity. Flachaire (2005) also examined performances of heteroskedasticity-robust tests. He showed that the wild bootstrap proposed in Davidson and Flachaire (2001) gives better results than pairs bootstrap. In particular, white noise used in the data generating process following the Rademacher distribution always gives better results than the two-point distribution proposed by Mammen (1993), used in Horowitz (1997) and recommended in the literature.

The white noise is modeled by the Rademacher distribution where

$$(4.3) \qquad PD_2: \ \varepsilon_t^* = \begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5 \end{cases}$$

so that $E(\varepsilon_t^*) = 0$, $E(\varepsilon_t^{*2}) = 1$, and $E(\varepsilon_t^{*3}) = 0$.

Thus, if we draw B pairs or wild bootstrap samples of size n with replacement yielding the bootstrap estimates $\hat{\beta}^*$ and $\tilde{\beta}^*$, the variance of the contrast of the two estimators can be estimated consistently by [Cameron and Trivedi (2005), p. 378]:

$$(4.4) \qquad \hat{V}_{Boot}\left(\hat{\beta}-\tilde{\beta}\right) = \frac{1}{B-1}\sum_{b=1}^{B}[(\hat{\beta}_b^* - \tilde{\beta}_b^*) - (\bar{\hat{\beta}}_b^* - \bar{\tilde{\beta}}_b^*)][(\hat{\beta}_b^* - \tilde{\beta}_b^*) - (\bar{\hat{\beta}}_b^* - \bar{\tilde{\beta}}_b^*)]'$$

where $\bar{\hat{\beta}}_b^* = B^{-1}\sum_{b=1}^{B}\hat{\beta}_b^*$ and $\bar{\tilde{\beta}}_b^* = B^{-1}\sum_{b=1}^{B}\tilde{\beta}_b^*$. The Hausman test statistic can then be computed as

$$(4.5) \qquad H_B = \left(\hat{\beta}-\tilde{\beta}\right)' \left(\hat{V}_{Boot}\left(\hat{\beta}-\tilde{\beta}\right)\right)^{-1} \left(\hat{\beta}-\tilde{\beta}\right)$$

Since this approach can be used for any chi-square distributed test, we analogously compute the bootstrap Wald test based on the auxiliary regression $y = X\beta + \hat{V}_1\eta + error$ as

(4.6) $$W_B = \hat{\eta}' \left(\hat{V}_{Boot}\left(\hat{\eta}\right)\right)^{-1} \hat{\eta}$$

where $\hat{\eta}$ is the OLS estimate of η and the bootstrap Wald test based on the Creel's (2004) system approach as

(4.7) $$W_{B2} = \left(R\hat{\beta}_{GMM2}\right)' \left[R\hat{V}_{Boot}\left(\hat{\beta}_{GMM2}\right)R'\right]^{-1} \left(R\hat{\beta}_{GMM2}\right)$$

or

(4.8) $$W_{B2a} = \left(R_1\hat{\beta}_{GMM2}\right)' \left[R_1\hat{V}_{Boot}\left(\hat{\beta}_{GMM2}\right)R_1'\right]^{-1} \left(R_1\hat{\beta}_{GMM2}\right)$$

where the two-step GMM estimator $\hat{\beta}_{GMM2}$ is defined as in (3.2), $R = \left[I_K \quad -I_K\right]$ and $R_1 = \left[I_{K_1} \quad 0 \quad -I_{K_1} \quad 0\right]$. We also compute the bootstrap Wald test based on the system and comparing the OLS and IV and the OLS and first-step GMM estimators. The covariance matrix for each estimator is computed analogously to (4.4). Furthermore, we examine the performance of the bootstrap robust *t* test based on the auxiliary regression. Size corrected powers of the bootstrapped tests are obtained by the percentile method described in Chapter 2. The critical values we use are the respective quantiles of the empirical distribution function of the bootstrapped test statistic.

4.2 A Monte Carlo Experiment - Bootstrapping under Heteroskedasticity

We use the same set up as in the previous sections. We perform 1,000 Monte Carlo simulations and 1,000 bootstrap replications. We consider the regression

(4.9) $$y = \beta_1 + \beta_2 x + u = 0 + x + u$$

Data are generated by specifying

(4.10) $$\begin{bmatrix} x \\ v \\ z_1 \\ z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & 0 & 0 \\ \rho_2 & 0 & 1 & 0 \\ \rho_3 & 0 & 0 & 1 \end{bmatrix}\right)$$

The key features are that

- endogeneity is controlled by the parameter ρ_1 which takes the values 0, .1, .2, .3, .4 and .5
- the strength of the instruments is controlled by ρ_2 and ρ_3 which take the values .1, .3 and .5
- the two instruments are uncorrelated with each other
- samples of sizes $n = 100$ and $n = 200$ are considered.

The error u in (4.9) is made heteroskedastic by specifying

(4.11) $u = (1 + \gamma x)v$

- The parameter γ controls the degree of heteroskedasticity and takes the values 0 (homoskedasticity), .5 and 1.

We perform pairs and wild bootstrapping. The two-point distributions that are used in wild bootstrapping are the Rademacher distribution (4.3) as proposed in Davidson and Flachaire (2001) and the distribution (4.2) as in Mammen (1993). We compute the bootstrap Wald test based on the system and comparing the OLS and IV, the OLS and first-step GMM estimators and the OLS and second-step GMM estimators. We also analyze the performance of the bootstrap contrast tests, the bootstrap *t* test and the bootstrap robust *t* test based on the auxiliary regression.

We summarize the notation we use:

- n = sample size
- ρ_1 controls endogeneity
- ρ_2 controls strength of IV #1
- ρ_3 controls strength of IV #2
- α = nominal level of significance
- **wb** = Wald test of the residual coefficient in auxiliary regression $K = 1$ df
- **wb0** = Wald test based on OLS and IV estimators $K = 2$ df
- **wb0a** = Wald test based on OLS and IV estimators $K = 1$ df
- **wb1** = Wald test based on OLS and GMM estimators $K = 2$ df
- **wb1a** = Wald test based on OLS and GMM estimators $K = 1$ df
- **wb2** = Wald test based on OLS and GMM2 estimators $K = 2$ df
- **wb2a** = Wald test based on OLS and GMM2 estimators $K = 1$ df

4.3 Discussion of Results on Bootstrap Tests under Heteroskedasticity

Since the performance of the classical Hausman tests is, as expected, not satisfactory and also the errors in levels of the *t* and robust *t* tests are enormous, we only report results for the Wald test based on the artificial regression and the Wald type tests based on the Creel's system comparing the OLS estimators to the IV, GMM and second round improvement of the GMM estimators. We consider the statistics that set the degrees of freedom of the chi-square test to one and to the nominal number of restrictions being tested (two).

4.3.1 The Effect of Sample Size

In Table 4.1 we consider the effect of sample size. We report the results for wild bootstrapping using the PD_1 distribution as in Mammen (1993) and two weak instruments. The results for the PD_1 distribution illustrate that the larger the sample size, the worse the size distortions except the Wald-type test with one degree of freedom comparing the OLS and both GMM estimators but this conclusion depends on the instrument strength and a bootstrapping type. If two strong instruments are used, the larger the sample size the better the test size properties. With pairs bootstrapping, a larger sample size reduces size distortions of most of the tests. With wild bootstrapping, the test size improved with a larger sample size for tests with one degree of freedom comparing the OLS and both GMM estimators with the exception of the case with Rademacher distribution under weak instruments where size distortions were increased for most of the tests. This unexpected result may be caused by a nonstandard finite sample distribution of the instrumental variable estimator or a small number of Monte Carlo samples in our bootstrap experiments. However, the same pattern (the larger sample size the lower the test size, in particular of the t- test) for models with weak instruments can be also observed in our Chapter 2 asymptotic experiments where the number of Monte Carlo samples is much higher, which seems to confirm the former. The differences in rejection frequencies of the tests for a size of 100 and 200 observations are remarkable. For example, if the correlation between the endogenous regressor and error term $\rho_1 = .4$, the size corrected power of each of the tests increases by at least twice as much.

TABLE 4.1. BOOTSTRAPPING - EFFECT OF SAMPLE SIZE

Table 104: Percent rejections (wild bootstr. - Mammen), n=100, gamma=1, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.056	0.056	0.056	0.056	0.029	0.107	0.056	0.030	0.107	0.058
0.1	0.051	0.051	0.051	0.051	0.029	0.090	0.057	0.029	0.097	0.061
0.2	0.098	0.098	0.098	0.098	0.047	0.149	0.102	0.052	0.148	0.102
0.3	0.168	0.168	0.168	0.168	0.088	0.221	0.165	0.091	0.221	0.164
0.4	0.215	0.215	0.215	0.215	0.136	0.304	0.226	0.138	0.303	0.227
0.5	0.326	0.326	0.326	0.326	0.219	0.396	0.339	0.223	0.398	0.340

Table 107: Percent rejections (wild bootstr. - Mammen), n=200, gamma=1, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.061	0.061	0.061	0.061	0.020	0.081	0.065	0.020	0.084	0.063
0.1	0.063	0.063	0.063	0.063	0.031	0.102	0.064	0.035	0.104	0.065
0.2	0.151	0.151	0.151	0.151	0.084	0.202	0.152	0.088	0.205	0.155
0.3	0.261	0.261	0.261	0.261	0.157	0.318	0.269	0.159	0.317	0.270
0.4	0.447	0.447	0.447	0.447	0.314	0.516	0.460	0.313	0.514	0.462
0.5	0.628	0.628	0.628	0.628	0.470	0.681	0.635	0.474	0.681	0.640

Table 4.1a. Size Corrected Power

Table 158: Size Corrected Power (wild bootstr. - Mammen), n=100, gamma=1, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb0a	wb1	wb1a	wb2	wb2a
0.0	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.1	0.051	0.051	0.051	0.046	0.051	0.051	0.053
0.2	0.097	0.097	0.097	0.086	0.091	0.093	0.093
0.3	0.166	0.166	0.166	0.132	0.150	0.141	0.148
0.4	0.214	0.214	0.214	0.197	0.210	0.203	0.209
0.5	0.325	0.325	0.325	0.285	0.318	0.292	0.319

Table 161: Size Corrected Power (wild bootstr. - Mammen), n=200, gamma=1, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb0a	wb1	wb1a	wb2	wb2a
0.0	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.1	0.064	0.064	0.064	0.077	0.068	0.070	0.070
0.2	0.154	0.154	0.154	0.168	0.165	0.156	0.167
0.3	0.268	0.268	0.268	0.278	0.289	0.263	0.295
0.4	0.453	0.453	0.453	0.459	0.485	0.444	0.487
0.5	0.633	0.633	0.633	0.646	0.650	0.630	0.659

4.3.2 The Effect of Instrument Strength

In Table 4.2 and Figure 4.1 we report three wild bootstrapping (using the PD_1 distribution as in Mammen (1993)) cases with *n* = 200. For the case with two strong instruments, all the Wald tests reject exogeneity of the variable in question in more than 92% of the samples if a correlation between the endogenous regressor and error $\rho_1 = .4$. After size correction, the differences among the Wald tests are negligible, however the **w** test based on artificial regression and the tests **wb0**, **wb01** and **wb0a** based on Creel's system comparing the OLS and IV estimators perform slightly better. Together with the Wald-type tests comparing only the potentially endogenous elements of the GMM and the OLS estimators have the best size properties. If we have two weaker instruments, all of the tests perform worse at all degrees of endogeneity, size distortions are also greater. The effect of instrument strength on individual tests is illustrated in Figure 4.2. After size correction, the Wald test **w2a** (comparing the potentially endogenous elements of the two step GMM and the OLS estimators) and **w1a** (comparing the potentially endogenous elements of the GMM and the OLS estimators) perform the best. In the third panel we report the results for one strong and one weak instrument. The size properties are better than in the case with two strong or two weak instruments. The rejection frequencies are higher than in the case with two weak instruments, the relative performance of the tests is almost the same, with the Wald tests comparing the potentially endogenous elements of the GMM and the OLS estimators performing the best.

4.3.3 The Effect of Heteroskedasticity

Table 4.3 and Figures 4.3 show effects of heteroskedasticity on test performance if we have two weak instruments available and *n* = 200. In the upper panel we present the size corrected power in the homoskedasticity case, the middle and lower panels show the effects of heteroskedasticity for $\gamma = .5$ and $\gamma = 1$. For any type of bootstrapping, under weak endogeneity ($\rho_1 = .1$) the size corrected power of all bootstrap tests under mild heteroskedasticity ($\gamma = .5$) is higher than under homoskedasticity. When the correlation between the regressor and the error term is higher than .1, we obtain evidence that the more pronounced heteroskedasticity, the lower the power of all Wald tests.

TABLE 4.2. BOOTSTRAPPING - EFFECT OF INSTRUMENT STRENGTH

Table 106: Percent rejections (wild bootstr. - Mammen), n=200, gamma=1, rho2=.5, rho3=.5

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.042	0.042	0.042	0.042	0.014	0.069	0.045	0.015	0.070	0.047
0.1	0.166	0.166	0.166	0.166	0.090	0.206	0.154	0.086	0.207	0.155
0.2	0.497	0.497	0.497	0.497	0.337	0.526	0.463	0.332	0.524	0.454
0.3	0.843	0.843	0.843	0.843	0.692	0.850	0.809	0.692	0.852	0.808
0.4	0.977	0.977	0.977	0.977	0.931	0.978	0.969	0.928	0.978	0.969
0.5	0.997	0.997	0.997	0.997	0.992	0.996	0.994	0.991	0.996	0.992

Table 107: Percent rejections (wild bootstr. - Mammen), n=200, gamma=1, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.061	0.061	0.061	0.061	0.020	0.081	0.065	0.020	0.084	0.063
0.1	0.063	0.063	0.063	0.063	0.031	0.102	0.064	0.035	0.104	0.065
0.2	0.151	0.151	0.151	0.151	0.084	0.202	0.152	0.088	0.205	0.155
0.3	0.261	0.261	0.261	0.261	0.157	0.318	0.269	0.159	0.317	0.270
0.4	0.447	0.447	0.447	0.447	0.314	0.516	0.460	0.313	0.514	0.462
0.5	0.628	0.628	0.628	0.628	0.470	0.681	0.635	0.474	0.681	0.640

Table 108: Percent rejections (wild bootstr. - Mammen), n=200, gamma=1, rho2=.5, rho3=.1

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.046	0.046	0.046	0.046	0.023	0.066	0.045	0.022	0.067	0.045
0.1	0.084	0.084	0.084	0.084	0.037	0.109	0.081	0.037	0.109	0.080
0.2	0.213	0.213	0.213	0.213	0.118	0.257	0.209	0.124	0.258	0.210
0.3	0.406	0.406	0.406	0.406	0.269	0.475	0.416	0.271	0.474	0.414
0.4	0.649	0.649	0.649	0.649	0.473	0.699	0.653	0.473	0.695	0.651
0.5	0.840	0.840	0.840	0.840	0.723	0.870	0.844	0.724	0.869	0.840

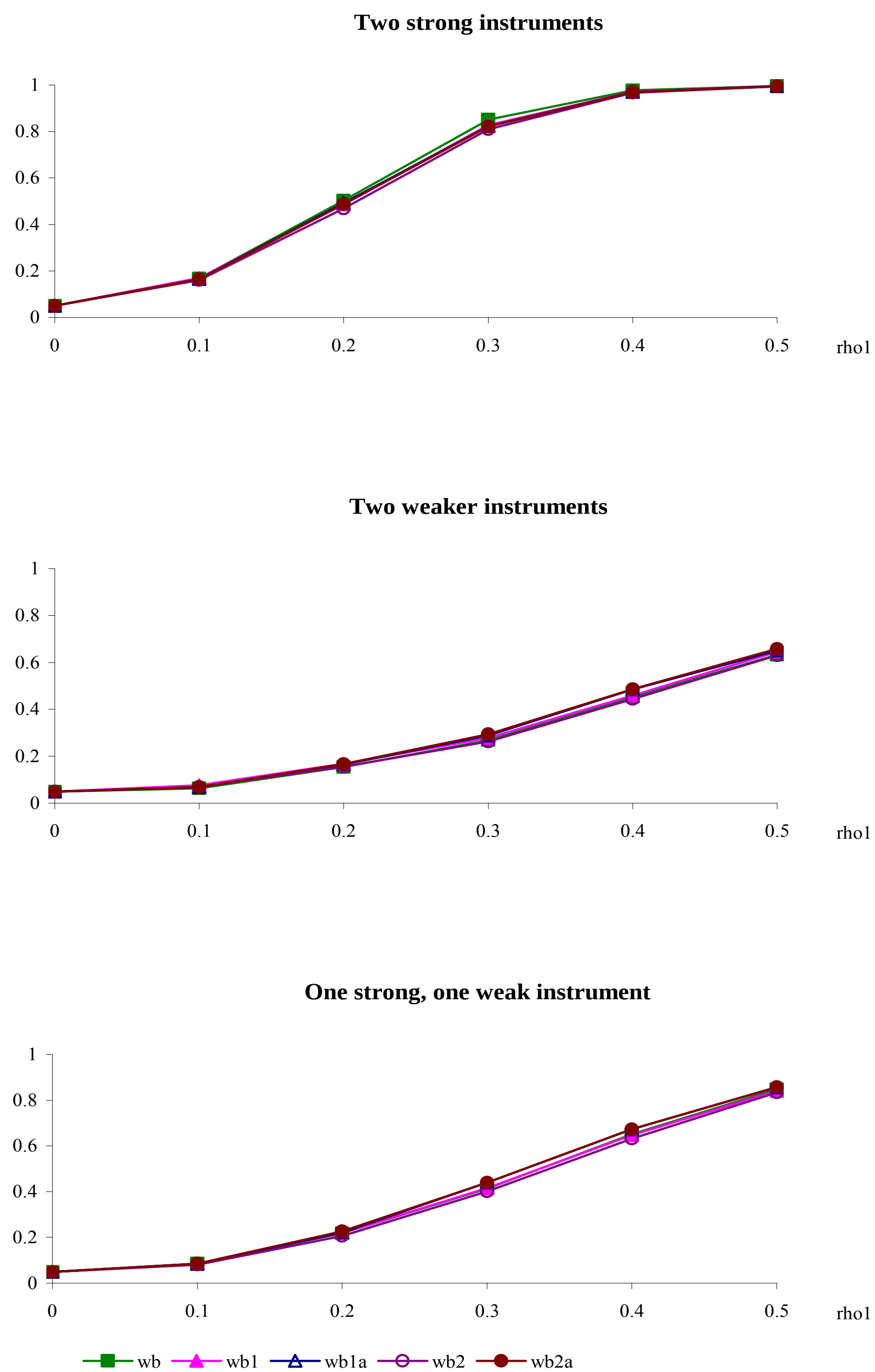


FIGURE 4.1. PERCENT REJECTIONS UNDER HETEROSKEDASTICITY (WILD BOOTSTRAPPING – MAMMEN, gamma = 1, n = 200)

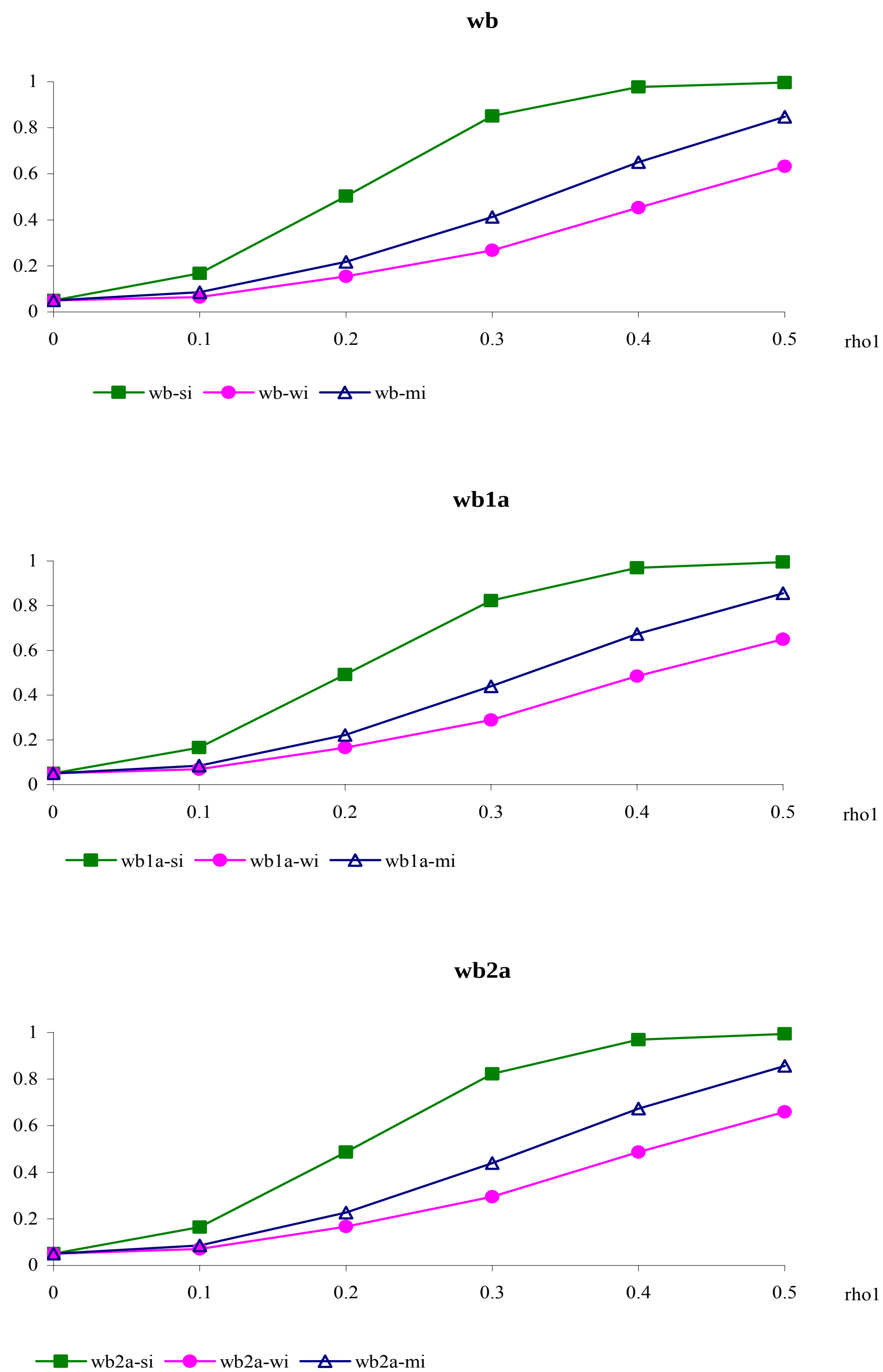


FIGURE 4.2. PERCENT REJECTIONS UNDER HETEROSKEDASTICITY (WILD BOOTSTRAPPING – MAMMEN) - EFFECT OF INSTRUMENT STRENGTH (gamma = 1, n = 200)

*si – two strong instruments, wi – two weaker instruments, mi - one strong, one weak instrument

TABLE 4.3. BOOTSTRAPPING - EFFECT OF HETEROSKEDASTICITY

Table 4.3.1. Pairs bootstrapping

Table 77: Percent rejections (pairs bootstrapping), n=200, gamma=0, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.042	0.004	0.036	0.036	0.004	0.041	0.033	0.004	0.039	0.034
0.1	0.071	0.016	0.056	0.055	0.016	0.056	0.054	0.016	0.055	0.052
0.2	0.230	0.061	0.192	0.188	0.058	0.202	0.192	0.056	0.201	0.190
0.3	0.450	0.181	0.401	0.397	0.183	0.404	0.389	0.183	0.403	0.387
0.4	0.774	0.448	0.723	0.717	0.447	0.734	0.716	0.449	0.733	0.717
0.5	0.960	0.775	0.940	0.937	0.766	0.938	0.932	0.762	0.937	0.932

Table 83: Percent rejections (pairs bootstrapping), n=200, gamma=.5, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.033	0.005	0.027	0.026	0.005	0.027	0.025	0.005	0.026	0.025
0.1	0.080	0.015	0.066	0.065	0.013	0.068	0.064	0.013	0.067	0.064
0.2	0.170	0.046	0.140	0.136	0.038	0.149	0.135	0.039	0.148	0.134
0.3	0.374	0.130	0.341	0.337	0.138	0.346	0.336	0.137	0.346	0.336
0.4	0.632	0.330	0.589	0.581	0.339	0.594	0.581	0.336	0.595	0.582
0.5	0.867	0.596	0.847	0.843	0.605	0.853	0.845	0.607	0.852	0.844

Table 89: Percent rejections (pairs bootstrapping), n=200, gamma=1, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.053	0.007	0.047	0.044	0.004	0.047	0.042	0.004	0.044	0.041
0.1	0.059	0.013	0.048	0.044	0.011	0.053	0.044	0.011	0.053	0.042
0.2	0.131	0.037	0.120	0.112	0.036	0.125	0.112	0.035	0.125	0.110
0.3	0.231	0.071	0.213	0.205	0.077	0.219	0.198	0.075	0.216	0.196
0.4	0.417	0.185	0.384	0.376	0.197	0.401	0.376	0.194	0.397	0.372
0.5	0.590	0.288	0.546	0.537	0.296	0.564	0.538	0.298	0.565	0.537

Table 4.3.2. Wild bootstrapping - PD_1 distribution

Table 95: Percent rejections (wild bootstr. - Mammen), n=200, gamma=0, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.060	0.060	0.060	0.060	0.020	0.079	0.061	0.020	0.081	0.061
0.1	0.088	0.088	0.088	0.088	0.037	0.116	0.095	0.038	0.115	0.095
0.2	0.268	0.268	0.268	0.268	0.150	0.308	0.273	0.150	0.306	0.273
0.3	0.513	0.513	0.513	0.513	0.329	0.554	0.512	0.329	0.553	0.512
0.4	0.824	0.824	0.824	0.824	0.669	0.845	0.826	0.669	0.845	0.827
0.5	0.969	0.969	0.969	0.969	0.910	0.979	0.966	0.911	0.978	0.966

Table 101: Percent rejections (wild bootstr. - Mammen), n=200, gamma=.5, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.041	0.041	0.041	0.041	0.015	0.061	0.046	0.015	0.061	0.046
0.1	0.094	0.094	0.094	0.094	0.048	0.130	0.095	0.049	0.129	0.095
0.2	0.201	0.201	0.201	0.201	0.104	0.238	0.208	0.103	0.238	0.207
0.3	0.422	0.422	0.422	0.422	0.256	0.473	0.426	0.258	0.469	0.425
0.4	0.679	0.679	0.679	0.679	0.516	0.706	0.679	0.515	0.709	0.678
0.5	0.896	0.896	0.896	0.896	0.809	0.919	0.902	0.808	0.918	0.900

Table 107: Percent rejections (wild bootstr. - Mammen), n=200, gamma=1, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.061	0.061	0.061	0.061	0.020	0.081	0.065	0.020	0.084	0.063
0.1	0.063	0.063	0.063	0.063	0.031	0.102	0.064	0.035	0.104	0.065
0.2	0.151	0.151	0.151	0.151	0.084	0.202	0.152	0.088	0.205	0.155
0.3	0.261	0.261	0.261	0.261	0.157	0.318	0.269	0.159	0.317	0.270
0.4	0.447	0.447	0.447	0.447	0.314	0.516	0.460	0.313	0.514	0.462
0.5	0.628	0.628	0.628	0.628	0.470	0.681	0.635	0.474	0.681	0.640

Table 4.3.3. Wild bootstrapping – PD_2 distribution (Rademacher)

Table 113: Percent rejections (wild bootstr. - Radem.), n=200, gamma=0, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.061	0.061	0.061	0.061	0.056	0.142	0.060	0.058	0.138	0.059
0.1	0.090	0.090	0.090	0.090	0.074	0.186	0.093	0.076	0.184	0.093
0.2	0.271	0.271	0.271	0.271	0.197	0.367	0.272	0.195	0.367	0.271
0.3	0.513	0.513	0.513	0.513	0.406	0.610	0.500	0.396	0.607	0.502
0.4	0.817	0.817	0.817	0.817	0.714	0.859	0.816	0.714	0.859	0.816
0.5	0.969	0.969	0.969	0.969	0.934	0.983	0.966	0.934	0.983	0.966

Table 119: Percent rejections (wild bootstr. - Radem.), n=200, gamma=.5, rho2=.3, rho3=.3

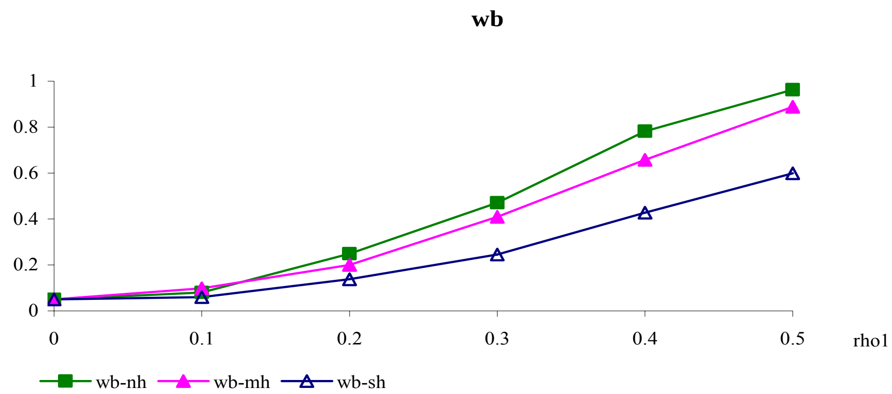
alpha=0.05

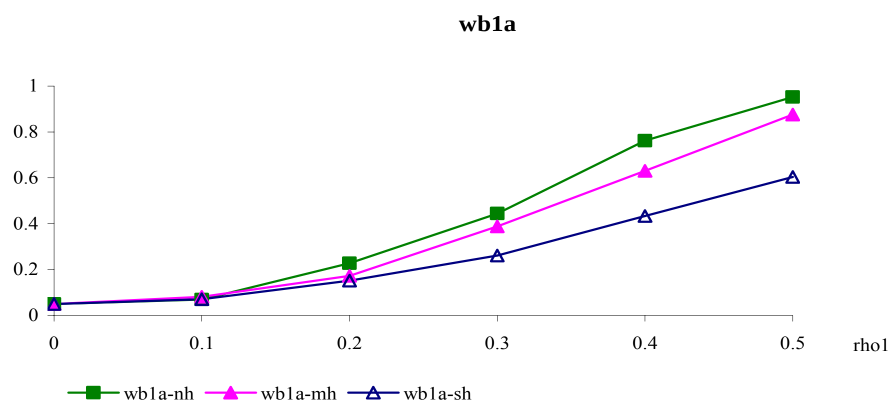
rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.042	0.042	0.042	0.042	0.037	0.110	0.045	0.038	0.112	0.046
0.1	0.097	0.097	0.097	0.097	0.086	0.201	0.094	0.086	0.201	0.094
0.2	0.202	0.202	0.202	0.202	0.149	0.302	0.205	0.146	0.302	0.205
0.3	0.417	0.417	0.417	0.417	0.309	0.538	0.414	0.309	0.529	0.412
0.4	0.674	0.674	0.674	0.674	0.566	0.748	0.667	0.566	0.747	0.666
0.5	0.895	0.895	0.895	0.895	0.842	0.936	0.906	0.840	0.936	0.905

Table 125: Percent rejections (wild bootstr. - Radem.), n=200, gamma=1, rho2=.3, rho3=.3

alpha=0.05

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.055	0.055	0.055	0.055	0.041	0.134	0.059	0.044	0.138	0.059
0.1	0.064	0.064	0.064	0.064	0.056	0.152	0.064	0.061	0.154	0.064
0.2	0.145	0.145	0.145	0.145	0.122	0.264	0.152	0.126	0.263	0.154
0.3	0.258	0.258	0.258	0.258	0.214	0.395	0.264	0.216	0.398	0.265
0.4	0.448	0.448	0.448	0.448	0.379	0.577	0.453	0.373	0.576	0.454
0.5	0.632	0.632	0.632	0.632	0.541	0.726	0.632	0.534	0.721	0.635





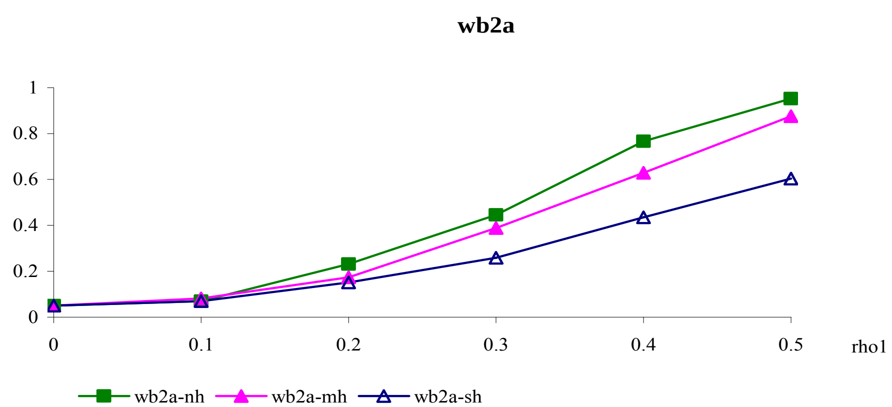


FIGURE 4.3.1. SIZE CORRECTED POWER - EFFECT OF HETEROSKEDASTICITY (n = 200, rho2 = .3, rho3 = .3)
Pairs bootstrapping
*nh – no heteroskedasticity (gamma = 0), mh – medium heteroskedasticity (gamma = .5), sh - strong heteroskedasticity (gamma = 1)

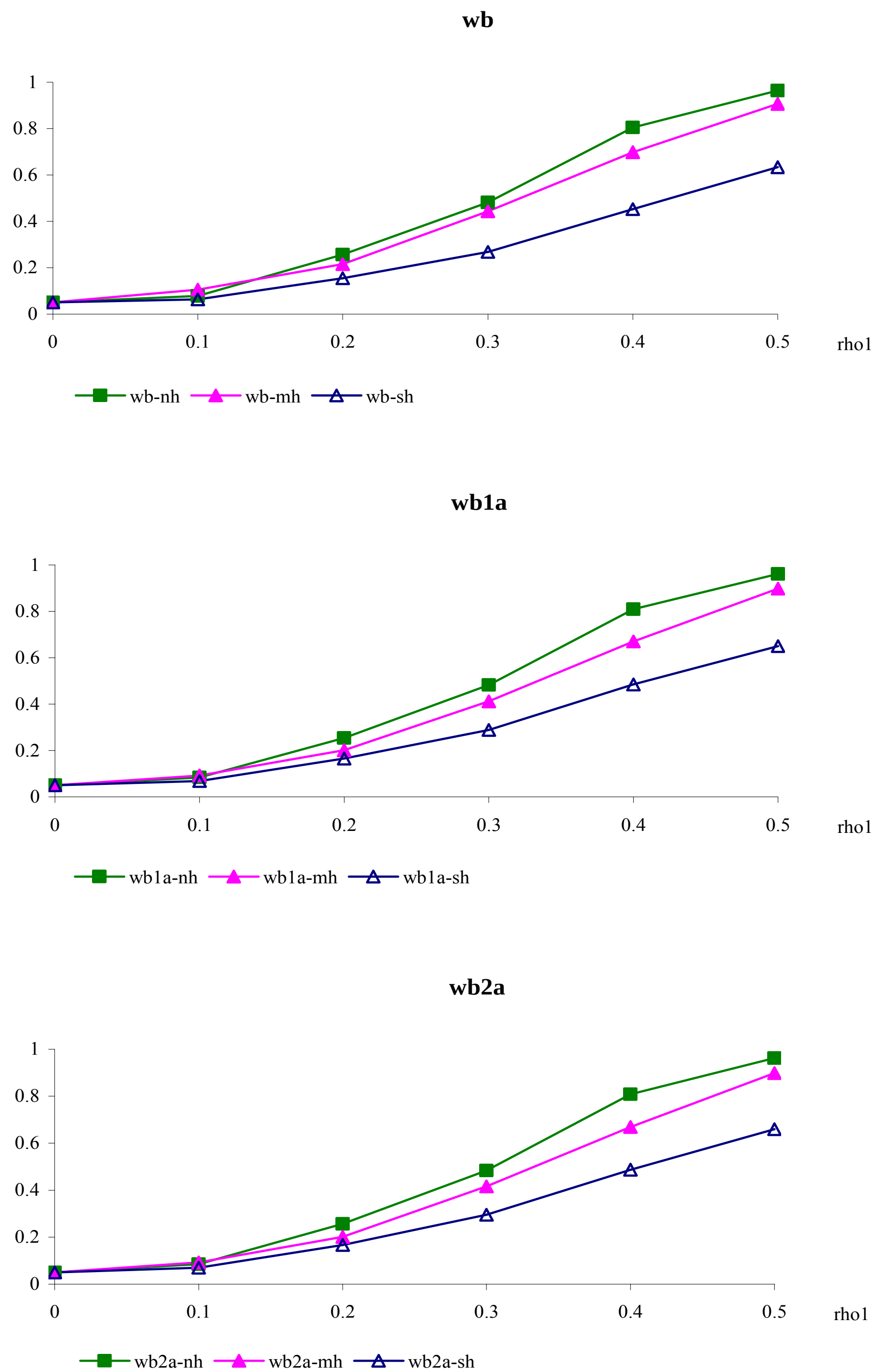
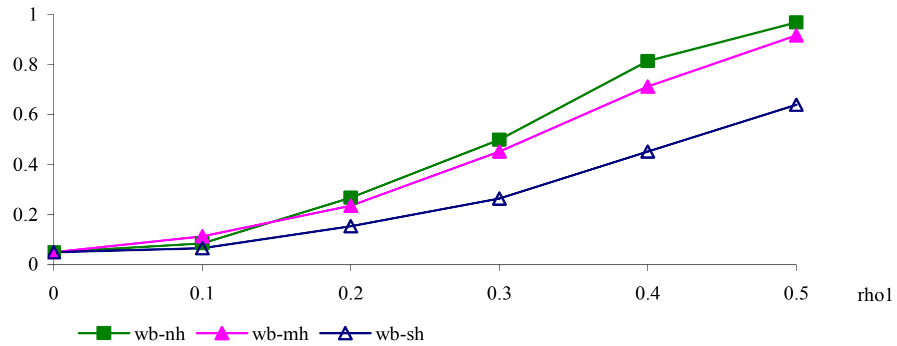
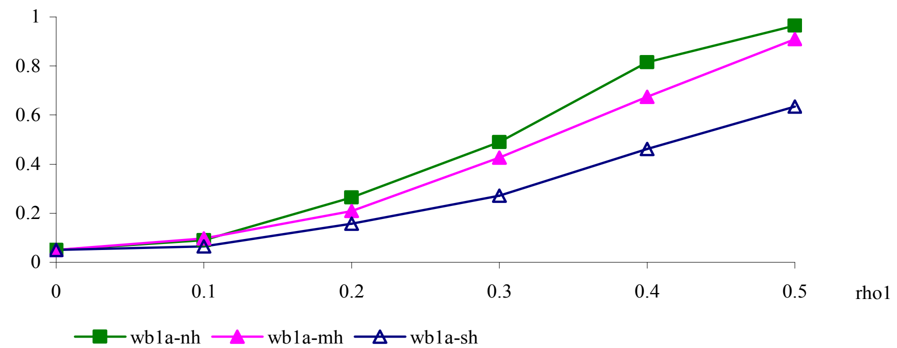


FIGURE 4.3.2. SIZE CORRECTED POWER - EFFECT OF HETEROSKEDASTICITY (n = 200, rho2 = .3, rho3 = .3)
Wild bootstrapping - PD_1 distribution
*nh – no heteroskedasticity (gamma = 0), mh – medium heteroskedasticity (gamma = .5), sh - strong heteroskedasticity (gamma = 1)

wb



wb1a



wb2a

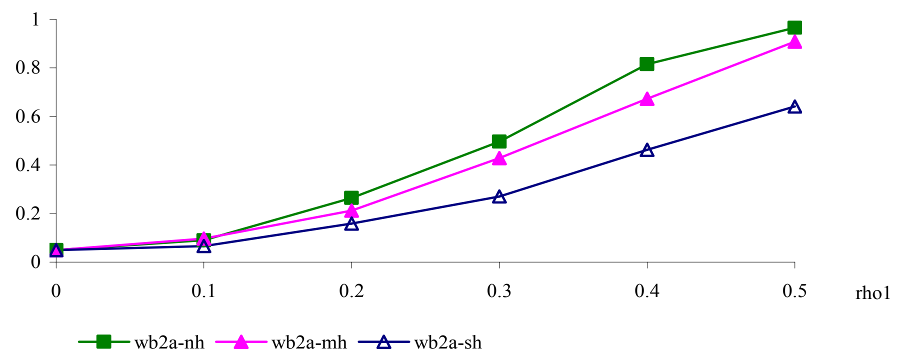


FIGURE 4.3.3. SIZE CORRECTED POWER - EFFECT OF HETEROSKEDASTICITY (n = 200, rho2 = .3, rho3 = .3)
Wild bootstrapping – PD_2 distribution (Rademacher)
*nh – no heteroskedasticity (gamma = 0), mh – medium heteroskedasticity (gamma = .5), sh - strong heteroskedasticity (gamma = 1)

4.3.4 The Effect of Bootstrapping under Heteroskedasticity

Table 4.4 and Figures 4.4 compare the effect of bootstrapping on the respective tests if the sample size n=100 and heteroskedasticity is present. In the case of two strong instruments, there are only small differences between the asymptotic and bootstrap tests. The Wald test based on the artificial regression that uses pairs bootstrapping has higher rejection rate than the test that uses wild bootstrapping but its size is also higher. The Wald tests based on the Creel's system that use wild bootstrapping have higher power than tests using pairs bootstrapping, wild bootstrapping using the Rademacher distribution performs slightly better that the one using the PD_1 distribution as in Mammen (1993). With pairs bootstrapping , the Wald tests based on the Creel's system comparing all the coefficients of the GMM and OLS or the IV and OLS estimators and having two degrees of freedom underreject the null, consequently their rejection frequencies are the lowest. After size correction, for low degrees of endogeneity pairs bootstrapping in comparison to asymptotics or to wild bootstrapping improves the performance of all tests. Size corrected Wald tests computed by pairs bootstrapping perform better than by both types of wild bootstrapping. Wild bootstrapping using the PD_1 distribution as in Mammen (1993) performs slightly better that the one using the Rademacher distribution for the size corrected Wald tests **wb1** and **wb2** based on the Creel's system comparing all the coefficients of the GMM and OLS estimators, the conclusions are reverse for the size corrected Wald test based on the Creel's system comparing only the coefficients of potentially endogenous elements of the GMM and OLS estimators.

In the case of weak instruments, the pairs bootstrapping tests perform significantly worse than the asymptotic and wild bootstrap tests. The bootstrap tests are not an improvement over the asymptotic tests except the wild bootstrap tests **wb11** and **wb21** based on the Creel's system comparing the GMM and OLS estimators and having one degree of freedom but for example their size is higher than .1 if the nominal level of significance is .05. The rejection rate of tests using the PD_1 distribution and the Rademacher distribution based on the Creel's system comparing the IV and OLS estimators is very similar. The evidence in the case of tests comparing the GMM and OLS estimators is mixed, Rademacher distribution helps to improve the rejection rate of the tests **wb1**, **wb11**, **wb2** and **wb21**; the PD_1 distribution does better in the case of the tests **wb1a** and **wb2a**.

If we compare the asymptotic and the wild bootstrapped tests in the case of one strong and one weak instrument, the only bootstrap tests that perform better than the asymptotic tests are the wild bootstrap tests **wb11** and **wb21**, the Rademacher distribution doing better than the PD_1 distribution. After size correction, the Rademacher distribution performs better than the PD_1 distribution for the Wald tests **wb1a** and **wb2a** based on the Creel's system comparing only the coefficients of potentially endogenous elements of the GMM and OLS estimators. Under weak endogeneity, pairs bootstrapping works better than wild bootstrapping for some of the size corrected Wald tests. If we compare corresponding tests (Figure 4.4.1 and Figure 4.4.2), the bootstrap **wb** is not an improvement over the asymptotic **t** test but the bootstrap **w21** test has higher power than the asymptotic **w21** test. However, after size correction, the bootstrap **wb2** test does not perform better than the asymptotic **w2** test.

Tables 4.5 provide evidence that if the sample size is higher (n = 200), wild bootstrapping is an improvement over asymptotics. Specifically, even under weak instruments, the bootstrap **w21** test has higher power than the asymptotic **w21** test and after the size correction (Figure 4.5.2) the bootstrap **wb2** test has higher power than the asymptotic **w2** test. The bootstrap **wb** is not an improvement over the asymptotic **t** test in the model with weak instruments (Figure 4.5.1).

Under two strong instruments (Table 4.5.1), wild bootstrapping does not improve test performance. After size adjustments, if the correlation between the regressor and the error term is low, pairs bootstrapping does well, it works better than wild bootstrapping. The distribution PD_1 performs better than the Rademacher distribution in the bootstrapped system tests **wb1** and **wb2** comparing GMM and OLS under weak endogeneity, differences among other tests are very small.

If only weak instruments are available (Table 4.5.2), there is no improvement over the asymptotic tests except the **wb11** and **wb21** test based on the Creel's system comparing the GMM and OLS estimators and using one degree of freedom, tests employing the Rademacher distribution doing better than the PD_1 distribution. The conclusions are reversed after size correction. The best performing are the **wb1a**

TABLE 4.4. EFFECT OF BOOTSTRAPPING - HETEROSKEDASTIC CASE (n = 100)

Table 4.4.1. Two strong instruments

alpha=0.05

Table 13: Percent rejections, n=100, gamma=1, rho2=.5, rho3=.5

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.051	0.053	0.051	0.054	0.051	0.045	0.061	0.061	0.054	0.035
0.1	0.103	0.107	0.104	0.112	0.107	0.096	0.112	0.112	0.105	0.062
0.2	0.283	0.290	0.285	0.295	0.284	0.264	0.280	0.280	0.276	0.166
0.3	0.561	0.569	0.563	0.568	0.555	0.530	0.536	0.536	0.538	0.376
0.4	0.809	0.815	0.810	0.803	0.794	0.773	0.768	0.768	0.777	0.631
0.5	0.939	0.942	0.940	0.925	0.919	0.907	0.908	0.908	0.912	0.827

Table 85: Percent rejections (pairs bootstrapping), n=100, gamma=1, rho2=.5, rho3=.5

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.061	0.012	0.052	0.043	0.010	0.047	0.034	0.009	0.040	0.032
0.1	0.102	0.026	0.089	0.076	0.018	0.065	0.053	0.016	0.057	0.048
0.2	0.314	0.109	0.283	0.254	0.079	0.248	0.224	0.073	0.239	0.207
0.3	0.553	0.267	0.491	0.475	0.204	0.460	0.429	0.187	0.443	0.408
0.4	0.796	0.501	0.755	0.737	0.424	0.701	0.673	0.402	0.685	0.654
0.5	0.922	0.721	0.904	0.883	0.626	0.859	0.830	0.595	0.841	0.815

Table 103: Percent rejections (wild bootstr. - Mammen), n=100, gamma=1, rho2=.5, rho3=.5

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.055	0.055	0.055	0.055	0.021	0.094	0.045	0.025	0.095	0.042
0.1	0.087	0.087	0.087	0.087	0.033	0.131	0.079	0.033	0.132	0.076
0.2	0.290	0.290	0.290	0.290	0.156	0.337	0.271	0.156	0.342	0.262
0.3	0.527	0.527	0.527	0.527	0.342	0.588	0.503	0.339	0.581	0.489
0.4	0.781	0.781	0.781	0.781	0.620	0.814	0.746	0.613	0.810	0.736
0.5	0.912	0.912	0.912	0.912	0.797	0.920	0.876	0.787	0.912	0.865

Table 121: Percent rejections (wild bootstr. - Radem.), n=100, gamma=1, rho2=.5, rho3=.5

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.056	0.056	0.056	0.056	0.042	0.137	0.055	0.045	0.141	0.054
0.1	0.088	0.088	0.088	0.088	0.060	0.181	0.082	0.061	0.183	0.084
0.2	0.294	0.294	0.294	0.294	0.190	0.403	0.275	0.189	0.398	0.271
0.3	0.531	0.531	0.531	0.531	0.406	0.636	0.501	0.407	0.634	0.501
0.4	0.783	0.783	0.783	0.783	0.659	0.846	0.767	0.655	0.844	0.758
0.5	0.915	0.915	0.915	0.915	0.846	0.938	0.886	0.838	0.936	0.882

Table 4.4.2. Two weak instruments

alpha=0.05

Table 14: Percent rejections, n=100, gamma=1, rho2=.3, rho3=.3

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.049	0.052	0.050	0.055	0.051	0.045	0.043	0.043	0.052	0.035
0.1	0.060	0.063	0.060	0.064	0.061	0.054	0.051	0.051	0.063	0.039
0.2	0.098	0.102	0.099	0.106	0.101	0.090	0.082	0.082	0.100	0.059
0.3	0.160	0.165	0.161	0.169	0.162	0.147	0.132	0.132	0.161	0.096
0.4	0.254	0.260	0.256	0.264	0.255	0.235	0.209	0.209	0.250	0.155
0.5	0.369	0.377	0.370	0.384	0.371	0.347	0.304	0.304	0.363	0.235

Table 86: Percent rejections (pairs bootstrapping), n=100, gamma=1, rho2=.3, rho3=.3

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.042	0.005	0.035	0.028	0.005	0.028	0.024	0.006	0.029	0.024
0.1	0.034	0.003	0.023	0.017	0.001	0.024	0.015	0.001	0.023	0.014
0.2	0.056	0.006	0.049	0.041	0.004	0.050	0.043	0.003	0.050	0.042
0.3	0.111	0.018	0.084	0.070	0.015	0.083	0.071	0.013	0.082	0.071
0.4	0.138	0.027	0.111	0.099	0.028	0.129	0.103	0.028	0.127	0.101
0.5	0.229	0.060	0.185	0.169	0.056	0.202	0.171	0.048	0.204	0.171

Table 104: Percent rejections (wild bootstr. - Mammen), n=100, gamma=1, rho2=.3, rho3=.3

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.056	0.056	0.056	0.056	0.029	0.107	0.056	0.030	0.107	0.058
0.1	0.051	0.051	0.051	0.051	0.029	0.090	0.057	0.029	0.097	0.061
0.2	0.098	0.098	0.098	0.098	0.047	0.149	0.102	0.052	0.148	0.102
0.3	0.168	0.168	0.168	0.168	0.088	0.221	0.165	0.091	0.221	0.164
0.4	0.215	0.215	0.215	0.215	0.136	0.304	0.226	0.138	0.303	0.227
0.5	0.326	0.326	0.326	0.326	0.219	0.396	0.339	0.223	0.398	0.340

Table 122: Percent rejections (wild bootstr. - Radem.), n=100, gamma=1, rho2=.3, rho3=.3

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.058	0.058	0.058	0.058	0.060	0.160	0.054	0.059	0.155	0.054
0.1	0.053	0.053	0.053	0.053	0.055	0.138	0.050	0.056	0.143	0.053
0.2	0.096	0.096	0.096	0.096	0.079	0.203	0.097	0.076	0.195	0.099
0.3	0.164	0.164	0.164	0.164	0.130	0.275	0.152	0.128	0.271	0.154
0.4	0.211	0.211	0.211	0.211	0.182	0.357	0.221	0.185	0.355	0.219
0.5	0.322	0.322	0.322	0.322	0.266	0.447	0.328	0.265	0.447	0.328

Table 4.4.3. One strong, one weak instrument

alpha=0.05

Table 15: Percent rejections, n=100, gamma=1, rho2=.5, rho3=.1

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.048	0.051	0.049	0.053	0.050	0.043	0.050	0.050	0.051	0.035
0.1	0.069	0.072	0.070	0.073	0.069	0.062	0.068	0.068	0.071	0.044
0.2	0.128	0.132	0.129	0.136	0.130	0.118	0.118	0.118	0.126	0.075
0.3	0.235	0.241	0.236	0.245	0.235	0.217	0.216	0.216	0.231	0.138
0.4	0.381	0.389	0.382	0.392	0.380	0.355	0.345	0.345	0.371	0.242
0.5	0.546	0.554	0.548	0.554	0.540	0.516	0.497	0.497	0.530	0.381

Table 87: Percent rejections (pairs bootstrapping), n=100, gamma=1, rho2=.5, rho3=.1

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.039	0.005	0.030	0.024	0.007	0.034	0.029	0.007	0.028	0.026
0.1	0.052	0.010	0.039	0.033	0.008	0.040	0.031	0.007	0.038	0.030
0.2	0.099	0.021	0.085	0.071	0.019	0.092	0.078	0.015	0.091	0.077
0.3	0.207	0.045	0.165	0.144	0.041	0.159	0.142	0.039	0.154	0.137
0.4	0.308	0.091	0.251	0.221	0.096	0.256	0.228	0.086	0.248	0.221
0.5	0.476	0.167	0.407	0.372	0.151	0.398	0.368	0.143	0.395	0.358

Table 105: Percent rejections (wild bootstr. - Mammen), n=100, gamma=1, rho2=.5, rho3=.1

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.050	0.050	0.050	0.050	0.023	0.090	0.054	0.029	0.100	0.055
0.1	0.052	0.052	0.052	0.052	0.030	0.087	0.052	0.032	0.085	0.053
0.2	0.115	0.115	0.115	0.115	0.071	0.181	0.129	0.080	0.183	0.126
0.3	0.228	0.228	0.228	0.228	0.137	0.276	0.238	0.139	0.276	0.234
0.4	0.333	0.333	0.333	0.333	0.204	0.394	0.342	0.208	0.396	0.342
0.5	0.509	0.509	0.509	0.509	0.334	0.572	0.510	0.335	0.573	0.508

Table 123: Percent rejections (wild bootstr. - Radem.), n=100, gamma=1, rho2=.5, rho3=.1

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.047	0.047	0.047	0.047	0.051	0.149	0.050	0.059	0.148	0.050
0.1	0.054	0.054	0.054	0.054	0.049	0.133	0.052	0.051	0.130	0.052
0.2	0.112	0.112	0.112	0.112	0.105	0.230	0.123	0.103	0.228	0.123
0.3	0.229	0.229	0.229	0.229	0.170	0.335	0.230	0.175	0.328	0.225
0.4	0.336	0.336	0.336	0.336	0.261	0.469	0.338	0.257	0.465	0.344
0.5	0.511	0.511	0.511	0.511	0.390	0.615	0.515	0.388	0.617	0.507

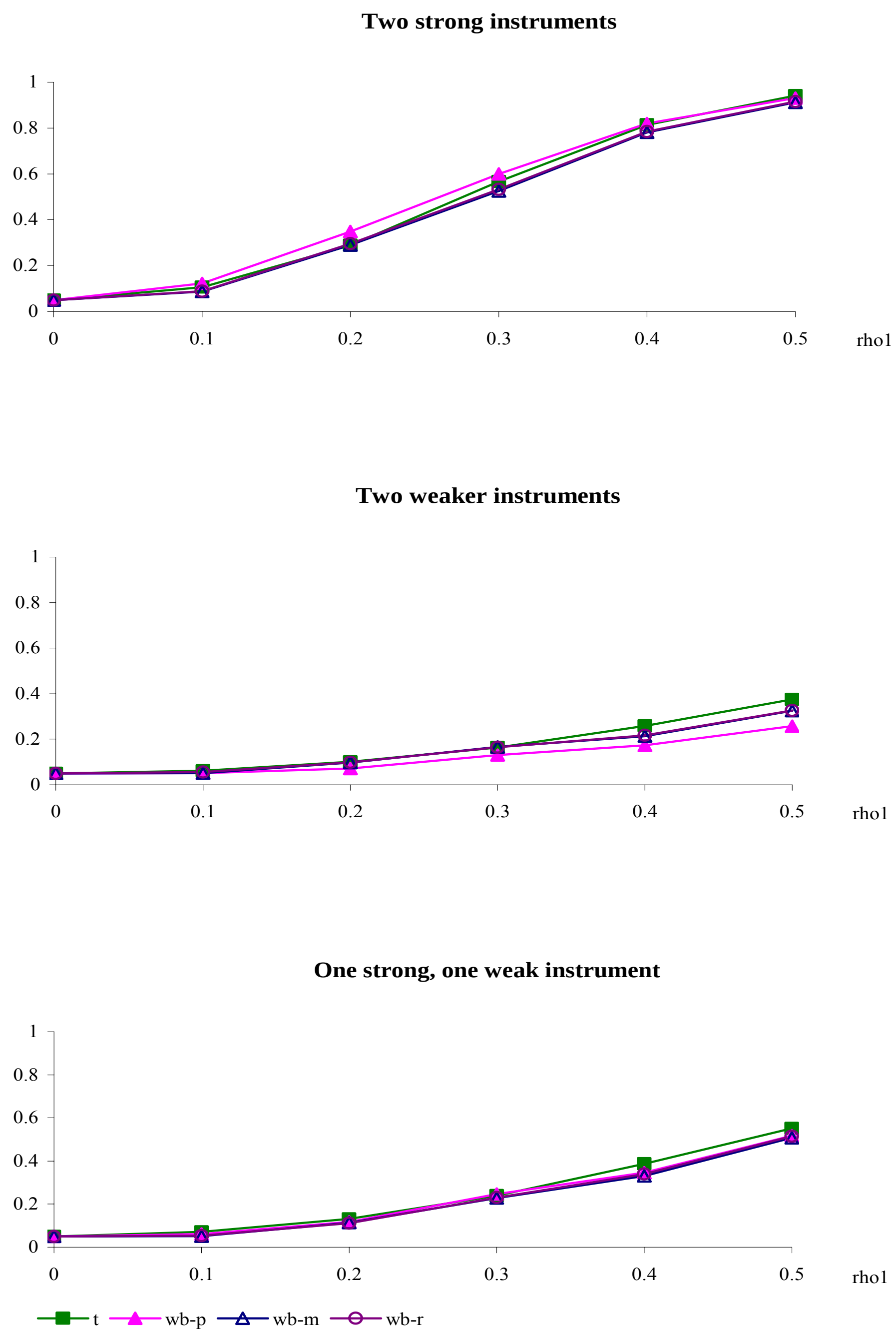


FIGURE 4.4.1.SIZE CORRECTED POWER - EFFECT OF BOOTSTRAPPING (gamma = 1, n = 100)
t/Wald test
*p – pairs bootstrapping, m – wild bootstrapping (Mammen), r – wild bootstrapping (Rademacher)

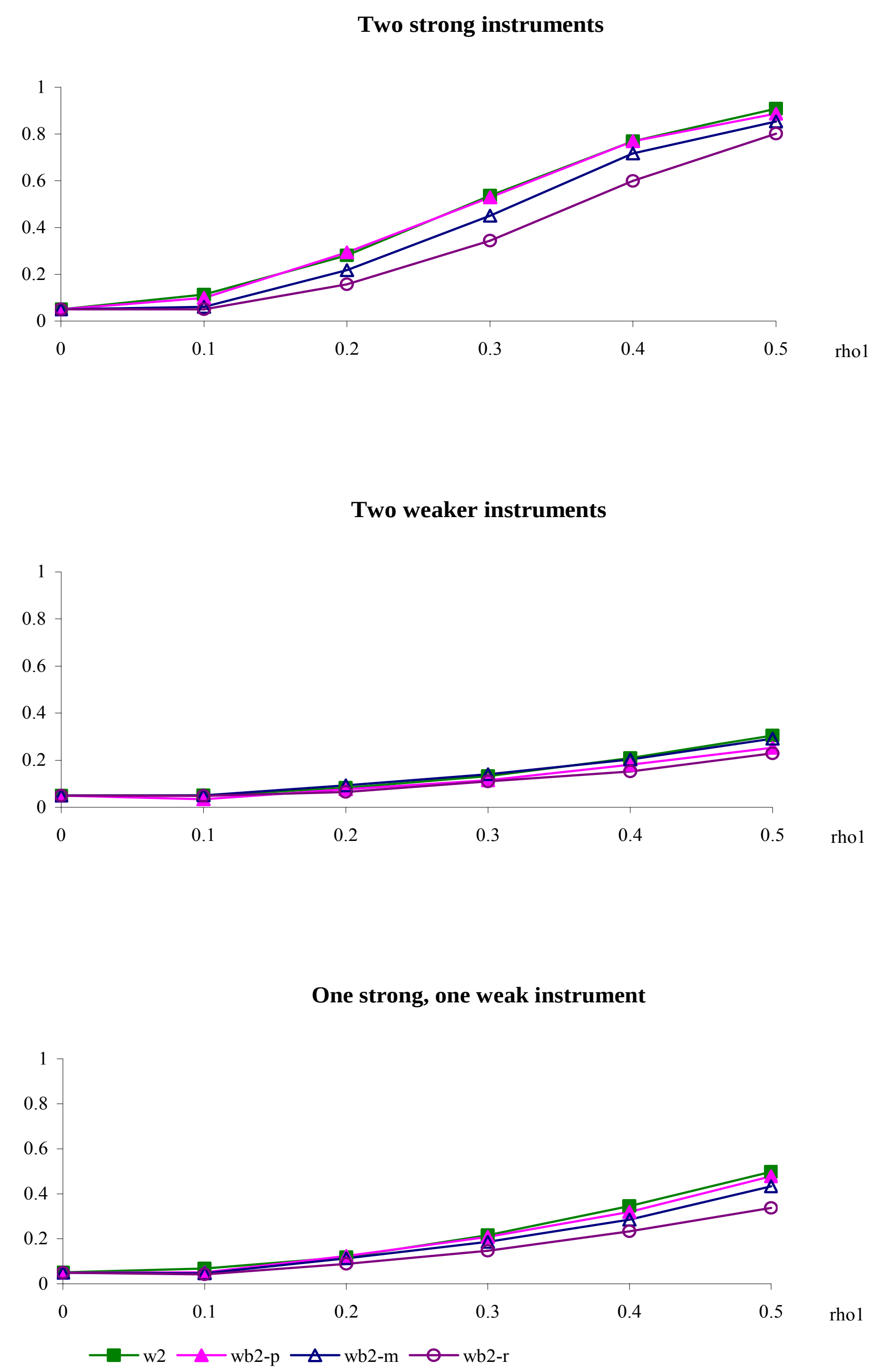


FIGURE 4.4.2. SIZE CORRECTED POWER - EFFECT OF BOOTSTRAPPING (gamma = 1, n = 100)

Creel system Wald test

*p – pairs bootstrapping, m – wild bootstrapping (Mammen), r – wild bootstrapping (Rademacher)

TABLE 4.5. EFFECT OF BOOTSTRAPPING - HETEROSKEDASTIC CASE (n = 200)

Table 4.5.1. Two strong instruments

alpha=0.05

Table 16: Percent rejections, n=200, gamma=1, rho2=.5, rho3=.5

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.051	0.052	0.051	0.054	0.052	0.049	0.056	0.056	0.052	0.044
0.1	0.167	0.170	0.168	0.173	0.169	0.161	0.170	0.170	0.166	0.113
0.2	0.516	0.520	0.517	0.520	0.513	0.501	0.506	0.506	0.506	0.377
0.3	0.853	0.855	0.854	0.852	0.848	0.840	0.838	0.838	0.841	0.737
0.4	0.980	0.980	0.980	0.975	0.974	0.972	0.971	0.971	0.972	0.941
0.5	0.998	0.998	0.998	0.996	0.996	0.995	0.995	0.995	0.996	0.989

Table 88: Percent rejections (pairs bootstrapping), n=200, gamma=1, rho2=.5, rho3=.5

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.054	0.013	0.046	0.041	0.010	0.050	0.044	0.010	0.048	0.042
0.1	0.179	0.065	0.167	0.163	0.052	0.162	0.132	0.051	0.157	0.128
0.2	0.528	0.287	0.503	0.486	0.256	0.469	0.431	0.247	0.461	0.421
0.3	0.850	0.655	0.830	0.822	0.612	0.810	0.776	0.602	0.809	0.774
0.4	0.978	0.917	0.978	0.975	0.887	0.967	0.957	0.880	0.966	0.953
0.5	0.997	0.988	0.997	0.997	0.982	0.994	0.992	0.979	0.994	0.991

Table 106: Percent rejections (wild bootstr. - Mammen), n=200, gamma=1, rho2=.5, rho3=.5

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.042	0.042	0.042	0.042	0.014	0.069	0.045	0.015	0.070	0.047
0.1	0.166	0.166	0.166	0.166	0.090	0.206	0.154	0.086	0.207	0.155
0.2	0.497	0.497	0.497	0.497	0.337	0.526	0.463	0.332	0.524	0.454
0.3	0.843	0.843	0.843	0.843	0.692	0.850	0.809	0.692	0.852	0.808
0.4	0.977	0.977	0.977	0.977	0.931	0.978	0.969	0.928	0.978	0.969
0.5	0.997	0.997	0.997	0.997	0.992	0.996	0.994	0.991	0.996	0.992

Table 124: Percent rejections (wild bootstr. - Radem.), n=200, gamma=1, rho2=.5, rho3=.5

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.038	0.038	0.038	0.038	0.040	0.131	0.050	0.043	0.125	0.052
0.1	0.167	0.167	0.167	0.167	0.123	0.273	0.162	0.122	0.271	0.160
0.2	0.501	0.501	0.501	0.501	0.383	0.587	0.483	0.383	0.585	0.479
0.3	0.844	0.844	0.844	0.844	0.735	0.880	0.822	0.734	0.877	0.820
0.4	0.976	0.976	0.976	0.976	0.940	0.982	0.971	0.939	0.981	0.970
0.5	0.997	0.997	0.997	0.997	0.994	0.997	0.995	0.994	0.997	0.995

Table 4.5.2. Two weak instruments

alpha=0.05

Table 17: Percent rejections, n=200, gamma=1, rho2=.3, rho3=.3

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.048	0.050	0.049	0.052	0.050	0.047	0.045	0.045	0.049	0.041
0.1	0.074	0.075	0.074	0.077	0.075	0.071	0.069	0.069	0.075	0.056
0.2	0.151	0.154	0.152	0.157	0.154	0.146	0.138	0.138	0.150	0.104
0.3	0.278	0.281	0.279	0.285	0.280	0.270	0.257	0.257	0.276	0.196
0.4	0.459	0.462	0.460	0.466	0.459	0.448	0.427	0.427	0.454	0.338
0.5	0.645	0.649	0.646	0.649	0.643	0.631	0.606	0.606	0.637	0.519

Table 89: Percent rejections (pairs bootstrapping), n=200, gamma=1, rho2=.3, rho3=.3

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.053	0.007	0.047	0.044	0.004	0.047	0.042	0.004	0.044	0.041
0.1	0.059	0.013	0.048	0.044	0.011	0.053	0.044	0.011	0.053	0.042
0.2	0.131	0.037	0.120	0.112	0.036	0.125	0.112	0.035	0.125	0.110
0.3	0.231	0.071	0.213	0.205	0.077	0.219	0.198	0.075	0.216	0.196
0.4	0.417	0.185	0.384	0.376	0.197	0.401	0.376	0.194	0.397	0.372
0.5	0.590	0.288	0.546	0.537	0.296	0.564	0.538	0.298	0.565	0.537

Table 107: Percent rejections (wild bootstr. - Mammen), n=200, gamma=1, rho2=.3, rho3=.3

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.061	0.061	0.061	0.061	0.020	0.081	0.065	0.020	0.084	0.063
0.1	0.063	0.063	0.063	0.063	0.031	0.102	0.064	0.035	0.104	0.065
0.2	0.151	0.151	0.151	0.151	0.084	0.202	0.152	0.088	0.205	0.155
0.3	0.261	0.261	0.261	0.261	0.157	0.318	0.269	0.159	0.317	0.270
0.4	0.447	0.447	0.447	0.447	0.314	0.516	0.460	0.313	0.514	0.462
0.5	0.628	0.628	0.628	0.628	0.470	0.681	0.635	0.474	0.681	0.640

Table 125: Percent rejections (wild bootstr. - Radem.), n=200, gamma=1, rho2=.3, rho3=.3

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.055	0.055	0.055	0.055	0.041	0.134	0.059	0.044	0.138	0.059
0.1	0.064	0.064	0.064	0.064	0.056	0.152	0.064	0.061	0.154	0.064
0.2	0.145	0.145	0.145	0.145	0.122	0.264	0.152	0.126	0.263	0.154
0.3	0.258	0.258	0.258	0.258	0.214	0.395	0.264	0.216	0.398	0.265
0.4	0.448	0.448	0.448	0.448	0.379	0.577	0.453	0.373	0.576	0.454
0.5	0.632	0.632	0.632	0.632	0.541	0.726	0.632	0.534	0.721	0.635

Table 4.5.3. One strong, one weak instrument

alpha=0.05

Table 18: Percent rejections, n=200, gamma=1, rho2=.5, rho3=.1

rho1	ho3	ho3a	t	tr	tr2	tr3	w21	d1	w2r1	DMF
0.0	0.051	0.053	0.052	0.053	0.051	0.048	0.052	0.052	0.052	0.042
0.1	0.089	0.091	0.089	0.092	0.090	0.085	0.087	0.087	0.089	0.066
0.2	0.214	0.217	0.214	0.219	0.215	0.206	0.207	0.207	0.212	0.147
0.3	0.425	0.429	0.426	0.430	0.425	0.412	0.407	0.407	0.419	0.305
0.4	0.655	0.659	0.655	0.659	0.653	0.642	0.628	0.628	0.645	0.516
0.5	0.839	0.841	0.839	0.835	0.830	0.822	0.813	0.813	0.825	0.726

Table 90: Percent rejections (pairs bootstrapping), n=200, gamma=1, rho2=.5, rho3=.1

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.046	0.009	0.041	0.039	0.007	0.037	0.030	0.006	0.037	0.030
0.1	0.079	0.017	0.072	0.069	0.017	0.074	0.071	0.017	0.073	0.069
0.2	0.209	0.090	0.191	0.186	0.081	0.196	0.181	0.079	0.193	0.179
0.3	0.395	0.190	0.369	0.358	0.184	0.377	0.355	0.181	0.372	0.344
0.4	0.637	0.366	0.607	0.589	0.357	0.611	0.594	0.350	0.602	0.587
0.5	0.826	0.589	0.807	0.795	0.579	0.809	0.790	0.570	0.807	0.788

Table 108: Percent rejections (wild bootstr. - Mammen), n=200, gamma=1, rho2=.5, rho3=.1

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.046	0.046	0.046	0.046	0.023	0.066	0.045	0.022	0.067	0.045
0.1	0.084	0.084	0.084	0.084	0.037	0.109	0.081	0.037	0.109	0.080
0.2	0.213	0.213	0.213	0.213	0.118	0.257	0.209	0.124	0.258	0.210
0.3	0.406	0.406	0.406	0.406	0.269	0.475	0.416	0.271	0.474	0.414
0.4	0.649	0.649	0.649	0.649	0.473	0.699	0.653	0.473	0.695	0.651
0.5	0.840	0.840	0.840	0.840	0.723	0.870	0.844	0.724	0.869	0.840

Table 126: Percent rejections (wild bootstr. - Radem.), n=200, gamma=1, rho2=.5, rho3=.1

rho1	wb	wb0	wb01	wb0a	wb1	wb11	wb1a	wb2	wb21	wb2a
0.0	0.047	0.047	0.047	0.047	0.050	0.130	0.046	0.050	0.129	0.044
0.1	0.085	0.085	0.085	0.085	0.068	0.174	0.081	0.069	0.170	0.081
0.2	0.217	0.217	0.217	0.217	0.162	0.317	0.216	0.164	0.316	0.216
0.3	0.411	0.411	0.411	0.411	0.328	0.530	0.416	0.326	0.526	0.413
0.4	0.646	0.646	0.646	0.646	0.531	0.744	0.650	0.527	0.741	0.648
0.5	0.846	0.846	0.846	0.846	0.764	0.896	0.844	0.763	0.894	0.842

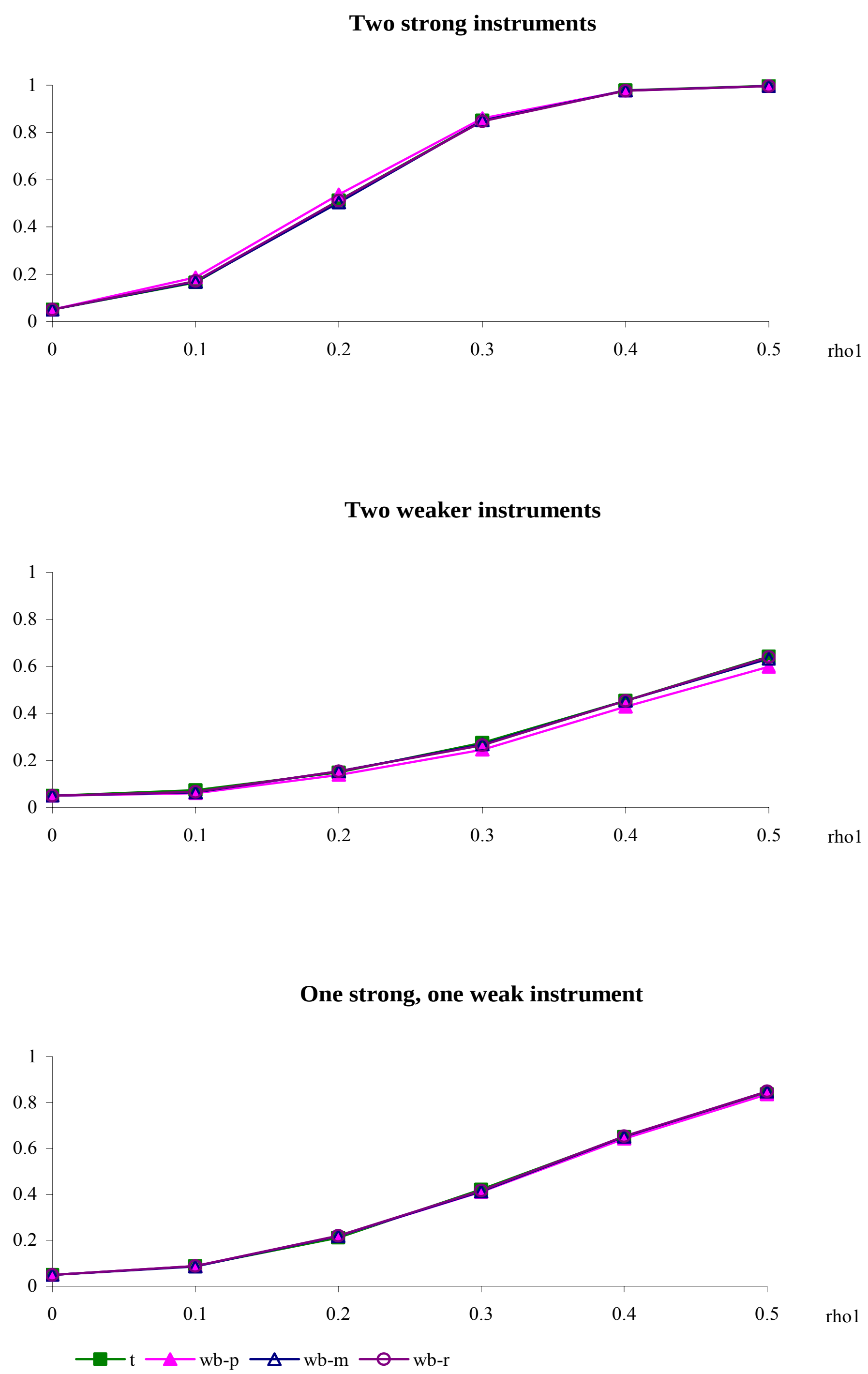
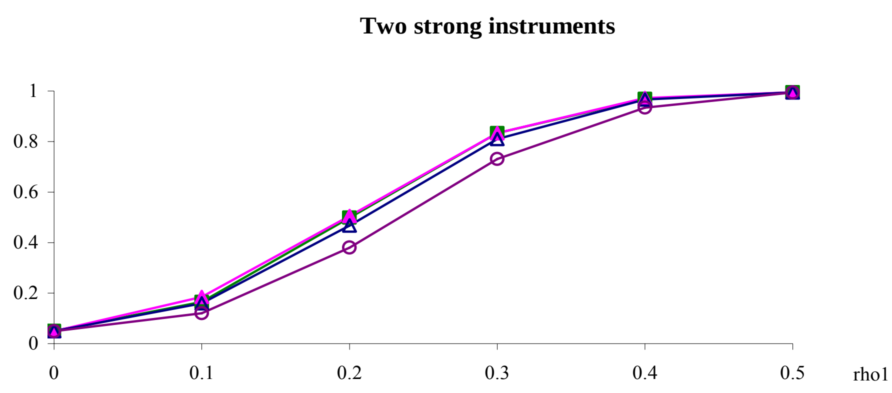
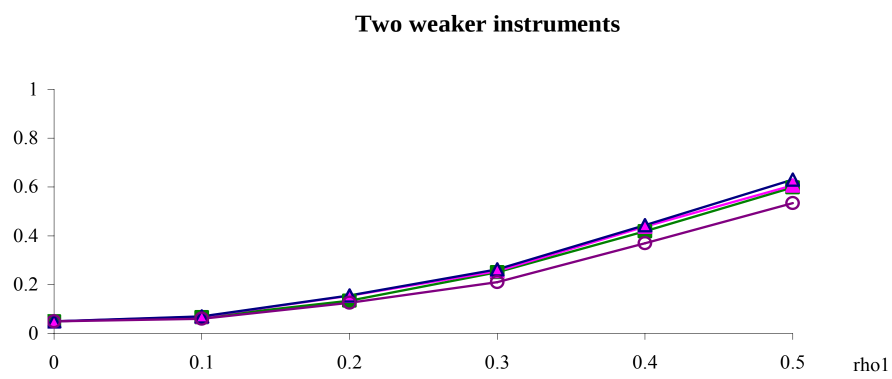


FIGURE 4.5.1. SIZE CORRECTED POWER - EFFECT OF BOOTSTRAPPING (gamma = 1, n = 200)
t/Wald test
*p – pairs bootstrapping, m – wild bootstrapping (Mammen), r – wild bootstrapping (Rademacher)





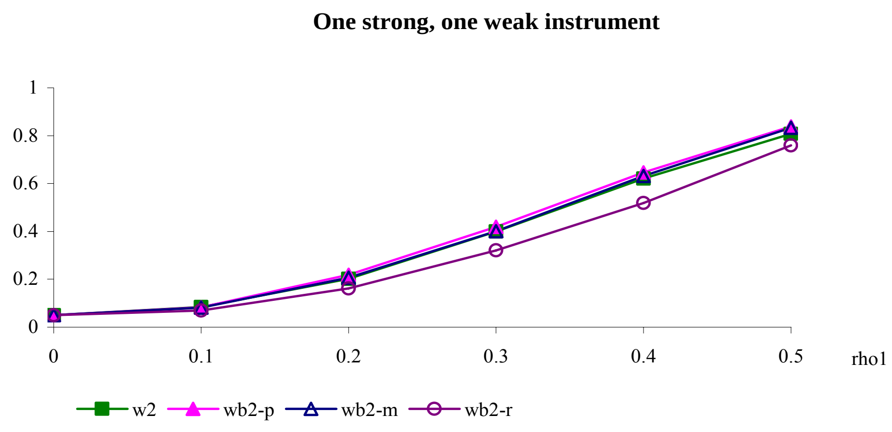


FIGURE 4.5.2. SIZE CORRECTED POWER - EFFECT OF BOOTSTRAPPING (gamma = 1, n = 200)
Creel system Wald test
*p – pairs bootstrapping, m – wild bootstrapping (Mammen), r – wild bootstrapping (Rademacher)

and `wb2a` tests based on the Creel's system comparing only the coefficients of potentially endogenous elements of the GMM and OLS estimators computed using the PD_1 distribution.

If we compare the size corrected powers of the asymptotic and the wild bootstrapping tests in the case of one strong and one weak instrument (Table 4.5.3), there is an improvement over the asymptotic equivalents for the `wb` test if wild bootstrapping is used and for the `wb2` test if pairs bootstrapping or wild bootstrapping with the PD_1 distribution are used. The best performing among the Wald tests, after level adjustments, are the `wb1a` and `wb2a` tests based on the Creel's system comparing only the coefficients of potentially endogenous elements of the GMM and OLS estimators computed using the PD_1 distribution.

4.4 Summary of Findings on Bootstrap Tests under Heteroskedasticity

We have bootstrapped the Wald test based on the artificial regression and the Wald tests based on the Creel's (2004) system approach if heteroskedastic errors were present.

We considered impacts of the sample size and instrument strength. The effect of sample size on power improvement was remarkable especially if only weak instruments were available. Under pairs bootstrapping, a larger sample size reduced size distortions of most of the tests in all scenarios but this conclusion could not be made under wild bootstrapping where instrument strength seemed to have an impact on a test size. The Rademacher distribution seemed to have better size properties than the PD_1 distribution used in Mammen (1993). As expected, the stronger the instruments the better the performance of the tests. Models with one strong and one weak instrument seem to have the best size properties. The size is more distorted under pairs than under wild bootstrapping.

With two strong instruments, after size adjustments, pairs bootstrapping of the tests performed better than both types of wild bootstrapping but in the case of weak instruments, the pairs bootstrap tests performed significantly worse than the asymptotic and wild bootstrap tests.

The errors used in the data generating process following the Rademacher distribution give better results than the distribution proposed by Mammen (1993) but after size correction, wild bootstrapping using the distribution as in Mammen (1993) performs slightly better than the one using the Rademacher distribution for the Wald test based on the Creel's system comparing all the coefficients of the GMM and

OLS estimators, the conclusions are reversed for the Wald test based on the Creel's system comparing only the coefficients of potentially endogenous elements of the GMM and OLS estimators.

The performance of the bootstrapped tests based on the Creel's (2004) system approach is not an improvement over the classical bootstrapped tests (the Wald test of the artificial regression coefficient) except a few scenarios with weak instruments under strong heteroskedasticity ($\gamma = 1$) and with one strong and one weak instrument when a sample size is larger.

We have provided evidence that for a larger sample size ($n = 200$) wild bootstrapping is a slight improvement over asymptotics, in particular for the Hausman test comparing potentially endogenous elements of the OLS and the second round GMM estimates. Bootstrapping increased the rejection rate even under strong heteroskedasticity and in the model with weak instruments. On the other hand, bootstrapping the Wald test based on the artificial regression did not provide improvements over first-order asymptotics when instruments were irrelevant.

Our Monte Carlo results are very limited. We only consider overidentified models with one endogenous regressor and two instruments. To evaluate the tests performance in different models, further study is needed.

5. EXOGENEITY TESTS IN COUNT DATA MODELS

5.1 Poisson Regression Model

There is a wide range of microeconomics examples with a nonnegative integer as the dependent variable. The number of occurrences of an event is usually modeled by a Poisson distribution, whose density is

$$\Pr[Y = y] = \frac{\exp(-\lambda)\lambda^y}{y!}, \qquad y = 0,1,2,....$$

The mean $E[Y] = \lambda$, the variance $V[Y] = \lambda$. The Poisson regression model is derived from the Poisson distribution by the exponential mean parameterization of the relation between the mean λ and K linearly independent regressors x_i:

$$\lambda_i = \exp(x_i'\beta), \quad i = 0,1,2,...,n.$$

Since $V(y_i | x_i) = \exp(x_i'\beta)$, the regression is heteroskedastic. The standard approach is to estimate the model by maximum likelihood. If we assume independence across observations, the log-likelihood function becomes

$$\ln L(\beta) = \sum_{i=1}^{n}\left[y_i x_i'\beta - \exp(x_i'\beta) - \ln y_i!\right]$$

The Poisson maximum likelihood estimator $\hat{\beta}_{ML}$ is the solution to the K first-order conditions: $\sum_{i=1}^{n}\left[y_i - \exp(x_i'\beta)\right] x_i = 0$. If the density of the number of occurrences is misspecified, the estimator is referred to as pseudo-ML estimator. The PML estimator $\hat{\beta}_{PML}$ is consistent, if the conditional mean is correctly specified and regressors x_i are exogenous.

Following Cameron and Trivedi (2005, p.669) , the variance matrix of the PML estimator is

$$V\left(\hat{\beta}_{PML}\right) = \left(\sum_{i=1}^{n}\lambda_i x_i x_i'\right)^{-1}\left(\sum_{i=1}^{n}\omega_i x_i x_i'\right)\left(\sum_{i=1}^{n}\lambda_i x_i x_i'\right)^{-1}$$

where $\omega_i = V(y_i | x_i)$ is the conditional variance of y_i. If $\hat{\lambda}_i \xrightarrow{p} \lambda_i$, then $N^{-1}\sum_{i=1}^{n}\left(y_i - \hat{\lambda}_i\right)^2 x_i x_i' \xrightarrow{p} \lim N^{-1}\sum_{i=1}^{n}\omega_i^2 x_i x_i'$ and a consistent estimator of $V\left(\hat{\beta}_{PML}\right)$ is thus

(5.1) $$\hat{V}\left(\hat{\beta}_{PML}\right)=\left(\sum_{i=1}^{n}\hat{\lambda}_i x_i x_i'\right)^{-1}\left(\sum_{i=1}^{n}\left(y_i-\hat{\lambda}_i\right)^2 x_i x_i'\right)\left(\sum_{i=1}^{n}\hat{\lambda}_i x_i x_i'\right)^{-1}$$

If all the regressors are exogenous, the (pseudo-)maximum likelihood estimator is consistent but inefficient. Endogeneity leads to maximum likelihood estimators that are inconsistent, so alternative estimators have to be used.

In the following, we describe alternative tests for endogeneity in the Poisson regression model. We compare the performance of the the classical contrast tests with uncorrected covariance matrix, the Wald test based on the artificial Poisson regression and the GMM based tests using the Creel's (2004) system approach. Further, we investigate how all these tests are affected by the sample size and by the strength of the instruments.

5.2 GMM Framework

To obtain estimators that are consistent under weak distributional assumptions, we can use the method of moments.

5.2.1 Additive Error Model

The exponential mean model with additive zero-mean error term assumes

$$y_i = E\left[y_i \middle| x_i\right]+u_i=\exp(x_i'\beta)+u_i$$

GMM estimation in the Poisson regression model is based on the conditional moments

$E\left[y_i-\exp(x_i'\beta)\middle|x_i\right]=0$. If the regressors are endogenous, instruments z_i have to be found such that $E\left[y_i-\exp(x_i'\beta)\middle|z_i\right]=0$ and $E\left[u_i\middle|z_i\right]=0$. The consistent estimator of β then mimimizes

$$Q(\beta)=\left[\sum_{i=1}^{n}z_i\left(y_i-\exp(x_i'\beta)\right)\right]' W_n\left[\sum_{i=1}^{n}z_i\left(y_i-\exp(x_i'\beta)\right)\right]$$

where W_n is a symmetric weighting matrix.

For the additive errors model define

(5.2) $$\hat{D}_a=\left.\frac{\partial\left(y-\exp(x'\beta)\right)}{\partial\beta'}\right|_{\hat{\beta}}=\sum_{i=1}^{n}\exp\left(x_i'\hat{\beta}\right)x_i' \text{ and}$$

(5.3) $$\hat{S}=\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2 z_i z_i' \text{ where } \hat{u}_i=y_i-\exp(x_i'\hat{\beta})$$

The nonlinear two stage least squares/instrumental variables estimator $\hat{\beta}_{NIV}$ solves

$$\min_{\beta}\left[\sum_{i=1}^{n} z_i\left(y_i-\exp(x_i'\beta)\right)\right]'\left(\frac{Z'Z}{n}\right)^{-1}\left[\sum_{i=1}^{n} z_i\left(y_i-\exp(x_i'\beta)\right)\right]$$

Its estimated asymptotic variance is [Cameron and Trivedi (2005, p. 105)]

$$\hat{V}\left(\hat{\beta}_{NIV}\right)=n\left(\hat{D}_a'Z(Z'Z)^{-1}Z'\hat{D}_a\right)^{-1}\left(\hat{D}_a'Z(Z'Z)^{-1}\hat{S}(Z'Z)^{-1}Z'\hat{D}_a\right)\left(\hat{D}_a'Z(Z'Z)^{-1}Z'\hat{D}_a\right)^{-1}$$

where $\hat{D}_a$ and $\hat{S}$ are as in (5.2) and (5.3), respectively and evaluated at $\hat{\beta}_{NIV}$.

The optimal GMM estimator $\hat{\beta}_{GMM}$ minimizes the objective function

$$Q(\beta)=\left[\sum_{i=1}^{n} z_i\left(y_i-\exp(x_i'\beta)\right)\right]'\hat{S}^{-1}\left[\sum_{i=1}^{n} z_i\left(y_i-\exp(x_i'\beta)\right)\right].$$

with the weighting matrix equal to the inverse of the asymptotic covariance matrix of the moment conditions evaluated at the first round estimate $\hat{\beta}_{NIV}$.

The optimal $\hat{\beta}_{GMM}$ estimated asymptotic variance is [Cameron and Trivedi (2005, p. 105)]

$$\hat{V}\left(\hat{\beta}_{GMM}\right)=n\left(\hat{D}_a'Z\hat{S}^{-1}Z'\hat{D}_a\right)^{-1}$$

where $\hat{D}_a$ and $\hat{S}$ are as in (5.2) and (5.3), respectively and evaluated at $\hat{\beta}_{GMM}$.

5.2.2 Multiplicative Error Model

The exponential mean model with multiplicative zero-mean error term assumes

$$y_i=\exp(x_i'\beta)\nu_i=\exp(x_i'\beta+u_i),\quad \nu_i=\ \exp(u_i).$$

It treats the observables and unobservables symmetrically. Mullahy (1997) shows that the multiplicative error model has some advantages over the additive one. Specifically, nonlinear IV estimators based on an additive residual function are not generally consistent whereas IV estimators based on multiplicative errors are.

The instrumental variables in the multiplicative error model satisfy $E\left[\nu_i \middle| z_i\right]=1$ and the IV estimator is based on the moment conditions $E\left[\frac{y_i}{\exp(x_i'\beta)}-1\middle|z_i\right]=0$. Windmeijer and Silva (1997) show that the same set of instruments will not, in general, be orthogonal to both error types. If Z are valid instruments in the additive model, then MZ (where M = diag (λ_i)) are valid instruments in the multiplicative model, giving the same estimation results for the two model specifications.

The consistent estimator of β in the multiplicative error model minimizes

$$Q(\beta) = \left[\sum_{i=1}^{n} z_i \left(\frac{y_i}{\exp(x_i'\beta)} - 1\right)\right]' W_n \left[\sum_{i=1}^{n} z_i \left(\frac{y_i}{\exp(x_i'\beta)} - 1\right)\right]$$

where W_n is a symmetric weighting matrix. For the multiplicative errors models define

$$(5.4) \qquad \hat{D}_m = \left.\frac{\partial\left(y/\exp(x'\beta)\right) - 1}{\partial\beta'}\right|_{\hat{\beta}} = \sum_{i=1}^{n} \frac{-y_i}{\exp\left(x_i'\hat{\beta}\right)} x_i' \text{ and}$$

$$(5.5) \qquad \hat{S} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 z_i z_i' \text{ where } \hat{u}_i = \frac{y_i}{\exp(x_i'\hat{\beta})} - 1$$

Thus, the optimal GMM estimator $\hat{\beta}_{GMM}$ for the multiplicative model minimizes the objective function

$$Q(\beta) = \left[\sum_{i=1}^{n} z_i \left(\frac{y_i}{\exp(x_i'\beta)} - 1\right)\right]' \hat{S}^{-1} \left[\sum_{i=1}^{n} z_i \left(\frac{y_i}{\exp(x_i'\beta)} - 1\right)\right]$$

where $\hat{S}$ is evaluated at $\hat{\beta}_{NIV}$. Its estimated asymptotic variance is

$$\hat{V}\left(\hat{\beta}_{GMM}\right) = n\left(\hat{D}_m' Z \hat{S}^{-1} Z' \hat{D}_m\right)^{-1}$$

where $\hat{D}_m$ and $\hat{S}$ are as in (5.4) and (5.5), respectively and evaluated at $\hat{\beta}_{GMM}$.

In the subsequent GMM applications, we assume multiplicative residual functions. Following Grogger (1990) and Creel (2004), we compute the usual Hausman test statistic that assumes one efficient estimator under the null hypothesis.

$$H_{NIV} = \left(\hat{\beta}_{NIV} - \hat{\beta}_{ML}\right)' \left[\hat{V}\left(\hat{\beta}_{NIV}\right) - \hat{V}\left(\hat{\beta}_{ML}\right)\right]^{+} \left(\hat{\beta}_{NIV} - \hat{\beta}_{ML}\right) \text{ and } H_{NIV} \overset{a}{\sim} \chi^2_{(K_1)}$$

where K_1 is the number of possibly endogenous regressors.

We also compute the test statistics H_{GMM} using the optimal GMM estimator $\hat{\beta}_{GMM}$.

$$H_{GMM} = \left(\hat{\beta}_{GMM} - \hat{\beta}_{ML}\right)' \left[\hat{V}\left(\hat{\beta}_{GMM}\right) - \hat{V}\left(\hat{\beta}_{ML}\right)\right]^{+} \left(\hat{\beta}_{GMM} - \hat{\beta}_{ML}\right) \text{ and } H_{GMM} \overset{a}{\sim} \chi^2_{(K_1)}$$

To avoid the estimator inefficiency problem under the null hypothesis in the usual Hausman test, we follow Creel's (2004) approach combining the sets of moment conditions ($E\left[y_i - \exp(x_i'\beta)\middle|x_i\right] = 0$ and $E\left[\frac{y_i}{\exp(x_i'\beta)} - 1\middle|z_i\right] = 0$) that lead to the ML and IV estimators into a single estimation problem. The procedure was already described in the previous chapter.

The two-step GMM system estimator $\hat{\beta}_{GMM2}$ obtained using Creel's approach minimizes

$$(5.6) \qquad Q(\beta) = \left\{ \begin{bmatrix} [y - \exp(x\beta)]'x & 0 \\ 0 & [y/\exp(x\beta) - 1]'Z \end{bmatrix} \hat{S}^{-1} \begin{bmatrix} x'[y - \exp(x\beta)] & 0 \\ 0 & Z'[y/\exp(x\beta) - 1] \end{bmatrix} \right\}$$

where the weighting matrix is the inverse of the asymptotic covariance matrix of the two sets of moment conditions evaluated at the first round estimate. The asymptotic covariance matrix

$$(5.7) \qquad V\left(\hat{\beta}_{GMM2}\right) = n \left\{ \begin{bmatrix} \hat{D}_a'X & 0 \\ 0 & \hat{D}_m'Z \end{bmatrix} \hat{S}^{-1} \begin{bmatrix} X'\hat{D}_a & 0 \\ 0 & Z'\hat{D}_m \end{bmatrix} \right\}^{-1}$$

where $\hat{D}_a$ and $\hat{D}_m$ are as in (5.2) and (5.4), respectively, $\hat{S}$ is the asymptotic covariance matrix of the two sets of moment conditions, all are evaluated at $\hat{\beta}_{GMM2}$.

The Wald test of the null hypothesis that $H_0 : plim\left(\hat{\beta}_{ML} - \hat{\beta}_{GMM}\right) = 0$ can then be based upon the two-step GMM system estimator $\hat{\beta}_{GMM2}$ and $V\left(\hat{\beta}_{GMM2}\right)$. If $R = \left[I_K \quad -I_K\right]$ then the test statistic is

$$(5.8) \qquad W_2 = \left(R\hat{\beta}_{GMM2}\right)' \left[RV\left(\hat{\beta}_{GMM2}\right)R'\right]^{-1} \left(R\hat{\beta}_{GMM2}\right)$$

Under the null hypothesis it has an asymptotic distribution with K degrees of freedom. If we use just the coefficients of the potentially endogenous regressors X_1 and define $R_1 = \left[I_{K_1} \quad 0 \quad -I_{K_1} \quad 0\right]$ the corresponding Wald statistic will be called W_{2a} and it will have K_1 degrees of freedom.

Based on Burnside and Eichenbaum (1996) finding that Wald tests have improved properties when restrictions of the null hypothesis of equality of the two estimators are imposed (we call the restricted GMM system estimator $\hat{\beta}_r$), we compute the restricted analogues of W_2 and W_{2a}, named W_{2r} and W_{2ar}. Their degrees of freedom under the null hypothesis are K and K_1, respectively.

Next test we perform is the test suggested by Cameron and Trivedi (2005, p. 245) comparing the objective function at the unrestricted estimates $\hat{\beta}_{GMM2}$ [from (3.2)] and restricted estimates $\hat{\beta}_r$:

(5.9) $$D = n\left[Q\left(\hat{\beta}_r\right) - Q\left(\hat{\beta}_u\right)\right]$$

Under the null hypothesis, this test statistic has an asymptotic chi-square distribution with number of degrees of freedom equal to *K*.

5.3 Artificial Regression

Tests based on auxiliary regression are typically robust to nonsphericality. We follow Wooldridge (2002, p. 663) who suggests the omitted variable approach to derive a test for endogeneity. Let the regressors x_i be partitioned into two parts, a $1 \times K_1$ vector of endogenous variables, x_{1i}, and $1 \times K_2$ vector of exogenous and predetermined variables, x_{2i}. Let c_1 be an unobserved variable potentially correlated with x_{1i}. The equation can then be rewritten as

(5.10) $$E\left[y_i \mid x_i\right] = \exp(x_{1i}'\beta_1 + x_{2i}'\beta_2 + c_{1i})$$

There are assumed to be $L \geq K$ instruments, of which K_2 are the columns of the matrix x_2. The values of the endogenous variables are assumed to be determined by a set of linear simultaneous equations. The reduced form equations for the endogenous explanatory variables can be written as

(5.11) $$x_{1i} = z_i \Pi_1 + v_i$$

where Π_1 is a $L \times K_1$ matrix and v_i is a $1 \times K_1$ vector of error terms.

We assume that $c_{1i} = v_i \rho_1 + error$ and that the *error* term is independent of v_i and z_i. Then x_{1i} is exogenous if and only if $\rho_1 = 0$. The null hypothesis $\rho_1 = 0$ can be tested using the Wald or LM statistic in the artificial regression

(5.12) $$E\left[y_i \mid x_{1i}, x_{2i}, \hat{v}_i\right] = \exp(x_{1i}'\beta_1 + x_{2i}'\beta_2 + \hat{v}_i\rho_1)$$

where $\hat{v}_i$ are the estimated residuals from the OLS first stage regression. The model can be estimated by pseudo-MLE. If the conditional mean is correctly specified, the resulting estimator ($\hat{\beta}_{2SCML}$) is consistent and no distributional assumptions about x_{1i} or x_{2i} are needed.

We compute the Wald statistics as

$$(5.13) \quad W = n\hat{\rho}_1{}'\hat{V}_0(\hat{\rho}_1)^{-1}\hat{\rho}_1$$

where $\hat{V}_0(\hat{\rho}_1)$ is a consistent estimator of the lower $K_1 \times K_1$ right-hand block of the covariance matrix of the two stage conditional maximum likelihood corresponding to $\hat{\rho}_1$. The covariance matrix of $\hat{\rho}_1$ is estimated under the null and thus is the usual covariance matrix. The test statistics is distributed chi-squared with the degrees of freedom equal to the number of variables specified as potentially endogenous. If the null hypothesis is rejected, a maximum likelihood estimator should not be employed.

We also compute the usual Hausman test statistic comparing the maximum likelihood and the two stage conditional maximum likelihood. The test can be based on the coefficients of just the endogenous variables, all the regressors and all the variables including the first stage reduced form residuals. Let $\theta = (\beta_1{}', \beta_2{}', \rho_1)$. Under the null of no endogeneity, θ is consistently though not efficiently estimated by the maximum likelihood estimator $\hat{\beta}_{ML} = (\hat{\beta}_{ML1}, \hat{\beta}_{ML2})$. Under the alternative, the two-stage conditional maximum likelihood estimator $\hat{\theta}_{2SCML} = (\hat{\beta}_{2SCML1}, \hat{\beta}_{2SCML2}, \hat{\rho}_1)$ is still consistent. Let $\hat{\theta}_{ML} = (\hat{\beta}_{ML1}, \hat{\beta}_{ML2}, 0)$. Then the alternative Hausman tests are

$$H_{CML1} = (\hat{\beta}_{2SCML1} - \hat{\beta}_{ML1})'\left[\hat{V}\left(\hat{\beta}_{2SCML1}\right) - \hat{V}\left(\hat{\beta}_{ML1}\right)\right]^{+}(\hat{\beta}_{2SCML1} - \hat{\beta}_{ML1})$$

$$H_{CML2} = T(\hat{\beta}_{2SCML} - \hat{\beta}_{ML})'\left[\hat{V}\left(\hat{\beta}_{2SCML}\right) - \hat{V}\left(\hat{\beta}_{ML}\right)\right]^{+}(\hat{\beta}_{2SCML} - \hat{\beta}_{ML})$$

$$H_{CML3} = T(\hat{\theta}_{2SCML} - \hat{\theta}_{ML})'\left[\hat{V}\left(\hat{\theta}_{2SCML}\right) - \hat{V}\left(\hat{\theta}_{ML}\right)\right]^{+}(\hat{\theta}_{2SCML} - \hat{\theta}_{ML})$$

All the test statistics have a number of degrees of freedom equal to the number of potentially endogenous variables. We also calculate the test statistics H_{CML2s} and H_{CML3s} using the degrees of freedom equal to the rank of the difference of covariance matrices.

5.4 A Monte Carlo Experiment in Count Data Models

We use the same count data model as in Creel (2004):

$$(5.14) \quad y \sim \text{Poisson}(\lambda), \text{ where } \lambda = \exp(\beta_1 + \beta_2 x + u) = \exp(-.5 + x + u)$$

but our instrumental variables estimator is based on different instruments, we use smaller sample sizes and perform more tests. Following our linear regression model the data are generated as

$$(5.15)\quad \begin{bmatrix} x \\ u \\ z_1 \\ z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & 0 & 0 \\ \rho_2 & 0 & 1 & \rho_4 \\ \rho_3 & 0 & \rho_4 & 1 \end{bmatrix} \right)$$

Similarly to the previous part:

- endogeneity is controlled by the parameter ρ_1 which takes the values 0, .1, .2, .3, .4 and .5
- the strength of the instruments is controlled by ρ_2 and ρ_3 which take the values .1, .3 and .5
- samples of sizes $n = 100$ and $n = 200$ are considered.

We perform 40,000 simulations. The instrumental variables estimator is based on the multiplicative residual function $y/\lambda - 1$ and the instruments z_1 and z_2.

The notation we use:

- n = sample size
- ρ_1 controls endogeneity
- ρ_2 controls strength of IV #1
- ρ_3 controls strength of IV #2
- ρ_4 correlation between instruments
- α = nominal level of significance
- **hc1** = H_1 contrast, 2SCMLE and MLE (endogenous regressor only), with $K = 1$ df
- **hc2** = H_1 contrast, 2SCMLE and MLE (original regressors), with $K = 1$ df
- **w** = Wald test of the residual coefficient in auxiliary regression
- **w2** = Creel system test using 2nd round estimator $K = 2$ df
- **w21** = Creel test **w2** with $K_1 = 1$ df
- **w2a** = **w2** (endogenous regressors only) with $K_1 = 1$ df
- **w2r** = Creel test **w2** with covariance evaluated at estimates restricted by null, $K = 2$ df
- **w2r1** = **w2r** with $K_1 = 1$ df

- **w2ar** = **w2r** (endogenous regressors only) with $K_1 = 1$ df
- **d** = Cameron & Trivedi (2005, p. 245) D with $K = 2$ df
- **d1** = **d** test with $K_1 = 1$ df

5.5 Monte Carlo Experiment Results for Count Data Models

As expected the performance of the contrast tests with uncorrected covariance matrix H_{NIV}, H_{GMM}, H_{CML1}, H_{CML2} and H_{CML3} was not satisfactory, so we report only the best performing **hc1** (H_{CML1}) and **hc2** (H_{CML2}) comparing the two-stage conditional maximum likelihood estimator and the maximum likelihood estimator of the endogenous regressors and all the original model regressors. The power of H_{CML3} was usually the highest but its size was higher than of the other two. All the test results can be found in an appendix that is available upon request.

5.5.1 The Effect of Sample Size

In Table 5.1 we compare the effect of sample size. Sample sizes $n = 100$ and 200 are examined and we have assumed one strong and one weak instrument. Under the null hypothesis on no endogeneity, the tests should reject with frequency α. When α = .05 and $n = 100$ the contrast tests **hc1** and **hc2**, the auxiliary regression test **w**, the system tests **w2** and **w21** (5.8) that are based on a second round improvement of the GMM estimator (5.6) and with usual covariance matrix (5.7), with the degrees of freedom equal to the nominal number of restrictions being tested and one respectively, have their size above .15. Also the size of the test **w2r1** that uses the restricted GMM covariance estimator and presumes one degree of freedom is close to .15. The system test (5.8) **w2a** that uses the coefficients of the potentially endogenous regressors and the usual covariance matrix, the test **w2r** that uses the restricted GMM covariance estimator and presumes the degrees of freedom equal to the nominal number of restrictions being tested and the test **d1** (5.9), based on the difference of the two GMM objective functions with one degree of freedom still over-reject but their size is below .10. The test **w2ar** that uses the coefficients of the potentially endogenous regressors and the restricted covariance matrix and the test **d2** (5.9), based on the difference of the two GMM objective functions with the number of degrees of freedom equal to the nominal number of restrictions both under-reject the correct null hypothesis. When $n = 200$, the test sizes remain very similar. The test **w** has the highest rejection rate (because of its too frequent rejection of the null), followed by the test **w21**. After size correction, the test **w2a** performs the best but its power is still very low. For example, when a correlation between the endogenous regressor and error is $\rho_1 = .3$, it rejects the null only 20.68% of the time for sample size *n* = 100 and 33.68% of the time for the sample size *n* = 100.

5.5.2 The Effect of Instrument Strength

In Table 5.2 we compare the effect of instrument strength for sample size *n* = 100 and α = .05. The rejection frequencies in the top panel correspond to the case with two strong instruments, the second panel to two weaker instruments, and the third panel to one strong and one weak instrument.

If we have two strong instruments, the test **w** based on the artificial regression performs the best followed by the tests **w21**, **hc1**, **hc2** and **w2r1**. For example, if a correlation between the endogenous regressor and error $\rho_1 = .3$, the test **w** rejects exogeneity 79.97% of the time. The power of the tests **w21**, **hc1**, **hc2** and **w2r1** ranges from .56 to .61. The test **w2ar** rejects exogeneity least frequently: 18.96% of the time.

If we have two weaker instruments, with correlation to the regressor $\rho_2 = \rho_3 = .3$, the performance of all tests worsens radically. The test **w** rejects exogeneity 70.22% of the time and power of the other tests is below .5 if $\rho_1 = .3$.

If we have one strong and one weak instrument, the power of all tests slightly improves in comparison to the case with two weak ones. But still, if $\rho_1 = .3$, the power of the test **w** is only .71 and the power of the other tests is between .51 (**w21**) and .10 (**w2ar**).

In Figure 5.1 we present the size corrected results of the effect of instrument strength for sample size *n* = 100 and α = .05. With two strong instruments, the test **w2a** rejects the null most frequently, however the differences among the tests **w2a**, **w2r**, **d** and **w** in case of stronger endogeneity are negligible, especially at significance levels above .1. If we have two weak instruments, the best performing test is the test **w** which is based on the artificial regression. With one strong and one weak instrument, the conclusions are the same as in the case with two strong instruments.

5.5.3 Estimator Bias Results

We also report the Monte Carlo means of the alternative estimators and their root-mean-squared-errors (rmse) for the parameter β_2. The estimators whose performance we report are the pseudo-ML

TABLE 5.1. PERCENT REJECTIONS (COUNT DATA) – EFFECT OF SAMPLE SIZE

Table 183: Percent rejections (Poisson model), n=100, rho2=.5, rho3=.1

alpha=0.05

rho1	hc1	w	w2	w21	w2a	w2r	w2r1	w2ar	d	d1
0.0	0.17165	0.53240	0.17645	0.28650	0.08150	0.06407	0.14122	0.03340	0.02760	0.08775
0.1	0.20038	0.56405	0.22012	0.33742	0.13003	0.08603	0.18045	0.04980	0.04718	0.12835
0.2	0.26190	0.62895	0.28935	0.41990	0.19737	0.12960	0.24710	0.07422	0.07760	0.18993
0.3	0.33810	0.70795	0.36917	0.50587	0.27858	0.18770	0.32560	0.10450	0.11905	0.26180
0.4	0.42805	0.78643	0.46410	0.60382	0.37983	0.27127	0.42838	0.14725	0.18388	0.35983
0.5	0.54183	0.86615	0.56210	0.70225	0.48785	0.36623	0.52568	0.19160	0.26143	0.46087

Table 186: Percent rejections (Poisson model), n=200, rho2=.5, rho3=.1

alpha=0.05

rho1	hc1	w	w2	w21	w2a	w2r	w2r1	w2ar	d	d1
0.0	0.16658	0.52053	0.21795	0.33222	0.08847	0.06500	0.14910	0.02688	0.03245	0.09903
0.1	0.22867	0.57770	0.29325	0.41287	0.17203	0.10300	0.21285	0.04793	0.07390	0.17790
0.2	0.31925	0.68437	0.41300	0.54150	0.29440	0.18230	0.32605	0.08007	0.14347	0.29657
0.3	0.43880	0.79297	0.54722	0.67550	0.44412	0.30043	0.46792	0.12695	0.24978	0.44270
0.4	0.55978	0.87857	0.67998	0.79217	0.59647	0.44095	0.61015	0.18108	0.38405	0.59578
0.5	0.69600	0.94303	0.79060	0.87810	0.73042	0.57890	0.73033	0.23435	0.53380	0.73628

Table 5.1a. Size Corrected Power

Table 189: Size Corrected Power (Poisson model), n=100, rho2=.5, rho3=.1

alpha=0.05

rho1	hc1	hc2	w	w2	w2a	w2r	w2ar	d
0.0	0.05000	0.05000	0.05000	0.05000	0.05000	0.05000	0.05000	0.05000
0.1	0.06423	0.06065	0.06170	0.07235	0.08778	0.06772	0.07068	0.08023
0.2	0.09555	0.08700	0.10073	0.10270	0.13847	0.10453	0.10180	0.12095
0.3	0.13878	0.12438	0.15673	0.15030	0.20680	0.15668	0.13667	0.17870
0.4	0.19275	0.17738	0.24063	0.20605	0.29502	0.23233	0.18815	0.25930
0.5	0.27545	0.27005	0.35997	0.27782	0.39582	0.32145	0.23342	0.35033

estimator ($\hat{\beta}_{PML}$), the two-stage conditional maximum likelihood estimator ($\hat{\beta}_{2SCML}$), the nonlinear two stage least squares/instrumental variables estimator ($\hat{\beta}_{NIV}$) and the optimal 2-step GMM ($\hat{\beta}_{GMM}$).

The true value of $\beta_2 = 1$. The sample means of the estimators are given in Table 5.3 for a case with weak instruments and sample sizes of $n = 100$ and $n = 200$. The bias in the ML estimator increases with the degree of endogeneity of the regressor *x*. The bias of the two-stage conditional maximum likelihood estimator decreases with higher degree of endogeneity, it also improves with a larger sample. If only weak instruments are available, under the null hypothesis of no endogeneity the instrumental variables and GMM estimators have almost no bias with $n = 100$, but their bias increases in larger samples. This seems counterintuitive but as Nelson and Starz (1990b) show, the finite sample distribution of the instrumental variable estimator is very different from its asymptotic counterpart. For a linear model with one regressor and one instrument, it is bimodal, fat-tailed and may be concentrated around a point which is closer to the least squares estimator than to the true parameter. Not only does it depend on the asymptotic variance but also on the degree of endogeneity. Another problem is that if instruments are irrelevant, the asymptotic variance grows without limit. And, the larger the asymptotic variance the poorer approximation the asymptotic density of finite sample density is. As the asymptotic variance rises (for example by decreasing a sample size), the small sample distribution becomes more, not less, concentrated. Thus, with weak instruments, the bias of the instrumental variable estimator may actually increase with a larger sample size. The expected result prevails for larger degrees of endogeneity: if $\rho_1 \geq .4$, the instrumental variables and the GMM estimator biases reduce with a larger sample size.

In Figure 5.2 we compare the estimators means under different instrument strength. Under the null of no endogeneity, the best performing are the instrumental variables and GMM estimators with weak instruments. However, their bias under the alternative is the largest with weak instruments, the estimators perform better if one weak and one strong instruments are available. If only strong instruments or one strong and one weak instrument were available, the ML estimator and the two-stage conditional maximum likelihood estimator would have the smallest bias under the null. The mean of the two-stage conditional maximum likelihood estimator increases with stronger endogeneity, the bias is the smallest at high degrees

TABLE 5.2. PERCENT REJECTIONS (COUNT DATA) – EFFECT OF INSTRUMENT STRENGTH (n = 100)

Table 181: Percent rejections (Poisson model), n=100, rho2=.5, rho3=.5

alpha=0.05

rho1	hc1	w	w2	w21	w2a	w2r	w2r1	w2ar	d	d1
0.0	0.24093	0.38435	0.10248	0.19478	0.08697	0.07895	0.16965	0.03938	0.03035	0.09325
0.1	0.30610	0.46505	0.16980	0.28058	0.17250	0.12068	0.23635	0.06915	0.06840	0.16212
0.2	0.42960	0.63205	0.29860	0.44083	0.32033	0.23020	0.38470	0.11930	0.14752	0.29607
0.3	0.57835	0.79970	0.45600	0.61250	0.48888	0.39387	0.56153	0.18962	0.26395	0.45913
0.4	0.72535	0.92027	0.61418	0.75805	0.65263	0.57108	0.72395	0.27387	0.41118	0.62382
0.5	0.86692	0.98195	0.74513	0.86393	0.78270	0.72288	0.84318	0.34373	0.56620	0.76425

Table 182: Percent rejections (Poisson model), n=100, rho2=.3, rho3=.3

alpha=0.05

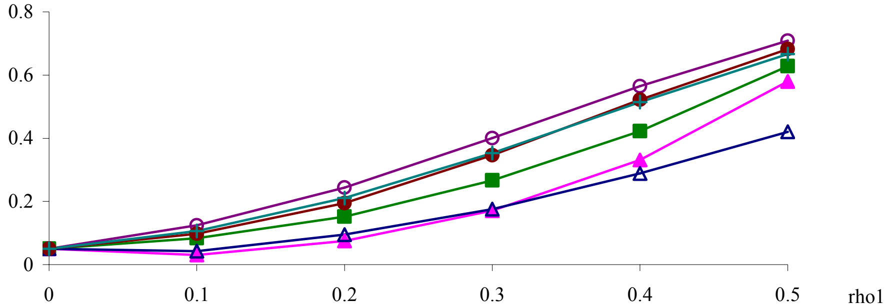
rho1	hc1	w	w2	w21	w2a	w2r	w2r1	w2ar	d	d1
0.0	0.14680	0.59105	0.20838	0.32370	0.07287	0.05532	0.12730	0.02962	0.02405	0.07863
0.1	0.16673	0.61500	0.24580	0.36212	0.10525	0.07477	0.16133	0.04445	0.03595	0.10798
0.2	0.20190	0.65453	0.29957	0.42040	0.15507	0.10663	0.20625	0.06115	0.05917	0.15180
0.3	0.24850	0.70217	0.35145	0.48157	0.20720	0.14480	0.25967	0.08198	0.08438	0.19938
0.4	0.31685	0.76055	0.42423	0.55780	0.27895	0.19470	0.32815	0.10670	0.11965	0.26527
0.5	0.40382	0.82132	0.50523	0.64097	0.36685	0.26637	0.41505	0.14113	0.17325	0.35213

Table 183: Percent rejections (Poisson model), n=100, rho2=.5, rho3=.1

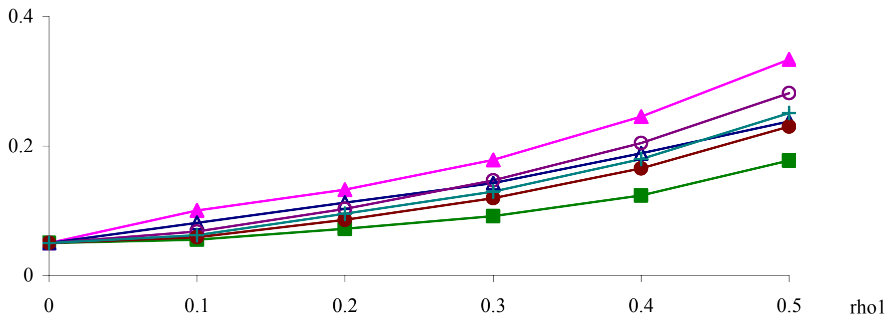
alpha=0.05

rho1	hc1	w	w2	w21	w2a	w2r	w2r1	w2ar	d	d1
0.0	0.17165	0.53240	0.17645	0.28650	0.08150	0.06407	0.14122	0.03340	0.02760	0.08775
0.1	0.20038	0.56405	0.22012	0.33742	0.13003	0.08603	0.18045	0.04980	0.04718	0.12835
0.2	0.26190	0.62895	0.28935	0.41990	0.19737	0.12960	0.24710	0.07422	0.07760	0.18993
0.3	0.33810	0.70795	0.36917	0.50587	0.27858	0.18770	0.32560	0.10450	0.11905	0.26180
0.4	0.42805	0.78643	0.46410	0.60382	0.37983	0.27127	0.42838	0.14725	0.18388	0.35983
0.5	0.54183	0.86615	0.56210	0.70225	0.48785	0.36623	0.52568	0.19160	0.26143	0.46087

Two strong instruments



Two weaker instruments



One strong, one weak instrument

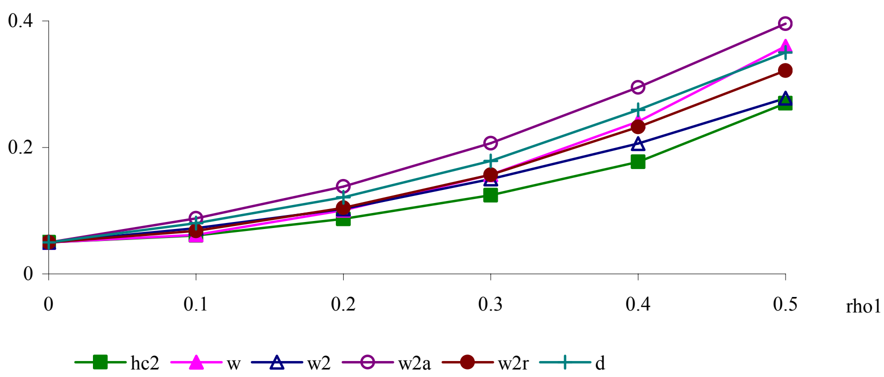


FIGURE 5.1. SIZE CORRECTED POWER - COUNT DATA (n = 100)

of endogeneity. Thus we confirm Nelson and Starz (1990a) result from a linear model that the bias of the instrumental variable estimator is inversely related to the degree of "feedback from the error to the regressor."

5.5.4 Estimator RMSE Comparisons

In Table 5.4 we compare the root-mse's of the alternative estimators for a case with weak instruments and sample sizes of *n* = 100 and *n* = 200. The RMSE of the ML estimator exceeds that of the two-stage conditional maximum likelihood estimator, the IV and GMM estimators when the degree of endogeneity $\rho_1 \geq .5$ (when *n* = 100) and if $\rho_1 \geq .4$ when *n* = 200, otherwise it is the lowest. When the degree of endogeneity is low, the two-stage conditional maximum likelihood estimator performs better than the IV or GMM estimator but worse than the ML estimator. The IV and the GMM estimator perform the best in a larger sample under strong endogeneity. The differences between them are negligible.

In Figures 5.3 and 5.4 we compare the root-mse's of the alternative estimators under different instrument strength in a sample of size *n* = 100. For low degrees of endogeneity, the RMSE of the MLE estimator is the smallest. The root-mse of the two-stage conditional maximum likelihood estimator is smaller than the root-mse's of the IV or GMM estimator for any degree of endogeneity. Each estimator performs the best if there are strong instruments available. The root-mse's decline with increasing endogeneity probably because of the impact of the estimator biases which, for linear models, are the

smaller the higher the correlation between the regressor and the error as demonstrated by Nelson and Starz (1990a).

5.6 Summary of Findings in Count Data Models

As expected, the performance of the contrast tests with uncorrected covariance matrix was not satisfactory. The tests **hc1** and **hc2** that are based on the two stage conditional maximum likelihood performed relatively better, often comparably to some of the tests based on the Creel's (2004) system approach. The Wald test **w** based on the artificial Poisson regression had the highest rejection rate of all tests, including also the tests based on the Creel's (2004) system approach but this was caused by its too frequent rejection of the null. Among the GMM based tests using the Creel's (2004) system approach, the test **w21** that is based on a two step optimal GMM estimator and uses the usual covariance matrix, with

TABLE 5.3. COUNT DATA - ESTIMATOR MEANS - EFFECT OF SAMPLE SIZE

Table 194: Estimator mean values (Poisson model), n=100, rho2=.3, rho3=.3

rho1	b2mle	b2tscmle	b2niv	b2gmm
0.0	0.96866	0.95798	1.00140	1.00983
0.1	1.06954	0.96450	1.00980	1.01936
0.2	1.16757	0.97604	1.01781	1.02514
0.3	1.26191	0.99039	1.03334	1.04107
0.4	1.35845	1.00696	1.04709	1.05563
0.5	1.45248	1.02450	1.06045	1.06900

Table 197: Estimator mean values (Poisson model), n=200, rho2=.3, rho3=.3

rho1	b2mle	b2tscmle	b2niv	b2gmm
0.0	0.98118	0.96455	1.02260	1.03007
0.1	1.08501	0.95896	0.99500	1.00076
0.2	1.18447	0.96077	0.97928	0.98434
0.3	1.27914	0.97193	0.97874	0.98400
0.4	1.36962	0.98216	0.98268	0.98808
0.5	1.45819	0.99541	0.99949	1.00517

TABLE 5.4. COUNT DATA - ESTIMATOR RMSE - EFFECT OF SAMPLE SIZE

Table 200: Estimator rmse (Poisson model), n=100, rho2=.3, rho3=.3

rho1	b2mle	b2tscmle	b2niv	b2gmm
0.0	0.21265	0.46820	0.52011	0.52639
0.1	0.23048	0.47530	0.52384	0.53457
0.2	0.27990	0.47718	0.51936	0.52526
0.3	0.34550	0.47399	0.50610	0.51255
0.4	0.42363	0.46710	0.49367	0.50083
0.5	0.50426	0.45052	0.47185	0.47674

Table 203: Estimator rmse (Poisson model), n=200, rho2=.3, rho3=.3

rho1	b2mle	b2tscmle	b2niv	b2gmm
0.0	0.16252	0.34551	0.41015	0.41484
0.1	0.18674	0.35281	0.40294	0.40874
0.2	0.25068	0.35221	0.38243	0.38519
0.3	0.32599	0.34827	0.35487	0.35758
0.4	0.40484	0.34532	0.32583	0.32608
0.5	0.48632	0.33153	0.30939	0.30960

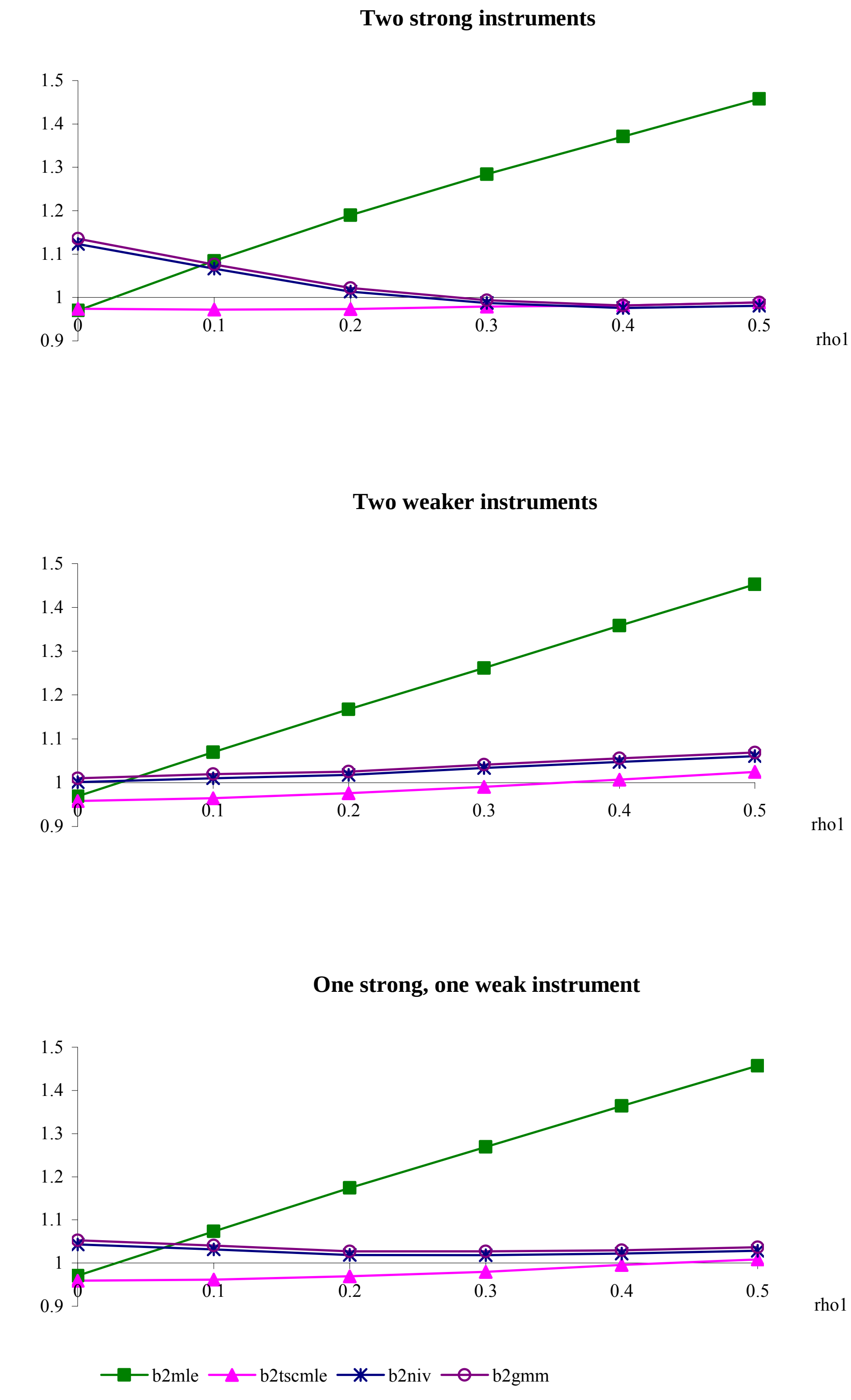


FIGURE 5.2. ESTIMATOR MEAN VALUES - COUNT DATA (n = 100)

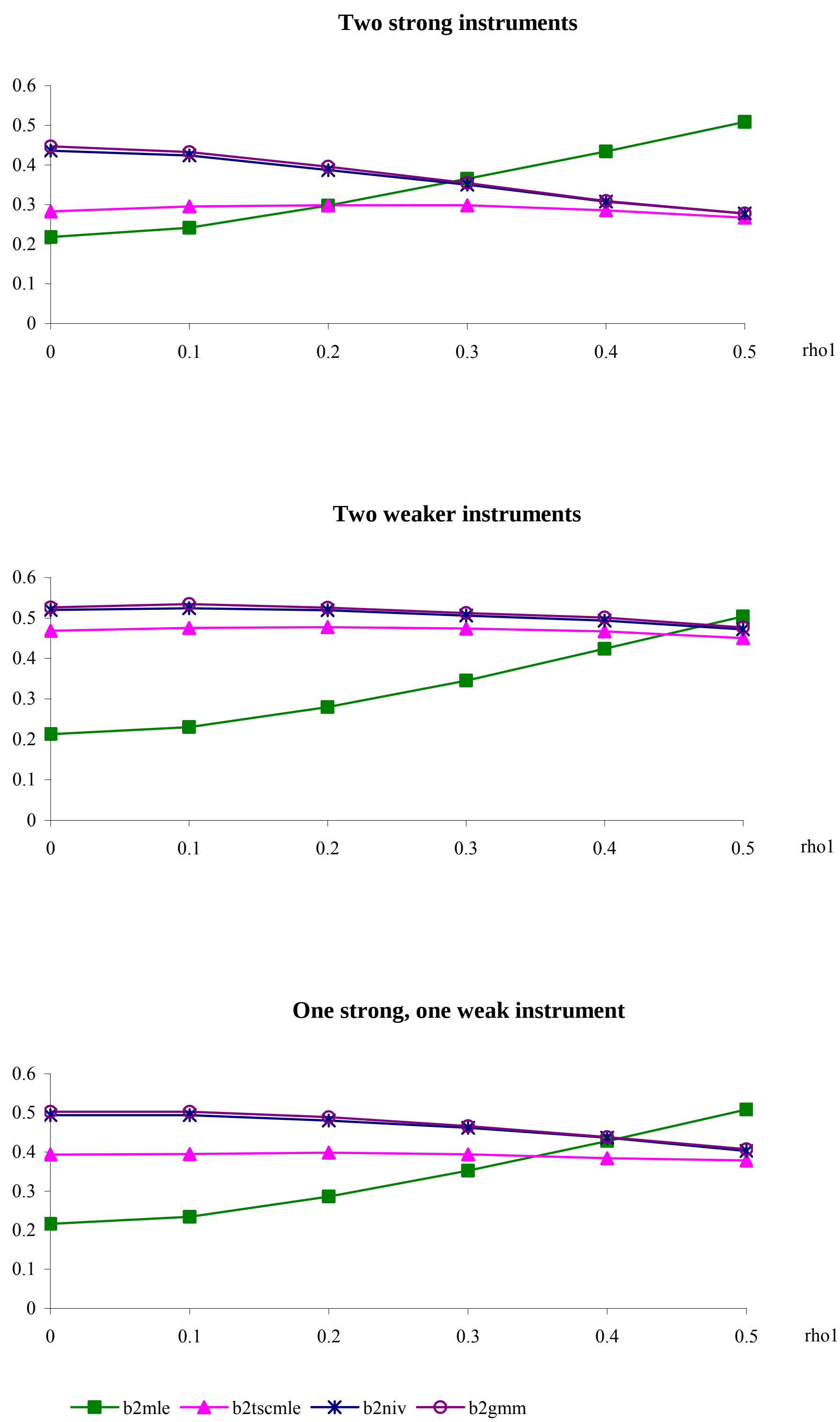


FIGURE 5.3. ESTIMATOR RMSE - COUNT DATA (n = 100)

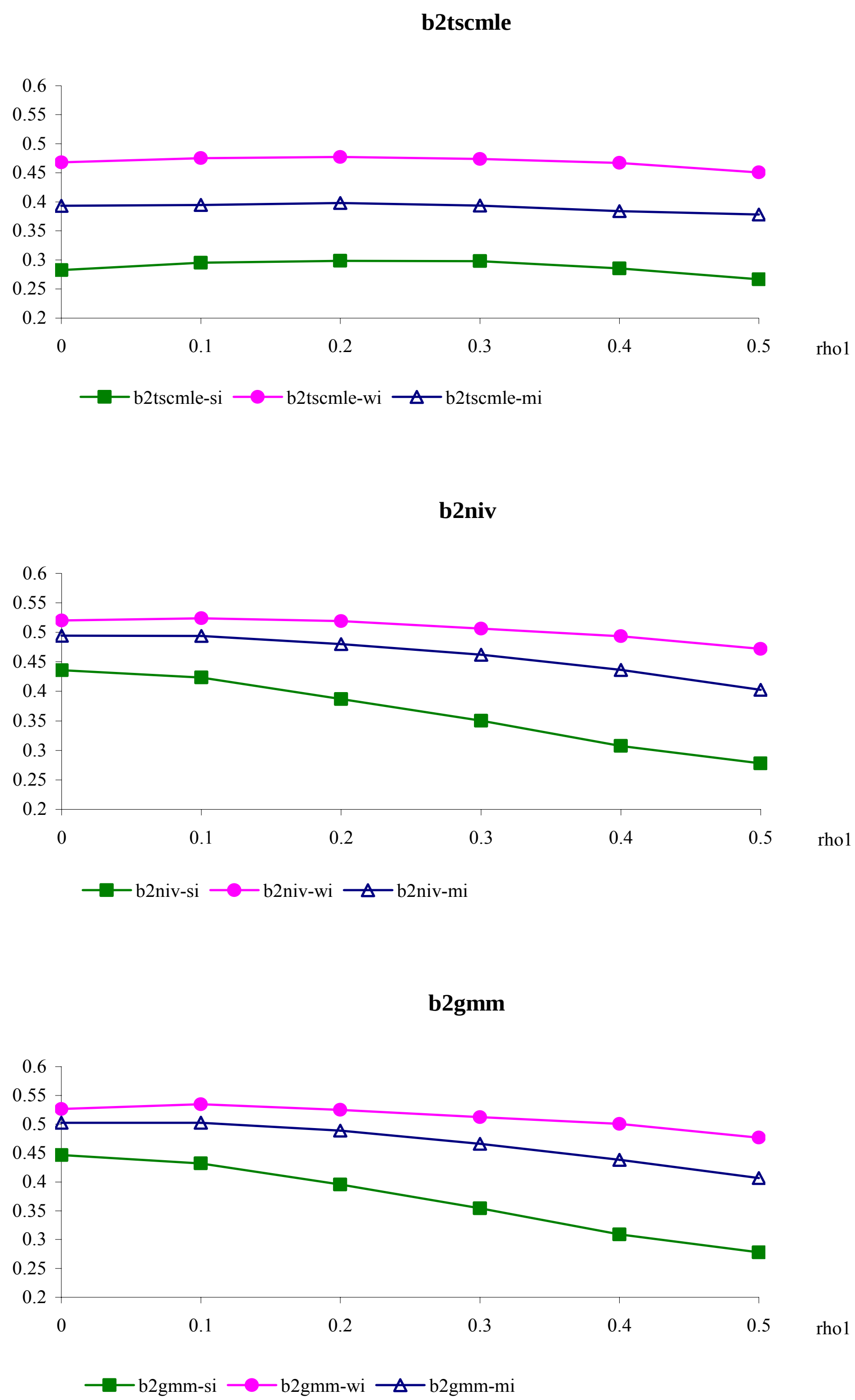


FIGURE 5.4. ESTIMATOR RMSE (COUNT DATA) - EFFECT OF INSTRUMENT STRENGTH (n =100)

*si – two strong instruments, wi – two weaker instruments, mi - one strong, one weak instrument

the degrees of freedom equal to one had the highest rejection frequency. The test **w2r** that uses the restricted GMM covariance estimator and presumes the degrees of freedom equal to the nominal number of restrictions being tested had size the closest to the correct value. After size correction, the test **w2a** based on the coefficients of the potentially endogenous regressors performed the best, but its power was still very low. Apparently, a sample size of n = 100 is too small to give a reliable results in GMM framework. The fact was also confirmed by comparing the root-mse's of the alternative estimators. The IV and the GMM estimator performed the best in a larger sample under strong endogeneity, otherwise the edge went to the two-stage conditional maximum likelihood estimator. Our results demonstrate that unless endogeneity is strong, the ML estimator has the lowest root-mse among all considered estimators. This suggests using pretest estimators. After size correction, the easy to implement Wald test (or t-test) of the Poisson regression augmented by the reduced form residuals performed the best if no strong instruments were available but its size was very unreliable. The test statistics properties could be improved by bootstrapping, specifically by wild bootstrapping since heteroskedasticity is present. Our future work will compare the tests performance under pairs and wild bootstrapping.

6. CONCLUSIONS

There are many applied economic cases where endogeneity is a potential issue and a researcher has to decide whether classical regression assumptions match the reality. The most commonly used to decide are the Hausman type tests. However, different versions of the Hausman test give sometimes contradictory conclusions. We examine characteristics (the size and power) of alternative versions of the Hausman test in models with homoskedastic and especially, heteroskedastic data. Options for carrying out a Hausman-like test in heteroskedastic cases include estimating an artificial regression and using robust standard errors, or bootstrapping the covariance matrix of the two estimators used in the contrast, or stacking moment conditions leading to two estimators and estimating them as a system. We conclude that the preferred test is based on an artificial regression, perhaps using a heteroskedasticity corrected covariance matrix estimator if heteroskedasticity is suspected. If instruments are weak, however, no test procedure is reliable. To obtain tests of closer to proper size in finite samples, we employ bootstrapping. Specifically, we compare pairs and wild bootstrapping as these alternatives may be effective when errors are heteroskedastic. Our results suggest that in large samples, wild bootstrapping is a slight improvement over asymptotics in models with weak instruments. However, the results are very sensitive to the type of bootstrapping used and whether we bootstrap the critical values or the covariance matrix of the difference of the estimators. Lastly, we consider another model where heteroskedasticity is present – the count data model. Our Monte Carlo experiment shows that the test using stacked moment conditions and the second round estimator has the best performance, but which could still be improved upon by bootstrapping. Our future work will compare the tests performance in count data models under pairs and wild bootstrapping. Another direction for further study could be evaluating pretest estimators performance in count data.

REFERENCES

Anderson T. W. and Herman Rubin (1949) “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations,” *The Annals of Mathematical Statistics*, 20(1), 46-63.

Andrews, D. W. K., and R. C. Fair (1988) “Inference in Nonlinear Econometric Models with Structural Change,” *Review of Economic Studies,* 55, 615-639.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman (2003) “Instrumental variables and GMM: Estimation and testing,” *The Stata Journal*, 3(1), 1-31.

Blomquist, Soren and Matz Dahlberg (1999) "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments," *Journal of Applied Econometrics*, 14(1), 69-88.

Bound, John, David A. Jaeger and Regina M. Baker (1995). “Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak,” *Journal of the American Statistical Association*, 90(430), 443-50.

Brown, B.W and W. K. Newey (2002) “Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference,” *Journal of Business and Economic Statistics*, 20 (4), 507-517.

Burnside, C. and M. S. Eichenbaum (1996) “Small-sample properties of GMM-based Wald tests,” *Journal of Business and Economic Statistics*, 14, 294-308.

Buse, A. (1992) "The Bias of Instrumental Variable Estimators," *Econometrica*, 60(1), 173-80.

Cameron, A. Colin and Pravin K. Trivedi (2005) *Microeconometrics: Methods and Applications*, Cambridge University Press.

Creel, Michael (2004) “Modified Hausman tests for inefficient estimators,” *Applied Economics*, 36, 2373-2376.

Dastoor, N.K. (2003) “A score-type and a Hausman-type statistic based on nonsingular generalized inverses,” *Journal of Econometrics*, 117, 313 – 330.

Davidson, Russell and Emmanuel Flachaire (2001) “The Wild Bootstrap, Tamed at Last,” *QED Working Paper*, No. 1000.

Davidson, Russell and James G. MacKinnon (1987). “Implicit Alternatives and the Local Power of Test Statistics,” *Econometrica*, 55(6),1305-1329.

Davidson, R. and J. G. MacKinnon (1990) “Specification tests based on artificial regression,” *Journal of the American Statistical Association*, 85(409), 220-227.

Davidson, Russell and James. G. MacKinnon (1993) *Estimation and Inference in Econometrics*, Oxford University Press.

Davidson, Russell and James G. MacKinnon (2006) “Bootstrap inference in a linear equation estimated by instrumental variables,” *QED Working Papers*, No. 1024.

Dufour, Jean-Marie (1997) “Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models,” *Econometrica*, 65(6), 1365-1387.

Durbin, J. (1954) “Errors in Variables,” *Review of the International Statistical Institute*, 22, 23-32.

Engle, Robert (1984) "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics, vol II*, ed. Griliches and Intrilligator (Amsterdam: North Holland), 775-826 (Chapter 13).

Flachaire, E. (1999) "A better way to bootstrap pairs," *Economics Letters,* 64, 257–262.

Flachaire, E. (2005) "Bootstrapping heteroskedastic regression models: Wild bootstrap vs pairs bootstrap," *Computational Statistics & Data Analysis*, 49(2), 361-376.

Freedman, D. A. (1981) "Bootstrapping regression models," *Annals of Statistics,* 9, 1218-1228.

Gaston, Noel and Daniel Trefler (1994) "Protection, Trade, and Wages: Evidence from U.S. Manufacturing," *Industrial and Labor Relations Review*, 47(4), 574-593.

Godfrey, L. G. and J. P. Hutton (1994) "Discriminating between Errors-in-Variables/Simultaneity and Misspecification in Linear Regression Models," *Economics Letters*, 44, 359-64.

Gourieroux, C., and A. Monfort (1989) "A general framework for testing a null hypothesis in a "mixed" form," *Econometric Theory*, 5 (1), 63–82.

Grogger, Jeffrey (1990) "A simple test for exogeneity in probit, logit and Poisson regression models," *Economics Letters*, 33, 329-332.

Hahn, J., J. Hausman, and G. Kuersteiner (2004) "Estimation with weak-instruments: accuracy of higher order bias and MSE approximations," *Econometrics Journal*, 7, 272–306.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics.

Hall, Alastair R., Glenn D. Rudebusch and David W. Wilcox (1996). "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37(2), 283-298.

Hall, Peter and Joel L. Horowitz (1996) "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, Vol. 64, No. 4, 891-916.

Hausman, J. (1978) "Specification tests in econometrics," *Econometrica*, 46(6), 1251-1271.

Hausman, J. and W. Taylor (1981) "A generalized specification test," *Economics Letters*, 8, 239-45.

Holly, A. (1982) "a Remark on Hausman's Specification Test," *Econometrica*, 50, 749–759.

Holly A. and A. Monfort (1986) "Some Useful Equivalence Properties of Hausman's Test," *Economics Letters,* 1, 39-40.

Horowitz, J.L. (1997) *Bootstrap methods in econometrics: theory and numerical performance*, in D.M. Kreps and K.F. Wallis, eds. *Advances in Economics and Econometrics: Theory and Applications*, Seventh World Congress, vol. 3, Ch. 7, Cambridge University Press.

Horowitz, Joel L. (2001) "The Bootstrap and Hypothesis Tests in Econometrics," *Journal of Econometrics*, 100, 37-40.

Horowitz, Joel L. and N. E. Savin (2000) "Empirically relevant critical values for hypothesis tests: A bootstrap approach," *Journal of Econometrics*, 95, 375-389.

Hwang (1980) "A Comparison of Tests of Overidentifying Restrictions, " *Econometrica*, 48(7), 1821-25.

Inoue, A.(2006) "A Bootstrap Approach to Moment Selection," *Econometrics Journal*, 9, 48–75.

Kariya, Takeaki and Jiro Hodoshima, (1980) "Finite Sample Properties of the Tests for Independence in Structural Systems and the LRT," *Economic Studies Quarterly*, 31(1), 45-56.

Kleibergen, F. (2002) "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70,1781–1803.

Krämer, W. and H. Sonnberger (1986) *The linear regression model under test*, Heidelberg : Physica-Verlag.

Liu, R. Y. (1988) "Bootstrap procedures under some non-i.i.d. models," *Annals of Statistics*, 16, 1696–1708.

MacKinnon, James G. (1992) "Model specification tests and artificial regressions," *Journal of Economic Literature*, 30(1), 102-146.

MacKinnon, J. G. (2002) "Bootstrap inference in econometrics," *Canadian Journal of Economics,* 35, 615–645.

MacKinnon, J. G. and H. L. White (1985) "Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 21, 53–70.

Maddala, G. S. and Jinook Jeong (1992) "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60(1), 181-83.

Mammen, E. (1993) "Bootstrap and wild bootstrap for high dimensional linear models," *Annals of Statistics*, 21, 255–285.

Moreira, M. J. (2003) "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027-1048.

Moreira, M. J., J. R. Porter, and G. A. Suarez (2004) "Bootstrap and Higher-Order Expansion Validity when Instruments may be Weak", *NBER Working Papers*, No. 302.

Mullahy, John (1997) "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *Review of Economics and Statistics*, 79, 586-593.

Nakamura, Alice and Masao Nakamura (1981). "On the Relationships Among Several Specification Error Tests Presented by Durbin, Wu, and Hausman," *Econometrica*, 49(6), 1583-1588.

Nelson, Charles R. and Richard Startz (1990a). "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator (in Notes and Comments)," *Econometrica*, 58(4), 967-976.

Nelson, Charles R. and Richard Startz (1990b). "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One," *Journal of Business*, 63(1), 125-40.

Newey, Whitney K. (1985) "Maximum Likelihood Specification Testing and Conditional Moment Tests,' *Econometrica*, 53(5), 1047-1070.

Newey, W., and D. McFadden (1994) "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle, and D. McFadden, North Holland, 2111–2245 (Chapter 36).

Rao, C. R. and S. K. Mitra (1971) *Generalized Inverse of Matrices and its Applications*, John Wiley & Sons, Inc.

Revankar, N.S. (1978) “Asymptotic relative efficiency analysis of certain tests of independence in a structural system,” *International Economic Review*, 19, 165-179.

Rothenberg, T.J. (1984) “Approximating the distributions of econometric estimators and test statistics,”. in: Griliches, Z.,Intriligator, M.D., (Eds.), Ch. 15 in *Handbook of Econometrics*, vol. II. North Holland, Amsterdam, 881–935.

Ruud, P. (1984) “Tests of specification in econometrics,” *Econometric Reviews*, 3, 211-242.

Shea, John (1997). “Instrument Relevance in Multivariate Linear Models: a Simple Measure,” *The Review of Economics and Statistics*, 79(2), 348-352.

Silvey, S. D. (1959) "The Lagrangian Multiplier Test," *Annals of Mathematical Statistics*, 30, 389-407.

Smith, R. and R. Blundell (1986) “An exogeneity test for a simultaneous equation tobit model with an application to labor supply,” *Econometrica*, 54, 679-685.

Staiger, D. and J. H. Stock (1997) “Instrumental variables regression with weak instruments,” *Econometrica*, 65, 557–586.

Vuong, Q. and D. Rivers (1988) “Limited information estimators and exogeneity tests for simultaneous probit models,” *Journal of Econometrics*, 39, 347-366.

Wang, Jiahui and Eric Zivot (1998) “Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 66(6), 1389-1404.

Wessie, J. (1999) “Seemingly unrelated estimation and the cluster-adjusted sandwich estimator,” *Stata Technical Bulletin* STB-52.

White, H. (1987) "Specification Testing in Dynamic Models," in Truman Bewley, ed., *Advances in Econometrics*, Cambridge University Press.

Windmeijer, F. a. G. and J. M. C. Santos Silva (1997) “Endogeneity in Count Data Models: An Application to Demand for Health Care,” *Journal of Applied Econometrics*, 12(3), 281-294.

Wong, K. (1996) “Bootstrapping Hausman’s exogeneity test,” *Economics Letters*, 53, 139-143.

Wooldridge, Jeffrey M. (2002) *Econometric analysis of cross section and panel data*, MIT Press.

Wu, De-Min (1973). “Alternative Tests of Independence between Stochastic Regressors and Distrurbances,” *Econometrica*, 41 (4), 733–750.

Wu, De-Min (1974) “Alternative Tests of Independence between Stochastic Regressors and Disturbances: Finite Sample Results,” *Econometrica*, 42(3), 529-546.

APPENDIX: HAUSMAN TEST DEGREES OF FREEDOM

To derive the Hausman test statistic, we start with the difference of the two estimators,

(A1.1) $\hat{q} = \tilde{\beta} - \hat{\beta} = [(Z'X)^{-1} Z' - (X'X)^{-1} X']y$

In (A1.2) to (A1.4), we show that $(Z'X) = (Z'Z) = (X'Z)$.

(A1.2)
$$Z'X = \begin{pmatrix} \hat{X}_1' \\ X_2' \end{pmatrix} \begin{pmatrix} X_1 & X_2 \end{pmatrix} = \begin{pmatrix} \hat{X}_1' X_1 & \hat{X}_1' X_2 \\ X_2' X_1 & X_2' X_2 \end{pmatrix}$$

Since $P_{\underline{Z}} = \underline{Z}(\underline{Z}'\underline{Z})^{-1}\underline{Z}'$ is idempotent,

(a.1.3) $\hat{X}_1' X_1 = X_1'\underline{Z}(\underline{Z}'\underline{Z})^{-1}\underline{Z}' X_1 = X_1'\underline{Z}(\underline{Z}'\underline{Z})^{-1}\underline{Z}'\underline{Z}(\underline{Z}'\underline{Z})^{-1}\underline{Z}' X_1 = \hat{X}_1'\hat{X}_1$

Using that $P_{\underline{Z}}$ is symmetric and idempotent and that $\underline{Z}$ includes X_2

$$X_2'\hat{X}_1 = X_2'\underline{Z}(\underline{Z}'\underline{Z})^{-1}\underline{Z}' X_1 = (\underline{Z}(\underline{Z}'\underline{Z})^{-1}\underline{Z}' X_2)' X_1 = (P_{\underline{Z}} X_2)' X_1 = X_2' X_1 .$$

Taking transposes, $\hat{X}_1' X_2 = X_1' X_2$.

Thus,

(A1.4)
$$Z'X = \begin{pmatrix} \hat{X}_1'\hat{X}_1 & \hat{X}_1' X_2 \\ X_2'\hat{X}_1 & X_2' X_2 \end{pmatrix} = \begin{pmatrix} \hat{X}_1' \\ X_2' \end{pmatrix} \begin{pmatrix} \hat{X}_1 & X_2 \end{pmatrix} = Z'Z = X'Z$$

Using the equality $(Z'X)^{-1} = (Z'Z)^{-1} = (X'Z)^{-1}$, the difference of the two estimators (A1.1) can be expressed as

(A1.5) $\hat{q} = \tilde{\beta} - \hat{\beta} = [(Z'Z)^{-1} Z' - (X'X)^{-1} X']y$

Using the inversion formula for partitioned matrices

$$(Z'Z)^{-1} = \left(\begin{array}{c|c} \left[\hat{X}_1' X_1 - X_1' X_2 (X_2' X_2)^{-1} X_2' X_1\right]^{-1} & -\left[\hat{X}_1' X_1 - X_1' X_2 (X_2' X_2)^{-1} X_2' X_1\right]^{-1} X_1' X_2 (X_2' X_2)^{-1} \\ -\left[X_1' X_2 (X_2' X_2)^{-1}\right]' \left[\hat{X}_1' X_1 - X_1' X_2 (X_2' X_2)^{-1} X_2' X_1\right]^{-1} & (X_2' X_2)^{-1} + \left[X_1' X_2 (X_2' X_2)^{-1}\right]' \left[\hat{X}_1' X_1 - X_1' X_2 (X_2' X_2)^{-1} X_2' X_1\right]^{-1} X_1' X_2 (X_2' X_2)^{-1} \end{array} \right)$$

This can be rewritten as

(A1.6) $$(Z'Z)^{-1} = \left(\begin{array}{c|c} \tilde{\Sigma} & -\tilde{\Sigma} A \\ -A'\tilde{\Sigma} & (X_2' X_2)^{-1} + A'\tilde{\Sigma} A \end{array} \right)$$

where

(A1.7) $A = X_1' X_2 (X_2' X_2)^{-1}$

(A1.8) $\tilde{\Sigma} = \left[\hat{X}_1' X_1 - X_1' X_2 (X_2' X_2)^{-1} X_2' X_1\right]^{-1}$

Similarly,

(A1.9) $$(X'X)^{-1} = \left(\begin{array}{c|c} \hat{\Sigma} & -\hat{\Sigma} A \\ -A'\hat{\Sigma} & (X_2' X_2)^{-1} + A'\hat{\Sigma} A \end{array} \right)$$

where A is the same as in (A1.7) and

(A1.10) $\hat{\Sigma} = \left[X_1{}'X_1 - X_1{}'X_2 \left(X_2{}'X_2 \right)^{-1} X_2{}'X_1 \right]^{-1}$

Thus, (A1.5) can be rewritten as

$$\begin{aligned}
\hat{q} &= [(Z'Z)^{-1} Z' - (X'X)^{-1} X']y \\
&= \left[\begin{pmatrix} \tilde{\Sigma} & -\tilde{\Sigma}A \\ -A'\tilde{\Sigma} & \left(X_2{}'X_2\right)^{-1} + A'\tilde{\Sigma}A \end{pmatrix} \left(\hat{X}_1 \mid X_2 \right)' - \begin{pmatrix} \hat{\Sigma} & -\hat{\Sigma}A \\ -A'\hat{\Sigma} & \left(X_2{}'X_2\right)^{-1} + A'\hat{\Sigma}A \end{pmatrix} \left(X_1 \mid X_2 \right)' \right] y \\
&= \left[\begin{pmatrix} \tilde{\Sigma}\hat{X}_1{}' - \tilde{\Sigma}AX_2{}' \\ -A'\tilde{\Sigma}\hat{X}_1{}' + \left(X_2{}'X_2\right)^{-1} X_2{}' + A'\tilde{\Sigma}AX_2{}' \end{pmatrix} - \begin{pmatrix} \hat{\Sigma}X_1{}' - \hat{\Sigma}AX_2{}' \\ -A'\hat{\Sigma}X_1{}' + \left(X_2{}'X_2\right)^{-1} X_2{}' + A'\hat{\Sigma}AX_2{}' \end{pmatrix} \right] y \\
&= \begin{bmatrix} \tilde{\Sigma}\hat{X}_1{}' - \tilde{\Sigma}AX_2{}' - \hat{\Sigma}X_1{}' + \hat{\Sigma}AX_2{}' \\ -A'\tilde{\Sigma}\hat{X}_1{}' + A'\tilde{\Sigma}AX_2{}' + A'\hat{\Sigma}X_1{}' - A'\hat{\Sigma}AX_2{}' \end{bmatrix} y
\end{aligned}$$

Hence, there is an exact linear relationship between the first K_1 elements of $\hat{q}$ and the remaining $(K - K_1)$ elements.

If we define $E = \tilde{\Sigma}\hat{X}_1{}' - \tilde{\Sigma}AX_2{}' - \hat{\Sigma}X_1{}' + \hat{\Sigma}AX_2{}'$ ($K_1 \times T$ matrix),

$$\hat{q} = [(Z'Z)^{-1} Z' - (X'X)^{-1} X']y = \begin{pmatrix} E \\ -A'E \end{pmatrix} y.$$

Thus, the difference of parameter coefficients can be rewritten as

(A1.11) $\hat{q} = \begin{pmatrix} Ey \\ -A'Ey \end{pmatrix} = \begin{pmatrix} \hat{q}_1 \\ -A'\hat{q}_1 \end{pmatrix}$

To construct the test, the variance of the asymptotic distribution of $\sqrt{T}\hat{q}$, $V(\hat{q})$, has to be determined. The estimators $\hat{\beta}$ and $\tilde{\beta}$ are correlated which could be a problem but Hausman (1978) presents the result that under the null hypothesis an asymptotically efficient estimator $\hat{\beta}$ must have zero asymptotic covariance with its difference from a consistent but asymptotically inefficient estimator $\tilde{\beta}$ and thus under the null hypothesis of no misspecification

(A1.12) $V(\hat{q}) = V(\tilde{\beta} - \hat{\beta}) = V(\tilde{\beta}) - V(\hat{\beta})$

The variance of the asymptotic distribution of $\sqrt{T}\hat{q}$

$$V(\hat{q}) = V(\tilde{\beta}) - V(\hat{\beta}) = \sigma^2 (T^{-1}Z'X)^{-1}(T^{-1}Z'Z)(T^{-1}X'Z)^{-1} - \sigma^2\left(T^{-1}X'X\right)^{-1}$$

is consistently estimated by

$$\begin{aligned}\hat{V}(\hat{q}) &= \tilde{\sigma}^2 (T^{-1}Z'X)^{-1}(T^{-1}Z'Z)(T^{-1}X'Z)^{-1} - \hat{\sigma}^2\left(T^{-1}X'X\right)^{-1}\\ &= T[\tilde{\sigma}^2 (Z'X)^{-1}(Z'Z)(X'Z)^{-1} - \hat{\sigma}^2\left(X'X\right)^{-1}]\\ &= T\left(\tilde{\sigma}^2\left(Z'Z\right)^{-1} - \hat{\sigma}^2\left(X'X\right)^{-1}\right)\end{aligned} \tag{A1.13}$$

Plugging in (A1.6) and (A1.9)

$$\begin{aligned}\hat{V}(\hat{q}) &= T\left(\tilde{\sigma}^2\left(Z'Z\right)^{-1} - \hat{\sigma}^2\left(X'X\right)^{-1}\right)\\ &= T\left[\tilde{\sigma}^2\begin{pmatrix}\tilde{\Sigma} & -\tilde{\Sigma}A\\ -A'\tilde{\Sigma} & \left(X_2{}'X_2\right)^{-1} + A'\tilde{\Sigma}A\end{pmatrix} - \hat{\sigma}^2\begin{pmatrix}\hat{\Sigma} & -\hat{\Sigma}A\\ -A'\hat{\Sigma} & \left(X_2{}'X_2\right)^{-1} + A'\hat{\Sigma}A\end{pmatrix}\right]\\ &= T\begin{pmatrix}\tilde{\sigma}^2\tilde{\Sigma} - \hat{\sigma}^2\hat{\Sigma} & -\left(\tilde{\sigma}^2\tilde{\Sigma} - \hat{\sigma}^2\hat{\Sigma}\right)A\\ -A'\left(\tilde{\sigma}^2\tilde{\Sigma} - \hat{\sigma}^2\hat{\Sigma}\right) & (\tilde{\sigma}^2 - \hat{\sigma}^2)\left(X_2{}'X_2\right)^{-1} + A'\left(\tilde{\sigma}^2\tilde{\Sigma} - \hat{\sigma}^2\hat{\Sigma}\right)A\end{pmatrix}\end{aligned}$$

where $A, \tilde{\Sigma}$ and $\hat{\Sigma}$ are as defined in (A1.7), (A1.8) and (A1.10) respectively.

Let

$$\hat{\Omega} = \tilde{\sigma}^2\tilde{\Sigma} - \hat{\sigma}^2\hat{\Sigma} \tag{A1.14}$$

$$C = (\tilde{\sigma}^2 - \hat{\sigma}^2)\left(X_2'X_2\right)^{-1} \tag{A1.15}$$

Then,

$$\hat{V}(\hat{q}) = T\begin{pmatrix}\hat{\Omega} & -\hat{\Omega}A\\ -A'\hat{\Omega} & C + A'\hat{\Omega}A\end{pmatrix} \tag{A1.16}$$

Using the inversion formula for partitioned matrices again (assuming that C^{-1} exists), we get that:

$$\hat{V}(\hat{q})^{-1} = T^{-1}\left(\begin{array}{c|c} \hat{\Omega}^{-1} + \hat{\Omega}^{-1}\hat{\Omega}A\left(C + A'\hat{\Omega}A - A'\hat{\Omega}\hat{\Omega}^{-1}\hat{\Omega}A\right)^{-1} A'\hat{\Omega}\hat{\Omega}^{-1} & \hat{\Omega}^{-1}\hat{\Omega}A\left(C + A'\hat{\Omega}A - A'\hat{\Omega}\hat{\Omega}^{-1}\hat{\Omega}A\right)^{-1} \\ \left(C + A'\hat{\Omega}A - A'\hat{\Omega}\hat{\Omega}^{-1}\hat{\Omega}A\right)^{-1} A'\hat{\Omega}\hat{\Omega}^{-1} & \left(C + A'\hat{\Omega}A - A'\hat{\Omega}\hat{\Omega}^{-1}\hat{\Omega}A\right)^{-1} \end{array}\right)$$

$$= T^{-1}\left(\begin{array}{c|c} \hat{\Omega}^{-1} + AC^{-1}A' & AC^{-1} \\ C^{-1}A' & C^{-1} \end{array}\right)$$

It follows that the Hausman statistic $H = T\hat{q}'[\hat{V}(\hat{q})]^{-1}\hat{q}$ can be reexpressed (using (A1.11)) as

$$H = T\left(\hat{q}_1' \;\vdots\; -\hat{q}_1'A\right)\left[T^{-1}\left(\begin{array}{c|c} \hat{\Omega}^{-1} + AC^{-1}A' & AC^{-1} \\ C^{-1}A' & C^{-1} \end{array}\right)\right]\begin{pmatrix} \hat{q}_1 \\ -A'\hat{q}_1 \end{pmatrix}$$

$$= \left(\hat{q}_1'\hat{\Omega}^{-1} + \hat{q}_1'AC^{-1}A' - \hat{q}_1'AC^{-1}A' \;\vdots\; \hat{q}_1'AC^{-1} - \hat{q}_1'AC^{-1}\right)\begin{pmatrix} \hat{q}_1 \\ -A'\hat{q}_1 \end{pmatrix}$$

$$= \left(\hat{q}_1'\hat{\Omega}^{-1} \;\vdots\; 0\right)\begin{pmatrix} \hat{q}_1 \\ -A'\hat{q}_1 \end{pmatrix}$$

$$= \hat{q}_1'\hat{\Omega}^{-1}\hat{q}_1$$

$$= (\tilde{\beta}_1 - \hat{\beta}_1)'\hat{\Omega}^{-1}(\tilde{\beta}_1 - \hat{\beta}_1)$$

That is, the test statistic can be written as a quadratic form in the first K_1 elements of $\hat{q}$:

(A1.17) $\quad H = T(\tilde{\beta}_1 - \hat{\beta}_1)'(T\hat{\Omega})^{-1}(\tilde{\beta}_1 - \hat{\beta}_1)$

where $K_1 \times K_1$ matrix $\hat{\Omega}$ is as defined in (A1.14). It means that asymptotically the test statistic has a chi-square distribution, with degrees of freedom equal to the number of potentially endogenous regressors at most.

VITA

Viera Chmelarova was born in Trstena, Czechoslovakia - currently Slovakia. In 1995, she completed her studies in mathematics and physics at Comenius University in Bratislava, Slovakia, and received a Magister (Master of Science) degree. In 1999, she received a Certificate in Applied Economics from Academia Instropolitana Nova in Svaty Jur, Slovakia, where she also worked as a research associate for a year. In the fall semester of 2000 she became a graduate student at the Economics Department of Louisiana State University. She received her Master of Science degree in economics from Louisiana State University in 2002. Viera taught undergraduate economics classes at Louisiana State University for two years. She will complete the degree of Doctor of Philosophy in May 2007.