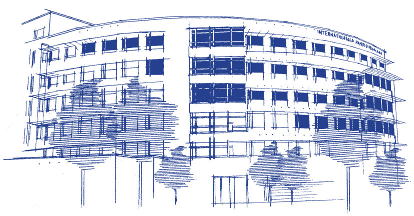
Robust Critical Values for the Jarque-bera Test for Normality

PANAGIOTIS MANTALOS

JÖNKÖPING INTERNATIONAL BUSINESS SCHOOL
JÖNKÖPING UNIVERSITY



Jönköping International Business School
Jönköping University
JIBS Working Papers No. 2010-8

ROBUST CRITICAL VALUES FOR THE JARQUE-BERA TEST FOR NORMALITY

PANAGIOTIS MANTALOS[1]

Department of Economics and Statistics
University of Jönköping

Sweden

ABSTRACT

We introduce the "sample" technique to generate robust critical values for the Jarque and Bera (JB) Lagrangian Multiplier (LM) test for normality, ***JBCV(***k_1, k_2***)***, by using improved critical values the true size of the test approaches its nominal value. Monte Carlo methods are used to study the size, and the power of the JB normality test with the "sample" critical values and compare with three alternatives to the Jarque and Bera LM test for normality: the Urzúa (1996) modification of the Jarque-Bera test, ***JBM***; the Omnibus K^2 statistic made by D'Agostino, Belanger and D'Agostino (1990), ***JBK***; and finally the, Jarque and Bera LM test for normality by using the quantities k_1 and k_2 are the definitions of sample skewness and kurtosis ***JB(***k_1, k_2***)***. The ***JBCV(***k_1, k_2***),*** shows superiority as it has the right size for all samples, small, medium and large, and at the same time has the higher power.

Keywords: Jarque and Bera LM test; Kurtosis; Omnibus K^2; Skewness; Test for normality.

[1] Contact: Panagiotis Mantalos. Sweden. E-mail: Mantalos03@yahoo.se

1. INTRODUCTION

In univariate data analysis, one of the most widely used assumptions is the assumption of "normality". Furthermore, the commonly assumed "mormality", helps us to estimate and make inferential comparisons and judgments.

However, violation of this assumption might produce misleading inferences and the result of using unreliable inferences is to produce misleading interpretations.

Testing for normality should be at least as important a step, or perhaps more, than the assumption for normality.

The most widely method, at least in econometrics, that has been suggested and used for testing whether the distribution underlying a sample is normal is the Bowman and Shenton (1975) statistic:

$$JB = n\left[\frac{skewness^2}{6} + \frac{(kurtosis - 3)^2}{24}\right] \tag{1.1}$$

which subsequently was derived by Bera and Jarque as the Lagrangian Multiplier (LM) test against the Pearson family distributions. For that reason, the *JB* test is also referred to as the Jarque-Bera test (Bowman and Shenton, 1975; Shenton and Bowman, 1977; Bera and Jarque, 1982; Jarque and Bera, 1987).

The *JB* statistic has an asymptotic chi-square distribution with two degrees of freedom.

Mantalos (2010) in a Monte Carlo study showed by using three different definitions (estimates) of the sample skewness and kurtosis, that the *JB* has rather poor small sample properties, the slow convergence of the test statistic to its limiting distribution, makes the test over-sized for small nominal level and under-sized for larger than 3% levels even in a reasonably large sample. Even the power of the tests shows the same erratic form.

A reason for this is that skewness and kurtosis are not independently distributed, and the sample kurtosis especially approaches normality very slowly. That is, the slow convergence of the test statistic to its limiting distribution, which makes the test behave erratically over under-sized even in a reasonably large sample.

However the *JB* test is simple to compute and its power has proved comparable to other powerful tests.

Urzúa (1996), D'Agostino, Belanger and D'Agostino (1990) and Doornik and Hansen (1994) are a few studies, as we see in the next chapter, that try to correct that problem by employing a small sample correction.

In recent years one of the new ways of dealing with, and solving this problem, has been to use the bootstrap technique.

By bootstrapping under the null hypothesis we approximate the distribution of the test statistic, thereby generating more robust critical values for our test statistic.

However, the issue of the bootstrap test, even if it is well applied, is not trivial. One of the basic problems in bootstrap testing is that one needs to resample the data, under the null hypothesis.

In our study the null hypothesis is that the data follows normal distribution and based on that we present one simple and easy way to apply "sample" under the null hypothesis.

By sampling under the null hypothesis we approximate the distribution of the test statistic, thereby generating more robust critical values for our test statistic.

The rest of the paper is organized as follows: Section 2 presents the skewness, kurtosis and Jarque and Bera test, while in section 3 we present our "sample" methodology. Section 4 presents the design of our Monte Carlo experiment. In Section 5 we describe the results concerning the size of the test, while power is analysed in Section 6. Finally, a brief summary and conclusions are presented in Section 7.

2. SKEWNESS, KURTOSIS AND JARQUE-BERA TEST

Let $\{x_i\}$ denote a sample of *n* observations, and let μ, σ_x^2 denote the mean and variance of $\{x_i\}$, and write $\mu_j = E\left[x_i - \mu\right]^j$, so that $\sigma_x^2 = \mu_2$. The skewness γ_1 and kurtosis γ_2 are defined as:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}, \gamma_2 = \frac{\mu_4}{\mu_2^2}.$$

Then the sample skewness and kurtosis are defined as:

$$skew = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x-\bar{x}\right)^3}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(x-\bar{x}\right)^2\right]^{3/2}} \tag{2.1}$$

$$kurt = \frac{m_4}{m_2^{2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x-\bar{x}\right)^4}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(x-\bar{x}\right)^2\right]^{2}} \tag{2.2}$$

These quantities are consistent estimates of the theoretical skewness γ_1 and kurtosis γ_2 of the distribution. Moreover, if the sample indeed comes from a normal population, then their exact finite sample distribution can also be calculated. Pearson (1931), by using the $g_1 = skew = \frac{m_3}{m_2^{3/2}}$ and $g_2 = kurt - 3 = \frac{m_4}{m_2^2} - 3$, showed that:

$$\mu_1(g_1) = 0 \tag{2.3}$$

$$\mu_2(g_1) = \frac{6\left(n-2\right)}{\left(n+1\right)\left(n+3\right)} \tag{2.4}$$

and

$$\mu_1(g_2) = -\frac{6}{n+1} \tag{2.5}$$

$$\mu_2(g_2) = \frac{24\left(n-2\right)\left(n-3\right)}{\left(n+1\right)^2\left(n+3\right)\left(n+5\right)} \tag{2.6}$$

The g_1 and g_2 are both asymptotically normal. Based on that and that the normal distribution will have skewness = 0 and kurtosis =3, Bowman and Shenton (1975) consider the follow test statistic based on equation (1.1) which subsequently was derived as an LM test by Jarque and Bera (1987):

$$JB = n\left[\frac{\left(g_1\right)^2}{6} + \frac{\left(g_2\right)^2}{24}\right] \tag{2.7}$$

JB is asymptotically chi-squared distributed with two degrees of freedom because *JB* is just the sum of the squares of two asymptotically independent standardized normal variables.

Based on Cramér (1946), and to remove the bias in g_2 and to achieve consistency at the same time Mantalos (2010) found that by using the follow estimates:

$$k_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1 \tag{2.8}$$

$$k_2 = \frac{n-1}{(n-2)(n-3)} \{(n+1) g_2 + 6\} \ , \tag{2.9}$$

we achieve better size with higher power. Note also that the quantities k_1 and k_2 are the definitions of sample skewness and kurtosis adopted by the computing packages SAS and SPSS, and also by the Excel spreadsheet programme (see Joanest and Gill, 1998).

Further, by using the (k_1, k_2) we have the following *JB statistic version*:

$$JB(k_1, k_2) = n\left[\frac{(k_1)^2}{6} + \frac{(k_2)^2}{24}\right] \tag{2.10}$$

However the rate of their convergence to the distribution limit is slow, especially for kurtosis. In order to solve this problem different solutions have been suggested.

Urzúa (1996) introduced a modification of the Jarque-Bera test by standardizing the skewness and kurtosis in the equation of *JB* (2.7), that is, by using the mean and variance for the skewness, (2.3), (2.4) and for the kurtosis (2.5),(2.6), appropriately in the following way:

$$JBM = n\left[\frac{(g_1)^2}{\mu_2(g_1)} + \frac{(g_2 - \mu_1(g_2))^2}{\mu_2(g_2)}\right] \tag{2.11}$$

Another modification is the Omnibus K2 statistic made by D'Agostino, Belanger and D'Agostino (1990). A variation of this Omnibus K2 is used by Doornik and Hansen (1994), who employ a small sample correction, and adapt the test for the multivariate case. Note, that this test statistic is used and reported by computing packages PcGive.

They suggested a transformation to the sample skewness g_1 and kurtosis g_2 in a way that makes their distribution as close to standard normal as possible.

In particular the Omnibus K^2 statistic suggested the following transformation for the sample skewness, a transformation that was derived by D'Agostino (1970):

$$z_1(g_1) = \delta . \ln\left(\frac{g_1}{\alpha\sqrt{\mu_2(g_1)}} + \sqrt{\frac{g_1^2}{\alpha^2\mu_2(g_1)} + 1}\right), \tag{2.12}$$

Where α and δ are calculated as

$$\delta = 1\big/\sqrt{\ln W}, \tag{2.13}$$

$$\alpha^2 = 2\big/\left(W^2 - 1\right), \tag{2.14}$$

with $W^2 = \left(\sqrt{2\gamma_2(g_1) + 4}\right) - 1$ (2.15)

Similarly for the sample kurtosis suggested the following transformation is made by Ancombe and Glynn (1983):

$$z_2(g_2) = \left(\sqrt{\frac{9A}{2}}\right)\left[1 - \frac{2}{9A} - \left(\frac{1 - 2/A}{1 + \frac{g_2 - \mu_1(g_2)}{\sqrt{\mu_2(g_2)}}\sqrt{2/(A-4)}}\right)\right]^{1/3}, \tag{2.16}$$

where $A = 6 + \frac{8}{\gamma_1(g_2)}\left(\frac{2}{\gamma_1(g_2)} + \sqrt{1 + \left[4\big/\left(\gamma_1(g_2)\right)^2\right]}\right)$ (2.17)

The statistics $z_1(g_1)$, (2.12) and $z_2(g_2)$, (2.16) produce the Omnibus K^2 statistic:

$$JBK = \left[z_1(g_1)\right]^2 + \left[z_2(g_2)\right]^2 \tag{2.18}$$

If the null hypothesis of normality is true it is asymptotically chi-squared distributed with two degrees of freedom because *JBK* is just the sum of the squares of two asymptotically independent standardized normal variables.

3. Computer intensive critical values

In *JB*, *JB*(k_1, k_2), *JBM* and *JBK* cases, however, the distributions of the test statistic we use are known only asymptotically and, unfortunately, unless the sample size is very large, the tests may not have the correct size. Inferential comparisons and judgements based on them might be misleading.

As mentioned earlier, one of the new ways to deal with this situation, and solve this problem, has been to use the bootstrap test.

By bootstrapping under the null hypothesis we approximate the distribution of the test statistic, thereby generating more robust critical values for our test statistic.

However, here we present one simple and easy way to test for normality by using only the JB statistic. Furthermore, by using the “sample” technique we generate robust critical values for our test statistic, so by using the improved critical values the true size of the test approaches its nominal value.

In our case it does not matter whether or not we know the nature of the theoretical distribution of the test statistic. What matters is that the technique well approximates these distributions.

The basic principle of generating critical values is to draw a number of “samples” from the model under the null hypothesis. In our case with the restriction that skewness be equal to zero, and kurtosis be equal to three, we use the computer to generate from the standard normal distribution samples with the same number of observations (n) as our data.

Then the procedure for calculating the critical values is given by the following steps:

a) We estimate the *JB*(k_1, k_2) test statistic as we have described in Section 2, (2.10).

b) We generate a sample of n i.i.d N(0,1) observations.

c) We then calculate the test statistic *JB**(k_1, k_2) as we have described in Section 2, (2.10) i.e., by calculating the sample skewness and kurtosis and then applying the Jarque and Bera test procedure by using the k_1, k_2 to the n i.i.d N(0,1) observations.

d) Repeating this step N_b times and taking the (1-α):*th* quintile of the distribution of *JB**, we obtain the α - level "sample critical values" (c_{α}^{*}).

e) We then reject Ho if $JB \geq c_{\alpha}^{*}$.

f) Finally an estimate of the P-value for testing is P*{ $JB^{*}\ k_1, k_2\ \geq JB\ k_1, k_2$ }.

The number of repetitions N_b that we use is 1000 but with today's computer power can easily be larger without noticing any time delay in the results.

Note we use the (2.10) instead of (2.7) for two reasons, the first to compare our results with the results of Mantalos (2010) but also because in (2.10) definitions of sample skewness and kurtosis adopted by many computing packages such as SAS and SPSS, and Excel are used. Moreover, the results by using (2.7) are almost identical to (2.10).

4. Monte Carlo Experiments

In this section we provide the characteristics of the Monte Carlo experiment undertaken. We calculate the *estimated size* by simply observing how many times the correct null hypothesis is rejected in repeated samples. By varying factors such as the number of observations 25, 50 (small sample) 75, 100 (medium sample) and 200, 500 (large sample); we obtain a succession of estimated percentages of the correct selection model under different conditions.

The Monte Carlo experiment has been performed by generating data according to the following data generating processes:

Model 0: is a sequence { x_i } of uncorrelated $N(0,1)$ random variables.

This model is used to estimate the size of the test while for the power we use the generalised lambda distribution suggested by Ramberg and Schmeiser (1974), that is an extension of Tukey's lambda distribution.

The inverse distribution functions formula is

$$F^{-1}(u) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2} \qquad (4.1)$$

Here lambda 1 is a location parameter, lambda 2 is a scale parameter and lambda 3 and lambda 4 jointly determine the shape of the distribution. In this way we are able to study the *JB* test under different shapes. Table 1 summarizes the different models with the different lambdas.

The number of replications per model used is 10,000. The calculations were performed using GAUSS 8.

Table 1: Different models for the estimated power of the *JB* test

Model	λ_1	λ_2	λ_3	λ_4
1: t-distrubution with df(14)	0.00000	0.05122	0.05122	0.078945
2: skwe=0.00 kurt=3.5	0.00000	0.06222	0.06222	0.10180
3: skwe=0.30 kurt=3.0	-0.36180	0.09255	0.18590	0.19910
4: skwe=-0.30 kurt=3.0	0.36180	0.18590	0.092550	0.19910
5: skwe=0.20 kurt=3.3	-0.16870	0.07651	0.10490	0.14160
6: skwe=-0.20 kurt=3.3	0.16870	0.10490	0.07651	0.14160

5. Analysis of the Size of the Tests

In this section we present the results of our Monte Carlo experiment concerning the size of the bootstrap tests. Simple graphical methods are used, methods developed and illustrated by Davidson and MacKinnon (1998) which are easy to interpret. The "P-*value plot*" is used to study the size, and the "*Size-Power curves*" to study the power of the tests. The graphs, the P-value plots and Size-Power curves are based on the empirical distribution function, EDF; the EDF of the P-values, denoted as $\hat{F}\ x_j$. For the P-value plots, if the distribution used to compute the p_s terms is correct, each of the p_s terms should be distributed uniformly on (0,1). Therefore the resulting graph should be close to the 45^o line. The P-value plots also make it possible and easy to distinguish between tests that systematically over-reject or under-reject, and tests that reject the null hypothesis about the right proportion of the time.

Figure 1 shows the truncated (up to 20% nominal level) P-value plots for the actual size of the *JB Tests,* for the small sample. Unfortunately the asymptotic *JB-Tests,* ***(JBK** = dash*, ***JBM**= point,* ***JB(** k_1, k_2 ***)**= point-dash*) show rather poor small sample properties, the tests over-sized for small up to 5% nominal level and under-sized for the rest of the levels, for 25 observations even in the larger sample of 50. Note that the best of those three tests is the Omnibus K^2 statistic made by D'Agostino, Belanger and D'Agostino (1990), *JBK.*

From the other side the ***JBCV(** k_1, k_2 **)*** (line in figures), as we see, Figure 1, tends to reject as much as the nominal size, in both small samples 25 and 50 observations. That is, the P-values lie between the confidence interval close to the 45^o line. That is, the ***JBCV(** k_1, k_2 **)*** which is the JB test with the use of the quantities k_1 and k_2 as estimations of sample skewness and kurtosis and with the "sample" critical values behaves very well.

In Figure 2, results are presented for the medium sample size. As noted previously, for the small samples the asymptotic *JB-Tests* over-sized for small up to 5% nominal level and under-sized for the rest of levels. Again the *JBK* is the best of the three tests and even now over-sized for small up to 5% nominal level but is lie on the line of the down limited of 95% confidence interval.

The ***JBCV(** k_1, k_2 **)*** behaves again well, the P-values lie between the confidence interval close to the 45^o line.

In large samples (Figure 3) we expected that all ***JB***-Tests should behave well however both ***JBM***, ***JB(** k_1, k_2 **)*** behave as before with small and medium samples, but now are near to the confidence interval. The *JBK* behaves well for more than 500 (Figure 3b) the P-values lie between the confidence interval close to the 45^o line.

Finally even here the ***JBCV(** k_1, k_2 **)*** has the right size the P-values lie between the confidence interval close to the 45^o line.

Figure 1 :Small sample P-value plots: Size of the Tests

Figure 1a: 25 observations

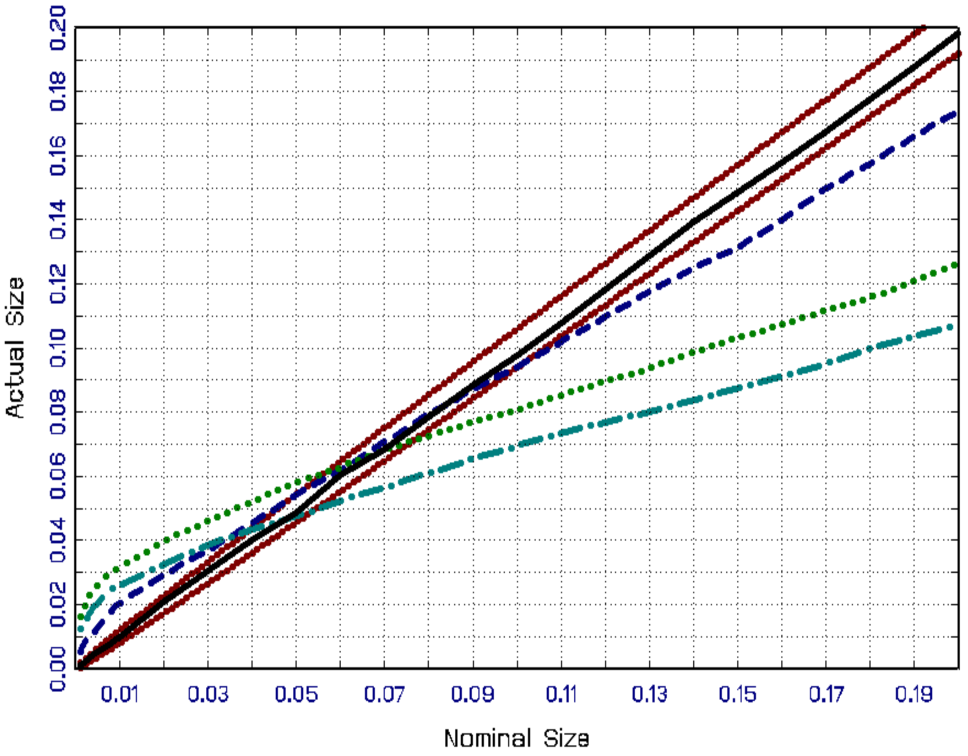
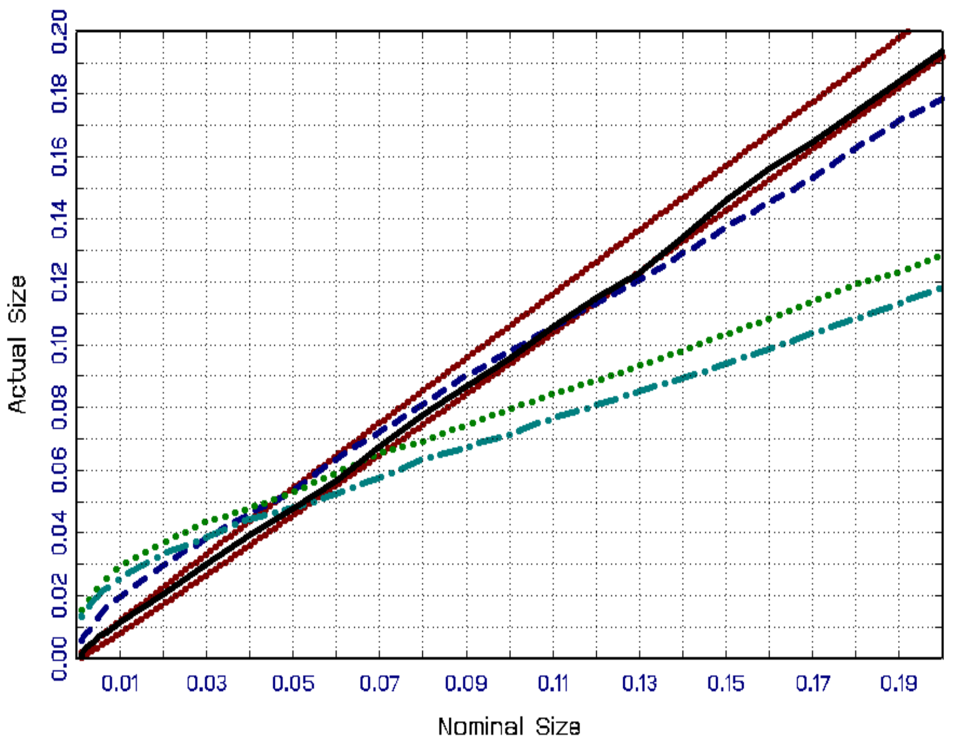


Figure 1b: 50 observations



JBCV(k_1, k_2***)*** *= line,* ***JBK =*** *dash,* ***JBM=*** *point,* ***JB(*** k_1, k_2***)=*** *point-dash,****95% confidence interval*** *= close-point*

Figure 2 :Medium sample P-value plots: Size of the Tests

Figure 2a: 75 observations

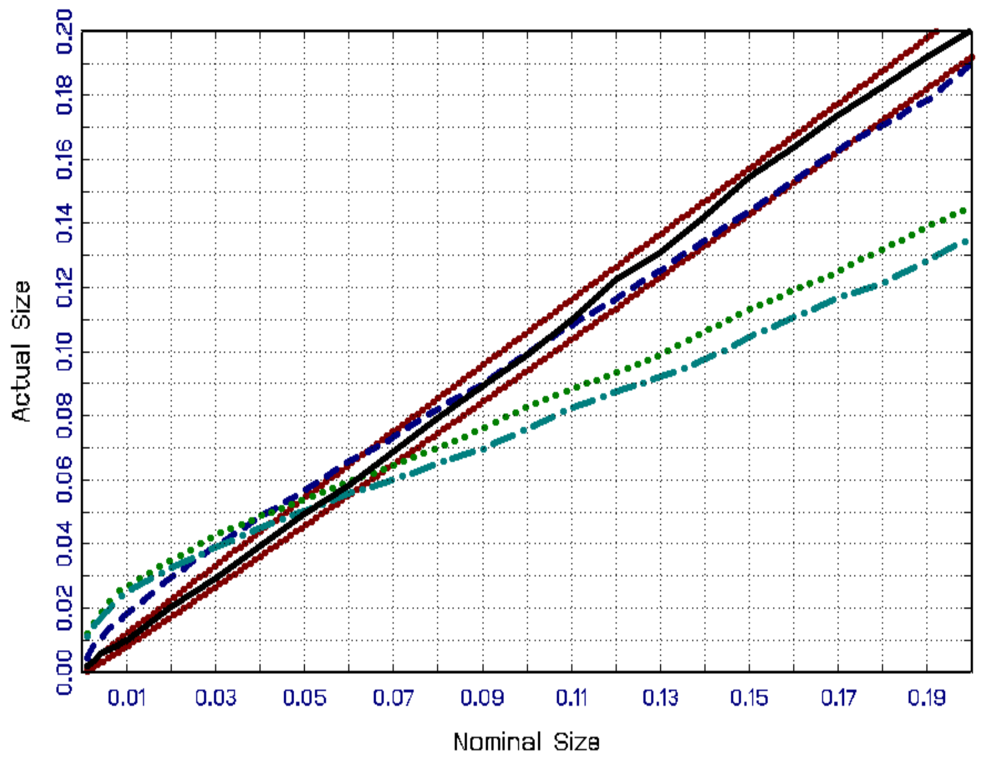
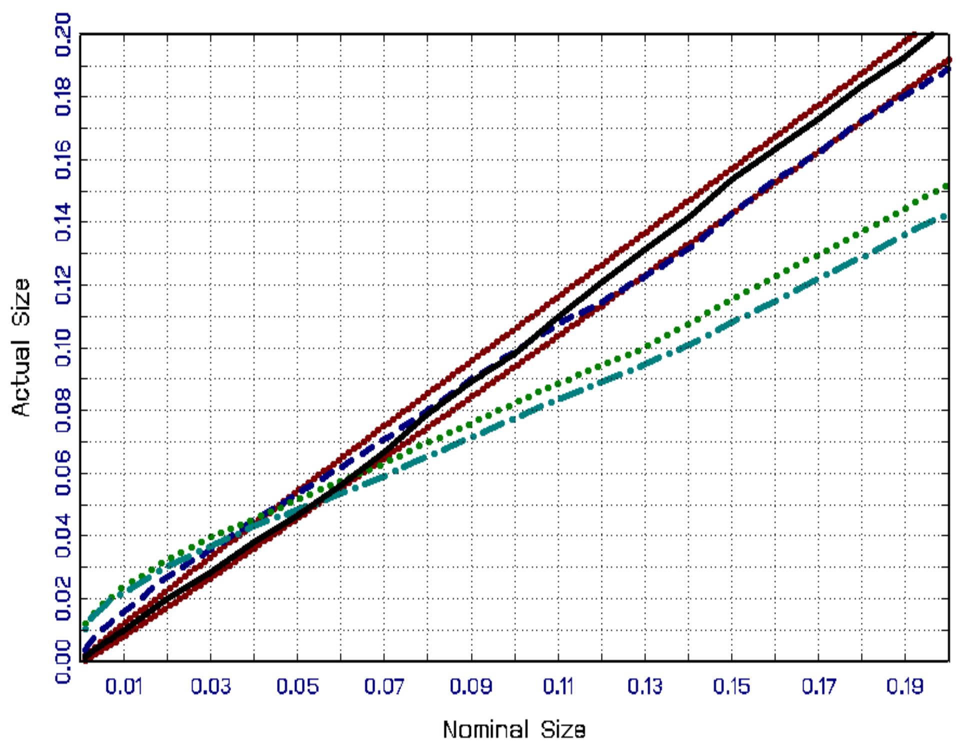


Figure 2b: 100 observations



JBK **=** dash, ***JBM*****=** *point*, ***JBCV(*** k_1, k_2 ***)*** *= line*, ***JB(*** k_1, k_2 ***)=*** *point-dash*,***95% confidence interval*** *= close-point*

Figure 3 :Large sample P-value plots: Size of the Tests

Figure 3a: 200 observations

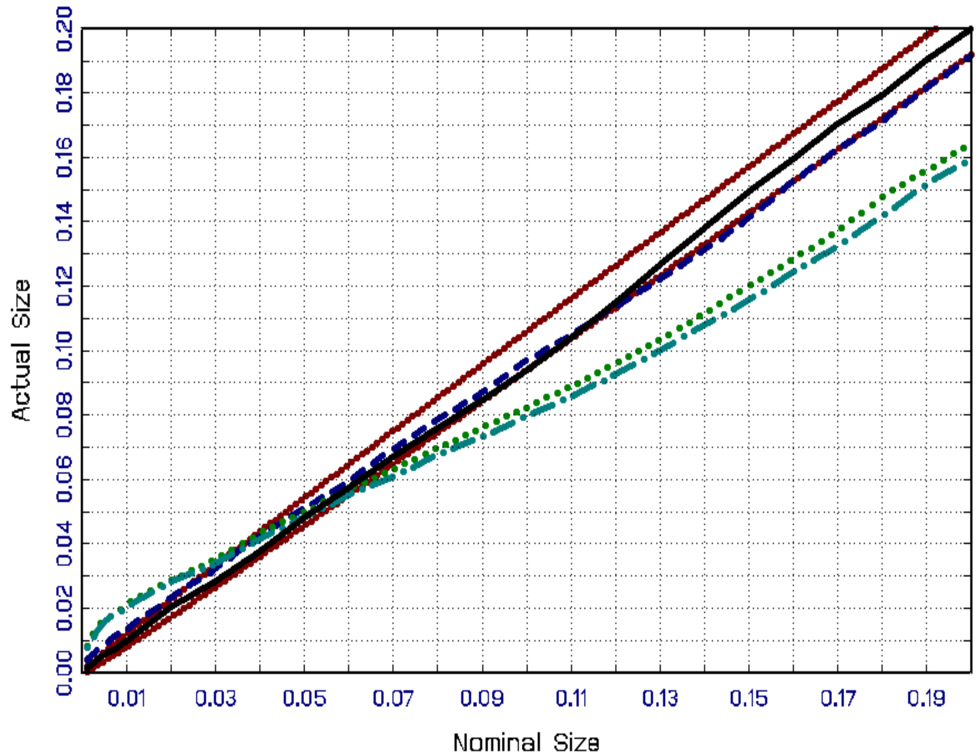
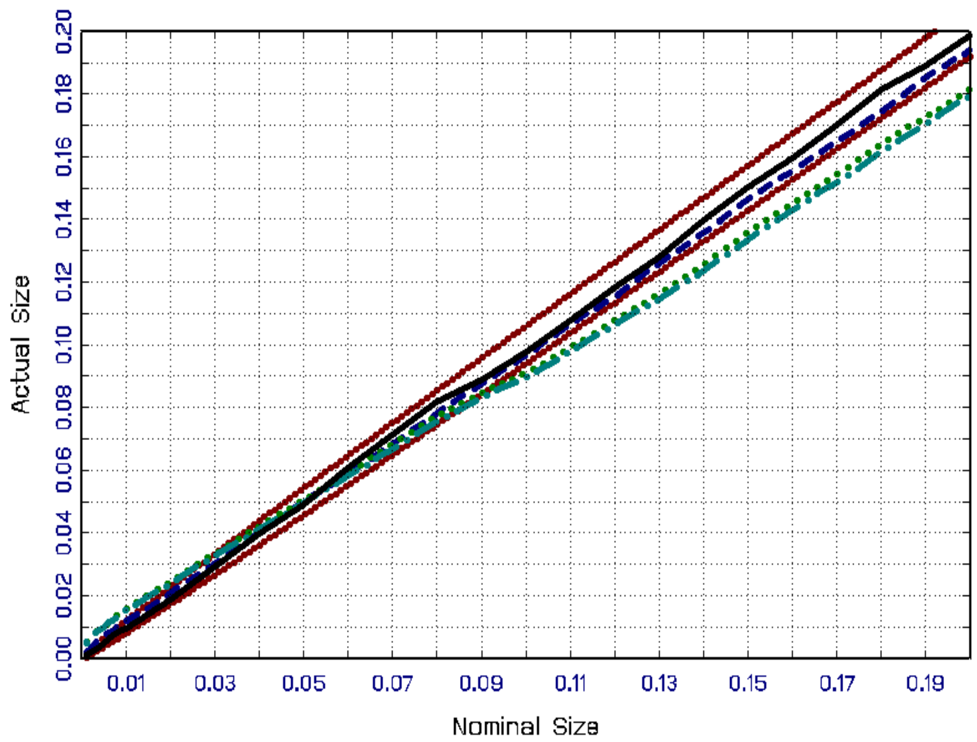


Figure 3b: 500 observations



JBK = dash, ***JBM***= *point*, ***JBCV***(k_1, k_2) = *line*, ***JB***(k_1, k_2)= *point-dash*,***95% confidence interval*** = *close-point*

To summarize our Monte Carlo experiment about the analysis of the size for the JB-Tests: we found that the rate of their convergence to the distribution limit is slow for the asymptotic *JB-Tests*, ***(JBK** = dash*, ***JBM**= point*, ***JB(***k_1, k_2***)**= point-dash*). We find rather poor small sample properties, that is, the tests over-sized for small up to 5% nominal level and under-sized for the rest of the levels, even for 200 observations. While the ***JBCV(***k_1, k_2***)***, the JB test with "sample" critical values, is robust and has the right size, for all samples, that is, the P-values lie between the confidence interval close to the 45^o line.

6. Analysis of the Power of the Tests

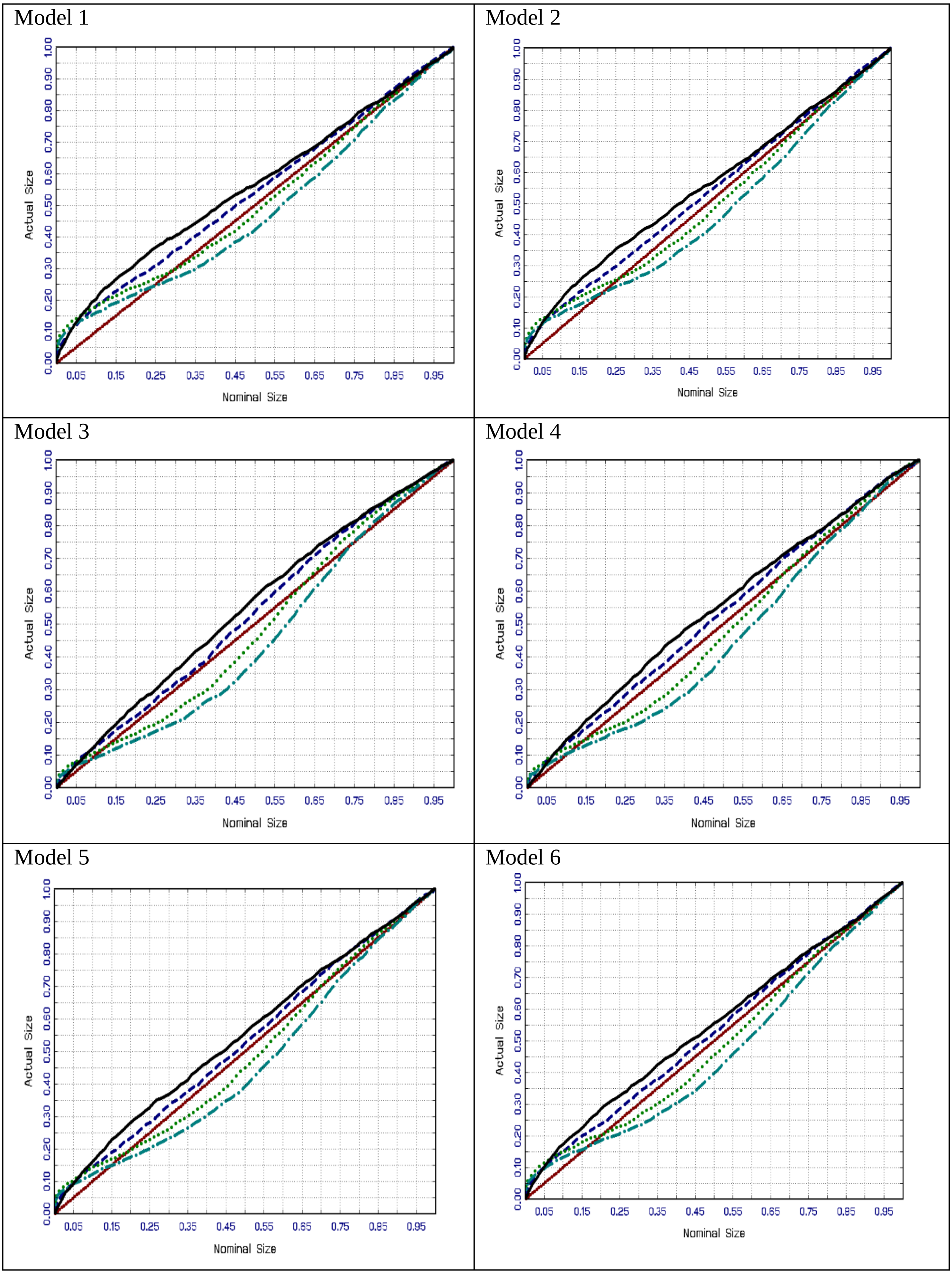
In this section, we analyse the power of the the *JB-tests* using small, medium and large sample sizes of observations. The power function is estimated by calculating the rejection frequencies in 1,000 replications using the different models of Table1.

We used the Size-Power curves to compare the estimated power functions of the alternative test statistics. This proved to be quite adequate, because those tests that gave reasonable results regarding size usually differed very little regarding power. Note in what follows figures with a *solid* curve are the estimated power of the ***JB(***g_1, g_2***)*** test as it was when we analysed the size of the tests. While those with a *dash* curve have the ***JB(***b_1, b_2***)***. Finally those with a *dot-dash* curve have the, ***JB(***k_1, k_2***)*** power.

Figure 4 shows the results of the small sample (25 observations) using the Size-Power curves for all three *JB*-tests for the six different models. We see the ***JB(***k_1, k_2***)*** test has higher power as we expected, because it was also the test with higher size from the size analysis section. Unfortunately *JB* shows rather poor small sample properties, and the same erratic form as with the size: the tests over- and under-rejection for the the 45^o line.

Even the ***JB(***k_1, k_2***)*** does not escape that bias. However, we are able to observe that the positive kurtosis (Models 1 and 2) has a larger effect than both the skewness and kurtosis together.

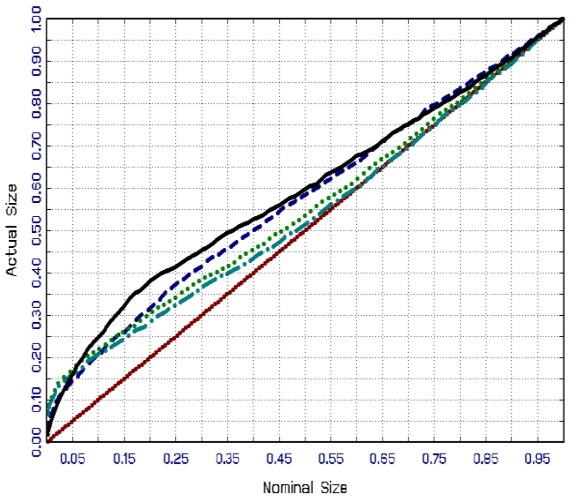
Figure 4: 25 Observations



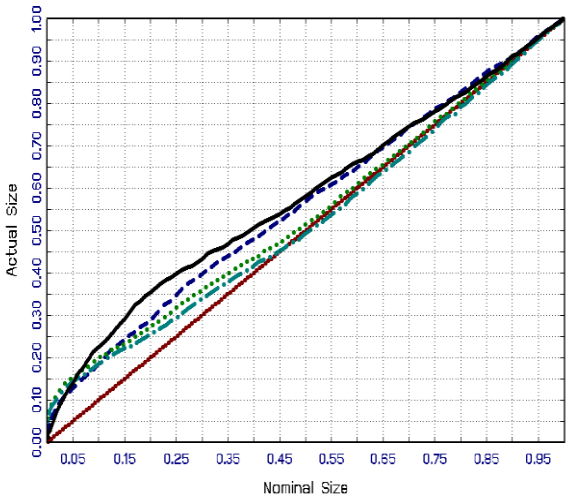
JBK = dash, ***JBM***= *point*, ***JBCV(*** k_1, k_2 ***)*** = *line*, ***JB(*** k_1, k_2 ***)***= *point-dash*, the 45^o line = *close-point*

Figure 5: 50 Observations

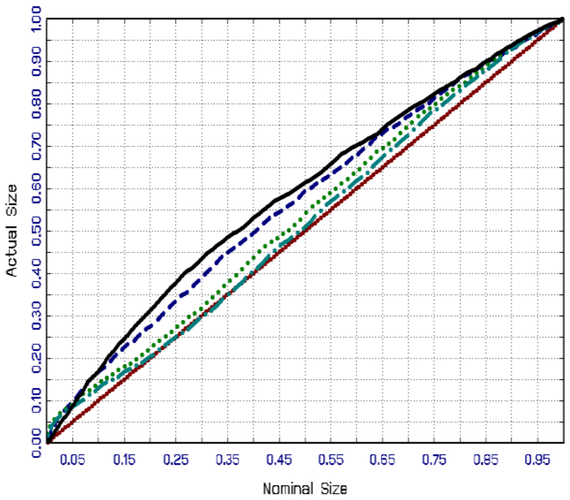
Model 1



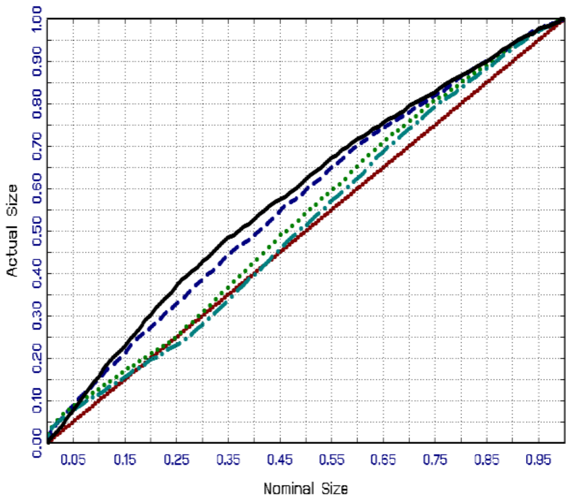
Model 2



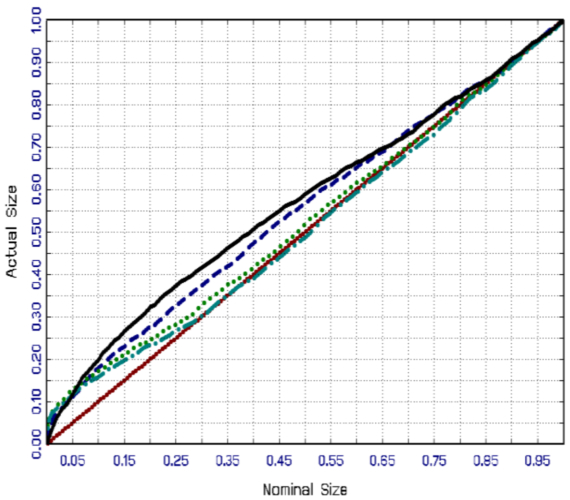
Model 3



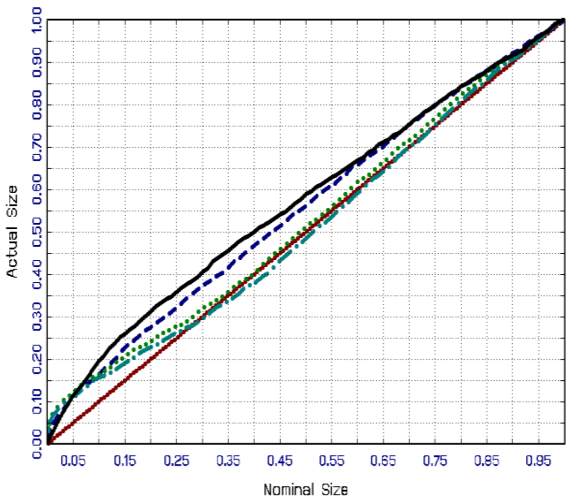
Model 4



Model 5

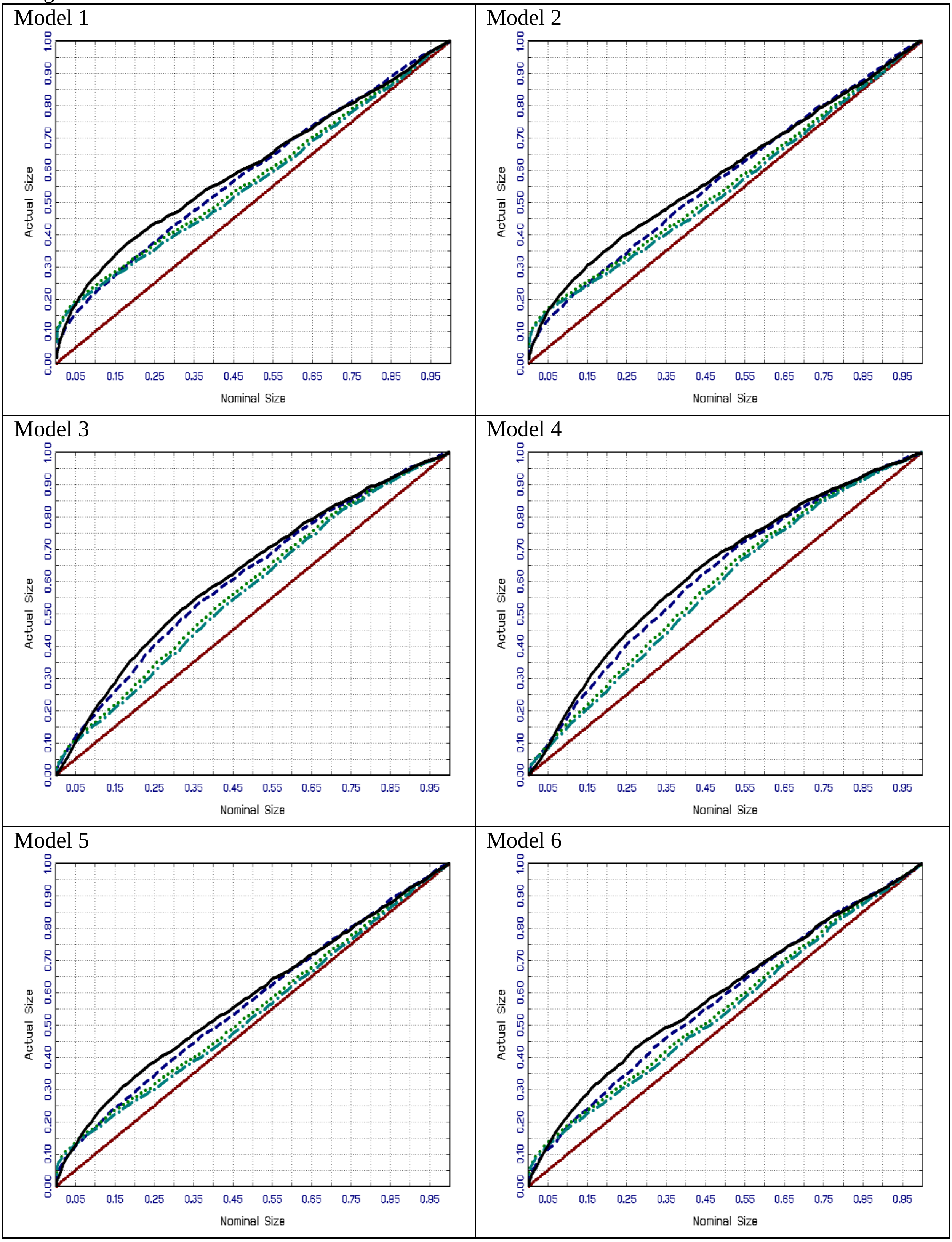


Model 6



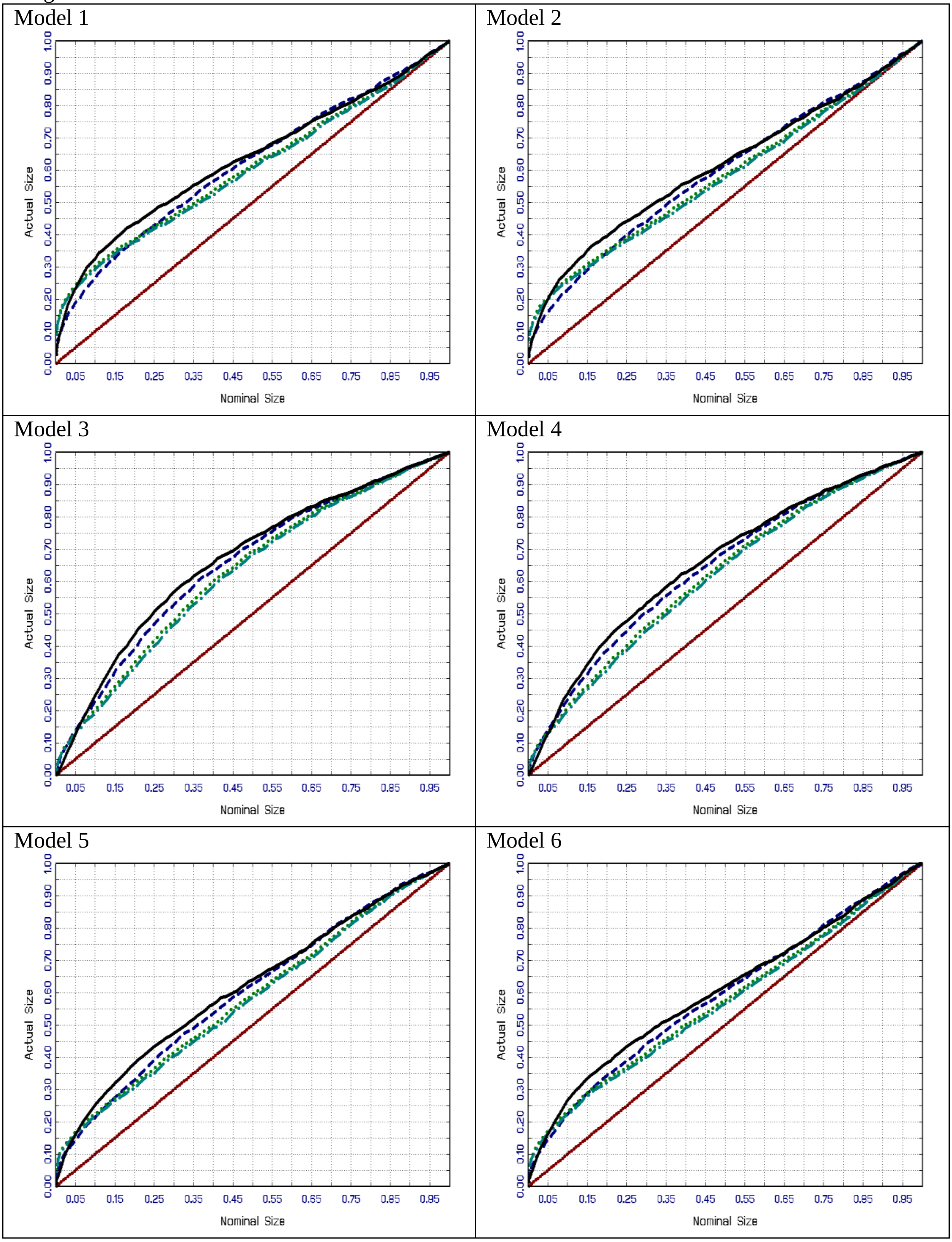
JBK* =** dash, ***JBM*=** *point*, ***JBCV(k_1, k_2 ***)*** = *line*, ***JB(*** k_1, k_2 ***)*=** *point-dash*, the 45^o line = *close-point*

Figure 6: 75 Observations



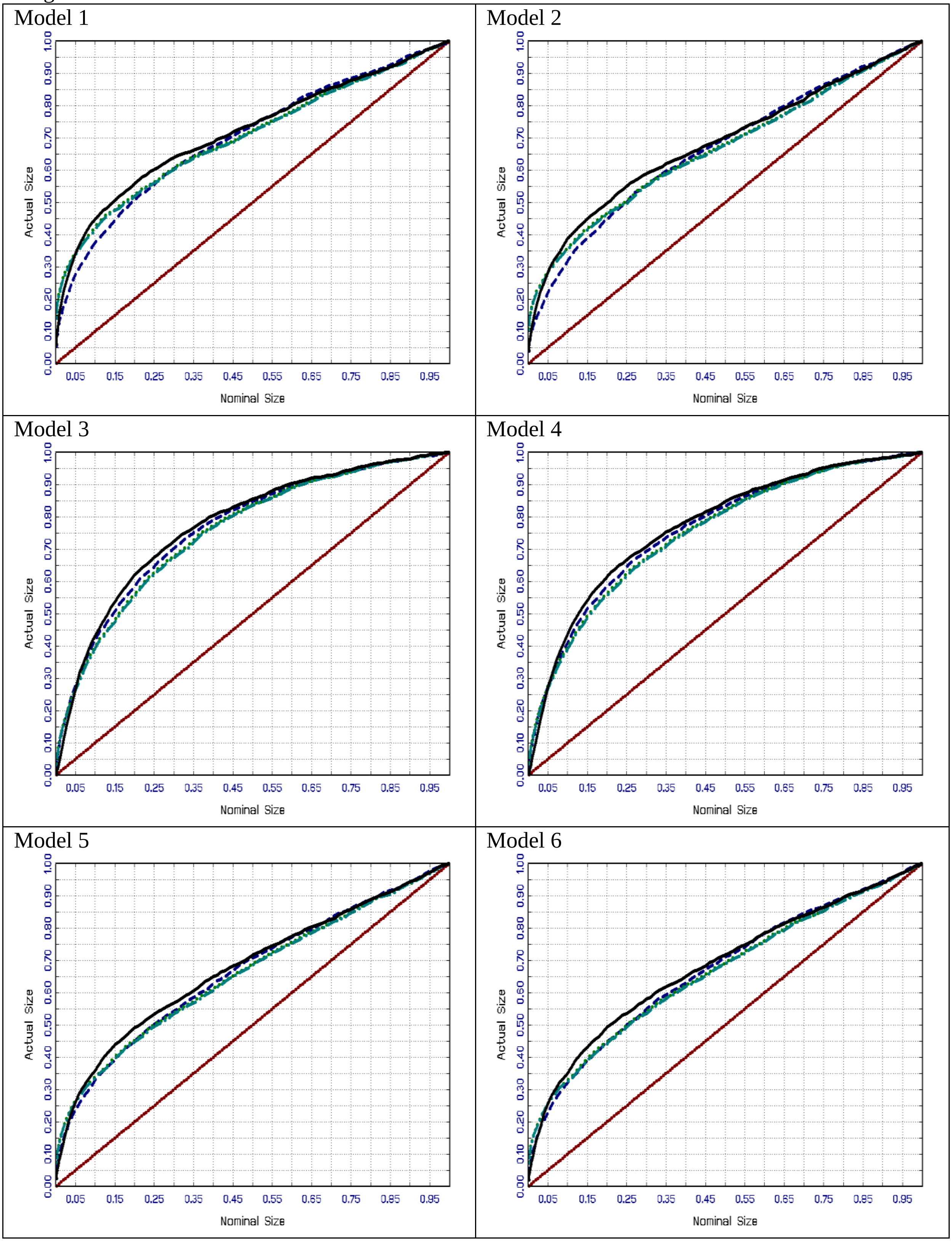
JBK = dash, ***JBM***= *point*, ***JBCV(*** k_1, k_2 ***)*** = *line*, ***JB(*** k_1, k_2 ***)***= *point-dash*, the 45^o line = *close-point*

Figure 7: 100 Observations



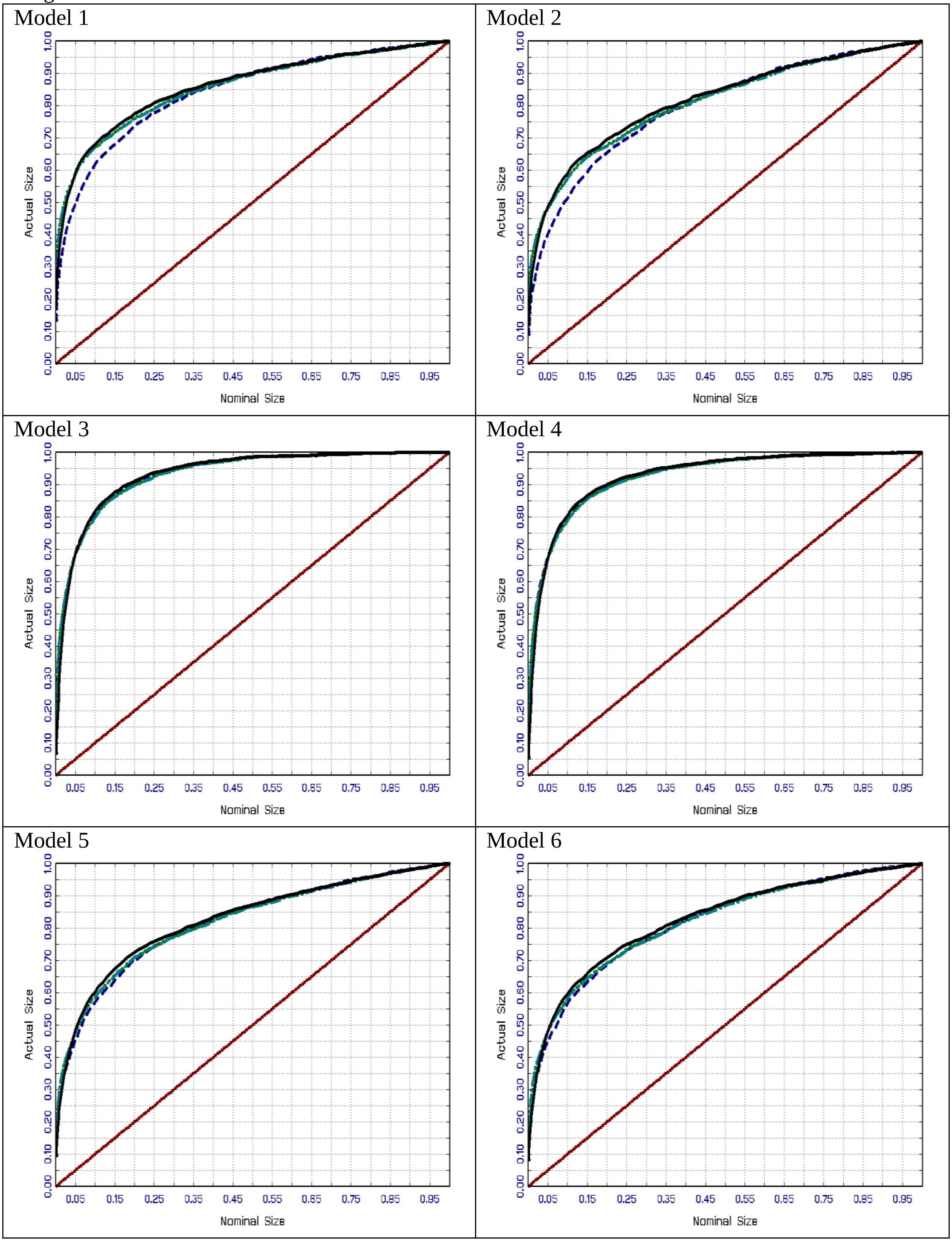
JBK = dash, ***JBM***= *point*, ***JBCV(*** k_1, k_2 ***)*** = *line*, ***JB(*** k_1, k_2 ***)***= *point-dash*, the 45^o line = *close-point*

Figure 8: 200 Observations



JBK = dash, ***JBM***= *point*, ***JBCV(*** k_1, k_2 ***)*** = *line*, ***JB(*** k_1, k_2 ***)***= *point-dash*, the 45^o line = *close-point*

Figure 9: 500 Observations



JBK = dash, ***JBM***= *point*, ***JBCV***(k_1, k_2) = *line*, ***JB***(k_1, k_2)= *point-dash*, the 45^o line = *close-point*

Figure 5 shows the results of the small sample of 50, the same results as before in Figure 4. Only after the sample of 75 observations (Figure 6) do the tests start to behave as they should.

In Figures 7–9 we observe the sample effects on the behaviour of the different versions of the *JB-tests,* that is, by increasing the sample we get higher power.

In small and medium samples the skewness effects more than the kurtosis the *JB-tests,* while in large samples the kurtosis effects more the power of the tests, see Figure 9 and Models 3 and 4.

7. SUMMARY AND CONCLUSIONS

The distributions of the test *JB* statistic and its modifications that we usually use are known only asymptotically and, unfortunately, unless the sample size is very large, the tests may not have the correct size. Inferential comparisons and judgements based on them might be misleading.

Here we presented one simple and easy way to test for normality by only using the *JB* statistic, but instead of the asymptotical critical, we generating robust critical values for our test statistic, by using the "sample" technique. That is, by using the improved critical values the true size of the test approaches its nominal value.

Monte Carlo methods and "P-*value plot*" are used to study the size, and the "*Size-Power curves*" to study the power of the JB normality test with the "sample" critical values and compare with three alternatives of the Jarque and Bera LM test for normality: the Urzúa (1996) modification of the Jarque-Bera test, ***JBM***; the Omnibus K^2 statistic made by D'Agostino, Belanger and D'Agostino (1990), ***JBK***; and finally the Jarque and Bera LM test for normality by using the quantities k_1 and k_2 are the definitions of sample skewness and kurtosis ***JB(*** k_1, k_2 ***)***.

About the size of the tests our analysis shows that our method of using the Jarque and Bera LM test for normality by using the "sample" critical values, ***JBCV(*** k_1, k_2 ***),*** is superior to the other modification of the JB test. The ***JBCV(*** k_1, k_2 ***),*** has the right size for all samples, small, medium and large. Moreover , in studied cases it has the higher power of the other comparing tests.

References

Anscombe, F.J. & Glynn, W. J. (1983). "Distribution of the kurtosis statistic b_2 for normal statistics". *Biometrika 70 (1): 227–234.*

Bera, A. & Jarque, C. (1982). “Model specification tests: A simultaneous approach”. *Journal of Econometrics 20: 59–82.*

Bowman, K. & Shenton, L.R. (1975). “Omnibus contours for departures from normality based on $\sqrt{b_1}$ and b_2 ”. *Biometrika 62: 243–250.*

Cramér, H. (1946). Mathematical Methods of Statistics, Princeton, N.J: Princeton University Press.

D’Agostino, Ralph B. (1970). "Transformation to normality of the null distribution of g_1". *Biometrika 57 (83):679–681.*

D’Agostino, R.B., Belanger, A. &; D’Agostino, R.B.Jr (1990). "A suggestion for using powerful and informative tests of normality". *The American 44 (4): 316–321*

Davidson, R & MacKinnon, J.G. (1998). “Graphical methods for investigating the size and power of test statistics.” *The Manchester School, 66: 1–26.*

Doornik, J. A. & Hansen, H. (1994). “An omnibus test for univariate and multivariate normality”, Discussion paper, Nuffield College.

Jarque, C. & Bera, A. (1987). “A test for normality of observations and regression residuals”. *International Statistical Review 55: 163–172*

Joanest, D.N. & Gill, C.A. (1998). “Comparing measures of sample skewness and kurtosis”. *The Statistician, 47:183 –189.*

Mantalos, P. (2010). “The three different measures of the sample skewness and kurtosis and the effects to the Jarque-Bera test for normality.” Discussion paper, Jönköping International Business School.

Pearson, E. S (1931). “Note on tests for normality”. *Biometrika 22 (3/4): 423–424.*

Ramberg, J. &. Schmeiser, B. (1974). “An approximate method for generating asymmetric random variables.” *Communications of the ACM, 17(2): 78–82.*

Shenton, L.R.& Bowman, K.O. (1977). “A bivariate model for the distribution of $\sqrt{b_1}$ and b_2”. *Journal of the American Statistical Association 72 (357): 206–211.*

Urzúa, C. (1996) “On the correct use of omnibus tests for normality”, *Economics Letters 53:247–51*