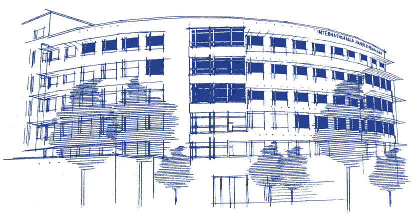
Three Different Measures of Sample Skewness and Kurtosis and their Effects on the Jarque-Bera Test for Normality

PANAGIOTIS MANTALOS

JÖNKÖPING INTERNATIONAL BUSINESS SCHOOL
JÖNKÖPING UNIVERSITY



Jönköping International Business School
Jönköping University
JIBS Working Papers No. 2010-9

Three Different Measures of Sample Skewness and Kurtosis and their Effects on the Jarque-Bera Test for Normality

Panagiotis Mantalos[1]

Departments of Economics and Statistics,

Jönköping International Business School, Sweden.

ABSTRACT

Monte Carlo methods are used to study the size and the power of three versions of the Jarque and Bera LM test for normality, $JB(g_1, g_2)$, $JB(b_1, b_2)$, and finally $JB(k_1, k_2)$. The difference between these tests comes from the different definitions (estimates) of sample skewness and kurtosis. The Jarque and Bera test 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. However the $JB(k_1, k_2)$ for a 5% nominal level shows good properties for all samples. The power of the tests shows the same erratic form.

Keywords: Jarque and Bera LM test ; Kurtosis; skewness; Test for normality.

[1] Communications sent to: Panagiotis Mantalos. Sweden. E-mail: Mantalos03@yahoo.se

1. Introduction

One of the most widely used instruments in the first step of analysing a data set is the assumption of normality. The commonly assumed normality 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 misleading interpretations.

That is, testing for normality should be at least an important step.

The most widely used method at least in econometrics, that has been suggested and used for testing whether the distribution underling a sample is normal is the Bowman and Shenton (Bowman and Shenton, 1975, Shenton and Bowman, 1977) statistic:

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

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

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

Simulation results comparing the power of the *JB* tests with other tests as the Shapiro-Wilk test (Shapiro and Wilk, 1965) or the Shapiro-Francia test (Shapiro and Francia, 1972) were reported by, among others, Pearson, D'Agostino and Bowman (1977), Jarque and Bera (1987), Mardia (1980), and Deb and Sefton (1996).

All these studies have showed that the *JB* test is simple to compute and its power has proved comparable to other powerful tests.

A very important ingredient for studying the *JB* test is that over the years, various measures of sample skewness and kurtosis have been proposed.

Three measures of skewness and kurtosis are studied by Joanest and Gill (1998); defined as g_1, b_1 and k_1 for skewness and g_2, b_2 and k_2 for kurtosis, and we will briefly describe them in the next chapter. Note also that the definitions of sample skewness and kurtosis that Joanest and Gill (1998) used, were the same as the definitions adopted by some computing packages, such as SAS, SPSS, MINITAB, BMDP, and also by the EXCEL spreadsheet program.

In our study we investigate what are the effects of these definitions on the *JB* test for normality.

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

2. Skewness, kurtosis and the 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 $\sigma_x^2 = \mu_2$. In the same way the central moments m_j are:

$$m_j = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^j$$

Then the skewness γ_1 and kurtosis γ_2 are defined as:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}, \gamma_2 = \frac{\mu_4}{\mu_2^2}. \qquad (2.1)$$

While the sample skewness and kurtosis are defined as:

$$skew = \frac{m_3}{m_2^{3/2}} = \frac{1/n \sum_{i=1}^{n} \left(x - \bar{x}\right)^3}{\left[1/n \sum_{i=1}^{n} \left(x - \bar{x}\right)^2\right]^{3/2}} \tag{2.2}$$

$$kurt = \frac{m_4}{m_2^2} = \frac{1/n \sum_{i=1}^{n} \left(x - \bar{x}\right)^4}{\left[1/n \sum_{i=1}^{n} \left(x - \bar{x}\right)^2\right]^{2}} \tag{2.3}$$

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 be calculated as well.

By defining $g_1 = skew = \frac{m_3}{m_2^{3/2}}$ the skewness, and $g_2 = kurt - 3 = \frac{m_4}{m_2^2} - 3$ the excess, Pearson (1931) showed that:

$$\mu_1(g_1) = 0 \tag{2.4}$$

$$\mu_2(g_1) = \frac{6\left(n-2\right)}{\left(n+1\right)\left(n+3\right)} \tag{2.5}$$

and

$$\mu_1(g_2) = -\frac{6}{n+1} \tag{2.6}$$

$$\mu_2(g_2) = \frac{24\left(n-2\right)\left(n-3\right)}{\left(n+1\right)^2\left(n+3\right)\left(n+5\right)} \tag{2.7}$$

Based on Cramér (1946), and to remove the bias in g_2 and achieve consistency at the same time, Joanest and Gill (1998) used the following estimates:

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

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

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 program. In contrast, MINITAB and BMDP define skewness and kurtosis by (see Joanest and Gill (1998)).

$$b_1 = \frac{m_3}{s^3} = \left(\frac{n-1}{n}\right)^{3/2} \frac{m_3}{m_2^{3/2}} \tag{2.10}$$

and

$$b_2 = \frac{m_4}{s^4} = \left(\frac{n-1}{n}\right)^{2} \frac{m_4}{m_2^{2}} \tag{2.11}$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Joanest and Gill (1998) found that for samples from a normal distribution, all three measures of skewness are unbiased, but in small samples the mean-squared error is less for MINITAB's b_1, than for g_1 and greater for SAS's k_1 than for g_1.

The kurtosis measure k_2 also has the largest mean-squared error whereas b_2 has a mean-squared error that is only slightly larger than that of g_2.

For large samples, there is very little difference between the three measures.

Moreover, Joanest and Gill (1998) found that the variances of k_1 and k_2 are the greatest, whereas b_1, and b_2 have the smallest variances, whatever distribution is being sampled.

Both g_1 and g_2 are 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, which subsequently was derived as an LM test by Jarque and Bera (1987):

$$JB(g_1,g_2)=n\left[\frac{(g_1)^2}{6}+\frac{(g_2)^2}{24}\right], \tag{2.12}$$

and by using (b_1, b_2) and (k_1, k_2) we have the following *JB statistic versions*:

$$JB(b_1,b_2)=n\left[\frac{(b_1)^2}{6}+\frac{(b_2)^2}{24}\right], \tag{2.13}$$

$$JB(k_1,k_2)=n\left[\frac{(k_1)^2}{6}+\frac{(k_2)^2}{24}\right]. \tag{2.14}$$

The *JB-statistics* are asymptotically chi-squared distributed with two degrees of freedom because *JB* is just the sum of squares of two asymptotically independent standardized normals.

3. 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

$$F^{-1}(u) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2} \qquad (3.1)$$

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

4. 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 that were developed and illustrated by Davidson and MacKinnon (1998) and 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, 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 *JB* shows rather poor small sample properties, the tests are over-sized for small 1% and 2% nominal levels and under-sized for the rest of the levels for 25 observations even in the larger sample of 50. For both the *b*s and *g*s show the same behaviour as the *JB* tests: over-sized for small nominal levels and under-sized for levels larger than 3%, with less bias for the *b*s. The only positive is that for the *k*s estimates the *JB* test, even if over-sized for small nominal level and under-sized for the larger, has the surprising result that for the 5% level, the confidence interval is in fact almost an exact 5% for both samples! Figure 2 shows the same result for the medium samples, 75 and 100. Now the estimated *JB* tests are nearer to the confidence interval. But there are still the same effects: the *JB* tests areover-sized for small nominal levesl and under-sized for levels larger than 3% with less bias for the *b*s. Even here, for the the *k*s estimates the *JB* test has surprisingly good results for the 5% level. The estimated size of the test is inside the confidence interval.

In large samples, more than 500, Figure 3b shows that the differences in definition are unimportant. However, in small or moderate samples and even in a sample as large as 200 observations, Figure 3a, the differences can be quite startling; the tests are over-sized for small, up to 3%, nominal levels, and under-sized for the rest of the levels.

Figure 1: Small sample P-value plots: size of the tests

Figure 1a: 25 observations

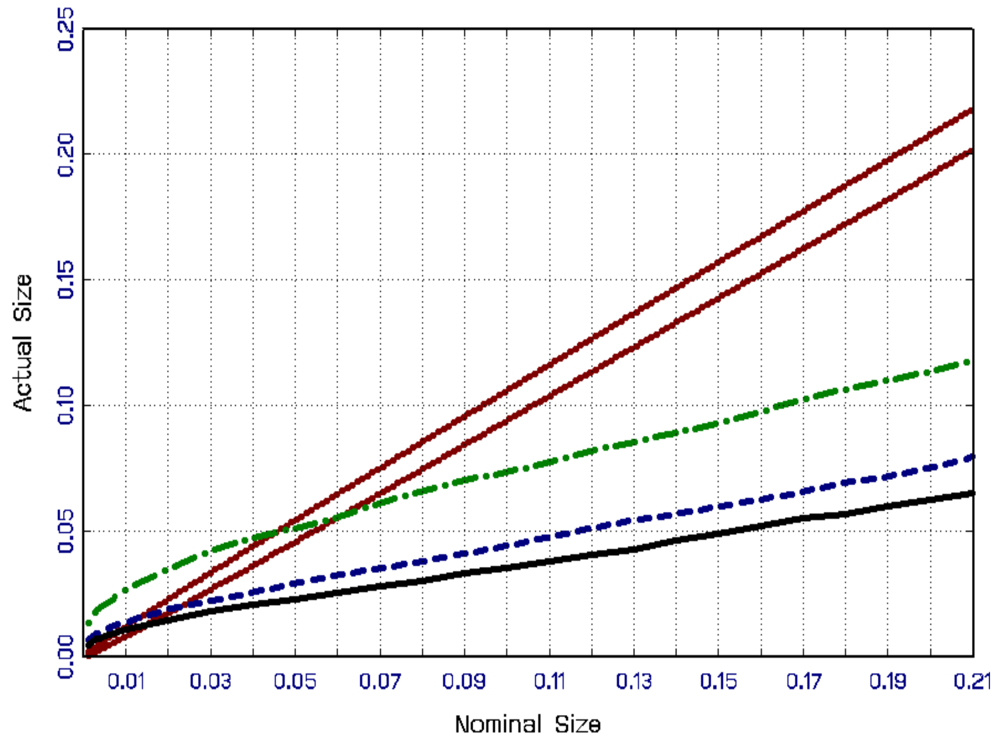
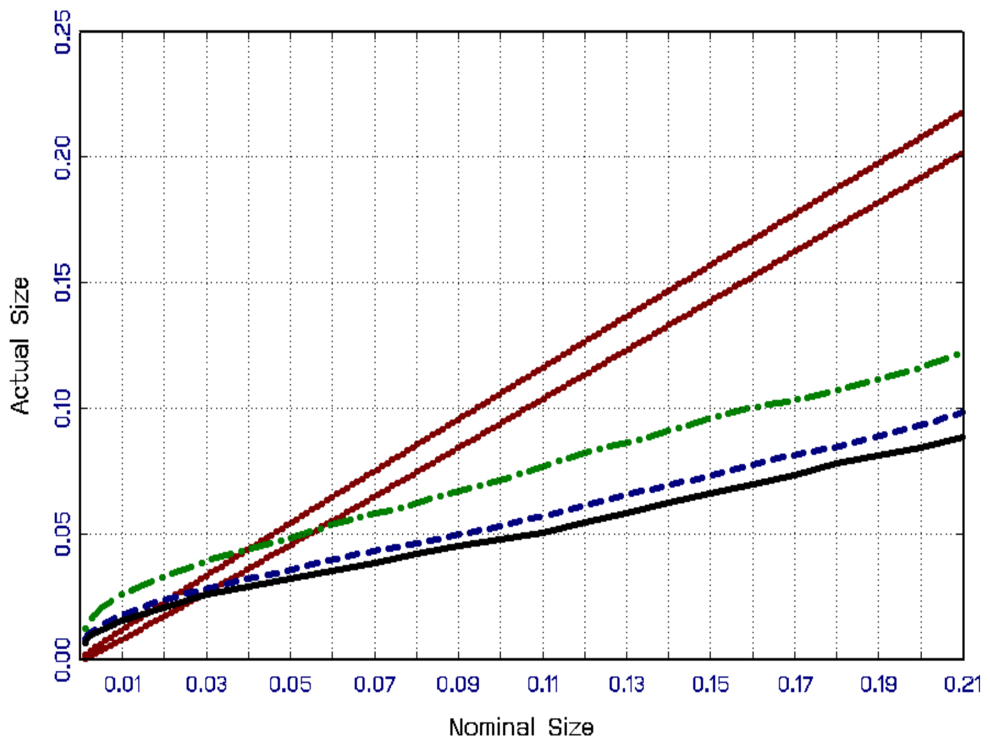


Figure 1b: 50 observations



JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *point-dash* and ***95% confidence interval*** *= point*

Figure 2: Medium sample P-value plots: size of the tests

Figure 2a: 75 observations

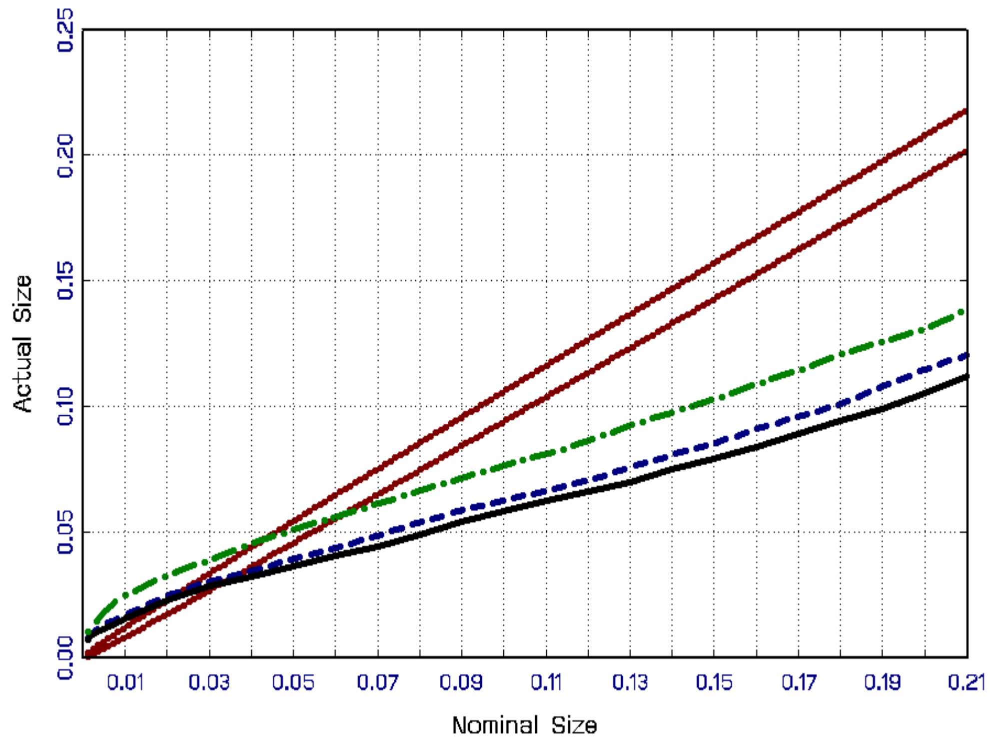
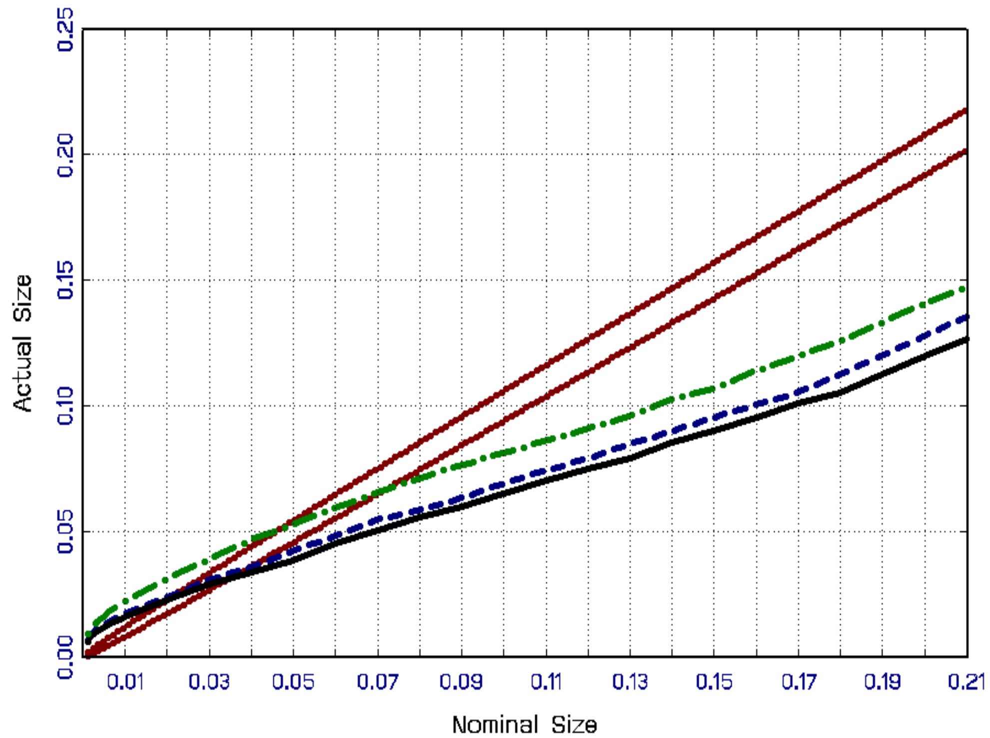


Figure 2b: 100 observations.



JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *point-dash* and ***95% confidence interval*** *= point*

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

Figure 3a: 200 observations.

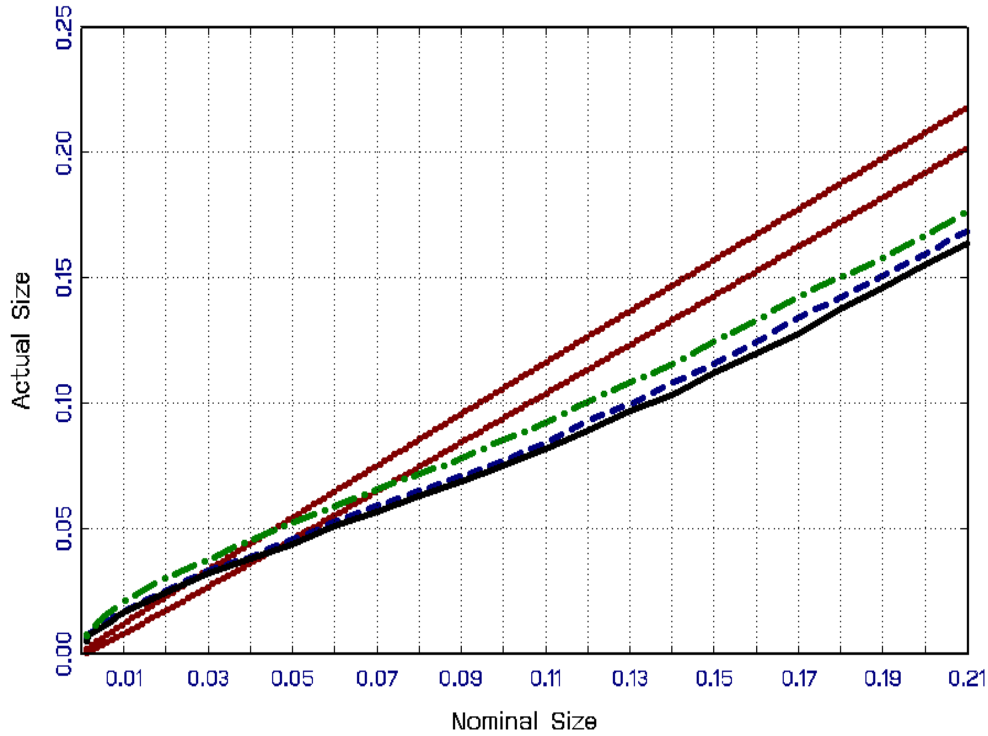
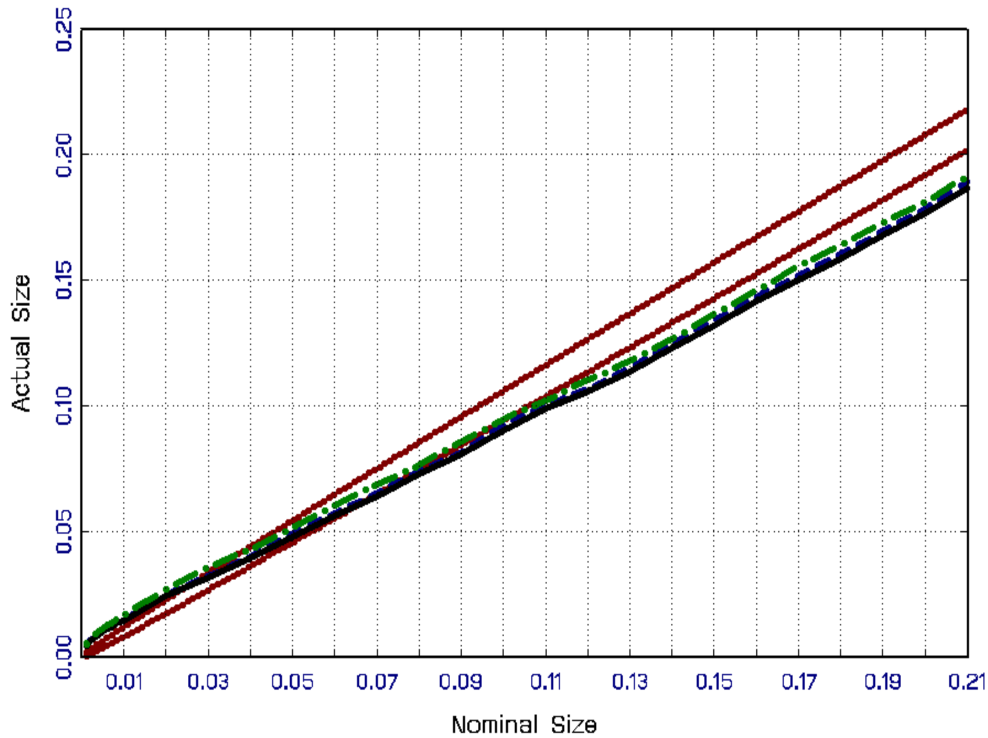


Figure 3b: 500 observations.



JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *point-dash* and ***95% confidence interval*** *= point*

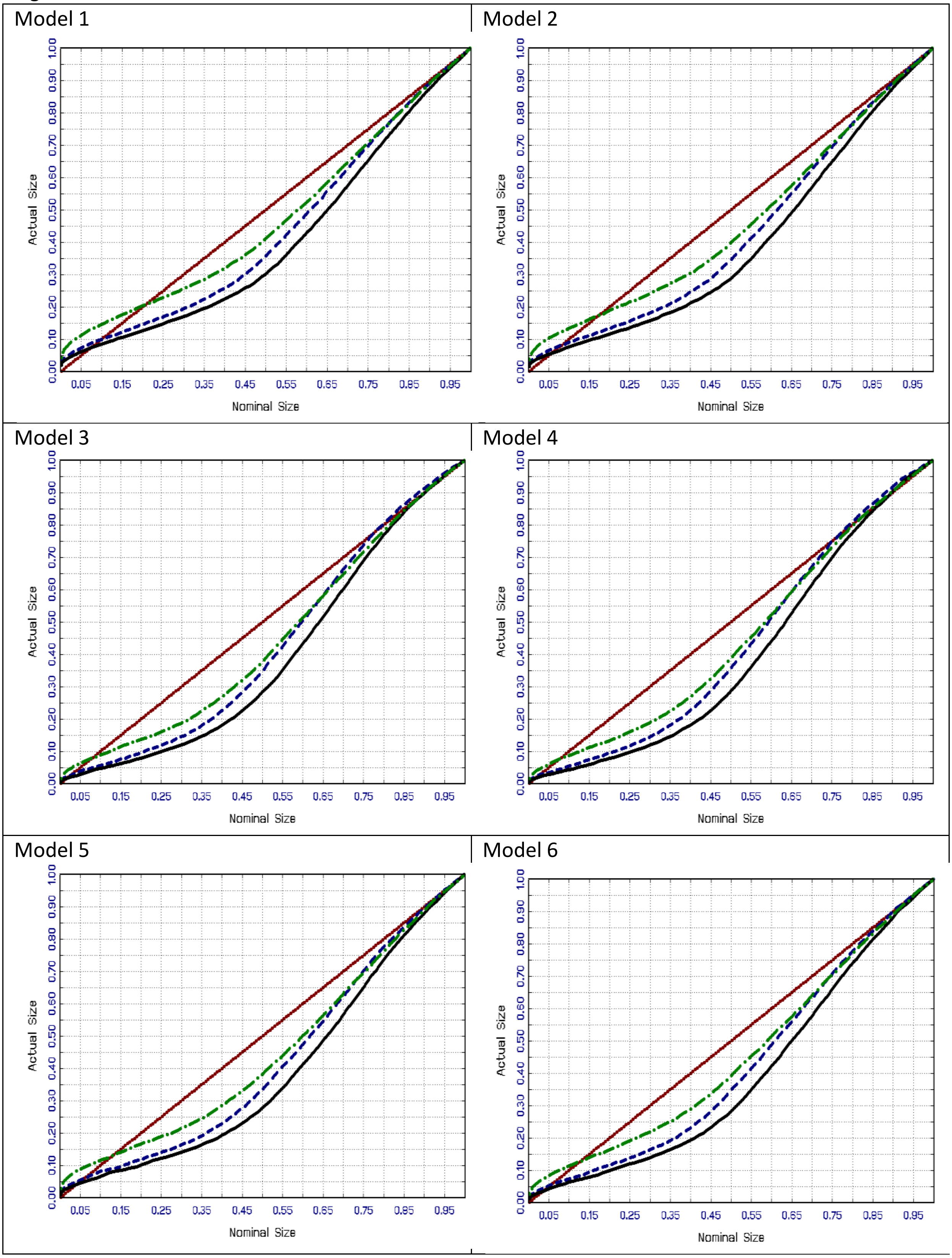
5. Analysis of the Power of the Tests.

In this section, we analyse the power of the *JB* tests using small, medium, and large sample sizes of observations. The power function is estimated by calculating the rejection frequencies in 1000 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 that, 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 with *dashed* curves we have the ***JB(*** b_1, b_2***)***. Finally with *dot-dash* curve is th*,* ***JB(*** k_1, k_2***)*** power . Now, 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 that 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-reject, when compared with the 45^o line.

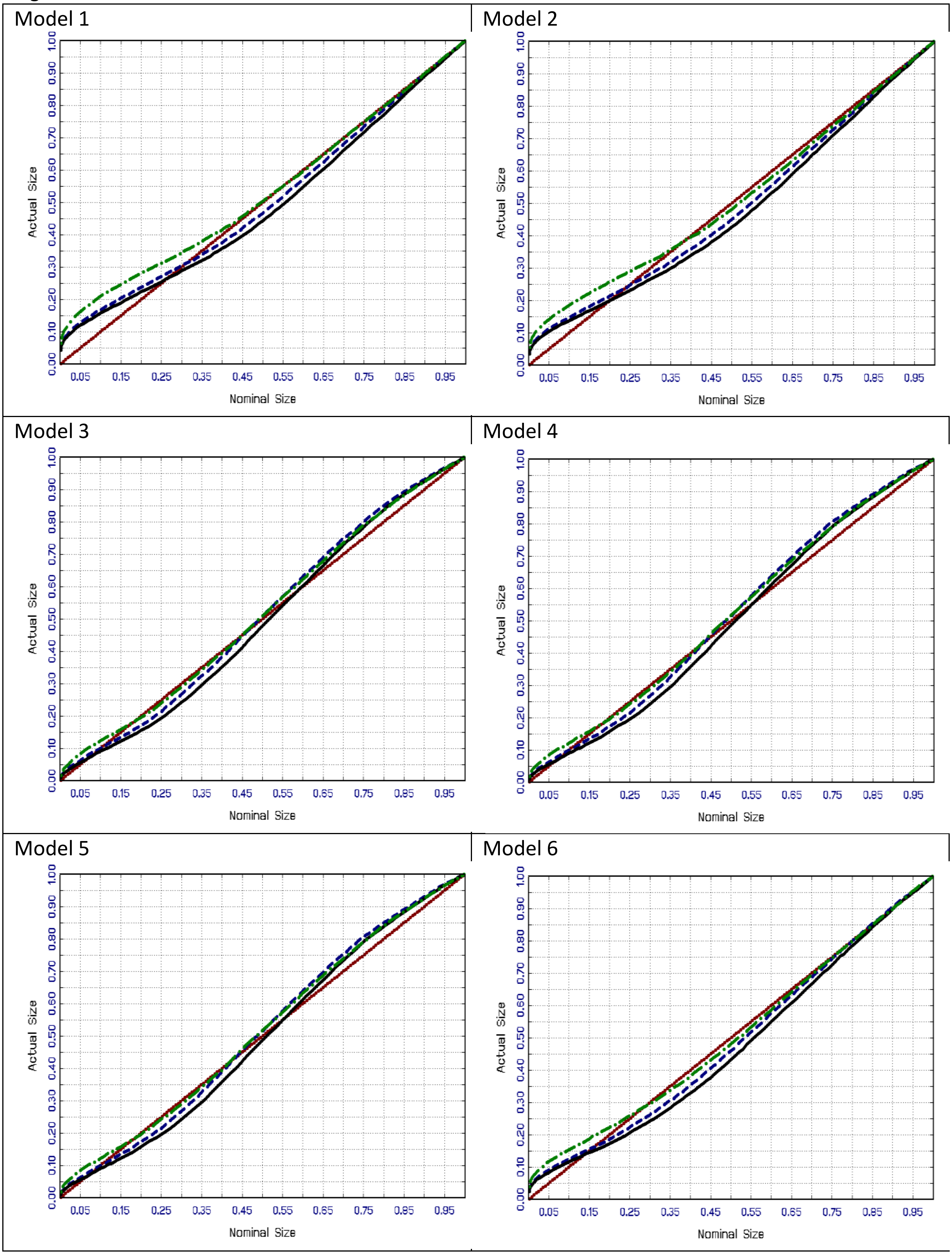
Even ***JB(*** k_1, k_2***)*** does not escape that bias. However, we are able to observe that positive kurtosis (models 1 and 2) has a larger effect than the both skewness and kurtosis together. Figure 5 shows the results of the small sample of 50. The same results as before in Figure 4, only after the sample of 75 observations, and in Figure 6 the tests start to behave as they should. In Figures 7–9 we observer 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 effect is more than the kurtosis on the *JB* tests, while in large samples the kurtosis is that which affects more the power of the tests, see Figure 9, models 3 and 4.

Figure 4: 25 observations



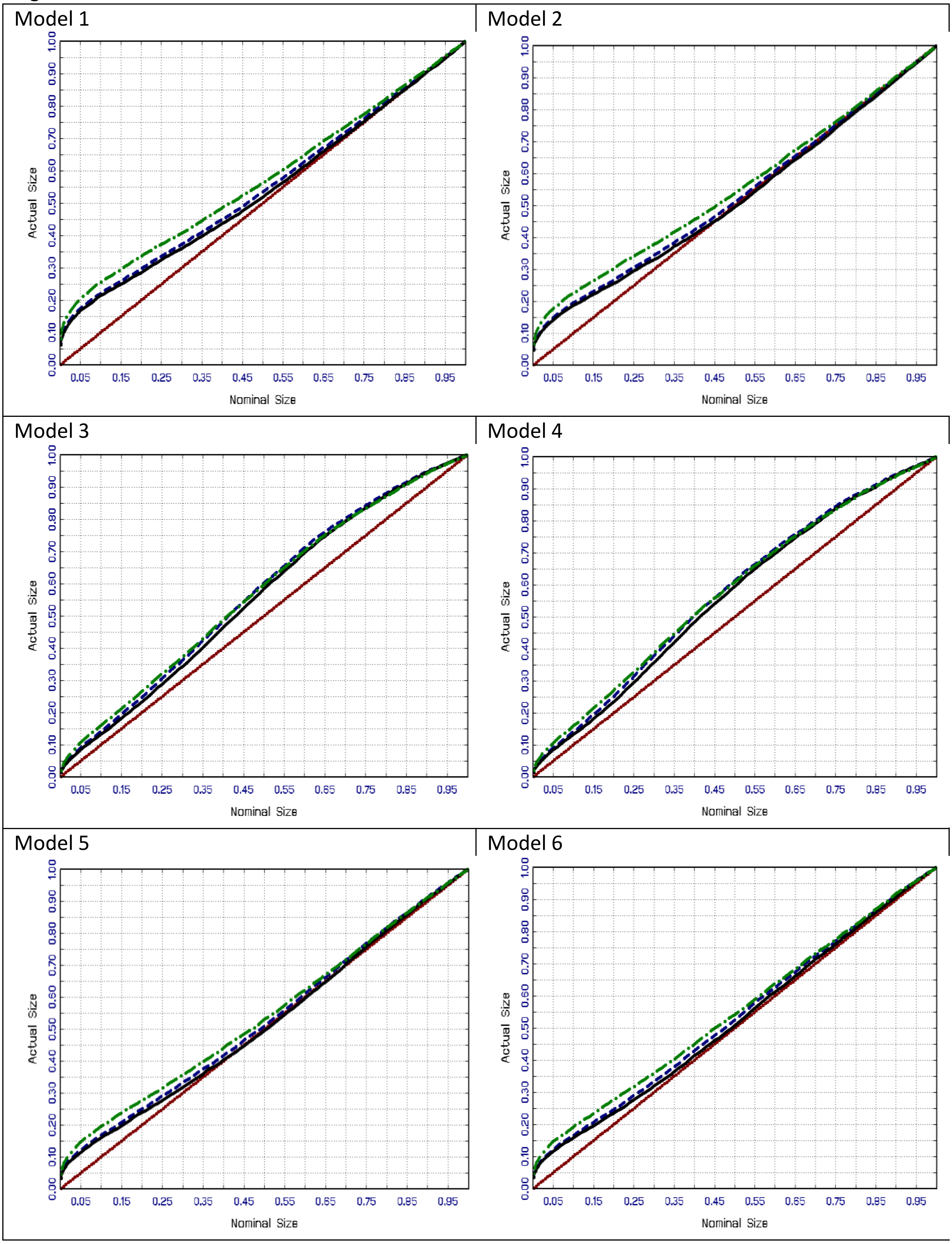
JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *dot-dash* and the 45^o line = *dots*

Figure 5: 50 observations



JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *dot-dash* and the 45^o line *= dots*

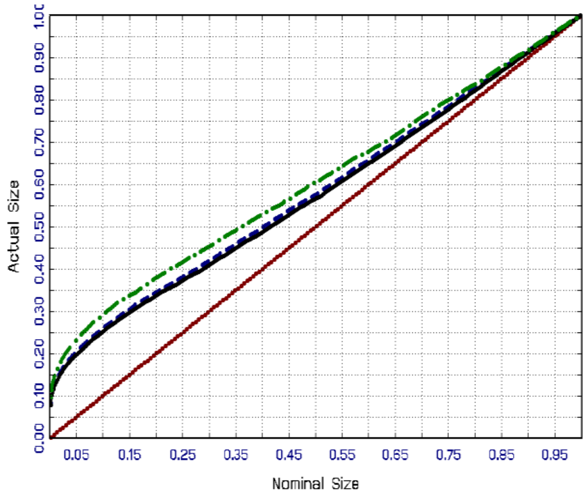
Figure 6: 75 observations



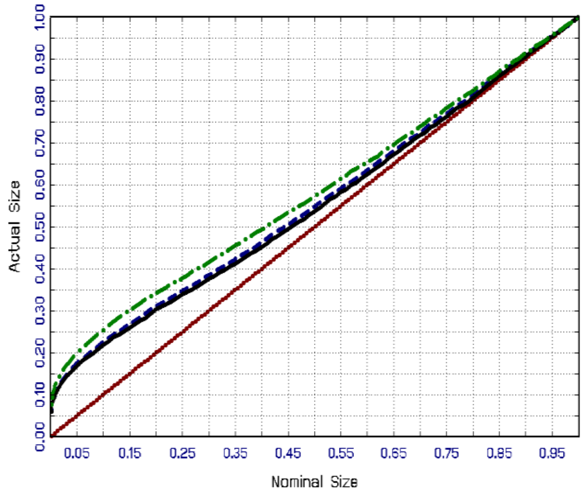
JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *dot-dash* and the 45^o line = *dots*

Figure 7: 100 observations

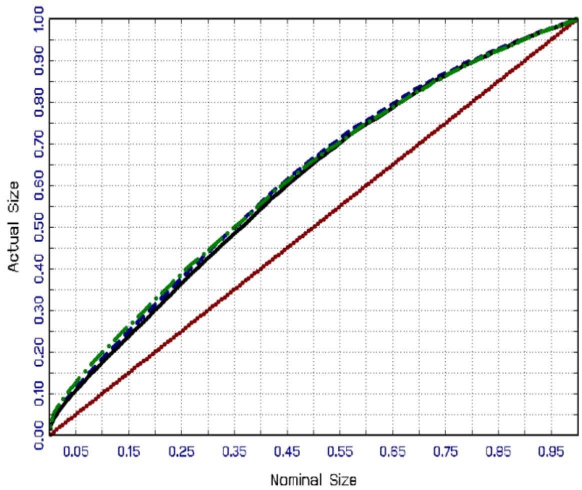
Model 1



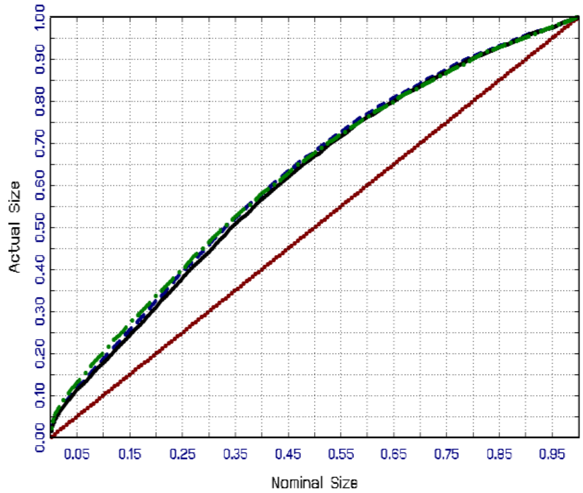
Model 2



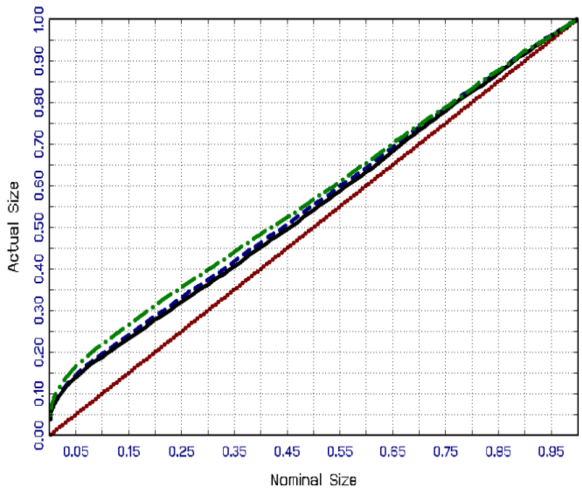
Model 3



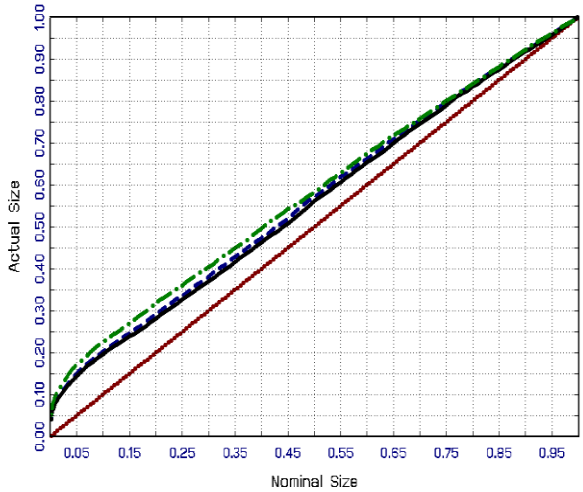
Model 4



Model 5

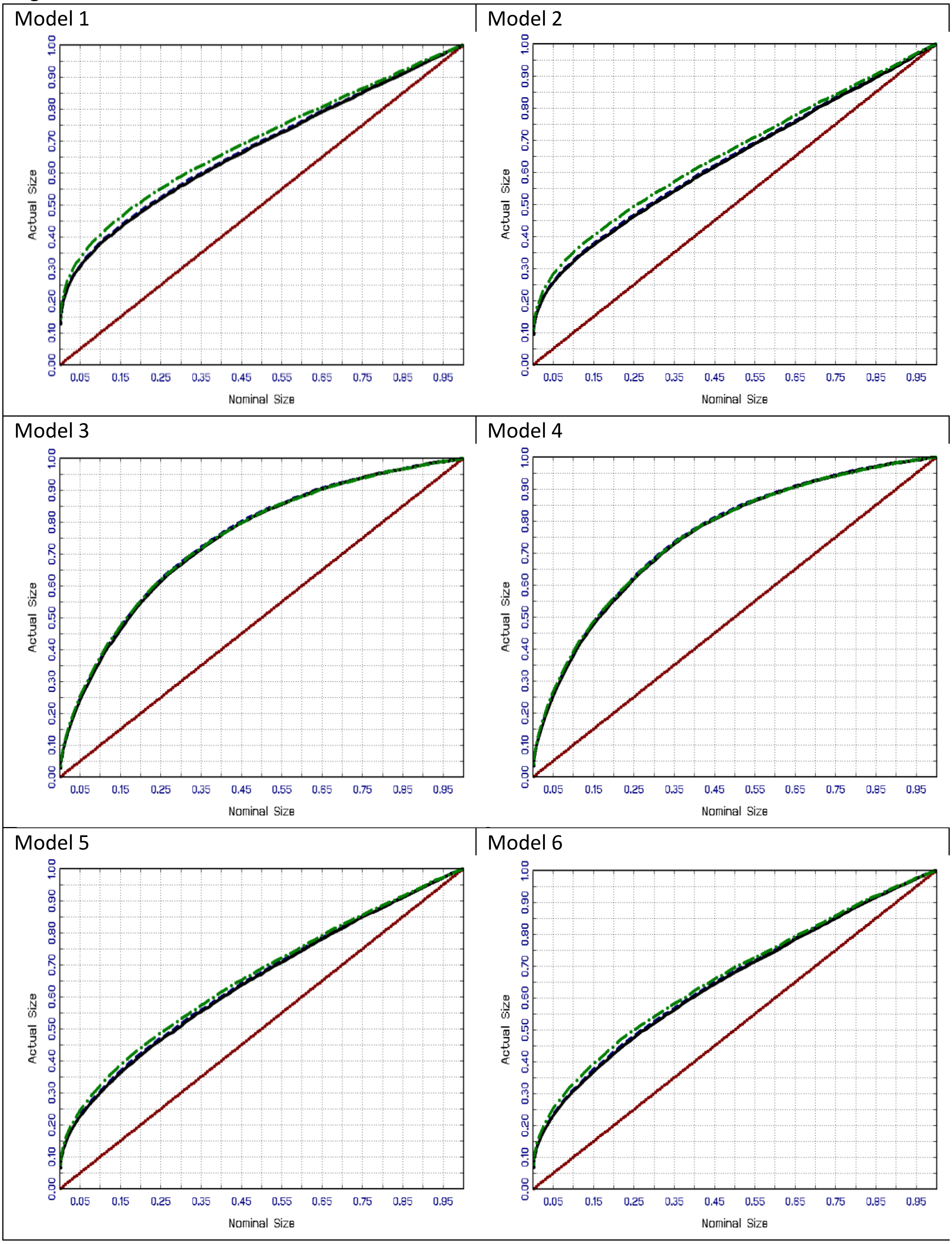


Model 6



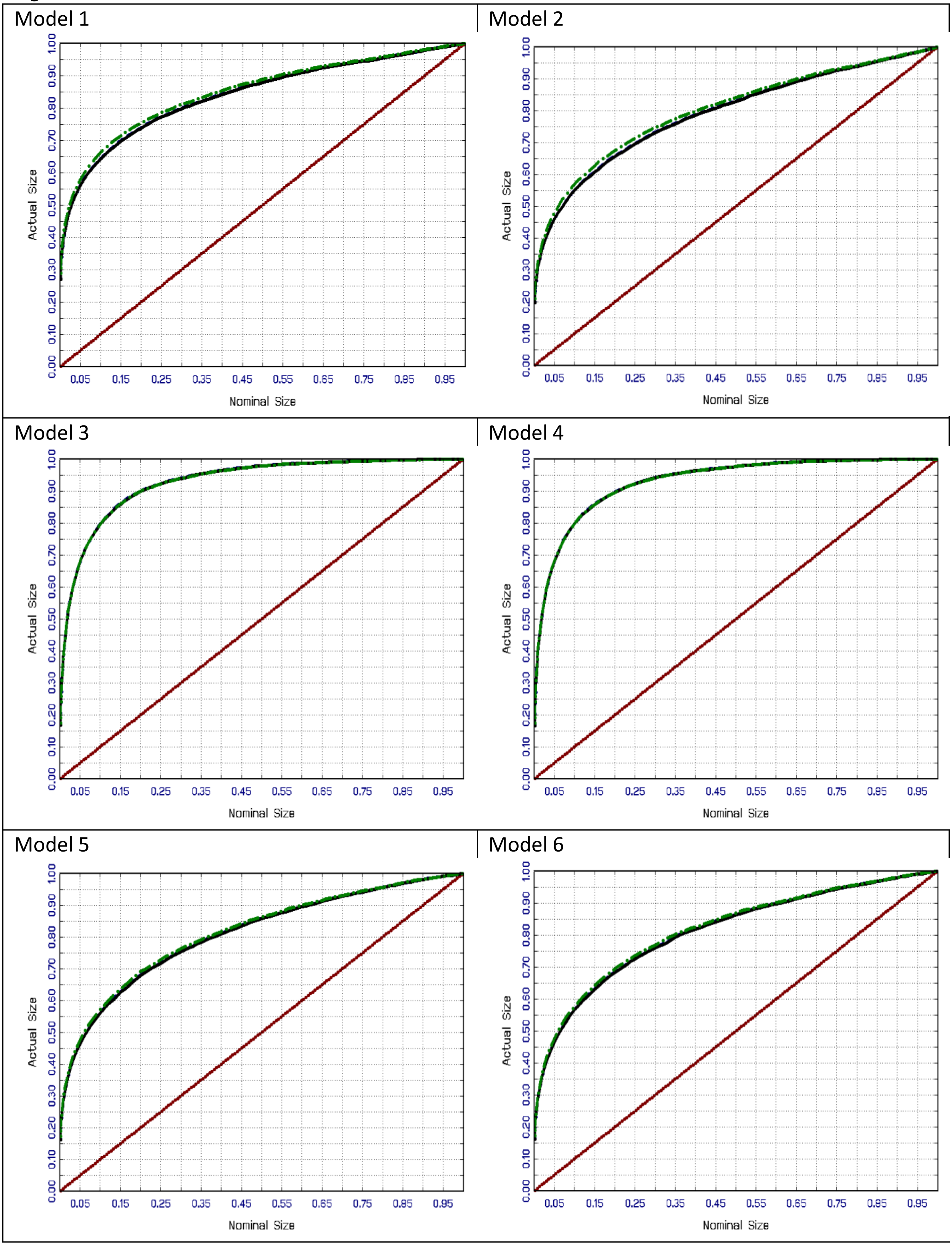
JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *dot-dash* and the 45^o line *= dots*

Figure 8: 200 observations



JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *dot-dash* and the 45^o line *= dots*

Figure 9: 500 observations



JB(b_1, b_2 ***)*** =small dash, ***JB(*** g_1, g_2 ***)*** *= line,* ***JB(*** k_1, k_2 ***)=*** *dot-dash* and the 45^o line = *dots*

6. Summary and Conclusions.

Monte Carlo methods and the P-*value plot* were used to study the size, and the *size-power curves*, to study the power of three alternative versions of the Jarque and Bera LM test for normality, *JB(* g_1, g_2 *)* and *JB(* b_1, b_2 *),* and finally *JB(* k_1, k_2 *)*. The difference between these tests comes from the different definitions (estimates) of sample skewness and kurtosis.

About the size of the tests our analysis shows that in large samples, more than 500, the differences in definition are unimportant. However, in small or moderate samples and even in a sample as large as 200, Figure 3a, the differences can be quite startling: the tests are over-sized for small, up to 3% nominal level, and under-sized for the rest of the levels.

Unfortunately *JB* has rather poor small sample properties: the slow convergence of the test statistic to its limiting distribution makes the test under-sized even in a reasonably large sample.

In spite of all the negative results of these three definitions of sample skewness and kurtosis, the quantities k_1 and k_2 as described in (2.8) and (2.9) yields a *JB*-statistic with surprisingly good results for the 5% nominal level: the estimated size of the test is inside the confidence interval, that is, near to its nominal.

The *JB*-statistic with the quantities k_1 and k_2 has also the higher power. In small and medium samples the *JB* shows rather poor small sample properties, and the same erratic form as with the size appears: the tests over- and under-rejection for the 45^o line.

Finally in small and medium samples the skewness affects the *JB* tests more than does the kurtosis, while in large samples the kurtosis is that which affects more the power of the tests.

References

Bera, A. and C. Jarque (1982). Model specification tests: A simultancous approach. *J. Econometrics* 20, 59-82.

Bowman, K. and L.R. Shenton (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.

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

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

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

Shapiro, S. and M.B.Wilk, (1965). An analysis of variance test for normality. *Biometrika* 52, 591-611.

Shapiro, S. and R. Francia, (1972). An approximate analysis of variance test for normality. *J. Am. Statist. Assoc.* 67, 215-216.

Shenton, L.R.; Bowman, K.O. (1977). A bivariate model for the distribution of $\sqrt{b_1}$ and b_2. *J.Am. Statist. Assoc.* 72(357): 206–211.