Properties of tests for heteroskedasticity

Vien Dinh and Jorrim Prins

December 2019

1 Introduction

An important assumption for various widely-used estimators is the assumption of homoskedasticity. As data points usually have dissimilar variabilities it is important to analyse the existing tests for this assumption. Three of the most important tests for this scenario are the LR, LM and Wald tests. In this compact research, Monte Carlo simulations will evaluate size and power properties of the tests and determine which test performs best.

2 Model

This research uses a standard linear regression model with errors that can be either homo- or hetero-skedastic. The model looks as follows:

$$y_i = x_i'\beta + u_i \tag{1}$$

$$u_i \sim N\left(0, \sigma^2 \exp\left(\gamma' z_i\right)\right) \tag{2}$$

This model is homoskedastic when $\gamma = 0$ and heteroskedastic otherwise, so this condition is used as the null hypothesis in the analysis. x_i is one observation of the X matrix (N x k) that includes a constant (the first column consists of ones) and z_i is one observation of the Z matrix (N x m) that (possibly) affects the variance of y_i. The full parameter vector of the model looks as follows: $\theta = (\beta', \sigma^2, \gamma')'$

3 Results & Analysis

The tests to compare for a possibly heteroskedastic model are the Wald test, Likelihood Ratio test (LR), Lagrange Multiplier/score test (LM) and LM outer-product-of-the-gradient test (LM OPG). The size and power properties of these tests will be evaluated and all results are obtained by using 5 percent significance levels. The specified model is simulated by Monte Carlo simulation with 1000 replications and varying sample size. The null hypothesis H_0 for the tests is homoskedasticity ($\gamma = 0$), while the alternative hypothesis H_a is $\gamma \neq 0$.

In Figure 1 the probability densities can be observed for $N = 100$ and $N = 1000$. We see that the LM and LM OPG tests have a higher peak than the Chi-squared distribution for $N = 100$, which is compensated by a lower density around the left side of the critical value of the test statistic. Asymtotically (for example with $N = 1000$) all the test statistics follow a Chi-squared ddistribution.

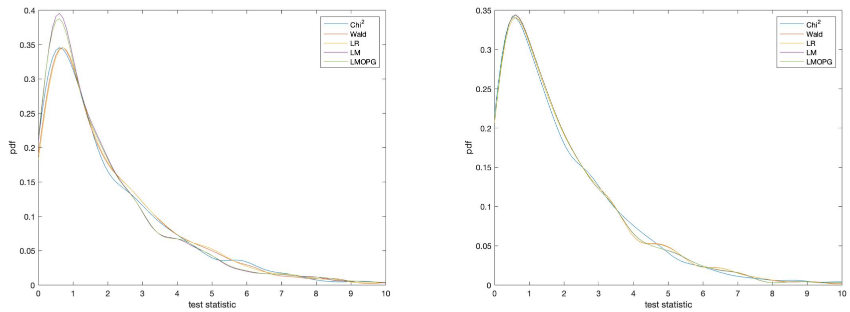


Figure 1: Probability distributions for N=100, 1000

3.1 Size

All four tests asymptotically converge to a chi-squared distribution with two degrees of freedom, so the size is found by using a critical value of 5.9915. Rejection probabilities under the null hypothesis, which defines the size, can be found in Figure 2. It shows convergence of the size to 5% when N gets large, as is expected for all tests. However, the tests behave different when N is not large enough. For N=15 and N=25 the size is very high for the Wald and LR test and very low for the LM and LM OPG test. Using the asymptotic critical value of 5.9915 to construct and compare power curves for the tests will cause problems as a result of this size distortion. An easy solution to this problem is adjusting the critical values per test to the value that ensures exactly 5% Type I errors and use these critical values to calculate the power curves. The adjusted critical values can be found in Figure 1 as well.

	Size			CV		
Test	N=15	N=25	N=100	N=15	N=25	N=100
Wald	19.4%	11.4%	5.7%	11.7208	8.7878	6.1071
LR	14.8%	11.8%	5.7%	9.4026	8.3173	6.1353
LM	1.8%	3.3%	4.6%	4.4525	5.0674	5.7765
LM OPG	1.7%	3.4%	4.8%	4.363	5.4544	5.9306

Figure 2: Size and size-adjusted critical values

3.2 Power

Size-adjusted power curves for $N = 15, 25, 100$ can be found in Figure 3 (larger, separate images per test can be found in Appendix B). The curves become smoother when N increases as the estimations of parameters and test-statistics are more reliable and asymptotic assumptions are closer to being true, so they behave exactly as expected. A power curve should increase with the distance between H_0 and H_a, this is the case for all three values of N with some exceptions that are caused by small sample size.

For large sample size, $N = 100$ in this case, all tests have similar power curves with power values near zero at $\gamma_a = 0$, values near one when $|\gamma_0 - \gamma_a| \approx 0.5$ or larger and smooth lines between them. This is not the case for smaller sample size as the curves of the tests moderately differ and have different properties. Obviously, the curves have larger shocks for low sample size ($N = 15$). However, when zooming in on the top right side of the graph in Figure 3, the LR test clearly has a higher power than the LM test. The Wald and LM OPG statistic are not shown to clarify the plot, as they have similar properties to the aforementioned tests.

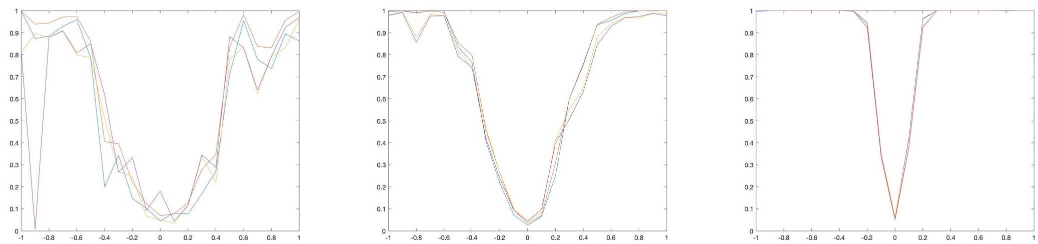


Figure 3: Size-adjusted power curves for N=15, 25, 100

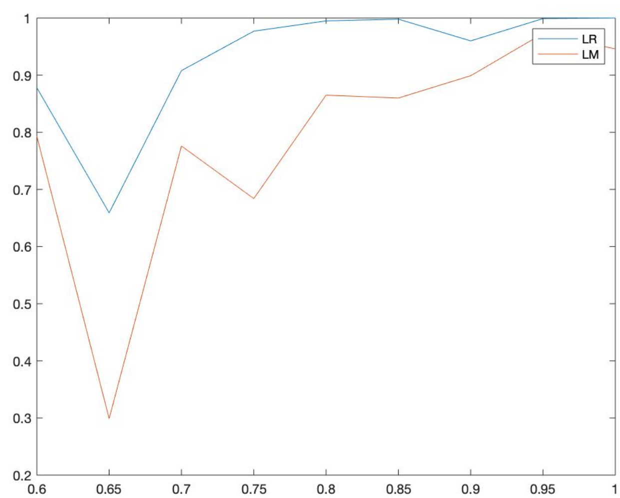


Figure 4: Size-adjusted power curves for N=15 for LR and LM test

4 Conclusion

Probability distributions in Figure 1, size properties in Figure 2 and power properties in Appendix B show very similar results between the LR and Wald on one side and LM and LM OPG on the other side, so this conclusion will only take into account the LR and LM test.

The size of the LM test is converging towards the desired value of 5 % faster than the size of the LR test in absolute sense. It is not recommended to compare these two absolute values as they move towards 5 % in a different direction and this has a different effect on test properties.

The larger sample models ($N = 25$ and $N = 100$) shows smooth size-adjusted power curves with the correct characteristics for all tests, this does not help distinguish between tests. Small sample properties of the two test do differ, as the LR test has higher size-adjusted power values than the LM test for $\gamma_a > 0.6$ and seems to perform better. Nevertheless, as the power envelope of these tests is unknown the LR test is not by definition better.

The big difference between the LR test and the LM test that has not been taken into account yet is the amount of estimation that is necessary. The LR test needs an estimation for two models (restricted and unrestricted), whereas the LM test only needs the restricted estimation. In conclusion, evaluating probability, size and power properties, the LR test seems to perform better in small sample models but if estimation of the unrestricted model is computationally exhausting the LM test might be beneficial.

Appendix A

A1 We have $y_i = x_i'\beta + u_i$ with $u_i \sim N\left(0, \sigma^2 \exp\{\gamma' z_i\}\right)$ so

$$f(y_i|x_i,\theta) = \frac{1}{\sqrt{2\pi\sigma^2 \exp\{\gamma' z_i\}}} e^{-\frac{1}{2}\left(\frac{(y_i - x_i'\beta)^2}{\sigma^2 \exp\{\gamma' z_i\}}\right)}$$

This gives a loglikelihood of

$$l(\theta) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\left(\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N}\gamma' z_i - \frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - x_i'\beta)^2}{\sigma^2 \exp\{\gamma' z_i\}}$$

A2 The estimator $\tilde{\theta}$ contains 3 variables $\tilde{\beta}, \widetilde{\sigma^2}$ and $\tilde{\gamma}$. The latter one is 0 by definition of the restriction, which results in a homoskedastic variance of all observations of y_i.

Filling this in the log-likehlihood gives:

$$\begin{aligned} l(\widetilde{\theta}_r) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i'\beta)^2 \\ &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \end{aligned}$$

Differentiating to β and σ^2 and setting these equal to 0 then leads to:

$$\begin{aligned} \frac{\partial l(\widetilde{\theta}_r)}{\partial \beta} &= \frac{1}{2\sigma^2}X'(y - X\beta) = 0 \\ X'y &= X'X\beta \\ \widetilde{\beta} &= (X'X)^{-1}X'y \\ \frac{\partial l(\widetilde{\theta}_r)}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{(y - X\beta)'(y - X\beta)}{2(\sigma^2)^2} = 0 \\ \widetilde{\sigma^2} &= \frac{1}{N}(y - X\beta)'(y - X\beta) \\ &= \frac{\widetilde{e}'\widetilde{e}}{N} \end{aligned}$$

This finally gives the restricted maximum likelihood estimator:

$$\widetilde{\theta}_r = \begin{pmatrix} (X'X)^{-1}X'y \\ \widetilde{e}'\widetilde{e}/N \\ 0 \end{pmatrix}$$

A3 The score vectors are equal to the derivative of the loglikelihood with respect to the first, second and third parameter of θ respectively. In this case, we calculate the score value and fill in the restricted values of $\tilde{\theta}$ afterwards to compare with given values.

First element:

$$s_1(\theta) = \frac{\partial l(\theta)}{\partial \beta} = \sum_{i=1}^{N} \frac{y_i - x_i'\beta}{\sigma^2 \exp{(\gamma' z_i)}} x_i$$

$$s_1(\tilde{\theta}) = \sum_{i=1}^{N} \frac{y_i - x_i'\tilde{\beta}}{\widetilde{\sigma^2} \exp{(0)}} x_i = \frac{1}{\widetilde{\sigma^2}} \sum_{i=1}^{N} \left(y_i - x_i'\widetilde{\beta} \right) x_i = \frac{1}{\widetilde{\sigma^2}} \widetilde{e}' X = 0$$

Second element:

$$s_2(\theta) = \frac{\partial l(\theta)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - x_i'\beta)^2}{(\sigma^2)^2 \exp{(\gamma' z_i)}}$$

$$s_2(\tilde{\theta}) = -\frac{N}{2\widetilde{\sigma^2}} + \frac{1}{2} \sum_{i=1}^{N} \frac{\left(y_i - x_i'\widetilde{\beta} \right)^2}{\left(\widetilde{\sigma^2} \right)^2 \exp(0)} \frac{\widetilde{e}'\widetilde{e}}{2 \left(\widetilde{\sigma^2} \right)^2} - \frac{N}{2\widetilde{\sigma^2}} = \frac{N\widetilde{\sigma^2}}{2 \left(\widetilde{\sigma^2} \right)^2} - \frac{N}{2\widetilde{\sigma^2}} = 0$$

Third element:

$$s_3(\theta) = \frac{\partial l(\theta)}{\partial \gamma} = -\frac{1}{2} \sum_{i=1}^{N} z_i + \frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - x_i'\beta)^2}{\sigma^2 \exp{(\gamma' z_i)}} z_i$$

$$s_3(\tilde{\theta}) = -\frac{1}{2} \sum_{i=1}^{N} \left(1 - \frac{\left(y_i - x_i'\widetilde{\beta} \right)^2}{\widetilde{\sigma^2} \exp 0} \right) z_i = -\frac{1}{2} \sum_{i=1}^{N} \left(1 - \frac{\widetilde{e_i}^2}{\widetilde{\sigma^2}} \right) z_i$$

A4 The Hessian exists of 9 elements:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta^2} & \frac{\partial^2 l}{\partial \beta \partial \sigma^2} & \frac{\partial^2 l}{\partial \beta \partial \theta} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 l}{\partial (\sigma^2)^2} & \frac{\partial^2 l}{\partial \sigma^2 \partial \theta} \\ \frac{\partial^2 l}{\partial \theta \partial \beta} & \frac{\partial^2 l}{\partial \theta \partial \sigma^2} & \frac{\partial^2 l}{\partial \theta^2} \end{bmatrix}$$

We derive these second derivatives under the null hypothesis

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta'} = -\sum_{i=1}^{N} \frac{\exp\{-\gamma' z_i\}}{\sigma^2} x_i x_i'$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \sigma^2} = -\sum_{i=1}^{N} (y_i - x_i'\beta) \frac{\exp\{-\gamma' z_i\}}{(\sigma^2)^2} x_i$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \gamma'} = -\sum_{i=1}^{N} (y_i - x_i'\beta) \frac{\exp\{-\gamma' z_i\}}{\sigma^2} x_i z_i'$$

$$\frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta'} = -\sum_{i=1}^{N} (y_i - x_i'\beta) \frac{\exp\{-\gamma' z_i\}}{\sigma^2} x_i'$$

$$\frac{\partial^2 \ell}{\partial \sigma^2 \partial \sigma^2} = -\sum_{i=1}^{N} (y_i - x_i'\beta)^2 \frac{\exp\{-\gamma' z_i\}}{(\sigma^2)^3} + \frac{1}{2}\frac{1}{(\sigma^2)^2}$$

$$\frac{\partial^2 \ell}{\partial \sigma^2 \partial \gamma'} = -\sum_{i=1}^{N} \frac{1}{2} (y_i - x_i'\beta)^2 \frac{\exp\{-\gamma' z_i\}}{(\sigma^2)^2} z_i'$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \beta'} = -\sum_{i=1}^{N} \frac{1}{2} (y_i - x_i'\beta)^2 \frac{\exp\{-\gamma' z_i\}}{\sigma^2} z_i' x_i$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \sigma^2} = -\sum_{i=1}^{N} \frac{1}{2} (y_i - x_i'\beta)^2 \frac{\exp\{-\gamma' z_i\}}{(\sigma^2)^2} z_i$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \gamma'} = -\sum_{i=1}^{N} \frac{1}{2} (y_i - x_i'\beta)^2 \frac{\exp\{-\gamma' z_i\}}{\sigma^2} z_i z_i'$$

A5 The information matrix under H_0 is:

$$I_N = \begin{pmatrix} \frac{1}{\sigma^2} X'X & 0 & 0 \\ 0 & \frac{1}{2}\frac{N}{\sigma^4} & \frac{1}{2\sigma^2}\iota' Z \\ 0 & \frac{1}{2\sigma^2} Z'\iota & \frac{1}{2} Z'Z \end{pmatrix}$$

Any matrix A can be written as:

$$A = \begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

With the inverse:

$$A^{-1} = \begin{pmatrix} E^{-1} + E^{-1}FS^{-1}GE^{-1} & -E^{-1}FS^{-1} \\ -S^{-1}GE^{-1} & S^{-1} \end{pmatrix}$$

Where $S = H - GE^{-1}F$.

In this case, the inverse of the information matrix can be written as:

$$A^{-1} = \begin{pmatrix} E^{-1} & 0 \\ 0 & H^{-1} \end{pmatrix}$$

Where $E = \frac{1}{\sigma^2} X'X$, so $E^{-1} = \sigma^2 (X'X)^{-1}$ and H is equal to

$$H = \begin{pmatrix} \frac{N}{2\sigma^4} & \frac{\iota' Z}{2\sigma^2} \\ \frac{Z'\iota}{2\sigma^2} & \frac{Z'Z}{2} \end{pmatrix}$$

Again, matrix H can be written as:

$$H = \begin{pmatrix} K & L \\ M & N \end{pmatrix}$$

with inverse:

$$H^{-1} = \begin{pmatrix} K^{-1} + K^{-1}LT^{-1}MK^{-1} & -K^{-1}LT^{-1} \\ -T^{-1}MK^{-1} & T^{-1} \end{pmatrix}$$

Where $T = N - MK^{-1}L$, so:

$$T = \frac{Z'Z}{2} - \frac{Z'\iota}{2\sigma^2} * \frac{2\sigma^4}{N} * \frac{\iota'Z}{2\sigma^2} = \frac{Z'Z}{2} - \frac{Z'\iota\iota'Z}{2N} = \frac{1}{2}Z'(I_N - \frac{1}{N}\iota\iota')Z = \frac{1}{2}Z'\bar{P}_\iota Z$$

So $T^{-1} = 2(Z'\bar{P}_\iota Z)^{-1}$.

$$\begin{aligned}
{H^{-1}}_{1,1} &= \frac{2\sigma^4}{N} + \frac{2\sigma^4}{N} * \frac{\iota'Z}{2\sigma^2} * 2(Z'\bar{P}_\iota Z)^{-1} * \frac{Z'\iota}{2\sigma^2} * \frac{2\sigma^4}{N} \\
&= \frac{2\sigma^4}{N} + \frac{1}{N}\iota'Z(Z'\bar{P}_\iota Z)^{-1}Z'\iota * \frac{2\sigma^4}{N} \\
&= \frac{2\sigma^4}{N}\left(1 + \frac{1}{N}\iota'Z(Z'\bar{P}_\iota Z)^{-1}Z'\iota\right) \\
{H^{-1}}_{1,2} &= -\frac{2\sigma^4}{N} * \frac{\iota'Z}{2\sigma^2} * 2(Z'\bar{P}_\iota Z)^{-1} \\
&= -\frac{2\sigma^2}{N} * \iota'Z * (Z'\bar{P}_\iota Z)^{-1} \\
{H^{-1}}_{2,1} &= -2(Z'\bar{P}_\iota Z)^{-1} * \frac{Z'\iota}{2\sigma^2} * \frac{2\sigma^4}{N} \\
&= -\frac{2\sigma^2}{N} * (Z'\bar{P}_\iota Z)^{-1} * Z'\iota \\
{H^{-1}}_{2,2} &= 2(Z'\bar{P}_\iota Z)^{-1}
\end{aligned}$$

This leads to the following inverser matrix:

$$H^{-1} = \begin{pmatrix} \frac{2\sigma^4}{N}\left(1 + \frac{1}{N}\iota'Z(Z'\bar{P}_\iota Z)^{-1}Z'\iota\right) & -\frac{2\sigma^2}{N} * \iota'Z * (Z'\bar{P}_\iota Z)^{-1} \\ -\frac{2\sigma^2}{N} * (Z'\bar{P}_\iota Z)^{-1} * Z'\iota & 2(Z'\bar{P}_\iota Z)^{-1} \end{pmatrix}$$

Fillig this in the A^{-1} matrix gives the inverse of the information matrix:

$$I_N^{-1} = \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 & 0 \\ 0 & \frac{2\sigma^4}{N}\left(1 + \frac{1}{N}\iota'Z(Z'\bar{P}_\iota Z)^{-1}Z'\iota\right) & -\frac{2\sigma^2}{N} * \iota'Z * (Z'\bar{P}_\iota Z)^{-1} \\ 0 & -\frac{2\sigma^2}{N} * (Z'\bar{P}_\iota Z)^{-1} * Z'\iota & 2(Z'\bar{P}_\iota Z)^{-1} \end{pmatrix}$$

From part 3 of the appendix we know that:

$$s'_N(\widetilde{\theta}_r) = \begin{pmatrix} 0 \\ 0 \\ -\frac{1}{2}(1-\frac{e'e}{\sigma^2})\iota' Z \end{pmatrix}$$

This is used to compute the LM statistic, $LM = (s'_N)\left(I_N^{-1}\right)(s_N)$. In the score vector and the inverse information matrix a number of 0's can be observed. The LM statistic then ends up being a product of the thrid element of the score vector and the right bottom element of the inverser information matrix:

$$\begin{aligned} LM &= -\frac{1}{2}(1-\frac{e'e}{\sigma^2})\iota' Z * 2(Z'\bar{P}_\iota Z)^{-1} * Z'\iota - \frac{1}{2}(1-\frac{e'e}{\sigma^2}) \\ &= \frac{1}{2}(1-\frac{e'e}{\sigma^2})\iota' Z * (Z'\bar{P}_\iota Z)^{-1} * Z'\iota(1-\frac{e'e}{\sigma^2}) \end{aligned}$$

This has the form:

$$\tfrac{1}{2} f' Z \left(Z'\overline{P_l} Z\right)^{-1} Z' f$$

with $f = \iota(1-\frac{e'e}{\sigma^2})$.

The LM statistic can be obtained as N times the uncentered R^2 with the auxiliary regression:

$$\iota = \tilde{\mathbf{s}}_i'\gamma + v_i$$

where $\widetilde{\mathbf{s}}_i = s_i\left(\widetilde{\boldsymbol{\theta}}_r\right)$, and computing:

$$\mathrm{LM}^* = NR_u^2$$

Appendix B

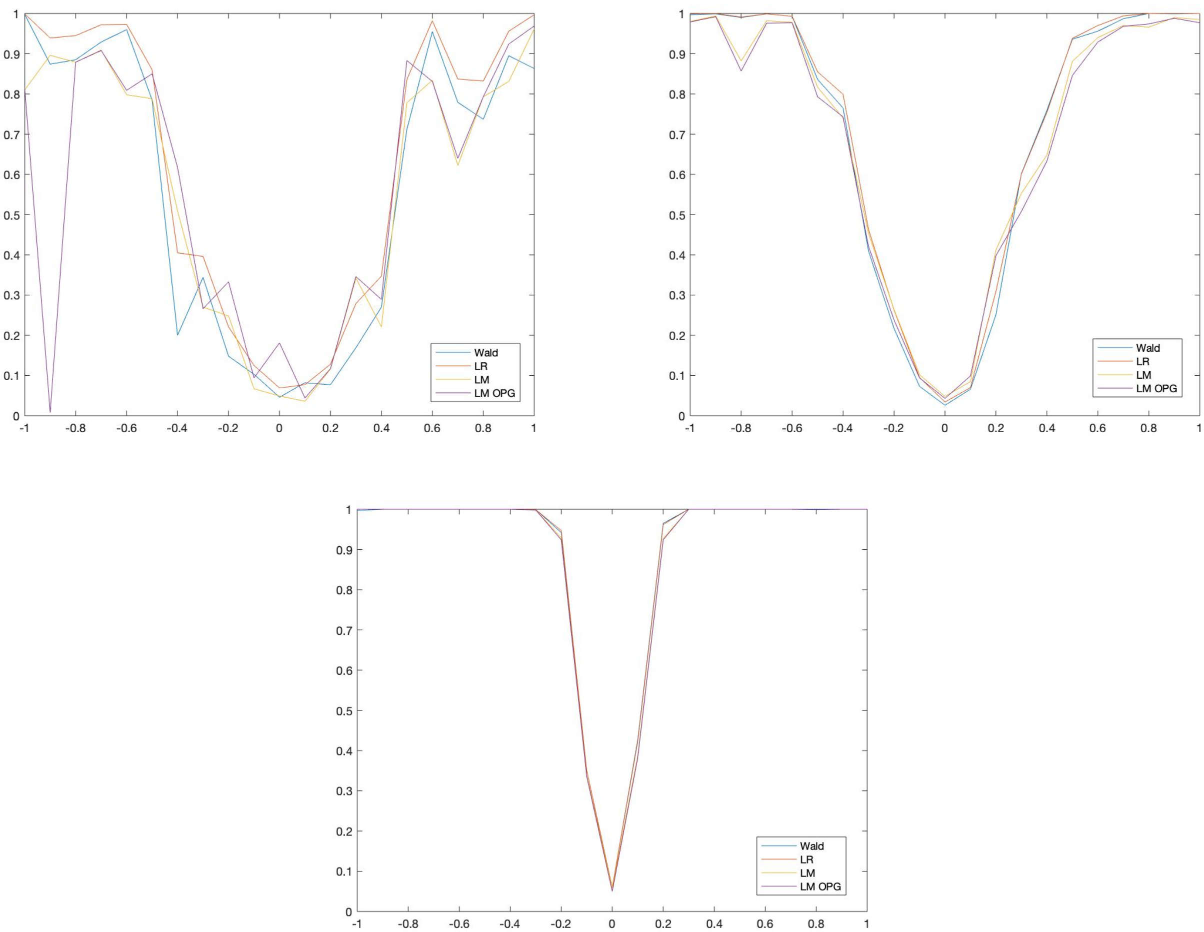


Figure 5: Size-adjusted power curves for N=15, 25 and 100

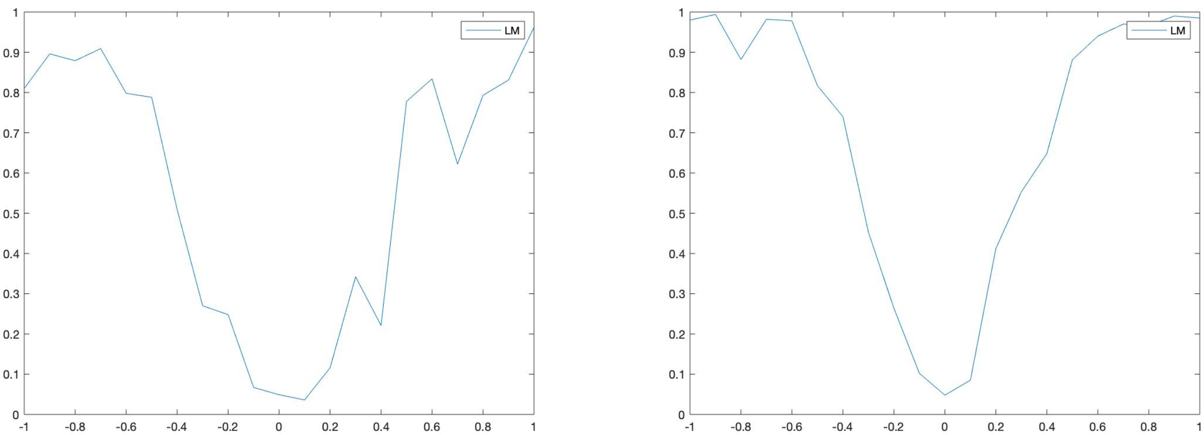


Figure 6: Size-adjusted power curves for LM statistic, N=15,25

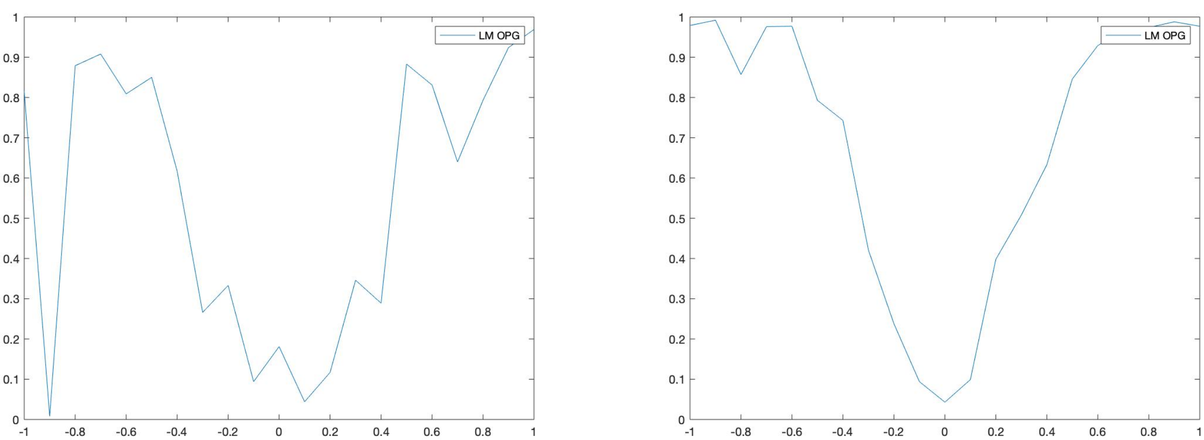


Figure 7: Size-adjusted power curves for LM OPG statistic, N=15,25

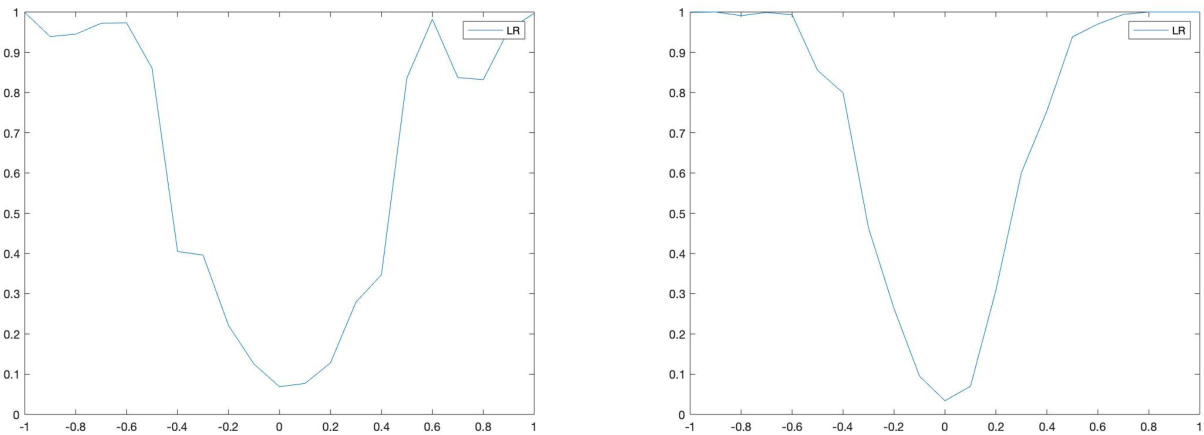


Figure 8: Size-adjusted power curves for LR statistic, N=15,25

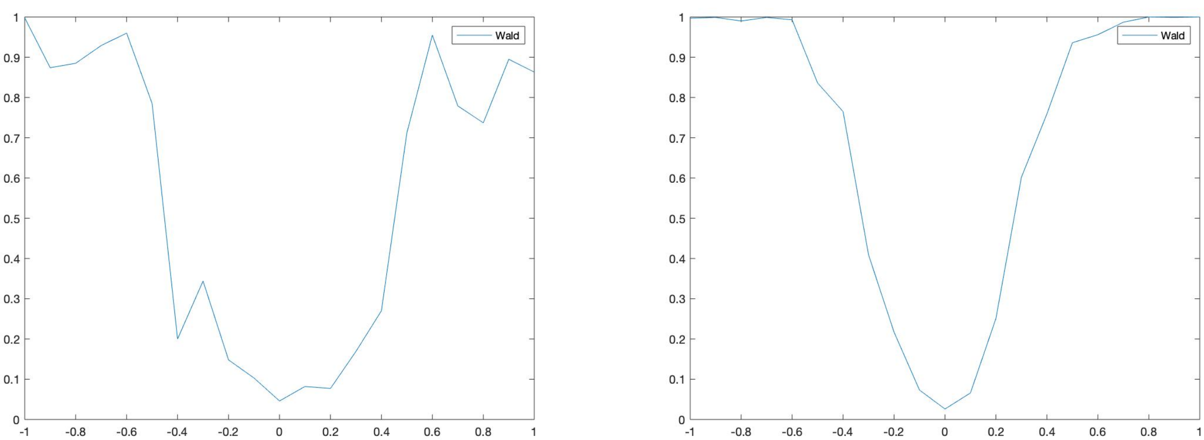
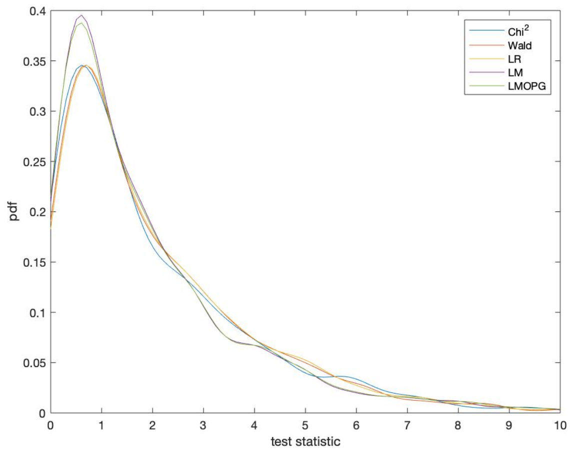


Figure 9: Size-adjusted power curves for Wald statistic, N=15,25



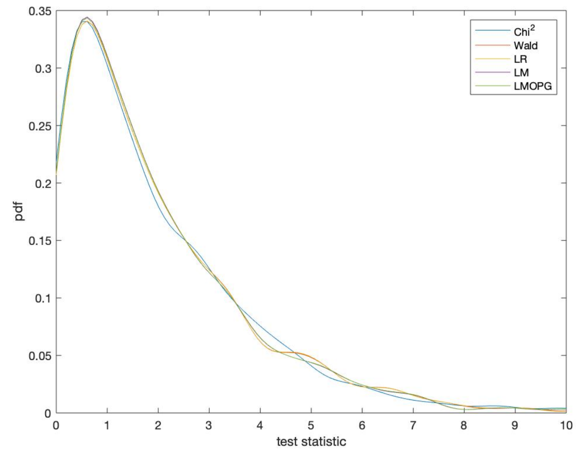


Figure 10: Probability distributions for N=100, 1000