EVALUATION OF THE GOLDFELD-QUANDT TEST AND ALTERNATIVES

A Thesis
Submitted to the Department of Economics
and the Institute of Economics and Social Sciences of
Bilkent University

In Partial Fullfillment of the Requirements
for the Degree of

MASTER OF ARTS IN ECONOMICS

by

Kerem Tomak

June, 1994

EVALUATION OF THE GOLDFELD-QUANDT TEST AND ALTERNATIVES

A Thesis
Submitted to the Department of Economics
and the Institute of Economics and Social Sciences of
Bilkent University
In Partial Fullfillment of the Requirements
for the Degree of

MASTER OF ARTS IN ECONOMICS

by

Kerem Tomak
June, 1994

KEREM TOMAK

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

Prof.Dr.Asad Zaman

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

Assoc.Prof.Dr.Osman Zaim

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

Assoc.Prof.Dr.Andrzej Sokolowski

Approved by the Institute of Social and Economic Sciences Acting Director:

Nur Bilge Criss

ABSTRACT

KEREM TOMAK
MA in Economics
Supervisor: Prof. Dr. Asad Zaman
June 1994

In this study, the widely used Goldfeld-Quandt test for heteroskedasticity in the linear regression model is evaluated. We reduce the dimension of the data space that is needed for the computation of the tests. We then compare the performances of the Likelihood Ratio and the Goldfeld-Quandt tests by using stringency measure. The problem of analytically non-tractable distribution function in the case of the Likelihood Ratio test is overcome by employing Monte Carlo methods.

It is observed that the Likelihood Ratio test is better in most of the cases than the Goldfeld-Quandt test.

Key Words: Heteroskedasticity, linear regression model, Goldfeld-Quandt test, stringency measure, Likelihood Ratio test, Monte Carlo methods, Kullback-Liebler Distance.

ÖZET

KEREM TOMAK
Yüksek Lisans Tezi, İktisat Bölümü
Tez Yöneticisi: Prof. Dr. Asad Zaman
Haziran 1994

Bu çalışmada, doğrusal regresyon modelinde heteroskedastisite durumunun test edilmesinde sık kullanılan Goldfeld-Quandt testi değerlendirilmektedir. Testler için gerekli veri uzayının büyüklüğü azaltılmıştır. Daha sonra Likelihood Ratio ve Goldfeld-Quandt testlerinin performansları sıkılık ölçüsü kullanılarak karşılaştırılmıştır. Likelihood Ratio testi hesaplanırken analitik olarak yazılamayan dağılım fonksiyonu problemi, Monte Carlo metodları kullanılarak çözülmüştür.

Hemen hemen bütün durumlarda Likelihood Ratio testinin Goldfeld-Quandt testinden daha iyi olduğu gözlemlenmiştir.

Anahtar Kelimeler: Heteroskedastisite, doğrusal regresyon modeli, Goldfeld-Quandt testi, sıkılık ölçümü, Likelihood Ratio testi, Monte Carlo metodları, Kullback-Liebler Uzaklığı.

Acknowledgements

I would like to express my gratitude to Prof. Dr. Asad Zaman for his inspiring and valuable supervision, his patience and for teaching me the subject.

I also would like to thank Assoc. Prof. Osman Zaim and Assoc. Prof. Andrzej Sokolowski for their valuable comments.

Special thanks go to Professor Sydney Afriat who convinced me about the necessity of completing a work started and to Professor Semih Koray who introduced me to the world of Economics.

I also thank my family for their understanding and care during the hard and sometimes frustrating learning period and Fethiye Akbulut for her valuable help during the literature survey.

Contents

1 Introduction

The linear regression model is the most frequently employed way of describing and analyzing data, and making predictions in economics, statistics and various other disciplines. Due to this popularity, as it is in other "popular" methods, it is commonly used in cases where its main assumptions are not satisfied. Very often certain predictions are made and prescriptions written for data at hand, which have very little, if not at all, to do with the real life.

> "... As a consequence, recent years have witnessed a remarkable growth of interest in testing - rather than estimating - econometric models. While it took more than a quarter of a century for the first serious article on testing to appear in Econometrica (the Chow test in 1960), the predominance of testing among articles in theoretical econometrics can hardly be overlooked in more recent volumes. In the 1980-1984 period alone, about fifty articles and notes appeared with a focus on testing econometric models."[13]

The disturbance structure of the regression models has been studied at length for a long time. Tests for normality, homoskedasticity and independence of disturbances and ways to overcome the difficulties caused by the existence or non existence of such concepts have been suggested and discussed in the literature and textbooks.

In our study, we will concentrate on one of the misbehaviours, namely heteroskedasticity of the error terms in the linear regression model. We are familiar with the standard regression model:

$$y_t = \beta' x_t + \epsilon_t,$$

where for $t = 1, 2, \ldots, T$, y_t are scalar observations on the dependent variable, β is $K \times 1$ vector of unknown parameters, x_t is a $K \times 1$ vector of the regressors, and ϵ_t is the unobserved error term. The problem of heteroskedasticity arises when the errors have different variances so that ϵ_t are independent but not identically distributed: $\epsilon_t \overset{\text{i.n.i.d.}}{\sim} N(0, \sigma_t^2)$. In this case, while the OLS estimates are consistent, they fail to be efficient. A more serious problem is that the usual regression statistics t and F are biased.

The first step in solving problems posed by heteroskedasticity is of course the detection of heteroskedasticity. For this purpose Goldfeld and Quandt suggested a test which has proven to be very popular and is widely used in applications today. Our goal in this paper is to suggest that this test is seriously deficient and to suggest improved alternatives and study the performance of these tests by Monte Carlo studies for a number of models.

2 Statistical and Econometric Background

In this section, we will follow the standard statistical definitions and theorems which can be found in any statistics textbook (see for example, [9]). A considerably large portion of econometric theory is devoted to the study of how a certain variable y_t is related to the values of some other variables x_t. The following relationship is central in many studies:

$$y_t = f\left(x_{1t}, x_{2t}, \ldots, x_{mt}\right) \text{ for } t = 1, 2, \ldots, T$$

If T is taken to be sufficiently large, and several of the variables are observed, we can estimate the function f. Hypotheses regarding the form of f is the goal of many econometric studies.

Throughout the text, we will employ the following model:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + \epsilon_t \text{ for } t = 1, 2, \ldots, T$$

In matrix form, we can write the above model as follows:

$$y = X\beta + \epsilon$$

where y is a $T \times 1$ vector of observations on the dependent variable, X is a $T \times k$ matrix of observations on the independent variables, β is a $k \times 1$

vector of unknown coefficients, and ϵ is an unobserved error term. Following assumptions about this model are standard:

1. ϵ_t are normally distributed, $\epsilon_t \sim N(0, \sigma^2 I_T)$
2. ϵ_t are independent,
3. ϵ_t have zero means and equal variances (*homoskedasticity* of the errors).
4. the regressors x_{kt} are independent of the error terms ϵ_t.

Under these assumptions, our model is called the linear regression model under the fully ideal conditions.

The implicit assumptions here are that the regression parameters do not change over time, that there are no measurement errors in x, that all relevant regressors are indeed included in the equation, and that the relationship between the dependent and the independent variables is indeed of a linear form. None of these assumptions is obvious in most applications, and the tests to be discussed in this study are supposed to determine whether in a given situation they hold or not [13].

The main concern of a regression problem is to estimate the parameters β and σ^2 and under the fully ideal conditions, these estimates are called the maximum likelihood estimates. Assuming that $X^t X$ is positive definite, the unique solution for β is given by

$$\widehat{\beta} = \left(X^t X\right)^{-1} X^t y$$

where $E\left(\widehat{\beta}\right) = \beta$ and $cov(\widehat{\beta}) = \sigma^2 \left(X^t X\right)^{-1}$.

The following definitions are in order:

Definition 1 *An estimator g is said to be unbiased for θ if $E(g) = \theta$.*

Definition 2 *If g is an unbiased estimator and it has the minimum variance in the class of unbiased estimators, g is said to be an efficient estimator.*

Definition 3 *Suppose that $\widetilde{\theta}_n$ is the estimator of θ based on a sample size of n. Then the sequence of estimators $\widetilde{\theta}_n$ is called a consistent sequence if*

$$\lim_{n\to\infty} P\left(\left|\widetilde{\theta}_n - \theta\right| < \epsilon\right) = 1.$$

The modified minimum variance unbiased estimate of σ^2 is given by:

$$\widehat{\sigma}^2_{MVU} = \frac{1}{T-K} \left\| y - X\widehat{\beta} \right\|^2.$$

The densities of these estimates are given in the following theorem:

Theorem 1 *The maximum likelihood estimators $\widehat{\beta}$ and $\widehat{\sigma}^2$ are independent with densities $\widehat{\beta} \sim N\left(\beta, \sigma^2 \left(X^t X\right)^{-1}\right)$ and $T\widehat{\sigma}^2/\sigma^2 \sim \chi^2_{T-k}$.*

Following facts are used in the proof of this theorem and we will also use them in our computations:

1. The notation $\alpha\chi_p^2$ is used to denote the density of a chi-square random variable after it has been multiplied by the scalar α. We say that Z has a Gamma density with parameters p and λ ($Z \sim G(p, \lambda)$) if

$$f^Z(z) = \frac{\lambda^p}{\Gamma(p)} z^{p-1} \exp(-\lambda z).$$

   In this notation, the chi-square density with p degrees of freedom, χ_p^2 is the same as $G(p/2, 1/2)$ and the scalar multiple $\alpha\chi_p^2$ is the same as $G(p/2, 1/(2\alpha))$. Thus the theorem states that $\widehat{\sigma}^2 \sim G((T-k)/2, 1/(2\sigma^2))$.

2. If $Y = AX + b$ then $Y \sim N(A\mu + b, A\Sigma A')$.

3. If $Y = AX$ and $Z = BX$ and $A\Sigma B' = 0$ then Y and Z are independent.

4. If $\mu = 0$ and $\Sigma = \mathbf{I}$ and M is an idempotent matrix, then $Y = X'MX \sim \chi_k^2$, where $k = trace(M)$.

Definition 4 *Suppose that we observe random variables (X, Y) with joint distribution $F^{(X,Y)}(x, y, \theta)$, and the distribution of $Y \mid X$ does not depend on the parameter of interest θ. In this case, we say that X is sufficient for θ.*

We can describe the hypothesis testing problem as follows. For the observations Y that come from a family of distributions $f^Y(y, \theta)$, indexed by the parameter $\theta \in \Theta$, the null hypothesis H_0 is that the parameter θ belongs to some subset Θ_0 of Θ, while the alternative is $H_1 : \theta \in \Theta_1$, where Θ_1 is a subset of Θ not intersecting with Θ_0. If we reject the null when it is true, this is called a type I error, while accepting the null when it is false is called a type II error. The probability of rejecting the null when the null is true is called the size or the significance level of the test. The probability of not making a type II error is called the power of the test.

The Neyman-Pearson Theorem describes the test of maximum power for any given size:

Theorem 2 *Suppose that the densities of the observation Y under the null and alternative hypothesis are $H_0 : Y \sim f^Y(y, \theta_0)$ and $H_1 : Y \sim f^Y(y, \theta_1)$. Fix $\alpha > 0$. Then there exists a constant c_α such that*

$$P(NP(Y) > c_\alpha \mid H_0) = \alpha,$$

where $NP(y) = \frac{f^Y(y,\theta_1)}{f^Y(y,\theta_0)}$. Define hypothesis test $\delta^*(y)$ by $\delta^*(y) = 1$ if $NP(y) > c_\alpha$ and $\delta^*(y) = 0$ if $NP(y) \leq c_\alpha$. Then this test has size α and if $\delta'(y)$ is a different test of smaller or equal size, then δ' has power less than the power of δ^*.

Except for the choice of level α or equivalently the constant c_α, the

Neyman-Pearson Theorem provides a complete solution to the hypothesis testing problem in the case of a simple versus simple hypothesis.

The concept of *stringency measure* is an important milestone in comparing the advantages of two hypothesis tests. We will use this entity in comparing various tests with each other so it is suitable to explain what it means and for that purpose, the following definitions are used throughout the text. Let X be an observation from a parametric family of densities $f(x, \theta)$. Suppose we wish to test $H_0 : \theta \in \Theta_0$ v.s. the alternative $H_1 : \theta \in \Theta_1$. Let T be the set of all hypothesis tests or equivalently, functions mapping X to the unit interval $[0, 1]$.

For any test $t \in T$, define $R(t, \theta)$ to be the probability of rejecting the null when θ is the true parameter. Then, the level (size) of the test $L(t)$ is defined to be the maximum probability of type I error:

$$L(t) = \sup_{\theta \in \Theta_0} R(t, \theta)$$

Let T_α be the set of all tests of size α. For any $\theta_1 \in \Theta_1$, the maximum possible power any test of size α can attain is given by the power envelope β_α^* defined as:

$$\beta_\alpha^*(\theta_1) = \sup_{T \in T_\alpha} R(T, \theta_1)$$

The shortcoming S of a test $T \in T_\alpha$ is measured with reference to $\beta^*_\alpha(\cdot)$:

$$S(T, \theta_1) = \beta^*_\alpha(\theta_1) - R(T, \theta_1)$$

It is convenient to define the deficiency of a test to be the negative of its maximum shortcoming:

$$S(T) = -\sup_{\theta_1 \in \Theta_1} R(T, \theta_1)$$

The deficiency of a test is the largest gap between the power curve of a test and the maximum possible power. The smaller the gap, the better the test. We will use the term stringency of a test for the negative of its deficiency.

Definition 5 *A test having the largest possible stringency in the set T_α is called a most stringent test of level α for testing H_0 against H_1.*

Having set a common ground for our analysis, it is now the time for explaining what heteroskedasticity means and how the previous work on detecting it evolved.

Heteroskedasticity problem arises when the error terms do not have constant variance for each observation in the regression problem, unlike one of our assumptions, namely homoskedasticity assumption, when we set the fully ideal conditions.

In that case, the $cov(\widehat{\beta}) = \Sigma \neq \sigma^2 \mathbf{I}$. Thus although $\widehat{\beta}$ is still unbiased and consistent, its covariance matrix is now:

$$cov\left(\widehat{\beta}\right) = \left(X^t X\right)^{-1} X^t \Sigma X \left(X^t X\right)^{-1}.$$

It is clear that one will make wrong inferences if he assumes that $\Sigma = \sigma^2 \mathbf{I}$. More than that, it is known that there are more efficient linear estimators when $\Sigma \neq \sigma^2 \mathbf{I}$.

A common example where heteroskedasticity is observed is the linear regression model where the dependent variable y represents consumption expenditures and the independent variable x represents income for several families. In this problem, if we estimate the problem by OLS, the residuals will be larger for larger values of x. Thus the error variances will not stay constant but increase with the value of x.

The consequences of heteroskedasticity are seen primarily on the least squares estimators. It is shown [15] that in the case of heteroskedasticity, the least squares estimators are still unbiased but inefficient. It is also shown that the estimates of the variances are also biased, thus invalidating the tests of significance. Thus it would be the case that under heteroskedasticity, we would get lower estimates than the true variance of the OLS estimator and our confidence intervals will be shorter than the true ones. Hypothesis for the regression coefficients would also be affected by this problem.

There have been various studies on testing for heteroskedasticity. The study of the presence of heteroskedasticity had mainly two directions. One direction led to only stating whether there was a problem of heteroskedasticity or not. Goldfeld and Quandt [7] named the tests suggested for this purpose as nonconstructive. The other direction also involved estimation by stating the form of heteroskedasticity. In our work, we will deal with the first direction and study the performance of the Goldfeld Quandt (GQ) test and try to suggest better alternatives.

A number of authors have studied the performance of the Goldfeld Quandt test under different forms of heteroskedasticity.

Evans and King [4] have recommended a new test and found that their suggested test was generally more powerful against medium and severe heteroskedasticity whereas the King and Szroeter tests performed better against weak heteroskedasticity.

They employed the linear regression model and tested $H_0 : \Sigma = I_n$ vs. $H_1 : \Sigma \neq I_n$. They study the form of heteroskedasticity given by

$$\sigma_t^2 = h\left(z_t'\alpha\right), \; t = 1, 2, \ldots, n$$

and they restrict the case to $s = 2$ in which σ_t^2 can be expressed as a kth power of some linear function of exogenous variable. They propose to reject

H_0 for small values of the statistic

$$s(\lambda^*) = \tilde{u}'\Sigma^{-1}(\lambda^*)\,\tilde{u}/\hat{u}'\hat{u}$$

where $\tilde{u}$ is the GLS residual vector assuming covariance matrix $\Sigma(\lambda^*)$. They assert that this test corresponds to the LR test against the alternative $\lambda = \lambda^*$.

In their experiments, they employed 9 different sets of data, 3 of which were artificial and the rest are real economic time series data. They ordered the regressors according to increasing values of the postulated deflator variable. They found out that against moderate or severe heteroskedasticity, tests based on OLS residuals alone were generally inferior. They also stated that

> "...the OLS based tests, the Harrison and McCabe and Goldfeld Quandt tests perform poorly because they largely ignore the form of the heteroskedasticity."

In another paper [5], they consider the problem of testing for heteroskedasticity in the linear regression model when one is willing to postulate only the ranking of the disturbance variances under the alternative hypothesis. They have compared the powers of various tests for heteroskedasticity of a given form. They computed the power of the GQ test by omitting 3 central observations for a sample size of 15, 4 for 20, 8 for 40 and 16 for 60.

They found out that the GQ test was better than Breusch and Pagan test and Harrison and McCabe test but worse than Szroeter and two other Evans and King tests.

Ali and Giacotto [1] asserted that the power of tests can be improved with the OLS residual estimates, the increased sample size and the variability of the regressors and it can be substantially reduced if the observations are not normally distributed. Each test that they studied was optimum to detect a specific form of heteroskedasticity. They also pointed out that a serious power loss might occur if the underlying heteroskedasticity assumption in the data generation deviated from it.

They carried out Monte Carlo experiments for power computations using 1000 replications. They performed the majority of their experiments with the following model:

$$y_t = \beta_0 + \beta_1 x_{t1} + u_t$$

where they set $\beta_0 = \beta_1 = 1$ (among all the authors that we have surveyed, they were the only ones to set equal regression coefficients across the two samples and they used this information for all the test statistic computations). They used 6 different data sets, three of which were stationary and the rest non-stationary. They experimented with 6 types of heteroskedasticity which had the following structure:

$$\sigma_t^2 = \sigma^2 g\left(z_t' \alpha\right).$$

They considered four distributions for u_t's. Finally, they experimented with three sample sizes: 10, 25 and 40 and three residual estimates, OLS, Recursive and BLUS.

Harvey and Phillips [8] proposed an exact parametric test against heteroskedasticity in the general linear model. They compared its power with Goldfeld Quandt and with BLUS. They found out that under a variety of circumstances all three tests were of comparable power. They employed two forms of heteroskedasticity and in their computations they used one variable for determining the variances of the disturbances and (an)other variable(s) for the constant term. They generated the variable associated with the disturbances once from a uniform and another time from a lognormal distribution and they kept the sample size at 20. They computed the powers of 3 tests including the Goldfeld Quandt test at the 5% significance level. They concluded that Goldfeld Quandt test was the worst among the three.

Griffiths and Surekha [16] found out that Szroeter's asymptotically normal test outperforms the Goldfeld Quandt test, the Breusch-Pagan Lagrange multiplier test and BAMSET, when it is possible to order the observations according to increasing variance. With no prior information on variance ordering, they concluded that BAMSET is the best. The highlight of their analysis was on the problem of ordering the observations according to increasing variances. They stated that when the observations are not ordered according to increasing variances, Goldfeld-Quandt, Szroeter and BAMSET tests loose their performances and BAMSET is the best among the three. Their findings deserve considerable interest for further studies on the other problem that we have side stepped, namely assuming strong prior information before calculating the GQ test. They set the regression coefficients to be different throughout their experiments and they used a linear regression model which included a constant term. They set the dimension of the regressor space to be equal to 2 and they considered two types variance structures referring to two types of heteroskedasticity, namely additive and multiplicative. They estimated the power of the tests by calculating the proportion of rejections in 5000 replications at a 5% level of significance.

Kadiyala and Oberhelman [12] suggested new tests for heteroskedasticity which were more favorable and they also quoted from Theil that "the Goldfeld Quandt test's use of the least squares residuals based on two separate subsets of the n observations implies that the analyst sacrifices twice the k degrees of freedom necessary to estimate the parameter vector, β. Thus it seems plausible that when it is assumed to be known that the two subsets have the same parameter vector, the power of the tests can be improved by taking this knowledge into account". But they still assume that the x's have been ordered in an ascending way and the disturbance variance changes monotonically with one of the independent variables or with a linear combination of them.

In general, we see that under different forms of heteroskedasticity, although Goldfeld Quandt is easy to compute - even without any computer work - it is deficient. Due to computational inefficiency, until now, there has been a tendency toward easily computed and at the same time more efficient tests for heteroskedasticity.

Almost all the authors cited above, have ignored the additional assumption of equal regression coefficients across the two samples and ignored the information contained in that assumption.

3 The Goldfeld-Quandt Test

The basic situation envisaged by Goldfeld and Quandt is as follows. Suppose that we are somehow able to order the observations so that the variances are increasing: $\sigma_1^2 \leq \sigma_2^2 \leq \ldots \leq \sigma_T^2$. GQ suggest that we should choose T_1 and T_2 such that $1 < T_1 < T_2 < T$, and divide the sample into two parts as follows. Define y_a, y_b to be the $T_1 \times 1$ and $(T - T_2 + 1) \times 1$ vectors $y_a = (y_1, y_2, \ldots, y_{T_1})'$ and $y_b = (y_{T_2}, \ldots y_T)'$. Let X_a and X_b be $T_1 \times K$ and $(T - T_2 + 1) \times K$ matrices of corresponding values of the regressors and ϵ_a and ϵ_b be the corresponding error vectors. Define $\widehat{\beta}_a = (X_a' X_a)^{-1} X_a' y_a$ and $\widehat{\beta}_b = (X_b' X_b)^{-1} X_b' y_b$ and let $SSR_a^2 = \left\| y_a - X_a \widehat{\beta}_a \right\|^2$ and $SSR_b^2 = \left\| y_b - X_b \widehat{\beta}_b \right\|^2$ be the sum of squared residuals for each half of the sample. While $T_1 \approx T/2$ and $T_2 \approx T_1 + 1$ are reasonable values, G & Q propose omitting a few observations from the middle to increase the contrast between the variances in the first half of the sample and that of the last half of the sample. They suggest that we should reject the null of homoskedasticity for large values of

$$GQ = \frac{SSR_b^2 / t_b}{SSR_a^2 / t_a}$$

This is an exact test since the GQ has an F distribution with $t_b \equiv T - (T_1 - K)$ and $t_a \equiv T_1 - K$ degrees of freedom under the null.

Two major sources of information are ignored in the GQ procedure. Taking these sources of information into account should improve the test. Our goal will be to study the extent of improvement that is possible.

The first information is that the regression coefficients are the same on both halves of the sample. In estimating β separately on each half of the regression, we are effectively using only half of the sample. This is obviously a major loss of information and will seriously reduce efficiency of estimation for β. Furthermore, this loss will persist in large samples, so that even asymptotically the efficiency of estimation will be impaired.

The ability to order the observations in order of increasing variance presupposes that we have some information about the source of the variance. Without such information it is impossible to order the observations in the manner required for the test. In applications, this information is nearly always in the form of some regressor say z_t and the GQ procedure is applied. It seems obvious however that if we have information about the source of the heteroskedasticity utilizing this information will also improve the testing procedure. This then is the second major source of inefficiency of the GQ test.

In our research we propose to isolate these two effects and study them separately since each problem has to be tackled in their own ways. Ordering the observations according to increasing variances attracts further investigation since it is by no means informative to accept some phenomenon before running the tests and draw conclusions out of these. But for our purposes, we will side step this problem and focus our attention to equal regression coefficients across the two samples case.

Consider the regression model separately on the two halves of the sample:

$$y_a = X_a\beta_a + \epsilon_a, \qquad y_b = X_b\beta_b + \epsilon_b.$$

Assume the $\epsilon_a \sim N(0, \sigma_a^2 I_{T_1})$ and $\epsilon_b \sim N(0, \sigma_b^2 I_{T-T_2+1})$. Zaman (1994) shows that the Goldfeld-Quandt statistic provides a UMP invariant test for the hypothesis $H_0 : \sigma_a^2 = \sigma_b^2$ versus the alternative $H_1 : \sigma_a^2 < \sigma_b^2$. This situation differs substantially from the one studied by Goldfeld and Quandt where the regression coefficients β_a and β_b are assumed to be different so there is no loss of information from not utilizing information relating to the source of change of variances.

The power of the Goldfeld Quandt test can easily be obtained from tables of the F distribution. Let $\alpha = \sigma_a^2/\sigma_b^2$ as before. Define $GQ = (taSSR_b^2)/(tbSSR_a^2)$. Then αGQ has an F distribution with ta and tb degrees of freedom. Choose c such that $P(GQ > c) = 0.01$ for example. Then the power of the level one percent test is obtained by calculating $P(GQ > c \mid \alpha) = P(\alpha GQ > \alpha c)$. This last probability is a tail probability of $F(ta, tb)$ and can be looked up in an F table.

4 The Model

In this section we propose to study the effect of adding the information that $\beta_a = \beta_b$. In fact, the GQ procedure is widely used in applications where this assumption is maintained, that the regression coefficients are the same across the two subsamples. It appears likely that utilizing this information will provide a test superior to the Goldfeld Quandt. We propose to study the amount of the gains available by comparing the power of GQ test with the likelihood ratio test using Monte Carlo methods.

Sufficient statistics for the model are $SSR_a^2 \sim \sigma_a^2 \chi_{ta}^2$ and $SSR_b^2 \sim \sigma_b^2 \chi_{tb}^2$ in addition to the OLS estimates $\widehat{\beta}_a \sim N\left(\beta_a, \sigma_a^2 \left(X_a' X_a\right)^{-1}\right)$ and also $\widehat{\beta}_b \sim N\left(\beta_b, \sigma_b^2 \left(X_b' X_b\right)^{-1}\right)$. The GQ test is based on the ratio $ta\, SSR_b^2 / tb\, SSR_a^2$ which is effectively a ratio of the estimates of the $\widehat{\sigma}_a^2$ and $\widehat{\sigma}_b^2$. However, when $\beta_a = \beta_b$, additional information about the variances can be obtained by looking at $\Delta \equiv \widehat{\beta}_a - \widehat{\beta}_b$. Then $\Delta \sim N\left(0, \sigma_a^2 \left(X_a' X_a\right)^{-1} + \sigma_b^2 \left(X_b' X_b\right)^{-1}\right)$. This can be simplified further as follows:

Let P be a nonsingular $K \times K$ matrix such that $PP' = \left(X_a' X_a\right)^{-1}$ so that $P^{-1} \left(X_a' X_a\right)^{-1} P'^{-1} = I_K$. Then if $\theta = P^{-1}\Delta$, we have $\theta \sim N\left(0, \sigma_a^2 I_K + \sigma_b^2 P^{-1} \left(X_b' X_b\right)^{-1} P'^{-1}\right)$. Let Q and Λ be the orthogonal and diagonal matrix of eigenvectors and eigenvalues of $P^{-1} \left(X_b' X_b\right)^{-1} P'^{-1}$ so that

$$P^{-1} \left(X_b' X_b\right)^{-1} P'^{-1} = Q \Lambda Q'$$

Let $\psi = Q^{-1}\theta$. Then we have $\psi \sim N\left(0, \sigma_a^2 + \sigma_b^2\Lambda\right)$ where Λ is a diagonal matrix. Thus in addition to $SSR_a^2 \sim \sigma_a^2\chi_{ta}^2$ and $SSR_b^2 \sim \sigma_b^2\chi_{tb}^2$ we have the following information for $i = 1, 2, \ldots, K$

$$\psi_i^2 \sim \left(\sigma_a^2 + \sigma_b^2\lambda_i\right)\chi_1^2,$$

where all of these are independent of each other and also of SSR_a^2 and SSR_b^2. The value of this additional information depends on the λ_i.

Lemma 3 *λ_i are eigenvalues of the matrix $X_a'X_a\left(X_b'X_b\right)^{-1}$.*

Proof. By definition, since Λ is the diagonal matrix of eigenvalues of $A = P^{-1}\left(X_b'X_b\right)^{-1}P'^{-1}$, the roots of the following polynomial yields the diagonal elements of Λ :

$$\det\left(\Lambda\mathbf{I} - P^{-1}\left(X_b'X_b\right)^{-1}P'^{-1}\right) = 0$$

multiplying by $\det(P)$ from the left and by $\det(P')$ from the right yields:

$$\begin{aligned}
\det\left(P\Lambda\mathbf{I}P' - PP^{-1}\left(X_b'X_b\right)^{-1}P'^{-1}P'\right) &= 0 \\
\det\left(\Lambda\mathbf{I}\left(X_a'X_a\right)^{-1} - \left(X_b'X_b\right)^{-1}\right) &= 0 \\
\det\left(\Lambda\mathbf{I} - \left(X_a'X_a\right)\left(X_b'X_b\right)^{-1}\right) &= 0
\end{aligned}$$

So that Λ has the eigenvalues of $X_a'X_a\left(X_b'X_b\right)^{-1}$ on its diagonal. □

Now, the joint density of $\psi_i^* = \psi_i^2/SSR_a^2$, $i = 1,\ldots,K$, $\psi_{K+1}^* = SSR_b^2/SSR_a^2$ and $\psi_{K+2}^* = SSR_a^2$ will be calculated.

Theorem 4 *The joint density of $\psi_i^* = \psi_i^2/SSR_a^2$, $i = 1,\ldots,K$, $\psi_{K+1}^* = SSR_b^2/SSR_a^2$ and $\psi_{K+2}^* = SSR_a^2$ is given by*

$$f^Y(y_1, y_2, \ldots, y_{K+1}, \alpha) = \frac{\left((y_{K+1})^{tb/2-1}/\alpha^{tb/2}\right) * \prod_{i=1}^{K} 1/\left(\psi_i^*(\alpha+\lambda_i)\right)^{1/2}}{\left(1 + (y_{K+1}/\alpha) + \sum_{i=1}^{K} \psi_i^*/(\alpha+\lambda_i)\right)^{k/2+tb}} \Gamma(tb + K/2)$$

Proof. For the proof, set

$$X_i \sim \sigma_i^2 \chi_{n_i}^2 \equiv G\left(\frac{n_i}{2}, \frac{1}{2\sigma_i^2}\right) \tag{1}$$

where $\sigma_i^2 = \sigma_a^2 + \sigma_b^2\lambda_i$ for $i = 1, 2, \ldots, K$ and $\sigma_{K+1}^2 = \sigma_a^2$, $\sigma_{K+2}^2 = \sigma_b^2$ with $n_i = 1$, for $i = 1, 2, \ldots, K$, $n_{K+1} = ta$, $n_{K+2} = tb$. In the context of our first model, $X_i = \psi_i^2$, $i = 1, 2, \ldots, K$, and $X_{K+1} = SSR_a^2$, $X_{K+2} = SSR_b^2$.

Now, transform these variables as follows:

$$Y_i = X_i/X_{K+2}, \quad i = 1, 2, \ldots, K+1 \quad Y_{K+2} = X_{K+2}$$

The inverse transformation is:

$$X_i = Y_i Y_{K+2}, \quad i = 1, 2, \ldots, K+1$$
$$X_{K+2} = Y_{K+2}$$

and the Jacobian of this transformation is:

$$J = |Y_{K+2}|^{K+1}$$

By 1 and facts from the second chapter, the joint density of $X = (X_i)_{i=1,\dots,K+2}$ is:

$$f^X(x_1, x_2, \dots, x_{k+2}) = \prod_{i=1}^{K+2} \frac{(x_i/2\sigma_i^2)^{n_i/2-1}}{2\sigma_i^2\Gamma(n_i/2)} \exp\left(-\frac{x_i}{2\sigma_i^2}\right)$$

Thus the joint density of the transformation is:

$$\begin{aligned} f^Y(y_1, y_2, \dots, y_{k+2}) \;&= \prod_{i=1}^{K+1} \frac{(y_i y_{K+2}/2\sigma_i^2)^{n_i/2-1}}{2\sigma_i^2\Gamma(\frac{n_i}{2})} \exp(-y_i y_{K+2}/2\sigma_i^2) \times \frac{(y_{K+2}/2\sigma_b^2)^{tb/2}}{2\sigma_b^2\Gamma(tb/2)} \\ &\times |y_{K+2}|^{K+1} \exp(-y_{K+2}/2\sigma_b^2) \end{aligned}$$

Now, we need to calculate $\int f^Y(y_1, y_2, \dots, y_{k+2})\ dy_{K+2}$. For that purpose, we make the following change of variables: $y_{K+2} = 2\sigma_b^2 z$. Then,

$$\begin{aligned} &\int_0^\infty f^Y(y_1, y_2, \dots, y_{k+2})\ dy_{K+2} \\ &= \int_0^\infty \prod_{i=1}^{K+1} \frac{(y_i z\sigma_b^2/\sigma_i^2)^{n_i/2-1}}{2\sigma_i^2\Gamma(\frac{n_i}{2})} \exp(-y_i z\sigma_b^2/\sigma_i^2) \times \frac{(z)^{tb/2}}{2\sigma_b^2\Gamma(tb/2)} |2\sigma_b^2 z|^{K+1} \times \exp(-z)\, 2\sigma_b^2 dz \\ &= \int_0^\infty \prod_{i=1}^{K+1} \frac{1}{\tilde{\sigma}_i^2} (zy_i/\tilde{\sigma}_i^2)^{n_i/2-1} \exp\left(-z\left(1 + \sum_{i=1}^{K+1} y_i/\tilde{\sigma}_i^2\right)\right) z^{k+1+tb/2-1}\, dz \end{aligned}$$

where $\tilde{\sigma}_i^2 = \sigma_i^2/\sigma_b^2$.

Now, we make the following change of variables: For $A = 1 + \sum_{i=1}^{K+1} y_i/\tilde{\sigma}_i^2$, set $w = z * A$. Then, $\frac{dw}{A} = dz$, and

$$\int_0^\infty \prod_{i=1}^{K+1} \frac{1}{\tilde{\sigma}_i^2} \left(zy_i/\tilde{\sigma}_i^2\right)^{n_i/2-1} \exp\left(-z\left(1+\sum_{i=1}^{K+1} y_i/\tilde{\sigma}_i^2\right)\right) z^{k+1+tb/2-1}\, dz$$
$$= \int_0^\infty \prod_{i=1}^{K+1} \frac{1}{\tilde{\sigma}_i^2} \left(wy_i/A\tilde{\sigma}_i^2\right)^{n_i/2-1} \times \exp(-w)\,(w/A)^{k+1+tb/2-1}\, \frac{dw}{A}$$

But then, taking out the constants and w-independent terms, we have

$$\begin{aligned}
&\int_0^\infty \prod_{i=1}^{K+1} \frac{1}{\tilde{\sigma}_i^2} \left(wy_i/A\tilde{\sigma}_i^2\right)^{n_i/2-1} \times \exp(-w)\,(w/A)^{k+1+tb/2-1}\, \frac{dw}{A} \\
&= \frac{\prod_{i=1}^{K+1} y_i^{\frac{n_i}{2}-1}/(\tilde{\sigma}_i^2)^{n_i/2}}{A^{tb+k/2}} \int_0^\infty w^{tb+k/2} \times \exp(-w)\, dw \qquad (2) \\
&= \frac{\prod_{i=1}^{K+1} y_i^{\frac{n_i}{2}-1}/(\tilde{\sigma}_i^2)^{n_i/2}}{A^{k/2+tb}} \Gamma(tb+K/2)
\end{aligned}$$

So that,

$$f^Y(y_1, y_2, \ldots, y_{K+1}, \alpha) = \frac{\left((y_{K+1})^{tb/2-1}/\alpha^{tb/2}\right) * \prod_{i=1}^{K} 1/\left(\psi_i^*(\alpha+\lambda_i)\right)^{1/2}}{\left(1 + (y_{K+1}/\alpha) + \sum_{i=1}^{K} \psi_i^*/(\alpha+\lambda_i)\right)^{k/2+tb}} \Gamma(tb+K/2)$$

This concludes the proof. □

5 The Power Envelope

As argued in Zaman (1994B) tests can be evaluated effectively using the stringency measure. In order to measure stringency, we need to calculate the power envelope. Since the problem is scale invariant, it is convenient to define $\alpha = \sigma_a^2/\sigma_b^2$ and calculate the envelope for $H_0 : \alpha = 1$ versus the alternative $H_1 : \alpha > 1$. For a fixed $\alpha > 1$, the power of the Neyman- Pearson test gives the maximum possible power attainable. The Neyman-Pearson test for this scale invariant situation is calculated as follows:

The Neyman-Pearson most powerful test for $H_0 : \alpha = 1$ versus $H_1 : \alpha = \alpha_1 > 1$ rejects for large values of the statistic, after dropping the constants,

$$NP = \frac{f^Y(y \mid \alpha_1)}{f^Y(y \mid 1)}$$

where f^Y is calculated in the previous chapter (2). From this, it is easily seen that the Neyman-Pearson test rejects for large values of the following statistic:

$$S = \frac{1 + \sum_{i=1}^{K+1} \frac{\psi_i^*}{1+\lambda_i}}{1 + \sum_{i=1}^{K+1} \frac{\psi_i^*}{\alpha_1+\lambda_i}}$$

Plotting the power of these tests for different values of α_1 will generate the power curve.

6 The Likelihood Ratio Test

In addition to comparing the GQ test to the power envelope, we also wish to compare the likelihood ratio test to the power envelope. To compute the likelihood ratio test, we need to obtain the ML estimator for α. The first order condition for a maximum of the joint log-likelihood function (computed in the previous section) can be written as follows:

$$\sum_{i=1}^{K+1} \frac{n_i/2}{(\alpha+\lambda_i)} = \left(tb + \frac{K}{2}\right) \frac{\sum_{i=1}^{K+1} \frac{\psi_i^*}{(\alpha+\lambda_i)^2}}{1 + \sum_{i=1}^{K+1} \frac{\psi_i^*}{(\alpha+\lambda_i)}} \tag{3}$$

where $n_i = 1$ for $i \leq K$, $\lambda_{K+1} = 0$, and $n_{K+1} = tb$.

This equation can be solved by standard root finding procedures to obtain ML estimate $\widehat{\alpha}$.

Once we have the ML estimate, we can obtain the likelihood ratio statistic by plugging this into the likelihood ratio. The log-likelihood function for the data can be written as:

$$\begin{aligned} f(\psi^*, \alpha) = c + \sum_{i=1}^{K+1} \left(\tfrac{n_i}{2} - 1\right) \ln(\psi_i^*) - \sum_{i=1}^{K+1} \left(\tfrac{n_i}{2}\right) \ln(\alpha + \lambda_i) \\ - \left(tb + \tfrac{K}{2}\right) \ln\left(1 + \sum_{i=1}^{K+1} \tfrac{\psi_i^*}{(\alpha+\lambda_i)}\right) \end{aligned}$$

From this, the likelihood ratio statistic can be written as

$$LR(\psi^*, \widehat{\alpha}) = \frac{f(\psi^*, \widehat{\alpha})}{f(\psi^*, 1)}$$

The distribution of this is not tractable, but can easily be obtained by Monte Carlo. Similarly, we can trace out the power curve of this test using Monte Carlo methods and obtain stringency by comparing the power with the power envelope.

7 Algorithm

The computations were based on the parameters K, ta, tb and λ chosen as follows:

First, λ was fixed at a $1 \times K$ vector of ones. Then, K was fixed at 4, and 6 where the corresponding values of ta and tb were adjusted according to the formulas $T = 2 * (K + ta)$ and $ta = tb$. The sample sizes T were chosen to be $T = 20, 40$ and 60. The main focus of the computations was on the form of the λ_i's and we experimented with different values. The more the difference of the λ_i's, the more favorable LR test became. The experiments were held with Monte Carlo sample size of 5000.

For the computation of the power of the tests, we used two programs for the PC: Mathematica for Windows 2.2 and Gauss 386. In Gauss, the algorithms used to find out the root of the first order condition were bisection, secant, fixed point, false position and Newton-Raphson methods. Among these, the most efficient ones were secant and Newton-Raphson methods. In Mathematica, the computations took more time than they did in Gauss but the results were more accurate due to Mathematica's built-in facilities.

Overall time that is needed on a 486DX-33 with 8MB of RAM was around 2 hours with Gauss for the generation of a power curve of 10 points of the LR test with Monte Carlo sample size of 10000. With Mathematica it was a couple of hours longer.

The following algorithm, outlined by Zaman [18], was used to compute the power of the tests (the algorithm is for the Likelihood Ratio Test computations. For other tests, just omit the part where the root finding procedures are called):

1. Set the values for K, ta, tb, λ.

2. Set the significance level (to %10, %5, %1).

3. Set $\alpha = 1$ in order to generate numbers according to LR statistic under the null hypothesis.

4. Call GEN() procedure.

5. Create random numbers with distributions specified in the model.

6. Call MLBIS [ML] [SEC] [MLA] procedure to calculate the ML estimate by finding the solution of the first order condition for the maximum of the log-likelihood function, by employing the bisection method [Newton-Raphson method] [Secant method] [Hooke and Jeeves method]. The last method directly maximizes the log-likelihood function.

7. Calculate the LR statistic.

8. Goto the 5th step until you reach the previously specified Monte Carlo sample size (MCSS).

9. Exit GEN() procedure which returns a MCSS×1 vector of LR statistic values under the null hypothesis.

10. Sort the returned vector in an ascending manner. If the significance level is %10, set the critical value index to be 0.90*MCSS. If it is %5, set the critical value index to be 0.95*MCSS etc.. The approximation to the real critical value is then the LR statistic value in the returned vector at that index level.

11. Set $\alpha = n$, where n is the value of α at which we want to compute the power of the LR test.

12. Repeat steps 4-9, where now, the GEN() procedure returns a MCSS×1 vector of LR statistic values under the alternative hypothesis.

13. Calculate the number of times the elements of this last vector exceeds the critical value calculated at step 10 and divide this number by MCSS to get the power of the LR statistic.

8 Example

The following example is given by Maddala, [15], p. 165. The given data are the consumption expenditures (y) and income (x) for 20 families in thousands of dollars. Maddala shows graphically (Figure 5.1, p. 161) that there is a heteroskedasticity problem. In order to apply the Goldfeld-Quandt test, he divides the observations into two groups of ten each. The first group is $X_a = \{22.3, 12.1, 6.2, 10.3, 8.1, 14.1, 16.4, 24.1, 18.2, 20.1\}$ and the second group is $X_b = \{32.3, 36.6, 42.3, 44.7, 26.1, 40.2, 34.5, 38, 30.1, 28.3\}$. The estimated equations are:

$$y_a = 1.0533 + 0.876x_a$$

$$y_b = 3.279 + 0.835x_b$$

where $R_a^2 = 0.985$, $R_b^2 = 0.904$, $\sigma_a^2 = 0.475$, $\sigma_b^2 = 3.154$. For this set of equations, F-ratio for the GQ test is 6.64 and it is significant at the 1% level thus we reject the null hypothesis of homoskedasticity. For the LR test, the estimated α-value is 6.64, $\lambda = 0.2059$ and the approximate critical values are 0.1753, 0.1982, 0.2384 for 10%, 5% and 1% significance levels. The value of the LR statistic is computed as follows:

1. Calculate $\Delta = \widehat{\beta}_a - \widehat{\beta}_b = 0.041$;

2. Compute $(X_a' X_a)^{-1}$;

3. Get the cholesky decomposition of the matrix computed at step 2,

which, for this example, is 0.019471068;

4. Calculate $\theta = P^{-1}\Delta = 2.1056883$;

5. Get Q and Λ as the eigenvector and eigenvalue of $P^{-1}(X_b'X_b)P'^{-1}$, which, again for this example, are 1.000 and 0.2059;

6. Finally compute $\psi = Q^{-1}\theta$ and divide by R_b^2. For our example, we got $\psi^* = 2.3293$.

When you substitute these values into the LR statistic formula given in section 6 and use the estimated $\alpha-$value, namely 6.64, we get the LR statistic value, 1.04782. Thus LR statistic rejects the null hypothesis of homoskedasticity at all significance levels.

When the logarithmic form is considered, the estimated equations become:

$$y_a = 0.128 + 0.934x_a$$

$$y_b = 0.276 + 0.902x_b$$

and $R_a^2 = 0.992$, $R_b^2 = 0.912$, $\sigma_a^2 = 0.001596$, $\sigma_b^2 = 0.002789$.

In this case, F-ratio for the test is 1.75 and if we use the 5% significance level, we do not reject the null hypothesis. But if we consider the LR test, estimated $\alpha-$value is now 1.75, $\theta = 1.6435$, $\psi^* = 1.8020$, λ is the same, the critical values are 0.1764, 0.2009, 0.2331 for 10%, 5% and 1% significance levels respectively and the value of the LR statistic is 0.6089. Thus LR test still rejects the null hypothesis whereas GQ test does not reject it at 5% significance level. We calculated the approximate critical values for the LR test by running Monte Carlo on the ψ^* variable in our model with a Monte Carlo sample size of 10000.

9 Results

As we have stated earlier in the section on Goldfeld-Quandt test, there are two difficulties one of which is the ordering of the observations according to increasing variances under the alternative. In order to avoid this, we transformed this problem so that we only study the effects of the maintained hypothesis that $\beta_a = \beta_b$ on testing the null hypothesis $H_0 : \sigma_a^2 = \sigma_b^2$ versus the one-sided alternative $H_1 : \sigma_a^2 < \sigma_b^2$. In this case, the difficulty of ordering the observations does not arise since the variances are not correlated with any other variable.

Once we reduced the number of difficulties that we faced from two to one, and the dimension of the data space from $T \times K$ to $1 \times K + 1$, we had to deal with λ_i which are constants depending on the regressor matrices and which contain additional information. The number of different combinations of λ_i was too high so we first experimented with supposedly worst, best and some intermediate cases. We kept the number of regressors at 6, sample size at 20, 40, 60 and 100 and changed α between 1 and 10 with step size equal to 1. As we employed a reduced form of the problem, we only dealt with one parameter α which simplified the computations dramatically. After we were convinced that the gains were actually possible in line with our intuition, we concentrated on a serious problem, namely the structure of the λ_i's.

It was of utmost importance to have an idea about the structure of the

λ_i's since it would tell us the conditions under which it would be reasonable to use Likelihood Ratio test instead of Goldfeld-Quandt test. For this purpose, we calculated the Kullback-Liebler Distance to alpha from the null hypothesis. We generated random $\lambda \sim N(0,1)/U(0,1)$ and calculated the intervals that would correspond to different α increments. Then, for each new λ set that is generated, we calculated the power curves, KL Distance and tried to find out the worst and the best cases out of 10 generations. The Figures 1-6 represent the worst and best cases that we encountered at this point. It is our finding that for λ values that correspond to certain KL Distance intervals that we have calculated, one will end up with a power curve somewhere between these two representative curves, i.e. λ values that have the same KL Distance value will generate very similar LR power curves. Thus we conclude that Likelihood Ratio test's power is higher than Goldfeld-Quandt test's power - although slightly at some "bad" configurations of λ- and the gains from using LR test for heteroskedasticity when the structure of the λ's is studied, is of considerable interest.

Figure 1. Maximum LR Power vs. GQ Power at 10% Significance Level for T=20 lambda={1.241,0.9768,1.575,3.776,16.81,1224}

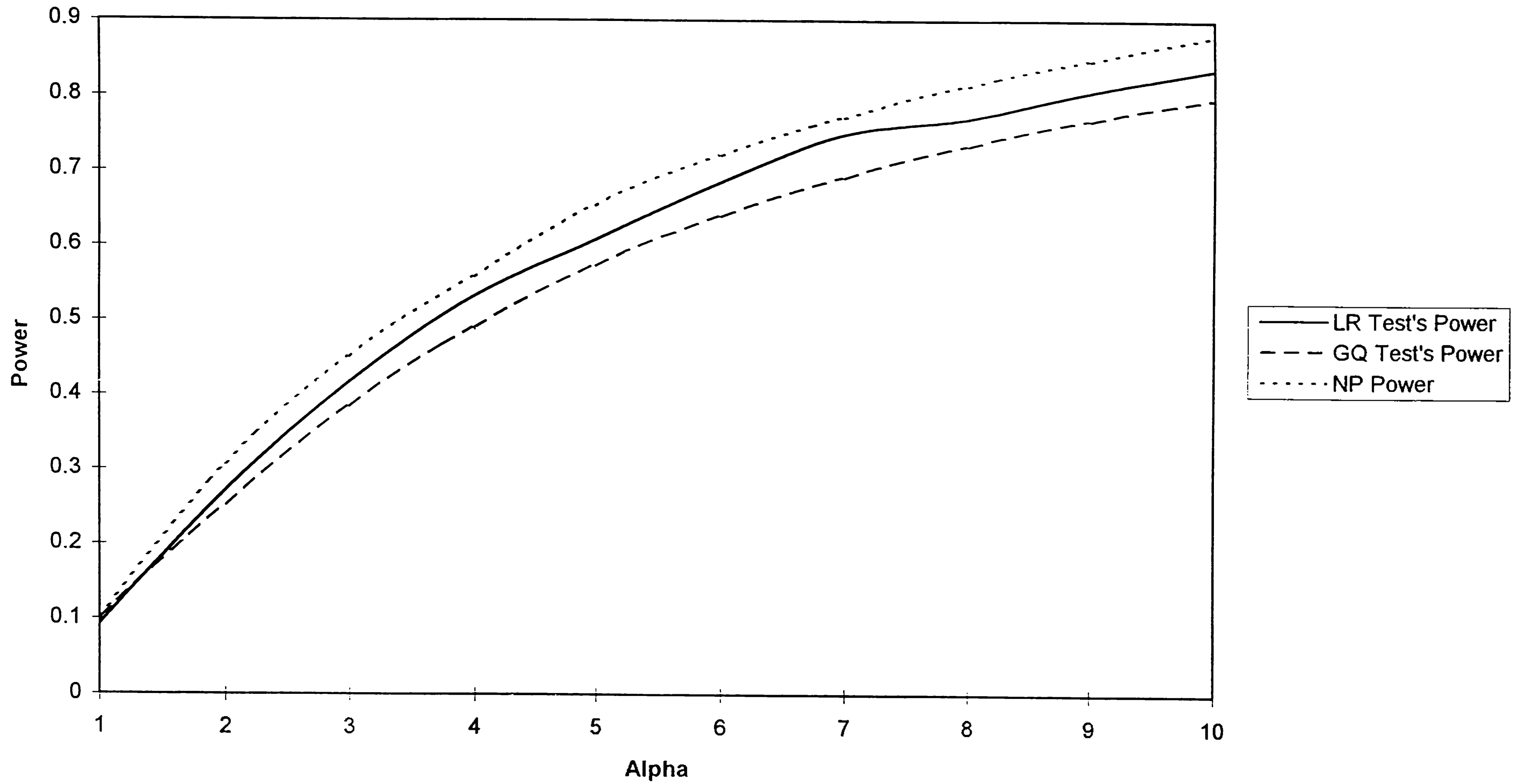


Figure 2. Maximum LR Power vs. GQ Power
at 10% Significance Level for T=40
lambda={1.241,0.9768,1.575,3.776,16.81,1224}

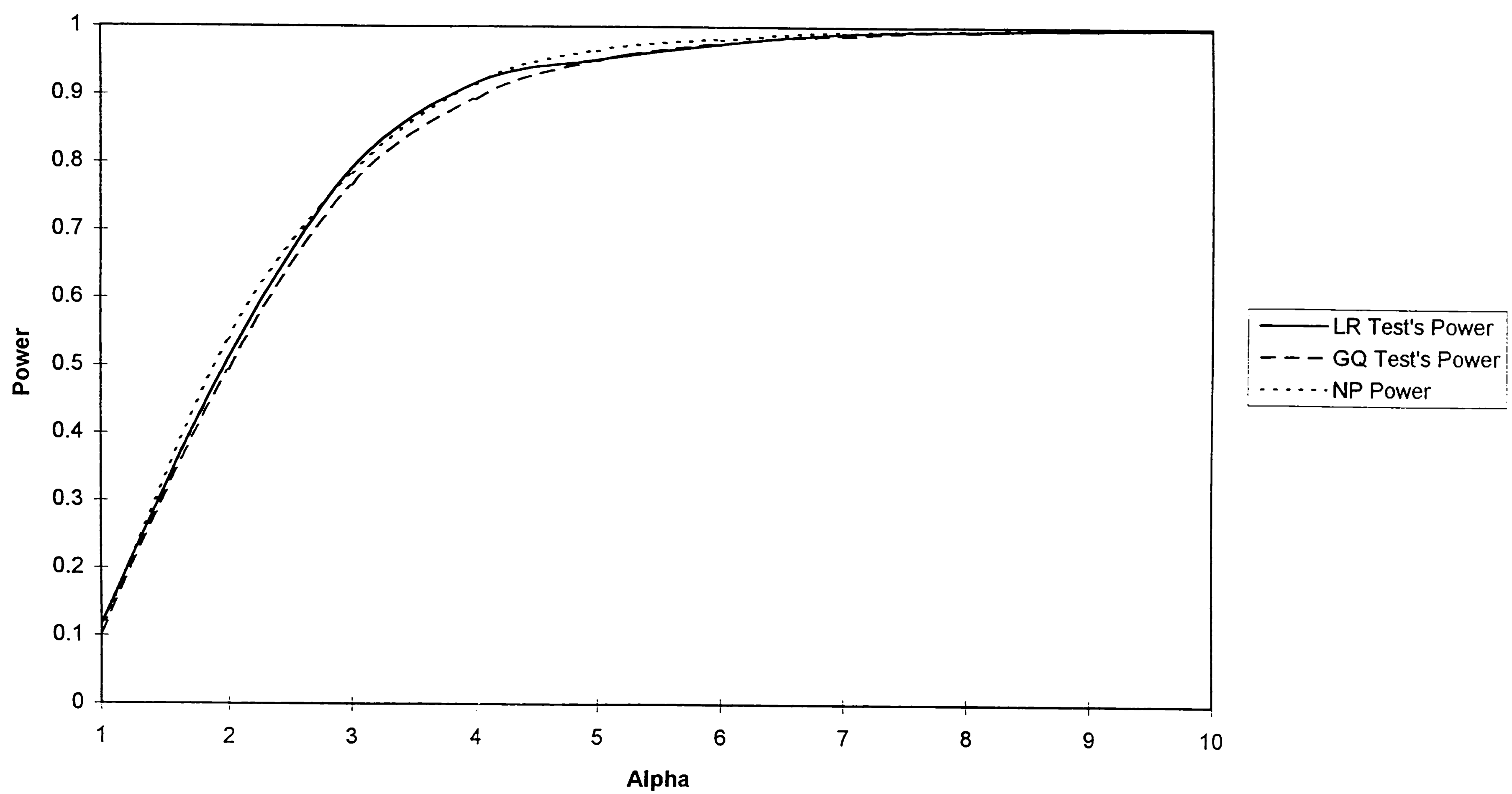


Figure 3. Maximum LR Power vs. GQ Power at 10% Significance Level for T=60 lambda={1.241,0.9768,1.575,3.776,16.81,1224}

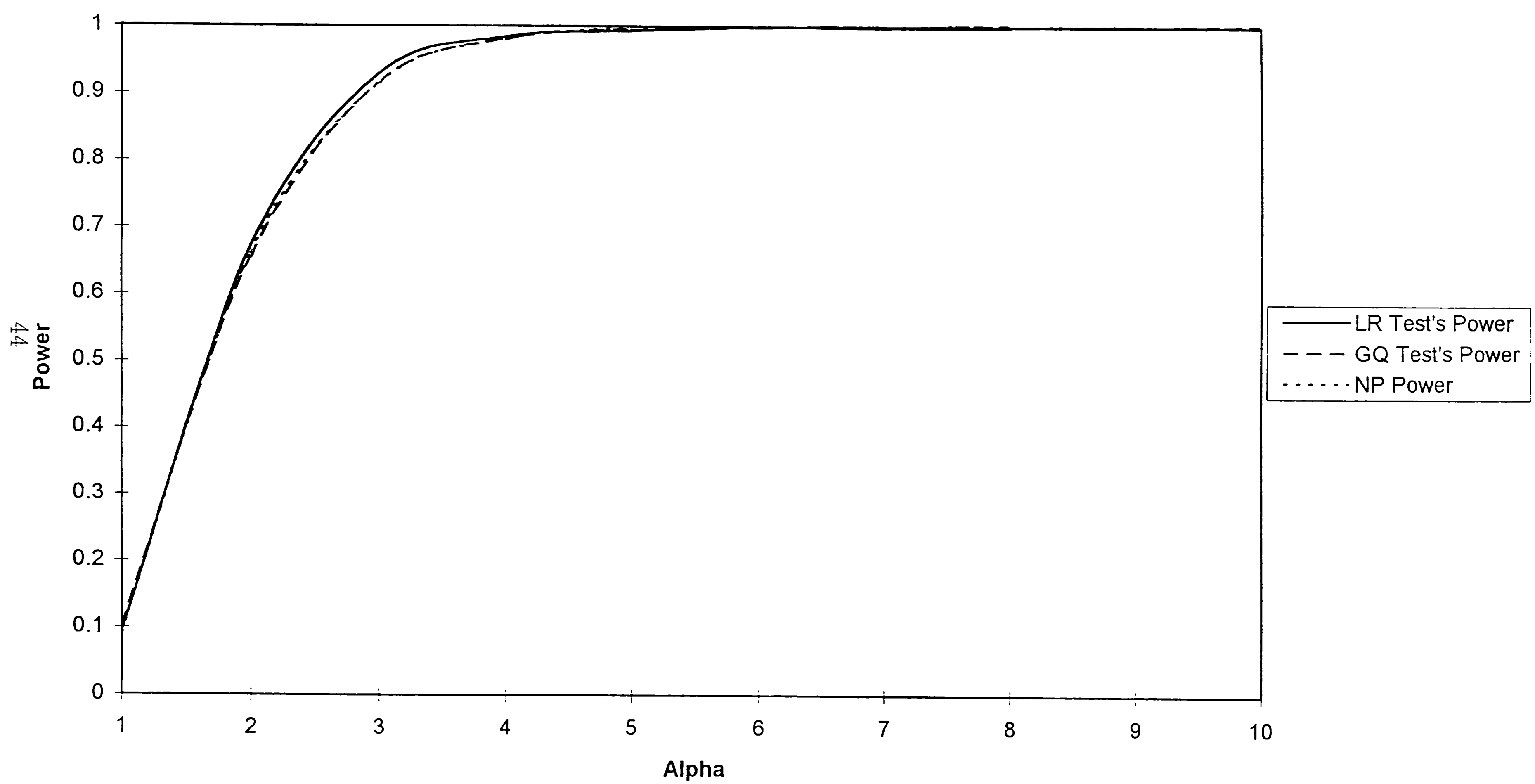


Figure 4. Minimum LR Power vs. GQ Power at 10% Significance Level for T=20 lambda={0.08712,6.949,1.297,1.117,3.394,0.3158}

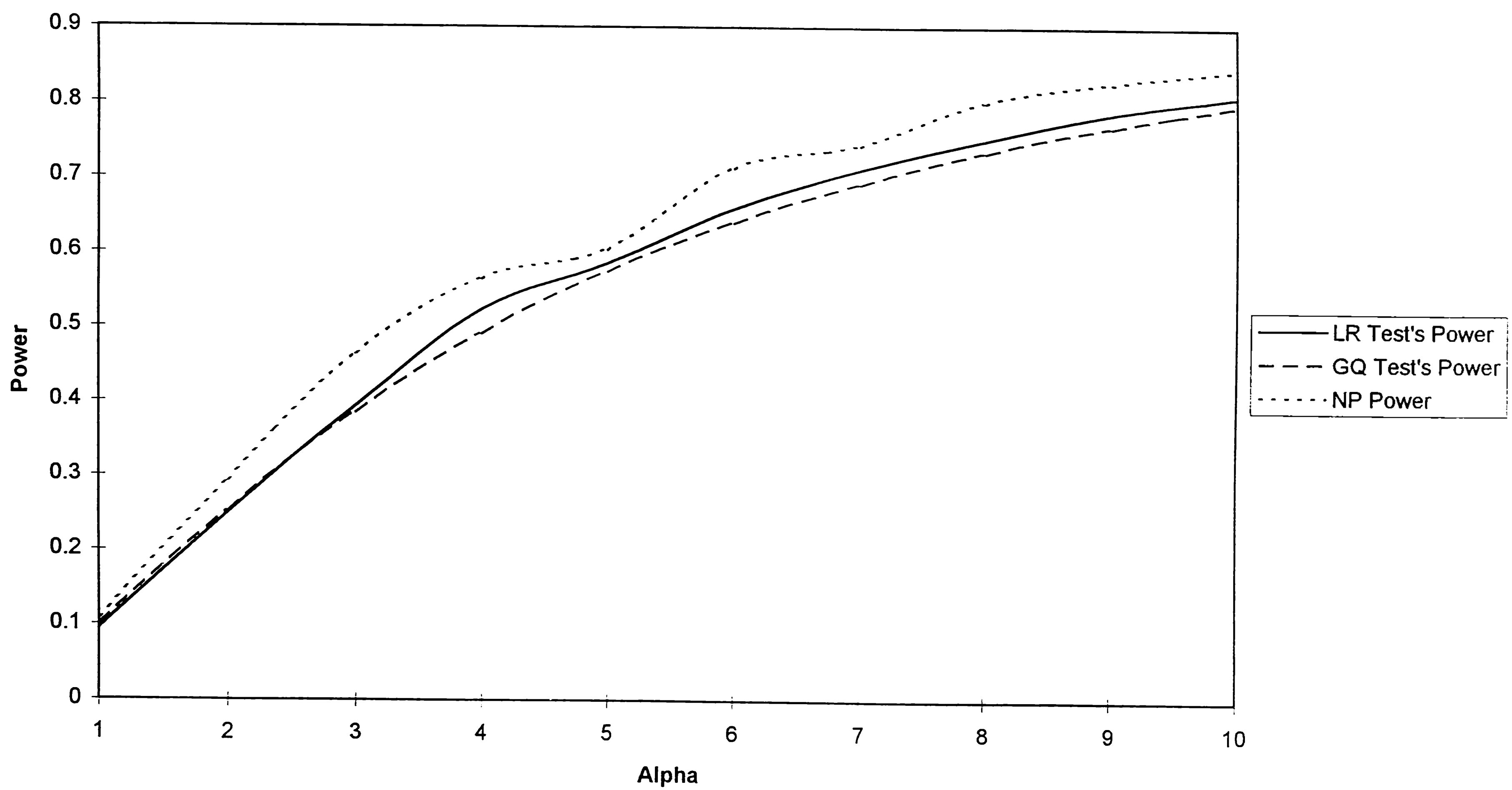


Figure 5. Minimum LR Power vs. GQ Power
at 10% Significance Level for T=40
lambda={0.08712,6.949,1.297,1.117,3.394,0.3158}

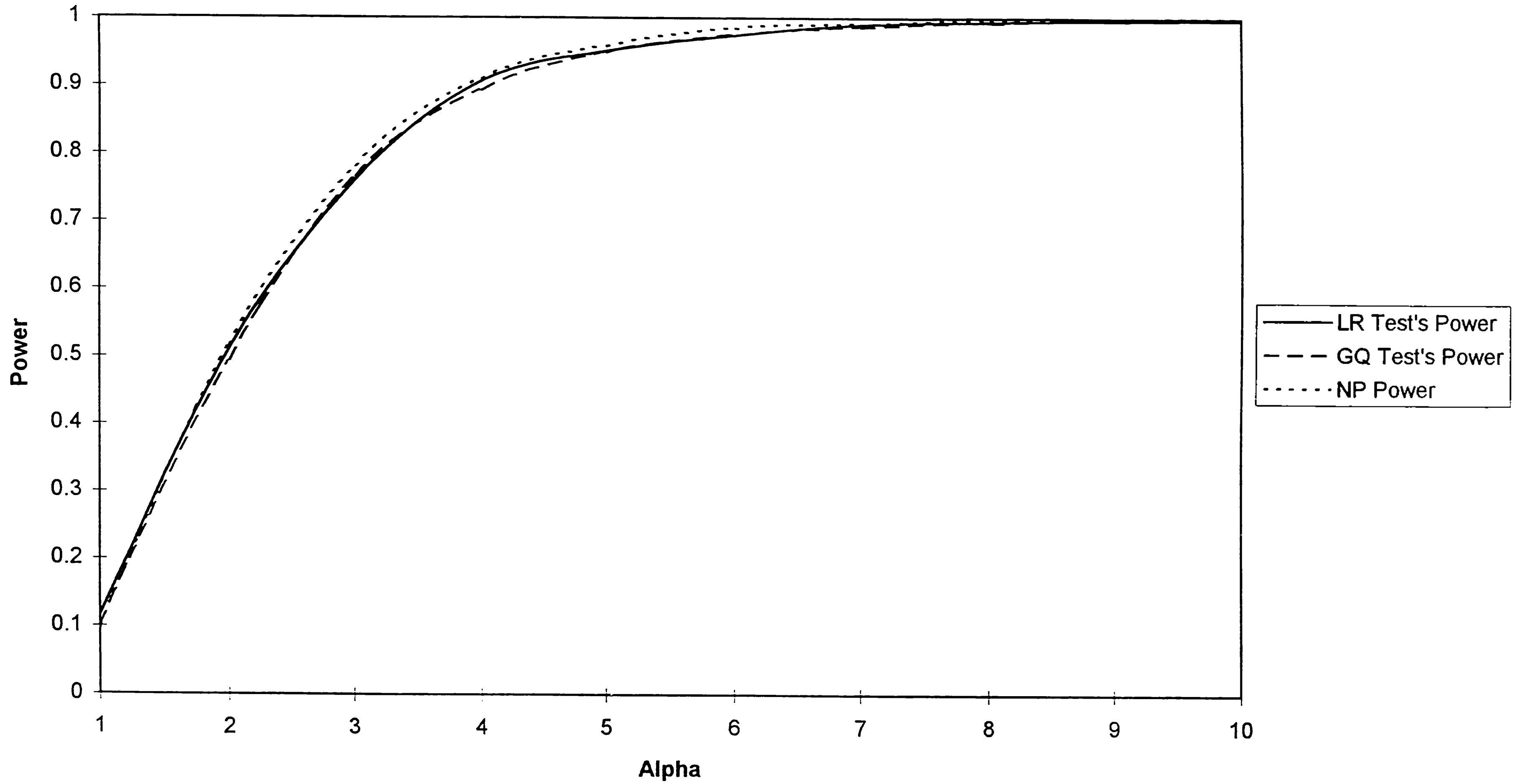
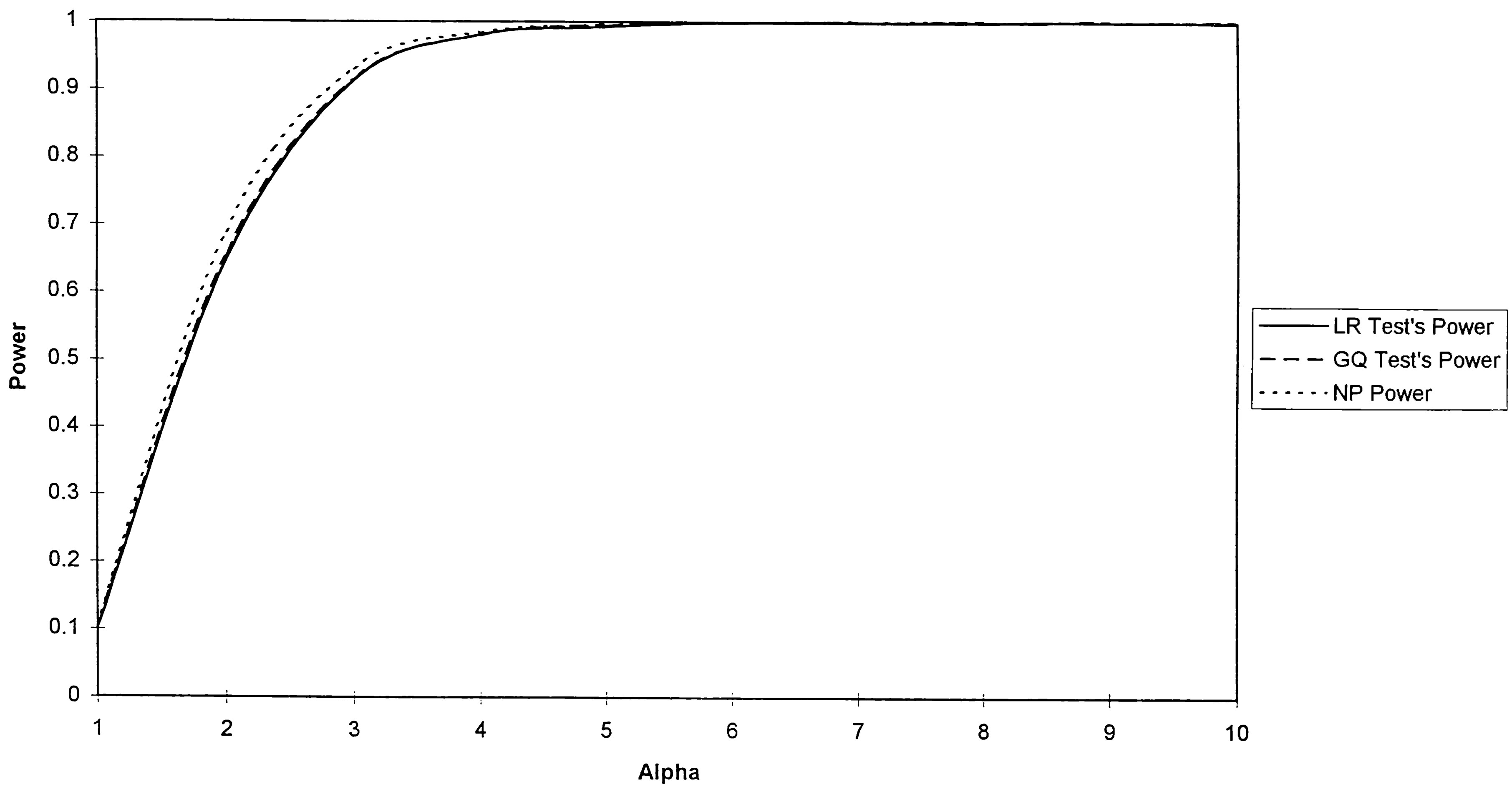


Figure 6. Minimum LR Power vs. GQ Power at 10% Significance Level for T=60 lambda={0.08712,6.949,1.297,1.117,3.394,0.3158}



References

[1] M. Ali and C. Giacotto, 1984, A study of several new and existing tests for heteroskedasticity in the general linear model. *Journal of Econometrics*, 26, 355-374.

[2] R.B. Davies, 1969, Beta-optimal tests and an application to the summary evaluation of experiments. *Journal of the Royal Statistical Society* B, 31:524-538.

[3] Jaysri Dutta and Asad Zaman, 1990, What do tests for heteroskedasticity detect?. Technical Report 9022, Center for Operational Research and Econometrics, Universite Catholique de Louvain, Belgium.

[4] M.A. Evans and M.L. King, 1985, A point optimal test for heteroskedastic disturbances. *Journal of Econometrics*, 27, 163-178.

[5] M.A. Evans and M.L. King, 1988, A further class of tests for heteroskedasticity. *Journal of Econometrics*, 37.

[6] S. M. Goldfeld and R. E. Quandt, 1965, Some tests for homoskedasticity. *Journal of the American Statistical Association*, 60, 539-547.

[7] S. M. Goldfeld and R. E. Quandt, 1972, *Nonlinear Methods in Econometrics.* North Holland, Amsterdam.

[8] A. Harvey and G. Phillips, 1974, A comparison of the power of some tests for heteroskedasticity in the general linear model. *Journal of Econometrics*, 2, 307-316.

[9] R. V. Hogg and A. T. Craig, 1978, *Introduction to Mathematical Statistics.* Macmillan Publishing Co., Inc., New York.

[10] G. G. Judge, W. E. Griffiths, R. Carter Hill, and Tsoung-Chao Lee, 1980, *The Theory and Practice of Econometrics.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

[11] M. L. King. A bounds test for heteroskedasticity, 1982, Technical Report 5/82, Monash University, Department of Economics and Operations Research, Monash University, Australia.

[12] K. Rao Kadiyala and H. D. Oberhelman, 1984, Alternative Tests for Heteroskedasticity of Disturbances: A Comparative Study. *Commun. Statist. Theor. Meth.*, 13 (8), 987-1001.

[13] W. Kramer and H. Sonnberger, 1986, *The Linear Regression Model Under Test.* Physica-Verlag Heidelberg Wien.

[14] E.L. Lehmann, 1980, *Testing Statistical Hypotheses.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

[15] Maddala, 1989, *Econometrics.* Macmillan Pub. Co., New York.

[16] K. Surekha and W. E. Griffiths, 1986, An evaluation of the power of some tests for heteroskedasticity. *Journal of Econometrics*, 31, 219-231.

[17] Asad Zaman, 1994, Most stringent and best invariant hypothesis tests for regression models. Technical Report 94-06, Bilkent University, Ankara, Turkey, February.

[18] Asad Zaman. Shortcomings of tests for autoregression. Technical Report 94-04, Bilkent University, Economics Department, Bilkent University, Ankara, Turkey, January 1994.