Robustifying Glejser test of heteroskedasticity

Kyung So Im

Department of Economics Wichita State University, 1845 Fairmount, Wichita, KS 67260-0078, USA

Received 1 September 1997; received in revised form 1 June 1999

Abstract

Verifying that the Glejser test for heteroskedasticity is asymptotically invalid unless the error density is symmetric, this paper proposes a simple modification to make the test robust to asymmetric disturbances. Simulation results demonstrate that the size of the modified test is correct for both symmetric and skewed errors. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C12; C20

Keywords: Heteroskedasticity; Glejser test; Size

1. Introduction

In least-squares analysis, the most widely used tests for the presence of heteroskedasticity examine whether the squared residuals are correlated with some other variables. Since they were advanced by Bickel (1978), Breusch and Pagan (1979), and White (1980), the squared residual-based tests have been adopted as a standard procedure in many applications. Much concern, however, has been focused on the potential loss of power when the density of the disturbances has fat tails, and has given rise to alternative approaches such as Koenker and Bassett (1982) and Newey and Powell (1987). The latter, in particular, focused their attention on the test proposed by Glejser (1969), which is designed to test whether the regression residuals in absolute value are

E-mail address: im@twsuvn.uc.wusu.edu (K.S. Im).
correlated with some other variables. Following the simulation result provided
by Newey and Powell (1987), the Glejser test is virtually identical to their more
elaborate test obtained by comparing two different expectile estimators, and is
much more powerful than the squared residual-based test when the error density
has fat tails. It also dominated the percentile-based test proposed by Koenker
and Bassett (1982).

It has been an issue in the literature whether the Glejser test is valid when the
error density is not symmetric. Several studies have addressed this issue through
simulation: Barone-Adesi and Talwar (1983) observed that the empirical size of
the Glejser test is quite different from the nominal size when the errors follow the
chi-square distribution with four degrees of freedom. Ali and Giaccotto (1984)
obtained the opposite result. The simulated size of the Glejser test is reported to
be close to the nominal size when the error density is lognormal, which is much
more skewed than the chi-square distribution with four degrees of freedom.
Trivedi and Long (1994) reported the over-rejection tendency of the Glejser test
when the error density is skewed.

It is verified in the following section that the Glejser test is not valid for
skewed errors. And we propose a simple modification to the Glejser test to
correct this non-robustness property. The modified test has correct size for
asymmetric disturbances as well as for symmetric ones.

Section 3 investigates the size of the modified test in finite samples.
The simulation results demonstrate that the modified test has correct size
both for the symmetric and for the skewed errors. Section 4 concludes this
paper.

2. Modification to Glejser test

Consider a linear model:

$$y_t = x_t \beta + \epsilon_t, \quad t = 1, \ldots, T,$$

(1)

where $x_t$ is the $1 \times k$ vector of the explanatory variables, and $\beta$ is the $k \times 1$
parameter vector of interest. \{(y_t, x_t): t = 1, \ldots, T\} is an independently and
identically distributed random sequence. Throughout the paper, the upper case
letters indicate the data matrix which stacks the observations for $t = 1, \ldots, T$.
For example, $X = (x_1', x_2', \ldots, x_T')'$ and $Y = (y_1, y_2, \ldots, y_T)'$.

A set of basic assumptions necessary for least-squares analysis is

**Assumption 1.** $E(x_t' \epsilon_t) = 0.$

**Assumption 2.** $E(x_t' x_t)$ is positive definite.

**Assumption 3.** $(y_t, x_t)$ has finite fourth moment.
These assumptions are central to the least-squares analysis, ensuring the consistency and the asymptotic normality of the OLS estimator of $\beta$ obtained as

$$\hat{\beta}_T = (X'X)^{-1}X'Y.$$  

(2)

Heteroskedasticity is present if

$$E(x'_t \epsilon_t^2) \neq \sigma^2 E(x'_t \epsilon_t),$$

where $\sigma^2 = E(\epsilon_t^2)$. We are interested in testing the null hypothesis

$$H_0: E[x'_t (\epsilon_t^2 - \sigma^2)] = 0.$$  

(3)

Therefore, the problem is whether the squared-error sequence $\{\epsilon_t^2\}$ is correlated with the sequence of the distinctive elements of $x_t \otimes x_t$. Test of $H_0$ leads, under suitable conditions, to a chi-square test with many degrees of freedom even for moderate values of $k$. It may often be the case that a test based on a reduced list of variables that are most suspect could raise the power significantly. Let the $1 \times h$ vector $z_t$ be the candidate variables chosen by a researcher. Then, the null hypothesis is

$$H_0': E[z'_t (\epsilon_t^2 - \sigma^2)] = 0.$$  

(4)

The standard approach for testing $H_0'$ examines whether $\gamma = 0$ in the regression

$$(\hat{\epsilon}_t^2 - \hat{\sigma}^2) = \alpha + z'_t \gamma + \text{error},$$  

(5)

where $\hat{\epsilon}_t = y_t - x_t \hat{\beta}$, $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_t^2$, which leads to the standard $TR^2$ statistic

$$\Psi_T = \frac{\sum_{t=1}^{T} (\hat{\epsilon}_t^2 - \hat{\sigma}^2) z_t (\sum_{t=1}^{T} \hat{z}_t^2)^{-1} \sum_{t=1}^{T} \hat{z}_t (\hat{\epsilon}_t^2 - \hat{\sigma}^2)}{T^{-1} \sum_{t=1}^{T} (\hat{\epsilon}_t^2 - \hat{\sigma}^2)^2},$$

(6)

where $\hat{z}_t = z_t - T^{-1} \sum_{t=1}^{T} z_t$. Then, we have the following well-known result.

**Theorem 1.** If Assumptions 1–3 and the additional conditions (i) $E(x'_t x_t \epsilon_t) = 0$ and (ii) $E(x'_t x_t \epsilon_t^4) = E(x'_t x_t) E(\epsilon_t^4)$ are satisfied, then, under the null hypothesis,

$$\Psi_T \overset{d}{\to} \chi^2(h),$$

where $\overset{d}{\to}$ denotes weak convergence in distribution and $\chi^2(h)$ is a chi-square random variable with $h$ degrees of freedom.

**Proof.** White (1980). □

---

1 This is a narrow definition, but would be relevant in many studies in the sense that the standard estimators of the asymptotic variance of the OLS estimators are consistent under this condition. See Judge et al. (1985) for a review on the studies about different types of heteroskedasticity.
Glejser (1969) proposed a test designed to see whether \(|e_i|\) is correlated with \(z_i\) as a test of heteroskedasticity. The statistic is obtained as \(TR^2\) in the regression;

\[
|\hat{e}_i| = \alpha + z_i \gamma + \text{error.} \tag{7}
\]

Letting \(\mu = E(|e_i|)\) and \(\hat{\mu} = T^{-1} \sum_{t=1}^{T} |\hat{e}_i|\), we have

\[
G_T = \frac{\sum_{t=1}^{T} (|\hat{e}_i| - \hat{\mu}) z_i (\sum_{t=1}^{T} z_i^2) T^{-1} \sum_{t=1}^{T} z_i^2 (|\hat{e}_i| - \hat{\mu})^2}{T^{-1} \sum_{t=1}^{T} (|\hat{e}_i| - \hat{\mu})^2}. \tag{8}
\]

The hypothesis that is directly tested by this statistic is

\[
H_0^*: E\left[z_i (|e_i| - \mu)\right] = 0. \tag{9}
\]

\(H_0^*\) is different from \(H_0^*\). This could be a subtle issue, the discussion of which is deferred to the final section. Instead, in this section we focus on our main results that the Glejser test requires symmetric error density, and that this non-robustness to the skewed errors could be amended by a simple modification.

To do this we need some preliminary results.

**Lemma 1.** Under Assumptions 1–3,

\[
\Pr[\text{sgn}(\hat{e}_i) \neq \text{sgn}(e_i)] = o(1). \nonumber
\]

**Proof.** \(\Pr[\text{sgn}(\hat{e}_i) \neq \text{sgn}(e_i)] = \Pr[|e_i| < |x_i(\hat{\beta} - \beta)|] = o(1).\) The second equality follows since \(|x_i(\hat{\beta} - \beta)| = O_p(T^{-1/2}).\) \(\Box\)

Let \(T_1\) and \(T_2\) be the number of residuals that take positive and negative values, respectively, and

\[
\hat{m}_T = \frac{T_1 - T_2}{T}.
\]

Also, let

\[
m = 2\Pr(e_i > 0) - 1.
\]

Then, we have

**Lemma 2.** Under Assumptions 1–3,

\[
\hat{m}_T = m + O_p(1). \nonumber
\]

**Proof.** \(T_1/T = \Pr(e_i > 0) + O_p(1)\) and \(T_2/T = 1 - \Pr(e_i > 0) + O_p(1)\) by Lemma 1 and the weak law of large numbers. \(\Box\)

For the subsequent argument, we need an additional assumption;
Assumption 4. \( E(x'_i x_i | e_t > 0) = E(x'_i x_i | e_t < 0) \).

This assumption may not be as strong as it looks, given that \( e_t \) is uncorrelated with \( x_t \). Let the subscripts \( \tau \) and \( s \) index the variables associated with the residuals that take positive and negative values, respectively. Assumption 4 implies
\[
\frac{1}{T} \sum_{t=1}^{T} \tilde{\varepsilon}_t x_t = \frac{1}{T} \sum_{s=1}^{T} \tilde{\varepsilon}_s x_s + o_p(1) = \frac{1}{T} \sum_{t=1}^{T} \tilde{\varepsilon}_t x_t + o_p(1),
\]
and we have the following lemma.

Lemma 3. Under Assumptions 1–4,
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{\varepsilon}_t (|\hat{e}_t| - |\hat{e}_t|) = -m \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{\varepsilon}_t x_t \sqrt{T} (\hat{\beta} - \beta) + o_p(1).
\]

Proof.
\[
\text{LHS} = \frac{1}{\sqrt{T}} \left[ \sum_{t=1}^{T} \tilde{\varepsilon}_t (\hat{e}_t - \varepsilon_t) - \sum_{s=1}^{T} \tilde{\varepsilon}_s (\hat{e}_s - \varepsilon_s) \right] + o_p(1)
\]
\[
= \frac{1}{\sqrt{T}} \left[ - \frac{1}{T} \sum_{t=1}^{T} \tilde{\varepsilon}_t x_t \sqrt{T} (\hat{\beta}_T - \beta) + \frac{1}{T} \sum_{s=1}^{T} \tilde{\varepsilon}_s x_s \sqrt{T} (\hat{\beta}_T - \beta) \right] + o_p(1)
\]
\[
= \left( - \frac{1}{T} \sum_{t=1}^{T} \frac{1}{T} \sum_{t=1}^{T} \tilde{\varepsilon}_t x_t + \frac{1}{T} \sum_{s=1}^{T} \tilde{\varepsilon}_s x_s \right) \sqrt{T} (\hat{\beta}_T - \beta) + o_p(1)
\]
\[
= - m \frac{1}{T} \sum_{t=1}^{T} \tilde{\varepsilon}_t x_t \sqrt{T} (\hat{\beta}_T - \beta) + o_p(1)
\]

The first equality follows from Lemma 1, and the fourth equality is directed by (10). □

The first term in the right-hand side of (11) becomes negligible as \( T \) grows only if \( m = 0 \), that is, when the median of the residuals converges to zero (or to the mean of the errors). This, in general, requires the symmetry of the error density. Therefore, the Glejser test is not valid asymptotically when the error density is skewed.

When the error density is symmetric, the residuals successfully replace the disturbances in the sense that the replacement does not alter the asymptotic
distribution of the score. Therefore, the asymptotic variance of the score \( \sum_{t=1}^{T} z_t |\hat{e}_t| \) is estimated consistently by \( T^{-1} \sum_{t=1}^{T} (|\hat{e}_t| - \tilde{\mu})^2 (\sum_{t=1}^{T} z_t^2 |\hat{e}_t|) \) under suitable conditions (see, e.g., the assumptions in Theorem 2 that follows).

On the other hand when the errors are skewed, \( (1/\sqrt{T}) \sum_{t=1}^{T} z_t |\hat{e}_t| \) and \( (1/\sqrt{T}) \sum_{t=1}^{T} z_t |\hat{e}_t| \) are not identical asymptotically and \( T^{-1} \sum_{t=1}^{T} (|\hat{e}_t| - \hat{\mu})^2 (\sum_{t=1}^{T} z_t^2 |\hat{e}_t|) \) is no longer a consistent estimator of the asymptotic variance of the score \( \sum_{t=1}^{T} z_t |\hat{e}_t| \) (while it does estimate the asymptotic variance of \( \sum_{t=1}^{T} z_t |e_t| \) consistently). The symmetry would be violated in many applications, and the direct application of the Glejser test is somehow limited. However, the Glejser test becomes robust to skewed errors through a simple modification. The clue is in the following Lemma.

**Lemma 4.** Under Assumptions 1–4,

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left| \hat{e}_t \right| \left( \hat{e}_t - \hat{\mu} \right) = o_p(1). \tag{12}
\]

**Proof.** Note that

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{\hat{m}_T} \sum_{t=1}^{T} \frac{1}{T} z_t (\hat{e}_t - e_t) = - \hat{m}_T \frac{1}{T} \sum_{t=1}^{T} z_t x_t \sqrt{T} (\hat{\beta}_T - \beta),
\]

which, combined with the results in Lemmas 2 and 3, complete the proof. \( \square \)

Observe that \( \sum_{t=1}^{T} z_t (\left| \hat{e}_t \right| - \hat{m}_T \hat{e}_t - \hat{\mu}) \) is the score in the regression

\[
\left| \hat{e}_t \right| - \hat{m}_T \hat{e}_t - \hat{\mu} = x + z_t \gamma + \text{error}. \tag{13}
\]

The \( TR^2 \) statistic from this regression is

\[
\Phi_T = \frac{\sum_{t=1}^{T} (\left| \hat{e}_t \right| - \hat{m}_T \hat{e}_t - \hat{\mu})^2 (\sum_{t=1}^{T} z_t^2 |\hat{e}_t|)^{-1} \sum_{t=1}^{T} z_t (\left| \hat{e}_t \right| - \hat{m}_T \hat{e}_t - \hat{\mu})}{T^{-1} \sum_{t=1}^{T} (\left| \hat{e}_t \right| - \hat{m}_T \hat{e}_t - \hat{\mu})^2}. \tag{14}
\]

The statistic, \( \Phi_T \), will converge to a chi-square random variable with \( h \) degrees of freedom under suitable conditions, and this result is no longer affected by the skewness of the errors. We state this result in the following theorem.

**Theorem 2.** Under Assumptions 1–4 and a set of additional assumptions: (i) \( E(x'_t x_t | e_t > 0) = E(x'_t x_t | e_t < 0) = E(x'_t x_t) \sigma^2 \), (ii) \( E(x'_t x_t | e_t) = E(x'_t x_t) \mu_t \), and (iii) \( E(x'_t x_t e_t) = 0 \),

\[
\Phi_T \overset{d}{\rightarrow} \chi^2(h).
\]
Proof. $T^{-1/2} \sum_{t=1}^{T} \hat{z}_t^2 (|\hat{e}_t| - \hat{\mu}_T \hat{e}_t - \hat{\mu})$ and $T^{-1/2} \sum_{t=1}^{T} \hat{z}_t^2 (|\hat{e}_t| - \mu) - \mu$ are identical asymptotically from Lemma 4. Also,

$$\text{Avar} \left[ T^{-1/2} \sum_{t=1}^{T} \hat{z}_t^2 (|\hat{e}_t| - \mu) - \mu \right] = E \left[ \hat{z}_t^2 (|\hat{e}_t| - \mu)^2 \right]$$

$$= E(\hat{z}_t^2)E(|\hat{e}_t| - \mu)^2.$$

The first equality follows from the independence of $\{(y_t, x_t): t = 1, \ldots, T\}$, and the second equality is ensured by assumptions (i)–(iii). Proof is complete by applying the multivariate Liapounov central limit theorem. □

Although the conditions (i)–(iii) of Theorem 2 may not be too restrictive under the null hypothesis of no heteroskedasticity, it would be useful sometimes to have a test that does not hinge on these conditions.

**Theorem 3.** Let

$$\Phi_T = \sum_{t=1}^{T} (|\hat{e}_t| - \hat{\mu}_T \hat{e}_t - \hat{\mu}) \hat{z}_t' \left[ \sum_{t=1}^{T} \hat{z}_t^2 (|\hat{e}_t| - \hat{\mu}_T \hat{e}_t - \hat{\mu})^2 \right]^{-1} \sum_{t=1}^{T} \hat{z}_t (|\hat{e}_t| - \hat{\mu}_T \hat{e}_t - \hat{\mu}).$$

(15)

If Assumptions 1–4 hold, and the sixth moment of $(y_t, x_t)$ exists, then under the null hypothesis $H_0'$ described in (9),

$$\Phi_T \stackrel{d}{\rightarrow} \chi^2(h).$$

Proof. The result follows from the multivariate Liapounov central limit theorem. □

Note that the score of the modified test is obtained by simply adding a term $\hat{\mu}_T \sum_{t=1}^{T} \hat{z}_t^2 \hat{e}_t$ to the score of the Glejser test, and this added term is $O_p(T^{-1/2})$ at most, under the alternative as well as under the null. In particular, when $z_t \subseteq x_t$, the additional term in the numerator is strictly zero. The modification adjusts only the estimators of the asymptotic variance of the score such that they are consistent for skewed errors as well. Therefore, the asymptotic power of the test should not be affected by this modification. Referring to Newey and Powell (1987) and Trivedi and Long (1994) for power comparison between the Glejser test and the standard residual-square-based test, we focus on the size issue in our simulation study.
3. Simulation results

The data-generation process is

\[ y_t = \alpha + x_t \beta + e_t, \quad t = 1, \ldots, T, \]

where \( \alpha = \beta = 1. \) \( x_t, \ t = 1, \ldots, T, \) are generated independently from chi-square distribution with one degree of freedom in each replication. Therefore, the regressor is not fixed across replications. The disturbances are serially independent, and are independent of \( x_t. \) We consider three different error densities: (i) standard normal, (ii) chi-square with four degrees of freedom, and (iii) lognormal. The chi-square distribution with four degrees of freedom and the lognormal distribution are chosen to make our results comparable with those reported earlier by Barone-Adesi and Talwar (1983) and Ali and Giaccotto (1984). It also fits the objective of this paper in studying the size of the Glejser test and its modified versions. Note that all the test statistics we study in this section are numerically invariant to the mean and the variance of the regressor, the variance of the disturbance, and to the different values of \( \alpha \) and \( \beta. \) For each error density, we consider the cases for \( T = 100 \) and 500.

Choosing \( z_t = x_t, \) we consider the four tests; the standard squared-residual-based tests \( (\psi_T), \) the Glejser test \( (G_T), \) the modified Glejser test \( (\Phi_T) \) and the robust version of the modified Glejser test \( (\Phi'_T), \) based on formula (6), (8), (14) and (15), respectively. The number of replications is 5000. Table 1 reports the empirical size obtained as the rejection ratio of the null hypothesis at the 5% significance level. Therefore, the asymptotic 95% confidence interval of the rejection rate is (0.044, 0.056).

As was predicted, the empirical sizes of the Glejser test \( (G_T) \) are far off from the nominal size when the density of the underlying disturbances is skewed. It tends to reject the null hypothesis too often, and this tendency increases with skewness. However, the empirical size of the modified test \( (\Phi_T) \) is close to the nominal size in every case. There is a slight over-rejection tendency for the robust version of the modified test \( (\Phi'_T). \) It is somewhat surprising to see that this tendency remains almost at the same level even when \( T = 500. \) The sluggish

<table>
<thead>
<tr>
<th>Densities</th>
<th>( T = 100 )</th>
<th>( T = 500 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.028</td>
<td>0.045</td>
</tr>
<tr>
<td>( \chi^2(4) )</td>
<td>0.033</td>
<td>0.110</td>
</tr>
<tr>
<td>Lognormal</td>
<td>0.043</td>
<td>0.151</td>
</tr>
</tbody>
</table>
convergence to its asymptotic distribution may have resulted from the estimation of higher moments involved with the computation of $\Phi_T$. Although there remains a slight under-rejection tendency, the empirical size of the standard squared residual-based test ($\Psi_T$) is proper in every case, as it ought to be.

4. Concluding remarks

The Glejser test examines whether the conditional first absolute moment of the error depends on $z_t$, while the interest centers on the possible dependence of the conditional error variance on $z_t$. Therefore, there would be situations where the Glejser test is lack of power against the variance dependence. Although we cannot prove analytically that these situations are unlikely in practice, it also seems difficult to bring a practical situation where some variable is correlated only with the squared error term without being correlated with the absolute error term. As a matter of fact, when we graphically examine the possible presence of heteroskedasticity, it is a common practice to look at the scatter plot of each of $z_t$ and the residuals, rather than the scatter plot of $z_t$ and the squared residuals. The Glejser test could be viewed as a formalization of this common graphic procedure.

Although the mean absolute deviation is not as popular as the variance, it could certainly serve better than the variance when the population density is leptokurtic. The variance measure is too sensitive to outliers that often show up in a leptokurtic population. In this regard, it is not surprising that the Glejser test is more powerful than the standard squared residual-based test when the error density has fat tails as was evidenced by numerous Monte Carlo studies. Hence, the Glejser test would serve in practice as a useful companion to the standard squared residual-based test. Asymmetry is another typical symptom of non-normality and is often detected in applications. The robustification developed in this paper makes the Glejser test more appealing in practice.

Acknowledgements

The first version of this paper was written while I was visiting Indiana University-Purdue University at Indianapolis. The great support from IUPUI is gratefully acknowledged. I am grateful to Pravin Trivedi who motivated me to study asymmetry problem in heteroskedasticity tests, and to the anonymous co-editor and the two referees who pointed out the several mistakes in the first version. After the acceptance of this paper, I learned of a similar test proposed by Machado and Santos Silva (2000).
References


