Testing under a special form of heteroscedasticity

Helmut Farbmacher\textsuperscript{a} and Heinrich Kögel\textsuperscript{b}

\textsuperscript{a}Max Planck Society, Munich Center for the Economics of Aging, Munich, Germany; \textsuperscript{b}Centre for European Economic Research (ZEW) Mannheim, Mannheim, Germany

\textbf{ABSTRACT}

In the presence of heteroscedasticity, conventional standard errors (which assume homoscedasticity) can be biased up or down. The most common form of heteroscedasticity leads to conventional standard errors that are too small. In this study, we discuss the conditions under which conventional standard errors are too large. In such settings, standard tests of heteroscedasticity may fail and leave the heteroscedasticity undetected. This is particularly problematic as power gains can be achieved when testing for the causal effect in such settings.

\textbf{KEYWORDS}

Heteroscedasticity; variance estimation; returns to education

\textbf{JEL CLASSIFICATION}

C12; C13; I26

\section{I. Introduction}

In the June 2011 issue of the \textit{American Economic Review}, Vikesh Amin commented on an article by Dorothe Bonjour et al. published in December 2003 also in the \textit{American Economic Review}. Bonjour et al. (2003) estimated the private return to education using a dataset containing 428 female monozygotic twins. One of their main findings was an estimated return to one additional year of education of 7.7%, which is statistically significant at the 5% level. Amin (2011) replicated their results and performed similar estimations where he excluded outliers. He found that many of Bonjour et al.’s within-twin pair estimates became smaller in magnitude and significant only at lower levels or insignificant when removing these extreme values. In this study, we show that the inference in Amin (2011) is mostly incorrect due to the presence of a special form of heteroscedasticity. The correct standard errors turn out to be around 15% lower, leading to different policy conclusions. In contrast to Amin (2011), we find a significant positive return to education for most of the within-twin pair regressions.

In Section II, we provide a theoretical background for the situation when an upward bias in conventional standard errors occurs. There we also discuss the difficulties in using standard tests for heteroscedasticity in such settings. We then propose a heteroscedasticity test which has better power properties. Section III presents the results of a series of Monte Carlo simulations based on data exhibiting this special form of heteroscedasticity. In Section IV, we use three test procedures to test for heteroscedasticity in Bonjour et al.’s dataset. The Koenker variant of the Breusch–Pagan test and the White test do not reject the hypothesis of homoscedasticity, which is as expected, due to the special form of heteroscedasticity present. However, our proposed test rejects the null hypothesis, in favour of the special form of heteroscedasticity. Also in Section IV, we present the within-twin pair regressions using the appropriate standard errors. Section V concludes.

\section{II. Inference problems}

In the presence of heteroscedasticity, conventional standard errors (which assume homoscedasticity) can be biased up or down. The most common form of heteroscedasticity, where the residual variance rises in increasing regressor values, usually leads to conventional standard errors that are too small. When Wald tests based on these standard errors are insignificant, heteroscedasticity robust standard errors do not change inference. On the other hand, inference is conservative in a setting with upward-biased conventional standard errors. Using heteroscedasticity robust standard errors may change inference in this case.

Angrist and Pischke (2010) derive the condition for such an upward bias in the classical bivariate linear regression model.\footnote{A similar insight can be derived in the multivariate regression model by partialling out all other covariates.}
\[ y_i = \alpha + \beta x_i + e_i \] (1)

Let \( \sigma^2_\beta \) denote the true sampling variance for the estimator \( \hat{\beta} \) and let \( [\sigma^2_{\hat{\beta}}]_{\text{conv}} \) denote the corresponding conventional standard error, assuming homoscedasticity. Then

\[ [\sigma^2_{\hat{\beta}}]_{\text{conv}} > \sigma^2_\beta \iff \text{Cov}[e_i^2, (x_i - \bar{x})^2] < 0. \] (2)

An upward bias in conventional standard errors occurs if there is a negative covariance between the squared residual \( e_i^2 \) and the squared deviation of \( x_i \) from its mean \( \bar{x} \). The further away the observation \( x_i \) is from \( \bar{x} \), the smaller becomes \( \text{Var}[e_i|x_i] = E[e_i^2|x_i] \), the conditional variance of residual \( e_i \). When \( \text{Cov}[e_i^2, (x_i - \bar{x})^2] < 0 \), the corresponding scatter plot of \( e_i \) on the regressor \( x_i \) often resembles an ellipse. That is why we refer to this form of heteroscedasticity as elliptical heteroscedasticity.

If the data exhibit elliptical heteroscedasticity, the usual Wald tests for hypotheses about \( \beta \) in the bivariate regression model using conventional standard errors give an actual size smaller than the nominal size. Policy conclusions based on estimates with conventional standard errors are thus conservative. Hence, Wald tests using heteroscedasticity robust standard errors lead to power gains compared with tests using conventional standard errors in this case.

By exploiting Equation (2), we can test specifically for elliptical heteroscedasticity in the classical bivariate regression model.

To derive our elliptical heteroscedasticity test, consider the regression

\[ e_i^2 = \delta_0 + \delta_1 (x_i - \bar{x})^2 + \xi_i, \] (3)

where the squared residuals \( e_i^2 \) are obtained from the regression of \( y_i \) on \( x_i \).

Under elliptical heteroscedasticity, we know that \( \text{Cov}[e_i^2, (x_i - \bar{x})^2] < 0 \) and therefore

\[ \delta_1 = \frac{\text{Cov}[e_i^2, (x_i - \bar{x})^2]}{\text{Var}[(x_i - \bar{x})^2]} < 0. \] (4)

Hence, by exploiting this knowledge we can test specifically for elliptical heteroscedasticity.

Our elliptical heteroscedasticity test conducts a one-sided Wald test for \( H_0 : \delta_1 \geq 0 \) against \( H_a : \delta_1 < 0 \) in the regression \( e_i^2 = \delta_0 + \delta_1 (x_i - \bar{x})^2 + \xi_i \).

The hypotheses are \( H_0: \text{no elliptical heteroscedasticity} \) and \( H_a: \text{elliptical heteroscedasticity} \).

### III. Monte Carlo simulations

To illustrate the testing issues arising from elliptical heteroscedasticity, we run a series of Monte Carlo simulations. The design of our Monte Carlo simulations is based on the following data generating process.

\[ y_i = 0.04 x_i + e_i \]

\[ e_i = \sqrt{\frac{1}{\{(x_i - \bar{x})^2 + 0.1\}}} \xi_i \]

\[ x_i = [x^*_i], x^*_i \sim N(0.04, 1.8^2), e_i \sim N(0, 1), \]

\[ a = 0, 0.01, 0.02, 0.03, \ldots, 0.5 \]

We chose the model so that the shape of the resulting \( y-x \) scatter plot resembles Panel A, Figure 1 in Amin (2011). For values of \( a \) between 0.15 and 0.3, the \( y-x \) scatter plot is most similar to Panel A. The operator \([\cdot]\) rounds \( x^*_i \) to the nearest integer. Hence, \( x_i \) is an integer, just as the within-twin difference in estimated years of schooling in Bonjour et al. (2003). Furthermore, also in accordance with the within-twin difference in estimated years of schooling, the values of \( x_i \) are centred around the mean \( \bar{x} \). The structure of the error term \( e_i \) implies that \( \text{Cov}[e_i^2, (x_i - \bar{x})^2] < 0 \) if \( a > 0 \). The larger is the parameter \( a \), the more negative is the covariance between \( e_i^2 \) and \( (x_i - \bar{x})^2 \), and therefore, the stronger is the upward bias caused by elliptical heteroscedasticity. For \( a = 0 \), the error term is homoscedastic. The number of observations is set to \( N = 214 \), as in Bonjour et al.’s dataset, and additionally to \( N = 2140 \). The number of replications is 10 000.

In each simulation, we evaluate the size and power of three different tests for heteroscedasticity: the Koenker (1981) variant of the Breusch–Pagan (1979) test, which drops the assumption of normality, with \( x \) as the independent variable, the White (1980) test and our elliptical heteroscedasticity test introduced in Section II. In addition, we compare the size for the parameter of interest in the causal model using Wald tests for the hypothesis \( H_0 : \beta = 0.04 \) against \( H_a : \beta \neq 0.04 \) in the
regression of $y_i$ on $x_i$ using robust and conventional standard errors.

**Figure 1.** Panel A shows power plots for the heteroscedasticity tests. The simulation with $a = 0$ gives the actual size of each test. While the rejection frequency of the Breusch–Pagan and White test is close to the given significance level of $\alpha = 5\%$ for $N = 214$, the actual size of the elliptical heteroscedasticity test is above this value, with 11.9%. However, the actual test size for the latter test approaches the theoretically given significance level for larger numbers of observations. The simulation with $N = 2140$ yields an actual size of 7.7% for the elliptical heteroscedasticity test.

For $a > 0$, **Figure 1.** Panel A displays the power of each test. The rejection frequency of the White test and our elliptical heteroscedasticity test increases with stronger elliptical heteroscedasticity, i.e. with increasing values of $a$. Compared with the elliptical heteroscedasticity test, the White test performs worse in detecting heteroscedasticity, although the difference in power gets smaller for larger values of $a$. In contrast to the elliptical heteroscedasticity test, the White test does not have elliptical heteroscedasticity as the alternative hypothesis, but rather heteroscedasticity in general. The less specific formulation of $H_a$ may explain the White test’s worse performance. The Breusch–Pagan test has considerably smaller rejection frequencies than the two other tests throughout the whole range of $a > 0$. For $N = 214$ and $N = 2140$, it does not reach a power of 5% for any given positive value of $a$. This result is related to the fact that the basic specification of the Breusch–Pagan test is for detecting linear
forms of heteroscedasticity, whereas elliptical heteroscedasticity implies a nonlinear form of heteroscedasticity.

Figure 1, Panel B displays the size of the Wald tests with $H_0 : \beta = 0.04$ for different values of $a$ using conventional and robust standard errors. Under homoscedasticity, $a = 0$, both test versions’ sizes are close to the given significance level of 5% for $N = 214$ and equal to $a = 5\%$ for $N = 2140$. In the presence of heteroscedasticity, $a > 0$, the Wald tests using robust standard errors yield also a size around 5%. The size of the Wald tests using conventional standard errors, however, decreases with increasing $a$. At $a = 0.5$, the Wald test with conventional standard errors has an actual size of 0.1% for $N = 214$ and $N = 2140$, respectively. Hence, $t$-tests with conventional standard errors do not reject the correct null hypothesis often enough for $a > 0$. This is due to the upward bias in conventional standard errors in this case. Similarly, the Wald tests with conventional standard errors lose power with increasing $a$. Conversely, the heteroscedasticity robust Wald tests become more powerful as $a$ gets larger.\(^2\)

### IV. Empirical illustration: returns to education

Amin (2011) excluded up to four twin pair outliers. These were determined on the basis of the absolute between-twin difference in hourly wages. Figure 1, in Amin (2011) illustrates which data points he removed. Figure 1, Panel A, in Amin (2011) already suggests that the data exhibit the elliptical heteroscedasticity discussed in Section II, which leads to an upward bias in conventional standard errors.

To test for the presence of heteroscedasticity, we performed the three tests outlined in Section II for all within-twin pair OLS and IV regressions in columns 3, 4 and 7, 8 of Table 1 in Amin (2011). In all regressions, the dependent variable is the within-twin difference in log hourly wages. The regressor of interest is the within-twin difference in self-reported education. In the IV regressions, this variable is instrumented by the within-twin difference in the co-twin’s report of the other twin’s education. The regressions in columns 7 and 8 include the covariates within-twin difference in marital status, current job tenure, part-time status and whether a person lives in London or the southeast of the UK.

Table 1 provides the $p$-values for the Koenker variant of the Breusch–Pagan test with within-twin difference in estimated years of schooling as the only independent variable, the White test and our proposed elliptical heteroscedasticity test from Section II. In the regressions including covariates, we partialled them out before testing. The elliptical heteroscedasticity test rejects for all regression specifications at least at the 10% level. In contrast, the Breusch–Pagan and White test do not reject the hypothesis of homoscedasticity in any regression. This can be attributed to the difficulties and lower power in detecting elliptical heteroscedasticity when using more general tests as discussed in Section II. Based on our proposed elliptical heteroscedasticity test, there is evidence for the presence of elliptical heteroscedasticity in the data. Hence, conventional standard errors are incorrect and may lead to false policy conclusions.

Table 1 also shows the causal effect of education on average wages. In addition to the replications using conventional standard errors, Table 1 reports robust standard error estimates and the corresponding significance levels. In all but two regressions, the robust standard error is smaller than the conventional one. This result is in line with the suspicion that elliptical heteroscedasticity is present in the data, which causes an upward bias in conventional standard errors. It also supports the conclusions from our elliptical heteroscedasticity test. In many regressions where the causal effect is insignificant using conventional standard errors, it becomes significant at the 5% or 10% level when using robust standard errors. With conventional standard errors, 13 out of the 20 regressions yield an insignificant parameter estimate. In contrast, only in 3 out of the 20 regressions do we fail to find a return to education significantly different from 0 when (correctly) using robust standard errors.

### V. Conclusion

In this study, we discuss the conditions under which conventional standard errors are upward biased. In such settings, standard tests of heteroscedasticity may fail and leave the heteroscedasticity undetected. When Wald tests based on downward-biased conventional standard errors are insignificant, heteroscedasticity robust standard errors do not change inference. On the other hand,

---

\(^2\)See Farbmacher and Kögel (2015) for more details on the simulation results.
Table 1. Replication and re-estimation of within-twin pair regressions in Table 1, Amin (2011) and corresponding heteroscedasticity tests.

| Sample                               | Within-twin pair  | Within-twin pair with covariates |
|--------------------------------------|-------------------|----------------------------------|
|                                      | OLS (3)          | IV (4)                           |
|                                      | OLS (7)          | IV (8)                           |
| Full Bonjour et al. dataset          | 0.039            | 0.077                            |
| Conventional SE                      | (0.023)*         | (0.033)**                        |
| Robust SE                            | (0.018)**        | (0.039)**                        |
| Breusch–Pagan test                   | 0.3645           | 0.3124                           |
| White test                           | 0.4805           | 0.4581                           |
| Elliptical Het. test                 | 0.0096           | 0.0234                           |
| Drop if abs. wage difference >90     | 0.032            | 0.050                            |
| Conventional SE                      | (0.021)          | (0.031)                          |
| Robust SE                            | (0.016)**        | (0.027)*                         |
| Breusch–Pagan test                   | 0.7221           | 0.6719                           |
| White test                           | 0.4982           | 0.4906                           |
| Elliptical Het. test                 | 0.0109           | 0.0107                           |
| Drop if abs. wage difference >75     | 0.032            | 0.050                            |
| Conventional SE                      | (0.021)          | (0.030)*                         |
| Robust SE                            | (0.016)**        | (0.027)*                         |
| Breusch–Pagan test                   | 0.7207           | 0.6737                           |
| White test                           | 0.5488           | 0.5421                           |
| Elliptical Het. test                 | 0.0176           | 0.0173                           |
| Drop if abs. wage difference >65     | 0.032            | 0.036                            |
| Conventional SE                      | (0.020)          | (0.029)                          |
| Robust SE                            | (0.016)**        | (0.022)                          |
| Breusch–Pagan test                   | 0.7143           | 0.7065                           |
| White test                           | 0.6147           | 0.6101                           |
| Elliptical Het. test                 | 0.0297           | 0.0286                           |
| Drop if abs. wage difference >60     | 0.028            | 0.036                            |
| Conventional SE                      | (0.019)          | (0.027)                          |
| Robust SE                            | (0.016)*         | (0.022)                          |
| Breusch–Pagan test                   | 0.9861           | 0.9466                           |
| White test                           | 0.6310           | 0.6237                           |
| Elliptical Het. test                 | 0.0586           | 0.0534                           |

N = 214(187) for the full Bonjour et al. (2003) dataset (including covariates). For further details on the regressions, see Section IV. The values for the heteroscedasticity tests are p-values. For further details on the tests, see Sections II and III. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Inference is conservative in a setting with upward-biased conventional standard errors. We discuss the power gains when using robust standard errors in this case and also potential problems of heteroscedasticity tests. In Monte Carlo simulations, we show that our proposed heteroscedasticity test has a far better actual size. In our application only this test detects the heteroscedasticity, and using then the appropriate standard errors leads to different test decisions.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Amin, V. 2011. “Returns to Education: Evidence from UK Twins: Comment.” *American Economic Review* 101 (4): 1629–1635. doi:10.1257/aer.101.4.1629.

Angrist, J., and J.-S. Pischke. 2010. *A Note on Bias in Conventional Standard Errors under Heteroskedasticity*. Accessed 01 May 2016. http://econ.lse.ac.uk/staff/spischke/mhe/josh/Notes%20on%20conv%20std%20error.pdf

Bonjour, D., L. F. Cherkas, J. E. Haskel, D. D. Hawkes, and T. D. Spector. 2003. “Returns to Education: Evidence from U.K. Twins.” *American Economic Review* 93 (5): 1799–1812. doi:10.1257/000282803322655554.

Breusch, T. S., and A. R. Pagan. 1979. “A Simple Test for Heteroscedasticity and Random Coefficient Variation.” *Econometrica* 47 (5): 1287–1294. doi:10.2307/1911963.

Farbmacher, H., and H. Kögel. 2015. Inference Problems Under a Special Form of Heteroskedasticity. *MEA Discussion Paper* 03-2015.

Koenker, R. 1981. “A Note on Studentizing a Test for Heteroskedasticity.” *Journal of Econometrics* 17 (1): 107–112. doi:10.1016/0304-4076(81)90062-2.

White, H. 1980. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica* 48 (4): 817–838. doi:10.2307/1912934.