MAX-PLANCK-INSTITUT FÜR SOZIALRECHT UND SOZIALPOLITIK
MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY

mea Munich Center for the Economics of Aging

Inference Problems under a Special Form of Heteroskedasticity

Helmut Farbmacher, Heinrich Kögel

03-2015

MEA DISCUSSION PAPERS

mea – Amalienstr. 33_D-80799 Munich_Phone +49 89 38602-355_Fax +49 89 38602-390_www.mea.mpisoc.mpg.de

Inference Problems under a Special Form of Heteroskedasticity

Helmut Farbmacher and Heinrich Kögel

Max Planck Society, Munich Center for the Economics of Aging

March 10, 2015

Abstract

In the presence of heteroskedasticity, conventional standard errors (which assume homoskedasticity) can be biased up or down. The most common form of heteroskedasticity leads to conventional standard errors that are too small. When Wald tests based on these standard errors are insignificant, heteroskedasticity robust standard errors do not change inference. On the other hand, 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 heteroskedasticity tests. As a solution for the poor performance of the usual heteroskedasticity tests in this setting, we propose a modification of the White test which has better properties. We illustrate our findings using a study in labor economics. The correct standard errors turn out to be around 15 percent lower, leading to different policy conclusions. Moreover, only our modified test is able to detect heteroskedasticity in this application.

1. Introduction

In the June 2011 issue of the *American Economic Review*, Vikesh Amin commented on an article by Dorothe Bonjour et al. published in December 2003 also in the *American Economic Review*. Bonjour et al. (2003) estimated the private return to education using a dataset containing 428 female monozygotic twins from the United Kingdom. One of their main findings was an estimated return to one additional year of education of 7.7 percent, which is statistically significant at the 5 percent 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 for the within-twin pair regressions in Amin (2011) is mostly incorrect due to the presence of a special form of heteroskedasticity, which we describe in Section 2. In contrast to Amin (2011), we find a significant positive return to education at conventional levels for most of his within-twin pair regressions. The majority of these regressions therefore support the conclusion in Bonjour et al. (2003) of a positive return to education that is significantly different from zero. In Section 2 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 the standard tests for heteroskedasticity to detect the form of heteroskedasticity relevant for this study. We then propose a variant of the White test which we develop specifically for detecting this kind of heteroskedasticity. Section 3 presents the results of a series of Monte Carlo simulations based on data exhibiting this special form of heteroskedasticity. In these simulations, we compare the power and size of the usual Wald tests in regressions using conventional and robust standard errors. In addition, we examine the power and size of three different tests for heteroskedasticity. In the main part, Section 4, we use the three test procedures to test for heteroskedasticity in Bonjour et al.'s dataset. The Koenker variant of the Breusch–Pagan test and the White test do not reject the hypothesis of homoskedasticity, which is as expected, due to the special form of heteroskedasticity present. However, our specific White test which we discuss in Section 2 rejects, at conventional significance levels, the null hypothesis, in favor of the special form of heteroskedasticity. Also in Section 4, we present our replication of the within-twin pair regressions in Table 1 of Amin (2011) and our re-estimated results using the appropriate standard errors.

2. Inference problems

In the presence of heteroskedasticity, conventional standard errors (which assume homoskedasticity) can be biased up or down. The most common form of heteroskedasticity, 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, heteroskedasticity robust standard errors do not change inference. On the other hand, inference is conservative in a setting with upward-biased conventional standard errors. Angrist and Pischke (2010) derive the condition for such an upward bias. In this setting heteroskedasticity robust standard errors are not only size-correct but also lead to power gains compared to the conventional standard errors.

Consider the classical bivariate linear regression model[1]

$$y_i = \alpha + \beta x_i + e_i$$

where the true sampling variance for the estimator $\hat{\beta}$ can be written as

$$\sigma^2_{\hat{\beta}} = \frac{1}{n}\frac{Var[e_i(x_i - \bar{x})]}{Var[x_i]^2}.$$

Under the assumption of homoskedasticity, the equation simplifies to the conventional standard error

$$[\sigma^2_{\hat{\beta}}]_{conv} = \frac{1}{n}\frac{\sigma^2_e}{Var[x_i]}.$$

Thus,

$$[\sigma^2_{\hat{\beta}}]_{conv} > \sigma^2_{\hat{\beta}} \iff \sigma^2_e > \frac{Var[e_i(x_i - \bar{x})]}{Var[x_i]}.$$

Since

$$\begin{aligned} Var[e_i(x_i - \bar{x})] &= E[e_i^2(x_i - \bar{x})^2] \\ &= E[e_i^2]E[(x_i - \bar{x})^2] + Cov[e_i^2, (x_i - \bar{x})^2] \\ &= \sigma^2_e Var[x_i] + Cov[e_i^2, (x_i - \bar{x})^2], \end{aligned}$$

the inequality can be rewritten as

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

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 $Var[e_i|x_i] = E[e_i^2|x_i]$, the conditional variance of residual e_i.[2]

[1] A similar insight can be derived in the multivariate regression model by partialling out all other covariates.

[2] In his blog, Chris Auld gives an intuition for why conventional standard errors are biased up in the situation described above. As the error variance does not remain constant under this special form of heteroskedasticity but rather decreases for larger deviations of x_i from $\bar{x}$, observations further away from $\bar{x}$ provide more information for the estimation of $\sigma^2_{\hat{\beta}}$ than is actually assumed under homoskedasticity. Hence, conventional standard errors are larger than the true sampling variance. http://chrisauld.com/2012/10/31/the-intuition-of-robust-standard-errors/ .

When $Cov[e_i^2, (x_i - \bar{x})^2] < 0$, 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 heteroskedasticity as *elliptical heteroskedasticity*. Figure 1 illustrates the elliptical shape of the residuals based on simulated data exhibiting elliptical heteroskedasticity.

A reverse 'U'-shaped relation between the squared residual e_i^2 and the regressor x_i often occurs when elliptical heteroskedasticity is present. Hence, statistical procedures testing for linear forms of heteroskedasticity based on e_i^2 as the dependent variable usually fail to detect elliptical heteroskedasticity. Figure 2 illustrates how the linear regression line from the regression of e_i^2 on x_i is close to zero, as the squared residuals first rise and then fall in an increasing x_i. Therefore, tests such as the Breusch–Pagan (1979) test with x_i as the only independent variable included usually do not reject the hypothesis of homoskedasticity. Furthermore, more general tests, e.g., the White (1980) test, to detect also non-linear heteroskedasticity, do not give information about the form of heteroskedasticity that is present. This is because such test procedures test the null hypothesis of homoskedasticity against the unspecific alternative of no homoskedasticity. Moreover, due to their open formulation of null and alternative hypothesis, more general tests can possess a lower power in detecting elliptical heteroskedasticity.

To be able to test specifically for the kind of heteroskedasticity in which our interest lies, we developed a variant of the White test, which we call *specific White test*.

In the bivariate regression model, the test procedure for the standard White test is as follows. The squared residuals e_i^2 are obtained from the regression of y_i on x_i. Then the test statistic nR^2 is calculated by multiplying the number of observations, n, by R squared in the regression

$$e_i^2 = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \eta_i.$$

In the case of one regressor, nR^2 is asymptotically χ^2 distributed with two degrees of freedom. The decision rule for the rejection of H_0 is, as usual, based on this test statistic. The null hypothesis is H_0: *homoskedasticity* and the alternative hypothesis is H_a: *heteroskedasticity*.

To derive our specific White test, consider the regression

$$e_i^2 = \delta_0 + \delta_1 (x_i - \bar{x})^2 + \xi_i.$$

Expanding the equation yields

$$e_i^2 = \delta_0 + \delta_1 \bar{x}^2 - 2\bar{x}\delta_1 x_i + \delta_1 x_i^2 + \xi_i$$

This is the same regression equation on which is based the standard White test, in the bivariate regression model with $\gamma_0 = \delta_0 + \delta_1 \bar{x}^2$, $\gamma_1 = -2\bar{x}\delta_1$ and $\gamma_1 = \delta_1$.
Under elliptical heteroskedasticity, we know that $Cov[e_i^2, (x_i - \bar{x})^2] < 0$ and therefore

$$\delta_1 = \gamma_1 = \frac{Cov[e_i^2, (x_i - \bar{x})^2]}{Var[(x_i - \bar{x})^2]} < 0\,.$$

Hence, by exploiting this knowledge we can alter the standard White test statistic to test specifically for elliptical heteroskedasticity.
Our specific White test conducts a one-sided Wald test for $H_0 : \gamma_1 \geq 0$ against $H_a : \gamma_1 < 0$ in the regression $e_i^2 = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \eta_i$. The hypotheses are H_0: *no elliptical heteroskedasticity* and H_a: *elliptical heteroskedasticity.*
Furthermore, if the data exhibit elliptical heteroskedasticity, the usual Wald tests for hypotheses about the slope parameter β in the OLS regression of y_i on x_i using the conventional standard error give an actual size smaller than the desired type I error. As a result, policy conclusions based on estimates using conventional standard errors are conservative. In contrast, robust standard errors yield the correct size and valid policy conclusions.

3. Monte Carlo Simulations

To illustrate the issues arising from elliptical heteroskedasticity described in Section 2, we run a series of Monte Carlo simulations.
The design of our Monte Carlo simulations is based on the following data generating

process.

$$
\begin{aligned}
& y_i = 0.04x_i + e_i \\
& e_i = \sqrt{\frac{1}{\{(x_i - \bar{x})^2 + 0.1\}^a}}\ \epsilon_i \\
& x_i = [x_i^*],\ x_i^* \sim N(0.04, 1.8^2),\ \epsilon_i \sim N(0,1), \\
& a = 0,\ 0.05,\ 0.1,\ 0.15,\ \ldots,\ 0.5
\end{aligned}
$$

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 [.] 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 centered around the mean $\bar{x}$. The structure of the error term e_i implies that $Cov[e_i^2, (x_i - \bar{x})^2] < 0$ 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 heteroskedasticity. For $a = 0$, the error term is homoskedastic. The number of observations is set to $N = 214$, as in the original dataset, and the number of replications is 10,000.

In each simulation, we evaluate the size and power of three different tests for heteroskedasticity: the Koenker (1981) variant of the Breusch–Pagan test, which drops the assumption of normality, with x as the independent variable, the White test, and our specific White test introduced in Section 2. In addition, we compare the power and size for the parameter of interest in the causal model using Wald tests for the hypotheses $H_0 : \beta = k$ against $H_a : \beta \neq k$ where $k \in 0, 0.01, 0.02, \ldots, 0.5$ in the regression of y_i on x_i using robust and conventional standard errors.

Figure 3 shows power plots for the heteroskedasticity tests. The simulation with $a = 0$ gives the size of each test. While the rejection frequency of the Breusch–Pagan and standard White test is close to the given significance level of $\alpha = 5\%$, the actual size of the specific White test is above this value, with 11.5%. However, we find that the actual test size for the latter test approaches the theoretically given significance level for larger numbers of observations. For example, running an analogous simulation with $N = 2,140$

(21, 400) yields an actual size of 7.6% (6.1%) for the specific White test.

For $a > 0$, Figure 3 displays the power of each test. The rejection frequency of the White test and our specific White test increases with stronger elliptical heteroskedasticity, i.e., with increasing values of a. Compared to the specific White test, the standard White test performs worse in detecting heteroskedasticity, although the difference in power gets smaller for larger values of a. At $a = 0.15$, the specific White test rejects H_0 about 75% of the time, while the standard White test has a rejection frequency of about 20%. At $a = 0.3$, our specific White test has a power of about 99% while the standard White test rejects H_0 roughly eight out of ten times. As mentioned before, in contrast to the specific White test, the standard White test does not have elliptical heteroskedasticity as the alternative hypothesis, but rather heteroskedasticity in general, which may explain its worse performance. The Breusch–Pagan test has considerably smaller rejection frequencies than the two other tests throughout the whole range of $a > 0$. It does not reach a power of 4% 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 heteroskedasticity.

Figure 4 displays the power and size of the Wald tests with $H_0 : \beta = k, k \in 0, 0.01, 0.02, \ldots, 0.5$ for different values of a using conventional and robust standard errors. The actual size of the tests is given at $H_0 : \beta = 0.04$. Under homoskedasticity, $a = 0$, both test versions' sizes are close to the given significance level of 5%. In the presence of heteroskedasticity, $a > 0$, 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: from 5.7% at $a = 0$ to 0.1% at $a = 0.5$. 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.

For $H_0 : \beta \neq 0.04$, Figure 4 shows graphically the power of the Wald tests. At $a = 0$, the power curves of both tests are almost the same. However, an ever increasing gap between them arises as a gets larger. The Wald test using robust standard errors becomes more powerful whereas the test using conventional standard errors loses power. The loss in power can be attributed to the increasing upward bias in conventional standard errors for rising values of $a > 0$. As expected, the tests' power gets larger the further away the null hypothesis is from the true parameter $\beta = 0.04$.

The same conclusions can be drawn from Figure 5. It displays the power curves of the

Wald tests in three dimensional space, where the x-axis corresponds to the exponent a and the y-axis specifies the hypothesis about the true parameter β. The decreasing power and size of the Wald tests using conventional standard errors can be seen by the fact that the object in the left-hand side panel is tilted to the right. The right-hand side panel shows an increase in power of the Wald tests using robust standard errors for larger values of a, as the three dimensional power curve is tilted to the left. Using robust instead of conventional standard errors thus implies a power gain for the Wald tests. The valley of both objects runs along the correctly specified hypothesis at $H_0 : \beta = 0.04$.

4. Re-Estimation of Within-Twin Pair Regressions

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, Panels A and B, in Amin (2011) illustrates which data points he removed. However, we noticed that outlier number 2 in Panel B does not correspond to the data point labelled 2 in Panel A. Figure 5 shows that observation number 2 is actually the data point with a difference in log hourly wages of approximately -2 instead of the point at approximately -3. Despite this graphical error, the correct observations were removed in his analysis.

The left-hand side panel in Figure 5 suggests that the data exhibit the elliptical heteroskedasticity discussed in Section 2, which leads to an upward bias in conventional standard errors. In order to test for the presence of heteroskedasticity, we performed the three tests outlined in Section 2 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 column (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 south-east of the UK.

Table I 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 standard White test, and our proposed specific White test from Section 2. In the regressions including covariates, we partialled them out before testing. The specific White test rejects H_0*: no elliptical heteroskedasticity* in favour of H_a*: elliptical heteroskedasticity* for all regression specifications at least at the 10% level. In contrast, the Breusch–Pagan and the standard White test do not reject the hypothesis of homoskedasticity in any regression. This can be attributed to the difficulties and lower power in detecting elliptical heteroskedasticity when using more general tests discussed in Section 2. Based on our proposed specific White test, there is evidence for the presence of elliptical heteroskedasticity in the data. Hence, conventional standard errors are incorrect and may lead to false policy conclusions. Instead, robust standard errors should be used for the calculation of test statistics.

Table II shows our replication results for the within-twin pair regressions in Table 1, Amin (2011). We use the dataset from the online appendix for Bonjour et al. (2003) that was also used by Amin. Our replication results for the regressions based on the full sample are the same as the ones by Amin (2011). These regressions were also performed by Bonjour et al. (2003).[3] Our replication results for the regressions using the restricted samples are very similar to the estimates in Amin (2011). In addition to the replications using conventional standard errors, Table II 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.[4] This result is in line with the suspicion that elliptical heteroskedasticity is present in the data, which causes an upward bias in conventional standard errors. It also supports the conclusions from our specific White test.

In many regressions where the parameter of interest 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 three out of the 20 regressions do we fail to find a return to education significantly different from zero when using robust

[3] However, we find the parameter estimate of interest for the OLS regression without covariates using the full sample, in column (3), to be significant at the 10% level instead of insignificant, as stated incorrectly in Amin (2011). This is because $\frac{0.039}{0.023} = 1.696 > 1.645$. Our conclusion remains valid when the values are rounded to more than three decimals. Bonjour et al. (2003) do not indicate significance at the 10% level.

[4] In the pooled OLS and IV regressions in Table 1, Amin (2011), robust standard errors are slightly larger than the conventional ones. Nevertheless, all point estimates remain highly significant at the 1% level.

standard errors. In particular, all point estimates based on the full sample as well as the sample excluding observations with an absolute wage difference of more than 90 and 75, respectively, are significant at the usual levels. Regarding the regressions based on samples with three or four outliers removed (row 3 and 4), three more estimates turn significant at least at the 10% level when using robust standard errors compared to the results using conventional standard errors. One of the main findings in Bonjour et al. (2003), a return to education of 7.7 percent (column (4)), remains significant at the 10% level after the exclusion of outliers with an absolute wage difference larger than 90 and 75, when using robust standard errors. However, dropping further observations causes the parameter estimate of interest to become insignificant even with robust standard errors. This is the same result as in Amin (2011), who used conventional standard errors.[5]

The point estimates for the parameter of interest generally decrease with the number of outliers removed. For example, after excluding all four extreme values, $\hat{\beta}$ falls from originally 0.082 to 0.041 in the IV regression with covariates (column (8)). Under conventional standard errors this effect and many other estimates are insignificant at the 10% level. Therefore, the conclusion in the case of conventional standard errors would falsely be that the return to education is not significantly different from zero. However, when using robust standard errors, which need to be used due to the presence of elliptical heteroskedasticity, most of these insignificant effects are significant at conventional levels. Thus, the correct conclusion based on the vast majority of regressions in Table II is that there is a positive return to education which is significant at conventional levels.

5. Conclusion

Bonjour et al. (2003) estimated the private returns to education using a sample of 428 female monozygotic twins from the UK. Most of their regression specifications yielded a positive return to one additional year of education, significant at conventional levels. Amin (2011) re-estimated their regressions but excluding up to four outliers from the dataset. By doing so, he found that many of Bonjour et al.'s within-twin pair estimates became smaller in magnitude and either insignificant or significant only at lower levels.

In this study we show that the inference for the within-twin pair regressions in Amin (2011) is mostly incorrect, due to a special non-linear form of heteroskedasticity present

[5] Although the IV regression based on the sample with outlier 1 removed is significant at the 10% level in Amin (2011) (row 2, column(4)), our corresponding estimate is insignificant at the 10% level

in the data. This special form of heteroskedasticity, referred to by us as *elliptical heteroskedasticity*, causes an upward bias in conventional standard errors, which are used by Amin (2011).

With the help of three tests for heteroskedasticity, we tested for the presence of elliptical heteroskedasticity. A basic variant of the Breusch–Pagan test, which is only able to detect linear forms of heteroskedasticity, and the standard White test, which can also detect non-linear heteroskedasticity, do not reject the null hypothesis of homoskedasticity for any regression. This is as expected, since the former procedure cannot detect non-linear heteroskedasticity and the latter test has often low power under this type of heteroskedasticity. However, our specific White test, developed for testing for elliptical heteroskedasticity, rejects the null hypothesis of *no elliptical heteroskedasticity* in favour of the alternative hypothesis *elliptical heteroskedasticity* for most regressions and samples. Our Monte Carlo simulations assessing the size and power of the three tests for heteroskedasticity confirm the problems in detecting elliptical heteroskedasticity with the usual tests and demonstrate the superiority of our own test.

We obtain the correct inference for the within-twin pair regressions in Amin (2011) by re-estimating these regressions using robust instead of conventional standard errors. In contrast to Amin (2011), we found a positive return to education, significant at conventional levels, for most of his within-twin pair regressions. For example, the IV regression including covariates based on the sample with all outliers removed yields an estimated return to education of 4.1 percent. The estimate is significant at the 10% level using robust standard errors as opposed to insignificant under conventional ones. However, our result for the favoured regression specification in Bonjour et al. (2003) is similar to Amin (2011). Removing outliers with an absolute wage difference of 90 and 75, respectively, leads to a decrease in the return to education from 7.7 percent to 5 percent. The latter estimate is significant at the 10% level. In summary, we come to the conclusion that most within-twin pair regressions in Amin (2011) support the notion of a significant positive return to education, although smaller than the results obtained by Bonjour et al. (2003).

Table I. Tests for Heteroskedasticity Based on Within-Twin Pair Regressions in Table 1, Amin (2011)

		Within-twin pair		Within-twin pair with covariates	
		OLS	IV	OLS	IV
Sample		(3)	(4)	(7)	(8)
Full Bonjour et al. dataset	Breusch–Pagan Test	0.3645	0.3124	0.6090	0.5435
	White Test	0.4805	0.4581	0.5300	0.5603
	Specific White Test	0.0257	0.0691	0.0129	0.0391
Drop if abs. wage difference > 90	Breusch–Pagan Test	0.7221	0.6719	0.8713	0.8276
	White Test	0.4982	0.4906	0.5451	0.5491
	Specific White Test	0.0105	0.0122	0.0219	0.0247
Drop if abs. wage difference > 75	Breusch–Pagan Test	0.7207	0.6737	0.7799	0.7437
	White Test	0.5488	0.5421	0.6034	0.6075
	Specific White Test	0.0171	0.0198	0.0389	0.0455
Drop if abs. wage difference > 65	Breusch–Pagan Test	0.7143	0.7065	0.8136	0.8126
	White Test	0.6147	0.6101	0.6562	0.6559
	Specific White Test	0.0294	0.0292	0.0518	0.0518
Drop if abs. wage difference > 60	Breusch–Pagan Test	0.9861	0.9466	0.8177	0.8014
	White Test	0.6310	0.6237	0.7247	0.7194
	Specific White Test	0.0572	0.0549	0.0948	0.0948

Table reports p values. For more details on the tests, see Section 2.

Table II. Replication and Re-Estimation Using Robust Standard Errors of Within-Twin Pair Regressions in Table 1, Amin(2011)

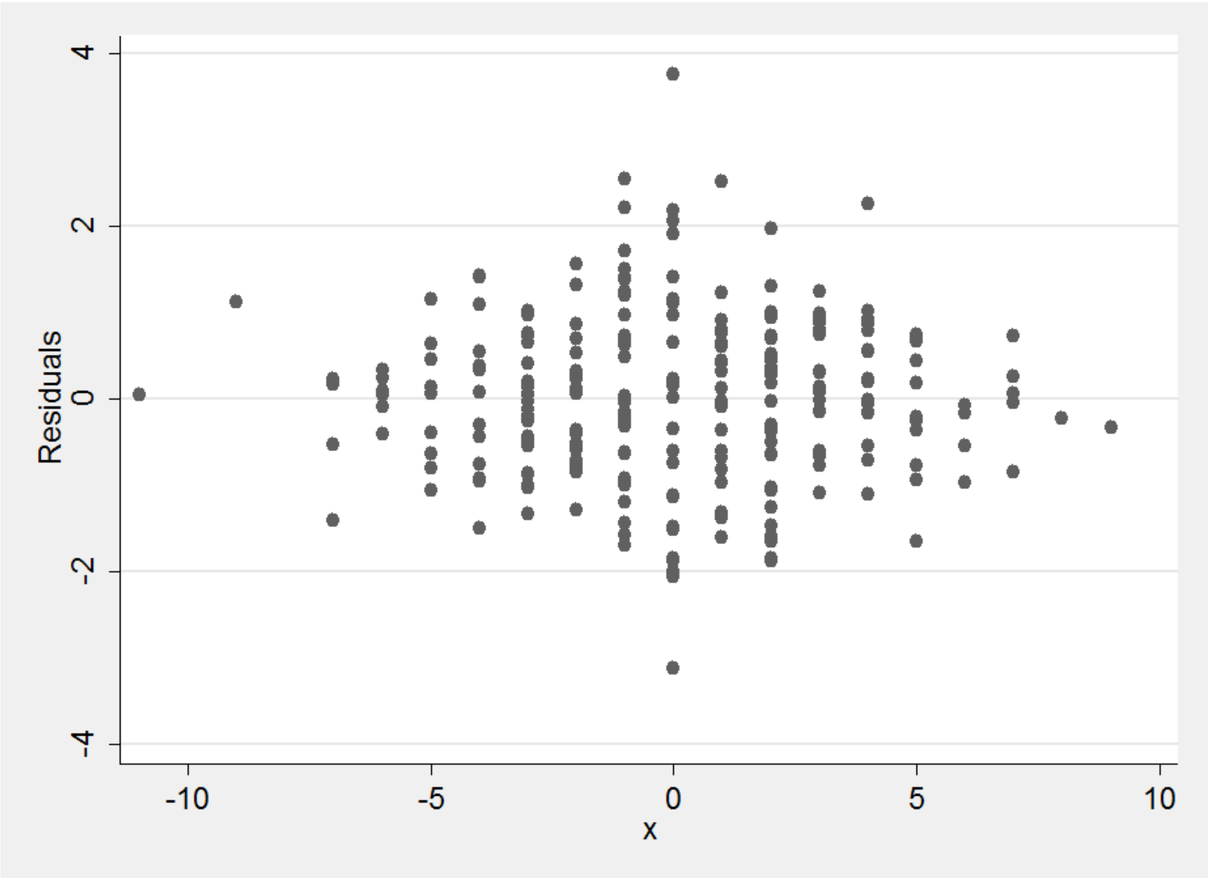
		Within-twin pair		Within-twin pair with covariates	
		OLS	IV	OLS	IV
Sample		(3)	(4)	(7)	(8)
Full Bonjour et al.dataset	$\hat{\beta}_{education}$	0.039	0.077	0.039	0.082
	Conventional SE	(0.023)*	(0.033)**	(0.024)	(0.036)**
	Robust SE	(0.018)**	(0.039)**	(0.018)**	(0.043)*
Drop if abs. wage difference > 90	$\hat{\beta}_{education}$	0.032	0.050	0.034	0.053
	Conventional SE	(0.021)	(0.031)	(0.023)	(0.033)
	Robust SE	(0.016)**	(0.027)*	(0.017)**	(0.030)*
Drop if abs. wage difference > 75	$\hat{\beta}_{education}$	0.032	0.050	0.036	0.055
	Conventional SE	(0.021)	(0.030)*	(0.022)	(0.032)*
	Robust SE	(0.016)**	(0.027)*	(0.017)**	(0.030)*
Drop if abs. wage difference > 65	$\hat{\beta}_{education}$	0.032	0.036	0.036	0.039
	Conventional SE	(0.020)	(0.029)	(0.021)*	(0.031)
	Robust SE	(0.016)**	(0.022)	(0.017)**	(0.024)
Drop if abs. wage difference > 60	$\hat{\beta}_{education}$	0.028	0.036	0.036	0.041
	Conventional SE	(0.019)	(0.027)	(0.019)*	(0.028)
	Robust SE	(0.016)*	(0.022)	(0.016)**	(0.023)*

*** Significant at the 1% level

** Significant at the 5% level

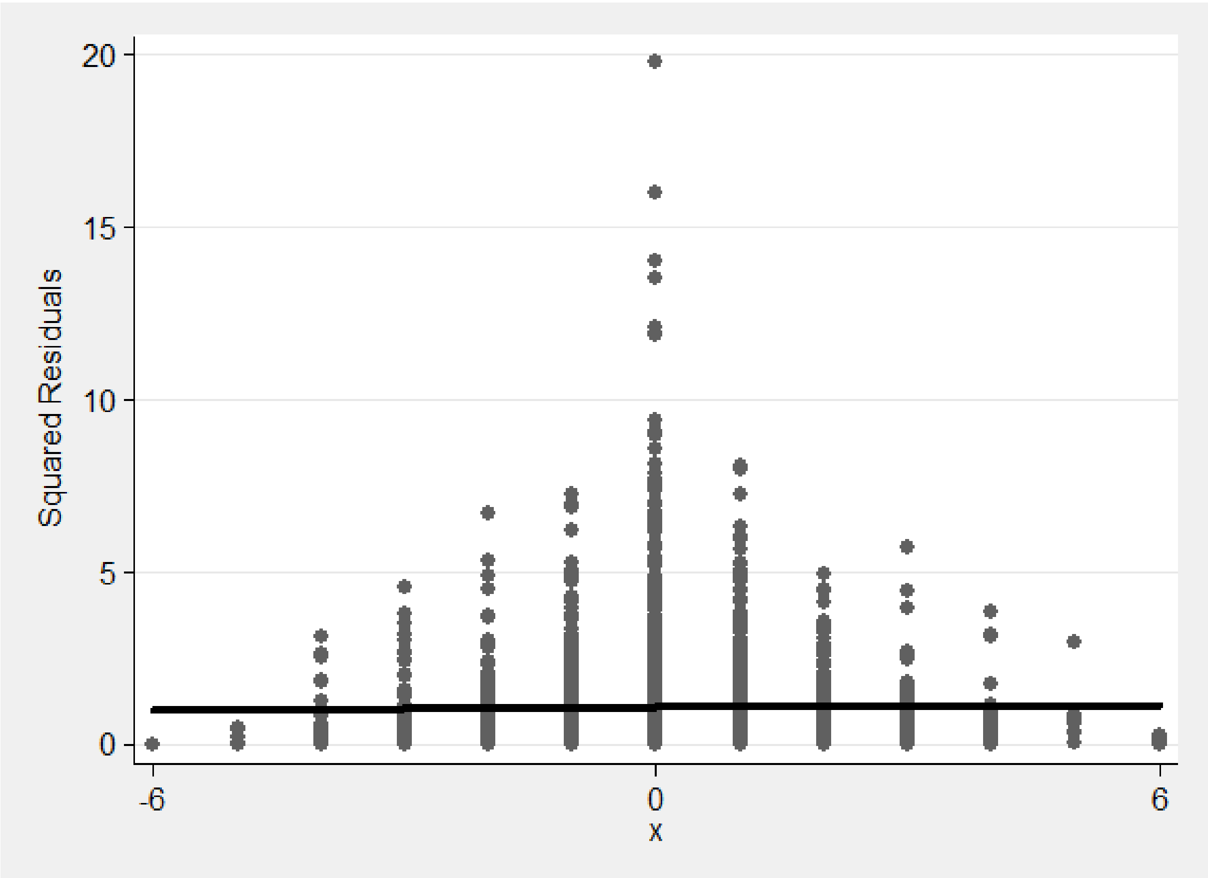
* Significant at the 10% level

Figure 1. Scatter Plot Illustrating Elliptical Heteroskedasticity



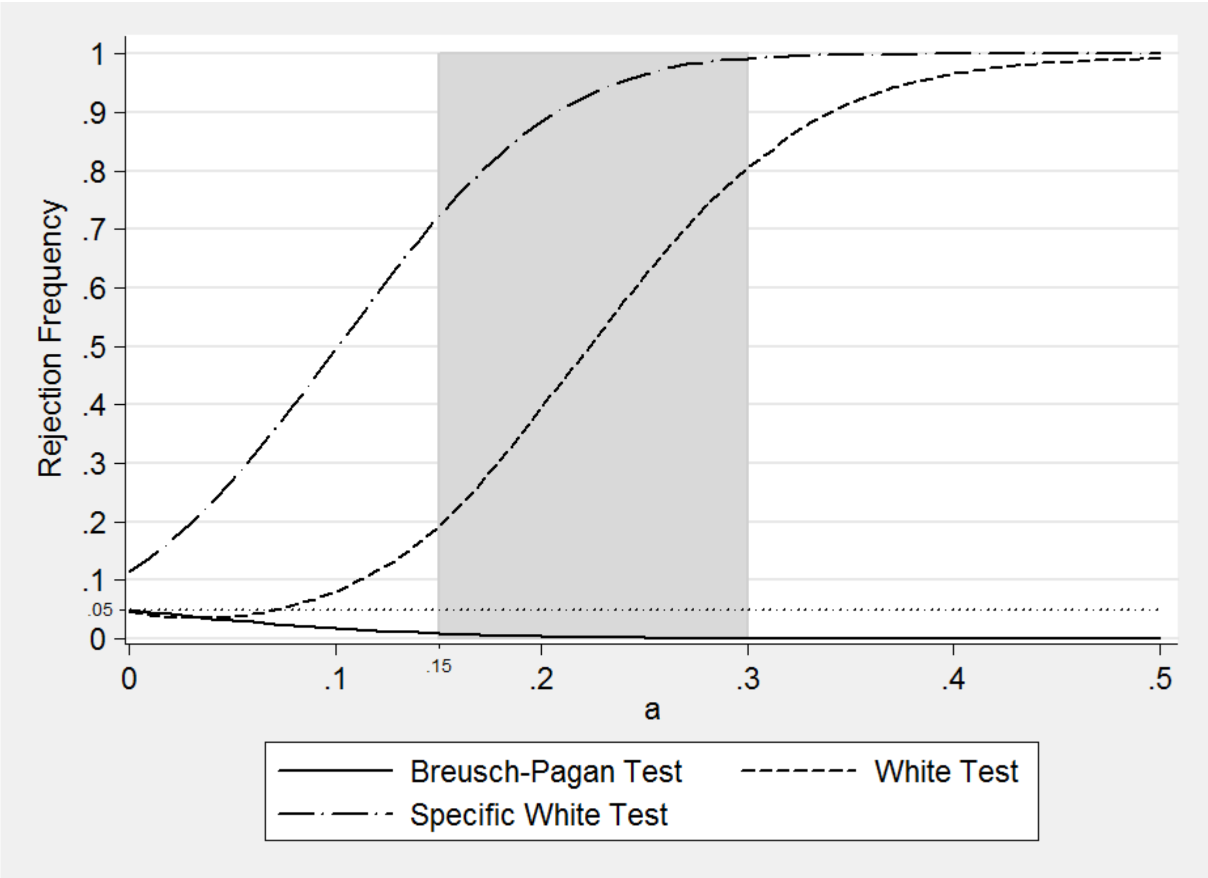
Graph shows simulated data based on the data generating process in Section 3. $N = 250$; $a = 0.2$.

Figure 2. Regression of Squared Residuals on x



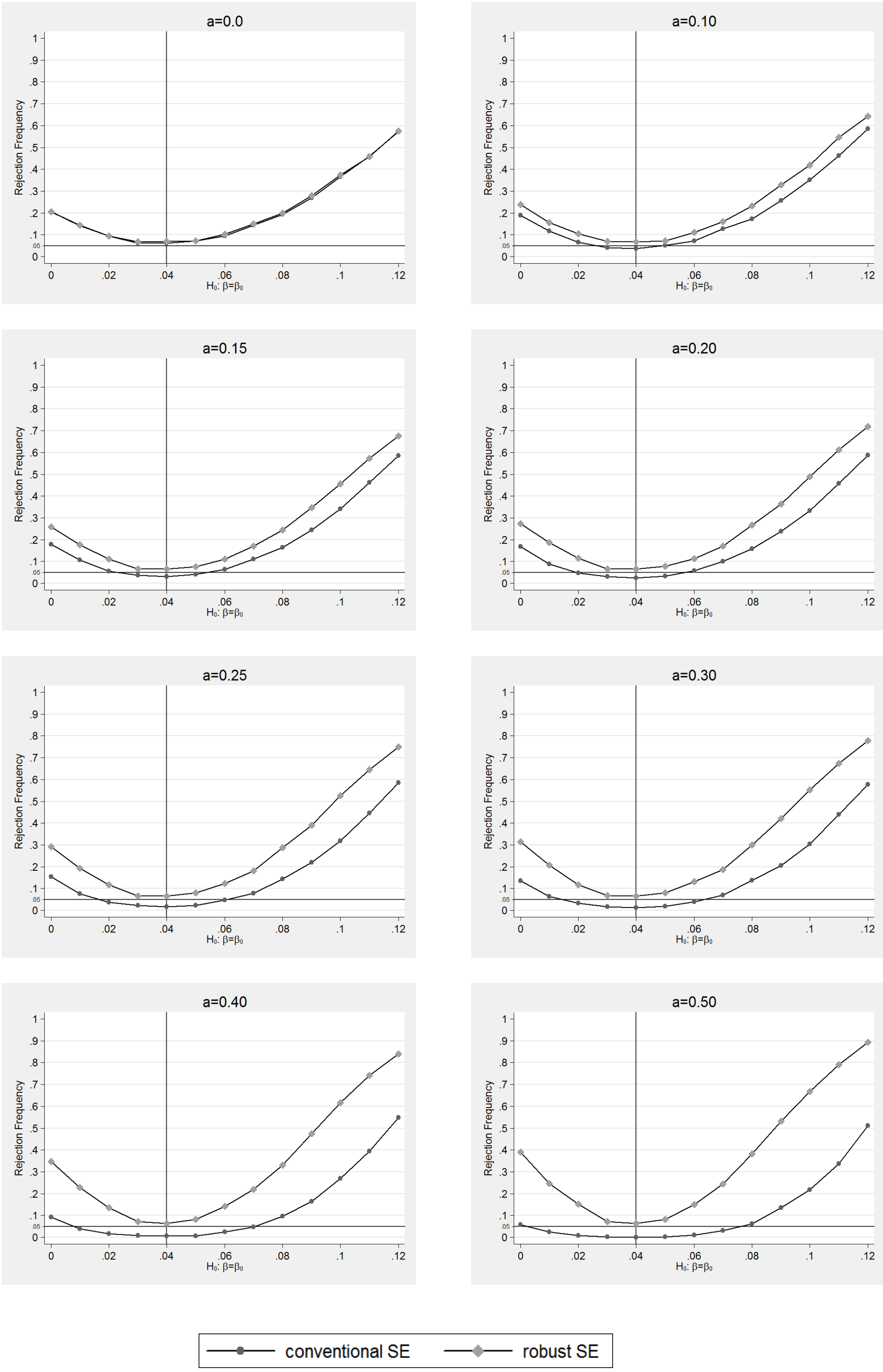
Graph shows simulated data based on the data generating process in Section 3. $N = 2140$; $a = 0.25$.

Figure 3. Power Plots for Heteroskedasticity Tests



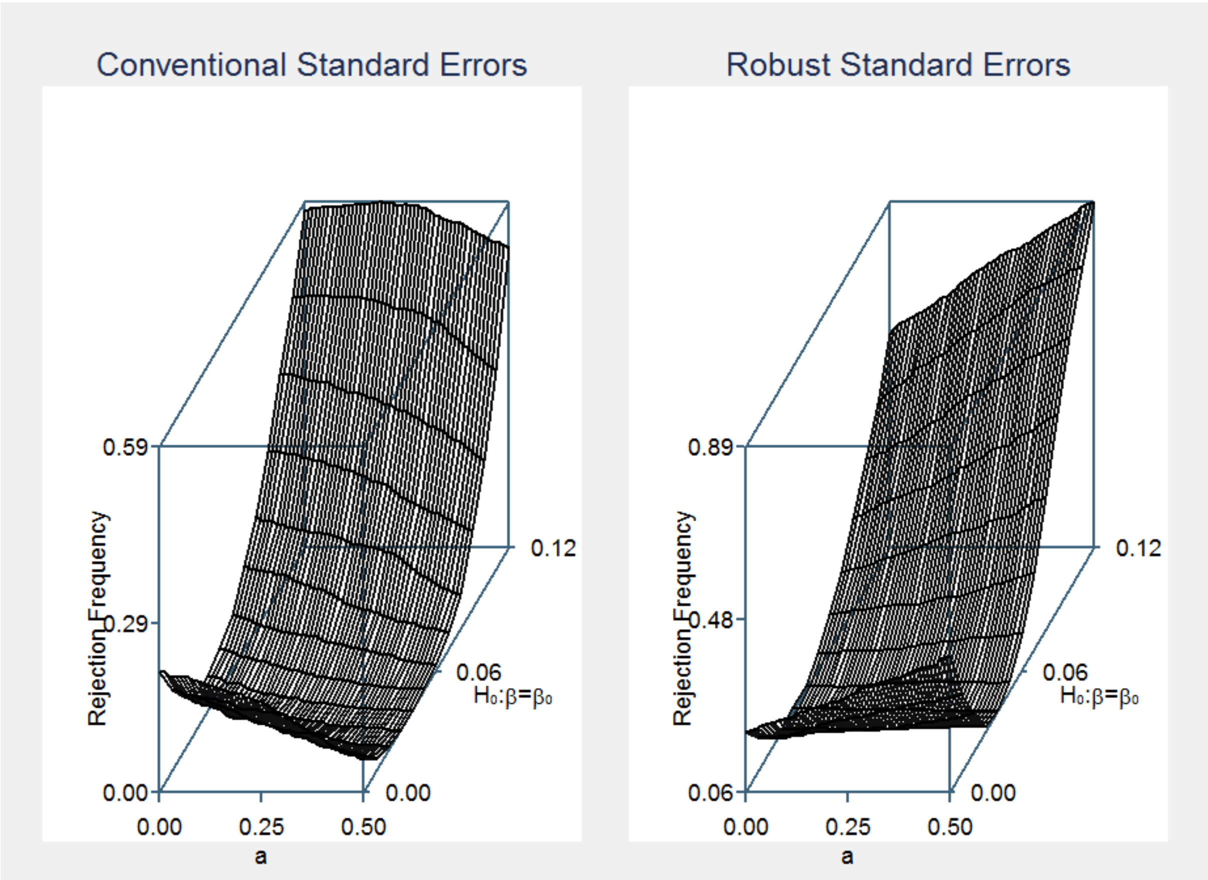
$N = 214$; $10,000$ replications; $a = 0$: homoskedasticity; $a > 0$: elliptical heteroskedasticity. Breusch–Pagan and White Test: Rejection frequencies for H_0 : *Homoskedasticity* and H_a : *Heteroskedasticity*; Specific White Test: Rejection frequencies for H_0: *No elliptical heteroskedasticity* and H_a: *Elliptical heteroskedasticity*, $\alpha = 5\%$.

Figure 4. Power Plots for Wald Tests Using Conventional and Robust Standard Errors



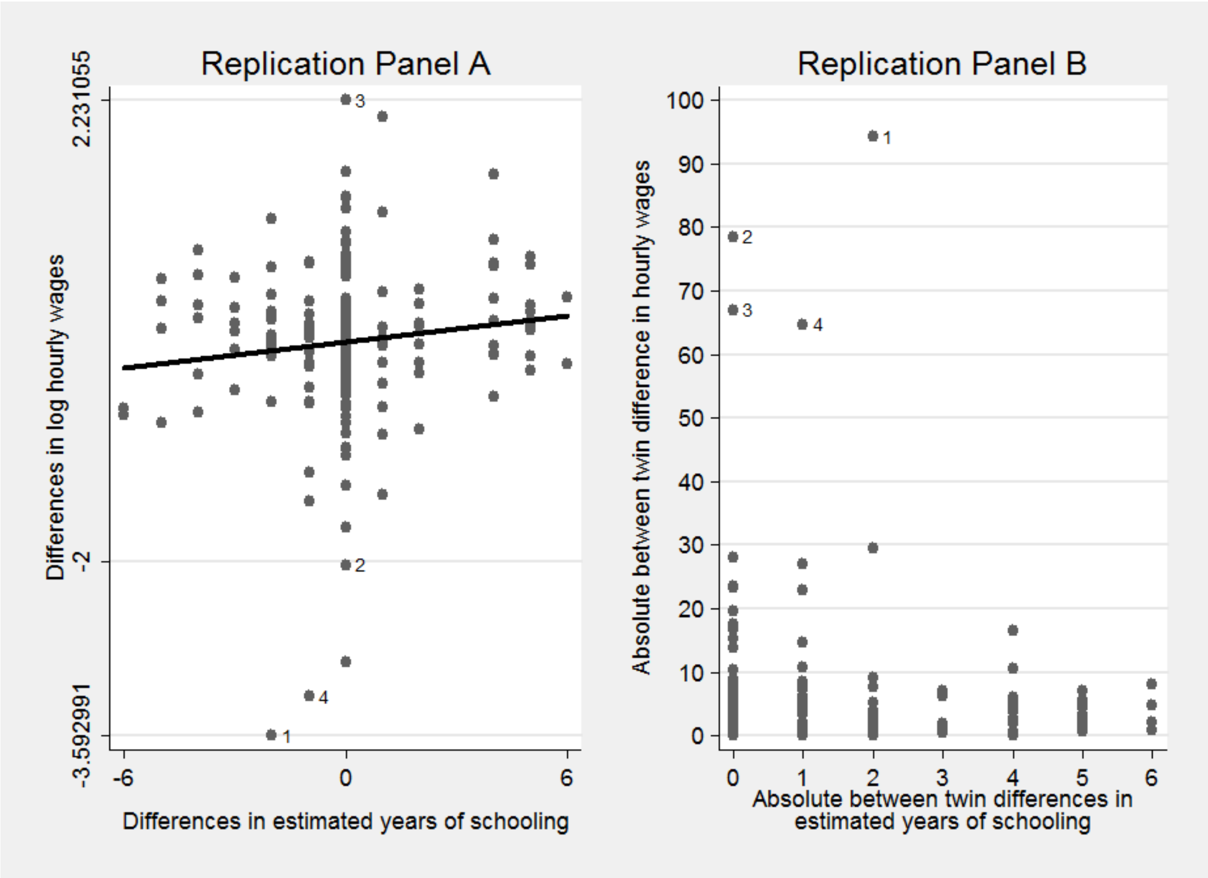
$N = 214$; $10,000$ replications; $\beta = 0.04$; $a = 0$: homoskedasticity; $a > 0$: elliptical heteroskedasticity. Rejection frequencies for $H_0 : \beta = k, k \in 0, 0.1, 0.2, \ldots, 0.5$ using Wald tests with conventional and robust standard errors; $\alpha = 0.05$.

Figure 5. Three Dimensional Power Plot for Wald Tests Using Conventional and Robust Standard Errors



$N = 214$; 10,000 replications; $\beta = 0.04$; $a = 0$: homoskedasticity; $a > 0$: elliptical heteroskedasticity. Rejection frequencies for $H_0 : \beta = k, k \in 0, 0.1, 0.2, \ldots, 0.5$ using Wald tests with conventional and robust standard errors; $\alpha = 0.05$.

Figure 6. Replication of Figure 1, Amin (2011)



References

Amin, V. (2011). Returns to Education: Evidence from UK Twins: Comment. *American Economic Review*, *101*(4), 1629–1635.

Angrist, J., & Pischke, J.-S. (2010). *A Note on Bias in Conventional Standard Errors under Heteroskedasticity*. Retrieved from `http://econ.lse.ac.uk/staff/spischke/mhe/josh/Notes%20on%20conv%20std%20error.pdf`

Auld, C. (2012). *The Intuition of Robust Standard Errors*. Retrieved from `http://chrisauld.com/2012/10/31/the-intuition-of-robust-standard-errors/`

Bonjour, D., Cherkas, L. F., Haskel, J. E., Hawkes, D. D., & Spector, T. D. (2003). Returns to Education: Evidence from the U.K. Twins. *American Economic Review*, *93*(5), 1799–1812.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 987–1007.

Koenker, R. (1981). A Note on Studentizing a Test for Heteroskedasticity. *Journal of Econometrics*, *17*(1), 107–112.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817–838.