2. Breusch-Pagan Extended

Consider a linear regression of the form

$$y = \alpha + \beta x + u, \tag{1}$$

with (y, x) both scalar random variables, where it is assumed that (a.i) $\mathbb{E}(u \cdot x) = \mathbb{E}u = 0$ and (a.ii) $\mathbb{E}(u^2|x) = \sigma^2$.

(1) The condition a.i is essentially untestable; explain why.
(2) Breusch and Pagan (1979) argue that one can test a.ii via an auxiliary regression $\hat{u}^2 = c+dx+e$, where the $\hat{u}$ are the residuals from the first regression, and the test of a.ii then becomes a test of $H_0 : d = 0$. Describe the logic of the test of a.ii.
(3) Use the two conditions a.i and a.ii to construct a GMM version of the Breusch-Pagan test.
(4) What can you say about the performance or relative merits of the Bruesch-Pagan test versus your GMM alternative?
(5) Suppose that in fact that x is distributed uniformly over the interval $[0, 2\pi]$, and $\mathbb{E}(u^2|x) = \sigma^2(x) = \sigma^2 \sin(2x)$, thus violating a.ii. What can you say about the performance of the Breusch-Pagan test in this circumstance? Can you modify your GMM test to provide a superior alternative?
(6) In the above, we've considered a test of a specific functional form for the variance of u. Suppose instead that we don't have any prior information regarding the form of $\mathbb{E}(u^2|x) = f(x)$. Discuss how you might go about constructing an extended version of the Breusch-Pagan test which tests for $f(x)$ non-constant.
(7) Show that you can use your ideas about estimating $f(x)$ to construct a more efficient estimator of β if $f(x)$ isn't constant. Relate your estimator to the optimal generalized least squares (GLS) estimator.

(1)
Since u is not directly observable, we cannot measure its covariance with x, which is necessary to test whether E(u · x) = 0. We can only estimate the covariance between y and x, which is the sum of the covariance between u and x and the covariance between βx and x, but this does not allow us to separate the effect of the error term from the effect of the independent variable.

Therefore, we cannot directly test whether a.i holds, but we can make assumptions or arguments about why it might be plausible. For example, if we believe that x is exogenously determined and does not depend on any unobservable factors that affect u, then we might argue that E(u · x) = 0 is a reasonable assumption. However, such assumptions or arguments cannot be empirically verified, and thus, the condition a.i is essentially untestable.

(2)
The logic of the test is as follows:

Estimate the regression model (1) and obtain the residuals u_hat = y - α - βx. Regress the squared residuals, u_hat_squared, on the independent variable, x, and obtain the residuals of this regression, e. Test the null hypothesis that the slope coefficient in this auxiliary regression is zero, that is, H0: d = 0.

If the null hypothesis is not rejected, then there is no evidence of conditional heteroskedasticity, and we can assume that the variance of the error term is constant conditional on x, and thus a.ii holds. On the other hand, if the null hypothesis is rejected, then there is evidence of conditional heteroskedasticity, and we cannot assume that the variance of the error term is constant conditional on x, and thus a.ii does not hold.

The intuition behind this test is that if a.ii holds, then the squared residuals should not be related to the independent variable, x, and the slope coefficient in the auxiliary regression should be zero. However, if there is conditional heteroskedasticity, then the variance of the error term depends on x, and the squared residuals should be related to x, and the slope coefficient in the auxiliary regression should be nonzero.

Therefore, the test proposed by Breusch and Pagan (1979) allows us to test for conditional heteroskedasticity in the linear regression model (1) by examining the relationship between the squared residuals and the independent variable, x.

(3)
Step 1: Identify the beliefs and assertions
Step 2: Translate it into data E((y - α + βx). x) = 0; E((y - α + βx)^2 | x) = σ2
Step 3: Going from expectations to yi, xi
G_hat(N) = [(1/N) Σ((yi - α + βxi). xi)), (1/N) Σ(xi(yi - α + βxi)^2) - σ2)]

GMM perfectly minimizes the objective function. Here, we want to construct a test to learn how well the minimization works. Right now, we have two unknowns and two equations. Only in the overidentified case, there is a minimum. But we can also add more moments to create an overidentified model. For example, we can add another condition that E(u^2) = σ2. That is σ does not vary with x at all.

(4)
Good: Our test is a little more general than the Bruesch-Pagan test.
Bad: Our test is checking for three things at once. One, is β a good estimate? Two, is σ2 a good estimate? Three, is the assumption that σ2 does not depend on x reasonable? If one of these conditions is off, for example is the last assumption does not hold, it can also affect how good my estimate of β is as well.

(5)
Performance of the BP test would be worse than GMM in this case. BP would predict dhat = 0 (i.e., flat slope) and lead us to accept the null hypothesis, which we shouldn't since we know $E(u^2|x)$ = σ2 (x)= σ2 sin(2x). GMM also uses the flat line, but leads us to reject because of the large errors between σ2 sin(2x) and the flat line.

We could modify our GMM test by adding in the moment condition $E(u^2|x)$ = σ2 (x)= σ2 sin(2x).

(6)
The standard Breusch-Pagan test assumes that the conditional variance of the error term u is a constant, $Var(u|x) = \sigma^2$ for all values of x. In this situation, we might go about constructing an extended version of the Breusch-Pagan test which tests for f(x) non-constant by:

1. Estimating the linear regression model y = a + Bx + u and obtaining the residuals u_hat

2. Calculating the squared residuals u_hat^2 and regressing them on the explanatory X variables

3. Estimating the following regression model: $u\_hat^2 = c + d1*x + d2*x^2 + ... + dk*x^k + v$

   where c is a constant, d1, d2, ..., dk are coefficients to be estimated, $x^2, x^3, ..., x^k$ are higher-order terms of x, and v is the error term

4. Calculating the test statistic: $LM = n*R^2$

   where n is the sample size, R^2 is the coefficient of determination from the regression model in step 3, and LM follows a chi-square distribution with k degrees of freedom under the null hypothesis of constant variance

5. Compare the test statistic LM with the critical value of the chi-square distribution and reject the null hypothesis (of constant variance) if LM exceeds the critical value. If the null hypothesis is rejected, it indicates that the variance of the error term u is *not* constant across different values of X.

(7)
We can construct a more efficient estimator of B if f(x) isn't constant by using weighted least squares (WLS), where the weights are inversely proportional to $var(u^2|x)$. We could estimate f(x) by regressing u^2 on X explanatory variables to get fhat(x). Then, use WLS to estimate B by weighting each observation by 1/fhat(x) using:

$B_{WLS} = (X'W^{-1}X)^{-1}X'W^{-1}y$

where X are explanatory variables, y is the vector of dependent variable values, and W is the diagonal matrix of weights, with the i-th diagonal element equal to $1/fhat(x_i)$

B_{WLS} is equivalent to the optimal GLS estimator of B, which assumes that the covariance matrix of the error term u is not constant but can be estimated from the data.