Learn How to Test for Heteroscedasticity with the Goldfeld-Quandt Test in Python

Authored by

Mohammed looti

October 26, 2025

In the crucial field of statistical modeling, particularly when employing linear regression techniques, the reliability of our conclusions rests heavily on satisfying several core assumptions. One of the most fundamental requirements is homoscedasticity. This condition dictates that the variance of the residuals--the differences between observed and predicted values--must remain constant across all observations and all levels of the predictor variables. When this assumption is violated, we encounter a severe methodological issue known as heteroscedasticity.

Heteroscedasticity describes a situation where the scatter or spread of the residuals changes systematically with the magnitude of the response variable or the predictors in a regression model. The presence of unequal variance does not bias the coefficient estimates themselves, but it fundamentally distorts the calculation of the standard errors. This, in turn, invalidates the resulting p-values and confidence intervals, thereby compromising the entire foundation of statistical inference drawn from the model. Therefore, identifying and addressing this violation is paramount for producing reliable research.

To formally diagnose this problem, statisticians rely on powerful diagnostic tools. Among the most well-established and robust methods is the Goldfeld-Quandt test. This comprehensive tutorial provides a detailed, practical walkthrough, demonstrating exactly how to execute the Goldfeld-Quandt test efficiently within the Python environment, utilizing the capabilities of the widely trusted statsmodels library. By following these steps, you will gain the ability to verify the homoscedasticity assumption in your own models and ensure the trustworthiness of your quantitative findings.

Understanding Heteroscedasticity and the Goldfeld-Quandt Test Mechanism

A deep understanding of heteroscedasticity is essential for anyone engaged in serious quantitative modeling. The ideal state in linear regression models is homoscedasticity, where the variance of the error terms (or residuals) remains unchanging across the full range of the independent variables. This stability is one of the core assumptions underpinning Ordinary Least Squares (OLS) regression, a method prized for its simplicity and the best linear unbiased estimator (BLUE) property it guarantees when assumptions hold.

When heteroscedasticity takes root, the residual variance systematically changes, often increasing or decreasing as the values of the independent variables or the predicted outcomes shift. Visually, this often manifests in a residual plot as a clear "funnel" or "cone" shape, indicating that the predictive accuracy of the model varies widely depending on where you are in the data distribution. While the coefficient estimates themselves are not systematically biased by this issue, the standard procedures for calculating the uncertainty around these estimates (the standard errors) are flawed, leading to misstated p-values and unreliable confidence intervals.

The Goldfeld-Quandt test is explicitly designed to detect this systematic change in variance. The test operates by strategically dividing the entire set of observations into two distinct subgroups.

These subgroups are typically defined by sorting the data based on a potential source of heteroscedasticity (usually one of the predictor variables or the fitted values) and then partitioning the data into lower and upper sections, often discarding a portion of the middle observations to enhance the statistical power of the comparison.

Once divided, the Goldfeld-Quandt test fits separate OLS regression models to each of the two sub-samples. The test then compares the ratio of the unexplained variance (the sum of residual squares) from the two models using an F-test. A large F-statistic, indicating a significant difference in variance between the two groups, serves as evidence to reject the null hypothesis of homoscedasticity and confirm the presence of heteroscedasticity.

Step 1: Preparing Your Dataset in Python

The successful execution of any statistical diagnostic, including the Goldfeld-Quandt test, begins with meticulous data preparation. For the purposes of this tutorial, we will construct a synthetic Pandas DataFrame in Python. This controlled dataset, consisting of 13 observations, simulates a scenario where we investigate the factors influencing student performance, allowing us to accurately demonstrate the subsequent statistical modeling and testing steps.

Our sample dataset is structured with three core variables: 'hours' (representing the hours studied by the student), 'exams' (representing the number of preparatory exams taken), and 'score' (representing the final exam score, our outcome of interest). The variables 'hours' and 'exams' will function as our predictor variables, while 'score' is designated as the response variable. We must ensure the data is properly imported and structured before proceeding to model fitting. The following code snippet initializes this data structure:

import pandas as pd

```
#create DataFrame
df = pd.DataFrame({'hours': ,
'exams': ,
'score': })

#view DataFrame
print(df)

hours exams score
0 1 1 76
1 2 3 78
2 2 3 85
3 4 5 88
```

4 2 2 72

5 1 2 69

6 5 1 94

7 4 1 94

8 2 0 88

9 4 3 92

10 4 4 90

11 3 3 75

12 6 2 96

The code successfully generates our `df` Pandas DataFrame, containing 13 rows of student data. It is vital to confirm that the data types and structure align with the expectations of the statsmodels library. This organized structure ensures that the subsequent regression analysis can be performed without structural errors, setting the stage for the diagnostic test.

Step 2: Fitting a Multiple Linear Regression Model

Once the dataset is prepared, the next logical step in the diagnostic process is to fit the primary multiple linear regression model. This model seeks to establish a linear relationship between our outcome, 'score', and our multiple predictors, 'hours' and 'exams'. We will leverage statsmodels, the preferred Python library for statistical modeling, specifically using its implementation of OLS regression.

A necessary precursor to fitting the OLS model is the inclusion of a constant term. This is handled easily in Python using the `sm.add_constant()` function. The constant term, or intercept, is mathematically crucial as it allows the regression plane to shift vertically, capturing the expected mean of the response variable when all predictor variables are held at zero. We define our response variable `y` and our augmented matrix of predictor variables `x` before passing them to the model fitting function.

The OLS method calculates the coefficients that minimize the sum of the squared residuals. This step is essential because the <a href="https://en.wikipedia.org/wiki/Goldfeld%E2%80%93Quandt_test relies entirely on the set of residuals generated by this fitted model. The quality of the regression itself (indicated by metrics like R-squared) provides context, but the primary output required here are the calculated residuals, which will be analyzed for patterns of unequal variance.

Executing the model fitting code provides us with a detailed summary:

import statsmodels.api as sm

```
#define predictor and response variables
y = df
x = df]

#add constant to predictor variables
x = sm.add_constant(x)

#fit linear regression model
model = sm.OLS(y, x).fit()

#view model summary
print(model.summary())

OLS Regression Results
==============================================================================
===
Dep. Variable: score R-squared: 0.718
Model: OLS Adj. R-squared: 0.661
Method: Least Squares F-statistic: 12.70
Date: Mon, 31 Oct 2022 Prob (F-statistic): 0.00180
Time: 09:22:56 Log-Likelihood: -38.618
No. Observations: 13 AIC: 83.24
Df Residuals: 10 BIC: 84.93
Df Model: 2
Covariance Type: nonrobust
==============================================================================
===
coef std err t P>|t|
------------------------------------------------------------------------------
const 71.4048 4.001 17.847 0.000 62.490 80.319
hours 5.1275 1.018 5.038 0.001 2.860 7.395
exams -1.2121 1.147 -1.057 0.315 -3.768 1.344
==============================================================================
===
Omnibus: 1.103 Durbin-Watson: 1.248
Prob(Omnibus): 0.576 Jarque-Bera (JB): 0.803
Skew: -0.289 Prob(JB): 0.669
Kurtosis: 1.928 Cond. No. 11.7
==============================================================================
===
```

The resulting summary confirms that our model explains 71.8% of the variability in student scores (R-squared = 0.718). However, we must view the standard errors and p-values presented in the coefficient table with caution. If heteroscedasticity were present, these uncertainty estimates would be invalid. The following step addresses this crucial reliability check using the Goldfeld-Quandt test.

Step 3: Executing the Goldfeld-Quandt Test in Python

With our linear regression model successfully fitted, we now perform the formal diagnostic test for heteroscedasticity. The statsmodels library provides the convenient function `het_goldfeldquandt()`, located within the `statsmodels.stats.diagnostic` module, specifically tailored for this rigorous statistical evaluation.

The Goldfeld-Quandt test requires sorting the dataset based on the variable suspected of causing the unequal variance, which is automatically handled by the function if the predictor matrix `x` is provided. The key methodological step is the splitting of the data into three segments: a lower group, a central group that is dropped, and an upper group. Dropping the central observations is a deliberate strategy to maximize the difference in variance between the remaining two extreme groups, thereby increasing the power of the subsequent F-test comparison.

The parameter `drop` specifies the proportion of central observations to exclude. While the optimal proportion can vary, removing approximately 20% to 25% of the data from the middle is a widely accepted practice to ensure adequate separation and distinctness between the low-variance and high-variance regions (if heteroscedasticity exists). For our analysis, we utilize `drop=0.2`, excluding 20% of the observations to clearly compare the variances of the residuals generated by the two separate OLS regression models fitted to the remaining sub-samples. The input parameters required are the response variable `y` and the predictor matrix `x` from our fitted model.

```
#perform Goldfeld-Quandt test
sm.stats.diagnostic.het_goldfeldquandt(y, x, drop=0.2)
```

(1.7574505407790355, 0.38270288684680076, 'increasing')

The execution of the `het_goldfeldquandt()` function yields a tuple containing three crucial elements: the calculated test statistic, the corresponding p-value, and an indication of the direction of the alternative hypothesis (in this case, 'increasing' variance). Our output shows an F-statistic of approximately 1.757 and a p-value of about 0.383. This quantitative result is what we will use to make our formal statistical determination regarding the presence of heteroscedasticity.

Interpreting the Goldfeld-Quandt Test Results

The final step in the diagnostic process is interpreting the output of the Goldfeld-Quandt test, which determines whether the assumption of homoscedasticity holds. The decision rests on comparing the calculated p-value against a predefined significance level, conventionally set at alpha = 0.05. The test is structured around the following formal hypotheses:

Null Hypothesis (H0): The model exhibits homoscedasticity. This means the variance of the residuals is constant across all observation groups.
Alternative Hypothesis (HA): Heteroscedasticity is present. This implies the variance of the residuals is not constant, and the 'increasing' output suggests the variance increases as the predictor variable values rise.

The decision rule is straightforward: if the p-value is less than alpha (0.05), we reject H0 and conclude that significant heteroscedasticity exists. Conversely, if the p-value is greater than alpha, we fail to reject H0, suggesting the assumption of constant variance is reasonable for the data.

For our student score model, the test statistic is **1.757**, and the corresponding p-value is **0.383**. Since 0.383 is substantially greater than the conventional significance level of 0.05, we must fail to reject the null hypothesis. The conclusion is that we lack sufficient statistical evidence to claim that heteroscedasticity is a significant issue in our multiple regression model. We can proceed with interpreting the original OLS regression output with confidence in the reliability of its standard errors.

Addressing Heteroscedasticity: Remedial Measures

While our example data supported the assumption of homoscedasticity, it is critical for any data practitioner to know how to proceed if the Goldfeld-Quandt test leads to the rejection of the null hypothesis. A significant finding of heteroscedasticity mandates remediation, as ignoring it invalidates the statistical inference drawn from the standard errors and p-values. Fortunately, several robust statistical methods exist to correct this issue without needing to discard the model entirely.

One primary method involves applying a transformation to the response variable (Y). Common transformations often used to stabilize variance include the logarithm (e.g., modeling log(Y)) or the square root. The goal of this transformation is to compress the spread of larger values in the response variable, thereby equalizing the variance of the residuals across the entire range of the predictors. Selecting the appropriate transformation often requires examining the pattern of heteroscedasticity observed in the residual plots.

A second, highly effective technique is utilizing **Weighted Least Squares (WLS) Regression**.

Unlike OLS, which treats all observations equally, WLS explicitly incorporates the unequal variance structure into the estimation process. It assigns a specific weight to each data point, typically defined as the inverse of the estimated variance of its error term. By giving less emphasis (lower weight) to observations associated with high variance and more emphasis (higher weight) to observations with low variance, WLS produces efficient coefficient estimates that are not susceptible to the distortion caused by heteroscedasticity.

The third widely accepted solution is the use of **Robust Standard Errors** (also known as Heteroscedasticity-Consistent Standard Errors). This approach offers a simple yet powerful fix: it corrects the standard errors without changing the original coefficient estimates derived from the OLS model. Robust standard errors mathematically adjust for the unequal scatter of the residuals, allowing for valid hypothesis testing and statistical inference even when the assumption of homoscedasticity is severely violated. This method is often preferred when the exact form of the heteroscedasticity is complex or unknown.

Further Exploration and Resources

Proficiency in regression diagnostics is an indispensable skill for rigorous quantitative analysis. The Goldfeld-Quandt test is a cornerstone tool, but it is part of a larger suite of diagnostic tests necessary to validate a model. Beyond testing for heteroscedasticity, analysts must also routinely check for issues such as autocorrelation (dependence between consecutive errors) and multicollinearity (high correlation among predictors).

To deepen your expertise in statistical testing using Python, you should explore alternative diagnostic methods. For instance, the White's Test is another powerful, general test for heteroscedasticity that does not require specifying the exact form of the non-constant variance, making it particularly versatile. For those interested in executing this complementary method, the following resource details how to perform White's Test in Python:

How to Perform White's Test in Python