Introduction to Heteroskedasticity

Non-Constant Variance and What to Do About It

Jake Anderson

May 16, 2026

Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

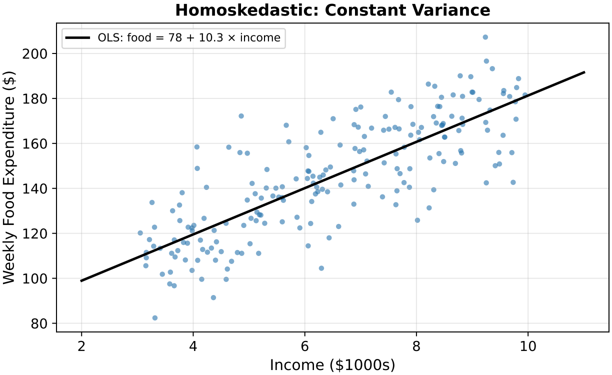
Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

Food Expenditure vs. Income: The Simple Case

Suppose we model weekly food expenditure as a function of household income:

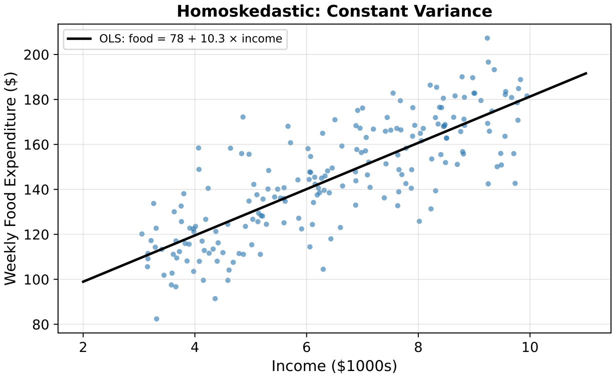
$$\text{food}_i = \beta_0 + \beta_1 \, \text{income}_i + e_i$$



Food Expenditure vs. Income: The Simple Case

Suppose we model weekly food expenditure as a function of household income:

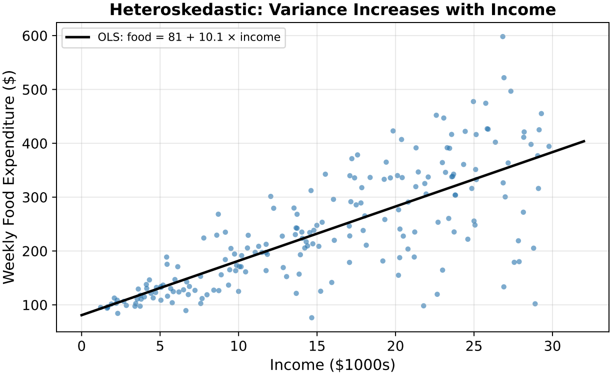
$$\text{food}_i = \beta_0 + \beta_1 \, \text{income}_i + e_i$$



The spread of points around the line looks roughly **constant**. This is **homoskedasticity**: $\text{Var}(e_i \mid x_i) = \sigma^2$ for all i.

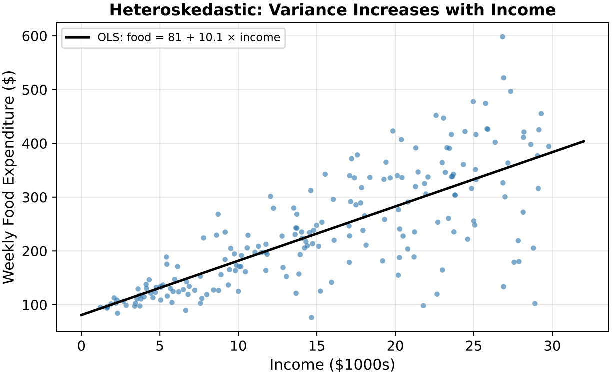
What If the Spread Changes?

Now consider the full income range:



What If the Spread Changes?

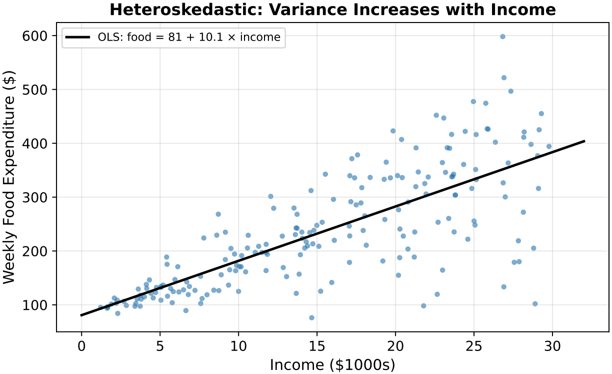
Now consider the full income range:



Higher-income households have **more variable** food spending. Low-income households are clustered tightly around the line.

What If the Spread Changes?

Now consider the full income range:



Higher-income households have **more variable** food spending. Low-income households are clustered tightly around the line.

$\Longrightarrow$ $\text{Var}(e_i \mid x_i)$ is **not constant**. This is **heteroskedasticity**.

Why Does Variance Change?

Heteroskedasticity is common whenever:

- **Discretion in spending:** Higher income $\implies$ more latitude to splurge, save, or invest
- **Returns to schooling:** At low education, wages cluster; at high education, they fan out by field and ability
- **Firm size:** Profits of large firms swing by millions; small firms by thousands
- **Experience and earnings:** Early-career workers cluster near an entry wage; paths diverge with experience

Why Does Variance Change?

Heteroskedasticity is common whenever:

- **Discretion in spending:** Higher income $\Longrightarrow$ more latitude to splurge, save, or invest
- **Returns to schooling:** At low education, wages cluster; at high education, they fan out by field and ability
- **Firm size:** Profits of large firms swing by millions; small firms by thousands
- **Experience and earnings:** Early-career workers cluster near an entry wage; paths diverge with experience

$\Longrightarrow$ Formally, $\text{Var}(e_i \mid x_i) = \sigma_i^2$ depends on x_i. The case $\sigma_i^2 = \sigma^2$ for all i is **homoskedasticity**.

Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

OLS Under Heteroskedasticity: The Good News

Even with heteroskedasticity, OLS is still:

- **Unbiased:** $\mathrm{E}[\hat{\beta}] = \beta$ (requires only $\mathrm{E}[e_i \mid x_i] = 0$)
- **Consistent:** $\hat{\beta} \xrightarrow{p} \beta$ as $n \to \infty$

OLS Under Heteroskedasticity: The Good News

Even with heteroskedasticity, OLS is still:

- **Unbiased:** $E[\hat{\beta}] = \beta$ (requires only $E[e_i \mid x_i] = 0$)
- **Consistent:** $\hat{\beta} \xrightarrow{p} \beta$ as $n \to \infty$

Why? The OLS formula:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_i (x_i - \bar{x}) e_i}{\sum_i (x_i - \bar{x})^2}$$

depends on $E[e_i \mid x_i] = 0$, not on $\text{Var}(e_i \mid x_i) = \sigma^2$.

OLS Under Heteroskedasticity: The Good News

Even with heteroskedasticity, OLS is still:

- **Unbiased:** $\text{E}[\hat{\beta}] = \beta$ (requires only $\text{E}[e_i \mid x_i] = 0$)
- **Consistent:** $\hat{\beta} \xrightarrow{p} \beta$ as $n \to \infty$

Why? The OLS formula:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_i (x_i - \bar{x}) e_i}{\sum_i (x_i - \bar{x})^2}$$

depends on $\text{E}[e_i \mid x_i] = 0$, not on $\text{Var}(e_i \mid x_i) = \sigma^2$.

$\Longrightarrow$ The coefficient estimates themselves are fine. The problem is elsewhere.

OLS Under Heteroskedasticity: The Bad News

The usual OLS standard error formula **assumes homoskedasticity**:

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_i (x_i - \bar{x})^2} \quad \text{where } \hat{\sigma}^2 = \frac{\sum_i \hat{e}_i^2}{n-2}$$

OLS Under Heteroskedasticity: The Bad News

The usual OLS standard error formula **assumes homoskedasticity**:

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_i (x_i - \bar{x})^2} \quad \text{where } \hat{\sigma}^2 = \frac{\sum_i \hat{e}_i^2}{n-2}$$

This averages all squared residuals into a **single** $\hat{\sigma}^2$. But if variance differs across observations, this average is wrong:

- Observations with **large** variance get too little weight in the SE
- Observations with **small** variance get too much weight

OLS Under Heteroskedasticity: The Bad News

The usual OLS standard error formula **assumes homoskedasticity**:

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_i (x_i - \bar{x})^2} \quad \text{where } \hat{\sigma}^2 = \frac{\sum_i \hat{e}_i^2}{n-2}$$

This averages all squared residuals into a **single** $\hat{\sigma}^2$. But if variance differs across observations, this average is wrong:

- Observations with **large** variance get too little weight in the SE
- Observations with **small** variance get too much weight

Consequences:

1. Standard errors are **inconsistent** (could be too small or too large; higher n doesn't help)
2. t-statistics and p-values are **unreliable**
3. Confidence intervals have **wrong coverage**
4. OLS is still linear and unbiased, but is no longer the "best" linear unbiased estimator $\Longrightarrow$ **no longer BLUE**.

OLS Under Heteroskedasticity: The Bad News

The usual OLS standard error formula **assumes homoskedasticity**:

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_i (x_i - \bar{x})^2} \quad \text{where } \hat{\sigma}^2 = \frac{\sum_i \hat{e}_i^2}{n-2}$$

This averages all squared residuals into a **single** $\hat{\sigma}^2$. But if variance differs across observations, this average is wrong:

- Observations with **large** variance get too little weight in the SE
- Observations with **small** variance get too much weight

Consequences:

1. Standard errors are **inconsistent** (could be too small or too large; higher n doesn't help)
2. t-statistics and p-values are **unreliable**
3. Confidence intervals have **wrong coverage**
4. OLS is still linear and unbiased, but is no longer the "best" linear unbiased estimator $\implies$ **no longer BLUE**.

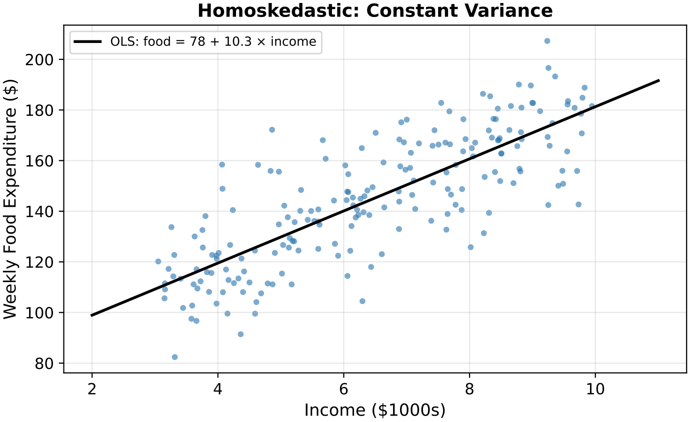
$\implies$ You cannot trust hypothesis tests from OLS when heteroskedasticity is present.

Outline

1 Motivation
2 What Goes Wrong
3 Visual Detection: Residual Plots
4 Breusch-Pagan Test
5 White Test
6 Goldfeld-Quandt Test
7 Fixing It: Robust Standard Errors
8 Fixing It: WLS
9 Fixing It: FGLS
10 End-to-End: Food Expenditure
11 Summary

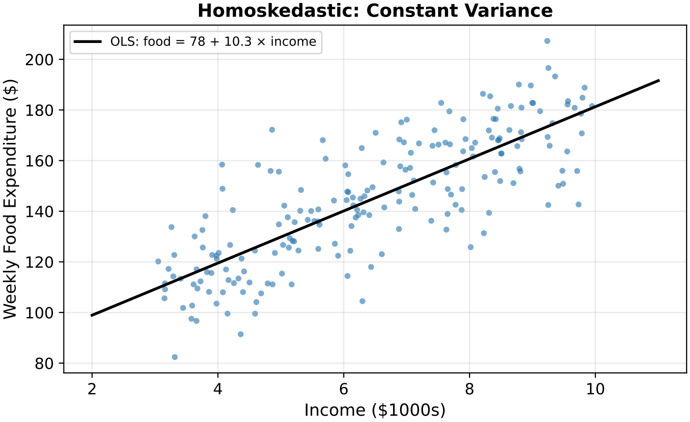
Visual Detection: Start with the Data

Consider this food expenditure regression with constant variance:



Visual Detection: Start with the Data

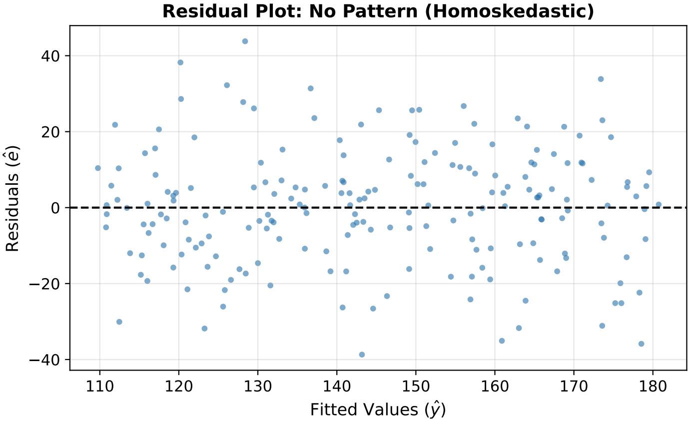
Consider this food expenditure regression with constant variance:



How do we confirm that the spread is truly constant? Plot the **residuals**.

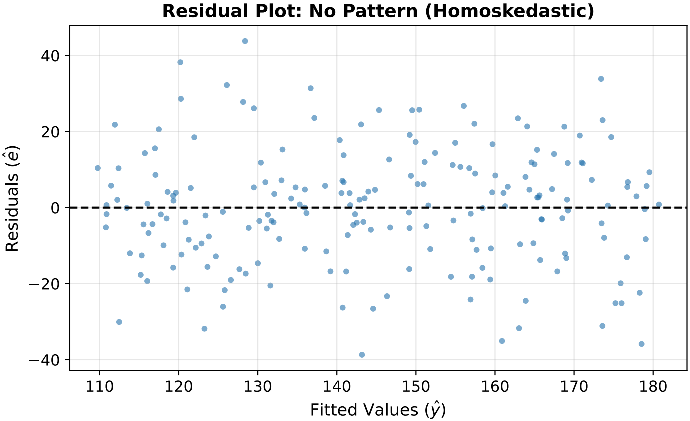
Residual Plots: Random Scatter

Plot residuals $\hat{e}_i$ against fitted values $\hat{y}_i$ (or against x_i):



Residual Plots: Random Scatter

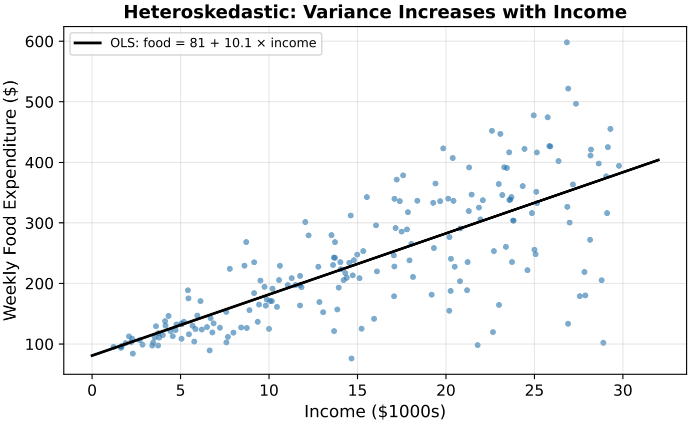
Plot residuals $\hat{e}_i$ against fitted values $\hat{y}_i$ (or against x_i):



Random scatter with no pattern $\implies$ no evidence of heteroskedasticity.

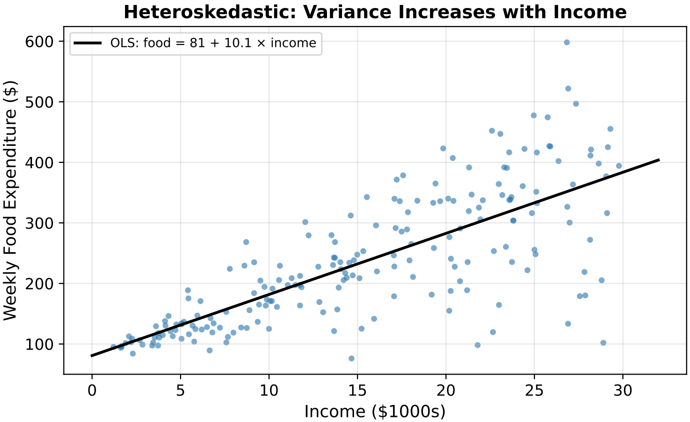
Now Consider Heteroskedastic Data

The same model, but with variance that increases with income:



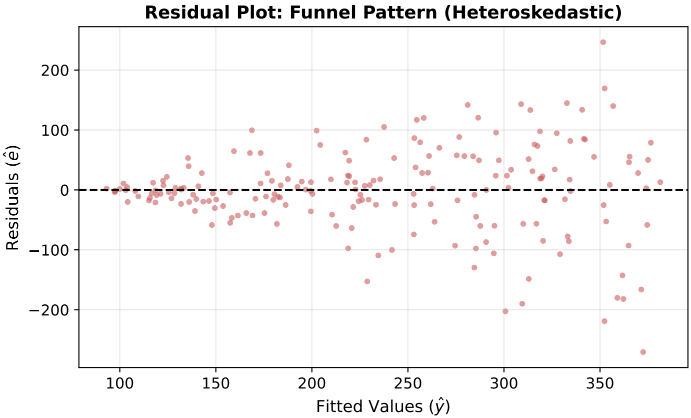
Now Consider Heteroskedastic Data

The same model, but with variance that increases with income:

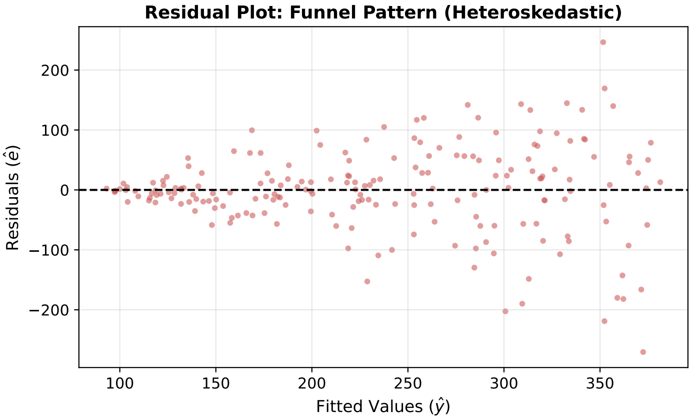


What does the residual plot look like here?

Residual Plot: The Funnel



Residual Plot: The Funnel



Residuals fan out as $\hat{y}$ increases $\implies$ strong evidence of heteroskedasticity. This "funnel" or "megaphone" pattern is the classic warning sign.

Formal Tests After the Residual Plot

The funnel plot is suggestive, but "looks heteroskedastic" is not a number we can put in a paper or defend in front of a referee.

Formal Tests After the Residual Plot

The funnel plot is suggestive, but "looks heteroskedastic" is not a number we can put in a paper or defend in front of a referee.

We need a formal test. Three are commonly taught:

- **Breusch-Pagan:** regress squared residuals on x (linear variance model)
- **White:** BP plus squares and cross-products (flexible variance model)
- **Goldfeld-Quandt:** compare residual variance in low- vs. high-x subsamples

Formal Tests After the Residual Plot

The funnel plot is suggestive, but "looks heteroskedastic" is not a number we can put in a paper or defend in front of a referee.

We need a formal test. Three are commonly taught:

- **Breusch-Pagan:** regress squared residuals on x (linear variance model)
- **White:** BP plus squares and cross-products (flexible variance model)
- **Goldfeld-Quandt:** compare residual variance in low- vs. high-x subsamples

$\Longrightarrow$ Each test makes a different assumption about how variance might depend on x. We work through them in turn.

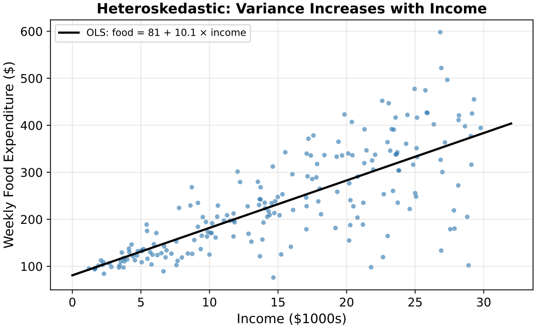
Outline

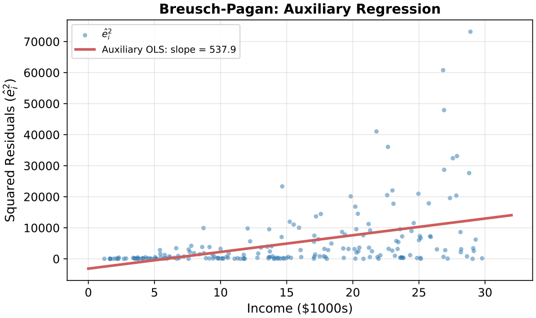
1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. **Breusch-Pagan Test**
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

BP Test: The Idea

Premise: If heteroskedasticity is present, squared residuals *will* be related to x.

The BP test assumes the relationship is **linear**: variance either grows or shrinks as x increases or decreases. Plot the original data on the left and squared residuals against x on the right; see whether a line through the right-hand cloud has nonzero slope.

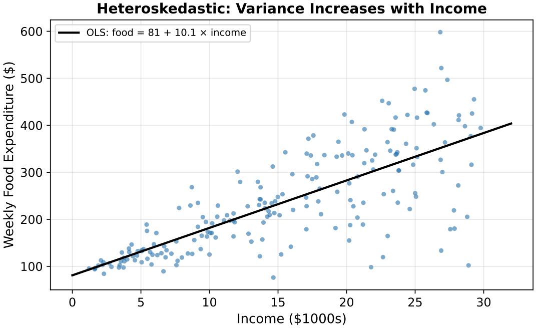


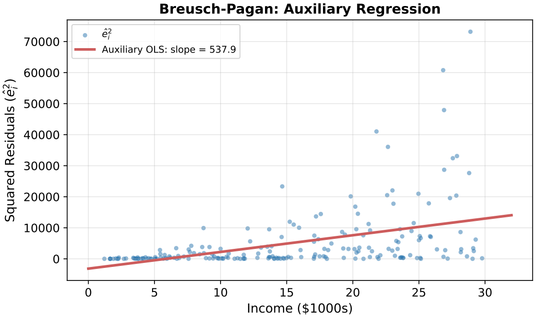


BP Test: The Idea

Premise: If heteroskedasticity is present, squared residuals *will* be related to x.

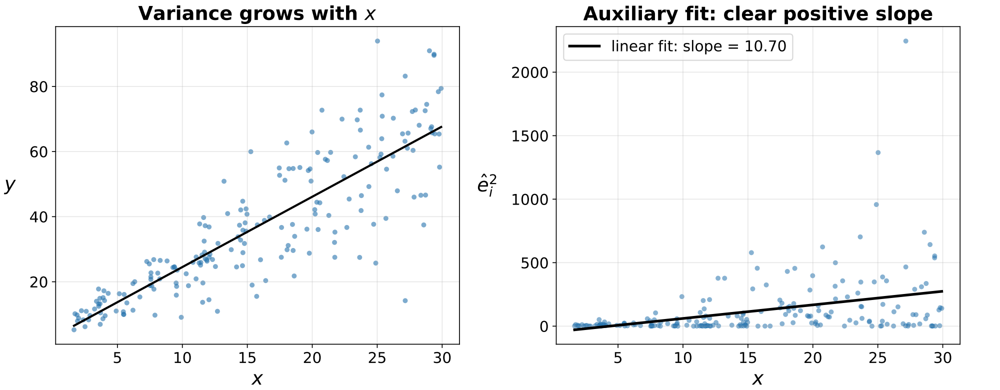
The BP test assumes the relationship is **linear**: variance either grows or shrinks as x increases or decreases. Plot the original data on the left and squared residuals against x on the right; see whether a line through the right-hand cloud has nonzero slope.



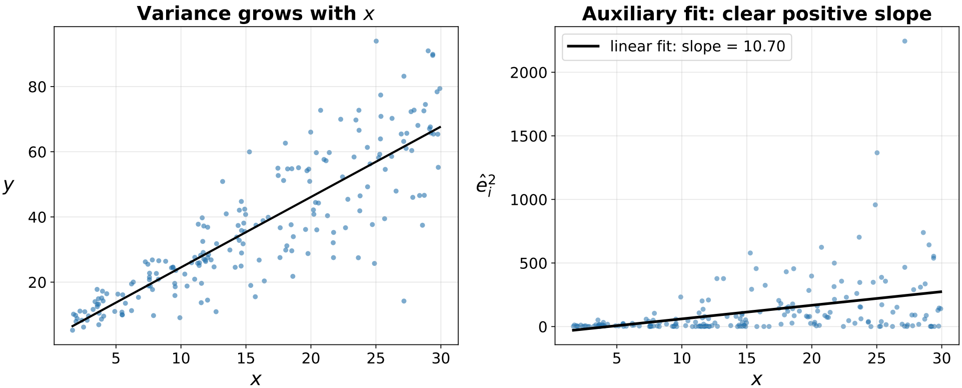


If the line on the right slopes up or down strongly $\Longrightarrow$ reject homoskedasticity.

When the BP Test Works: Variance Grows with x



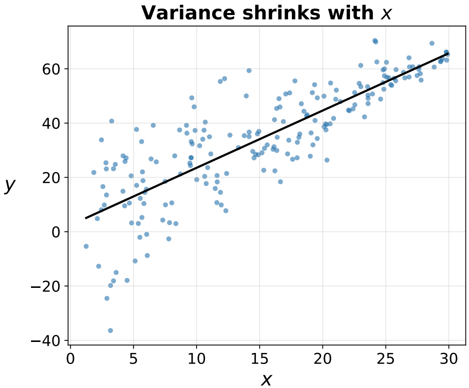
When the BP Test Works: Variance Grows with x

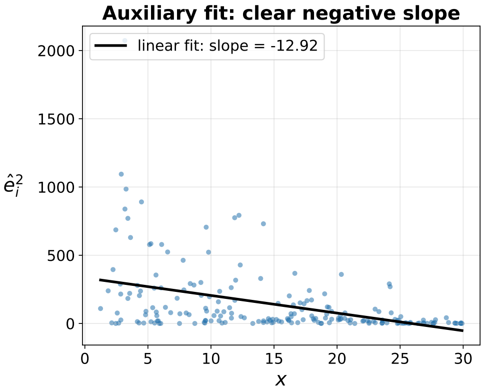


The funnel widens to the right. The linear auxiliary fit picks up a strong positive slope.

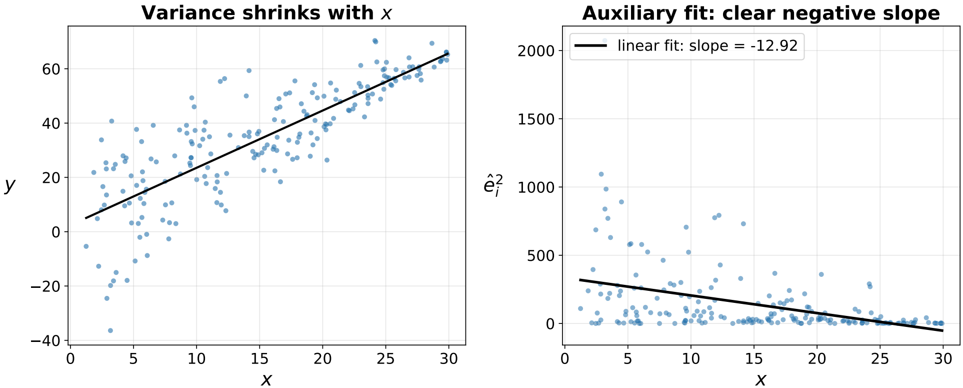
BP Test successfully detects heteroskedasticity!

When the BP Test Works: Variance Shrinks with x





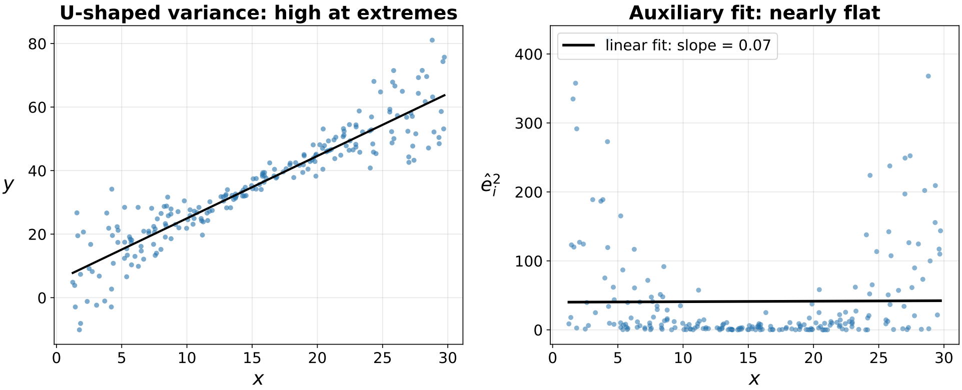
When the BP Test Works: Variance Shrinks with x



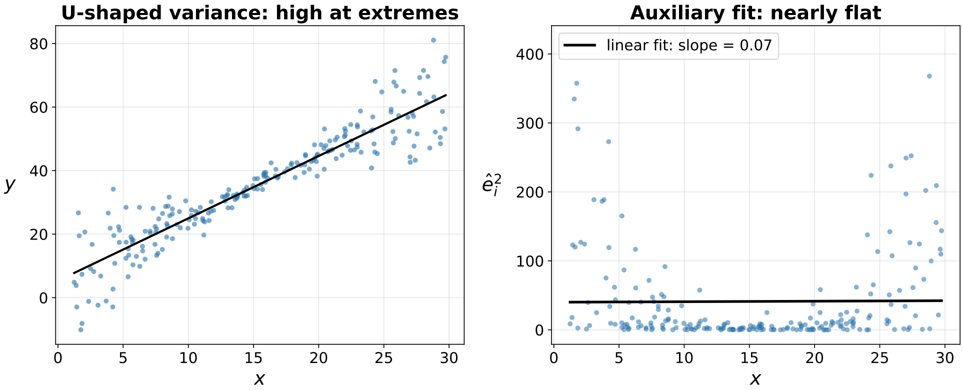
The funnel narrows to the right. The linear auxiliary fit picks up a strong negative slope.

BP Test successfully detects heteroskedasticity!

When the BP Test Fails: U-Shaped Variance



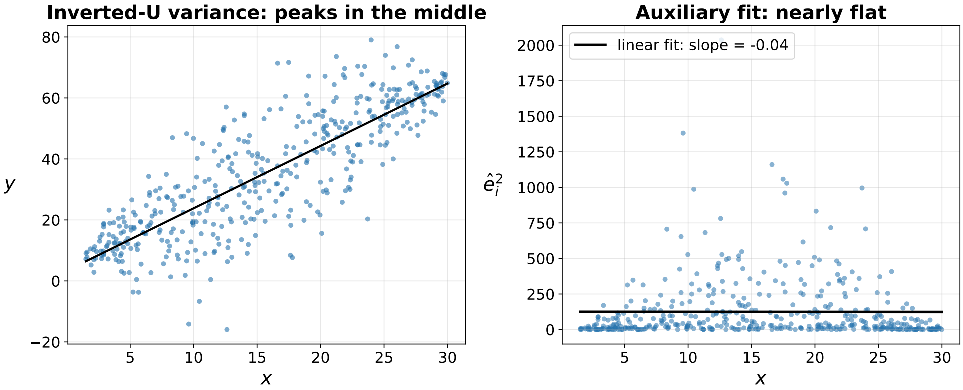
When the BP Test Fails: U-Shaped Variance



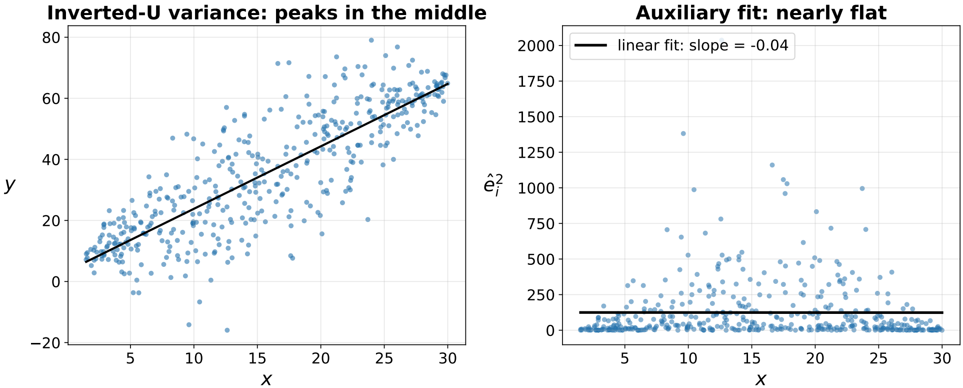
Variance is high at both extremes of x and low in the middle. A line drawn through this cloud is nearly flat $\implies$ the BP slope estimate is close to zero.

BP Test fails to detect this kind of heteroskedasticity

When the BP Test Fails: Inverted-U Variance



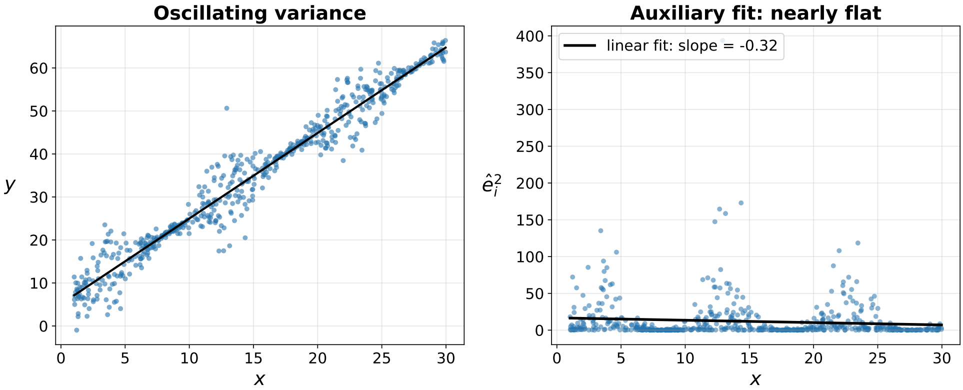
When the BP Test Fails: Inverted-U Variance



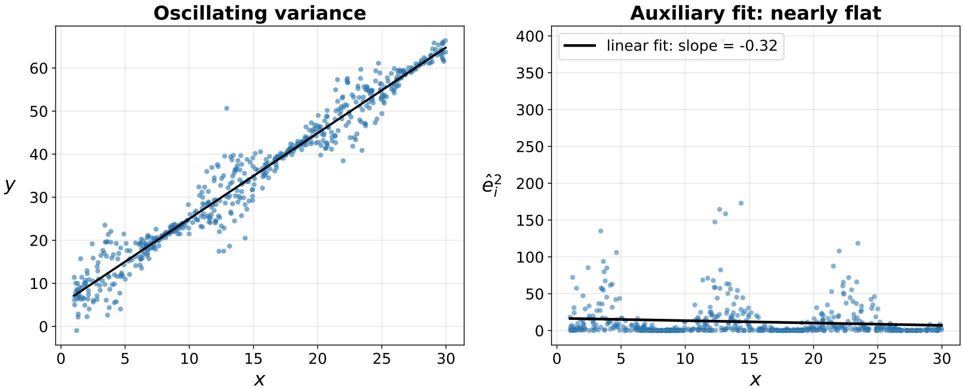
Variance peaks in the middle and falls off at both ends. Same pathology as the U-shape: the linear fit averages the rise and fall to roughly zero.

BP Test fails to detect this kind of heteroskedasticity

When the BP Test Fails: Oscillating Variance



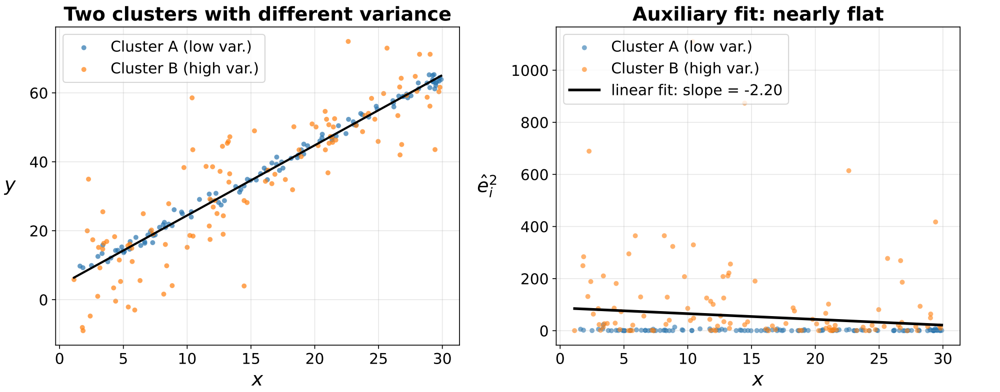
When the BP Test Fails: Oscillating Variance



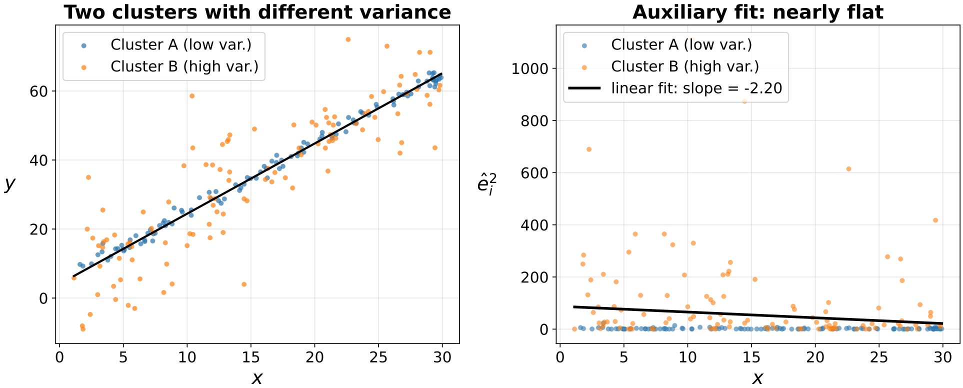
Variance moves up and down across x. Any monotone summary collapses the wave to a flat trend.

BP Test fails to detect this kind of heteroskedasticity

When the BP Test Fails: Clustered Variance



When the BP Test Fails: Clustered Variance



Two groups with different variance levels but the same range of x. Variance depends on *group*, not on x, so a regression of $\hat{e}_i^2$ on x alone reveals nothing.

BP Test fails to detect this kind of heteroskedasticity

Cluster-robust standard errors are designed for this case (covered separately).

BP Test: Procedure

Auxiliary regression: a second OLS regression run on the *output* of the first. The dependent variable is $\hat{e}_i^2$; the regressors are the variables we suspect drive the variance.

BP Test: Procedure

Auxiliary regression: a second OLS regression run on the *output* of the first. The dependent variable is $\hat{e}_i^2$; the regressors are the variables we suspect drive the variance.

Steps:

1. Run OLS: $y_i = \beta_0 + \beta_1 x_i + e_i$. Obtain residuals $\hat{e}_i$.

BP Test: Procedure

Auxiliary regression: a second OLS regression run on the *output* of the first. The dependent variable is $\hat{e}_i^2$; the regressors are the variables we suspect drive the variance.

Steps:

1. Run OLS: $y_i = \beta_0 + \beta_1 x_i + e_i$. Obtain residuals $\hat{e}_i$.
2. Run the auxiliary regression: $\hat{e}_i^2 = \gamma_0 + \gamma_1 x_i + v_i$. If x has no relation to the squared residual, $\gamma_1 = 0$.

BP Test: Procedure

Auxiliary regression: a second OLS regression run on the *output* of the first. The dependent variable is $\hat{e}_i^2$; the regressors are the variables we suspect drive the variance.

Steps:

1. Run OLS: $y_i = \beta_0 + \beta_1 x_i + e_i$. Obtain residuals $\hat{e}_i$.
2. Run the auxiliary regression: $\hat{e}_i^2 = \gamma_0 + \gamma_1 x_i + v_i$. If x has no relation to the squared residual, $\gamma_1 = 0$.
3. Under the assumed (linear) variance model, $\gamma_1 = 0 \iff$ homoskedasticity. Test $H_0 : \gamma_1 = 0$.

BP Test: Procedure

Auxiliary regression: a second OLS regression run on the *output* of the first. The dependent variable is $\hat{e}_i^2$; the regressors are the variables we suspect drive the variance.

Steps:

1. Run OLS: $y_i = \beta_0 + \beta_1 x_i + e_i$. Obtain residuals $\hat{e}_i$.
2. Run the auxiliary regression: $\hat{e}_i^2 = \gamma_0 + \gamma_1 x_i + v_i$. If x has no relation to the squared residual, $\gamma_1 = 0$.
3. Under the assumed (linear) variance model, $\gamma_1 = 0 \iff$ homoskedasticity. Test $H_0 : \gamma_1 = 0$.
4. Test statistic: $\text{BP} = n \cdot R^2_{\text{aux}} \sim \chi^2(1)$ under H_0. The R^2 from the auxiliary regression measures how much of the variation in $\hat{e}_i^2$ is explained by x. Bigger $R^2 \implies$ stronger evidence of heteroskedasticity.

BP Test: Procedure

Auxiliary regression: a second OLS regression run on the *output* of the first. The dependent variable is $\hat{e}_i^2$; the regressors are the variables we suspect drive the variance.

Steps:

1. Run OLS: $y_i = \beta_0 + \beta_1 x_i + e_i$. Obtain residuals $\hat{e}_i$.
2. Run the auxiliary regression: $\hat{e}_i^2 = \gamma_0 + \gamma_1 x_i + v_i$. If x has no relation to the squared residual, $\gamma_1 = 0$.
3. Under the assumed (linear) variance model, $\gamma_1 = 0 \iff$ homoskedasticity. Test $H_0 : \gamma_1 = 0$.
4. Test statistic: $\text{BP} = n \cdot R^2_{\text{aux}} \sim \chi^2(1)$ under H_0. The R^2 from the auxiliary regression measures how much of the variation in $\hat{e}_i^2$ is explained by x. Bigger $R^2 \implies$ stronger evidence of heteroskedasticity.

With multiple regressors, include all x's in the auxiliary regression: $\text{BP} \sim \chi^2(k)$, where k is the number of *slope* regressors. **The intercept is never counted in k.**

Note: this is the studentized (Koenker) form of the BP test, the version reported by R's `bptest` default and taught in HGL.

Example: Wages and Experience

We have data on 1,000 workers, with observed experience (years) and hourly wage ($/hour).

$$\text{wage}_i = \beta_0 + \beta_1 \, \text{experience}_i + e_i$$

Example: Wages and Experience

We have data on 1,000 workers, with observed experience (years) and hourly wage ($/hour).

$$\text{wage}_i = \beta_0 + \beta_1\,\text{experience}_i + e_i$$

The fitted line slopes up: more experience, higher wages on average. But notice the spread of points around the line widens with experience.

Step 1: Run OLS

Coefficients

	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
(Intercept)	8.8372	0.5234	16.88	$< 2 \cdot 10^{-16}$ ***
experience	1.0840	0.0303	35.81	$< 2 \cdot 10^{-16}$ ***

Residual standard error: 8.37 on 998 degrees of freedom
Multiple R^2*:* 0.5623 *Adjusted* R^2*:* 0.5619
F*-statistic:* 1282 on 1 and 998 DF, $p < 2 \cdot 10^{-16}$

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Step 1: Run OLS

Coefficients

	Estimate	Std. Error	t-value	$\Pr(>\lvert t\rvert)$
(Intercept)	8.8372	0.5234	16.88	$< 2 \cdot 10^{-16}$ ***
experience	1.0840	0.0303	35.81	$< 2 \cdot 10^{-16}$ ***

Residual standard error: 8.37 on 998 degrees of freedom
Multiple R^2: 0.5623 *Adjusted R^2:* 0.5619
F-statistic: 1282 on 1 and 998 DF, $p < 2 \cdot 10^{-16}$

Signif. codes: `0 *** 0.001 ** 0.01 * 0.05 . 0.1 1`

Slope is extremely significant under OLS *assumptions*. **But are those SEs valid given the data?** BP will tell us.

Step 2: State the Hypothesis

The BP test asks whether $\hat{e}_i^2$ is linearly related to experience.

Step 2: State the Hypothesis

The BP test asks whether $\hat{e}_i^2$ is linearly related to experience.

Null and alternative:

$$H_0 : \gamma_1 = 0 \quad \text{(homoskedasticity, under the linear variance model)}$$
$$H_1 : \gamma_1 \neq 0 \quad \text{(heteroskedasticity in experience)}$$

Step 2: State the Hypothesis

The BP test asks whether $\hat{e}_i^2$ is linearly related to experience.

Null and alternative:

$$H_0 : \gamma_1 = 0 \quad \text{(homoskedasticity, under the linear variance model)}$$
$$H_1 : \gamma_1 \neq 0 \quad \text{(heteroskedasticity in experience)}$$

Test statistic:

$$\text{BP} = n \cdot R^2_{\text{aux}} \sim \chi^2(k) \text{ under } H_0$$

where k is the number of *slope* regressors in the auxiliary (the intercept is not counted). Here the auxiliary has one slope regressor (experience), so $k = 1$.

Step 3: Auxiliary Regression

Compute squared residuals $\hat{e}_i^2$ and regress them on experience:

$$\hat{e}_i^2 = \gamma_0 + \gamma_1 \,\text{experience}_i + v_i$$

Estimated: $\hat{e}_i^2 = -18.59 + 5.93\,\text{experience}_i, \quad R^2_{\text{aux}} = 0.158.$

Step 3: Auxiliary Regression

Compute squared residuals $\hat{e}_i^2$ and regress them on experience:

$$\hat{e}_i^2 = \gamma_0 + \gamma_1 \, \text{experience}_i + v_i$$

Estimated: $\hat{e}_i^2 = -18.59 + 5.93 \, \text{experience}_i, \quad R^2_{\text{aux}} = 0.158.$

Squared residuals climb with experience: clear positive slope.

Step 4: Compute the Test Statistic

Plug in:

$$\begin{aligned} \text{BP} &= n \cdot R^2_{\text{aux}} \\ &= 1{,}000 \cdot 0.1584 \\ &= \boxed{158.41} \end{aligned}$$

Step 4: Compute the Test Statistic

Plug in:

$$\begin{aligned}\text{BP} &= n \cdot R^2_{\text{aux}} \\ &= 1{,}000 \cdot 0.1584 \\ &= \boxed{158.41}\end{aligned}$$

Compare to the critical value of $\chi^2(1)$ at $\alpha = 0.05$.

Step 5: Compare to χ^2 Critical Value

df	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
1	2.706	**3.841**	5.024	6.635
2	4.605	5.991	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

Step 5: Compare to χ^2 Critical Value

df	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
1	2.706	**3.841**	5.024	6.635
2	4.605	5.991	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

One *slope* regressor in the auxiliary (experience) $\Longrightarrow$ df $= 1$. The intercept is not counted. At $\alpha = 0.05$, the highlighted cell gives the cutoff:

$$\chi^2_{0.05,\,1} = 3.841$$

Step 5: Compare to χ^2 Critical Value

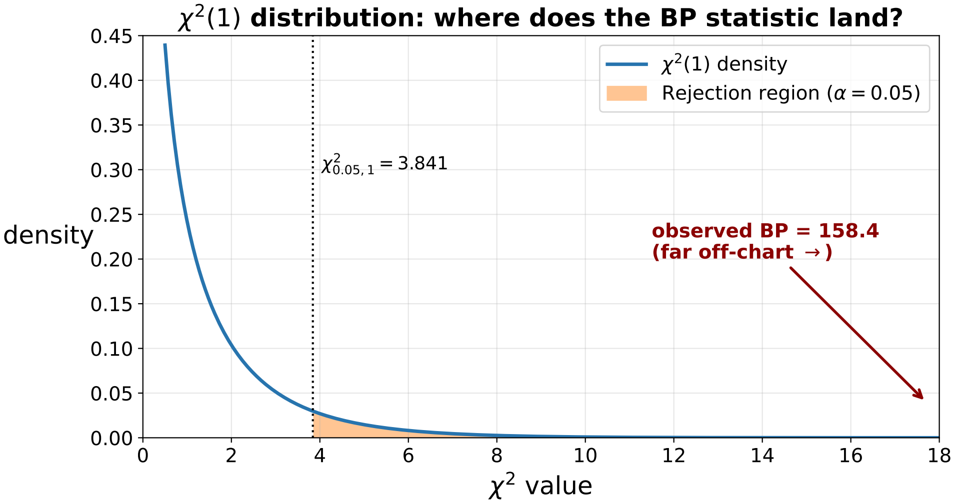
df	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
1	2.706	**3.841**	5.024	6.635
2	4.605	5.991	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

One *slope* regressor in the auxiliary (experience) $\Longrightarrow$ df $= 1$. The intercept is not counted. At $\alpha = 0.05$, the highlighted cell gives the cutoff:

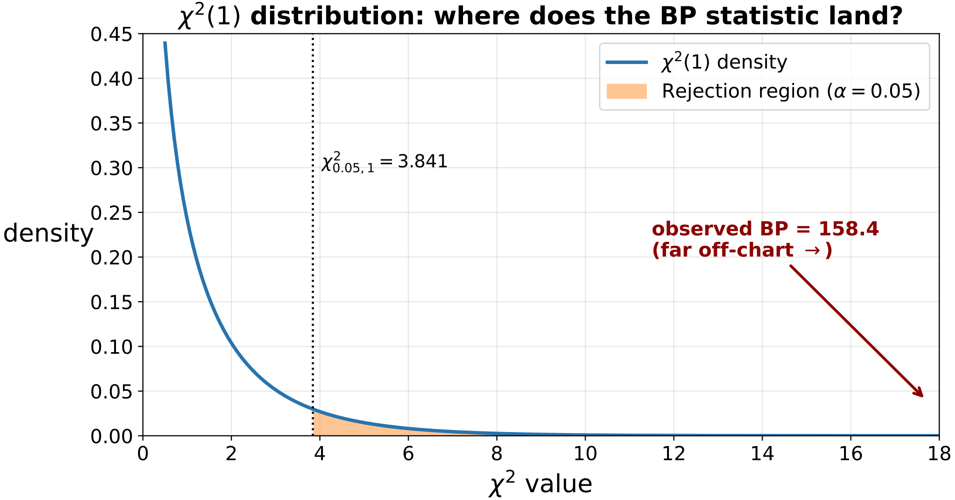
$$\chi^2_{0.05,\,1} = 3.841$$

Reject H_0 if BP > 3.841.

Step 6: Visualize the Decision



Step 6: Visualize the Decision



The orange shaded region is the upper 5% tail of $\chi^2(1)$. Our BP $= 158.4$ is far past the right edge of the chart; the p-value is essentially zero.

Step 7: Conclusion

Decision: $\text{BP} = 158.41 \gg 3.841 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Step 7: Conclusion

Decision: $\text{BP} = 158.41 \gg 3.841 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Plain language: the variance of wage residuals depends strongly on experience. As workers gain experience, the spread of wages around the regression line widens.

Step 7: Conclusion

Decision: $\text{BP} = 158.41 \gg 3.841 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Plain language: the variance of wage residuals depends strongly on experience. As workers gain experience, the spread of wages around the regression line widens.

What this means for our results:

- The OLS coefficients ($\hat{\beta}_0 = 8.84$, $\hat{\beta}_1 = 1.08$) are still unbiased and consistent.
- The OLS standard errors (e.g., 0.030 on experience) are **inconsistent** estimators of the true sampling variance of $\hat{\beta}$. The discrepancy does not shrink with n, so OLS t-statistics, p-values, and confidence intervals do not have their stated size / coverage.
- Use *robust* (HC) standard errors instead. The point estimates do not change; HC SEs are consistent for the true sampling variance, so the corresponding t-statistics have the correct asymptotic standard-normal distribution and 95% CIs cover β approximately 95% of the time.

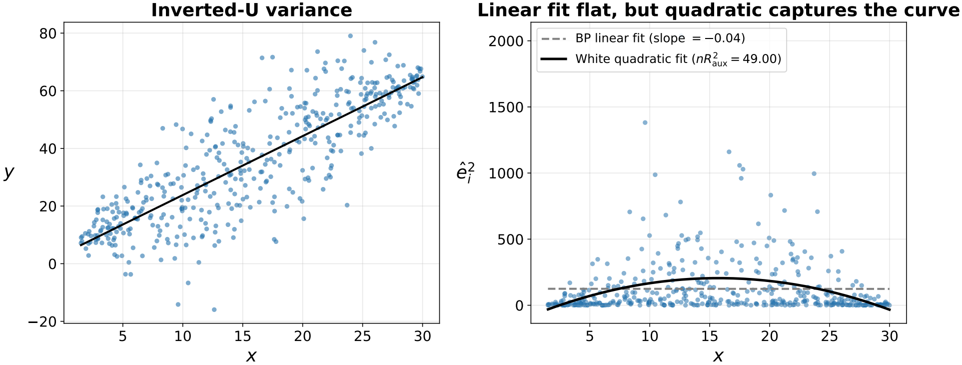
Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

White Test: The Idea

Premise: BP assumes σ_i^2 is a *linear* function of x. The White test relaxes that: σ_i^2 may depend on x in non-linear ways too.

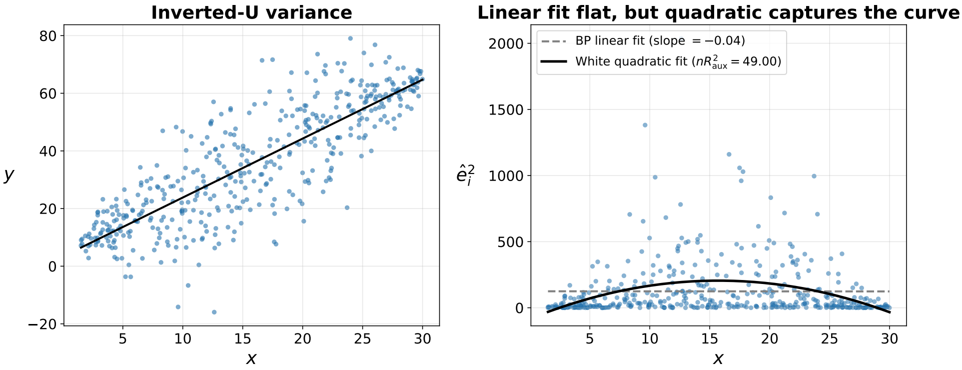
Auxiliary regression: regress $\hat{e}_i^2$ on the original regressors, their *squares*, and their *cross-products*.



White Test: The Idea

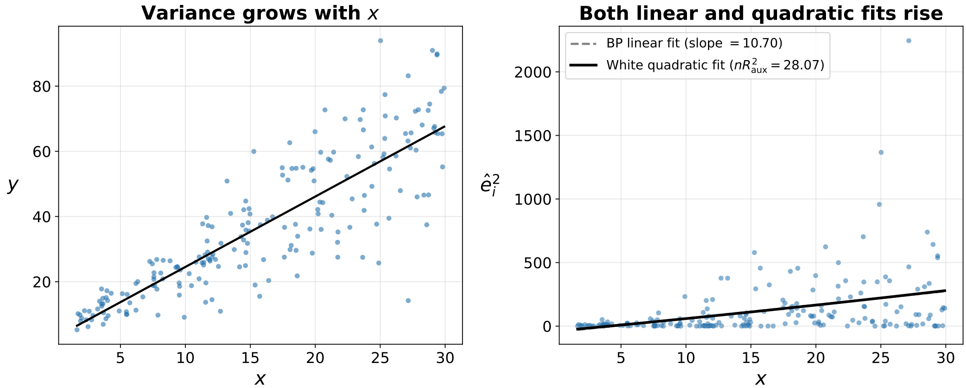
Premise: BP assumes σ_i^2 is a *linear* function of x. The White test relaxes that: σ_i^2 may depend on x in non-linear ways too.

Auxiliary regression: regress $\hat{e}_i^2$ on the original regressors, their *squares*, and their *cross-products*.

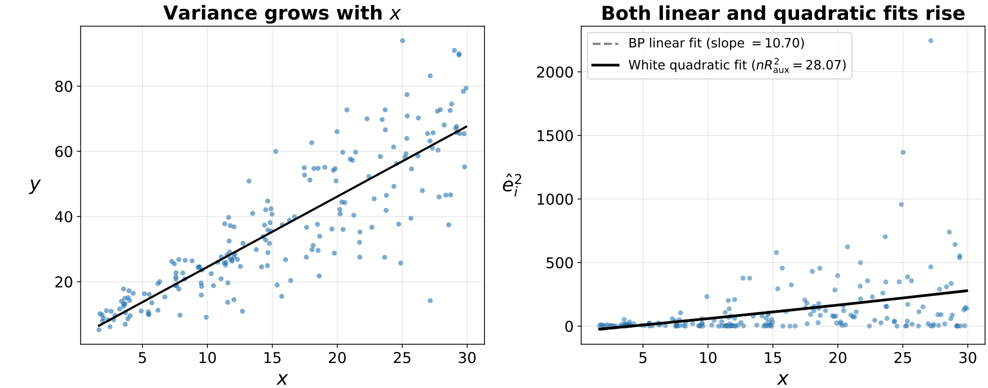


If the quadratic fit explains a large fraction of $\hat{e}_i^2$ $\Longrightarrow$ reject homoskedasticity.

When the White Test Works: Variance Grows with x



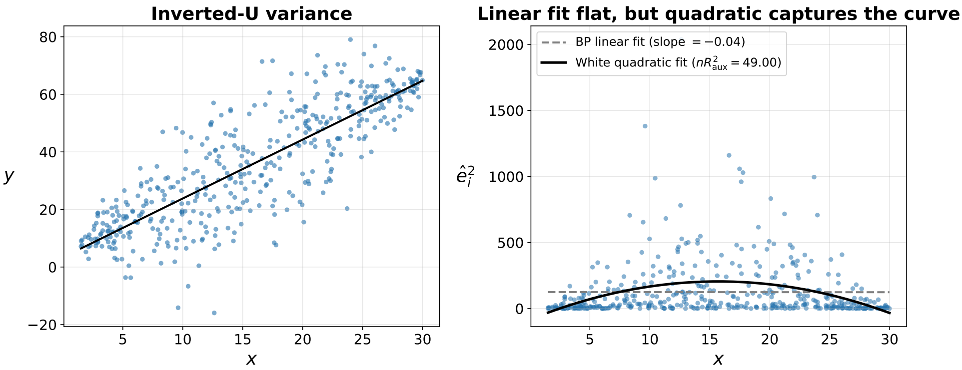
When the White Test Works: Variance Grows with x



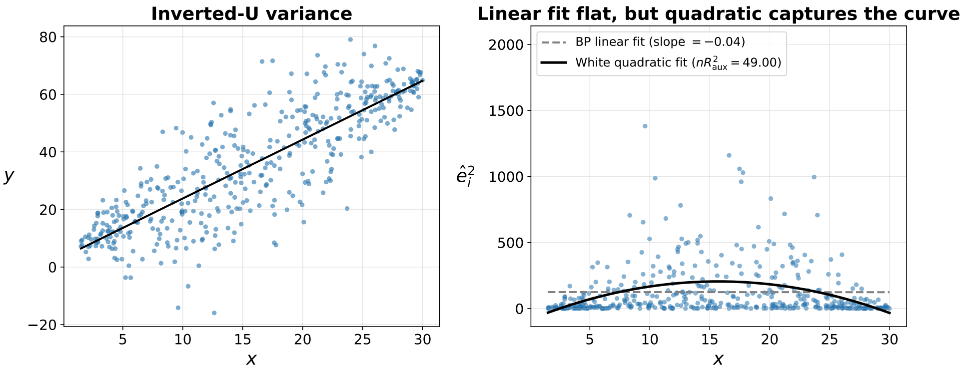
Linear and quadratic fits both rise. Both BP and White reject; White is more general but doesn't lose anything here.

White Test successfully detects heteroskedasticity!

When the White Test Works: Inverted-U Variance



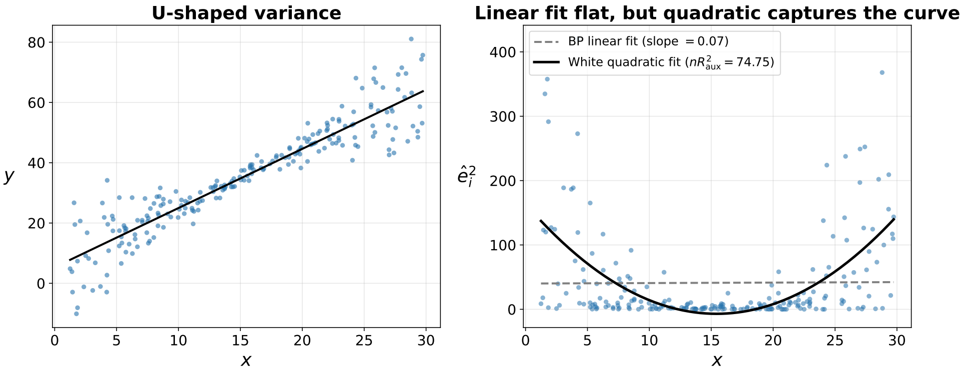
When the White Test Works: Inverted-U Variance



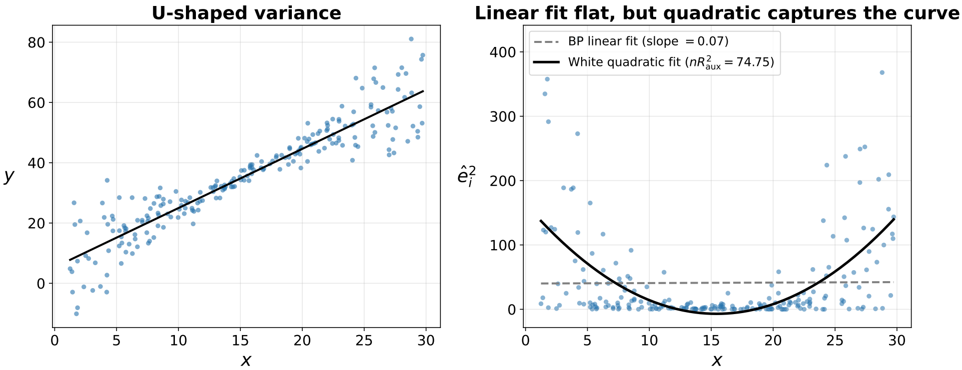
Variance peaks in the middle. The BP linear fit averages to roughly flat (BP fails), but the quadratic captures the curve clearly. **This is where White outperforms BP.**

White Test successfully detects heteroskedasticity!

When the White Test Works: U-Shaped Variance



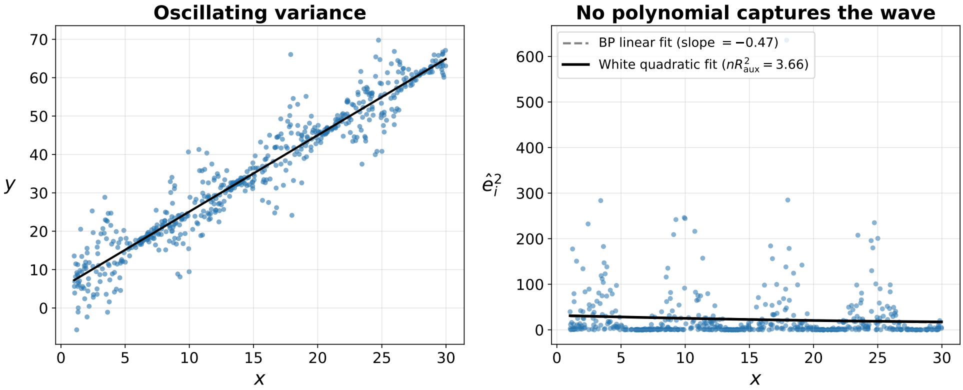
When the White Test Works: U-Shaped Variance



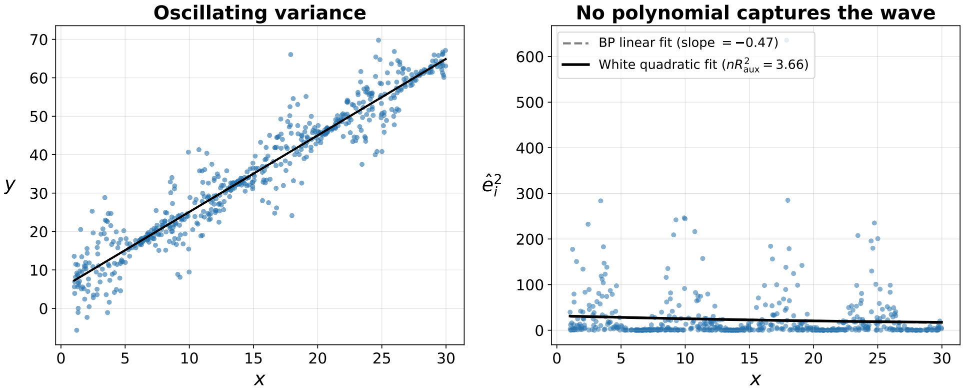
Variance is high at both extremes. Same mechanism as the inverted-U: the linear fit is flat (BP fails), but the quadratic curls upward at both ends.

White Test successfully detects heteroskedasticity!

When the White Test Fails: Oscillating Variance



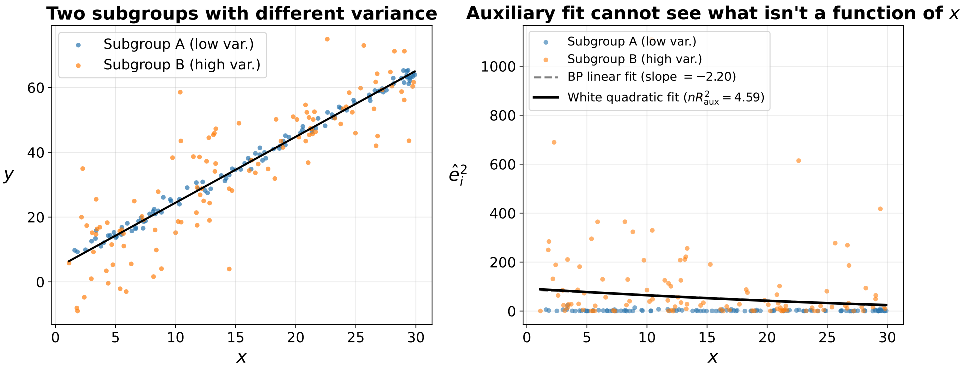
When the White Test Fails: Oscillating Variance



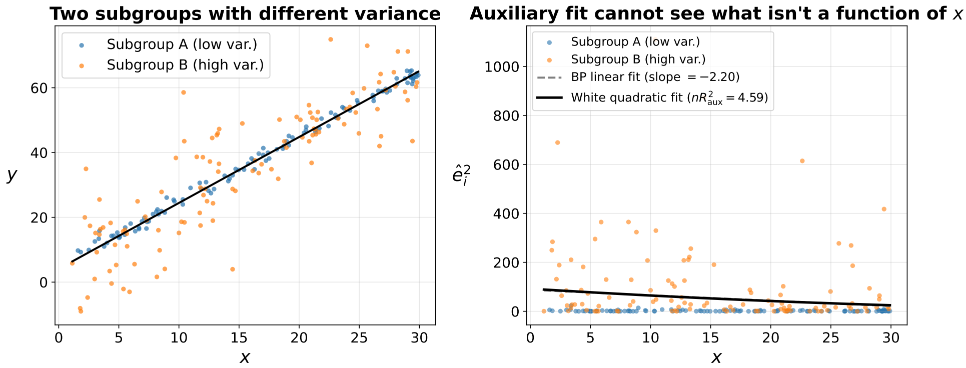
Variance rises and falls multiple times across x. A quadratic in x has only one extremum, so it can't track multiple peaks and troughs; the auxiliary R^2 is small.

White Test fails to detect this kind of heteroskedasticity

When the White Test Fails: Two Subgroups



When the White Test Fails: Two Subgroups



Variance depends on *subgroup*, not on x. No polynomial in x can capture subgroup membership.

White Test fails to detect this kind of heteroskedasticity

Cluster-robust standard errors are designed for this case (covered separately).

White Test: Procedure

Auxiliary regression (single regressor case):

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + v_i$$

For multiple regressors, include all x's, their squares, and pairwise cross-products.

White Test: Procedure

Auxiliary regression (single regressor case):

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + v_i$$

For multiple regressors, include all x's, their squares, and pairwise cross-products.

Steps:

1. Run OLS on the original model. Save residuals $\hat{e}_i$.

White Test: Procedure

Auxiliary regression (single regressor case):

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + v_i$$

For multiple regressors, include all x's, their squares, and pairwise cross-products.

Steps:

1. Run OLS on the original model. Save residuals $\hat{e}_i$.
2. Run the auxiliary regression. Compute R^2_{aux}.

White Test: Procedure

Auxiliary regression (single regressor case):

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + v_i$$

For multiple regressors, include all x's, their squares, and pairwise cross-products.

Steps:

1. Run OLS on the original model. Save residuals $\hat{e}_i$.
2. Run the auxiliary regression. Compute R^2_{aux}.
3. Test statistic: White $= n \cdot R^2_{\mathrm{aux}} \sim \chi^2(q)$ under $H_0 : \alpha_1 = \cdots = \alpha_q = 0$.

White Test: Procedure

Auxiliary regression (single regressor case):

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + v_i$$

For multiple regressors, include all x's, their squares, and pairwise cross-products.

Steps:

1. Run OLS on the original model. Save residuals $\hat{e}_i$.
2. Run the auxiliary regression. Compute R^2_{aux}.
3. Test statistic: White $= n \cdot R^2_{\mathrm{aux}} \sim \chi^2(q)$ under $H_0 : \alpha_1 = \cdots = \alpha_q = 0$.

q is the number of *slope* regressors in the auxiliary regression (linearly independent ones; the intercept is not counted). With one regressor in the original model, $q = 2$ (one for x, one for x^2). With k regressors, $q = k + k + \binom{k}{2} = k(k+3)/2$.

Tradeoff: the flexibility comes at a cost. With many regressors, the auxiliary regression has many terms, and power drops.

Example: Wages and Experience

We have data on 1,000 workers, with observed experience (years) and hourly wage ($/hour).

$$\text{wage}_i = \beta_0 + \beta_1\,\text{experience}_i + e_i$$

Step 1: Run OLS

Coefficients

	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
(Intercept)	8.8372	0.5234	16.88	$< 2 \cdot 10^{-16}$ ***
experience	1.0840	0.0303	35.81	$< 2 \cdot 10^{-16}$ ***

Residual standard error: 8.37 on 998 degrees of freedom
Multiple R^2*:* 0.5623 *Adjusted* R^2*:* 0.5619
F-statistic: 1282 on 1 and 998 DF, $p < 2 \cdot 10^{-16}$

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Step 1: Run OLS

Coefficients

	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
(Intercept)	8.8372	0.5234	16.88	$< 2 \cdot 10^{-16}$ ***
experience	1.0840	0.0303	35.81	$< 2 \cdot 10^{-16}$ ***

Residual standard error: 8.37 on 998 degrees of freedom
Multiple R^2*:* 0.5623 *Adjusted* R^2*:* 0.5619
F-statistic: 1282 on 1 and 998 DF, $p < 2 \cdot 10^{-16}$

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Slope highly significant under OLS *assumptions*. White will test the variance specification.

Step 2: State the Hypothesis

The White test asks whether $\hat{e}_i^2$ is related to experience or experience^2.

Step 2: State the Hypothesis

The White test asks whether $\hat{e}_i^2$ is related to experience or experience^2.

Null and alternative:

$$H_0 : \alpha_1 = \alpha_2 = 0 \quad \text{(homoskedasticity)}$$
$$H_1 : \text{at least one of } \alpha_1, \alpha_2 \neq 0 \quad \text{(heteroskedasticity)}$$

Step 2: State the Hypothesis

The White test asks whether $\hat{e}_i^2$ is related to experience or experience^2.

Null and alternative:

$$H_0 : \alpha_1 = \alpha_2 = 0 \quad \text{(homoskedasticity)}$$
$$H_1 : \text{at least one of } \alpha_1, \alpha_2 \neq 0 \quad \text{(heteroskedasticity)}$$

Test statistic:

$$\text{White} = n \cdot R^2_{\text{aux}} \sim \chi^2(q) \text{ under } H_0$$

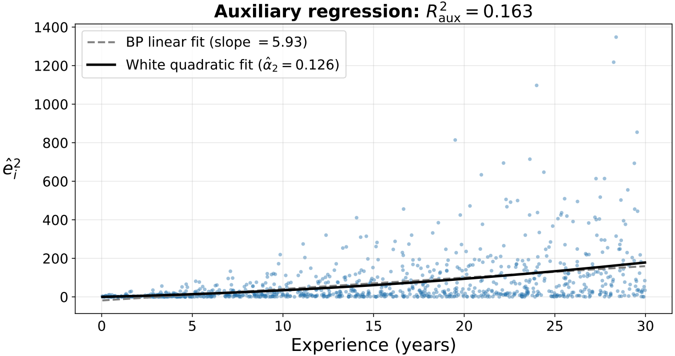
where q is the number of *slope* regressors in the auxiliary (intercept not counted). Here $q = 2$: one for experience, one for experience^2.

Step 3: Auxiliary Regression

Compute squared residuals $\hat{e}_i^2$ and regress on experience and experience^2:

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 \,\text{experience}_i + \alpha_2 \,\text{experience}_i^2 + v_i$$

Estimated: $\hat{\alpha}_0 = -0.11$ (11.15), $\hat{\alpha}_1 = 2.16$ (1.74), $\hat{\alpha}_2 = 0.126$ (0.057), $R^2_{\text{aux}} = 0.163$.

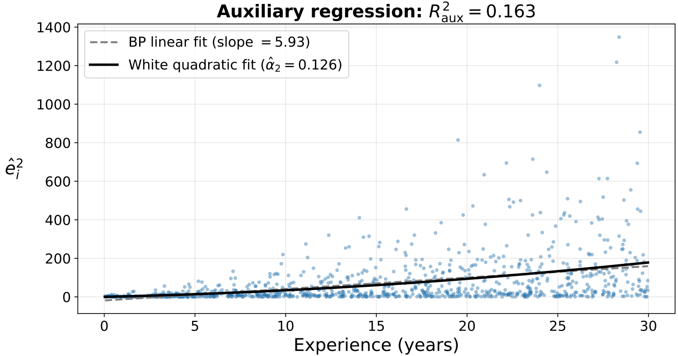


Step 3: Auxiliary Regression

Compute squared residuals $\hat{e}_i^2$ and regress on experience and experience^2:

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 \text{experience}_i + \alpha_2 \text{experience}_i^2 + v_i$$

Estimated: $\hat{\alpha}_0 = -0.11$ (11.15), $\hat{\alpha}_1 = 2.16$ (1.74), $\hat{\alpha}_2 = 0.126$ (0.057), $R^2_{\text{aux}} = 0.163$.



The White picks up both the linear and the curvature components.

Step 4: Compute the Test Statistic

Plug in:

$$\begin{aligned} \text{White} &= n \cdot R^2_{\text{aux}} \\ &= 1{,}000 \cdot 0.1626 \\ &= \boxed{162.59} \end{aligned}$$

Step 4: Compute the Test Statistic

Plug in:

$$\begin{aligned} \text{White} &= n \cdot R^2_{\text{aux}} \\ &= 1{,}000 \cdot 0.1626 \\ &= \boxed{162.59} \end{aligned}$$

Compare to the critical value of $\chi^2(2)$ at $\alpha = 0.05$.

Step 5: Compare to χ^2 Critical Value

df	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
1	2.706	3.841	5.024	6.635
2	4.605	**5.991**	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

Step 5: Compare to χ^2 Critical Value

df	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
1	2.706	3.841	5.024	6.635
2	4.605	**5.991**	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

The auxiliary has 2 *slope* regressors (experience and experience^2) $\implies$ df $= 2$. The intercept is not counted. At $\alpha = 0.05$:

$$\chi^2_{0.05,\,2} = 5.991$$

Step 5: Compare to χ^2 Critical Value

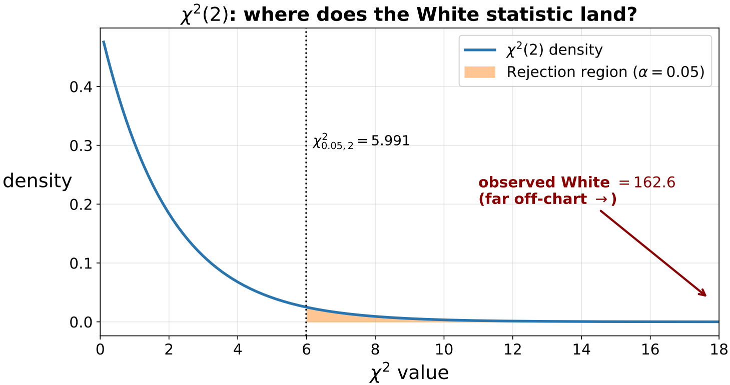
df	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
1	2.706	3.841	5.024	6.635
2	4.605	**5.991**	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

The auxiliary has 2 *slope* regressors (experience and experience^2) $\implies$ df $= 2$. The intercept is not counted. At $\alpha = 0.05$:

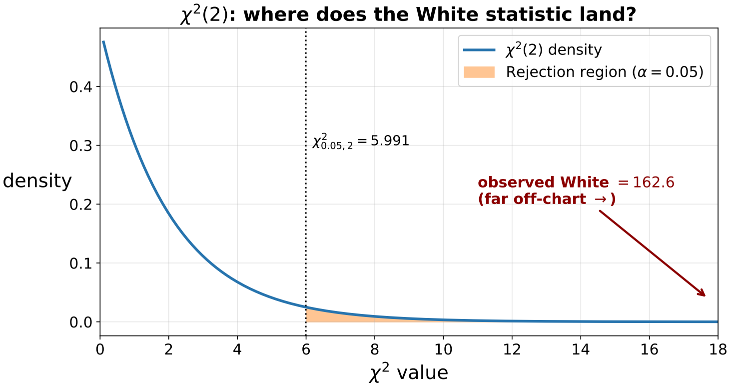
$$\chi^2_{0.05,\,2} = 5.991$$

Reject H_0 if White > 5.991.

Step 6: Visualize the Decision



Step 6: Visualize the Decision



The orange region is the upper 5% tail of $\chi^2(2)$. Our White $= 162.59$ is far past the right edge; the p-value is essentially zero.

Step 7: Conclusion

Decision: White $= 162.59 \gg 5.991 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Step 7: Conclusion

Decision: White $= 162.59 \gg 5.991 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Plain language: the variance of the wage residuals is strongly related to experience, including non-linear (quadratic) components.

Step 7: Conclusion

Decision: White $= 162.59 \gg 5.991 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Plain language: the variance of the wage residuals is strongly related to experience, including non-linear (quadratic) components.

What this means for our results:

- The OLS coefficients ($\hat{\beta}_0 = 8.84$, $\hat{\beta}_1 = 1.08$) are still unbiased and consistent.
- The OLS standard errors are **inconsistent** for the true sampling variance of $\hat{\beta}$; the discrepancy does not shrink with n.
- Use *robust* (HC) standard errors. White is a more general test than BP, but the inference fix is the same: HC SEs are consistent for the true sampling variance, so t-statistics and CIs have correct asymptotic distribution / coverage.

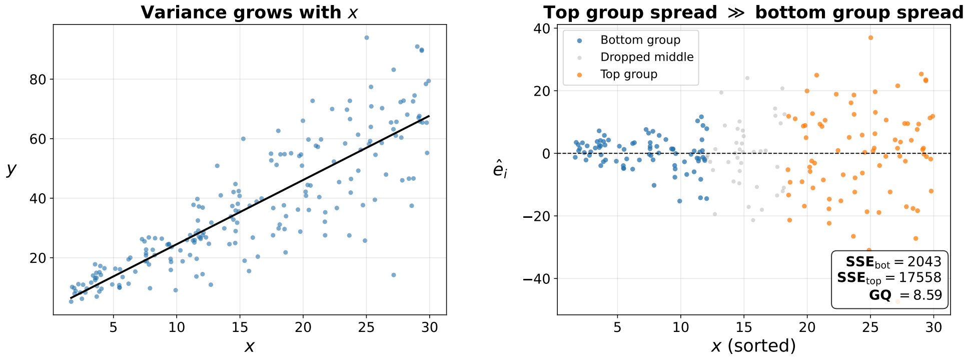
Outline

1 Motivation
2 What Goes Wrong
3 Visual Detection: Residual Plots
4 Breusch-Pagan Test
5 White Test
6 Goldfeld-Quandt Test
7 Fixing It: Robust Standard Errors
8 Fixing It: WLS
9 Fixing It: FGLS
10 End-to-End: Food Expenditure
11 Summary

GQ Test: The Idea

Premise: If variance changes with x, the residual variance in a high-x subsample should differ from the residual variance in a low-x subsample.

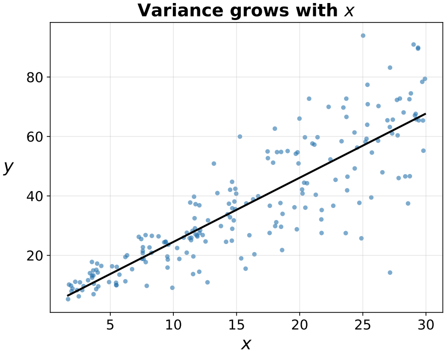
Procedure (sketch): sort the data by the suspect variable, drop the middle observations, run separate OLS on the bottom and top groups, compare the residual sums of squares.

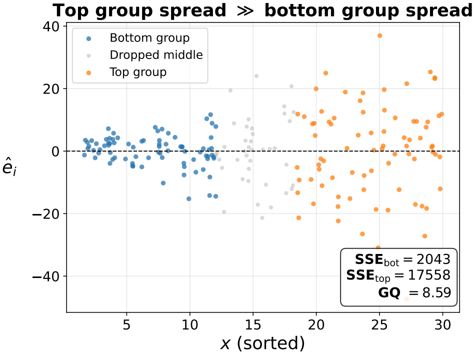


GQ Test: The Idea

Premise: If variance changes with x, the residual variance in a high-x subsample should differ from the residual variance in a low-x subsample.

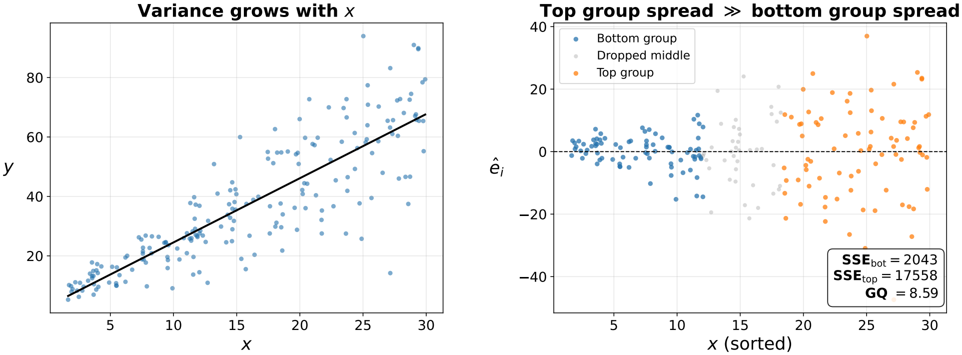
Procedure (sketch): sort the data by the suspect variable, drop the middle observations, run separate OLS on the bottom and top groups, compare the residual sums of squares.



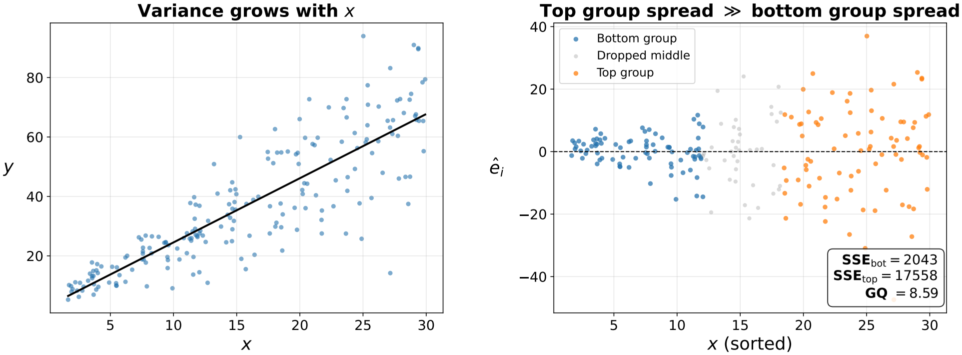


If SSE_{top} and SSE_{bot} differ enough $\Longrightarrow$ reject homoskedasticity.

When the GQ Test Works: Variance Grows with x



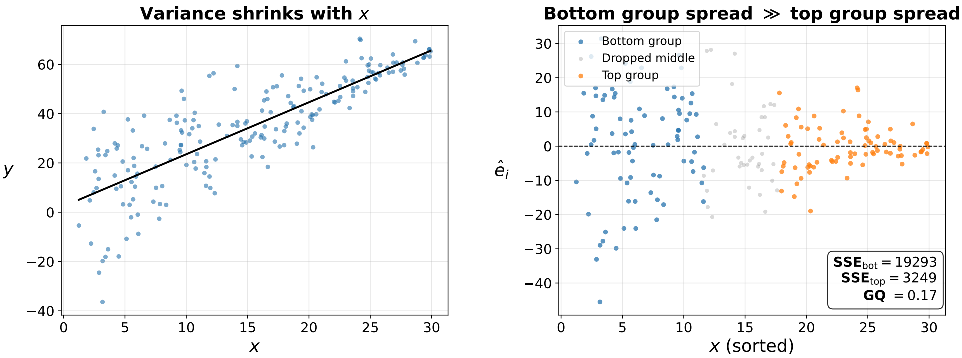
When the GQ Test Works: Variance Grows with x



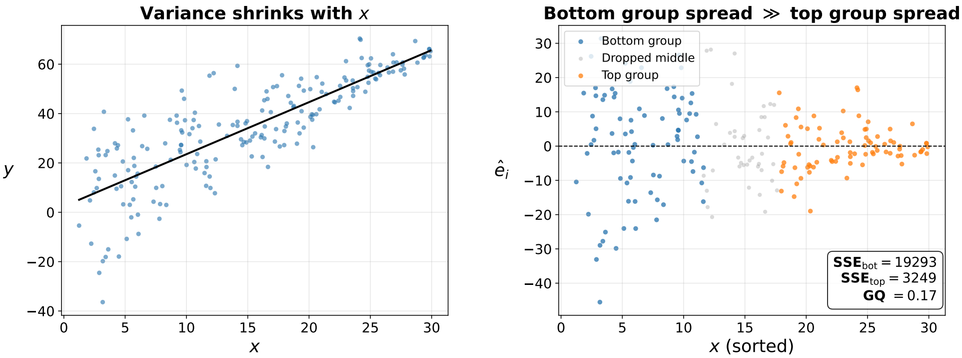
The top group's residuals fan out; SSE_{top} is several times SSE_{bot}. The GQ ratio is large.

GQ Test successfully detects heteroskedasticity!

When the GQ Test Works: Variance Shrinks with x



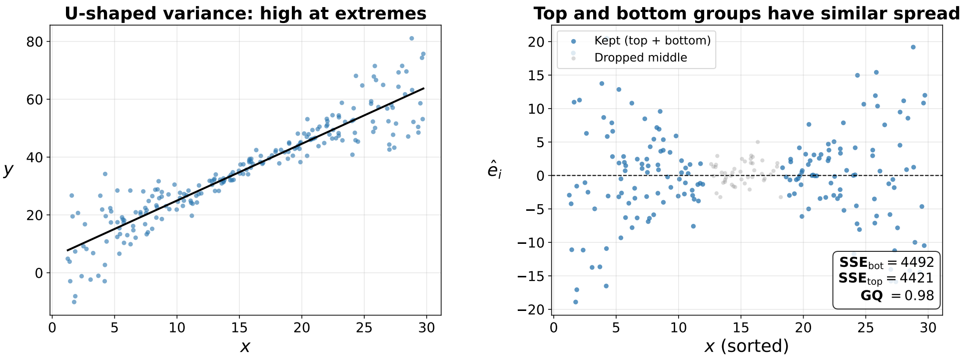
When the GQ Test Works: Variance Shrinks with x



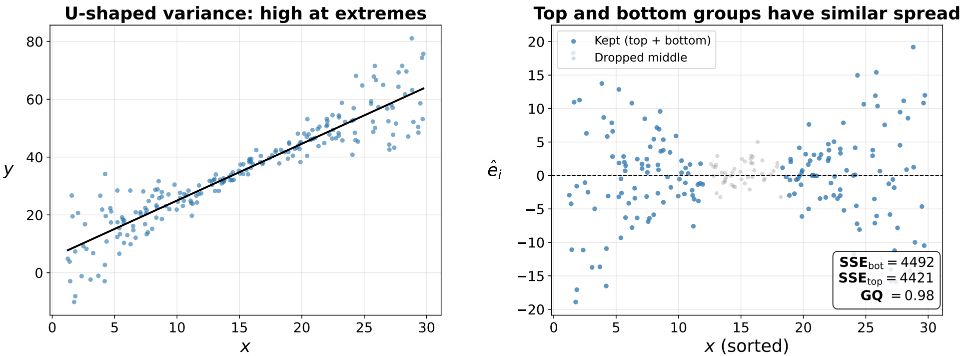
The bottom group's residuals fan out instead; SSE_{bot} dominates SSE_{top}. To run a one-sided right-tail F test the practitioner should put the suspected high-variance group on top (here: low-x).

GQ Test successfully detects heteroskedasticity!

When the GQ Test Fails: U-Shaped Variance



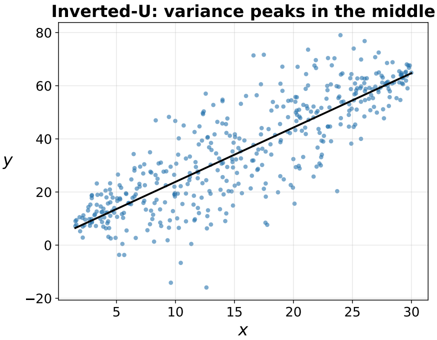
When the GQ Test Fails: U-Shaped Variance

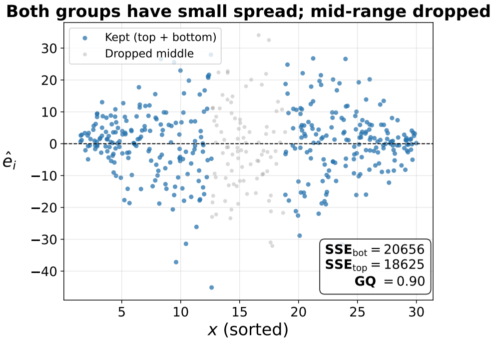


Variance is high at both extremes. Both top and bottom groups have similar (high) spread; the middle (low-variance) observations are exactly the ones GQ throws away. $\text{SSE}_{top} \approx \text{SSE}_{bot}$.

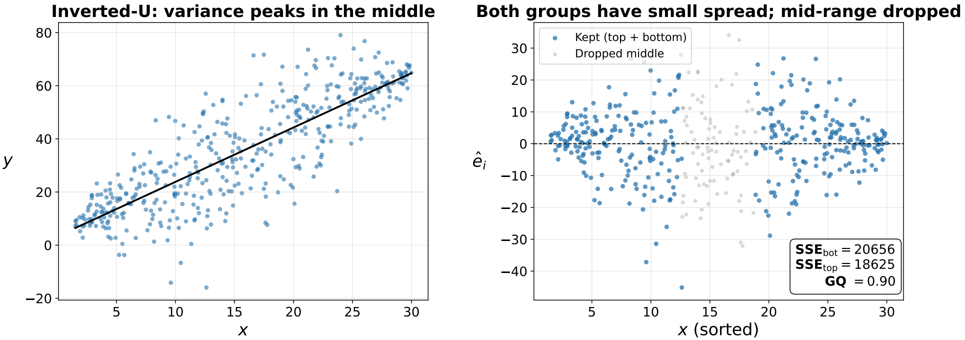
GQ Test fails to detect this kind of heteroskedasticity

When the GQ Test Fails: Inverted-U Variance





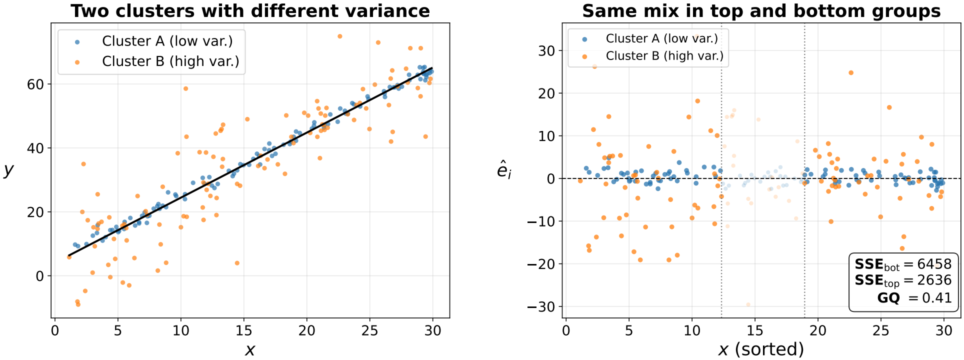
When the GQ Test Fails: Inverted-U Variance



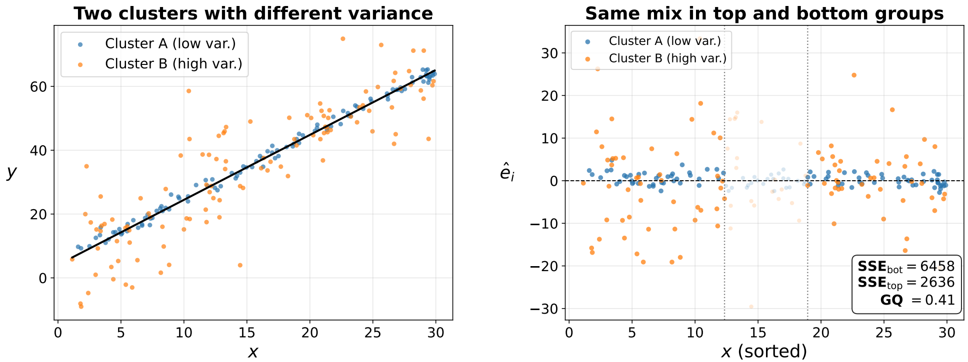
Variance peaks in the middle. The dropped middle is exactly the high-variance region; both retained groups have similar small spread. $\text{SSE}_{\text{top}} \approx \text{SSE}_{\text{bot}}$.

GQ Test fails to detect this kind of heteroskedasticity

When the GQ Test Fails: Clustered Variance



When the GQ Test Fails: Clustered Variance



Variance depends on *group*, not on x. After sorting by x, both top and bottom groups have a roughly equal mix of clusters, so SSE values are similar. Picking the wrong sort variable disarms GQ.

GQ Test fails to detect this kind of heteroskedasticity

GQ Test: Procedure

Steps:

1. Sort observations by the suspect variable x_i.

GQ Test: Procedure

Steps:

1. Sort observations by the suspect variable x_i.
2. Drop the middle c observations (typically $c \approx n/5$, to sharpen the contrast between low- and high-variance groups).

GQ Test: Procedure

Steps:

1. Sort observations by the suspect variable x_i.
2. Drop the middle c observations (typically $c \approx n/5$, to sharpen the contrast between low- and high-variance groups).
3. Run **separate** OLS on the bottom group and the top group. Save SSE_{bot} and SSE_{top}.

GQ Test: Procedure

Steps:

1. Sort observations by the suspect variable x_i.
2. Drop the middle c observations (typically $c \approx n/5$, to sharpen the contrast between low- and high-variance groups).
3. Run **separate** OLS on the bottom group and the top group. Save SSE_{bot} and SSE_{top}.
4. By convention put the group you suspect has *larger* variance on top. Compute

$$\text{GQ} = \frac{\text{SSE}_{\text{top}}/(n_{\text{top}} - k)}{\text{SSE}_{\text{bot}}/(n_{\text{bot}} - k)}$$

   where k is the number of regressors (including intercept).

GQ Test: Procedure

Steps:

1. Sort observations by the suspect variable x_i.
2. Drop the middle c observations (typically $c \approx n/5$, to sharpen the contrast between low- and high-variance groups).
3. Run **separate** OLS on the bottom group and the top group. Save SSE_{bot} and SSE_{top}.
4. By convention put the group you suspect has *larger* variance on top. Compute

$$\text{GQ} = \frac{\text{SSE}_{\text{top}}/(n_{\text{top}} - k)}{\text{SSE}_{\text{bot}}/(n_{\text{bot}} - k)}$$

   where k is the number of regressors (including intercept).

Under H_0 (homoskedasticity): $\text{GQ} \sim F(n_{\text{top}} - k,\ n_{\text{bot}} - k)$. **One-sided test:** reject H_0 if $\text{GQ} > F_{1-\alpha}(n_{\text{top}} - k,\ n_{\text{bot}} - k)$.

Example: Wages and Experience

We have data on 1,000 workers, with observed experience (years) and hourly wage (\$/hour).

$$\text{wage}_i = \beta_0 + \beta_1 \, \text{experience}_i + e_i$$

Example: Wages and Experience

We have data on 1,000 workers, with observed experience (years) and hourly wage ($/hour).

$$\text{wage}_i = \beta_0 + \beta_1\,\text{experience}_i + e_i$$

The fitted line slopes up; the spread of points around it widens with experience.

Step 1: Run OLS

Coefficients

	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
(Intercept)	8.8372	0.5234	16.88	$< 2 \cdot 10^{-16}$ ***
experience	1.0840	0.0303	35.81	$< 2 \cdot 10^{-16}$ ***

Residual standard error: 8.37 on 998 degrees of freedom
Multiple R^2*:* 0.5623 *Adjusted* R^2*:* 0.5619
F*-statistic:* 1282 on 1 and 998 DF, $p < 2 \cdot 10^{-16}$

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Step 1: Run OLS

Coefficients

	Estimate	Std. Error	t-value	$\Pr(>\lvert t \rvert)$
(Intercept)	8.8372	0.5234	16.88	$< 2 \cdot 10^{-16}$ ***
experience	1.0840	0.0303	35.81	$< 2 \cdot 10^{-16}$ ***

Residual standard error: 8.37 on 998 degrees of freedom
Multiple R^2: 0.5623 *Adjusted R^2:* 0.5619
F-statistic: 1282 on 1 and 998 DF, $p < 2 \cdot 10^{-16}$

Signif. codes: `0 *** 0.001 ** 0.01 * 0.05 . 0.1 1`

Slope is highly significant under OLS *assumptions*. **Are those SEs valid given the data?** GQ will tell us.

Step 2: State the Hypothesis

The GQ test asks whether the residual variance in the high-experience subsample exceeds the residual variance in the low-experience subsample.

Step 2: State the Hypothesis

The GQ test asks whether the residual variance in the high-experience subsample exceeds the residual variance in the low-experience subsample.

Null and alternative:

$$H_0 : \sigma^2_{\text{top}} = \sigma^2_{\text{bot}} \quad \text{(homoskedasticity)}$$
$$H_1 : \sigma^2_{\text{top}} > \sigma^2_{\text{bot}} \quad \text{(variance grows with experience)}$$

Step 2: State the Hypothesis

The GQ test asks whether the residual variance in the high-experience subsample exceeds the residual variance in the low-experience subsample.

Null and alternative:

$$H_0 : \sigma^2_{\text{top}} = \sigma^2_{\text{bot}} \quad \text{(homoskedasticity)}$$
$$H_1 : \sigma^2_{\text{top}} > \sigma^2_{\text{bot}} \quad \text{(variance grows with experience)}$$

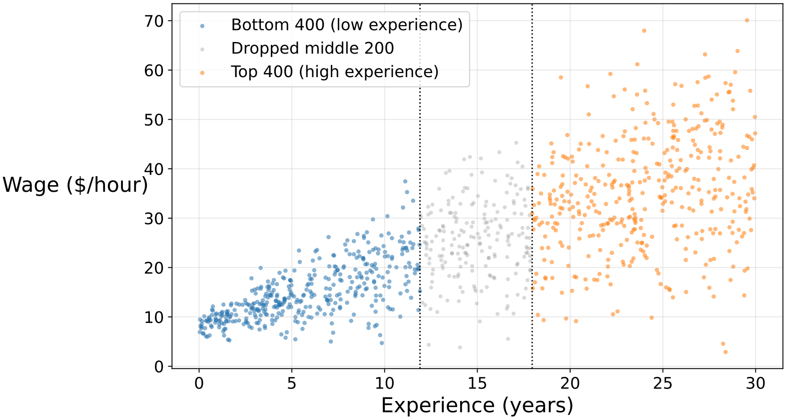
Test statistic:

$$\text{GQ} = \frac{\text{SSE}_{\text{top}}/(n_{\text{top}} - k)}{\text{SSE}_{\text{bot}}/(n_{\text{bot}} - k)} \sim F(n_{\text{top}} - k,\ n_{\text{bot}} - k) \text{ under } H_0$$

where k is the number of regressors (here, $k = 2$: intercept and experience).

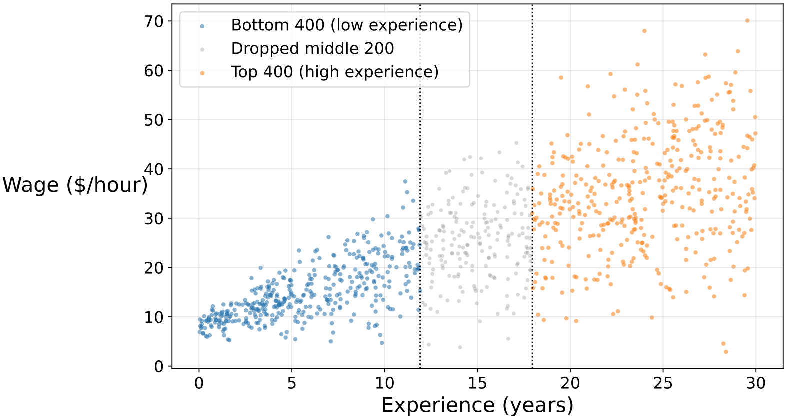
Step 3: Sort and Split

Sort the 1,000 observations by experience, drop the middle 200 (20%), keep 400 in each group.



Step 3: Sort and Split

Sort the 1,000 observations by experience, drop the middle 200 (20%), keep 400 in each group.



The dotted lines mark the split. The bottom group (low experience, blue) clearly has tighter spread than the top group (high experience, orange). Run separate OLS on each group and save SSE.

Step 4: Compute the Test Statistic

From the separate OLS regressions:

$$\text{SSE}_{\text{bot}} = 6{,}962, \qquad \text{SSE}_{\text{top}} = 50{,}159$$

With $n_{\text{bot}} = n_{\text{top}} = 400$ and $k = 2$, both groups have $n - k = 398$ degrees of freedom.

Step 4: Compute the Test Statistic

From the separate OLS regressions:

$$\text{SSE}_{\text{bot}} = 6{,}962, \qquad \text{SSE}_{\text{top}} = 50{,}159$$

With $n_{\text{bot}} = n_{\text{top}} = 400$ and $k = 2$, both groups have $n - k = 398$ degrees of freedom.

$$\begin{aligned} \text{GQ} &= \frac{\text{SSE}_{\text{top}}/398}{\text{SSE}_{\text{bot}}/398} \\ &= \frac{50{,}159}{6{,}962} \\ &= \boxed{7.21} \end{aligned}$$

Step 4: Compute the Test Statistic

From the separate OLS regressions:

$$\text{SSE}_{\text{bot}} = 6{,}962, \qquad \text{SSE}_{\text{top}} = 50{,}159$$

With $n_{\text{bot}} = n_{\text{top}} = 400$ and $k = 2$, both groups have $n - k = 398$ degrees of freedom.

$$\begin{aligned} \text{GQ} &= \frac{\text{SSE}_{\text{top}}/398}{\text{SSE}_{\text{bot}}/398} \\ &= \frac{50{,}159}{6{,}962} \\ &= \boxed{7.21} \end{aligned}$$

Compare to the critical value of $F(398, 398)$ at $\alpha = 0.05$.

Step 5: Compare to F Critical Value

For $F(\mathrm{df}_1, \mathrm{df}_2)$ at $\alpha = 0.05$ (one-sided right tail):

$\mathrm{df}_1 \backslash \mathrm{df}_2$	100	200	300	398	∞
100	1.392	1.320	1.296	1.282	1.243
200	1.320	1.243	1.215	1.199	1.155
300	1.296	1.215	1.184	1.169	1.119
398	1.282	1.199	1.169	**1.180**	1.094
∞	1.243	1.155	1.119	1.094	1.000

Step 5: Compare to F Critical Value

For $F(\text{df}_1, \text{df}_2)$ at $\alpha = 0.05$ (one-sided right tail):

$\text{df}_1 \backslash \text{df}_2$	100	200	300	398	∞
100	1.392	1.320	1.296	1.282	1.243
200	1.320	1.243	1.215	1.199	1.155
300	1.296	1.215	1.184	1.169	1.119
398	1.282	1.199	1.169	**1.180**	1.094
∞	1.243	1.155	1.119	1.094	1.000

With $\text{df}_1 = \text{df}_2 = 398$ and $\alpha = 0.05$:

$$F_{0.05}(398,\ 398) \approx 1.180$$

Step 5: Compare to F Critical Value

For $F(\mathrm{df}_1, \mathrm{df}_2)$ at $\alpha = 0.05$ (one-sided right tail):

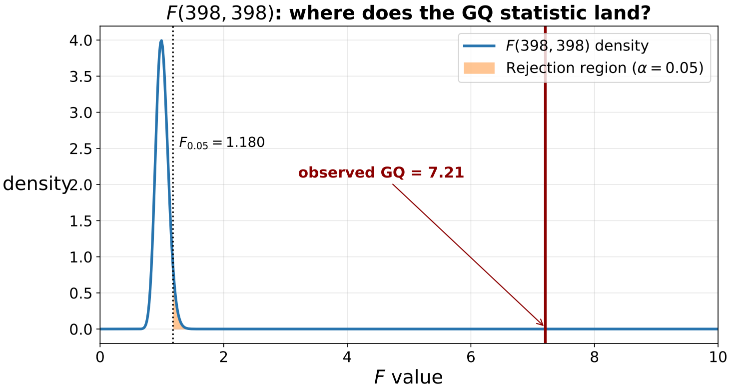
$\mathrm{df}_1 \backslash \mathrm{df}_2$	100	200	300	398	∞
100	1.392	1.320	1.296	1.282	1.243
200	1.320	1.243	1.215	1.199	1.155
300	1.296	1.215	1.184	1.169	1.119
398	1.282	1.199	1.169	**1.180**	1.094
∞	1.243	1.155	1.119	1.094	1.000

With $\mathrm{df}_1 = \mathrm{df}_2 = 398$ and $\alpha = 0.05$:

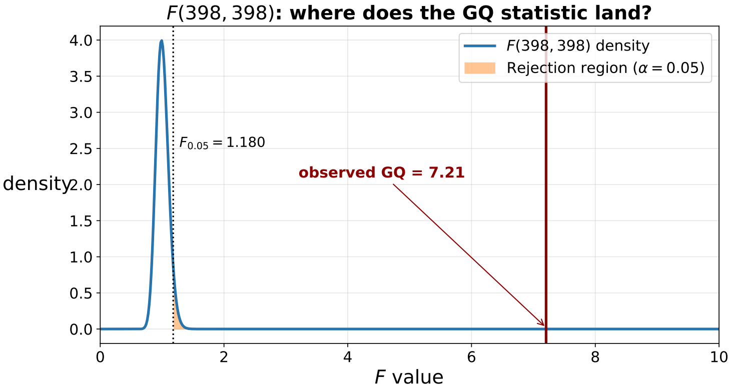
$$F_{0.05}(398, 398) \approx 1.180$$

Reject H_0 if $\mathrm{GQ} > 1.180$.

Step 6: Visualize the Decision



Step 6: Visualize the Decision



The orange region is the upper 5% tail of $F(398, 398)$. Our $\text{GQ} = 7.21$ is far past the right edge; the p-value is essentially zero.

Step 7: Conclusion

Decision: $\text{GQ} = 7.21 \gg 1.180 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Step 7: Conclusion

Decision: $\text{GQ} = 7.21 \gg 1.180 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Plain language: the residual variance in the high-experience group is many times the residual variance in the low-experience group. Wages are much more spread out for senior workers than for junior workers.

Step 7: Conclusion

Decision: $\text{GQ} = 7.21 \gg 1.180 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Plain language: the residual variance in the high-experience group is many times the residual variance in the low-experience group. Wages are much more spread out for senior workers than for junior workers.

What this means for our results:

- The OLS coefficients ($\hat{\beta}_0 = 8.84$, $\hat{\beta}_1 = 1.08$) are still unbiased and consistent.
- The OLS standard errors are **inconsistent** for the true sampling variance of $\hat{\beta}$; the discrepancy does not shrink with n.
- Use *robust* (HC) standard errors (consistent for the true sampling variance, so t-statistics and CIs have correct asymptotic distribution / coverage), or use WLS / FGLS if you have a credible model for the variance.

Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. **Fixing It: Robust Standard Errors**
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

We Found Heteroskedasticity. Now What?

A test rejects homoskedasticity. We have two strategies:

We Found Heteroskedasticity. Now What?

A test rejects homoskedasticity. We have two strategies:

1. **Keep OLS, fix the SE.** Use a heteroskedasticity-consistent variance estimator. Inference becomes valid; the point estimate is unchanged.

We Found Heteroskedasticity. Now What?

A test rejects homoskedasticity. We have two strategies:

1. **Keep OLS, fix the SE.** Use a heteroskedasticity-consistent variance estimator. Inference becomes valid; the point estimate is unchanged.
2. **Re-estimate with weights.** Use WLS (if you know σ_i^2) or FGLS (if you have to estimate it). Both inference and efficiency improve, but only when the variance model is correct.

We Found Heteroskedasticity. Now What?

A test rejects homoskedasticity. We have two strategies:

1. **Keep OLS, fix the SE.** Use a heteroskedasticity-consistent variance estimator. Inference becomes valid; the point estimate is unchanged.
2. **Re-estimate with weights.** Use WLS (if you know σ_i^2) or FGLS (if you have to estimate it). Both inference and efficiency improve, but only when the variance model is correct.

$\Longrightarrow$ Robust SEs are the default. WLS/FGLS pay off when you have a credible model for the variance.

Robust SEs: The Idea

Instead of assuming $\text{Var}(e_i \mid x_i) = \sigma^2$, estimate a **heteroskedasticity-consistent** (HC) variance:

Robust SEs: The Idea

Instead of assuming $\text{Var}(e_i \mid x_i) = \sigma^2$, estimate a **heteroskedasticity-consistent** (HC) variance:

HC0 (White):

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\sum_i (x_i - \bar{x})^2 \, \hat{e}_i^2}{\left[\sum_i (x_i - \bar{x})^2\right]^2}$$

Robust SEs: The Idea

Instead of assuming $\text{Var}(e_i \mid x_i) = \sigma^2$, estimate a **heteroskedasticity-consistent** (HC) variance:

HC0 (White):

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\sum_i (x_i - \bar{x})^2 \, \hat{e}_i^2}{\left[\sum_i (x_i - \bar{x})^2\right]^2}$$

Each observation contributes its **own** squared residual $\hat{e}_i^2$ rather than the pooled $\hat{\sigma}^2$.

Robust SEs: The Idea

Instead of assuming $\mathrm{Var}(e_i \mid x_i) = \sigma^2$, estimate a **heteroskedasticity-consistent** (HC) variance:

HC0 (White):

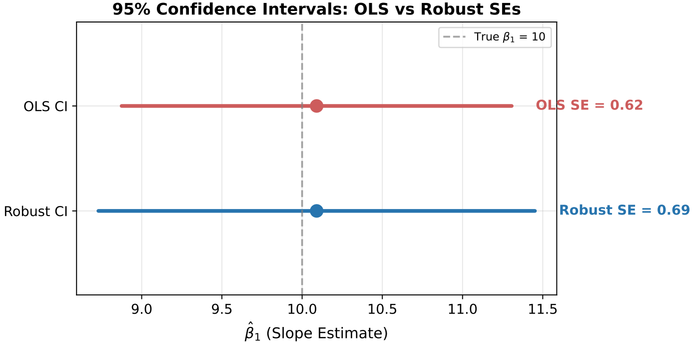
$$\widehat{\mathrm{Var}}(\hat{\beta}_1) = \frac{\sum_i (x_i - \bar{x})^2 \, \hat{e}_i^2}{\left[\sum_i (x_i - \bar{x})^2\right]^2}$$

Each observation contributes its **own** squared residual $\hat{e}_i^2$ rather than the pooled $\hat{\sigma}^2$.

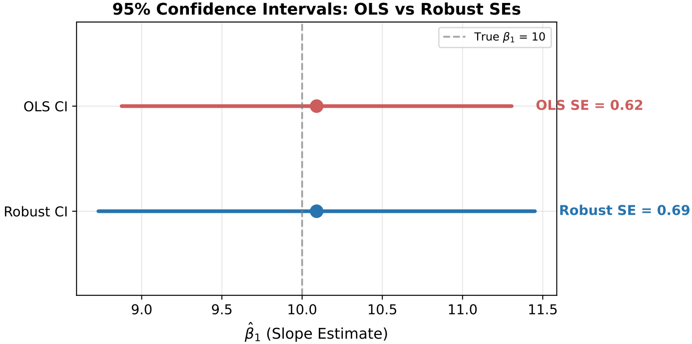
HC1 (small-sample correction): multiply by $\dfrac{n}{n-k}$

HC1 is the default "robust" SE in most software. With large n, HC0 and HC1 are nearly identical.

Robust SEs: What Changes?



Robust SEs: What Changes?



- The **point estimate** $\hat{\beta}_1$ does not change (same OLS regression)
- Only the **standard error** (and therefore the CI width) changes
- When variance grows with $|x_i - \bar{x}|$ (the typical case in food/income, wage/experience), OLS SEs are **too small** $\implies$ robust CIs are wider. The reverse can happen but is uncommon.

Robust SEs in R

Using the car and `lmtest` packages:

```
library(car) library(lmtest)
model <- lm(food   income, data = food_data)
# Usual OLS standard errors summary(model)
# Robust (HC1) standard errors coeftest(model, vcov. = hccm(model, type = "hc1"))
```

Robust SEs in R

Using the car and lmtest packages:

```
library(car) library(lmtest)
model <- lm(food   income, data = food_data)
# Usual OLS standard errors summary(model)
# Robust (HC1) standard errors coeftest(model, vcov. = hccm(model, type = "hc1"))
```

$\Longrightarrow$ Always report robust SEs unless you have strong reason to believe homoskedasticity holds. Many applied researchers use robust SEs by default.

Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

Weighted Least Squares: The Idea

Robust SEs fix inference but don't improve **efficiency**. Can we do better?

Weighted Least Squares: The Idea

Robust SEs fix inference but don't improve **efficiency**. Can we do better?

Intuition: Observations with small variance carry more information about the regression line. We should trust them more.

Weighted Least Squares: The Idea

Robust SEs fix inference but don't improve **efficiency**. Can we do better?

Intuition: Observations with small variance carry more information about the regression line. We should trust them more.

WLS weights: $w_i = 1/\sigma_i^2$

- Low-variance observations $\implies$ large weight
- High-variance observations $\implies$ small weight

Weighted Least Squares: The Idea

Robust SEs fix inference but don't improve **efficiency**. Can we do better?

Intuition: Observations with small variance carry more information about the regression line. We should trust them more.

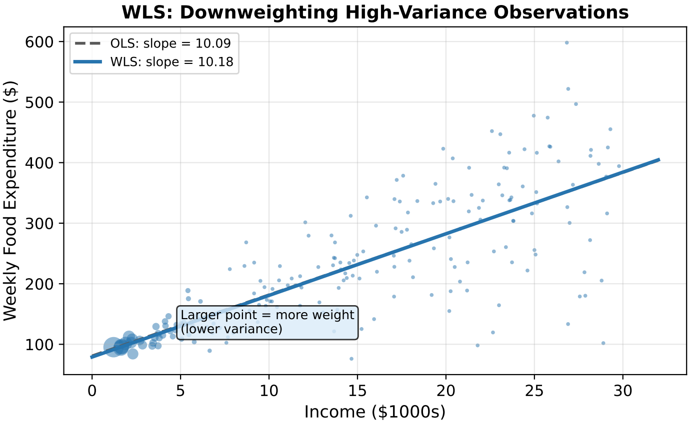
WLS weights: $w_i = 1/\sigma_i^2$

- Low-variance observations $\implies$ large weight
- High-variance observations $\implies$ small weight

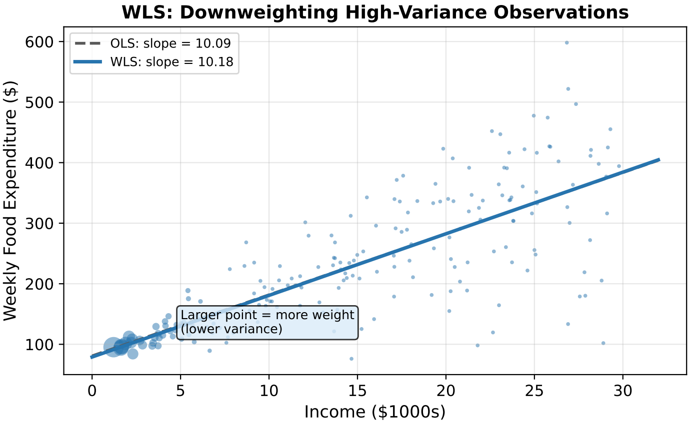
WLS minimizes:

$$\sum_i w_i(y_i - \beta_0 - \beta_1 x_i)^2 = \sum_i \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma_i^2}$$

WLS: Visual Intuition



WLS: Visual Intuition



Larger points have more weight (lower variance). The WLS line is pulled toward the low-income, low-variance observations where the data is most informative.

WLS as a Transformation

Goal: transform the model into one with a homoskedastic error.

Since $\text{Var}(e_i \mid x_i) = \sigma_i^2$, we want to divide e_i by σ_i (its standard deviation). Doing so gives a unit-variance error. We apply the same scaling to every term in the equation:

$$\frac{y_i}{\sigma_i} = \beta_0 \cdot \frac{1}{\sigma_i} + \beta_1 \cdot \frac{x_i}{\sigma_i} + \frac{e_i}{\sigma_i}$$

WLS as a Transformation

Goal: transform the model into one with a homoskedastic error.

Since $\text{Var}(e_i \mid x_i) = \sigma_i^2$, we want to divide e_i by σ_i (its standard deviation). Doing so gives a unit-variance error. We apply the same scaling to every term in the equation:

$$\frac{y_i}{\sigma_i} = \beta_0 \cdot \frac{1}{\sigma_i} + \beta_1 \cdot \frac{x_i}{\sigma_i} + \frac{e_i}{\sigma_i}$$

Define the transformed error $e_i^* \equiv e_i/\sigma_i$. Since σ_i^2 is a function of x_i, conditioning on x_i lets us pull it out of the variance:

$$\text{Var}(e_i^* \mid x_i) = \frac{\text{Var}(e_i \mid x_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

WLS as a Transformation

Goal: transform the model into one with a homoskedastic error.

Since $\text{Var}(e_i \mid x_i) = \sigma_i^2$, we want to divide e_i by σ_i (its standard deviation). Doing so gives a unit-variance error. We apply the same scaling to every term in the equation:

$$\frac{y_i}{\sigma_i} = \beta_0 \cdot \frac{1}{\sigma_i} + \beta_1 \cdot \frac{x_i}{\sigma_i} + \frac{e_i}{\sigma_i}$$

Define the transformed error $e_i^* \equiv e_i/\sigma_i$. Since σ_i^2 is a function of x_i, conditioning on x_i lets us pull it out of the variance:

$$\text{Var}(e_i^* \mid x_i) = \frac{\text{Var}(e_i \mid x_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

$\implies$ OLS on the transformed model is **homoskedastic**. This is exactly WLS.

WLS as a Transformation

Goal: transform the model into one with a homoskedastic error.

Since $\text{Var}(e_i \mid x_i) = \sigma_i^2$, we want to divide e_i by σ_i (its standard deviation). Doing so gives a unit-variance error. We apply the same scaling to every term in the equation:

$$\frac{y_i}{\sigma_i} = \beta_0 \cdot \frac{1}{\sigma_i} + \beta_1 \cdot \frac{x_i}{\sigma_i} + \frac{e_i}{\sigma_i}$$

Define the transformed error $e_i^* \equiv e_i/\sigma_i$. Since σ_i^2 is a function of x_i, conditioning on x_i lets us pull it out of the variance:

$$\text{Var}(e_i^* \mid x_i) = \frac{\text{Var}(e_i \mid x_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

$\implies$ OLS on the transformed model is **homoskedastic**. This is exactly WLS.

WLS is **BLUE** *when the weights* $1/\sigma_i^2$ *are correct* (Aitken's theorem) and is strictly more efficient than OLS in that case. With incorrect weights the efficiency advantage can disappear or reverse.

WLS: The Catch

WLS requires knowing σ_i^2 (or at least σ_i^2 up to a positive scalar; multiplying every w_i by the same constant gives identical $\hat{\beta}_{\text{WLS}}$, so only the relative weights matter).

WLS: The Catch

WLS requires knowing σ_i^2 (or at least σ_i^2 up to a positive scalar; multiplying every w_i by the same constant gives identical $\hat{\beta}_{\text{WLS}}$, so only the relative weights matter).

Common assumptions:

- $\text{Var}(e_i) = \sigma^2 x_i \implies w_i = 1/x_i$
- $\text{Var}(e_i) = \sigma^2 x_i^2 \implies w_i = 1/x_i^2$
- $\text{Var}(e_i) = \sigma^2/n_i$ (grouped data) $\implies w_i = n_i$

WLS: The Catch

WLS requires knowing σ_i^2 (or at least σ_i^2 up to a positive scalar; multiplying every w_i by the same constant gives identical $\hat{\beta}_{\text{WLS}}$, so only the relative weights matter).

Common assumptions:

- $\text{Var}(e_i) = \sigma^2 x_i \implies w_i = 1/x_i$
- $\text{Var}(e_i) = \sigma^2 x_i^2 \implies w_i = 1/x_i^2$
- $\text{Var}(e_i) = \sigma^2/n_i$ (grouped data) $\implies w_i = n_i$

R note: The `weights` argument in `lm()` expects $w_i = 1/\sigma_i^2$, not σ_i^2:

```
# If Var(e_i) proportional to income: wls <- lm(food   income, data = food_data, weights = 1 / income)
```

WLS: The Catch

WLS requires knowing σ_i^2 (or at least σ_i^2 up to a positive scalar; multiplying every w_i by the same constant gives identical $\hat{\beta}_{\text{WLS}}$, so only the relative weights matter).

Common assumptions:

- $\text{Var}(e_i) = \sigma^2 x_i \implies w_i = 1/x_i$
- $\text{Var}(e_i) = \sigma^2 x_i^2 \implies w_i = 1/x_i^2$
- $\text{Var}(e_i) = \sigma^2 / n_i$ (grouped data) $\implies w_i = n_i$

R note: The `weights` argument in `lm()` expects $w_i = 1/\sigma_i^2$, not σ_i^2:

```
# If Var(e_i) proportional to income: wls <- lm(food   income, data = food_data, weights = 1 / income)
```

$\implies$ If we know the variance form, WLS gives efficient estimates. But what if we don't know σ_i^2?

Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

Feasible GLS: The Idea

GLS (generalized least squares) is the broader family that includes WLS. WLS requires knowing the σ_i^2. **Feasible** GLS replaces them with an estimate $\hat{\sigma}_i^2$ obtained from the data in a first stage.

Feasible GLS: The Idea

GLS (generalized least squares) is the broader family that includes WLS. WLS requires knowing the σ_i^2. **Feasible** GLS replaces them with an estimate $\hat{\sigma}_i^2$ obtained from the data in a first stage.

Three steps:

1. Estimate a variance function $\hat{\sigma}_i^2$ from OLS residuals.
2. Decompose $\hat{\sigma}_i^2 = \hat{\sigma}^2 \cdot \hat{h}_i$ and divide each observation by $\sqrt{\hat{h}_i}$.
3. Run OLS on the transformed data.

Feasible GLS: The Idea

GLS (generalized least squares) is the broader family that includes WLS. WLS requires knowing the σ_i^2. **Feasible** GLS replaces them with an estimate $\hat{\sigma}_i^2$ obtained from the data in a first stage.

Three steps:

1. Estimate a variance function $\hat{\sigma}_i^2$ from OLS residuals.
2. Decompose $\hat{\sigma}_i^2 = \hat{\sigma}^2 \cdot \hat{h}_i$ and divide each observation by $\sqrt{\hat{h}_i}$.
3. Run OLS on the transformed data.

The next three frames work through the algebra of why this works.

Step 1: Estimate the Variance Function

Run OLS, save residuals $\hat{e}_i$. Postulate a model for the variance:

$$\sigma_i^2 = \exp(\alpha_1 + \alpha_2 z_i)$$

where z_i is any observable suspected to drive the variance (x_i, $\ln x_i$, or another variable). Estimate by OLS on the log of squared residuals:

$$\ln(\hat{e}_i^2) = \alpha_1 + \alpha_2 z_i + v_i$$

Step 1: Estimate the Variance Function

Run OLS, save residuals $\hat{e}_i$. Postulate a model for the variance:

$$\sigma_i^2 = \exp(\alpha_1 + \alpha_2 z_i)$$

where z_i is any observable suspected to drive the variance (x_i, $\ln x_i$, or another variable). Estimate by OLS on the log of squared residuals:

$$\ln(\hat{e}_i^2) = \alpha_1 + \alpha_2 z_i + v_i$$

Logging $\hat{e}_i^2$ has two purposes:

- Keeps $\hat{\sigma}_i^2 > 0$ after exponentiating.
- Linearizes a multiplicative variance structure into something OLS can fit.

Step 1: Estimate the Variance Function

Run OLS, save residuals $\hat{e}_i$. Postulate a model for the variance:

$$\sigma_i^2 = \exp(\alpha_1 + \alpha_2 z_i)$$

where z_i is any observable suspected to drive the variance (x_i, $\ln x_i$, or another variable). Estimate by OLS on the log of squared residuals:

$$\ln(\hat{e}_i^2) = \alpha_1 + \alpha_2 z_i + v_i$$

Logging $\hat{e}_i^2$ has two purposes:

- Keeps $\hat{\sigma}_i^2 > 0$ after exponentiating.
- Linearizes a multiplicative variance structure into something OLS can fit.

$\Longrightarrow$ The fitted variance is $\hat{\sigma}_i^2 = \exp(\hat{\alpha}_1 + \hat{\alpha}_2 z_i)$.

Decomposing the Variance: Baseline × Skedastic Function

Rewrite the estimated variance as a product:

$$\hat{\sigma}_i^2 = \underbrace{e^{\hat{\alpha}_1}}_{\hat{\sigma}^2} \cdot \underbrace{e^{\hat{\alpha}_2 z_i}}_{\hat{h}_i}$$

- $\hat{\sigma}^2$: a common "baseline" variance shared by every observation.
- $\hat{h}_i$: the **skedastic function** $\Longrightarrow$ how much observation i's variance deviates from the baseline.

Decomposing the Variance: Baseline × Skedastic Function

Rewrite the estimated variance as a product:

$$\hat{\sigma}_i^2 = \underbrace{e^{\hat{\alpha}_1}}_{\hat{\sigma}^2} \cdot \underbrace{e^{\hat{\alpha}_2 z_i}}_{\hat{h}_i}$$

- $\hat{\sigma}^2$: a common "baseline" variance shared by every observation.
- $\hat{h}_i$: the **skedastic function** $\Longrightarrow$ how much observation i's variance deviates from the baseline.

This decomposition is what makes the transformation work: $\hat{\sigma}^2$ is a constant we can drop, and $\hat{h}_i$ is the only thing that varies across i.

Decomposing the Variance: Baseline × Skedastic Function

Rewrite the estimated variance as a product:

$$\hat{\sigma}_i^2 = \underbrace{e^{\hat{\alpha}_1}}_{\hat{\sigma}^2} \cdot \underbrace{e^{\hat{\alpha}_2 z_i}}_{\hat{h}_i}$$

- $\hat{\sigma}^2$: a common "baseline" variance shared by every observation.
- $\hat{h}_i$: the **skedastic function** $\implies$ how much observation i's variance deviates from the baseline.

This decomposition is what makes the transformation work: $\hat{\sigma}^2$ is a constant we can drop, and $\hat{h}_i$ is the only thing that varies across i.

$\implies$ We only need $\hat{h}_i$, not $\hat{\sigma}^2$, to rescale the data.

Why Dividing by $\sqrt{\hat{h}_i}$ Restores Homoskedasticity

Define the rescaled error $e_i^* = e_i/\sqrt{h_i}$. Its conditional variance is:

$$\text{Var}(e_i^* \mid x_i) = \text{Var}\left(\frac{e_i}{\sqrt{h_i}} \,\middle|\, x_i\right) = \frac{1}{h_i}\,\text{Var}(e_i \mid x_i) = \frac{1}{h_i}\,\sigma^2 h_i = \sigma^2.$$

Why Dividing by $\sqrt{\hat{h}_i}$ Restores Homoskedasticity

Define the rescaled error $e_i^* = e_i/\sqrt{h_i}$. Its conditional variance is:

$$\text{Var}(e_i^* \mid x_i) = \text{Var}\left(\frac{e_i}{\sqrt{h_i}} \,\middle|\, x_i\right) = \frac{1}{h_i}\,\text{Var}(e_i \mid x_i) = \frac{1}{h_i}\,\sigma^2 h_i = \sigma^2.$$

- Pull $1/\sqrt{h_i}$ out as a constant (it's a function of x_i, which we condition on).
- Square the constant: $(1/\sqrt{h_i})^2 = 1/h_i$.
- Substitute $\text{Var}(e_i \mid x_i) = \sigma^2 h_i$ from the variance model.
- The h_i's cancel $\implies$ the rescaled error has constant variance σ^2.

Why Dividing by $\sqrt{\hat{h}_i}$ Restores Homoskedasticity

Define the rescaled error $e_i^* = e_i/\sqrt{h_i}$. Its conditional variance is:

$$\text{Var}(e_i^* \mid x_i) = \text{Var}\left(\frac{e_i}{\sqrt{h_i}} \,\middle|\, x_i\right) = \frac{1}{h_i}\,\text{Var}(e_i \mid x_i) = \frac{1}{h_i}\,\sigma^2 h_i = \sigma^2.$$

- Pull $1/\sqrt{h_i}$ out as a constant (it's a function of x_i, which we condition on).
- Square the constant: $(1/\sqrt{h_i})^2 = 1/h_i$.
- Substitute $\text{Var}(e_i \mid x_i) = \sigma^2 h_i$ from the variance model.
- The h_i's cancel $\implies$ the rescaled error has constant variance σ^2.

$\implies$ On the rescaled data, the homoskedasticity assumption holds and OLS is BLUE again.

Step 2: Transform the Data

Apply the rescaling to every variable in the regression:

$$y_i^* = \frac{y_i}{\sqrt{\hat{h}_i}}, \qquad x_{i1}^* = \frac{1}{\sqrt{\hat{h}_i}}, \qquad x_{i2}^* = \frac{x_{i2}}{\sqrt{\hat{h}_i}}.$$

Step 2: Transform the Data

Apply the rescaling to every variable in the regression:

$$y_i^* = \frac{y_i}{\sqrt{\hat{h}_i}}, \qquad x_{i1}^* = \frac{1}{\sqrt{\hat{h}_i}}, \qquad x_{i2}^* = \frac{x_{i2}}{\sqrt{\hat{h}_i}}.$$

Why divide *every* column (including the intercept's column of 1's)? Because the transformation must apply to both sides of the regression:

$$\frac{y_i}{\sqrt{\hat{h}_i}} = \beta_1 \cdot \frac{1}{\sqrt{\hat{h}_i}} + \beta_2 \cdot \frac{x_{i2}}{\sqrt{\hat{h}_i}} + \frac{e_i}{\sqrt{\hat{h}_i}}.$$

Step 2: Transform the Data

Apply the rescaling to every variable in the regression:

$$y_i^* = \frac{y_i}{\sqrt{\hat{h}_i}}, \qquad x_{i1}^* = \frac{1}{\sqrt{\hat{h}_i}}, \qquad x_{i2}^* = \frac{x_{i2}}{\sqrt{\hat{h}_i}}.$$

Why divide *every* column (including the intercept's column of 1's)? Because the transformation must apply to both sides of the regression:

$$\frac{y_i}{\sqrt{\hat{h}_i}} = \beta_1 \cdot \frac{1}{\sqrt{\hat{h}_i}} + \beta_2 \cdot \frac{x_{i2}}{\sqrt{\hat{h}_i}} + \frac{e_i}{\sqrt{\hat{h}_i}}.$$

Aside: why not also divide by $\hat{\sigma}$? Multiplying every observation by the same constant $1/\hat{\sigma}$ doesn't change OLS estimates or their relative weighting $\implies$ we only need $\sqrt{\hat{h}_i}$.

Step 3: Run OLS on the Transformed Data

Apply OLS to $(y_i^*, x_{i1}^*, x_{i2}^*)$. The resulting estimator $\hat{\beta}^{\text{FGLS}}$ has the same form as any OLS coefficient.

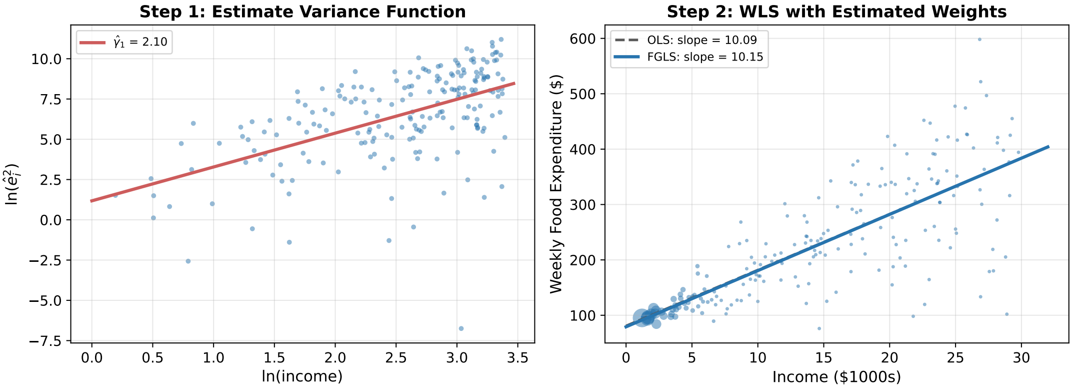
Step 3: Run OLS on the Transformed Data

Apply OLS to $(y_i^*, x_{i1}^*, x_{i2}^*)$. The resulting estimator $\hat{\beta}^{\text{FGLS}}$ has the same form as any OLS coefficient.

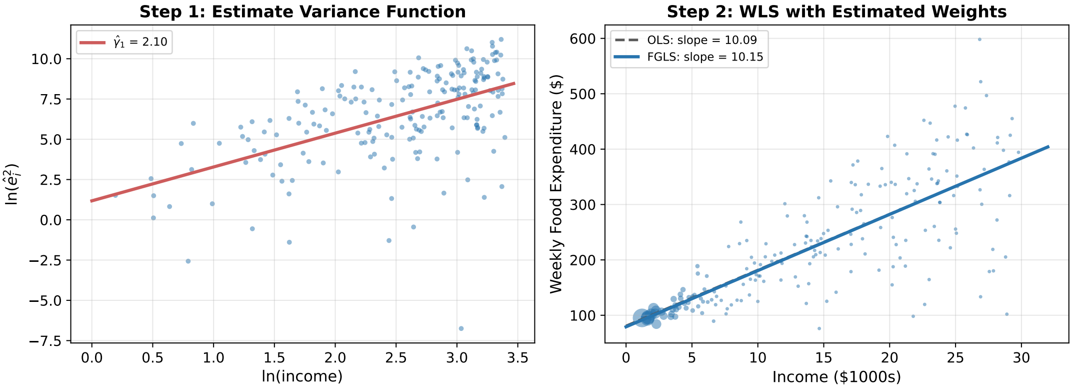
Properties:

- If the variance model is correctly specified, FGLS is consistent for β and asymptotically as efficient as true GLS.
- If the variance model is mis-specified, FGLS is still consistent but is not guaranteed to beat OLS.
- Robust SEs can be combined with FGLS for an additional layer of insurance.

FGLS in Two Pictures



FGLS in Two Pictures



Left: the auxiliary regression reveals how variance scales with income. Right: FGLS uses these estimated weights to refit the model.

Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

Setup: The Food Expenditure Model

Note on indexing: HGL labels the intercept β_1 and the slope β_2.

Model: $\text{food}_i = \beta_1 + \beta_2\,\text{income}_i + e_i \quad (n = 40$ households; income in \$100s.)

Setup: The Food Expenditure Model

Note on indexing: HGL labels the intercept β_1 and the slope β_2.

Model: $\text{food}_i = \beta_1 + \beta_2\,\text{income}_i + e_i \quad (n = 40$ households; income in \$100s.)

OLS results (assuming homoskedasticity):

$$\widehat{\text{food}} = 83.42 + 10.21\,\text{income}$$

	$\hat{\beta}_1$ (intercept)	$\hat{\beta}_2$ (income)
Estimate	83.42	10.21
OLS SE	43.41	2.09

Setup: The Food Expenditure Model

Note on indexing: HGL labels the intercept β_1 and the slope β_2.

Model: $\text{food}_i = \beta_1 + \beta_2\,\text{income}_i + e_i \quad (n = 40$ households; income in \$100s.)

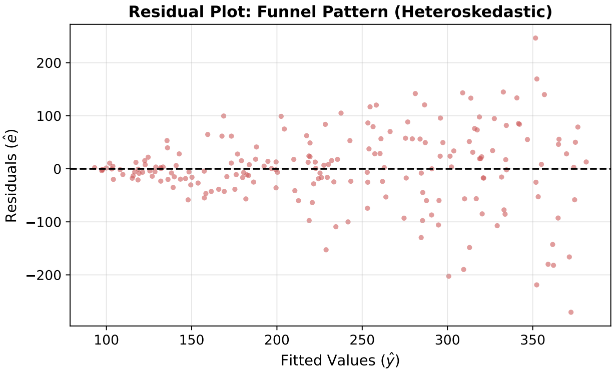
OLS results (assuming homoskedasticity):

$$\widehat{\text{food}} = 83.42 + 10.21\,\text{income}$$

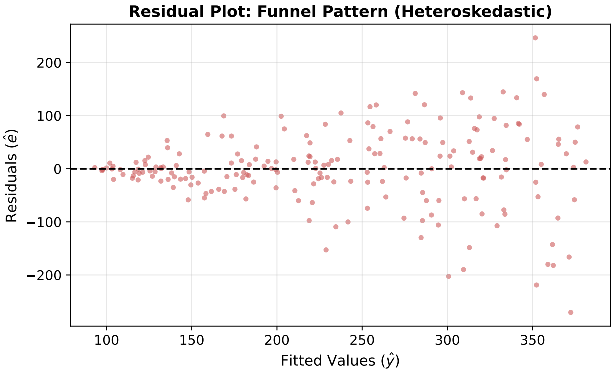
	$\hat{\beta}_1$ (intercept)	$\hat{\beta}_2$ (income)
Estimate	83.42	10.21
OLS SE	43.41	2.09

$t = 10.21/2.09 = 4.88$ on $\hat{\beta}_2 \implies$ significant at any conventional level. But the OLS SE is only valid if errors are homoskedastic.

Diagnose: Is There Heteroskedasticity?

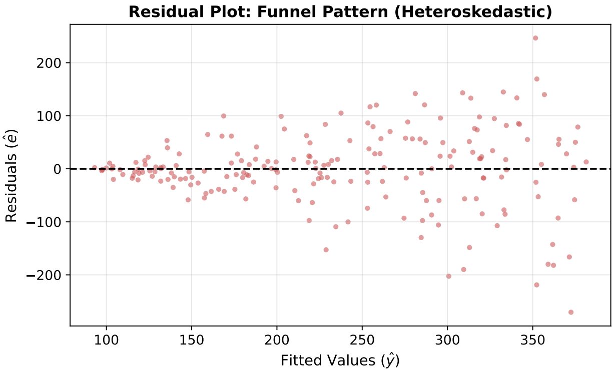


Diagnose: Is There Heteroskedasticity?



Residual plot: clear funnel pattern $\implies$ variance grows with income.

Diagnose: Is There Heteroskedasticity?



Residual plot: clear funnel pattern $\implies$ variance grows with income.

Formal tests (covered in earlier sections):

- Goldfeld–Quandt: $\text{GQ} = 3.61 > F_{0.05}(14, 14) = 2.48 \implies$ reject. (BP test also rejects)
- OLS coefficients are still unbiased, but the $\text{SE} = 2.09$ is wrong.

Two Routes Forward

Heteroskedasticity has been detected. Two options:

Two Routes Forward

Heteroskedasticity has been detected. Two options:

Route A: Robust standard errors (HC1).

- Keep OLS coefficient estimates.
- Replace the SE formula with the sandwich estimator.
- Pro: no need to model the variance. Con: still uses OLS weighting, so it's not efficient.

Two Routes Forward

Heteroskedasticity has been detected. Two options:

Route A: Robust standard errors (HC1).

- Keep OLS coefficient estimates.
- Replace the SE formula with the sandwich estimator.
- Pro: no need to model the variance. Con: still uses OLS weighting, so it's not efficient.

Route B: FGLS.

- Estimate a variance function, rescale the data, refit by OLS on transformed variables.
- Pro: efficient if the variance model is correctly specified. Con: requires a variance specification.

Two Routes Forward

Heteroskedasticity has been detected. Two options:

Route A: Robust standard errors (HC1).

- Keep OLS coefficient estimates.
- Replace the SE formula with the sandwich estimator.
- Pro: no need to model the variance. Con: still uses OLS weighting, so it's not efficient.

Route B: FGLS.

- Estimate a variance function, rescale the data, refit by OLS on transformed variables.
- Pro: efficient if the variance model is correctly specified. Con: requires a variance specification.

$\Longrightarrow$ Below: run both, then compare.

Route A: Robust SEs

Compare OLS vs. HC1 robust standard errors for $\hat{\beta}_2$:

Route A: Robust SEs

Compare OLS vs. HC1 robust standard errors for $\hat{\beta}_2$:

	SE($\hat{\beta}_2$)	95% CI for β_2	CI Width
OLS	2.09	$(5.97, 14.45)$	8.47
Robust (HC1)	1.81	$(6.55, 13.87)$	7.33

Route A: Robust SEs

Compare OLS vs. HC1 robust standard errors for $\hat{\beta}_2$:

	$SE(\hat{\beta}_2)$	95% CI for β_2	CI Width
OLS	2.09	$(5.97, 14.45)$	8.47
Robust (HC1)	1.81	$(6.55, 13.87)$	7.33

- In *this* sample, the robust SE is slightly **smaller** than the OLS SE. The OLS SE is invalid under heteroskedasticity, but whether it over- or under-states the truth depends on how σ_i^2 correlates with $(x_i - \bar{x})^2$ in the realized data.
- Both intervals reject $H_0: \beta_2 = 0$.

Route A: Robust SEs

Compare OLS vs. HC1 robust standard errors for $\hat{\beta}_2$:

	SE($\hat{\beta}_2$)	95% CI for β_2	CI Width
OLS	2.09	$(5.97, 14.45)$	8.47
Robust (HC1)	1.81	$(6.55, 13.87)$	7.33

- In *this* sample, the robust SE is slightly **smaller** than the OLS SE. The OLS SE is invalid under heteroskedasticity, but whether it over- or under-states the truth depends on how σ_i^2 correlates with $(x_i - \bar{x})^2$ in the realized data.
- Both intervals reject $H_0 : \beta_2 = 0$.

$\Longrightarrow$ The coefficient stays at 10.21; only the SE changes.

Route B, Step 1: Specify the Variance Function

Inspecting the residual plot suggests the variance scales with income. Two common specifications:

- **Proportional:** $\sigma_i^2 = \sigma^2 \cdot \text{income}_i \quad (h_i = \text{income}_i)$.
- **Exponential:** $\sigma_i^2 = \exp(\alpha_1 + \alpha_2 \ln \text{income}_i)$.

Route B, Step 1: Specify the Variance Function

Inspecting the residual plot suggests the variance scales with income. Two common specifications:

- **Proportional:** $\sigma_i^2 = \sigma^2 \cdot \text{income}_i \quad (h_i = \text{income}_i)$.
- **Exponential:** $\sigma_i^2 = \exp(\alpha_1 + \alpha_2 \ln \text{income}_i)$.

Estimate the exponential form by regressing log squared residuals on $\ln \text{income}_i$:

$$\ln(\hat{e}_i^2) = \hat{\alpha}_1 + \hat{\alpha}_2 \ln(\text{income}_i) + v_i$$

Fitted on the HGL food data: $\hat{\alpha}_1 = 0.938$, $\hat{\alpha}_2 = 2.329 \implies$ variance grows roughly with the square of income.

Route B, Step 1: Specify the Variance Function

Inspecting the residual plot suggests the variance scales with income. Two common specifications:

- **Proportional:** $\sigma_i^2 = \sigma^2 \cdot \text{income}_i \quad (h_i = \text{income}_i)$.
- **Exponential:** $\sigma_i^2 = \exp(\alpha_1 + \alpha_2 \ln \text{income}_i)$.

Estimate the exponential form by regressing log squared residuals on $\ln \text{income}_i$:

$$\ln(\hat{e}_i^2) = \hat{\alpha}_1 + \hat{\alpha}_2 \ln(\text{income}_i) + v_i$$

Fitted on the HGL food data: $\hat{\alpha}_1 = 0.938$, $\hat{\alpha}_2 = 2.329 \implies$ variance grows roughly with the square of income.

Decompose:

$$\hat{\sigma}_i^2 = \underbrace{e^{\hat{\alpha}_1}}_{\hat{\sigma}^2} \cdot \underbrace{\text{income}_i^{\hat{\alpha}_2}}_{\hat{h}_i}$$

Route B, Step 2: Transform and Refit

Divide every variable by $\sqrt{\hat{h}_i}$:

$$\text{food}_i^* = \frac{\text{food}_i}{\sqrt{\hat{h}_i}}, \qquad 1_i^* = \frac{1}{\sqrt{\hat{h}_i}}, \qquad \text{income}_i^* = \frac{\text{income}_i}{\sqrt{\hat{h}_i}}.$$

Run OLS on $(\text{food}^*, 1^*, \text{income}^*)$.

Route B, Step 2: Transform and Refit

Divide every variable by $\sqrt{\hat{h}_i}$:

$$\text{food}_i^* = \frac{\text{food}_i}{\sqrt{\hat{h}_i}}, \qquad 1_i^* = \frac{1}{\sqrt{\hat{h}_i}}, \qquad \text{income}_i^* = \frac{\text{income}_i}{\sqrt{\hat{h}_i}}.$$

Run OLS on $(\text{food}^*, 1^*, \text{income}^*)$.

FGLS results on the HGL food data:

$$\widehat{\text{food}}^{\text{FGLS}} = 76.05 + 10.63\,\text{income}, \qquad \text{SE}(\hat{\beta}_2^{\text{FGLS}}) = 0.97.$$

Route B, Step 2: Transform and Refit

Divide every variable by $\sqrt{\hat{h}_i}$:

$$\text{food}_i^* = \frac{\text{food}_i}{\sqrt{\hat{h}_i}}, \qquad 1_i^* = \frac{1}{\sqrt{\hat{h}_i}}, \qquad \text{income}_i^* = \frac{\text{income}_i}{\sqrt{\hat{h}_i}}.$$

Run OLS on $(\text{food}^*, 1^*, \text{income}^*)$.

FGLS results on the HGL food data:

$$\widehat{\text{food}}^{\text{FGLS}} = 76.05 + 10.63\,\text{income}, \qquad \text{SE}(\hat{\beta}_2^{\text{FGLS}}) = 0.97.$$

$\implies$ FGLS slope is close to OLS (10.21), but its SE is roughly half of both the OLS SE (2.09) and the robust SE (1.81) $\implies$ large efficiency gain from down-weighting the noisy high-income observations.

Comparison: OLS vs Robust vs FGLS

Method	$\hat{\beta}_2$	SE($\hat{\beta}_2$)	95% CI
OLS (homosk. SE)	10.21	2.09	(5.97, 14.45)
OLS + Robust HC1	10.21	1.81	(6.55, 13.87)
FGLS ($h_i = \text{income}_i^{2.33}$)	10.63	0.97	(8.67, 12.60)

Comparison: OLS vs Robust vs FGLS

Method	$\hat{\beta}_2$	SE($\hat{\beta}_2$)	95% CI
OLS (homosk. SE)	10.21	2.09	(5.97, 14.45)
OLS + Robust HC1	10.21	1.81	(6.55, 13.87)
FGLS ($h_i = \text{income}_i^{2.33}$)	10.63	0.97	(8.67, 12.60)

Reading the comparison:

- Coefficient estimate barely moves between OLS and FGLS $\Longrightarrow$ OLS was unbiased.
- OLS vs. robust SE differ only modestly here; in other samples the gap can be larger and run in either direction.
- FGLS SE is roughly half of the OLS/robust SEs $\Longrightarrow$ efficiency gain when the variance model is correctly specified.

Comparison: OLS vs Robust vs FGLS

Method	$\hat{\beta}_2$	SE($\hat{\beta}_2$)	95% CI
OLS (homosk. SE)	10.21	2.09	(5.97, 14.45)
OLS + Robust HC1	10.21	1.81	(6.55, 13.87)
FGLS ($h_i = \text{income}_i^{2.33}$)	10.63	0.97	(8.67, 12.60)

Reading the comparison:

- Coefficient estimate barely moves between OLS and FGLS $\Longrightarrow$ OLS was unbiased.
- OLS vs. robust SE differ only modestly here; in other samples the gap can be larger and run in either direction.
- FGLS SE is roughly half of the OLS/robust SEs $\Longrightarrow$ efficiency gain when the variance model is correctly specified.

When to use which:

- Unsure about the variance structure $\Longrightarrow$ robust SEs.
- Confident in a variance specification and want efficiency $\Longrightarrow$ FGLS.

R Code: All Three Estimators

```r
library(car); library(lmtest); library(sandwich)

# 1. OLS
m_ols <- lm(food ~ income, data = food_data)
summary(m_ols)

# 2. OLS with HC1 robust SEs
V_hc1 <- vcovHC(m_ols, type = "HC1")
coeftest(m_ols, vcov. = V_hc1)

# 3. FGLS: estimate variance function, then refit with weights
log_resid_sq <- log(resid(m_ols)^2)
var_fit      <- lm(log_resid_sq ~ log(income), data = food_data)
h_hat        <- exp(fitted(var_fit))
m_fgls       <- lm(food ~ income, data = food_data, weights = 1 / h_hat)
summary(m_fgls)
```

Outline

1. Motivation
2. What Goes Wrong
3. Visual Detection: Residual Plots
4. Breusch-Pagan Test
5. White Test
6. Goldfeld-Quandt Test
7. Fixing It: Robust Standard Errors
8. Fixing It: WLS
9. Fixing It: FGLS
10. End-to-End: Food Expenditure
11. Summary

Summary: Which Fix to Use?

Method	**When to Use**	**Limitation**
Robust SEs	Always safe; default in applied work	No efficiency gain
WLS	You know $\text{Var}(e_i) \propto h(x_i)$	Requires correct variance form
FGLS	Can estimate the variance function	Relies on auxiliary regression

Summary: Which Fix to Use?

Method	When to Use	Limitation
Robust SEs	Always safe; default in applied work	No efficiency gain
WLS	You know $\text{Var}(e_i) \propto h(x_i)$	Requires correct variance form
FGLS	Can estimate the variance function	Relies on auxiliary regression

Practical advice:

1. Always check residual plots after running OLS
2. Report robust SEs by default
3. Use WLS/FGLS only when you have a good model for the variance structure

Summary: Which Fix to Use?

Method	When to Use	Limitation
Robust SEs	Always safe; default in applied work	No efficiency gain
WLS	You know $\text{Var}(e_i) \propto h(x_i)$	Requires correct variance form
FGLS	Can estimate the variance function	Relies on auxiliary regression

Practical advice:

1. Always check residual plots after running OLS
2. Report robust SEs by default
3. Use WLS/FGLS only when you have a good model for the variance structure

$\Longrightarrow$ Robust SEs are the minimum standard. WLS/FGLS can do better if you model the variance correctly.

Thank you!

jakeanderson@g.ucla.edu