Elements of Econometrics. Lecture 10. Heteroscedasticity.

FCS, 2022-2023

ASSUMPTIONS FOR MODEL A
VIOLATION:

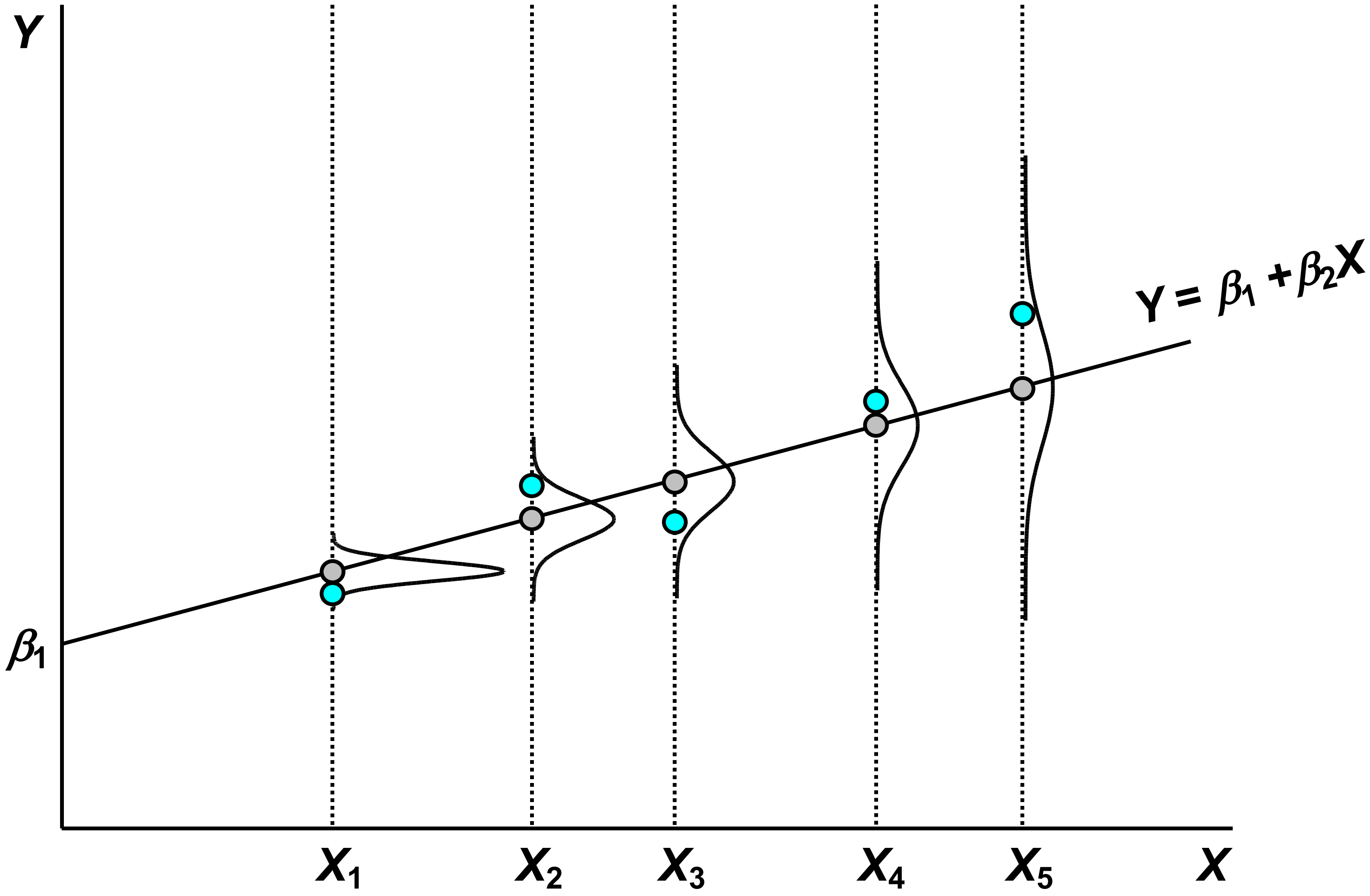
1. Reasons
   Consequences
   Detection
   Remedial measures

A.4 The disturbance term is homoscedastic: $\sigma_{u_i}^2 = \sigma_u^2$ **for all *i* (Gauss-Markov 2 condition)**

Heteroscedasticity: $\sigma_{u_i}^2 = \sigma_i^2$

Reasons: the factors combined in the disturbance term may change in scale together with X.

HETEROSCEDASTICITY

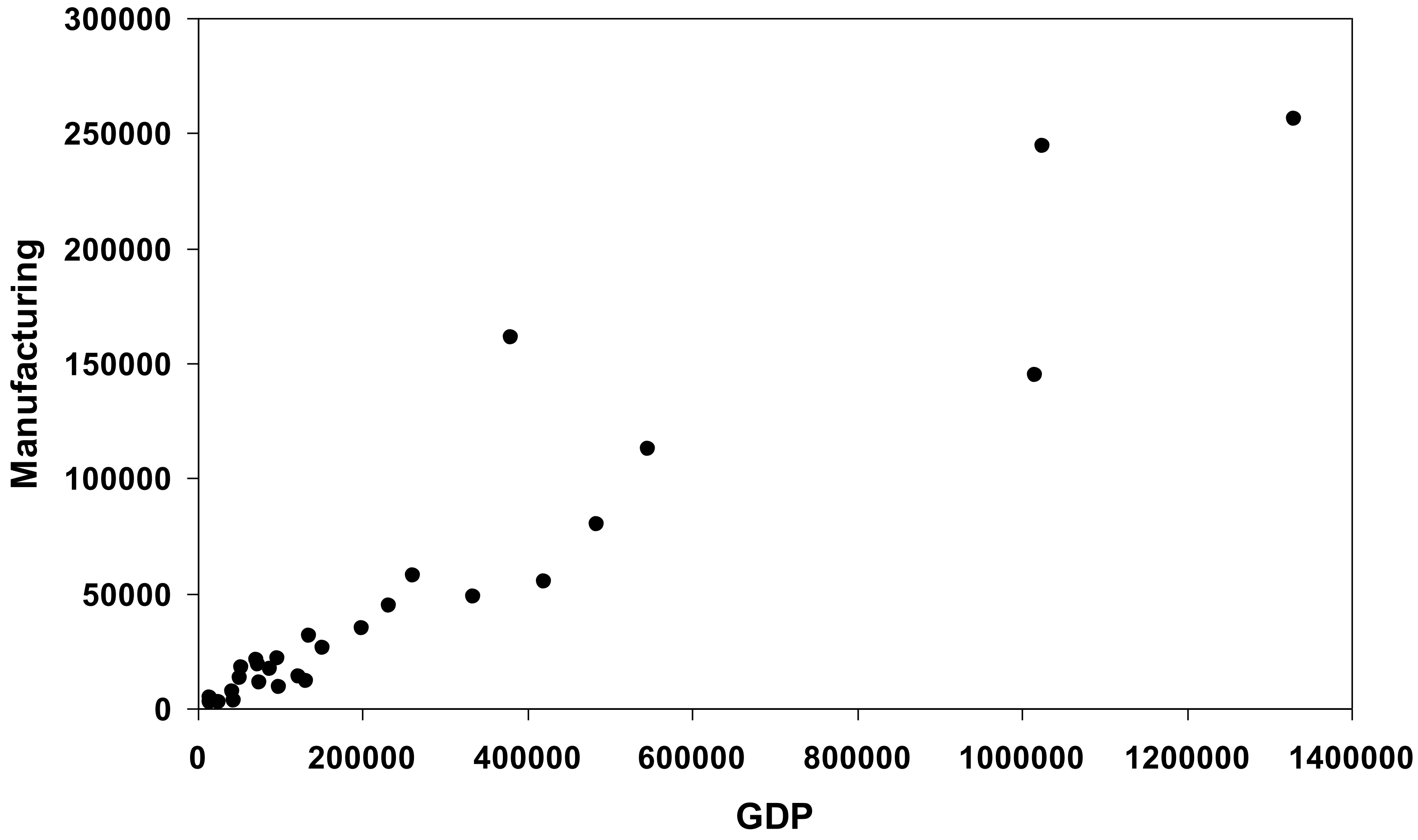


Consequences of heteroscedasticity:

1. Standard errors of the regression coefficients are estimated wrongly and the *t* tests (and *F* test) are invalid.

2. OLS estimators are inefficient (though still unbiased).

HETEROSCEDASTICITY: EXAMPLE



The regression of Manufacturing on GDP (28 large developed countries, without USA and Japan)

UNBIASEDNESS OF THE REGRESSION COEFFICIENTS:

NO NEED OF HOMOSCEDASTICITY

$$\hat{\beta}_2 = \beta_2 + \frac{\sum(X_i - \bar{X})(u_i - \bar{u})}{\sum(X_i - \bar{X})^2} = \beta_2 + \sum a_i u_i$$

$$\sum(X_i - \bar{X})(u_i - \bar{u}) = \sum(X_i - \bar{X})\, u_i$$

$$\bar{u}\sum(X_i - \bar{X}) = 0 \qquad \text{since} \qquad \sum(X_i - \bar{X}) = 0$$

$$\mathrm{E}(\hat{\beta}_2) = \mathrm{E}(\beta_2) + \mathrm{E}\left(\sum a_i u_i\right)$$

$$= \beta_2 + \sum \mathrm{E}(a_i u_i) = \beta_2 + \sum a_i \mathrm{E}(u_i)$$

HETEROSCEDASTICITY: PRECISION OF b_2 COEFFICIENT, SLR

Heteroscedasticity-consistent (Robust) standard errors

$$\sigma^2_{\hat{\beta}_2} = \mathrm{E}\left\{\left(\hat{\beta}_2 - \mathrm{E}(\hat{\beta}_2)\right)^2\right\} = \mathrm{E}\left\{\left(\hat{\beta}_2 - \beta_2\right)^2\right\} = \mathrm{E}\left\{\left(\sum_{i=1}^{n} a_i u_i\right)^2\right\} =$$

$$= \mathrm{E}\left\{\sum_{i=1}^{n} a_i^2 u_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} a_i a_j u_i u_j\right\} = \sum_{i=1}^{n} a_i^2 \mathrm{E}(u_i^2) + \sum_{i=1}^{n}\sum_{j\neq i} a_i a_j \mathrm{E}(u_i u_j) =$$

$$= \sum_{i=1}^{n} a_i^2 \sigma_i^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2\, \sigma_i^2}{\left(\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2\right)^2} = \frac{\sum_{i=1}^{n} {x_i}^2\, \sigma_i^2}{\left(\sum_{j=1}^{n} {x_j}^2\right)^2}$$

$$Hense\quad {s_{\hat{\beta}_2}}^2 = \frac{\sum_{i=1}^{n} {x_i}^2\, \hat{u}_i^2}{\left(\sum_{j=1}^{n} {x_j}^2\right)^2} = \sum_{i=1}^{n} {a_i}^2\, \hat{u}_i^2$$

The formula for $\hat{\beta}_2$ population variance differs from the standard one.

It can be greater or less, depending on σ_i^2 behaviour.

HETEROSCEDASTICITY DETECTION:

GENERAL CASE

General: White test for heteroscedasticity (White, 1980), for detection of any form of association between σ_i^2 and the regressors.

Since σ_i^2 are unobservable, $\hat{u}_i^{\,2}$ are used as proxies.

The White test consists of two steps:

1. Regressing the squared residuals on the explanatory variables in the model, their squares, and their cross-products, omitting any duplicative variables.
2. Test statistic nR^2 is calculated, using R^2 from this regression. Under the null hypothesis of no association (homoscedasticity), it is distributed as a chi-squared statistic with degrees of freedom equal to the number of regressors, including the constant, minus one, in large samples.

   F-test can be also applied. It is also a large-samples test here, and the results are usually the same.

So, after estimating the model $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$

$$\textit{the regression}$$

$$\hat{u}_i^{\,2} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i}^{\;2} + \beta_5 X_{3i}^{\;2} + \beta_6 X_{2i} X_{3i} + v_i$$

$$\textit{is estimated}$$

HETEROSCEDASTICITY DETECTION:

Breusch-Pagan Test

The Breusch-Pagan (or Breusch-Pagan-Godfrey) test for heteroscedasticity, for detection of between σ_i^2 and the regressors. Comparing to the White test, includes less parameters to estimate.

Since σ_i^2 is unobservable, $\hat{u}_i^{\,2}$ is used as a proxy.

The test consists of two steps:

1. Regressing the squared residuals on the explanatory variables in the model.
2. Test statistic nR^2 is calculated, using R^2 from this regression. Under the null hypothesis of no association (homoscedasticity), it is distributed as a chi-squared statistic with degrees of freedom equal to the number of regressors, including the constant, minus one, in large samples.

F-test can be also applied. It is also a large-samples test here, and the results are usually the same.

After estimating the model $$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

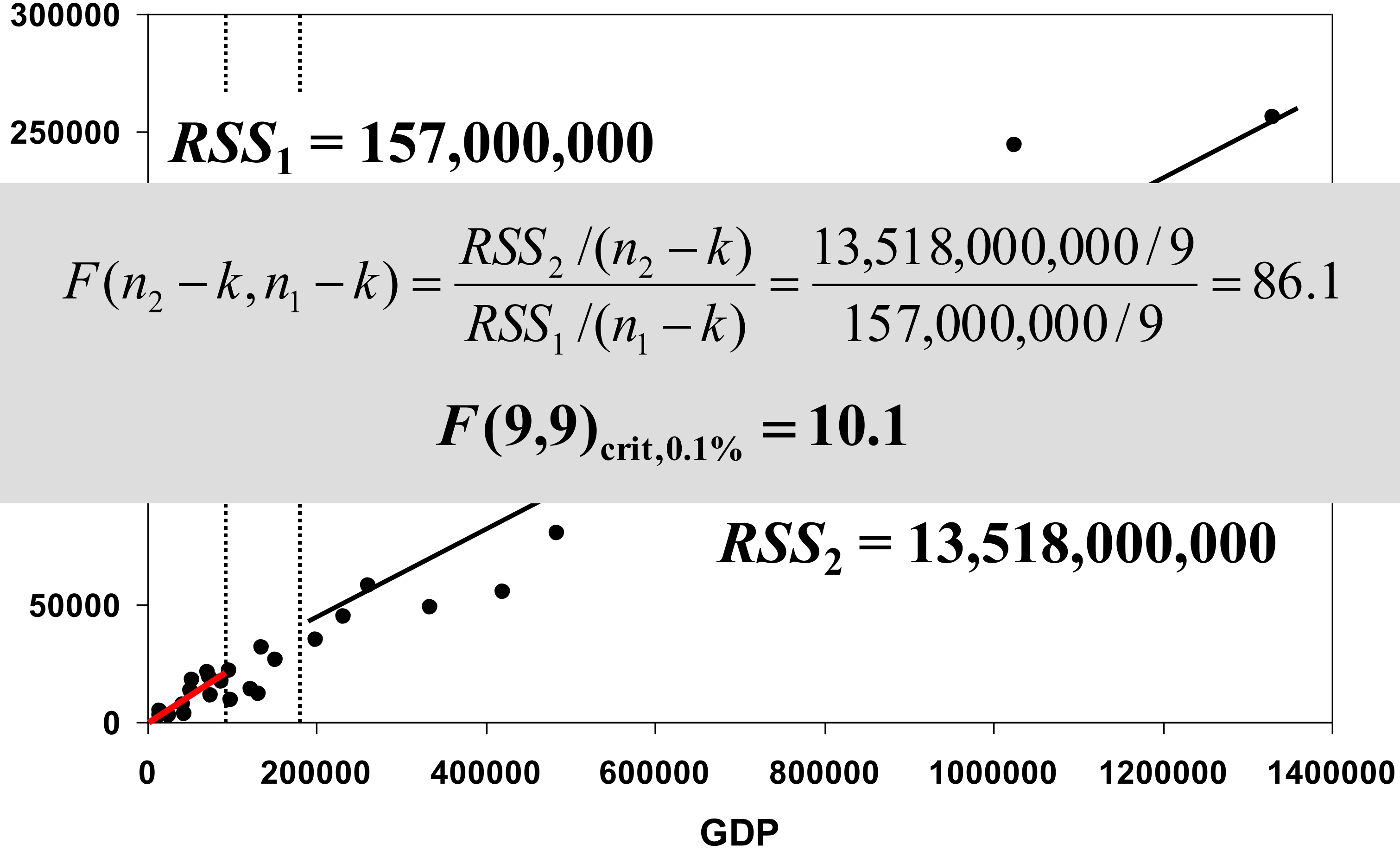
$$\textit{the regression}$$

$$\hat{u}_i^{\,2} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + v_i$$

$$\textit{is estimated}$$

HETEROSCEDASTICITY DETECTION: GOLDFELD–QUANDT TEST:

SPECIAL CASE WHEN $\sigma_i = \gamma X_i$



$$F(n_2 - k, n_1 - k) = \frac{RSS_2/(n_2 - k)}{RSS_1/(n_1 - k)} = \frac{13{,}518{,}000{,}000/9}{157{,}000{,}000/9} = 86.1$$

$$F(9,9)_{\text{crit},0.1\%} = 10.1$$

The Goldfeld–Quandt test is a test for this type of heteroscedasticity. The sample is divided into three ranges containing n_1 observations with the smallest values of the X and n_2 observations with the largest values, the rest in the middle. Then the model is estimated for two extreme subsamples, with RSS_1 and RSS_2. Normally $n_1/n = n_2/n = 3/8$ is recommended.

Here the null hypothesis of homoscedasticity is rejected at the 0.1% level.

WHITE AND BREUSCH-PAGAN TESTS EXAMPLE

$$LGEARN = \beta_1 + \beta_2 S + \beta_3 ASVABC + u$$

Dependent Variable: LGEARN Method: Least Squares Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.777056	0.132301	5.873384	0.0000
S	0.077404	0.010007	7.734779	0.0000
ASVABC	0.012379	0.002662	4.650030	0.0000

R-squared	0.227830	Mean dependent var	2.456463		
S.D. dependent var	0.541347	S.E. of regression	0.476537		
Sum squared resid	128.7586	F-statistic	83.64712	Durbin-Watson stat	1.728273

$$RES1_i^2 = \beta_1 + \beta_2 S_i + \beta_3 ASVABC_i + \beta_4 S_i^2 + \beta_5 ASVABC_i^2 + \beta_6 S_i ASVABC_i + v_i$$

$$RES1_i^2 = \beta_1 + \beta_2 S_i + \beta_3 ASVABC_i + v_i$$

We will apply the White test in the Earnings Function (EAEF 40). Assume that in the true model *LGEARN* depends only on *S* and *ASVABC*.

EViews: View→Residual Diagnostics → Heteroscedasticity Tests → White (with or without cross terms); or Breusch-Pagan-Godfrey

WHITE TEST EXAMPLE

$$RES1_i^2 = \beta_1 + \beta_2 S_i + \beta_3 ASVABC_i + \beta_4 S_i^2 + \beta_5 ASVABC_i^2 + \beta_6 S_i ASVABC_i + v_i$$

Dependent Variable: RES2 Method: Least Squares Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.504124	0.608615	2.471388	0.0138
S	-0.088943	0.074242	-1.198013	0.2314
ASVABC	-0.033729	0.020207	-1.669187	0.0956
S2	0.002982	0.003686	0.808916	0.4189
ASVABC2	0.000385	0.000244	1.573253	0.1162
SASV	8.13E-05	0.001411	0.057587	0.9541

R-squared	0.019295	Mean dependent var	0.225892
S.D. dependent var	0.404860	S.E. of regression	0.402709
Sum squared resid	91.46633	F-statistic	2.219325
Durbin-Watson stat	1.846331	Prob(F-statistic)	0.051024

$nR^2=570*0.019295=11.0$ $chi^2_{crit,5\%}(5)=11.07$

The null-hypothesis of homoscedasticity is not rejected here at the 5% level, but marginally.
The same conclusion would be done using F-statistic.

BREUSCH-PAGAN TEST EXAMPLE

$$RES1_i^2 = \beta_1 + \beta_2 S_i + \beta_3 ASVABC_i + v_i$$

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	2.076642	Prob. F(2,567)	0.1263
Obs*R-squared	4.144898	Prob. Chi-Square(2)	0.1259
Scaled explained SS	6.575749	Prob. Chi-Square(2)	0.0373

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/12/22 Time: 17:31
Sample: 1 570
Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005250	0.112189	0.046795	0.9627
S	0.005146	0.008486	0.606448	0.5445
ASVABC	0.002987	0.002257	1.323162	0.1863

R-squared	0.007272	Mean dependent var	0.225892
Adjusted R-squared	0.003770	S.D. dependent var	0.404860
S.E. of regression	0.404097	Akaike info criterion	1.030923
Sum squared resid	92.58771	Schwarz criterion	1.053795
Log likelihood	-290.8132	Hannan-Quinn criter.	1.039847
F-statistic	2.076642	Durbin-Watson stat	1.805433
Prob(F-statistic)	0.126303		

The null-hypothesis of homoscedasticity is not rejected here at the 5% level.
The same conclusion would be done using F-statistic.

HETEROSCEDASTICITY: REMEDIAL MEASURES.

Generalised Least Squares (GLS). Weighted Regressions.

GLS is OLS on the transformed variables that satisfy the standard LS assumptions.
GLS estimators are BLUE.

WLS (Weighted Least Squares) is a special case of GLS for heteroscedasticity case.

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad \Rightarrow \quad \frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

$\mathrm{Var}(u_i) = \sigma_i^2$ **, not constant for all *i***

$$\mathrm{Var}\left\{\frac{u_i}{\sigma_i}\right\} = \frac{1}{\sigma_i^2}\mathrm{Var}(u_i) = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

$$Y' = \beta_1 H + \beta_2 X' + u'$$

$$Y' = \frac{Y_i}{\sigma_i}, \quad H = \frac{1}{\sigma_i}, \quad X' = \frac{X_i}{\sigma_i}, \quad u' = \frac{u_i}{\sigma_i}$$

In the revised model, we regress *Y'* on *X'* and *H*, as defined. Note that there is no intercept in the revised model. β_1 becomes the slope coefficient of the variable $1/\sigma_i$.

HETEROSCEDASTICITY: WEIGHTED AND LOGARITHMIC REGRESSIONS

$$\boldsymbol{Y_i = \beta_1 + \beta_2 X_i + u_i}$$

$\mathrm{var}(u_i) = \sigma_i^2$ **, not constant for all** $\boldsymbol{i}$

$$\sigma_i \; unknown, but \; possibly \; \sigma_i = \lambda Z_i$$

$$\boldsymbol{\frac{Y_i}{Z_i} = \beta_1 \frac{1}{Z_i} + \beta_2 \frac{X_i}{Z_i} + \frac{u_i}{Z_i}}$$

$$\mathrm{var}\left\{\frac{u_i}{Z_i}\right\} = \frac{1}{Z_i^2}\sigma_i^2 = \frac{\sigma_i^2}{\sigma_i^2 / \lambda^2} = \lambda^2$$

$$\boldsymbol{Y' = \beta_1 H + \beta_2 X' + u' \qquad Y' = \frac{Y_i}{Z_i}, \;\; H = \frac{1}{Z_i}, \;\; X' = \frac{X_i}{Z_i}, \;\; u' = \frac{u_i}{Z_i}}$$

The disturbance term in the revised model has constant variance λ^2.

We do not need to know the value of λ^2.

HETEROSCEDASTICITY-CONSISTENT STANDARD ERRORS: EViews Estimation

$$LGEARN = \beta_1 + \beta_2 S + \beta_3 ASVABC + u$$

Dependent Variable: LGEARN Method: Least Squares Included observations: 570
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.777056	0.141987	5.472718	0.0000
S	0.077404	0.010639	7.275155	0.0000
ASVABC	0.012379	0.002689	4.602784	0.0000

R-squared	0.227830	Mean dependent var	2.456463
S.D. dependent var	0.541347	S.E. of regression	0.476537
Sum squared resid	128.7586	F-statistic	83.64712

The standard errors are greater but asymptotically valid

$$\text{Standard errors}: \; s_{\hat{\beta}_2} = \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2} \hat{u}_i^2}{\left(\sum_{j=1}^{n} x_j^{\,2}\right)^2}} = \sqrt{\sum_{i=1}^{n} a_i^{\,2} \hat{u}_i^2}$$

EViews: Proc → Specify/Estimate→ LS (specify equation) → Options →

Coefficient Covariance Matrix: White.

Weighted Least Squares: EViews Estimation

$$LGEARN = \beta_1 + \beta_2 S + \beta_3 ASVABC + u$$

Dependent Variable: LGEARN/S (weights are 1/S):

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.073486	0.010766	6.825850	0.0000
1/S	0.918090	0.127407	7.205937	0.0000
ASVABC/S	0.010606	0.002558	4.145702	0.0000

R-squared	0.159513	Mean dependent var	0.182388

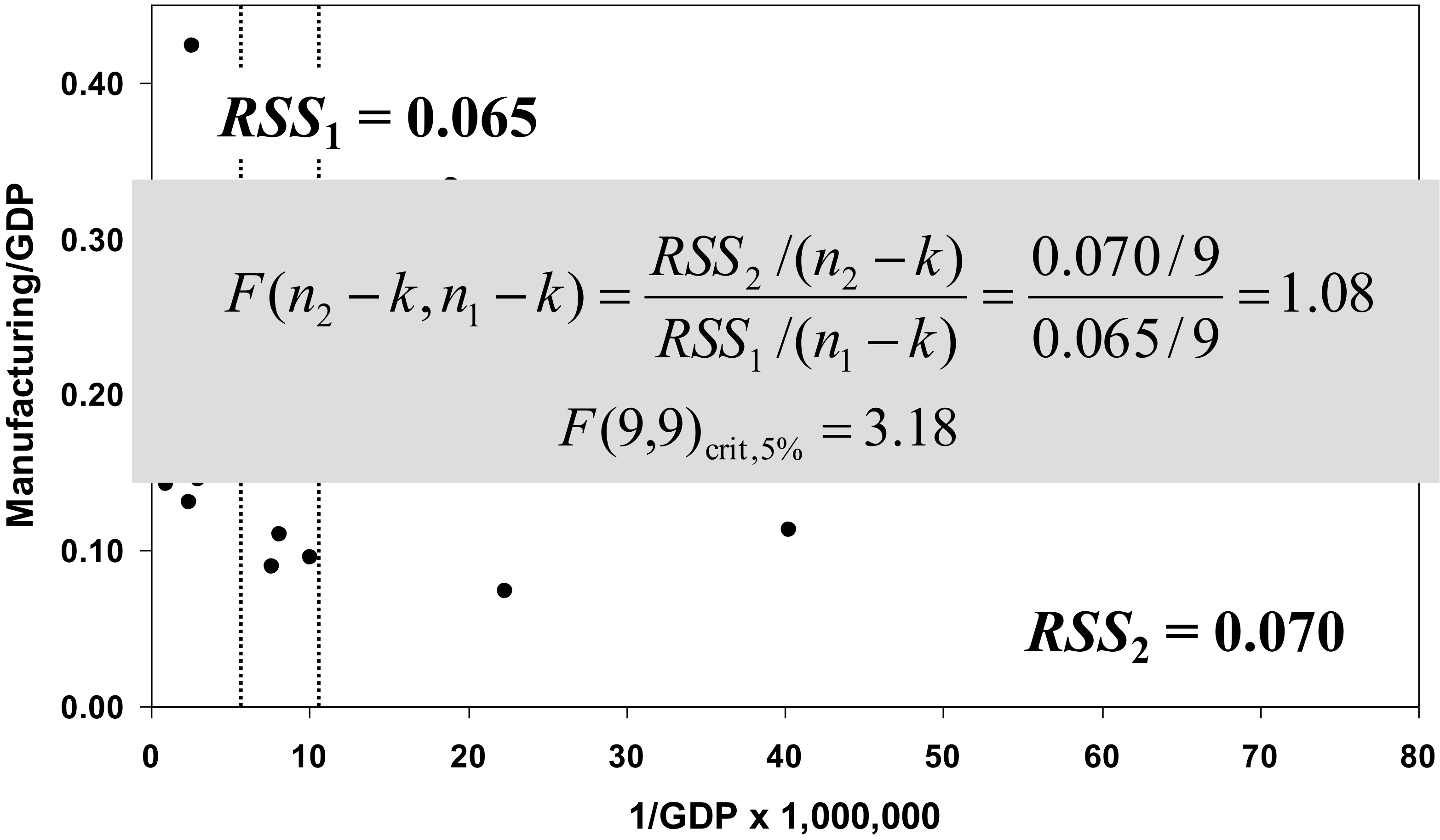
Heteroskedasticity Test: White

F-statistic	5.484209	Prob. F(4,565)	0.0002
Obs*R-squared	21.30382	Prob. Chi-Square(4)	0.0003

Heteroscedasticity has not been removed. The same result if the weights are 1/ASVABC. So the heteroscedasticity was of some other type.

EViews: Proc → Specify/Estimate – LS (specify equation)→Options →Weights: Type – inverse standard deviation; weighting series: 1/S.

HETEROSCEDASTICITY: WEIGHTED REGRESSION, EXAMPLE

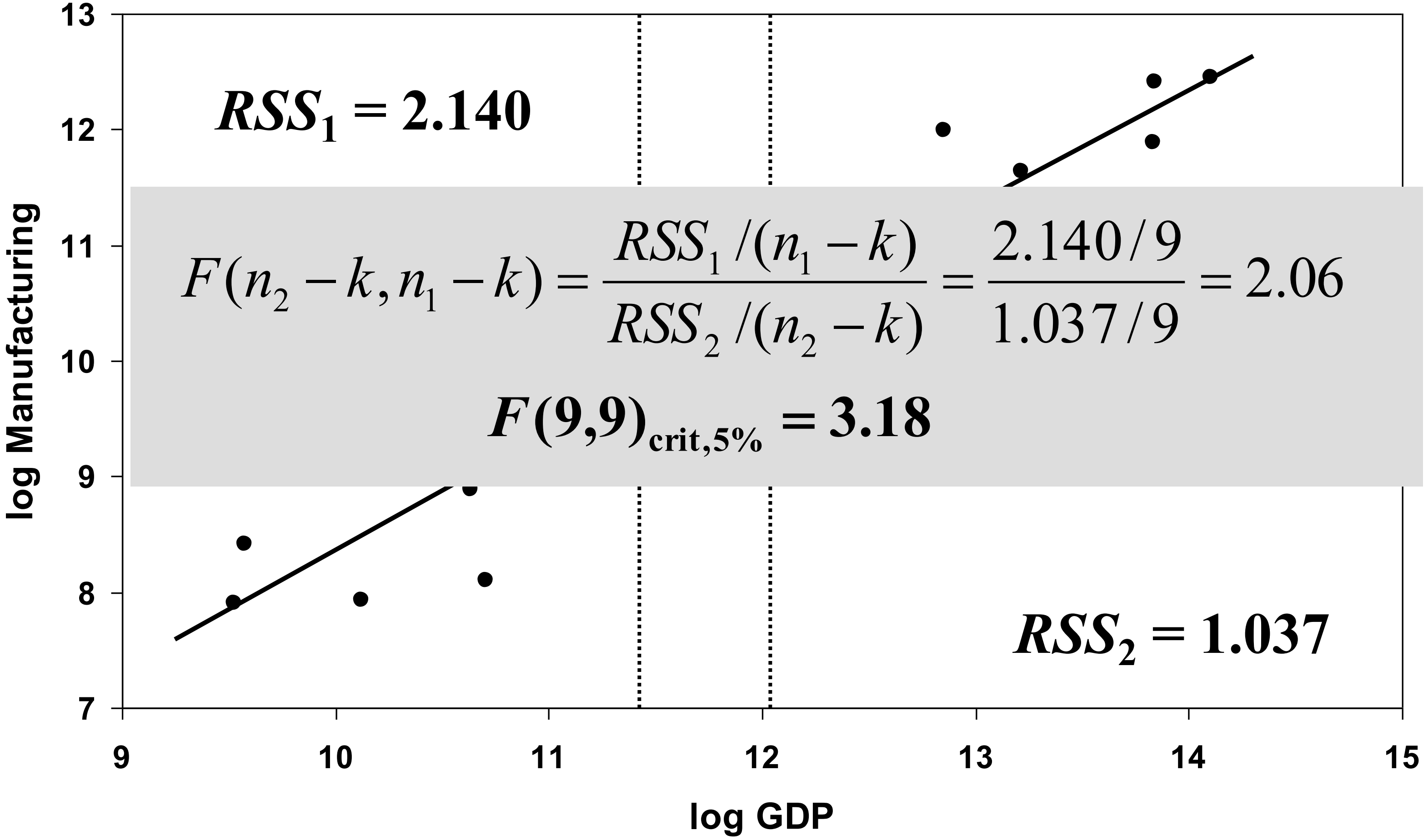


$$F(n_2 - k, n_1 - k) = \frac{RSS_2/(n_2 - k)}{RSS_1/(n_1 - k)} = \frac{0.070/9}{0.065/9} = 1.08$$

$$F(9,9)_{\text{crit},5\%} = 3.18$$

In MANU – GDP model Goldfeld-Quandt test found heteroscedasticity of the type $\sigma_i = \gamma GDP_i$. After dividing by GDP, the F statistic is not significant. The heteroscedasticity has been eliminated.

HETEROSCEDASTICITY: DOUBLE LOGARITHMIC REGRESSION



$$F(n_2 - k, n_1 - k) = \frac{RSS_1/(n_1 - k)}{RSS_2/(n_2 - k)} = \frac{2.140/9}{1.037/9} = 2.06$$

$$F(9,9)_{crit,5\%} = 3.18$$

Heteroscedasticity can be also removed in logarithmic model.

The null hypothesis of homoscedasticity is not rejected.

HETEROSCEDASTICITY: WEIGHTED AND LOGARITHMIC REGRESSIONS

$$\underset{}{M\hat{A}NU} = \underset{(5700)}{604} + \underset{(0.013)}{0.194}GDP \qquad R^2 = 0.89$$

$$\frac{M\hat{A}NU}{POP} = \underset{(1371)}{612}\frac{1}{POP} + \underset{(0.016)}{0.182}\frac{GDP}{POP} \qquad R^2 = 0.70$$

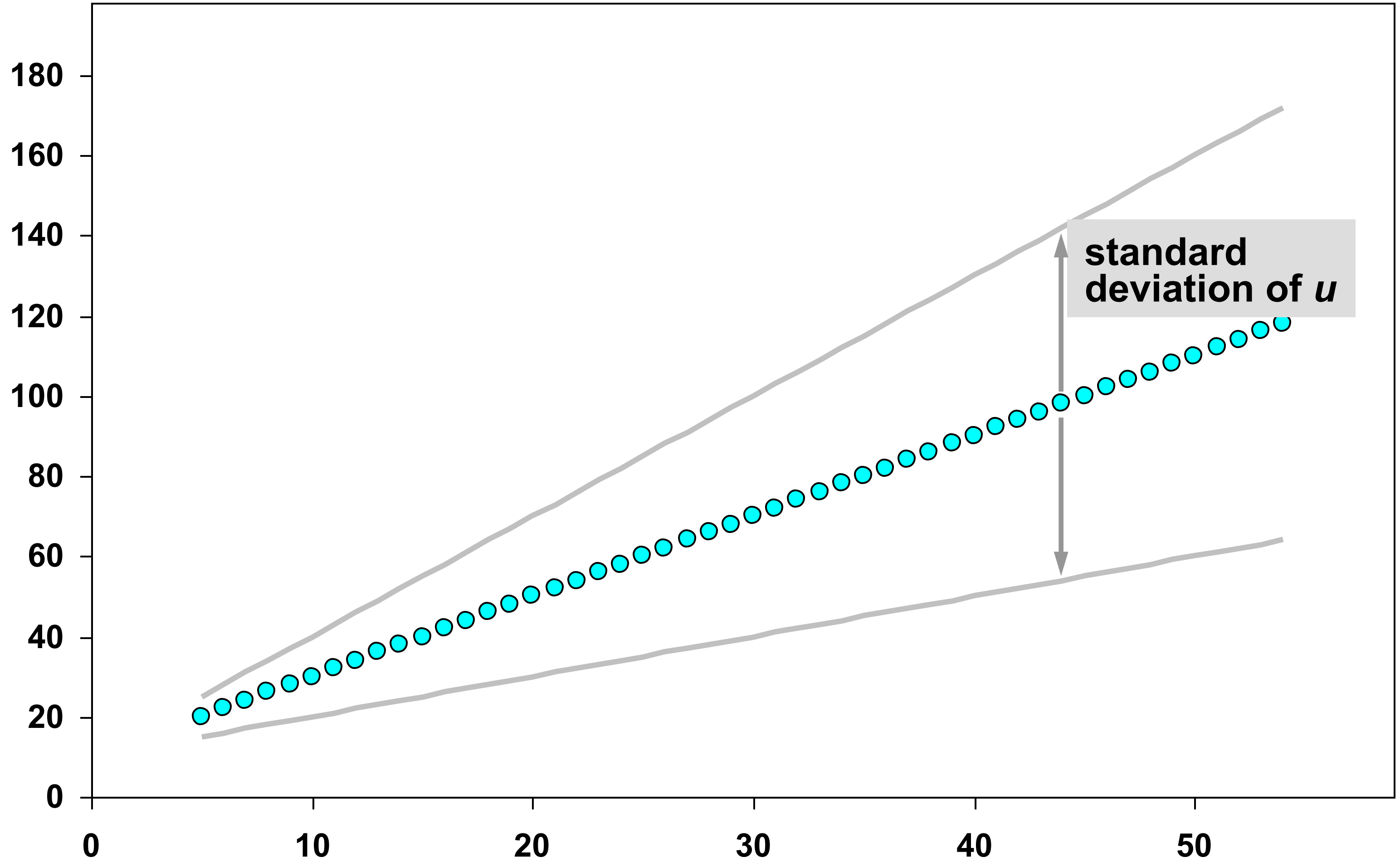
$$\frac{M\hat{A}NU}{GDP} = \underset{(0.019)}{0.189} + \underset{(841)}{533}\frac{1}{GDP} \qquad R^2 = 0.02$$

$$\log M\hat{A}NU = \underset{(0.785)}{-1.694} + \underset{(0.066)}{0.999}\log GDP \qquad R^2 = 0.90$$

Here is a summary of the regressions using the four alternative specifications of the model.

HETEROSCEDASTICITY: MONTE CARLO ILLUSTRATION

$$Y_i = 10 + 2.0X_i + u_i \quad X_i = \{5,6, ..., 54\} \quad u_i = X_i\varepsilon_i \quad \varepsilon_i \sim N(0,1)$$

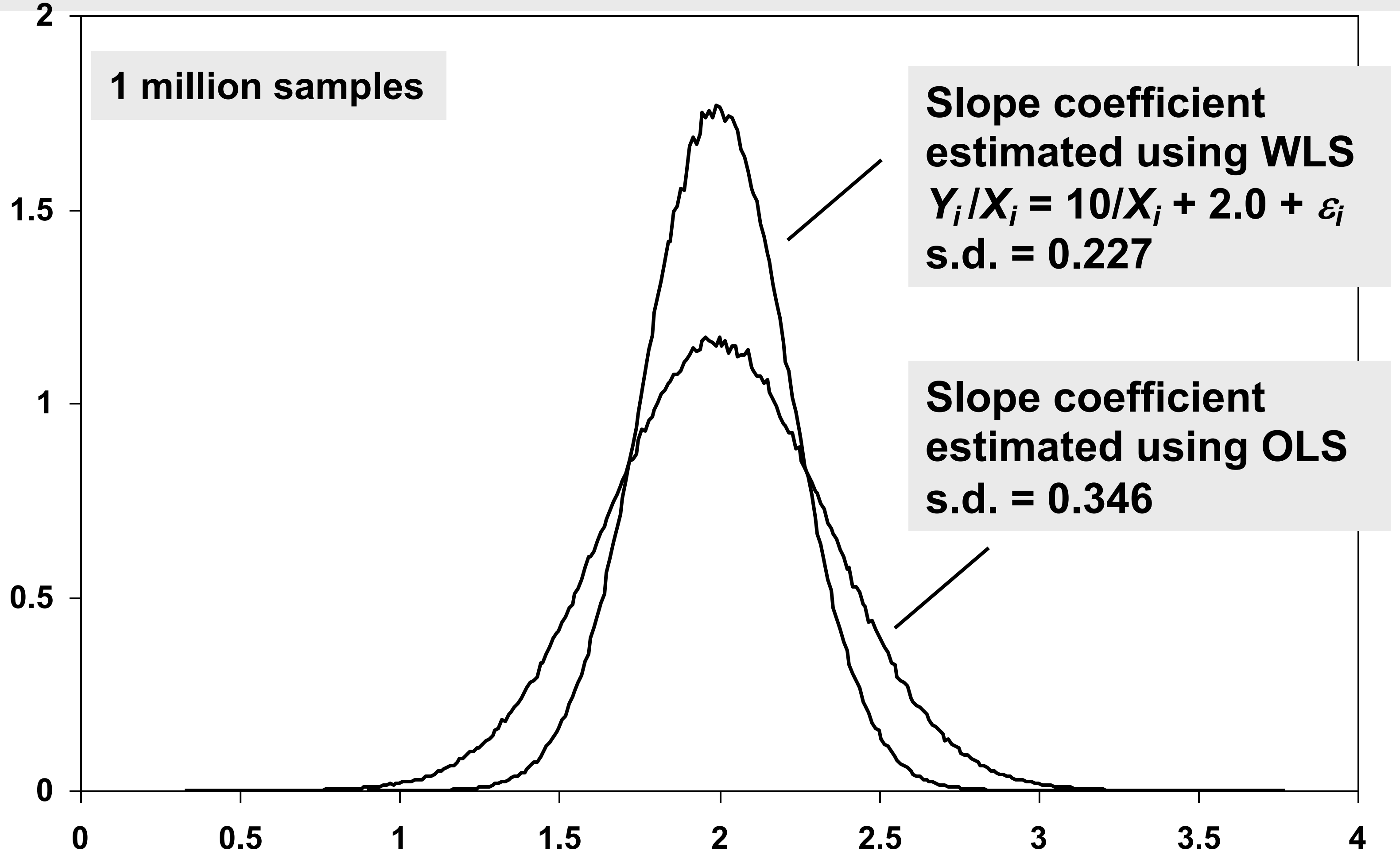


Monte Carlo simulation: $Y = 10 + 2X$, $X = 5, \ldots, 54$, $u_i = X\varepsilon_i$,

where ε_i is iid N(0,1)

HETEROSCEDASTICITY: MONTE CARLO ILLUSTRATION

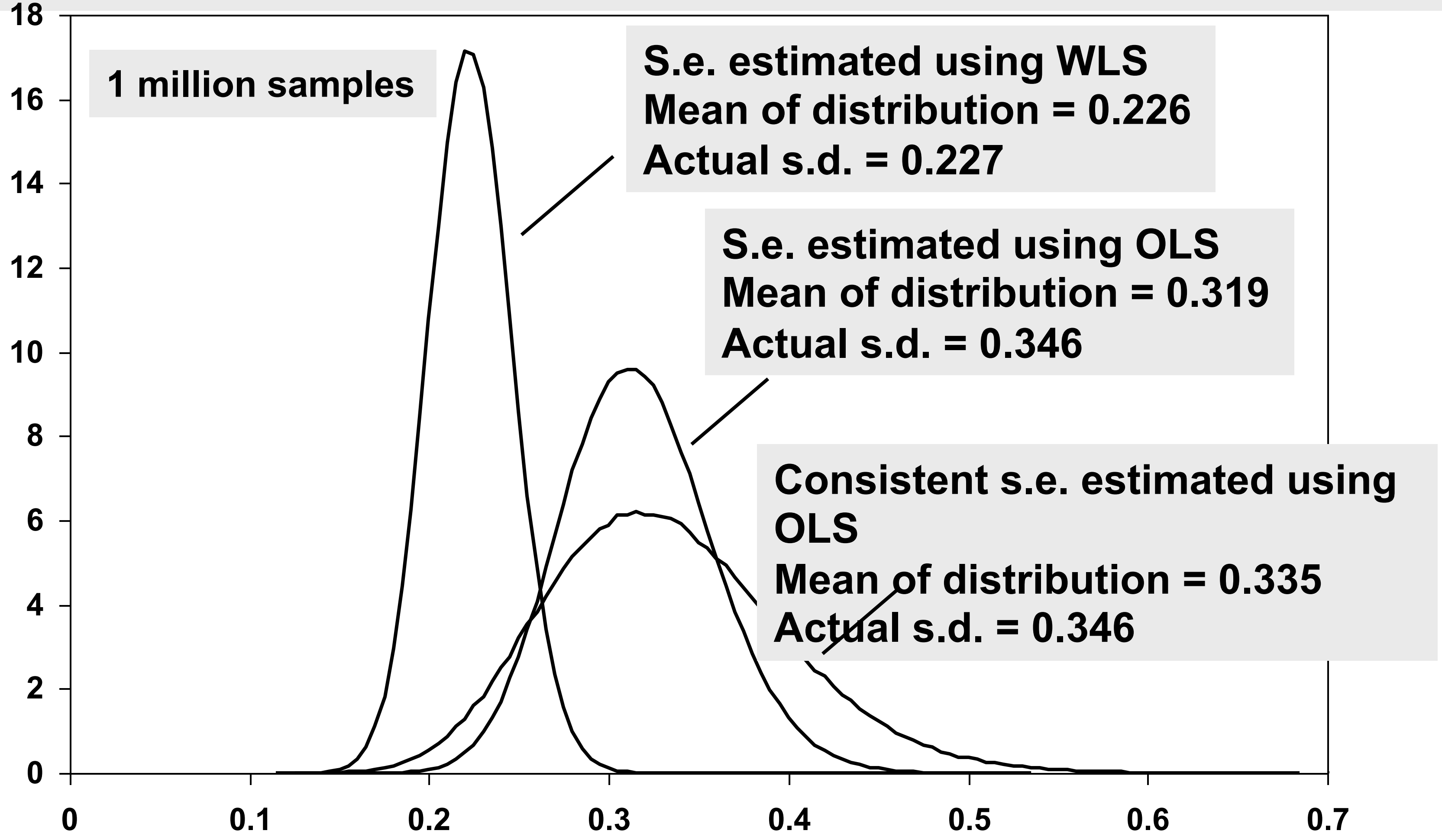
$$Y_i = 10 + 2.0X_i + u_i \quad X_i = \{5,6, ..., 54\} \quad u_i = X_i\varepsilon_i \quad \varepsilon_i \sim N(0,1)$$



Both OLS and WLS are unbiased, but the WLS estimates distribution has smaller variance. This illustrates OLS inefficiency with heteroscedasticity.

HETEROSCEDASTICITY: MONTE CARLO ILLUSTRATION

$$Y_i = 10 + 2.0X_i + u_i \quad X_i = \{5,6, ..., 54\} \quad u_i = X_i\varepsilon_i \quad \varepsilon_i \sim N(0,1)$$



Consistent standard errors are valid only in large samples. For small samples, their properties are unknown. They can mislead even more than the OLS standard errors.