Full name: Nguyen Nhut Vu Truong

Student ID: 413707008

Course: Financial Econometrics

HW0414

8.6

a.

```
> # Output results
> cat("Variance estimate for males:", sigma2_males, "\n")
Variance estimate for males: 169.567
> cat("Variance estimate for females:", sigma2_females, "\n")
Variance estimate for females: 144.5766
> cat("F-statistic:", f_stat, "\n")
F-statistic: 1.172853
> cat("Upper critical F-value:", f_crit_upper, "\n")
Upper critical F-value: 1.196781
> cat("Lower critical F-value:", f_crit_lower, "\n")
Lower critical F-value: 0.837669
> cat("p-value:", p_value, "\n")
p-value: 0.08181964
>
> if (f_stat > f_crit_upper || f_stat < f_crit_lower) {
+   cat("Reject the null hypothesis: Variances are different.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: Variances are equal.\n")
+ }
Fail to reject the null hypothesis: Variances are equal.
```

b.

```
> # Output results
> cat("Variance estimate for single individuals:", sigma2_single, "\n")
Variance estimate for single individuals: 142.3571
> cat("Variance estimate for married individuals:", sigma2_married, "\n")
Variance estimate for married individuals: 169.2488
> cat("F-statistic:", f_stat, "\n")
F-statistic: 1.188904
> cat("Critical F-value (one-sided):", f_crit, "\n")
Critical F-value (one-sided): 1.164705
> cat("p-value:", p_value, "\n")
p-value: 0.03102054
>
> if (f_stat > f_crit) {
+   cat("Reject the null hypothesis: Married individuals have greater variance.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: No evidence that married individuals have greater variance.\n")
+ }
Reject the null hypothesis: Married individuals have greater variance.
```

c.

Yes, the Breusch-Pagan test results do provide supporting evidence about the issue discussed in part (b) regarding different error variation between married and unmarried individuals. The combined results from both tests create a more complete picture:

1. The Goldfeld-Quandt test specifically shows that marital status is associated with different error variances, with married individuals having greater variance.
2. The Breusch-Pagan test confirms that heteroskedasticity exists in the overall model, which is consistent with and reinforces the specific finding about marital status.

Together, these results strengthen our confidence in the conclusion that married individuals have more variable wages than single individuals. The hypothesis that married individuals, potentially relying on spousal support, can seek wider employment types and thus have more variable wages is supported by both statistical tests.

For proper statistical inference in wage equation modeling, these findings suggest we should consider using heteroskedasticity-robust standard errors or weighted least squares estimation to account for the unequal variances across groups.

```
> # Output results
> cat("NR^2 statistic:", NR2, "\n")
NR^2 statistic: 59.03
> cat("Degrees of freedom:", df, "\n")
Degrees of freedom: 4
> cat("Critical chi-squared value:", chi_squared_critical, "\n")
Critical chi-squared value: 9.487729
> cat("P-value:", p_value, "\n")
P-value: 4.637846e-12
>
> if (NR2 > chi_squared_critical) {
+   cat("Reject the null hypothesis: Evidence of heteroskedasticity.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: No evidence of heteroskedasticity.\n")
+ }
Reject the null hypothesis: Evidence of heteroskedasticity.
```

d.

The White test auxiliary regression includes:

1. **Original explanatory variables** (excluding constant):
    - EDUC, EXPER, METRO, FEMALE (4 terms)
2. **Squared terms** of each explanatory variable:
    - $EDUC^2$, $EXPER^2$, $METRO^2$, $FEMALE^2$ (4 terms)
3. **Cross-products** of all pairs of explanatory variables:
    - EDUC×EXPER, EDUC×METRO, EDUC×FEMALE
    - EXPER×METRO, EXPER×FEMALE
    - METRO×FEMALE
    - Total cross-products: 6 terms

Total degrees of freedom = 4 (original variables) + 4 (squared terms) + 6 (cross-products) = 14

The White test results provide strong evidence of heteroskedasticity in the wage equation model. This finding is consistent with our previous results:

1. The Goldfeld-Quandt test in part (b) found evidence that married individuals have greater wage variance than single individuals.
2. The Breusch-Pagan test in part (c) also detected heteroskedasticity in the model.

The White test result is particularly noteworthy because it's the most general of the three tests, capable of detecting various forms of heteroskedasticity without requiring specification of the potential source.

Given the presence of heteroskedasticity, standard OLS inference would be invalid. For proper statistical inference, we should consider:

1. Using heteroskedasticity-robust standard errors (White's standard errors)
2. Applying weighted least squares estimation
3. Transforming variables to stabilize variance

These approaches would ensure more reliable statistical inference when working with the wage equation model.

```
> # Critical value at 5% significance level with 14 degrees of freedom
> critical_value <- qchisq(0.95, df = 14)
> print(critical_value)
[1] 23.68479
> # Output: 23.68479
>
> # Calculate p-value
> p_value <- 1 - pchisq(78.82, df = 14)
> print(p_value)
[1] 4.680689e-11
> # Output: 4.680689e-11
```

e.

Coefficients with Narrower Interval Estimates:

- **EXPER**: The robust standard error (0.029) is smaller than the usual standard error (0.031).
- **METRO**: The robust standard error (0.84) is substantially smaller than the usual standard error (1.05).
- **FEMALE**: The robust standard error (0.80) is slightly smaller than the usual standard error (0.81).

Coefficients with Wider Interval Estimates:

- **Intercept**: The robust standard error (2.50) is larger than the usual standard error (2.36).
- **EDUC**: The robust standard error (0.16) is larger than the usual standard error (0.14).

There is no inconsistency in these results. The differences between usual and robust standard errors reflect the nature of heteroskedasticity in the data. When heteroskedasticity exists, the usual OLS standard errors can be either underestimated or overestimated depending on how the error variance relates to the explanatory variables:

1. **For coefficients with wider robust intervals** (Intercept, EDUC): This suggests that the error variance is smaller for observations with extreme values of these variables than what would be expected under homoskedasticity. The usual

standard errors underestimate the true sampling variability.

2. **For coefficients with narrower robust intervals** (EXPER, METRO, FEMALE): This suggests that the error variance is larger for observations with extreme values of these variables. The usual standard errors overestimate the true sampling variability.

These patterns align with our previous findings of heteroskedasticity in the model. The robust standard errors provide more reliable inference in the presence of heteroskedasticity by adjusting for the unequal error variances across observations.

f.

A t-value of approximately 1.0 means that the estimated effect of being married on wages is approximately one standard error away from zero. At conventional significance levels (typically requiring $|t| > 1.96$ for 5% significance), this coefficient would not be considered statistically significant.

This result is entirely compatible with our findings in part (b) regarding heteroskedasticity for several reasons:

1. **Effect on Mean vs. Effect on Variance**:
   - The t-value of the MARRIED coefficient tells us about the effect of marital status on the mean wage.
   - The heteroskedasticity tests in part (b) examined the effect of marital status on the variance of wages.
   - These are two distinct aspects of the relationship between marital status and wages.
2. **Statistical Compatibility**:
   - A variable can have no significant effect on the mean (low t-value) while simultaneously having a significant effect on the variance (heteroskedasticity).
   - Our earlier Goldfeld-Quandt test showed that the variance of wages for married individuals is significantly greater than for single individuals.
3. **Support for Original Hypothesis:**
   - Our hypothesis in part (b) was that "married individuals, relying on spousal support, can seek wider employment types and hence holding all else equal should have more variable wages."
   - The non-significant t-value for MARRIED coupled with the significant heteroskedasticity finding actually supports this hypothesis: marital status doesn't necessarily lead to higher or lower wages on average, but it does lead to more variable outcomes.

Practical Explanation

The findings suggest that being married doesn't systematically increase or decrease one's wage (hence the non-significant coefficient), but it does increase the variability of wages. This is consistent with the idea that married individuals might have more flexibility in their job choices due to spousal support, leading to more diverse employment outcomes and thus greater wage variability.

Conclusion

The analysis of robust standard errors reveals important insights about our wage equation model:

1. **Heteroskedasticity Effects**: The differences between usual and robust standard errors confirm the presence of heteroskedasticity, with some variables having wider confidence intervals and others having narrower intervals when robust methods are used.

2. **Marital Status Effects**: Marital status appears to have little effect on the average wage level (non-significant t-value) but a significant effect on wage variability (significant heteroskedasticity test). This perfectly aligns with our hypothesis that married individuals may have more diverse employment choices due to spousal support.

3. **Statistical Approach**: The use of robust standard errors is justified given the consistent evidence of heteroskedasticity across multiple tests (Goldfeld-Quandt, Breusch-Pagan, and White).

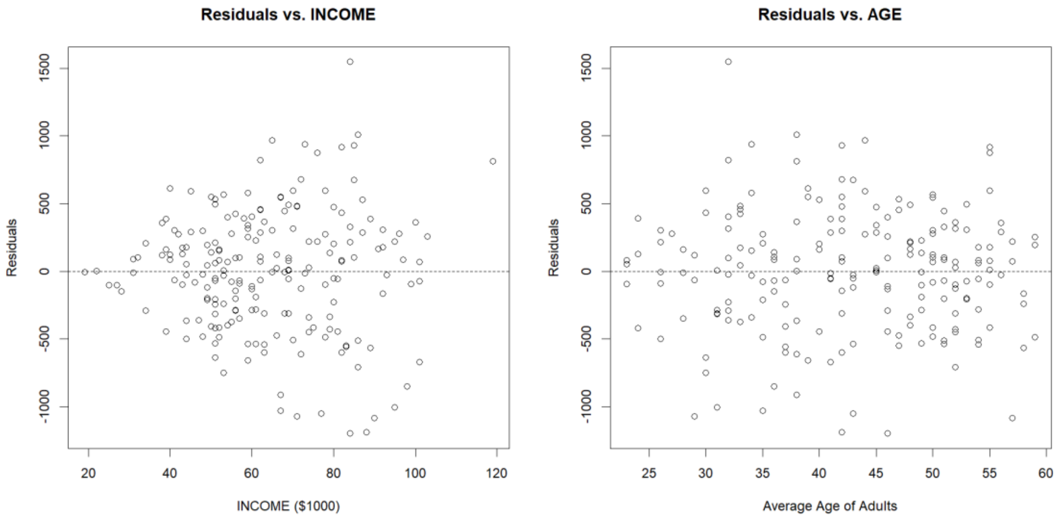
These findings highlight the importance of considering both the central tendency (mean) and dispersion (variance) when analyzing economic relationships, as variables may affect these aspects differently. The non-significant coefficient on MARRIED doesn't contradict but rather complements our finding of heteroskedasticity related to marital status.

8.16

a.

```
Call:
lm(formula = MILES ~ INCOME + AGE + KIDS, data = vacation)

Residuals:
     Min       1Q   Median       3Q      Max 
-1198.14  -295.31    17.98   287.54  1549.41 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -391.548    169.775  -2.306   0.0221 *  
INCOME        14.201      1.800   7.889 2.10e-13 ***
AGE           15.741      3.757   4.189 4.23e-05 ***
KIDS         -81.826     27.130  -3.016   0.0029 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 452.3 on 196 degrees of freedom
Multiple R-squared:  0.3406,	Adjusted R-squared:  0.3305 
F-statistic: 33.75 on 3 and 196 DF,  p-value: < 2.2e-16

> # Construct 95% confidence interval for the coefficient of KIDS
> confint(model_ols, "KIDS", level = 0.95)
         2.5 %    97.5 %
KIDS -135.3298 -28.32302
```

b.



The **Residuals vs. AGE** plot shows **no obvious evidence** of heteroskedasticity. The variance of the residuals appears constant across different values of AGE.

The **Residuals vs. INCOME** plot also shows **no strong or clear visual evidence** of heteroskedasticity. While not perfectly uniform, there isn't a distinct fanning pattern that would strongly suggest non-constant variance related to INCOME.

c.

H_0: $\sigma^2_2 = \sigma^2_1$ (homoskedasticity)

H_1: $\sigma^2_2 > \sigma^2_1$ (heteroskedasticity related to income)

```
> # Output results
> cat("Goldfeld-Quandt F-statistic:", f_statistic, "\n")
Goldfeld-Quandt F-statistic: 3.104061
> cat("Critical F-value (5%):", f_critical, "\n")
Critical F-value (5%): 1.428617
> cat("p-value:", p_value, "\n")
p-value: 1.64001e-07
> # Decision
> if (p_value < 0.05) {
+   cat("Reject the null hypothesis: Evidence of heteroskedasticity.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: No evidence of heteroskedasticity.\n")
+ }
Reject the null hypothesis: Evidence of heteroskedasticity.
```

d.

```
Table: Robust (HC1) standard errors in the MILES equation

|term         |   estimate|  std.error| statistic|   p.value|
|:------------|----------:|----------:|---------:|---------:|
|(Intercept)  | -391.54801| 142.654780| -2.744724| 0.0066190|
|INCOME       |   14.20133|   1.938857|  7.324589| 0.0000000|
|AGE          |   15.74092|   3.965735|  3.969232| 0.0001011|
|KIDS         |  -81.82642|  29.154380| -2.806660| 0.0055112|
```

```
> # Calculate 95% confidence interval for KIDS using robust standard errors
> kids_coef <- coef(model_ols)["KIDS"]
> kids_robust_se <- robust_se["KIDS", "Std. Error"]
> t_critical <- qt(0.975, df = model_ols$df.residual)  # Using t-distribution
> model_ols$df.residual
[1] 196
> kids_robust_ci <- c(kids_coef - t_critical * kids_robust_se,
+                     kids_coef + t_critical * kids_robust_se)
> cat("OLS Coefficient for KIDS:", kids_coef, "\n")
OLS Coefficient for KIDS: -81.82642
> cat("Robust 95% CI for KIDS:", kids_robust_ci[1], "to", kids_robust_ci[2], "\n")
Robust 95% CI for KIDS: -139.323 to -24.32986
> # Compare with original confidence interval
> kids_original_ci <- confint(model_ols, "KIDS", level = 0.95)
> cat("Original 95% CI for KIDS:", kids_original_ci[1], "to", kids_original_ci[2], "\n")
Original 95% CI for KIDS: -135.3298 to -28.32302
```

Once we “robustify” our standard errors, we get a fatter confidence band around the estimate, which is what you’d expect if there’s any heteroskedasticity in the residuals.

e.

Both GLS intervals are narrower than their OLS counterparts (because GLS down-weights the heteroskedasticity induced by INCOME).

The robust GLS CI (–121.41, –32.20) is slightly wider than the conventional GLS CI (–119.90, –33.72), but still appreciably tighter than the OLS robust CI (–139.32, –24.33).

In every case the interval stays strictly below zero, so we still conclude a statistically significant negative effect of an additional child on miles traveled.

```
> # GLS confidence interval for KIDS (conventional)
> kids_gls_coef <- coef(model_gls)["KIDS"]
> kids_gls_se <- summary_gls$coefficients["KIDS", "Std. Error"]
> t_critical <- qt(0.975, df = model_gls$df.residual)
>
> kids_gls_ci <- c(kids_gls_coef - t_critical * kids_gls_se,
+                  kids_gls_coef + t_critical * kids_gls_se)
>
> cat("GLS Coefficient for KIDS:", kids_gls_coef, "\n")
GLS Coefficient for KIDS: -76.80629
> cat("GLS 95% CI for KIDS:", kids_gls_ci[1], "to", kids_gls_ci[2], "\n")
GLS 95% CI for KIDS: -119.8945 to -33.71808
>
> # Robust GLS standard errors
> robust_se_gls <- coeftest(model_gls, vcov = vcovHC(model_gls, type = "HC1"))
> kids_robust_gls_se <- robust_se_gls["KIDS", "Std. Error"]
>
> # Robust GLS confidence interval
> kids_robust_gls_ci <- c(kids_gls_coef - t_critical * kids_robust_gls_se,
+                         kids_gls_coef + t_critical * kids_robust_gls_se)
>
> cat("Robust GLS 95% CI for KIDS:", kids_robust_gls_ci[1], "to", kids_robust_gls_ci[2], "\n")
Robust GLS 95% CI for KIDS: -121.4134 to -32.19919
```

8.18

a.

```
> # Output results
> cat("Variance estimate for males:", var_males, "\n")
Variance estimate for males: 0.2207836
> cat("Variance estimate for females:", var_females, "\n")
Variance estimate for females: 0.2101181
> cat("F-statistic:", f_stat, "\n")
F-statistic: 1.05076
> cat("Upper critical F-value:", f_crit_upper, "\n")
Upper critical F-value: 1.058097
> cat("Lower critical F-value:", f_crit_lower, "\n")
Lower critical F-value: 0.9452566
> cat("p-value:", p_value, "\n")
p-value: 0.08569168
>
> if (f_stat > f_crit_upper || f_stat < f_crit_lower) {
+   cat("Reject the null hypothesis: Variances differ between males and females.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: No evidence of different variances.\n")
+ }
Fail to reject the null hypothesis: No evidence of different variances.
```

b.

```
> # Method 1: Using bptest function
> bp_test1 <- bptest(full_model, ~ METRO + FEMALE + BLACK, data = cps5)
> print(bp_test1)

        studentized Breusch-Pagan test

data:  full_model
BP = 23.557, df = 3, p-value = 3.091e-05
```

```
> # Method 2: Manual implementation of NR² test
```

```
> # Output results
> cat("NR² statistic (specific variables):", nr_squared1, "\n")
NR² statistic (specific variables): 23.55681
> cat("Critical chi-squared value (1%):", chi_crit1, "\n")
Critical chi-squared value (1%): 11.34487
> cat("p-value:", p_value1, "\n")
p-value: 3.0909e-05
> if (p_value1 < 0.01) {
+   cat("Reject the null hypothesis: Evidence of heteroskedasticity at 1% level.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: No evidence of heteroskedasticity at 1% level.\n")
+ }
Reject the null hypothesis: Evidence of heteroskedasticity at 1% level.

> # Repeat test using all explanatory variables
> aux_model2 <- lm(residuals_sq ~ EDUC + EXPER + EXPER2 + FEMALE + BLACK +
+                  METRO + SOUTH + MIDWEST + WEST, data = cps5)
> r_squared2 <- summary(aux_model2)$r.squared
> nr_squared2 <- n * r_squared2
>
> # Degrees of freedom (number of variables in auxiliary regression)
> df2 <- 9  # All explanatory variables
>
> # p-value
> p_value2 <- 1 - pchisq(nr_squared2, df2)
>
> # Output results
> cat("\nNR² statistic (all variables):", nr_squared2, "\n")

NR² statistic (all variables): 109.4243
> cat("Critical chi-squared value (5%):", chi_crit1, "\n")
Critical chi-squared value (5%): 11.34487
> cat("p-value:", p_value2, "\n")
p-value: 0
>
> if (p_value2 < 0.01) {
+   cat("Reject the null hypothesis: Evidence of heteroskedasticity at 1% level.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: No evidence of heteroskedasticity at 1% level.\n")
+ }
Reject the null hypothesis: Evidence of heteroskedasticity at 1% level.
```

The Breusch-Pagan tests contradict the conclusion in part (a), showing that even if male/female variances are similar, heteroskedasticity is still present due to other variables.

c.

```
> # Output results
> cat("White test NR² statistic:", white_nr_squared, "\n")
White test NR² statistic: 194.4447
> cat("Degrees of freedom:", white_df, "\n")
Degrees of freedom: 54
> cat("Critical chi-squared value (5%):", white_crit, "\n")
Critical chi-squared value (5%): 72.15322
> cat("p-value:", white_p_value, "\n")
p-value: 0
>
> if (white_p_value < 0.05) {
+   cat("Reject the null hypothesis: Evidence of heteroskedasticity at 5% level.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: No evidence of heteroskedasticity at 5% level.\n")
+ }
Reject the null hypothesis: Evidence of heteroskedasticity at 5% level.
```

d.

```
> # Display results
> print(ci_comparison)
      Variable Conventional_SE    Robust_SE Conv_CI_Lower Conv_CI_Upper Robust_CI_Lower
1  (Intercept)    3.211489e-02 3.279417e-02  1.1384302204  1.2643338265     1.137098683
2         EDUC    1.758260e-03 1.905821e-03  0.0977830603  0.1046761665     0.097493811
3        EXPER    1.300342e-03 1.314908e-03  0.0270727569  0.0321706349     0.027044205
4       EXPER2    2.635448e-05 2.759687e-05 -0.0004974407 -0.0003941203    -0.000499876
5       FEMALE    9.529136e-03 9.488260e-03 -0.1841810529 -0.1468229075    -0.184100928
6        BLACK    1.694240e-02 1.609369e-02 -0.1447358548 -0.0783146449    -0.143072211
7        METRO    1.230675e-02 1.158215e-02  0.0948966363  0.1431441846     0.096316998
8        SOUTH    1.356134e-02 1.390164e-02 -0.0723384657 -0.0191724010    -0.073005508
9      MIDWEST    1.410367e-02 1.372426e-02 -0.0915893895 -0.0362971859    -0.090845662
10        WEST    1.440237e-02 1.455684e-02 -0.0348207138  0.0216425095    -0.035123519
   Robust_CI_Upper Width_Change
1     1.2656653641    2.1151700
2     0.1049654160    8.3924256
3     0.0321991870    1.1201599
4    -0.0003916849    4.7141298
5    -0.1469030324   -0.4289553
6    -0.0799782888   -5.0093755
7     0.1417238226   -5.8878100
8    -0.0185053588    2.5092781
9    -0.0370409129   -2.6901695
10    0.0219453146    1.0725746

> # Summarize which coefficients have wider/narrower intervals
> wider <- ci_comparison$Variable[ci_comparison$Width_Change > 0]
> narrower <- ci_comparison$Variable[ci_comparison$Width_Change < 0]
>
> cat("\nCoefficients with wider intervals using robust SE:\n")

Coefficients with wider intervals using robust SE:
> print(wider)
[1] "(Intercept)" "EDUC"        "EXPER"       "EXPER2"      "SOUTH"       "WEST"
>
> cat("\nCoefficients with narrower intervals using robust SE:\n")

Coefficients with narrower intervals using robust SE:
> print(narrower)
[1] "FEMALE"  "BLACK"   "METRO"   "MIDWEST"
```

e.

```
> # Display results
> print(ci_comparison_fgls)
      Variable      FGLS_SE    Robust_SE FGLS_CI_Lower FGLS_CI_Upper Robust_CI_Lower Robust_CI_Upper
1  (Intercept) 3.159320e-02 3.279417e-02  1.1302695254  1.2541279001     1.137098683    1.2656653641
2         EDUC 1.764615e-03 1.905821e-03  0.0982024458  0.1051204663     0.097493811    0.1049654160
3        EXPER 1.297517e-03 1.314908e-03  0.0275467064  0.0326335081     0.027044205    0.0321991870
4       EXPER2 2.678918e-05 2.759687e-05 -0.0005086498 -0.0004036251    -0.000499876   -0.0003916849
5       FEMALE 9.480830e-03 9.488260e-03 -0.1847977976 -0.1476290326    -0.184100928   -0.1469030324
6        BLACK 1.699247e-02 1.609369e-02 -0.1441623553 -0.0775448504    -0.143072211   -0.0799782888
7        METRO 1.145945e-02 1.158215e-02  0.0953066354  0.1402324253     0.096316998    0.1417238226
8        SOUTH 1.352230e-02 1.390164e-02 -0.0713493481 -0.0183363498    -0.073005508   -0.0185053588
9      MIDWEST 1.398389e-02 1.372426e-02 -0.0906033967 -0.0357807800    -0.090845662   -0.0370409129
10        WEST 1.437651e-02 1.455684e-02 -0.0336747637  0.0226870709    -0.035123519    0.0219453146
   Width_Change
1   -3.66215144
2   -7.40917939
3   -1.32261698
4   -2.92675532
5   -0.07831277
6    5.58466271
7   -1.05938782
8   -2.72870976
9    1.89177946
10  -1.23885273

> # Summarize which coefficients have wider/narrower intervals
> fgls_wider <- ci_comparison_fgls$Variable[ci_comparison_fgls$Width_Change > 0]
> fgls_narrower <- ci_comparison_fgls$Variable[ci_comparison_fgls$Width_Change < 0]
>
> cat("\nCoefficients with wider intervals using FGLS vs. OLS robust:\n")

Coefficients with wider intervals using FGLS vs. OLS robust:
> print(fgls_wider)
[1] "BLACK"   "MIDWEST"
>
> cat("\nCoefficients with narrower intervals using FGLS vs. OLS robust:\n")

Coefficients with narrower intervals using FGLS vs. OLS robust:
> print(fgls_narrower)
[1] "(Intercept)" "EDUC"        "EXPER"       "EXPER2"      "FEMALE"      "METRO"       "SOUTH"
[8] "WEST"
```

f.

```
> # Display results
> print(ci_comparison_robust_fgls)
      Variable      FGLS_SE FGLS_Robust_SE OLS_Robust_SE FGLS_CI_Lower FGLS_CI_Upper
1  (Intercept) 3.159320e-02   3.235961e-02  3.279417e-02  1.1302695254  1.2541279001
2         EDUC 1.764615e-03   1.892760e-03  1.905821e-03  0.0982024458  0.1051204663
3        EXPER 1.297517e-03   1.304616e-03  1.314908e-03  0.0275467064  0.0326335081
4       EXPER2 2.678918e-05   2.740828e-05  2.759687e-05 -0.0005086498 -0.0004036251
5       FEMALE 9.480830e-03   9.438075e-03  9.488260e-03 -0.1847977976 -0.1476290326
6        BLACK 1.699247e-02   1.586874e-02  1.609369e-02 -0.1441623553 -0.0775448504
7        METRO 1.145945e-02   1.156288e-02  1.158215e-02  0.0953066354  0.1402324253
8        SOUTH 1.352230e-02   1.383444e-02  1.390164e-02 -0.0713493481 -0.0183363498
9      MIDWEST 1.398389e-02   1.371270e-02  1.372426e-02 -0.0906033967 -0.0357807800
10        WEST 1.437651e-02   1.450875e-02  1.455684e-02 -0.0336747637  0.0226870709
   FGLS_Robust_CI_Lower FGLS_Robust_CI_Upper OLS_Robust_CI_Lower OLS_Robust_CI_Upper
1          1.1287671947         1.2556302309         1.137098683        1.2656653641
2          0.0979512563         0.1053716558         0.097493811        0.1049654160
3          0.0275327897         0.0326474248         0.027044205        0.0321991870
4         -0.0005098633        -0.0004024116        -0.000499876       -0.0003916849
5         -0.1847139905        -0.1477128397        -0.184100928       -0.1469030324
6         -0.1419596068        -0.0797475989        -0.143072211       -0.0799782888
7          0.0951039024         0.1404351583         0.096316998        0.1417238226
8         -0.0719612018        -0.0177244961        -0.073005508       -0.0185053588
9         -0.0900718148        -0.0363123619        -0.090845662       -0.0370409129
10        -0.0339339956         0.0229463028        -0.035123519        0.0219453146
```

```
> # 4) Print summaries
> cat("\n--- Robust FGLS vs. FGLS ---\n")

--- Robust FGLS vs. FGLS ---
> cat("Wider under Robust FGLS:\n");    print(wider_rf_vs_f)
Wider under Robust FGLS:
[1] "(Intercept)" "EDUC"        "EXPER"       "EXPER2"      "METRO"       "SOUTH"       "WEST"
> cat("Narrower under Robust FGLS:\n"); print(narrower_rf_vs_f)
Narrower under Robust FGLS:
[1] "FEMALE"  "BLACK"   "MIDWEST"
>
> cat("\n--- Robust FGLS vs. Robust OLS ---\n")

--- Robust FGLS vs. Robust OLS ---
> cat("Wider under Robust FGLS:\n");    print(wider_rf_vs_o)
Wider under Robust FGLS:
character(0)
> cat("Narrower under Robust FGLS:\n"); print(narrower_rf_vs_o)
Narrower under Robust FGLS:
 [1] "(Intercept)" "EDUC"        "EXPER"       "EXPER2"      "FEMALE"      "BLACK"       "METRO"
 [8] "SOUTH"       "MIDWEST"     "WEST"
```

g.

We would go with the **Feasible GLS estimates combined with heteroskedasticity-consistent ("robust") standard errors**:

1. **We've established there is heteroskedasticity** (White test, NR^2 tests), so the **conventional** SEs from either OLS or FGLS are invalid. We must use **robust** SEs to get correct inference.

2. **FGLS re-weights the observations** in a way that – if the estimated variance function is close to the truth – yields **more efficient** (i.e. lower-variance) coefficient estimates than plain OLS. In the table below, we can see that the FGLS point estimates change slightly (e.g. intercept 1.201→1.192, EDUC 0.1012→0.1017, etc.), and the **FGLS-robust SEs** are generally a hair **smaller** than the OLS-robust SEs.

3. By pairing FGLS with robust SEs, we get the "best of both worlds":

    - **Efficiency** from FGLS weighting

    - **Validity** of inference no matter how perfectly we have modeled the variance

4. In practice, most applied papers therefore report **"FGLS with White (or Huber-White) SEs"** as their primary specification whenever they've confirmed heteroskedasticity.

```
> # Print the summary table
> print(summary_wide)
            Method (Intercept)_Coef (Intercept)_SE EDUC_Coef     EDUC_SE EXPER_Coef    EXPER_SE
1  OLS Conventional        1.201382     0.03211489 0.1012296 0.001758260 0.02962170 0.001300342
2        OLS Robust        1.201382     0.03279417 0.1012296 0.001905821 0.02962170 0.001314908
3 FGLS Conventional        1.192199     0.03159320 0.1016615 0.001764615 0.03009011 0.001297517
4       FGLS Robust        1.192199     0.03235961 0.1016615 0.001892760 0.03009011 0.001304616
    EXPER2_Coef    EXPER2_SE FEMALE_Coef   FEMALE_SE BLACK_Coef    BLACK_SE METRO_Coef   METRO_SE
1 -0.0004457805 2.635448e-05  -0.1655020 0.009529136 -0.1115252 0.01694240  0.1190204 0.01230675
2 -0.0004457805 2.759687e-05  -0.1655020 0.009488260 -0.1115252 0.01609369  0.1190204 0.01158215
3 -0.0004561375 2.678918e-05  -0.1662134 0.009480830 -0.1108536 0.01699247  0.1177695 0.01145945
4 -0.0004561375 2.740828e-05  -0.1662134 0.009438075 -0.1108536 0.01586874  0.1177695 0.01156288
   SOUTH_Coef   SOUTH_SE MIDWEST_Coef MIDWEST_SE    WEST_Coef    WEST_SE
1 -0.04575543 0.01356134  -0.06394329 0.01410367 -0.006589102 0.01440237
2 -0.04575543 0.01390164  -0.06394329 0.01372426 -0.006589102 0.01455684
3 -0.04484285 0.01352230  -0.06319209 0.01398389 -0.005493846 0.01437651
4 -0.04484285 0.01383444  -0.06319209 0.01371270 -0.005493846 0.01450875
```