1.

a)

VARIABLES	(1) Fixed Effects
lnwg	0.1561***
	(0.0322)
Constant	7.2386***
	(0.0853)
Observations	2,662
Number of id	532
R-squared	0.0169
Individual FE	YES
Year FE	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The coefficient β in this fixed effect model represents the elasticity of hours worked $lnhr_{it}$ with respect to wages $lnwg_{it}$, controlling for individual-specific characteristics (that do not vary over time) and common time effects. If we are estimating this using the model provided, β measures the percentage change in hours worked for a 1% change in wages.

b)

lnhr	(1) Pooled OLS	(2) Two-way FE	(3) within (FE)	(4) between
lnwg	0.0832***	0.1561***	0.1587***	0.0723***
	(0.0141)	(0.0322)	(0.0321)	(0.0223)
Constant	7.4351***	7.2386***	7.2375***	7.4668***
	(0.0374)	(0.0853)	(0.0843)	(0.0588)
Observations	2,662	2,662	2,662	2,662
R-squared	0.0129	0.0169	0.0113	0.0195
Number of id		532	532	532
Individual FE		YES	YES	
Year FE		YES		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	first diff	re GLS	re MLE (i)	re MLE (i)	re MLE (i)	re MLE (iii)	re MLE (iii)	re MLE (iii)
d_lnwg	0.1418***							
	(0.0346)							
lnwg		0.1002***	0.1012***			0.1012***		
		(0.0183)	(0.0186)			(0.0324)		
Constant	-0.0013	7.3919***	7.3893***	0.1672***	0.2491***	7.3893***	0.1672***	0.2491***
	(0.0071)	(0.0487)	(0.0494)	(0.0075)	(0.0038)	(0.0872)	(0.0211)	(0.0207)
Observations	2,130	2,662	2,662	2,662	2,662	2,662	2,662	2,662
R-squared	0.0078							
Number of id		532	532	532	532	532	532	532

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The estimates are similar in direction and magnitude but vary depending on assumptions about unobserved heterogeneity, dynamics, and fixed effects. These differences are informative for understanding the underlying relationships and the robustness of the results.

The discrepancy in these estimates arises from methodological differences:

Pooled OLS assumes no individual or time heterogeneity, leading to a potentially biased β due to omitted variable bias.

Two-way FE and Within (FE) adjust for individual and time-fixed effects, likely increasing β, as unobserved heterogeneity (e.g., individual ability or firm-specific traits) that might otherwise bias the relationship downward is now controlled for.

Between estimator uses cross-sectional differences, yielding a lower β, likely because it doesn't fully account for within-individual variations over time.

c)

While there are similarities in the range, the models differ, especially between the random effects and fixed effects or OLS estimates. The random effects MLE results are generally within the lower range of the coefficients compared to the fixed effects and pooled OLS estimates although all of the estimates of β are significant in all the models estimated in the last part.

d)

lnhr	(1) Pooled OLS	(2) Pooled OLS	(3) within (FE)	(4) within (FE)	(5) re GLS	(6) re GLS
lnwg	0.0832***	0.0832***	0.1587***	0.1587**	0.1002***	0.1002***
	(0.0141)	(0.0241)	(0.0321)	(0.0765)	(0.0183)	(0.0379)
Constant	7.4351***	7.4351***	7.2375***	7.2375***	7.3919***	7.3919***
	(0.0374)	(0.0662)	(0.0843)	(0.2003)	(0.0487)	(0.1029)
Observations	2,662	2,662	2,662	2,662	2,662	2,662
R-squared	0.0129	0.0129	0.0113	0.0113		
Robust	NO	YES	NO	YES	NO	YES
Number of id			532	532	532	532
Individual FE			YES	YES		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Yes, there is a systematic difference between default standard errors and panel robust standard errors as shown in the table:

In columns (1) and (3), the default standard errors are smaller compared to the robust ones. For example:

In column (1) for ln(wg), the standard error is 0.0141 (default).

In column (2), which applies robust standard errors, the standard error for the same coefficient increases to 0.0241.

When robust standard errors are applied, they account for potential heteroskedasticity and autocorrelation within the panel data. This adjustment typically increases the magnitude of the standard errors, as observed in the robust versions (e.g., column (2)) compared to the default versions (e.g., column (1)).

The differences indicate that default standard errors might understate the uncertainty in the coefficient estimates, especially in the presence of heteroskedasticity or serial correlation. Panel robust standard errors provide a more reliable measure of variability under these conditions, which is essential for valid statistical inference in panel data models.

The systematic difference arises because default standard errors may underestimate variability in the presence of heteroskedasticity or serial correlation, while robust standard errors provide a more conservative and accurate estimate.

e)

In a fixed effects model, the relationship between the dependent variable $lnhr_{it}$ and the independent variable $lnwg_{it}$ is specified as:

$$lnhr_{it} = \alpha_i + \beta lnwg_{it} + u_{it},$$

where:

- α_i is the individual-specific fixed effect (unobserved time-invariant heterogeneity),
- u_{it} is the idiosyncratic error term.

Pooled OLS **does not account for individual-specific effects $\boldsymbol{\alpha_i}$.** As a result, if $lnwg_{it}$ is correlated with α_i, which is common in most economic panel data settings (e.g., unobserved factors such as individual skill levels affecting both hours worked and wages), the pooled OLS estimator will suffer from **omitted variable bias**. This bias occurs because the unobserved fixed effects α_i are part of the error term in pooled OLS, violating the assumption that the explanatory variable $lnwg_{it}$ is uncorrelated with the error.

Therefore, the **pooled OLS estimator will not be consistent for β in a fixed effects model** unless $lnwg_{it}$ is uncorrelated with α_i, which is rare in practice.

In a random effects model, the relationship is given by:

$$lnhr_{it} = \alpha_i + \beta lnwg_{it} + u_{it},$$

with the critical assumption that:

$$\alpha_i \sim i.i.d.\ , and\ uncorrelated\ with\ lnwg_{it}$$

If this assumption holds, then $lnwg_{it}$ is uncorrelated with the composite error term $(\alpha_i + u_{it})$, and the pooled OLS estimator will provide a consistent estimate of β. However, while pooled OLS would be consistent in this case, it is not efficient because it does not exploit the panel structure of the data. Random effects GLS would be more efficient as it accounts for the variance structure introduced by the random effects.

Thus, **pooled OLS will be consistent for β in a random effects model** as long as the random effects assumption (no correlation between $lnwg_{it}$ and α_i) is satisfied.

f)

	Coefficients (b) fixed_effe~s	(B) random_eff~s	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
lnwg	.1586654	.1002087	.0584567	.0263989

b = Consistent under H0 and Ha; obtained from **xtreg**.
B = Inconsistent under Ha, efficient under H0; obtained from **xtreg**.

Hausman Test Procedure:

1. **Calculate the Test Statistic:**

$$H = (b - B)' \left[\mathrm{Var}(b - B)\right]^{-1} (b - B)$$

where:

b is the coefficient vector from the fixed effects model (0.1586654 for lnwg).

B is the coefficient vector from the random effects model (0.1002087 for lnwg).

Var(b−B) is the variance of the difference, given as Std. err where Std. err.=0.0263989.

Difference: b−B=0.1586654−0.1002087=0.0584567.

Variance of the difference: (0.0263989)2=0.0006969.

Test statistic: H=((b−B)^2)/Var(b−B)=(0.0584567)^2/0.0006969.

Compare the Test Statistic to a Critical Value: The test statistic follows a chi-squared (χ2) distribution with degrees of freedom equal to the number of coefficients tested (1 in this case). Use the critical value for χ2(1) at the desired significance level (e.g., χ0.05,12=3.841).

Let's compute the statistic:

The Hausman test statistic is H=4.903

Decision Rule:

The test statistic follows a χ2 distribution with 1 degree of freedom.

At a significance level of 0.05, the critical value is χ0.05,12=3..

Conclusion:

Since H=4.903 exceeds the critical value of 3.841, we reject the null hypothesis that the random effects model is consistent.

The fixed effects model is preferred because it does not rely on the assumption that the unobserved individual-specific effects are uncorrelated with the regressors.

g)

Pooled OLS (Column 1): (The table in part b)

Coefficient on lnwg: 0.0832, significant at $p<0.01$.

This indicates that, on average, a 1% increase in wages is associated with an 8.32% increase in hours worked. However, pooled OLS may suffer from omitted variable bias due to unobserved heterogeneity.

Two-Way Fixed Effects (Column 2):

Coefficient on lnwg: 0.1561, significant at $p<0.01$.

Adding individual and year fixed effects accounts for unobserved time-invariant factors and common shocks, suggesting a stronger positive relationship between wages and labor supply when these are controlled for.

Within Fixed Effects (Column 3):

Coefficient on lnwg 0.1587, significant at $p<0.01$.

This further supports a positive and statistically significant relationship between wages and labor supply.

Conclusion:

Across all models, the coefficient on lnwg is positive and significant, suggesting that an increase in wages leads to an increase in hours worked. This provides strong evidence that the labor supply curve is upward sloping, indicating that higher wages incentivize individuals to work more hours, consistent with the theoretical expectations of a positively sloped labor supply curve in the context of wage increases.

2)

lcrime	(1) Pooled OLS	(2) Fixed Effects	(3) Fixed Effects
d78	-0.0547	0.0857	0.0857
	(0.0945)	(0.0638)	(0.0552)
clrprc1	-0.0185***	-0.0040	-0.0040
	(0.0053)	(0.0047)	(0.0042)
clrprc2	-0.0174***	-0.0132**	-0.0132***
	(0.0054)	(0.0052)	(0.0047)
Constant	4.1812***	3.3510***	3.3510***
	(0.1879)	(0.2325)	(0.2599)
Observations	106	106	106
R-squared	0.4710	0.4209	0.4209
Number of district		53	53
Individual FE		YES	YES
Robust		NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

a)

Intercept: 4.1812, which represents the baseline log crime rate when all other variables are zero.

Year Dummy (d78) Coefficient: -0.0547, indicating no statistically significant change in log crime rates between 1972 and 1978 (p=0.564)

Clear-Up Percentage (clrprc) Coefficient: -0.0185, which is statistically significant (p=0.001). A 1% increase in the current clear-up percentage reduces the log crime rate by 0.0185.

Lagged Clear-Up Percentage (clrprc2) Coefficient: -0.0174, also statistically significant (p=0.002). A 1% increase in the previous period's clear-up percentage reduces the log crime rate by 0.0174.

This suggests that both current and past clear-up percentages have significant deterrent effects on crime.

The Durbin-Watson statistic is 1.223, which is below the benchmark value of 2, indicating potential positive serial correlation in the residuals. This suggests that the strict exogeneity assumption may not fully hold. A fixed-effects model with heteroskedasticity-robust standard errors will help address this issue. Let me proceed with Part b by estimating the model using fixed effects.

residuals	(1) Test For Serial Correlation
lag_resid	0.581***
	(0.0994)
o.d78	-
clrprc1	0.00347
	(0.00528)
clrprc2	-0.00656
	(0.00544)
Constant	0.113
	(0.173)
Observations	53
R-squared	0.417

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A significant value indicates the presence of serial correlation.

b)

For the variable clrprc1, the Pooled OLS coefficient is -0.0185 (significant at the 1% level), whereas the Fixed Effects estimate reduces to -0.0040 and becomes statistically insignificant. This suggests that the effect estimated by Pooled OLS for clrprc1 may be driven by unobserved factors, which are controlled in the FE model.

For the variable clrprc2, the Pooled OLS coefficient is -0.0174 (significant at the 1% level), while the FE model coefficient is smaller in magnitude at -0.0132, but still significant. The smaller FE coefficient indicates potential omitted variable bias in the Pooled OLS estimates, as FE controls for unobserved, time-invariant characteristics.

The d78 coefficient in the FE model is positive (0.0857) and significant in some specifications, whereas the Pooled OLS coefficient is negative (-0.0547) and insignificant. This divergence underscores the potential role of unobserved heterogeneity that biases the Pooled OLS estimates.

The standard errors in the FE specification are not robust to serial correlation in columns (2) and are only adjusted in column (3). The presence of individual fixed effects and the structure of the data (likely panel) raise the possibility of serial correlation, especially since repeated observations on the same units over time can introduce correlation in the residuals. For instance, unmodeled temporal effects could bias standard errors, leading to incorrect inference. Testing for serial correlation is essential to ensure that the assumptions of the FE model are valid. Serial correlation can be tested using methods like the Breusch-Godfrey test or examining residual patterns.

Serial correlation is a concern given the structure of panel data and should be tested and accounted for, particularly when individual FE are used, as is evident in column (3) where robust standard errors are reported.

c)

```
( 1)  clrprc1 - clrprc2 = 0

      F(  1,    52) =    1.84
           Prob > F =    0.1804
```

The p-value (0.1804) is greater than the common significance level (e.g., 0.05).

Therefore, we fail to reject the null hypothesis. This means there is no sufficient evidence to conclude that $\beta_1 = \beta_2$. In other words, the coefficients for clrprc1 and clrprc2 can be assumed to be statistically equal.

Since $\beta 1$ and $\beta 2$ are not significantly different, a more parsimonious model can be used by combining clrprc1 and clrprc2 into a single variable (e.g., their sum or average) and re-estimating the model. This reduces model complexity without losing significant information.

VARIABLES	(1) Fixed Effects
d78	0.0993*
	(0.0555)
clrprc_combined	-0.0083***
	(0.0029)
Constant	3.3164***
	(0.2529)
Observations	106
Number of district	53
R-squared	0.4076
Individual FE	YES
Robust	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Comparing this models goodness-of-fit metrics R-squared with the previous model we see that it hasn't changed significantly and this model can be used which is less complex.

3)

a)

VARIABLES	(1) Pooled OLS	(2) Pooled OLS	(3) Pooled OLS
educ	0.0994***	0.0994***	0.0994***
	(0.0047)	(0.0046)	(0.0092)
black	-0.1438***	-0.1438***	-0.1438***
	(0.0236)	(0.0244)	(0.0501)
hisp	0.0157	0.0157	0.0157
	(0.0208)	(0.0197)	(0.0392)
exper	0.0892***	0.0892***	0.0892***
	(0.0101)	(0.0101)	(0.0124)
expersq	-0.0028***	-0.0028***	-0.0028***
	(0.0007)	(0.0007)	(0.0009)
married	0.1077***	0.1077***	0.1077***
	(0.0157)	(0.0153)	(0.0261)
union	0.1801***	0.1801***	0.1801***
	(0.0171)	(0.0162)	(0.0276)
Constant	-0.0347	-0.0347	-0.0347
	(0.0646)	(0.0647)	(0.1201)
Observations	4,360	4,360	4,360
R-squared	0.1866	0.1866	0.1866
Robust	NO	YES	
Cluster			YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The usual OLS standard errors are not fully reliable in this context, even if (c_i) (the individual-specific fixed effect) is uncorrelated with the explanatory variables. This unreliability arises primarily due to the presence of unobserved heterogeneity and serial correlation in the data.

The term (c_i) represents time-invariant, unobserved characteristics of individuals, such as innate ability or personality traits, which influence wages. While pooled OLS assumes that (c_i) is uncorrelated with the explanatory variables, it introduces serial correlation in the residuals (u_{it}) across observations for the same individual. This means that for any given individual (i), residuals are not independent over time because they share the same unobserved (c_i). As a result, the OLS standard errors may underestimate the true variability, leading to overstated statistical significance of the coefficients.

Moreover, wage data often exhibits heteroscedasticity, where the variance of the error terms is not constant across observations. This can arise from differences in characteristics such as education, race, or union status. Under heteroscedasticity, the usual OLS standard errors are invalid because they rely on the assumption of homoscedasticity (constant error variance).

To mitigate these issues, robust standard errors can be used to account for heteroscedasticity. However, robust errors alone do not address serial correlation, which remains a concern in panel data. A better solution is to use clustered standard errors at the individual level, which account for within-individual correlation of residuals over time. Clustering adjusts the standard errors for both heteroscedasticity and serial correlation, providing more reliable inference.

Even if (c_i) (unobserved individual heterogeneity) is uncorrelated with the explanatory variables, pooled OLS does not account for potential clustering or correlation in the panel structure of the data. Therefore, standard errors from pooled OLS are not reliable in the presence of unobserved heterogeneity. This justifies the use of panel-specific methods like RE or FE.

In conclusion, even if (c_i) is uncorrelated with the explanatory variables, the usual OLS standard errors are unreliable due to serial correlation and potential heteroscedasticity. For this dataset, using clustered standard errors at the individual level would be the most appropriate approach to obtain valid and reliable results.

b)

VARIABLES	(1) Pooled OLS	(2) Random Effects	(3) Random Effects
educ	0.0994***	0.1012***	0.1012***
	(0.0092)	(0.0089)	(0.0089)
black	-0.1438***	-0.1441***	-0.1441***
	(0.0501)	(0.0476)	(0.0503)
hisp	0.0157	0.0202	0.0202
	(0.0392)	(0.0426)	(0.0399)
exper	0.0892***	0.1121***	0.1121***
	(0.0124)	(0.0083)	(0.0105)
expersq	-0.0028***	-0.0041***	-0.0041***
	(0.0009)	(0.0006)	(0.0007)
married	0.1077***	0.0628***	0.0628***
	(0.0261)	(0.0168)	(0.0190)
union	0.1801***	0.1074***	0.1074***
	(0.0276)	(0.0178)	(0.0209)
Constant	-0.0347	-0.1075	-0.1075
	(0.1201)	(0.1107)	(0.1152)
Observations	4,360	4,360	4,360
R-squared	0.1866		
Cluster	YES		
Number of nr		545	545
Robust		NO	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The coefficients in the RE model are almost identical to the Pooled OLS estimates. This is expected because the RE model effectively averages out unobserved individual heterogeneity (c_i), but assumes it is uncorrelated with the explanatory variables. In this dataset, the lack of significant changes suggests minimal correlation between c_i and the predictors.

The standard errors in the RE model are slightly larger than those in the Pooled OLS regression (the version without Robust and Clustered standard errors) but is close to the Clustered version of the pooled OLS. This adjustment accounts for individual-level unobserved heterogeneity and potential dependence in errors across time for the same individual.

The RE model provides more reliable estimates when c_i is uncorrelated with the predictors, as it accounts for panel data structure. Pooled OLS, in contrast, ignores the panel nature of the data, making its inference less robust.

c)

VARIABLES	(1) Random Effects	(2) Random Effects	(3) two-way FE	(4) two-way FE
expersq	-0.0041***	-0.0041***	-0.0052***	-0.0052***
	(0.0006)	(0.0007)	(0.0007)	(0.0008)
married	0.0628***	0.0628***	0.0467**	0.0467**
	(0.0168)	(0.0190)	(0.0183)	(0.0210)
union	0.1074***	0.1074***	0.0800***	0.0800***
	(0.0178)	(0.0209)	(0.0193)	(0.0227)
educ	0.1012***	0.1012***		
	(0.0089)	(0.0089)		
black	-0.1441***	-0.1441***		
	(0.0476)	(0.0503)		
hisp	0.0202	0.0202		
	(0.0426)	(0.0399)		
exper	0.1121***	0.1121***		
	(0.0083)	(0.0105)		
Constant	-0.1075	-0.1075	1.4260***	1.4260***
	(0.1107)	(0.1152)	(0.0183)	(0.0210)
Observations	4,360	4,360	4,360	4,360
R-squared			0.1806	0.1806
Number of nr	545	545	545	545
Robust	NO	YES	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Because fixed effects control for all time-invariant characteristics of the individuals. Experience, in a typical panel data setting, is often perfectly collinear with individual fixed effects and the time variable since it increases linearly over time for each individual. This collinearity makes it impossible to estimate the effect of $exper_{it}$ separately from the fixed effects. Thus, $exper_{it}$ cannot be included in the FE model.

In the RE model (columns 1 and 2), the coefficient on "married" is approximately 0.0628. In the two-way FE model (columns 3 and 4), this drops to about 0.0467. This suggests that some of the observed marriage premium in the RE model was due to time-invariant individual characteristics correlated with marriage (e.g., personality traits or preferences).

The union premium decreases as well, moving from 0.1074 in the RE models to 0.0800 in the FE models. Similar to the marriage premium, this reduction indicates that part of the union premium in the RE models is explained by time-invariant unobserved heterogeneity that the FE model accounts for. By switching to the FE model, the analysis provides estimates that are less likely to be biased by unobserved individual-level heterogeneity, offering a more accurate assessment of the causal impact of marriage and union status on the outcome variable.

d)

	(1)
VARIABLES	FE
exper	0.1705***
	(0.0273)
expersq	-0.0060***
	(0.0009)
married	0.0475***
	(0.0183)
union	0.0794***
	(0.0193)
d81_educ	-0.0010
	(0.0026)
d82_educ	-0.0062
	(0.0041)
d83_educ	-0.0114**
	(0.0057)
d84_educ	-0.0136*
	(0.0072)
d85_educ	-0.0162*
	(0.0087)
d86_educ	-0.0170*
	(0.0101)
d87_educ	-0.0167
	(0.0115)
Constant	0.9200***
	(0.0766)
Observations	4,360
Number of nr	545
R-squared	0.1812

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The coefficients of these interaction terms represent the deviation in the return to education for each year relative to the baseline. Most of the interaction terms have negative coefficients, and some are statistically significant at the 10% level. This suggests that the return to education has decreased slightly over time for those specific years. The coefficients for the interaction terms are relatively small in magnitude, indicating that the changes in the return to education over time are modest. The pattern of negative coefficients suggests a downward trend in the return to education during the years under study. This could imply that the premium for education diminished over the 1981-1987 period, possibly due to changes in labor market conditions or other factors affecting wage determination.

In summary, based on the results, the return to education appears to have slightly declined over time, particularly in the mid-1980s, as indicated by the significant and negative coefficients for the interaction terms in 1984, 1985, and 1986.

e)

	(1)
VARIABLES	FE
exper	0.1213***
	(0.0100)
expersq	-0.0050***
	(0.0008)
married	0.0436**
	(0.0209)
union	0.0757***
	(0.0218)
lead_union	0.0515**
	(0.0223)
Constant	1.0521***
	(0.0303)
Observations	3,815
Number of nr	545
R-squared	0.1458

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The coefficient of *lead_union* is positive (0.0515) and statistically significant at the 5% level. This indicates that future union membership is positively associated with the current log of wages (increases it by 5 percent).

The significance of $lead\_union$ suggests that future union membership is correlated with the current error term, violating the assumption of strict exogeneity in the fixed effects model. Strict exogeneity requires that the explanatory variables (here, $union_{it}$) are uncorrelated with the error term for all time periods, including future ones. The positive and significant $lead\_union$ indicates that this assumption does not hold, possibly because unobserved factors influencing wages (e.g., individual career progression, employer preferences) also affect the likelihood of future union membership.