Exam

Please discuss each of the 3 problems on a separate sheet of paper, not just on a separate page!

Problem 1: (15 points)
A researcher has data for the year 2000 from the US National Longitudinal Survey of Youth on the following characteristics of the respondents: hourly earnings, EARNINGS, measured in dollars; years of schooling, S; years of work experience, EXP; sex; and ethnicity (blacks, hispanics, and 'whites' (those not classified as black or hispanic)). She drops the hispanics from the sample, leaving 2,135 'whites' and 273 blacks, and defines dummy variables MALE and BLACK. MALE is defined to be 1 for males and 0 for females. BLACK is defined to be 1 for blacks and 0 for 'whites'. She defines LGEARN to be the natural logarithm of EARNINGS. She fits the following ordinary least squares regressions, each with LGEARN as the dependent variable:

(1) Explanatory variables S, EXP, and MALE, whole sample

(2) Explanatory variables S, EXP, MALE, and BLACK, whole sample

(3) Explanatory variables S, EXP, and MALE, 'whites' only

(4) Explanatory variables S, EXP, and MALE, blacks only

She then defines interactive terms SB = S*BLACK, EB = EXP*BLACK, and MB = MALE*BLACK, and runs a fifth regression, still with LGEARN as the dependent variable:

(5) Explanatory variables S, EXP, MALE, BLACK, SB, EB, MB, whole sample.

The results are shown in the following table. Unfortunately some of those for Regression 4 are missing from the table. RSS = residual sum of squares. Standard errors are given in parentheses.

	1 whole sample	2 whole sample	3 'whites' only	4 blacks only	5 whole sample
S	0.124 (0.004)	0.121 (0.004)	0.122 (0.004)	V	0.122 (0.004)
EXP	0.033 (0.002)	0.032 (0.002)	0.033 (0.003)	W	0.033 (0.003)
MALE	0.278 (0.020)	0.277 (0.020)	0.306 (0.021)	X	0.306 (0.021
BLACK	—	-0.144 (0.032)	—	—	0.205 (0.225)
SB	—	—	—	—	-0.009 (0.016)
EB	—	—	—	—	-0.006 (0.007)
MB	—	—	—	—	-0.280 (0.065)
constant	0.390 (0.075)	0.459 (0.076)	0.411 (0.084)	Y	0.411 (0.082)
R2	0.335	0.341	0.332	0.321	0.347
RSS	610.0	605.1	555.7	Z	600.0
n	2,408	2,408	2,135	273	2,408

1. Calculate the missing coefficients V, W, and X in Regression 4 (just the coefficients, not the standard errors), giving an explanation of your computations.
2. Give an interpretation of the coefficient of BLACK in Regression 2.
3. Perform an F-test of the joint explanatory power of BLACK, SB, EB, and MB in Regression (5).

Solution:

1. Since Regression 5 includes a complete set of black intercept and slope dummy variables, the basic coefficients will be the same as for a regression using the 'whites' only subsample and the coefficients modified by the dummies will give the counterparts for the blacks only subsample. Hence $V = 0.122 - 0.009 = 0.113$, $W = 0.033 - 0.006 = 0.027$, and $X = 0.306 - 0.280 = 0.026$.
2. It suggests that blacks earn 14.4 percent less than whites, controlling for other characteristics.
   Or: $(e^{-0.144} - 1) \cdot 100 = 13.4\%$
3. Write the model as
   $LGEARN = \beta_1 + \beta_2 S + \beta_3 EXP + \beta_4 MALE + \beta_5 BLACK + \beta_6 SB + \beta_7 EB + \beta_8 MB + u.$

The null hypothesis for the test is $H_0 : \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$, and the alternative hypothesis is H_1: at least one coefficient is different from 0. The F statistic is

* $F(4, 2400) = \frac{(610.0-600.0)/4}{600.0/2400} = \frac{2400}{240} = 10.0$, or
* $F(4, 2400) = \frac{(0.347-0.335)/4}{(1-0.347)/2400} = 11.02$

In either case (what ever statistic you used) this is highly significant and so the null hypothesis is rejected.

Problem 2: (20 points)
Baltagi and Griffin (1983) considered the following gasoline demand equation:

$$log\frac{Gas}{Car} = \beta_0 + \beta_1 \cdot log\frac{Y}{N} + \beta_2 \cdot log\frac{PMG}{PGDP} + \beta_3 \cdot log\frac{Car}{N} + u$$

where Gas/Car is motor gasoline consumption per auto, Y/N is real income per capita, $PMG/PGDP$ is real motor gasoline price and Car/N denotes the stock of cars per capita. This panel consists of annual observations across eighteen OECD countries, covering the period 1960-1978.
The tables below show some computations done with the data.

1. Regarding the pooled regression, the random effects model and the fixed effects model which of these models do you suggest? Argue carefully and do all possible (necessary) tests.
2. How high is the price elasticity in your final model? How large is the corresponding standard error?
3. State your final model with all computed estimates.

Table 1: Pooled OLS regression.

```
 reg  lgaspcar lincomep lrpmg lcarpcap

      Source |       SS       df       MS              Number of obs =     342
-------------+------------------------------           F(  3,   338) =  664.00
       Model |  87.8386024     3  29.2795341           Prob > F      =  0.0000
    Residual |  14.9043581   338  .044095734           R-squared     =  0.8549
-------------+------------------------------           Adj R-squared =  0.8536
       Total |  102.742961   341  .301299005           Root MSE      =  .20999

------------------------------------------------------------------------------
    lgaspcar |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lincomep |   .8899616   .0358058    24.86   0.000     .8195313    .9603919
       lrpmg |  -.8917979   .0303147   -29.42   0.000    -.9514272   -.8321685
    lcarpcap |  -.7633727   .0186083   -41.02   0.000    -.7999754   -.7267701
       _cons |   2.391326   .1169343    20.45   0.000     2.161315    2.621336
------------------------------------------------------------------------------
```

Table 2: Random Effects Estimator using Gasoline Demand Data

```
Random-effects GLS regression                   Number of obs      =       342
Group variable: cntry                           Number of groups   =        18

R-sq:  within  = 0.8363                         Obs per group: min =        19
       between = 0.7099                                        avg =      19.0
       overall = 0.7309                                        max =        19
Random effects u_i ~ Gaussian                   Wald chi2(3)       =   1642.20
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
    lgaspcar |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lincomep |   .5549858   .0591282     9.39   0.000     .4390967    .6708749
       lrpmg |  -.4203893   .0399781   -10.52   0.000     -.498745   -.3420336
    lcarpcap |  -.6068402    .025515   -23.78   0.000    -.6568487   -.5568316
       _cons |   1.996699    .184326    10.83   0.000     1.635427    2.357971
-------------+----------------------------------------------------------------
     sigma_u |  .19554468
     sigma_e |  .09233034
         rho |  .81769856   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Table 3: Fixed Effects Estimator using Gasoline Demand Data (Regression using demeaned variables.)

```
Fixed-effects (within) regression               Number of obs      =       342
Group variable: cntry                           Number of groups   =        18

R-sq:  within  = 0.8396                         Obs per group: min =        19
       between = 0.5755                                        avg =      19.0
       overall = 0.6150                                        max =        19
                                                F(3,321)           =    560.09
corr(u_i, Xb)  = -0.2468                        Prob > F           =    0.0000
------------------------------------------------------------------------------
    lgaspcar |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lincomep |   .6622498    .073386     9.02   0.000     .5178715    .8066282
       lrpmg |  -.3217025   .0440992    -7.29   0.000    -.4084626   -.2349424
    lcarpcap |  -.6404829   .0296788   -21.58   0.000    -.6988726   -.5820933
       _cons |    2.40267   .2253094    10.66   0.000     1.959401     2.84594
```

```
-------------+----------------------------------------------------------------
     sigma_u |  .34841289
     sigma_e |  .09233034
         rho |  .93438173   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(17, 321) =    83.96             Prob > F = 0.0000
```

Table 4: Breusch and Pagan Lagrangian multiplier test for random effects.

```
        lgaspcar[cntry,t] = Xb + u[cntry] + e[cntry,t]
        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                lgaspcar |    .301299       .5489071
                       e |   .0085249       .0923303
                       u |   .0382377       .1955447

        Test:   Var(u) = 0
                              chi2(1) =  1465.55
                          Prob > chi2 =     0.0000
```

Table 5: Hausman test.

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       .       random_eff~s    Difference          S.E.
-------------+----------------------------------------------------------------
    lincomep |    .6622498     .5549858          .107264        .0434669
       lrpmg |   -.3217025    -.4203893         .0986868        .0186143
    lcarpcap |   -.6404829    -.6068402        -.0336428        .0151597
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      302.80
                Prob>chi2 =      0.0000
```

Table 6: 'areg' (Linear regression with the dummy-variable set)

```
Linear regression, absorbing indicators                Number of obs =     342
                                                       F(  3,    17) =   15.12
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9734
                                                       Adj R-squared =  0.9717
                                                       Root MSE      =  .09233

                               (Std. Err. adjusted for 18 clusters in country)
------------------------------------------------------------------------------
             |               Robust
    lgaspcar |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lincomep |   .6622498   .1625623     4.07   0.001     .3192733    1.005226
       lrpmg |  -.3217025   .1296806    -2.48   0.024    -.5953046   -.0481004
    lcarpcap |  -.6404829   .1025072    -6.25   0.000    -.8567543   -.4242116
       _cons |    2.40267   .6132217     3.92   0.001     1.108886    3.696455
-------------+----------------------------------------------------------------
     country |   absorbed                                     (18 categories)
```

Solution:

1. * Pooled versus Fixed Effects Model:
   $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_n$
   $F(n-1, nT-n-K) = \frac{(R^2_{LSDV} - R^2_{pooled})/(n-1)}{(1-R^2_{LSDV})/(nT-n-K)} = \frac{(0.9734-0.8549)/17}{(1-0.9734)/(342-18-3)} =$ 84.12
   The group individual effects are jointly significant, therefore reject $H_0 \Rightarrow$ fixed effects model.

   * Breusch Pagan LM Test
   $H_0 : \sigma_u^2 = 0$ (Pooled Model)
   $H_1 : \sigma_u^2 \neq 0$ (Error component is present)
   $LM = 1,465.55$ with $p - value = 0.000 < \alpha = 0.05 \Rightarrow$ reject H_0 and do not use the pooled model.

   * Hausmann Test
   $H_0 : \beta_{FE} - \beta_{RE} = 0$ (Random Effects Model)
   $H_1 : \beta_{FE} - \beta_{RE} \neq 0$ (Fixed Effects Model)

$$H = 302.8 \text{ with } p - value = 0.000 < \alpha = 0.05 \Rightarrow \text{reject } H_0$$

Final model: Fixed Effects Model.

2. How high is the price elasticity in your final model? How large is the corresponding standard error?
   Estimate $b_2 = -0.322$: 1% of price increase reduces the demand per 0.322%. The standard error of the demeaned regression has to be corrected with $\sqrt{\frac{nT-K}{nT-n-K}} = \sqrt{\frac{342-3}{342-18-3}}$ and therefore is 0.0453. Or if heteroscedasticity is still a problem one could use the robust standard errors from the dummy regression: 0.1297.

3. State your final model with all computed estimates.

$$log\left(\widehat{\frac{Gas}{Car}}\right) = 2.403 + \hat{\alpha}_i + 0.662log\left(\frac{Y}{N}\right) - 0.322log\left(\frac{PMG}{PGDP}\right) - 0.640log\left(\frac{Car}{N}\right)$$

Problem 3: (15 points)
A researcher correctly believes that a time series process can be adequately represented by a deterministic trend, $Y_t = \beta_1 + \beta_2 t + u_t$, where $t = 1, 2, ..., T$ is the time trend, and u_t is an identically and independently distributed (i.i.d) random variable with zero mean and finite variance (σ^2).

1. Demonstrate that Y_t is nonstationary.
2. Demonstrate that the first difference in Y_t is stationary.
3. Given a sample of size T, and noting that $\bar{t} = 0.5T$, demonstrate that the ordinary least squares (OLS) estimator of the slope coefficient may be decomposed as

$$b_2^{OLS} = \beta_2 + \frac{\sum_{t=1}^{T}(t-0.5T)(u_t - \bar{u})}{\sum_{t=1}^{T}(t-0.5T)^2}$$

4. Hence demonstrate that b_2^{OLS} is unbiased.

Solution:

1. $E(Y_t) = \beta_1 + \beta_2 t$ is dependent of t and therefore violating the assumption of covariance stationarity.
2. $\Delta Y_t = Y_t - Y_{t-1} = \beta_1 + \beta_2 t + u_t - (\beta_1 + \beta_2(t-1) + u_{t-1}) = \beta_2 + u_t - u_{t-1}$
$E(\Delta Y_t) = \beta_2$, $Var(\Delta Y_t) = 2\sigma^2$, and $\rho_1 = -\frac{1}{2}$ and $\rho_k = 0, \text{if } |k| > 1$. Therefore ΔY_t is covariance stationary.
3.

$$\begin{aligned} b_2^{OLS} &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \\ &= \frac{\sum_t (t - 0.5T)(Y_t - \bar{Y})}{\sum_t (t - 0.5T)^2} \\ &= \frac{\sum_t (t - 0.5T)(\beta_1 + \beta_2 t + u_t - \beta_1 - \beta_2 \bar{t} - \bar{u})}{\sum_t (t - 0.5T)^2} \\ &= \frac{\sum_t (t - 0.5T)(\beta_2 (t - \bar{t}) + (u_t - \bar{u}))}{\sum_t (t - 0.5T)^2} \\ &= \beta_2 + \frac{\sum_{t=1}^{T}(t - 0.5T)(u_t - \bar{u})}{\sum_{t=1}^{T}(t - 0.5T)^2} \end{aligned}$$

4.

$$
\begin{aligned}
E[b_2^{OLS}] &= \beta_2 + E[\frac{\sum_{t=1}^{T}(t-0.5T)(u_t-\bar{u})}{\sum_{t=1}^{T}(t-0.5T)^2}] \\
&= \beta_2 + \frac{\sum_{t=1}^{T}(t-0.5T)E[(u_t-\bar{u})]}{\sum_{t=1}^{T}(t-0.5T)^2} \\
&= \beta_2
\end{aligned}
$$