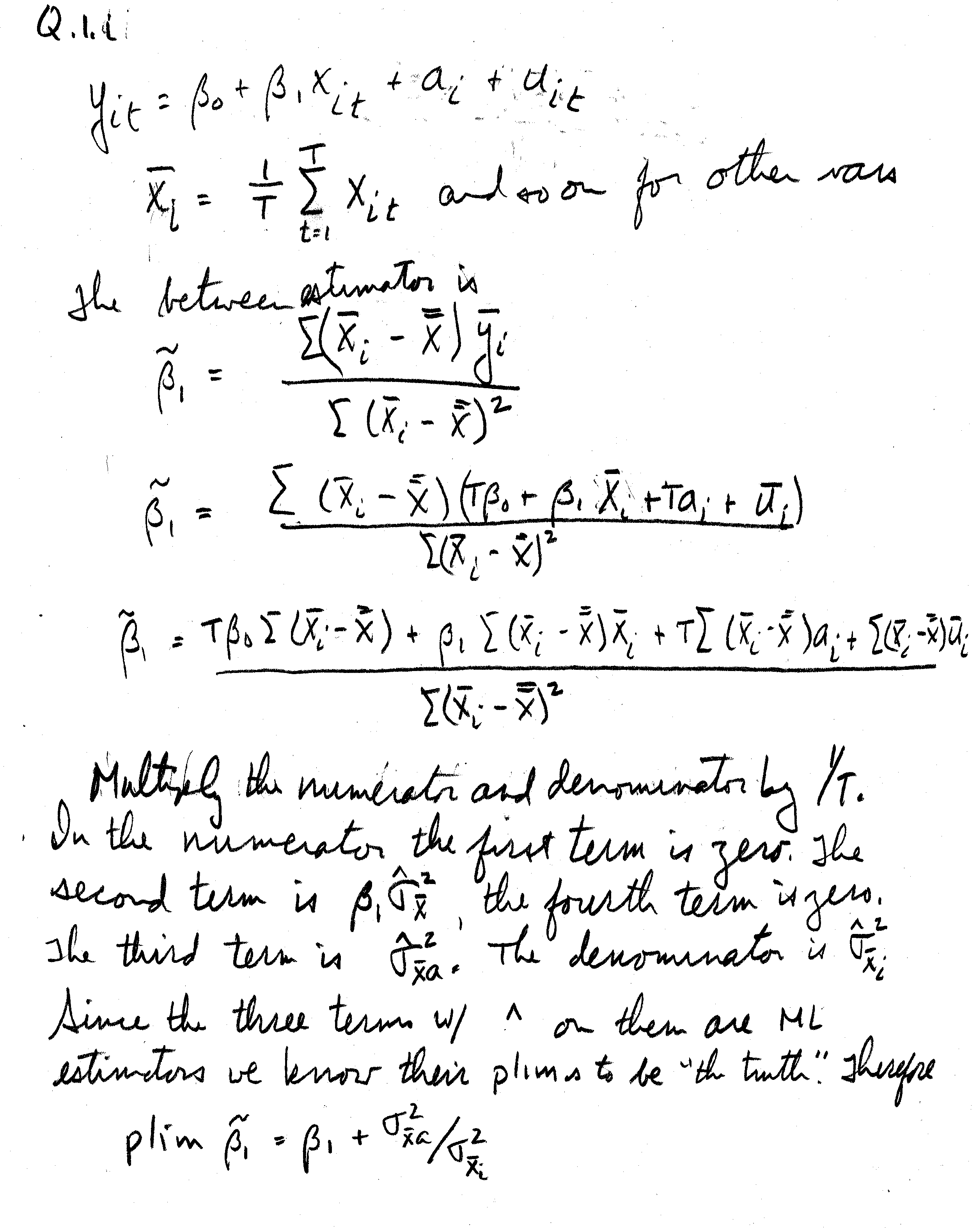
Temple University

Department of Economics

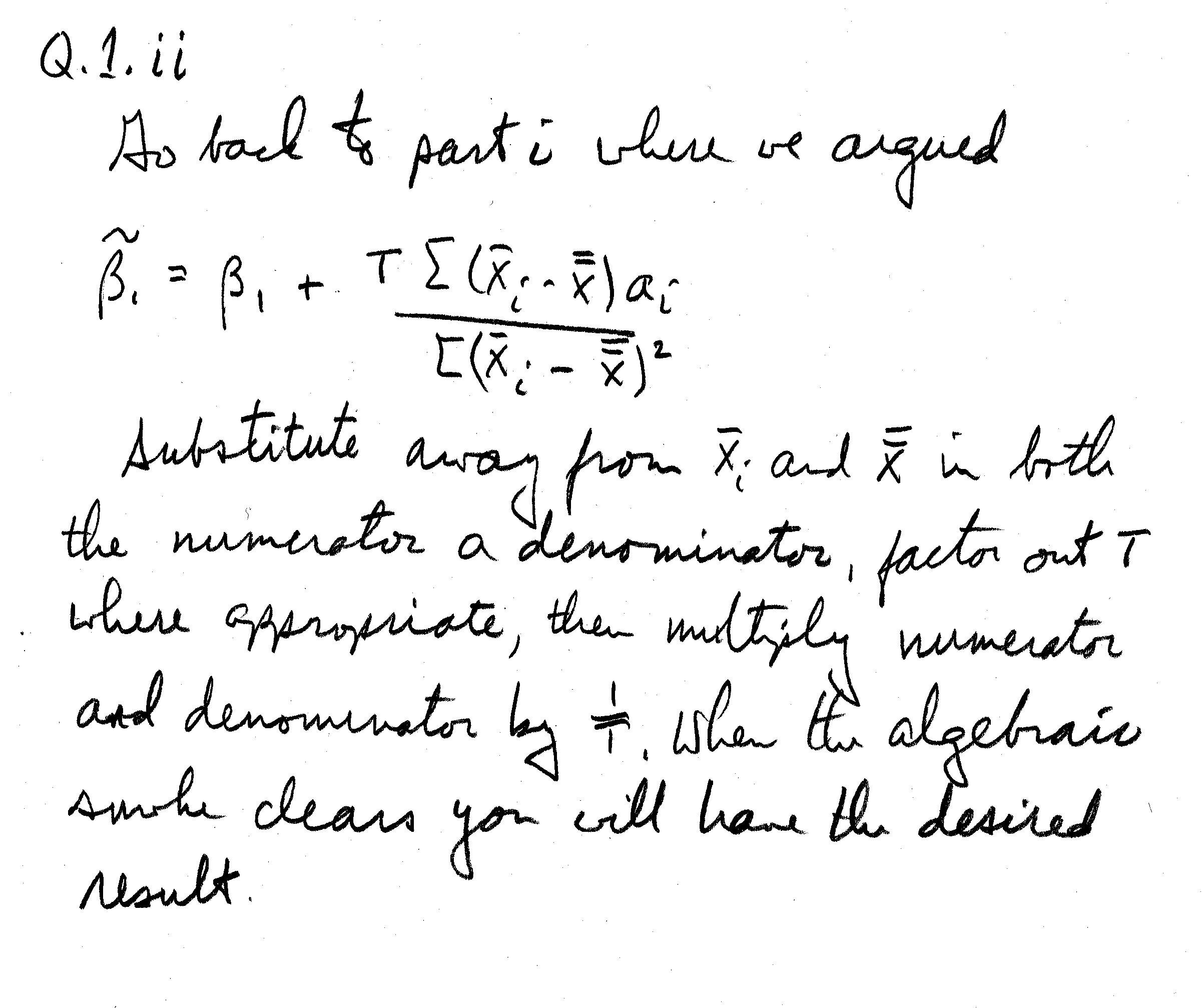
Economics 8190

Fall 2011

1. The REM model is for a panel data set with i = 1,…,n and t = 1,…,T. The equation used to obtain the between estimator is where the overbar represents the average over time. We can assume that E(ai) = 0 because we have included an intercept in the model. Suppose that is uncorrelated with but for all i,t.
2. Let be the between estimator. Show that .



1. Assume further that the for t = 1,2,…,T are uncorrelated and with constant variance . Show that .



1. In this exercise you will estimate a wage equation for men using the dataset wagepan.raw, the description of which is in wagepan.des. The basic model has as its dependent variable log(wage). The independent variables are educ, black, hispan, exper, exper2, married, and union.

There are a total of 4360 observations: 545 individuals observed for each of 8 years.

* 1. Which independent variables are time varying and which are not?

|  |  |
| --- | --- |
| Time Varying | NOT Time Varying |
| Log(wage)  Exper  Expersq  Married  Union | Educ  Black  Hispanic |

There are some caveats.

* + 1. Exper is increasing by one for each passing year for everyone in the sample, although many people have different starting values for this variable. The effect is that a set of time dummies would be perfectly correlated with exper for each person. This is broken by expersq because the transformation is quadratic rather than linear.
    2. Marriage and divorce over the sample period result in this series being time varying.
    3. People join and leave the union, so the series is time varying.
    4. In this sample no one acquired additional education after joining the study, so education is constant over time for each person, although not everyone has the same amount of education.
    5. One’s race does not change over time.
  1. Use OLS to estimate a model with all of the independent variables and report your results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: LWAGE | | |  |  |
| Method: Panel Least Squares | | |  |  |
| Date: 09/19/11 Time: 10:58 | | |  |  |
| Sample: 1980 1987 | | |  |  |
| Periods included: 8 | | |  |  |
| Cross-sections included: 545 | | |  |  |
| Total panel (balanced) observations: 4360 | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | -0.034706 | 0.064569 | -0.537498 | 0.5910 |
| EDUC | 0.099388 | 0.004678 | 21.24762 | 0.0000 |
| BLACK | -0.143842 | 0.023560 | -6.105465 | 0.0000 |
| HISP | 0.015698 | 0.020811 | 0.754305 | 0.4507 |
| EXPER | 0.089179 | 0.010111 | 8.819962 | 0.0000 |
| EXPERSQ | -0.002849 | 0.000707 | -4.027155 | 0.0001 |
| MARRIED | 0.107666 | 0.015696 | 6.859221 | 0.0000 |
| UNION | 0.180073 | 0.017121 | 10.51793 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.186587 | Mean dependent var | | 1.649147 |
| Adjusted R-squared | 0.185278 | S.D. dependent var | | 0.532609 |
| S.E. of regression | 0.480744 | Akaike info criterion | | 1.374868 |
| Sum squared resid | 1005.810 | Schwarz criterion | | 1.386575 |
| Log likelihood | -2989.212 | Hannan-Quinn criter. | | 1.378999 |
| F-statistic | 142.6132 | Durbin-Watson stat | | 0.863095 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Each additional year of education increases the log of one’s earnings.

There is an earnings penalty for being black, but there is no penalty for being Hispanic.

There are diminishing marginal returns to more education.

Union membership and being married are positively correlated with wages.

* 1. Estimate the coefficients of a REM equation with all of the independent variables and report your results.

Do it assuming there are random effects across people

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: LWAGE | | |  |  |
| Method: Panel EGLS (Cross-section random effects) | | | | |
| Date: 09/19/11 Time: 11:02 | | |  |  |
| Sample: 1980 1987 | | |  |  |
| Periods included: 8 | | |  |  |
| Cross-sections included: 545 | | |  |  |
| Total panel (balanced) observations: 4360 | | | |  |
| Swamy and Arora estimator of component variances | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | -0.107464 | 0.110411 | -0.973312 | 0.3305 |
| EDUC | 0.101225 | 0.008890 | 11.38692 | 0.0000 |
| BLACK | -0.144131 | 0.047488 | -3.035096 | 0.0024 |
| HISP | 0.020151 | 0.042488 | 0.474281 | 0.6353 |
| EXPER | 0.112119 | 0.008239 | 13.60860 | 0.0000 |
| EXPERSQ | -0.004069 | 0.000590 | -6.893449 | 0.0000 |
| MARRIED | 0.062795 | 0.016728 | 3.753850 | 0.0002 |
| UNION | 0.107379 | 0.017783 | 6.038446 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Effects Specification | |  |  |
|  |  |  | S.D. | Rho |
|  |  |  |  |  |
|  |  |  |  |  |
| Cross-section random | | | 0.324567 | 0.4606 |
| Idiosyncratic random | | | 0.351255 | 0.5394 |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Weighted Statistics | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.178240 | Mean dependent var | | 0.589338 |
| Adjusted R-squared | 0.176918 | S.D. dependent var | | 0.388204 |
| S.E. of regression | 0.352193 | Sum squared resid | | 539.8226 |
| F-statistic | 134.8502 | Durbin-Watson stat | | 1.588405 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Unweighted Statistics | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.180806 | Mean dependent var | | 1.649147 |
| Sum squared resid | 1012.958 | Durbin-Watson stat | | 0.846488 |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. Estimate the coefficients of an LSDV model that has just exper2, married and union on the RHS.
     1. Why have some independent variables been dropped from the specification? Make sure you understand why exper is not included.

Those variables which do not vary over time have been dropped. Exper is not included because it has the effect, although not perfect, of being a fixed time effects variable. Furthermore, the change in exper is constant across time. The presence of the person dummy accounts for the differences across people in their years of experience in the initial period, but the effect of a subsequent year of experience cannot be distinguished from the aggregate benefit of experience.

* + 1. Report your results.

**LSDV / Fixed Effects results:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: LWAGE | | |  |  |
| Method: Panel Least Squares | | |  |  |
| Date: 09/19/11 Time: 11:43 | | |  |  |
| Sample: 1980 1987 | | |  |  |
| Periods included: 8 | | |  |  |
| Cross-sections included: 545 | | |  |  |
| Total panel (balanced) observations: 4360 | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | 1.395302 | 0.012294 | 113.4964 | 0.0000 |
| EXPERSQ | 0.003699 | 0.000189 | 19.56038 | 0.0000 |
| MARRIED | 0.107343 | 0.018196 | 5.899163 | 0.0000 |
| UNION | 0.082762 | 0.019770 | 4.186373 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Effects Specification | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Cross-section fixed (dummy variables) | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.600523 | Mean dependent var | | 1.649147 |
| Adjusted R-squared | 0.543201 | S.D. dependent var | | 0.532609 |
| S.E. of regression | 0.359974 | Akaike info criterion | | 0.911490 |
| Sum squared resid | 493.9646 | Schwarz criterion | | 1.713408 |
| Log likelihood | -1439.048 | Hannan-Quinn criter. | | 1.194508 |
| F-statistic | 10.47621 | Durbin-Watson stat | | 1.750265 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. Use the appropriate statistical tests to decide which is the preferred specification.

This last part should have had more clarification. In order to choose the preferred specification with respect to the error structure we need to estimate models that have the same RHS variables. If the RHS differs across specifications then we may be fooled by the significant differences between RHS variables. We’ll go with the sparse model that includes just expersq, married and union.

We’ll need the **REM results for the sparse model:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: LWAGE | | |  |  |
| Method: Panel EGLS (Cross-section random effects) | | | | |
| Date: 09/19/11 Time: 11:55 | | |  |  |
| Sample: 1980 1987 | | |  |  |
| Periods included: 8 | | |  |  |
| Cross-sections included: 545 | | |  |  |
| Total panel (balanced) observations: 4360 | | | |  |
| Swamy and Arora estimator of component variances | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | 1.405208 | 0.019082 | 73.64196 | 0.0000 |
| EXPERSQ | 0.003192 | 0.000180 | 17.77294 | 0.0000 |
| MARRIED | 0.131555 | 0.016718 | 7.868846 | 0.0000 |
| UNION | 0.103451 | 0.018307 | 5.650828 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Effects Specification | |  |  |
|  |  |  | S.D. | Rho |
|  |  |  |  |  |
|  |  |  |  |  |
| Cross-section random | | | 0.349990 | 0.4859 |
| Idiosyncratic random | | | 0.359974 | 0.5141 |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Weighted Statistics | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.114094 | Mean dependent var | | 0.563588 |
| Adjusted R-squared | 0.113484 | S.D. dependent var | | 0.386055 |
| S.E. of regression | 0.363490 | Sum squared resid | | 575.5354 |
| F-statistic | 186.9996 | Durbin-Watson stat | | 1.506674 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Unweighted Statistics | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.042514 | Mean dependent var | | 1.649147 |
| Sum squared resid | 1183.960 | Durbin-Watson stat | | 0.732410 |
|  |  |  |  |  |
|  |  |  |  |  |

We’ll also need the **pooled OLS results** for the sparse model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: LWAGE | | |  |  |
| Method: Panel Least Squares | | |  |  |
| Date: 09/19/11 Time: 11:59 | | |  |  |
| Sample: 1980 1987 | | |  |  |
| Periods included: 8 | | |  |  |
| Cross-sections included: 545 | | |  |  |
| Total panel (balanced) observations: 4360 | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | 1.464776 | 0.013893 | 105.4363 | 0.0000 |
| EXPERSQ | 0.001177 | 0.000197 | 5.969530 | 0.0000 |
| MARRIED | 0.190004 | 0.016222 | 11.71284 | 0.0000 |
| UNION | 0.170411 | 0.018151 | 9.388550 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.068292 | Mean dependent var | | 1.649147 |
| Adjusted R-squared | 0.067650 | S.D. dependent var | | 0.532609 |
| S.E. of regression | 0.514278 | Akaike info criterion | | 1.508813 |
| Sum squared resid | 1152.084 | Schwarz criterion | | 1.514666 |
| Log likelihood | -3285.212 | Hannan-Quinn criter. | | 1.510879 |
| F-statistic | 106.4283 | Durbin-Watson stat | | 0.766792 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Correlated Random Effects - Hausman Test | | | |  |
| Equation: Untitled | | |  |  |
| Test cross-section random effects | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Test Summary | | Chi-Sq. Statistic | Chi-Sq. d.f. | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| Cross-section random | | 88.494170 | 3 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Cross-section random effects test comparisons: | | | | |
|  |  |  |  |  |
| Variable | Fixed | Random | Var(Diff.) | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| EXPERSQ | 0.003699 | 0.003192 | 0.000000 | 0.0000 |
| MARRIED | 0.107343 | 0.131555 | 0.000052 | 0.0007 |
| UNION | 0.082762 | 0.103451 | 0.000056 | 0.0056 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |
| --- |
| Due to the large test statistic for any of the coefficients we do not reject the null that the FE model is the correct model.  You should have noticed that the chi-square stat based on the vector of coefficients was negative. The difference between the coefficient variances was not positive definite. This should have raised a red flag. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Redundant Fixed Effects Tests** | | |  |  |
| Equation: Untitled | | |  |  |
| Test cross-section fixed effects | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Effects Test | | Statistic | d.f. | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| Cross-section F | | 9.336050 | (544,3812) | 0.0000 |
| Cross-section Chi-square | | 3692.328295 | 544 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |

We reject the null that the fixed effects are all zero. We do not pick pooled OLS.

Since we prefer LSDV to OLS and we prefer LSDV to REM we can conclude that LSDV / Fixed Effects is the preferred model.