# Exercise Chapter 4

Using ‘Labour\_Force\_SA\_SALDRU\_1993’ answer the following questions:

1. What was the average rate of broad unemployment in South Africa in 1993?

To answer this we simply need the mean of our data.

|  |
| --- |
| Average unemployment in South Africa in 1993 across all racial groups |
| . sum unemp  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  unemp | 13461 .3102296 .4626047 0 1 |

2. How did this unemployment rate differ between white and black South Africans? Present as a table and as a regression of unemployment on race and interpret the coefficients on the race dummy.

|  |
| --- |
| Average unemployment for White and Black South Africans in 1993 |
| . keep if african==1 | white==1  (2925 observations deleted)  . sum unemp  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  unemp | 11701 .3293736 .4700058 0 1  . bysort african:sum unemp  ---------------------------------------------------------------------------------------------------------------------------------------------------------------  -> african = 0  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  unemp | 2013 .0452062 .2078076 0 1  ---------------------------------------------------------------------------------------------------------------------------------------------------------------  -> african = 1  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  unemp | 9688 .3884187 .4874158 0 1 |

Note above that we have confined the data to Black and White South Africans as defined in the data. The average unemployment rate is now slightly higher. The unemployment rate of White South Africans is 4.52%, while the unemployment rate of the Black South African is 38.84%.

|  |
| --- |
| A Linear Probability Model of Unemployment on Race |
| . reg unemp african  Source | SS df MS Number of obs = 11701  -------------+------------------------------ F( 1, 11699) = 961.72  Model | 196.327495 1 196.327495 Prob > F = 0.0000  Residual | 2388.26681 11699 .204142817 R-squared = 0.0760  -------------+------------------------------ Adj R-squared = 0.0759  Total | 2584.59431 11700 .220905496 Root MSE = .45182  ------------------------------------------------------------------------------  unemp | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  african | .3432125 .0110672 31.01 0.000 .3215189 .3649062  \_cons | .0452062 .0100704 4.49 0.000 .0254666 .0649458  ------------------------------------------------------------------------------ |

The regression gives the same results as the summary statistics. For White South Africans (i.e. "*african*"=0), the predicted unemployment rate is equal to 4.52%. And for Black South Africans (i.e. "*african*"=1), the predicted unemployment rate is equal to 38.84% (0.3432+0.0452). What the parameter estimates imply is that the unemployment rates significantly differ across race. This could be due to discrimination, different employment opportunities in different areas or mismatch between the potential employee’s skills and those demanded by the firm.

3. Estimate a linear probability model for unemployment (the linear probability model is simply OLS with a binary dependent variable) as a function of “otherinc” only.

|  |
| --- |
|  |
| . reg unemp otherinc  Source | SS df MS Number of obs = 11600  -------------+------------------------------ F( 1, 11598) = 17.86  Model | 3.94262539 1 3.94262539 Prob > F = 0.0000  Residual | 2560.13703 11598 .220739527 R-squared = 0.0015  -------------+------------------------------ Adj R-squared = 0.0015  Total | 2564.07966 11599 .221060407 Root MSE = .46983  ------------------------------------------------------------------------------  unemp | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  otherinc | .0000153 3.63e-06 4.23 0.000 8.22e-06 .0000224  \_cons | .3272431 .0044049 74.29 0.000 .3186087 .3358775  ------------------------------------------------------------------------------ |

From the pooled regression estimates, an additional unit of other income results in a 0.00153 percentage points increase in the probability of being unemployed. As we noticed the large difference across race, it would be better to model this with a race control.

4. Test for normality and if the errors are homoskedastic and report your results.

|  |
| --- |
| These are tests run after the regression reported in the table above of:  reg unemp otherinc |
| . predict unemp\_res,resid  (12907 missing values generated)  . estat hettest, rhs  Breusch-Pagan / Cook-Weisberg test for heteroskedasticity  Ho: Constant variance  Variables: otherinc  chi2(1) = 32.73  Prob > chi2 = 0.0000  . sum unemp\_res, detail  Residuals  -------------------------------------------------------------  Percentiles Smallest  1% -.3456341 -1.487577  5% -.3333734 -.9969821  10% -.3317642 -.5418048 Obs 11600  25% -.3272431 -.5418048 Sum of Wgt. 11600  50% -.3272431 Mean -3.29e-17  Largest Std. Dev. .469809  75% .667776 .6727569  90% .6727569 .6727569 Variance .2207205  95% .6727569 .6727569 Skewness .7206875  99% .6727569 .6727569 Kurtosis 1.534274  . display r(skewness)  .72068747  . display r(kurtosis)  1.5342736  . scalar S=r(skewness)  . scalar KK=r(kurtosis)  . scalar N=r(N)  . scalar JB=N\*(S^2/6 + (KK-3)^2/24)  . scalar list JB  JB = 2042.5259  . disp chiprob(2,JB) 0  . sktest unemp\_res  Skewness/Kurtosis tests for Normality  ------- joint ------  Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2  -------------+---------------------------------------------------------------  unemp\_res | 1.2e+04 0.0000 . . .  . swilk unemp\_res  Shapiro-Wilk W test for normal data  Variable | Obs W V z Prob>z  -------------+--------------------------------------------------  unemp\_res | 11600 0.68680 1771.667 20.102 0.00000 |

The skewness-kurtosis test result is missing, which is an indication of non-normality of the residuals. Similarly, the Shapiro-Wilk test of normality rejects the null of normality at the 1% significance level. The Breusch-Pagan test rejects the null hypothesis of homoskedasticity at the 1% significance level.

5. What were the significant (in both the statistical and economic sense of this term) determinants of unemployment for black South Africans in 1993?

This question is asking you to use the data you have been given to investigate factors associated with unemployment. It is specific that you should confine the sample to Black South Africans as our exploration of the data has shown that unemployment rates are very much higher for Black than White South Africans. So, you can interpret this question as one asking you to think about what might explain this outcome. This question anticipates material which is covered in chapter 17, section17.1 (see Table 17.1) where the sample is additionally restricted to men.

|  |
| --- |
|  |
| . reg unemp age agesq educ educ\_sq otherinc i.impass totm\_rec if african==1  Source | SS df MS Number of obs = 9641  -------------+------------------------------ F( 7, 9633) = 203.44  Model | 294.773781 7 42.1105401 Prob > F = 0.0000  Residual | 1993.93216 9633 .20698974 R-squared = 0.1288  -------------+------------------------------ Adj R-squared = 0.1282  Total | 2288.70594 9640 .237417629 Root MSE = .45496  ------------------------------------------------------------------------------  unemp | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  age | -.0400094 .0022694 -17.63 0.000 -.044458 -.0355608  agesq | .0003504 .0000285 12.29 0.000 .0002945 .0004063  educ | .0137682 .0037034 3.72 0.000 .0065087 .0210277  educ\_sq | -.0017748 .0002802 -6.33 0.000 -.002324 -.0012256  otherinc | .0000313 4.76e-06 6.58 0.000 .000022 .0000407  1.impass | .0758939 .0093458 8.12 0.000 .0575742 .0942137  totm\_rec | .0003092 .000033 9.37 0.000 .0002445 .0003739  \_cons | 1.274752 .0441837 28.85 0.000 1.188143 1.361361  ------------------------------------------------------------------------------ |

|  |
| --- |
| Description of variables used in the above regression |
| . des unemp age agesq educ educ\_sq otherinc impass totm\_rec    variable name variable label  ---------------------------------------------------------------------------------------------------------------------------------------------------------------  unemp Person is unemployed or not? Y=1, n=0  age Age of person  agesq Square of age  educ Years of education  educ\_sq Square of educ  otherinc Monthly value of non-earned income received from all sources and by all members  impass Are there any roads in the community that become impassable at certain times of  totm\_rec Monthly income received in remittances by household (cash+kind), i.e., yearly s |

We regress individual unemployment status (which is a dummy variable) on age, education, other income, road conditions and household remittances for the sample of black South Africans only. As Age increases the probability of being unemployed decreases at a decreasing rate. In contrast for education as years of education increase the rate of increase is decreasing. We have used non-linear specifications for both age and education as that it an easy way of seeing if the relationship is convex (it is with age) or concave (it is with education). The result should puzzle you as it appears to show, over a certain range, that the probability of being unemployed increases with education. Surely you would expect the opposite? If you want to see what is going on look at Figure 17.1 on page 246. The effects of other income, road conditions, and household remittance are statistically significant at the 1% significance level.

Considering the economic sense of significance, we firstly look at the descriptive statistics of some variables.

|  |
| --- |
|  |
| . sum unemp age agesq educ educ\_sq otherinc impass totm\_rec if e(sample)==1  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  unemp | 9641 .3877191 .4872552 0 1  age | 9641 35.26201 11.63039 16 89  agesq | 9641 1378.661 928.363 256 7921  educ | 9641 6.770356 4.006421 0 15  educ\_sq | 9641 61.88746 52.35956 0 225  -------------+--------------------------------------------------------  otherinc | 9641 147.4281 974.3407 0 43700  impass | 9641 .4910279 .4999454 0 1  totm\_rec | 9641 49.74493 141.8395 0 5301.638 |

Despite of the very small coefficients on other income and household remittances, the variations of these variables are very large. This implies that they might be significant in economic sense even if their mean values are low or moderate. However, for those with 500 additional units of other income the increase in the probability of being unemployed is only 1.5 percentage points. In terms of economic significance it is clearly age and education which are the economically significant determinants of unemployment for black South Africans on the basis of the above LPM.

Using ‘Macro\_1980\_2000\_PENN61.dta’

6. Provide and interpret tests for normality and heteroskedasticity for:

|  |
| --- |
|  |
| . reg lrgdpch lkp time  Source | SS df MS Number of obs = 164  -------------+------------------------------ F( 2, 161) = 802.82  Model | 189.750784 2 94.8753921 Prob > F = 0.0000  Residual | 19.0266931 161 .118178218 R-squared = 0.9089  -------------+------------------------------ Adj R-squared = 0.9077  Total | 208.777477 163 1.28084342 Root MSE = .34377  ------------------------------------------------------------------------------  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .7027657 .0176802 39.75 0.000 .6678508 .7376807  time | -.0032893 .0541328 -0.06 0.952 -.1101912 .1036126  \_cons | 1.959418 .1650695 11.87 0.000 1.633437 2.285398  ------------------------------------------------------------------------------  . predict prod\_res,resid  .  .  . estat hettest, rhs  Breusch-Pagan / Cook-Weisberg test for heteroskedasticity  Ho: Constant variance  Variables: lkp time  chi2(2) = 46.27  Prob > chi2 = 0.0000  .  . sum prod\_res, detail  Residuals  -------------------------------------------------------------  Percentiles Smallest  1% -1.08468 -1.380555  5% -.646329 -1.08468  10% -.3614077 -1.033204 Obs 164  25% -.1193987 -1.024396 Sum of Wgt. 164  50% .0173679 Mean -7.58e-17  Largest Std. Dev. .3416551  75% .1631383 .6961473  90% .4217149 .7102066 Variance .1167282  95% .5010367 .8163975 Skewness -.8881458  99% .8163975 .8236672 Kurtosis 5.458331  . display r(skewness)  -.88814584  . display r(kurtosis)  5.4583314  .  . scalar S=r(skewness)  . scalar KK=r(kurtosis)  . scalar N=r(N)  .  . scalar JB=N\*(S^2/6 + (KK-3)^2/24)  . scalar list JB  JB = 62.857138  . disp chiprob(2,JB)  2.243e-14  . sktest prod\_res  Skewness/Kurtosis tests for Normality  ------- joint ------  Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2  -------------+---------------------------------------------------------------  prod\_res | 164 0.0000 0.0002 24.80 0.0000  . swilk prod\_res  Shapiro-Wilk W test for normal data  Variable | Obs W V z Prob>z  -------------+--------------------------------------------------  prod\_res | 164 0.93248 8.483 4.870 0.00000 |

|  |
| --- |
|  |
| . tsset nwbcode year  panel variable: nwbcode (strongly balanced)  time variable: year, 1980 to 2000, but with gaps  delta: 1 unit  . reg d20\_lrgdpch d20\_lkp  Source | SS df MS Number of obs = 82  -------------+------------------------------ F( 1, 80) = 96.77  Model | 5.40953702 1 5.40953702 Prob > F = 0.0000  Residual | 4.47227136 80 .055903392 R-squared = 0.5474  -------------+------------------------------ Adj R-squared = 0.5418  Total | 9.88180838 81 .121997634 Root MSE = .23644  ------------------------------------------------------------------------------  d20\_lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  d20\_lkp | .5443869 .055341 9.84 0.000 .4342548 .6545189  \_cons | .0587519 .0339368 1.73 0.087 -.0087845 .1262883  ------------------------------------------------------------------------------  . predict prod\_res\_20,resid  (82 missing values generated)  .  .  . estat hettest, rhs  Breusch-Pagan / Cook-Weisberg test for heteroskedasticity  Ho: Constant variance  Variables: d20\_lkp  chi2(1) = 4.28  Prob > chi2 = 0.0386  .  . sum prod\_res\_20, detail  Residuals  -------------------------------------------------------------  Percentiles Smallest  1% -.7831116 -.7831116  5% -.4201175 -.5957912  10% -.2842925 -.5775757 Obs 82  25% -.1281319 -.5169526 Sum of Wgt. 82  50% .0244685 Mean 1.57e-17  Largest Std. Dev. .2349749  75% .1342547 .4188301  90% .2654286 .4302549 Variance .0552132  95% .3534512 .4440705 Skewness -.6218389  99% .5064333 .5064333 Kurtosis 4.121693  . display r(skewness)  -.62183895  . display r(kurtosis)  4.1216933  .  . scalar S=r(skewness)  . scalar KK=r(kurtosis)  . scalar N=r(N)  .  . scalar JB=N\*(S^2/6 + (KK-3)^2/24)  . scalar list JB  JB = 9.5835129  . disp chiprob(2,JB) .00829787  . sktest prod\_res\_20  Skewness/Kurtosis tests for Normality  ------- joint ------  Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2  -------------+---------------------------------------------------------------  prod\_res\_20 | 82 0.0204 0.0518 8.11 0.0173  . swilk prod\_res\_20  Shapiro-Wilk W test for normal data  Variable | Obs W V z Prob>z  -------------+--------------------------------------------------  prod\_res\_20 | 82 0.96924 2.155 1.685 0.04603 |

7. The key point to notice in comparing tests for heteroskedasticity and normality is how much they differ between the two specifications. It is clear that the properties of the residual in the levels specification are very different from those of the differenced specification. While for the former we can reject the null of homoscedasticity and normality at the 1 per cent significance level this is no longer true for the differenced specification.

Now while we do not need homoscedasticity or normality to argue that our estimates are unbiased, for that all we need is the zero conditional mean assumption – see the discussion in Chapter 2. However, it is entirely possible, indeed likely, that problems of heteroscedastic residuals and non-normality are indicative that there are problems with our specification which we could usefully address. In this case we have evidence that differencing is removing some key variable or variables from the specification. What these might be and why differencing may be a very useful (and when it may not) are subjects we take up fully later in the book.

7. Compare these tests and provide reasons why they might differ.

One way of addressing this question is to graph the two dependent variables and the residuals from the regression.

|  |  |
| --- | --- |
| Histogram of lrgdpch | Histogram of d20\_lrgdpch |
|  |  |
| Histogram of residuals from production function for lrgdpch | Histogram of residuals from regression using d20\_lrgdpch |
|  |  |

The reasons why these will differ have already been covered in the answer to Chapter 3 question 7. The difference equation has allowed for time invariant unobservables while the levels equation has not. Doing so makes the residuals (a bit) more normal.