**Exercise Chapter 5**

1. Explain how to calculate the returns to education from a macro production function

and use ‘Macro\_1980\_2000\_PENN61’ to derive an estimate from that data for cross sections of 1980 and 2000 and the pooled cross section.

In Chapter 1 Section 1.4 we showed the link between the production function and the earnings function. Both the Mincerian earnings function and the human capital augmented production function can be interpreted as examples of technical relationships. The Mincerian earnings function links wages to skills while the production function is, in principle, simply a description of the technology that shows how inputs determine outputs.

The Mincerian earnings function is given in equation (1.20):

The human capital augmented production function is given as (1.23):

So given an estimate for the production function we can retrieve the implied coefficients for the earnings function. Clearly this depends on the specification we have chosen for the production function.

If you run this specification on the data you will find that while the relationship between labour productivity and education is linear the quadratic term is not significantly different from zero so in the runs reported below we confine attention to the linear specification and present three specifications.

The first two show cross section regressions for 1980 and 2000. The third regression is a pooled cross section regression controlling for time. You are asked in the question to estimate the implied Mincerian returns to education for the regressions. The calculation of ‘delta’ given below each regression calculates the implied return which is 0.19 for the cross section of 1980, 0.17 for the cross section of 2000 and 0.19 for the pooled regressions.

|  |
| --- |
| . /\*Runs for answering Exercise 5 Question 1\*/  .  . reg lrgdpch lkp tyr15 if year==1980,robust  Linear regression Number of obs = 82  F( 2, 79) = 354.08  Prob > F = 0.0000  R-squared = 0.8996  Root MSE = .34401  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .53678 .0524281 10.24 0.000 .4324245 .6411354  tyr15 | .0900477 .0299484 3.01 0.004 .030437 .1496584  \_cons | 3.018438 .360643 8.37 0.000 2.300596 3.73628  ------------------------------------------------------------------------------  . gen delta\_80=\_b[tyr15]/(1-\_b[lkp])  . sum delta\_80  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  delta\_80 | 164 .1943952 0 .1943952 .1943952  .  . reg lrgdpch lkp tyr15 if year==2000,robust  Linear regression Number of obs = 82  F( 2, 79) = 474.71  Prob > F = 0.0000  R-squared = 0.9358  Root MSE = .30255  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .6195039 .0598026 10.36 0.000 .5004698 .738538  tyr15 | .0663426 .0311635 2.13 0.036 .0043131 .1283721  \_cons | 2.323293 .4072022 5.71 0.000 1.512778 3.133809  ------------------------------------------------------------------------------  . gen delta\_00=\_b[tyr15]/(1-\_b[lkp])  . sum delta\_00  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  delta\_00 | 164 .1743582 0 .1743582 .1743582  .  . reg lrgdpch lkp tyr15 time,robust  Linear regression Number of obs = 164  F( 3, 160) = 535.49  Prob > F = 0.0000  R-squared = 0.9195  Root MSE = .3242  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .5737218 .0401488 14.29 0.000 .4944318 .6530118  tyr15 | .0814435 .0219312 3.71 0.000 .0381315 .1247555  time | -.0645375 .0501323 -1.29 0.200 -.1635439 .0344689  \_cons | 2.725696 .2769911 9.84 0.000 2.178666 3.272727  ------------------------------------------------------------------------------  . gen delta\_com=\_b[tyr15]/(1-\_b[lkp])  . sum delta\_com  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  delta\_com | 164 .1910572 0 .1910572 .1910572 |

2. Compare the Mincerian rate of return to education from the micro data ‘Labour\_Force\_SA\_SALDRU\_1993’ with that which you obtained from the macro data.

You can find this in Table 3.3.

|  |
| --- |
|  |
| . /\*EDE Page 41 Table 3.3\*/  .  . reg logwphy educ  Source | SS df MS Number of obs = 6968  -------------+------------------------------ F( 1, 6966) = 2680.34  Model | 2368.42412 1 2368.42412 Prob > F = 0.0000  Residual | 6155.34858 6966 .883627416 R-squared = 0.2779  -------------+------------------------------ Adj R-squared = 0.2778  Total | 8523.77271 6967 1.22344951 Root MSE = .94001  ------------------------------------------------------------------------------  logwphy | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  educ | .1353827 .002615 51.77 0.000 .1302565 .1405088  \_cons | .4581331 .0238719 19.19 0.000 .4113368 .5049294  ------------------------------------------------------------------------------ |

3. Do you expect the returns to education to be higher in the macro data set and, if so, why?

If at the macro level the economy wide effects of increased education can be captured then yes one would expect the returns to be higher at the macro than the micro level. If we compare the implied returns to education from our macro production function with the micro evidence from South Africa then the average returns at the macro level are indeed higher. The ‘Mincerian’ return at the macro level for the pooled cross section production function is 0.19 which compares with 0.14 with the South African data.

Of course, one must not push too far any comparison of micro data for one country with a world macro production function. However, the returns in the macro production are substantial and larger than that suggested by most micro data sets.

4. Does your answer to the previous question provide any support for the view that there are externalities to education?

The answer would be yes if we believed the point estimate on the macro production function. As has been stressed that depends on our believing the zero conditional mean assumption. As we will discuss in detail in Section III with a panel we can test one aspect of that zero conditional mean assumption namely that the regressors are uncorrelated with some time invariant factor.

5. Does the Mincerian return to education return to education obtained from ‘Labour\_Force\_SA\_SALDRU\_1993’ differ by gender and by race?

It is important to remember there is not ‘a’ return to education. Our estimates for any measure of the return to education will depend on, at least, three factors. The first is what we control for in the regression, the second is the choice of functional form and the third is how we address the endogeneity issue that education may be correlated with some unobserved variables in the error term. The third of these issues is the one to which most attention has been given in research on this topic and one method of addressing the problem is to use instrumental variables which will be introduced in Chapter 11.

In the runs below we control for experience but present both the linear and the non-linear specification. The specification below interacts the African dummy with the education variable. Below the regression we calculate the rates of return by gender and race.

As you will see the returns clearly do differ by gender and by race. While black South Africans are paid

|  |
| --- |
|  |
| . reg logwphy educ exper exper\_sq educ\_african educ\_male african male  Source | SS df MS Number of obs = 5891  -------------+------------------------------ F( 7, 5883) = 763.37  Model | 3571.81955 7 510.259936 Prob > F = 0.0000  Residual | 3932.39818 5883 .668434163 R-squared = 0.4760  -------------+------------------------------ Adj R-squared = 0.4754  Total | 7504.21773 5890 1.27406074 Root MSE = .81758  ------------------------------------------------------------------------------  logwphy | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  educ | .059705 .007065 8.45 0.000 .0458551 .073555  exper | .0486841 .003375 14.43 0.000 .0420679 .0553002  exper\_sq | -.0006202 .0000622 -9.97 0.000 -.0007422 -.0004982  educ\_african | .0849157 .0072098 11.78 0.000 .0707819 .0990495  educ\_male | -.0171486 .0048901 -3.51 0.000 -.026735 -.0075622  african | -1.972546 .0823292 -23.96 0.000 -2.133942 -1.811151  male | .5177499 .0444937 11.64 0.000 .4305259 .6049739  \_cons | 1.204856 .0936375 12.87 0.000 1.021292 1.38842  ------------------------------------------------------------------------------  .  . gen ror\_black\_male=\_b[educ]+\_b[educ\_african]+\_b[educ\_male] if e(sample)==1  (18616 missing values generated)  . gen ror\_white\_male=\_b[educ]+\_b[educ\_male] if e(sample)==1  (18616 missing values generated)  . gen ror\_black\_female=\_b[educ]+\_b[educ\_african] if e(sample)==1  (18616 missing values generated)  . gen ror\_white\_female=\_b[educ] if e(sample)==1  (18616 missing values generated)  .  . sum ror\*  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  ror\_black\_male | 5891 .1274721 0 .1274721 .1274721  ror\_white\_male | 5891 .0425564 0 .0425564 .0425564  ror\_black\_female | 5891 .1446207 0 .1446207 .1446207  ror\_white\_female | 5891 .059705 0 .059705 .059705  . |

much less than white South African, if we condition on education and experience, the returns to education are much higher for black than for White South Africans.

6. Is education or trade a more significant (in both the statistical and economic sense of this term) determinant of productivity in the macro production function?

In answering this question you need to decide on the functional form. We have for reasons already covered argued for the semi-logarithmic specification for how education is related to productivity. How should trade be modelled? The data we are given taken from the PENN World Tables 6.1 is for the share of imports and exports in GDP. This is a measure of how open the economy is in the sense that the higher is the number the greater the openness to trade of the economy. The table below shows the summary statistics for the variables.

|  |
| --- |
| Average years of Education in Population aged over 15  -------------------------------------------------------------  Percentiles Smallest  1% .535 .261  5% 1.018 .535  10% 2.172 .545 Obs 164  25% 3.4255 .764 Sum of Wgt. 164  50% 5.412 Mean 5.674268  Largest Std. Dev. 2.888071  75% 7.672 11.737  90% 9.604 11.848 Variance 8.340953  95% 10.837 11.865 Skewness .2431177  99% 11.865 12.049 Kurtosis 2.295535  Openness (Imports+Exports/GDP in current prices)  -------------------------------------------------------------  Percentiles Smallest  1% 11.71014 11.50686  5% 20.69277 11.71014  10% 30.27279 15.34849 Obs 164  25% 44.26226 17.18601 Sum of Wgt. 164  50% 62.62534 Mean 69.2741  Largest Std. Dev. 39.09166  75% 86.94586 175.5566  90% 112.5906 181.3853 Variance 1528.158  95% 129.5876 230.3342 Skewness 2.0024  99% 230.3342 295.1855 Kurtosis 10.45278 |

These statistics will be important when we come to assess the economic significance of our econometric results. So what should be our assumption about functional form for the trade variable and indeed is the trade variable the one we want?

In the Table below we show histogram for both the levels and 20 year differences of our trade variable. In the top left hand corner is the openc Penn measure in the top right is its log. In the bottom left is the 20 year difference of the openness measure and in the bottom right the 20 year difference of the log on the openness measure.

|  |  |
| --- | --- |
|  |  |
|  |  |

The logarithmic transformation has some effect in making the distribution closer to the normal and given the small sample size we will use the logarithmic transformation which will mean our coefficient can be treated as an elasticity. With that in mind we now report the summary statistics for the logarithmic transformation.

|  |
| --- |
| . sum ln\_openc d20\_ln\_openc, d  ln\_openc  -------------------------------------------------------------  Percentiles Smallest  1% 2.460455 2.442943  5% 3.029784 2.460455  10% 3.410249 2.731017 Obs 164  25% 3.790123 2.844095 Sum of Wgt. 164  50% 4.13717 Mean 4.095213  Largest Std. Dev. .5498596  75% 4.465285 5.167961  90% 4.723758 5.200624 Variance .3023456  95% 4.864357 5.439531 Skewness -.3574651  99% 5.439531 5.687604 Kurtosis 3.59281    d20\_ln\_openc  -------------------------------------------------------------  Percentiles Smallest  1% -.6369031 -.6369031  5% -.3197879 -.4860824  10% -.2126947 -.3967283 Obs 82  25% -.0062264 -.3280003 Sum of Wgt. 82  50% .1945864 Mean .2209396  Largest Std. Dev. .3978757  75% .399151 .8508254  90% .6707603 1.162873 Variance .1583051  95% .8207166 1.418321 Skewness 1.217643  99% 1.907923 1.907923 Kurtosis 6.501405 |

The regressions are given in the table below:

|  |
| --- |
| /\*Runs to answer Chapter 5 Exercise Question 6\*/  . reg lrgdpch lkp tyr15 ln\_openc time,robust  Linear regression Number of obs = 164  F( 4, 159) = 402.46  Prob > F = 0.0000  R-squared = 0.9201  Root MSE = .32396  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .5715609 .040452 14.13 0.000 .4916684 .6514534  tyr15 | .0808485 .0221879 3.64 0.000 .0370274 .1246696  ln\_openc | .053018 .0430771 1.23 0.220 -.0320591 .138095  time | -.074588 .0501815 -1.49 0.139 -.1736964 .0245203  \_cons | 2.537035 .3317645 7.65 0.000 1.881801 3.192269  ------------------------------------------------------------------------------  . reg d20\_lrgdpch d20\_lkp d20\_tyr15 d20\_ln\_openc,robust  Linear regression Number of obs = 82  F( 3, 78) = 35.23  Prob > F = 0.0000  R-squared = 0.5858  Root MSE = .22908  ------------------------------------------------------------------------------  | Robust  d20\_lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  d20\_lkp | .5043905 .0745621 6.76 0.000 .3559487 .6528323  d20\_tyr15 | .0417427 .0348874 1.20 0.235 -.0277128 .1111982  d20\_ln\_openc | .1619756 .0701848 2.31 0.024 .0222485 .3017027  \_cons | -.0186678 .054976 -0.34 0.735 -.1281167 .0907811  ------------------------------------------------------------------------------  . reg d20\_lrgdpch d20\_tyr15 d20\_ln\_openc,robust  Linear regression Number of obs = 82  F( 2, 79) = 5.40  Prob > F = 0.0063  R-squared = 0.1564  Root MSE = .32484  ------------------------------------------------------------------------------  | Robust  d20\_lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  d20\_tyr15 | .1123064 .0482326 2.33 0.022 .0163018 .208311  d20\_ln\_openc | .277294 .1171291 2.37 0.020 .0441544 .5104337  \_cons | .0565737 .0710219 0.80 0.428 -.0847919 .1979392  ------------------------------------------------------------------------------ |

There are several points that need to be noted about these regression results. The first simply exploits the cross-section variation in the data. In doing so it controls for time which is, in this specification, a measure of total factor productivity. While this is not significantly different from zero the point estimate is negative which is rather surprising. You need to consider why this may have arisen.

The second regression is a cross section of differences. As we will see later in the book (Section III) such differencing is a way of removing the time invariant unobservables from the cross-section regression. Intuitively, if there is some unobserved factor affecting both inputs and output – say the quality of skills in the economy – then this unobserved factor will be biasing our point estimates. Differencing the data removes in fact all unobservable factors that are time invariant. Thus, this regression has a very different set of controls as a result of the differencing.

As you will see the point estimates change markedly as a result of the differencing. In the differenced specification the education term is not significantly different from zero at conventional levels and the point estimate is halved. In contrast the point estimate of trade increases from 0.05 to 0.16 and the estimate is now significantly different from zero at the 5 per sent significance level.

In the final regression shown in the table we drop the capital variable. One interpretation of the resulting regression is that it shows the total effect of education and trade on GDP, some of which effect may operate through changing the stock of physical capital. As you will see the point estimates now rise and both variables are significant.

The question asks you not only about the statistical significance but also the relative economic importance of trade and education. To answer that part of the question you need to consider how large a change in those variables is shown in the data.

For education a move from one standard deviation below the means to one standard deviation above the mean is an increase in the average years of education of about six years. Our point estimates suggest a range of exp ((0.08 or 0.04)\*6)-1, ie a range from 61 per cent to 27 per cent increase in GDP when we condition on the physical capital stock.

For our measure of trade as both dependent and independent variable are in logs our estimates in the table are elasticities which imply that a 1 percentage increase in trade openness increases GDP, again controlling for physical capital input, by between 0.05 and 0.16 per cent which might sound small. However, note that there are substantial changes in the trade measure

The estimates for both education and trade need to be seen in the context of the range of GDP per capita shown in the data. The summary statistics are given below.

|  |
| --- |
| Ln of Real GDP in US$(1996 PPP)  -------------------------------------------------------------  Percentiles Smallest  1% 6.103621 6.093689  5% 6.7689 6.103621  10% 6.907778 6.17768 Obs 164  25% 7.518515 6.40638 Sum of Wgt. 164  50% 8.471683 Mean 8.480849  Largest Std. Dev. 1.131744  75% 9.566365 10.19236  90% 10.01333 10.20005 Variance 1.280843  95% 10.11767 10.20583 Skewness -.1588518  99% 10.20583 10.4131 Kurtosis 1.927214 |

If we go from one standard deviation below to one standard deviation above there is a 9 fold increase in GDP per capita (exp(2.2)). The implication is that neither education nor trade are explaining very much of this variation. It is the physical capital stock which is far more closely correlated with GDP per capita.

Does this mean that capital per capita ‘explains’ differences in GDP per capita while trade and human capital does not? The answer to that question is yes if by ‘explain’ we mean high correlation in that by far the greater part of the variance of GDP per capita is correlated with physical capital per capita. The answer to the question is no if we wish to understand the underlying factors that drive GDP per capita.

What we have sought to identify here is a production function. You can think of that as the ‘how’ of how GDP per capita is changed (with the qualification of course which is the answer to the next question that we have addressed endogeneity issues). It does not answer the ‘what’ question in the sense of ‘What factors drive GDP per capita’ as we then need to know not only what drives physical capital but also the other determinants such as human capital or trade.

It is common to seek an answer to the ‘what’ question by simply regressing GDP per capita on what are seen as such fundamental determinants of GDP. Examples include ‘institutions’ and ‘trust’ or, in the case of Africa slavery. If such regression too can address the endogeneity issues that arise, a big and important qualification, they answer the ‘what’ question. However there remains the ‘how’ question and for that the production function will be useful.

7. Discuss whether any of your regressions can be given a causal interpretation.

The (very) short answer to that question, which has already been indicated in the outline answer given to the last question, is no. The reason is that to make any causal statement in the sense that economists mean the term we need the zero conditional mean assumption which has been stressed when discussing the interpretation of regression results.

The slightly longer answer is that the issues that arise of endogeneity differ somewhat between micro and macro data sets. At the macro level it can make sense for example to think of labour as exogenous in a way that would not make sense for micro data. A firm clearly chooses how much labour it will employ. While, in some sense, total labour supply is in the long run endogenous to how fast the economy grows, in the short term it may not be seriously misleading to treat it as exogenous. The same, at the macro level, will probably not be true of the capital stock which changes as a result of investment which may well be driven by income and by expected changes in income.

In summary addressing causality issues requires us to address the zero conditional mean assumption and how that can be done is very much the subject of the later parts of the book.