**Exercise Chapter 6**

This Stata do file generates the random walks shown in this chapter. Run this program and answer the questions below:

clear

set mem 200m

set more off

capture log close

log using "*your address*\spurious.log", replace

\*Base model: without time trend\*

set mem 200m

set matsize 300

set obs 500

set seed 3533

gen eps1a=invnorm(uniform())

gen eps2a=invnorm(uniform())

sum eps1a eps2a

gen lylsa=0 if \_n==1

gen lklsa=0 if \_n==1

local X=2

while `X'<=500 {

qui replace lylsa=lylsa[\_n-1]+eps1a if \_n==`X'

qui replace lklsa=lklsa[\_n-1]+eps2a if \_n==`X'

local X=`X'+1

}

drop if \_n==1

reg lylsa lklsa

gen time=\_n

scatter lylsa lklsa time

exit

**Part A:**

1. Control for time in the regression reg lylsa lklsa. Why does this control not solve the problem posed by the random walk?

In the Table below we control for time and there still appears to be a relationship between the two variables, even though we know, as we have created the data, that there is no such relationship. Why does not allowing for time correct the problem? The answer is that the correlation we observe is not due to time. It is due to the fact that two random walks (note that time play no role in the generation of the variables that is entirely due to a stochastic effect on the change in the variable) will look as though they are correlated as both will be trending variables. See the chart below the regressions.

|  |
| --- |
| . reg lylsa lklsa  Source | SS df MS Number of obs = 499  -------------+------------------------------ F( 1, 497) = 222.62  Model | 16030.6243 1 16030.6243 Prob > F = 0.0000  Residual | 35788.8259 497 72.00971 R-squared = 0.3094  -------------+------------------------------ Adj R-squared = 0.3080  Total | 51819.4502 498 104.055121 Root MSE = 8.4859  ------------------------------------------------------------------------------  lylsa | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lklsa | -.8254878 .0553262 -14.92 0.000 -.9341899 -.7167857  \_cons | 6.221213 1.211204 5.14 0.000 3.8415 8.600925  ------------------------------------------------------------------------------  . gen time=\_n  . reg lylsa lklsa time  Source | SS df MS Number of obs = 499  -------------+------------------------------ F( 2, 496) = 330.33  Model | 29598.1015 2 14799.0508 Prob > F = 0.0000  Residual | 22221.3486 496 44.8011061 R-squared = 0.5712  -------------+------------------------------ Adj R-squared = 0.5694  Total | 51819.4502 498 104.055121 Root MSE = 6.6934  ------------------------------------------------------------------------------  lylsa | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lklsa | .352561 .0805422 4.38 0.000 .194315 .5108071  time | -.0668089 .0038391 -17.40 0.000 -.0743518 -.059266  \_cons | -1.565194 1.054945 -1.48 0.139 -3.637906 .5075188  ------------------------------------------------------------------------------  . scatter lylsa lklsa time |



2. Add a time trend of 0.01 to both equations and establish how correlated are lyls and lkls.

## The answer is best seen in a figure:

## 

## The stata code that gives this figure is as follows (it is given to you in the do file for this exercise):

|  |
| --- |
| replace lylsa=0 if \_n==1replace lklsa=0 if \_n==1local X=2while `X'<=500 {qui replace lylsa=lylsa[\_n-1]+eps1a+0.01\*time if \_n==`X'qui replace lklsa=lklsa[\_n-1]+eps2a+0.01\*time if \_n==`X'local X=`X'+1}reg lylsa lklsareg lylsa lklsa timescatter lylsa lklsa time |

The output of this stata code is:

|  |
| --- |
| . reg lylsa lklsa  Source | SS df MS Number of obs = 500  -------------+------------------------------ F( 1, 498) = .  Model | 67040296.1 1 67040296.1 Prob > F = 0.0000  Residual | 33256.6408 498 66.7804033 R-squared = 0.9995  -------------+------------------------------ Adj R-squared = 0.9995  Total | 67073552.7 499 134415.937 Root MSE = 8.1719  ------------------------------------------------------------------------------  lylsa | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lklsa | .966222 .0009643 1001.94 0.000 .9643273 .9681167  \_cons | -16.8034 .5599332 -30.01 0.000 -17.90353 -15.70328  ------------------------------------------------------------------------------  . reg lylsa lklsa time  Source | SS df MS Number of obs = 500  -------------+------------------------------ F( 2, 497) = .  Model | 67048835.5 2 33524417.8 Prob > F = 0.0000  Residual | 24717.19 497 49.7327766 R-squared = 0.9996  -------------+------------------------------ Adj R-squared = 0.9996  Total | 67073552.7 499 134415.937 Root MSE = 7.0521  ------------------------------------------------------------------------------  lylsa | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lklsa | 1.010778 .0035006 288.74 0.000 1.0039 1.017656  time | -.1204385 .0091912 -13.10 0.000 -.1384969 -.1023801  \_cons | -6.233952 .9402629 -6.63 0.000 -8.081332 -4.386572  ------------------------------------------------------------------------------ |

As you can see the *t* statistics is exploding. The two series seem to be highly correlated.

3. Does adding a time trend to the regression solve the problem in this case? If not, why not?

The regression in the box above shows that including time does not solve the problem. It is true time is now playing a role (we are making it do so) but the random walk of the two series dominates our results and is the source of the correlation. To repeat time is not the source of the correlation.

4. Present results which show that these variables are, in fact, unrelated?

|  |
| --- |
| . tsset time  time variable: time, 1 to 500  delta: 1 unit  . reg d.lylsa d.lklsa time  Source | SS df MS Number of obs = 499  -------------+------------------------------ F( 2, 496) = 545.63  Model | 1024.54271 2 512.271357 Prob > F = 0.0000  Residual | 465.673601 496 .938858067 R-squared = 0.6875  -------------+------------------------------ Adj R-squared = 0.6863  Total | 1490.21631 498 2.99240224 Root MSE = .96895  ------------------------------------------------------------------------------  D.lylsa | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lklsa |  D1. | -.0487213 .0432497 -1.13 0.260 -.1336965 .0362538  |  time | .0104065 .0005109 20.37 0.000 .0094027 .0114104  \_cons | -.0261456 .0874172 -0.30 0.765 -.1978993 .1456081  ------------------------------------------------------------------------------ |

5. Derive the residuals from the models and test if these residual are autocorrelated.

|  |
| --- |
| . reg lylsa lklsa time  Source | SS df MS Number of obs = 500  -------------+------------------------------ F( 2, 497) = 332.27  Model | 29715.124 2 14857.562 Prob > F = 0.0000  Residual | 22223.7391 497 44.7157728 R-squared = 0.5721  -------------+------------------------------ Adj R-squared = 0.5704  Total | 51938.8631 499 104.085898 Root MSE = 6.687  ------------------------------------------------------------------------------  lylsa | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lklsa | .350179 .0798032 4.39 0.000 .1933857 .5069722  time | -.0667503 .0038271 -17.44 0.000 -.0742695 -.0592311  \_cons | -1.460505 1.039487 -1.41 0.161 -3.502836 .5818269  ------------------------------------------------------------------------------  . predict res\_1, resid  . reg res\_1 l.res\_1, robust  Linear regression Number of obs = 499  F( 1, 497) =19708.91  Prob > F = 0.0000  R-squared = 0.9756  Root MSE = 1.0446  ------------------------------------------------------------------------------  | Robust  res\_1 | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  res\_1 |  L1. | .9878276 .0070364 140.39 0.000 .9740029 1.001652  |  \_cons | .0022187 .0467654 0.05 0.962 -.0896635 .0941009  ------------------------------------------------------------------------------  . dfuller res\_1, trend regress lags(2)  Augmented Dickey-Fuller test for unit root Number of obs = 497  ---------- Interpolated Dickey-Fuller ---------  Test 1% Critical 5% Critical 10% Critical  Statistic Value Value Value  ------------------------------------------------------------------------------  Z(t) -1.778 -3.980 -3.420 -3.130  ------------------------------------------------------------------------------  MacKinnon approximate p-value for Z(t) = 0.7154  ------------------------------------------------------------------------------  D.res\_1 | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  res\_1 |  L1. | -.0125944 .0070846 -1.78 0.076 -.0265143 .0013255  LD. | .038977 .0453033 0.86 0.390 -.0500348 .1279888  L2D. | -.0053344 .0455593 -0.12 0.907 -.0948492 .0841805  \_trend | .0000969 .000328 0.30 0.768 -.0005476 .0007414  \_cons | -.0224347 .0948233 -0.24 0.813 -.2087432 .1638738  ------------------------------------------------------------------------------ |

|  |
| --- |
| . tsset time  time variable: time, 1 to 500  delta: 1 unit  . reg d.lylsa d.lklsa time  Source | SS df MS Number of obs = 499  -------------+------------------------------ F( 2, 496) = 545.63  Model | 1024.54271 2 512.271357 Prob > F = 0.0000  Residual | 465.673601 496 .938858067 R-squared = 0.6875  -------------+------------------------------ Adj R-squared = 0.6863  Total | 1490.21631 498 2.99240224 Root MSE = .96895  ------------------------------------------------------------------------------  D.lylsa | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lklsa |  D1. | -.0487213 .0432497 -1.13 0.260 -.1336965 .0362538  |  time | .0104065 .0005109 20.37 0.000 .0094027 .0114104  \_cons | -.0261456 .0874172 -0.30 0.765 -.1978993 .1456081  ------------------------------------------------------------------------------  . predict res\_2, resid  (1 missing value generated)  . dfuller res\_2, trend regress lags(2)  Augmented Dickey-Fuller test for unit root Number of obs = 496  ---------- Interpolated Dickey-Fuller ---------  Test 1% Critical 5% Critical 10% Critical  Statistic Value Value Value  ------------------------------------------------------------------------------  Z(t) -12.510 -3.980 -3.420 -3.130  ------------------------------------------------------------------------------  MacKinnon approximate p-value for Z(t) = 0.0000  ------------------------------------------------------------------------------  D.res\_2 | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  res\_2 |  L1. | -.945139 .0755505 -12.51 0.000 -1.093581 -.7966968  LD. | .00765 .0620077 0.12 0.902 -.1141833 .1294833  L2D. | -.0199857 .0454196 -0.44 0.660 -.1092266 .0692551  \_trend | .0000353 .0003043 0.12 0.908 -.0005625 .0006332  \_cons | -.0121922 .087796 -0.14 0.890 -.1846945 .16031  ------------------------------------------------------------------------------ |

6. Using relevant Dickey-Fuller tests establish if these are spurious regressions.

These are given as part of the answer to the last question. The residuals from the levels regression are clearly a random walk while that is not the case for the differenced regression.

**Part B:**

The following are two macro production functions; the first is a static model, the second a dynamic model with a lagged dependent variable:

Static Model:

Dynamic Model:

*ln(y)* is the natural log of GDP per capita and *ln(k)* is the natural log of capital per capita, *Year* is the trend variable.

The data for this exercise can be found in ‘Macro\_PEBLIF’. This macro data set is a merged file with data from PENN6.1, Barro and Lee (2000), the IMF Financial Statistics and measures of political and civil rights from the Freedom House annual ‘Freedom in the World’ survey over the period 1950 to 2000.

1. Estimate both the static and dynamic models using the time series data over the period from 1950 to 2000 from ‘Macro\_PEBLIF’ for Ghana, Argentina, South Korea and Australia.

|  |
| --- |
| .  . tsset nwbcode year  panel variable: nwbcode (unbalanced)  time variable: year, 1950 to 2000  delta: 1 unit  . reg lrgdpch lkp year if country=="GHANA",robust  Linear regression Number of obs = 36  F( 2, 33) = 2.34  Prob > F = 0.1116  R-squared = 0.1788  Root MSE = .10307  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .7942225 .379288 2.09 0.044 .0225553 1.56589  year | .018905 .0087532 2.16 0.038 .0010965 .0367135  \_cons | -36.57828 20.26778 -1.80 0.080 -77.81338 4.656823  ------------------------------------------------------------------------------  . reg lrgdpch lkp year if country=="AUSTRALIA", robust  Linear regression Number of obs = 36  F( 2, 33) = 662.75  Prob > F = 0.0000  R-squared = 0.9782  Root MSE = .02943  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .3917586 .1281314 3.06 0.004 .1310732 .6524439  year | .0101316 .0026492 3.82 0.001 .0047419 .0155213  \_cons | -14.64836 3.863365 -3.79 0.001 -22.50844 -6.788284  ------------------------------------------------------------------------------  . reg lrgdpch lkp year if country=="ARGENTINA", robust  Linear regression Number of obs = 36  F( 2, 33) = 7.54  Prob > F = 0.0020  R-squared = 0.1690  Root MSE = .10767  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .2437151 .1793201 1.36 0.183 -.1211143 .6085445  year | .0012622 .0031984 0.39 0.696 -.0052449 .0077693  \_cons | 4.145156 4.749613 0.87 0.389 -5.518005 13.80832  ------------------------------------------------------------------------------  . reg lrgdpch lkp year if country=="KOREA, REP.",robust  Linear regression Number of obs = 36  F( 2, 33) = 6635.06  Prob > F = 0.0000  R-squared = 0.9955  Root MSE = .04593  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lkp | .6140602 .1389213 4.42 0.000 .3314227 .8966976  year | .0051243 .0132718 0.39 0.702 -.0218774 .032126  \_cons | -7.332347 24.99232 -0.29 0.771 -58.1796 43.51491  ------------------------------------------------------------------------------  .  . reg lrgdpch l.lrgdpch lkp year if country=="GHANA",robust  Linear regression Number of obs = 36  F( 3, 32) = 7.91  Prob > F = 0.0004  R-squared = 0.5570  Root MSE = .07687  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lrgdpch |  L1. | .6875792 .1548702 4.44 0.000 .3721189 1.003039  |  lkp | .2011699 .3900436 0.52 0.610 -.593323 .9956628  year | .0049745 .0091812 0.54 0.592 -.0137269 .0236759  \_cons | -9.210153 20.95464 -0.44 0.663 -51.89336 33.47305  ------------------------------------------------------------------------------  . gen lr\_lkp\_gha=\_b[lk]/(1-\_b[l.lrgdpch])  . gen lr\_year\_gha=\_b[year]/(1-\_b[l.lrgdpch])  .  . reg lrgdpch l.lrgdpch lkp year if country=="AUSTRALIA", robust  Linear regression Number of obs = 36  F( 3, 32) = 1230.04  Prob > F = 0.0000  R-squared = 0.9892  Root MSE = .021  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lrgdpch |  L1. | .7799375 .0871064 8.95 0.000 .6025076 .9573673  |  lkp | -.0483535 .1273699 -0.38 0.707 -.3077975 .2110905  year | .0049932 .0024312 2.05 0.048 .0000411 .0099453  \_cons | -7.191459 3.48596 -2.06 0.047 -14.29213 -.090791  ------------------------------------------------------------------------------  . gen lr\_lkp\_aus=\_b[lk]/(1-\_b[l.lrgdpch])  . gen lr\_year\_aus=\_b[year]/(1-\_b[l.lrgdpch])  .  . reg lrgdpch l.lrgdpch lkp year if country=="ARGENTINA", robust  Linear regression Number of obs = 36  F( 3, 32) = 56.28  Prob > F = 0.0000  R-squared = 0.7709  Root MSE = .05741  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lrgdpch |  L1. | .8814115 .106229 8.30 0.000 .6650301 1.097793  |  lkp | -.222252 .1338548 -1.66 0.107 -.4949053 .0504014  year | .002881 .0017741 1.62 0.114 -.0007328 .0064947  \_cons | -2.331379 2.301809 -1.01 0.319 -7.020012 2.357253  ------------------------------------------------------------------------------  . gen lr\_lkp\_arg=\_b[lk]/(1-\_b[l.lrgdpch])  . gen lr\_year\_arg=\_b[year]/(1-\_b[l.lrgdpch])  .  . reg lrgdpch l.lrgdpch lkp year if country=="KOREA, REP.",robust  Linear regression Number of obs = 36  F( 3, 32) = 4406.61  Prob > F = 0.0000  R-squared = 0.9968  Root MSE = .03907  ------------------------------------------------------------------------------  | Robust  lrgdpch | Coef. Std. Err. t P>|t| [95% Conf. Interval]  -------------+----------------------------------------------------------------  lrgdpch |  L1. | .5995471 .167996 3.57 0.001 .2573506 .9417437  |  lkp | .240774 .2071211 1.16 0.254 -.181118 .6626659  year | .0021343 .0129953 0.16 0.871 -.0243362 .0286048  \_cons | -3.013887 24.62976 -0.12 0.903 -53.18307 47.1553  ------------------------------------------------------------------------------  . gen lr\_lkp\_kor=\_b[lk]/(1-\_b[l.lrgdpch])  . gen lr\_year\_kor=\_b[year]/(1-\_b[l.lrgdpch])  .  . sum lr\_\*  Variable | Obs Mean Std. Dev. Min Max  -------------+--------------------------------------------------------  lr\_lkp\_gha | 10208 .6439069 0 .6439069 .6439069  lr\_year\_gha | 10208 .0159225 0 .0159225 .0159225  lr\_lkp\_aus | 10208 -.219726 0 -.219726 -.219726  lr\_year\_aus | 10208 .02269 0 .02269 .02269  lr\_lkp\_arg | 10208 -1.874144 0 -1.874144 -1.874144  -------------+--------------------------------------------------------  lr\_year\_arg | 10208 .0242938 0 .0242938 .0242938  lr\_lkp\_kor | 10208 .6012542 0 .6012542 .6012542  lr\_year\_kor | 10208 .0053298 0 .0053298 .0053298 |

2. What are the assumptions that ensure the OLS estimates are unbiased and consistent and how can they be tested?

The answer to this question can be found in Section 6.4 pp81ff in EDE. Two important points distinguish cross section from time series data. The first is the source of any trend we find in the data and the second is how strong is the strict exogeneity assumption that ensures our OLS estimates are unbiased. The conditions for consistency are discussed on pages 87ff.

3. What are the short and long run coefficients on capital and technical progress for both South Korea and Ghana?

These can be seen in the table above.

4. Establish the time series properties of the arguments for the production function

This is done for you in the stata do file ‘Exercise\_Chapter\_6\_Part B.do’

5. Estimate the constant returns to scale production function and assess if the variables are cointegrated.

Again this is done in the stata do file ‘Exercise\_Chapter\_6\_Part\_B.do’. You will find that for all four countries the GDP per capita and Capital per capita are not cointegrated. This implies that the production function regression of GDP per capita on Capital per capita is likely to be a spurious regression. Should you conclude from this that capital does not play a role in determining output? Clearly not. Something is going wrong and what that something is needs to be tracked down. The most obvious problem is in the nature of the macro data we are using. Rather than use the PENN comparative data a place to start would be in the use of the time series from the national accounts. It would then be necessary to investigate the specific factors that may underlie the behaviour of the data series. While such analysis means that the general simplicity of our aggregate production function is lost, there is little virtue in erroneous simplicity.

The four countries which you have been asked to analyse were chosen for a reason. They represent four kinds of economy which have had very different growth experiences since the 1960s. Countries which have been poor over a long period of time have either never grow sufficiently fast to drive up levels of per capita income, China and India before the 1980s are the most important examples of such economies, or growth has begun and not been sustained of which both Ghana and Argentina are good examples. In contrast South Korea which in the 1950s was at the same per capita income level as Ghana has grown rapidly with only very small declines in the late 1970s and late 1990s. This suggests that to understand differing pattern of long run income change we need to understand not only how the process begins but how it can be sustained.

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