Time series (1)

1. Regression between two variables that have a trend: spurious correlation

Source: EARNS in Wooldridge. *Economic Report of the President*, 1989, Table B-47. The data are for the nonfarm business sector.

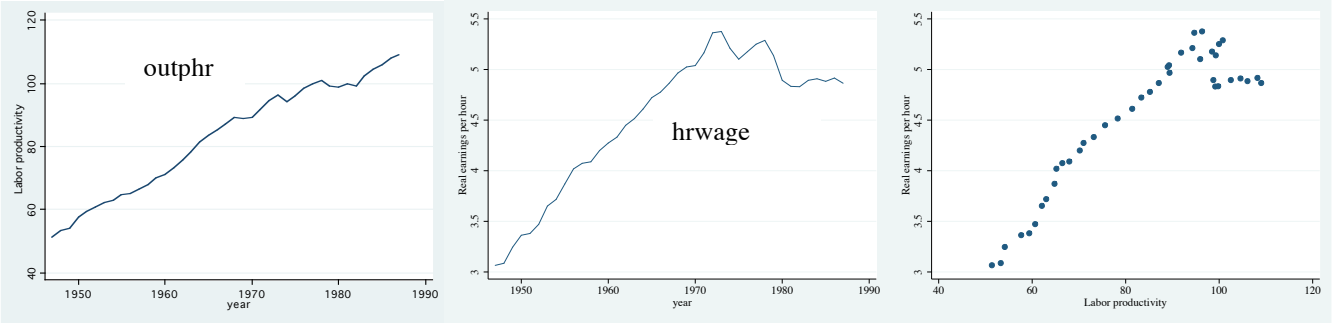
```
  1. year                    1947 to 1987
  4. outphr                  output per labor hour
  5. hrwage                  average real earnings / hour
```

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/EARNS
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
year	41	1967	11.97915	1947	1987
outphr	41	82.91463	17.5342	51.4	109
hrwage	41	4.509252	.6854282	3.065012	5.37533941

Looking at observations before 1972, both series have a regular trend.
If you represent a scatter diagram hourly wage against labor productivity), you may have a spurious correlation:

(`twoway scatter hrwage outphr`




```
. reg hrwage outph if year<=1972
```

Source	SS	df	MS		
Model	11.4674843	1	11.4674843	Number of obs	= 26
Residual	.234573166	24	.009773882	F(1, 24)	= 1173.28
Total	11.7020574	25	.468082297	Prob > F	= 0.0000
				R-squared	= 0.9800
				Adj R-squared	= 0.9791
				Root MSE	= .09886

hrwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
outphr	.0511744	.001494	34.25	0.000	.0480909	.0542578
_cons	.492528	.1100935	4.47	0.000	.2653061	.7197498

```
. reg hrwage outph
```

Source	SS	df	MS		
Model	15.8941543	1	15.8941543	Number of obs	= 41
Residual	2.89831975	39	.074315891	F(1, 39)	= 213.87
Total	18.7924741	40	.469811852	Prob > F	= 0.0000
				R-squared	= 0.8458
				Adj R-squared	= 0.8418
				Root MSE	= .27261

hrwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
outphr	.0359503	.0024582	14.62	0.000	.030978	.0409226
_cons	1.528445	.2082233	7.34	0.000	1.107274	1.949616

a. You could add the (omitted) time variable

```
. reg hrwage outphr t if year<=1972

      Source |       SS       df       MS              Number of obs =      26
-------------+------------------------------           F(  2,    23) = 2104.28
       Model |  11.6384528     2  5.81922639           Prob > F      =  0.0000
    Residual |  .063604645    23  .002765419           R-squared     =  0.9946
-------------+------------------------------           Adj R-squared =  0.9941
       Total |  11.7020574    25  .468082297           Root MSE      =  .05259

------------------------------------------------------------------------------
      hrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      outphr |   -.010874   .0079313    -1.37   0.184    -.0272811    .0055331
           t |   .1079083   .0137239     7.86   0.000     .0795183    .1362983
       _cons |   3.536657   .3915594     9.03   0.000     2.726654    4.346659
------------------------------------------------------------------------------
```

b. General method is to use first difference:

```
. sort year
. g dwage=hrwage-hrwage[_n-1]
(1 missing value generated)
. g doutphr=outphr-outphr[_n-1]
(1 missing value generated)
. reg dwage doutphr if year<=1972

      Source |       SS       df       MS              Number of obs =      25
-------------+------------------------------           F(  1,    23) =    1.39
       Model |  .003598404     1  .003598404           Prob > F      =  0.2497
    Residual |  .059336242    23  .002579837           R-squared     =  0.0572
-------------+------------------------------           Adj R-squared =  0.0162
       Total |  .062934646    24  .002622277           Root MSE      =  .05079

------------------------------------------------------------------------------
       dwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     doutphr |   .0133027   .0112637     1.18   0.250     -.009998    .0366034
       _cons |   .0688566   .0219951     3.13   0.005     .0233563    .1143568
------------------------------------------------------------------------------
```

This also works for time dependences that are not necessarily a regular sustained time trend, as in our example when we consider all the years:

```
. reg dwage doutphr

      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  1,    38) =   22.09
       Model |  .124788428     1  .124788428           Prob > F      =  0.0000
    Residual |  .214659545    38  .005648935           R-squared     =  0.3676
-------------+------------------------------           Adj R-squared =  0.3510
       Total |  .339447973    39  .008703794           Root MSE      =  .07516

------------------------------------------------------------------------------
       dwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     doutphr |   .0463143    .009854     4.70   0.000      .026366    .0662626
       _cons |  -.0217087   .0185087    -1.17   0.248    -.0591776    .0157602
------------------------------------------------------------------------------
```

```
. twoway scatter dwage doutphr
```