# Lecture 2: Regression analysis with time-series data

**Prof. dr. Wolter Hassink**
**Utrecht University School of Economics**
**w.h.j.hassink@uu.nl**

**Contents:**

- The nature of time series
- Interpretation of dynamic regression model
- Trends
- Seasonality
- Spurious regression
- Properties of OLS-estimators – consistent estimator
- Contemporaneous exogeneity
- Weak dependency
- Implications of a unit root model
- Autocorrelation

**Material:**

Wooldridge:
Chapter 10: 10.1, 10.2, 10.3, 10.5
Chapter 11: 11.1, 11.2
Chapter 12: 12.1, 12.2, 12.3, 12.4

## Motivation: economic intertemporal transmission mechanisms

- Central question: how long does it take before a change of a key economic variable has an influence on economic processes?

- For instance the interest increase by the ECB has influence on the economy. There are several intertemporal transmission mechanisms through increased lending, etc.

- Any further examples on transmission mechanisms in your master programme?

- Other example: effect of interest on inflation.

- Thus: A change of inflation in period t has an effect on unemployment in $t + 1$ (or $t + 2$), etc.

- What if a researcher ignores the intertemporal transmission mechanism? She specifies the following regression equation:

$$inflation_t = \beta_0 + \beta_1 interest_t + u_t$$

- RESULT 1: It leads to correlation of the error term over time: autocorrelation.
- RESULT 2: autocorrelation will lead to biased and inconsistent parameter estimates
- RESULT 3: t-statistics and F-statistics are wrong!

- What if a researcher correctly describes the intertemporal transmission mechanism? She specifies the following equation:

$$inflation_t = \beta_0 + \beta_1 interest_{t-1} + \beta_2 interest_{t-2} + u_t$$

- RESULT: there is hardly any autocorrelation.

# The nature of time series

## The nature of time series

*Aim: to introduce the specific features of a times-series data set as well as the time-series operators of Stata.*

- Differences between cross-sectional data and time-series data:
  - Logical (temporal) ranking of the observations. Time-series data do not stem from a random sample.
- A sequence of random variables indexed by time is referred to as a stochastic process or a time-series process $(y_1,...,y_t,...,y_n)$. The subscript refers to the particular time period.
- When a time-series data set is collected, only one possible outcome (realization) of the stochastic process may be obtained.
- Time series commands in Stata:

  Begin the do-file with the following command:
  - `tsset year` (*year* is in this data set the variable that denotes the time period; in other data sets a different variable name may be used)
  - Stata time-series operators for:

$$x_{t-1}, x_{t-2}, x_{t+1}, x_{t+2}, x_t - x_{t-1}, x_{t-1} - x_{t-2}$$

```
Operator  Meaning
---------------------------------
L.        lag (x_t-1)
L1.       lag (x_t-1)
L2.       2-period lag (x_t-2)
...
F.        lead (x_t+1)
F1.       lead (x_t+1)
F2.       2-period lead (x_t+2)
...
D.        difference (x_t - x_t-1)
D1.       difference (x_t - x_t-1)
L.D.      1-period lagged difference
           (x_t-1 - x_t-2)
```

## Definition: error term is i.i.d.: identically and independently distributed

Consider the linear regression equation

$$inflation_t = \beta_0 + \beta_1 interest_{t-1} + \beta_2 interest_{t-2} + u_t$$

We usually assume the error term $u_t$ is i.i.d.: identically and independently distributed:

| | |
|---|---|
| $E(u_t) = 0$ | the expected value of the error term is zero |
| $Var(u_t) = \sigma^2$ | a constant variance of the error term |
| $Cov(u_t, u_{t-1}) = 0$ | no association between error terms across time |

In case of autocorrelation (see below), the assumption that $u_t$ is i.i.d. is violated (*t*-statistics and *F*-statistics are wrong).

# Interpretation of dynamic regression model

## Interpretation of dynamic regression model

*Aim: to calculate the long-run effect of a model with both lagged dependent and lagged independent variables.*

- $y_t$: outcome of $y$ (e.g. inflation) in period $t$: contemporaneous variable
- $y_{t-1}$: lag of 1 period: outcome of $y$ in period $t$-1: lagged variable
- $y_{t+1}$: lead of 1 period: outcome of $y$ in period $t$+1: lead variable

- Dynamic model:

$$inf_t = \beta_0 + \beta_1 inf_{t-1} + \delta_1 unemp_t + \delta_2 unemp_{t-1} + u_t \quad t=1949,\ldots,2003$$

- It should be read as follows:

$$inf_{2003} = \beta_0 + \beta_1 inf_{2002} + \delta_1 unemp_{2003} + \delta_2 unemp_{2002} + u_{2003}$$
$$inf_{2002} = \beta_0 + \beta_1 inf_{2001} + \delta_1 unemp_{2002} + \delta_2 unemp_{2001} + u_{2002}$$
$$\ldots$$
$$inf_{1949} = \beta_0 + \beta_1 inf_{1948} + \delta_1 unemp_{1949} + \delta_2 unemp_{1948} + u_{1949}$$

- The right-hand side variables do not contain any lead variables as a result of causality (in other words, the current dependent variable cannot be explained by the future independent variables), but they may incorporate future expectations regardless.
- This model is referred to as finite distributed lag.
- The contemporaneous effect (or short-run effect) is the parameter that registers the effect of $unemp_t$ on $inf_t$
  - It refers to the same period $t$. Short-run effect: $\delta_1$
- Long-run effect:
  - Includes the change in the dependent variable in all periods as result of changes in the independent variables

- Long-run effect of *unemp* on *inf*: $\dfrac{(\delta_1+\delta_2)}{(1-\beta_1)}$
    - First, start with long-run values (equilibrium values)
    - $E(unemp_t)=E(unemp_{t-1})=E(unemp_{t-2})=...=unemp*$
      $E(inf_t)=E(inf_{t-1})=E(inf_{t-2})=...=inf*$

    - In the long run: equilibrium relationship:

$$inf* = \beta_0 + \beta_1 inf* + \delta_1 unemp* + \delta_2 unemp*$$

$$inf* - \beta_1 inf* = \beta_0 + (\delta_1+\delta_2)unemp*$$

$$(1-\beta_1)inf* = \beta_0 + (\delta_1+\delta_2)unemp*$$

$$inf* = \frac{\beta_0}{(1-\beta_1)} + \frac{(\delta_1+\delta_2)}{(1-\beta_1)}unemp*$$

- The long-run effect of *unemp** on *inf** is $\dfrac{(\delta_1+\delta_2)}{(1-\beta_1)}$

- Lags of dependent and independent variables were introduced because:
    - They have a clear economic interpretation
    - It help us to mitigate the statistical problems of autocorrelation (see last part of this lecture)

## A static regression model

**Example 1:** Data set: PHILLIPS.DTA

. tsset year

time variable: year, 1948 to 2003

delta: 1 unit

. reg inf unem

```
      Source |       SS       df       MS              Number of obs =      56
-------------+------------------------------           F(  1,    54) =    3.58
       Model |   31.599858     1   31.599858           Prob > F      =  0.0639
    Residual |  476.815691    54   8.8299202           R-squared     =  0.0622
-------------+------------------------------           Adj R-squared =  0.0448
       Total |  508.415549    55  9.24391907           Root MSE      =  2.9715
------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   .5023782   .2655624     1.89   0.064    -.0300424    1.034799
       _cons |   1.053566   1.547957     0.68   0.499    -2.049901    4.157033
------------------------------------------------------------------------------
```

- The effect of unemployment on inflation in the above model is 0.50.

- The model does not distinguish between long-run effects and short-run effects.

## A dynamic regression model

**Example 2:** Regression of inflation(*t*) on inflation(*t*-1), unemployment(*t*) and unemployment(*t*-1)

. reg inf l.inf unem l.unem

```
      Source |       SS       df       MS              Number of obs =      55
-------------+------------------------------           F(  3,    51) =   17.26
       Model |  246.994365     3  82.331455            Prob > F      =  0.0000
    Residual |  243.322736    51  4.77103403           R-squared     =  0.5037
-------------+------------------------------           Adj R-squared =  0.4746
       Total |  490.317101    54  9.07994631           Root MSE      =  2.1843
------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inf |
         L1. |   .7878368   .1199723     6.57   0.000     .5469823    1.028691
        unem |
         --. |  -.6746869   .3591467    -1.88   0.066    -1.395704    .0463301
         L1. |   .5117387    .308468     1.66   0.103    -.1075366    1.131014
       _cons |   1.668624   1.233995     1.35   0.182    -.8087242    4.145972
------------------------------------------------------------------------------
```

- Note that the estimated parameter on inflation(*t*-1) is statistically significant and that its value is 0.79. Furthermore, the estimated parameter on unemployment(*t*) and unemployment(*t-1*) are statistically significant at the 10-percent level (*p*-value of 0.103 is considered to be significant at 10-percent level).

- The short-run effect is: -0.675
- The long-run effect is: (-0.675 + 0.512)/(1 - 0.788) = -0.768
- However, the estimated effect is statistically insignificant according to the delta method (for delta method, see example 4 of lecture 1: the delta method is a transformation of the estimated parameters, so that the std.error of the transformation can be calculated):

```
. nlcom (_b[ unem] + _b[l.unem])/(1-_b[l.inf])

       _nl_1:  (_b[ unem] + _b[l.unem])/(1-_b[l.inf])
------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |  -.7680326   1.379379    -0.56   0.580    -3.537252    2.001187
------------------------------------------------------------------------------
```

# Trends

## Trends

*Aim: to introduce the use of a trend in a time-series model.*

- Many economic time series have a common tendency to grow over time.
- In order to draw causal inference, we must recognize that some time series contain a time trend.
- What kind of model adequately captures trending behaviour?
- Linear trend (see Figure below for *stockprice*) is the effect of the variable $t$ in an equation with a level variable as the dependent variable.

$stockprice_t = \alpha_0 + \alpha_1 t + e_t$ $\quad$ $t$=1,…,$n$ with

$Ee_t = 0, Var(e_t) = \sigma^2$

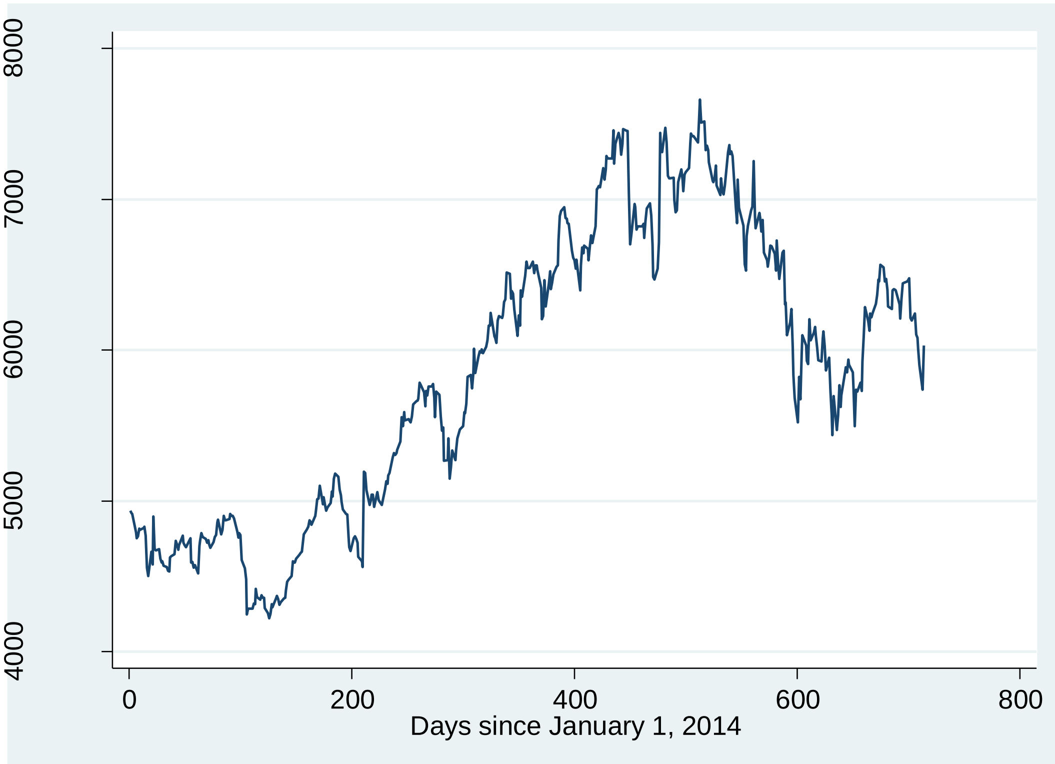
- Exponential trend (see Figure below for log(*stockprice*)). Is the effect of the variable $t$ in an equation with a logarithmic variable is the level variable.

$\log(stockprice_t) = \alpha_0 + \alpha_1 t + e_t$ $\quad$ $t$=1,…,$n$ with

$Ee_t = 0, Var(e_t) = \sigma^2$

# Example: Linear trend

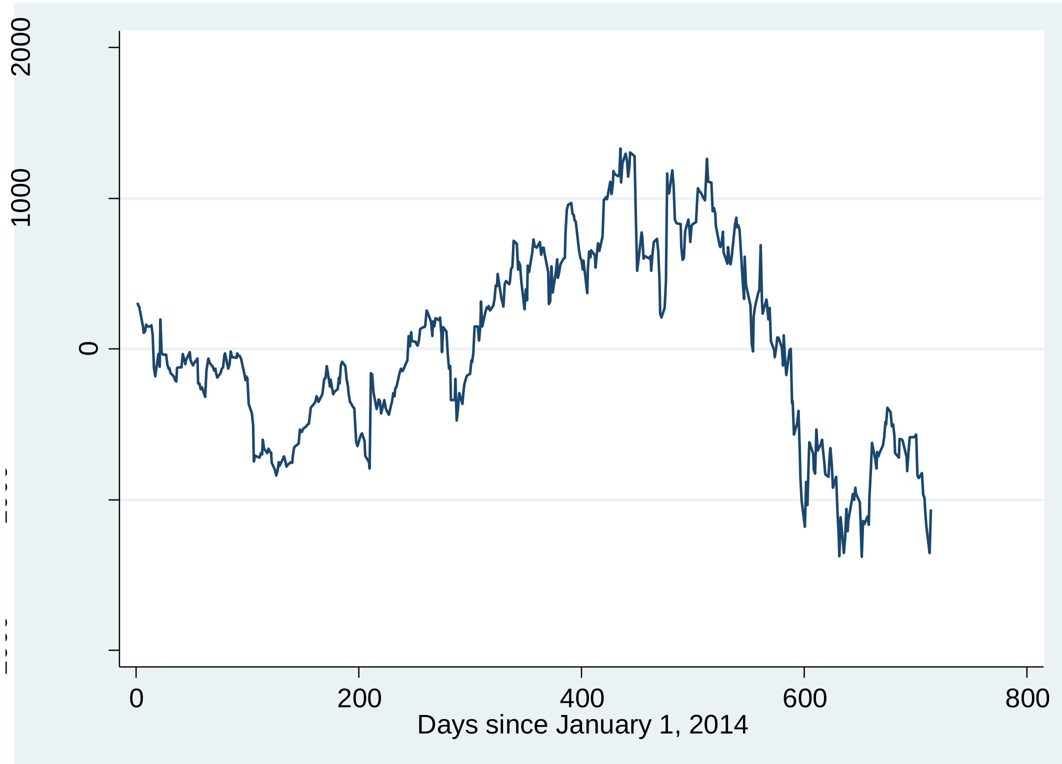
```
use "F:\stockdata.dta", clear
* We make use of stockdata of ASML *
. keep if stock == 1
(14259 observations deleted)
. tsset time
        time variable:  time, 1 to 714, but with gaps
                delta:  1 unit
graph twoway line stockprice time
```




```
. sum stockprice
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
  stockprice |       510    5859.759    945.7895    4219.45     7661.5
-------------+--------------------------------------------------------


. reg stockprice time

      Source |       SS       df       MS              Number of obs =     510
-------------+------------------------------           F(  1,   508) =  670.24
       Model |  259001715     1   259001715           Prob > F      =  0.0000
    Residual |  196307841   508  386432.758           R-squared     =  0.5688
-------------+------------------------------           Adj R-squared =  0.5680
       Total |  455309556   509  894517.792           Root MSE      =  621.64
------------------------------------------------------------------------------
  stockprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |    3.45742   .1335481    25.89   0.000     3.195045    3.719794
       _cons |   4624.423   55.08719    83.95   0.000     4516.197     4732.65
------------------------------------------------------------------------------
```

- Conclusion: increase of stockprice is 3.45 per day

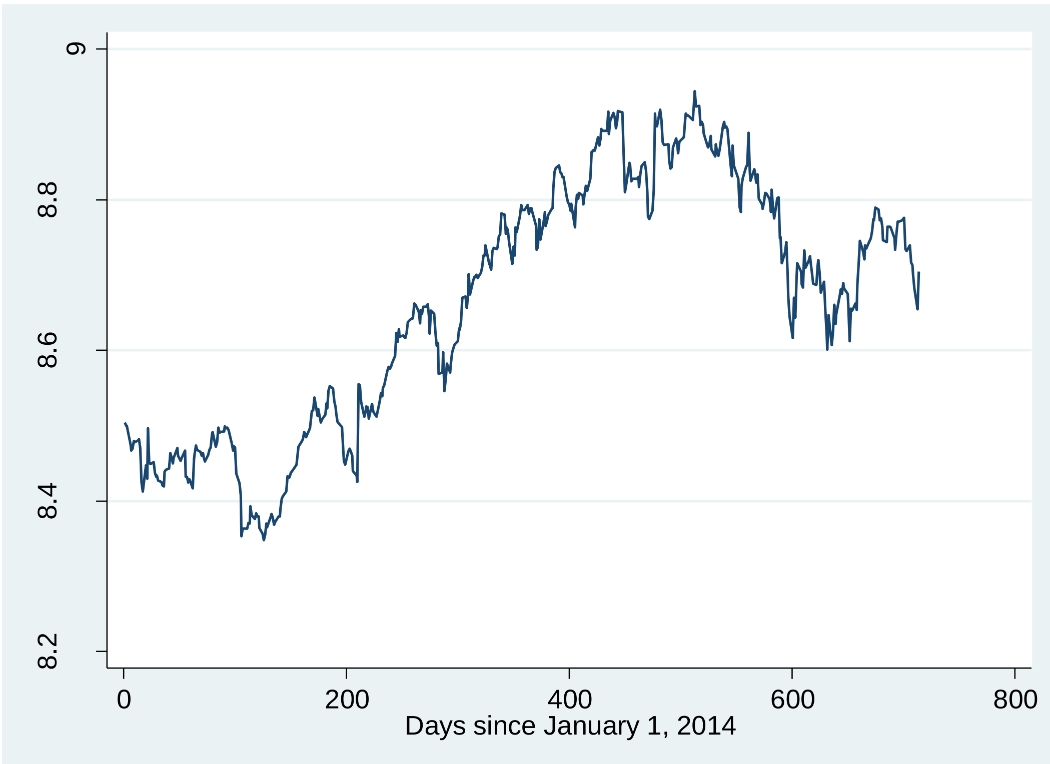
```
. predict residu_ASML, resid
. graph twoway line residu_ASML time
```



## Exponential trend

For the natural logarithm of the stockprice:

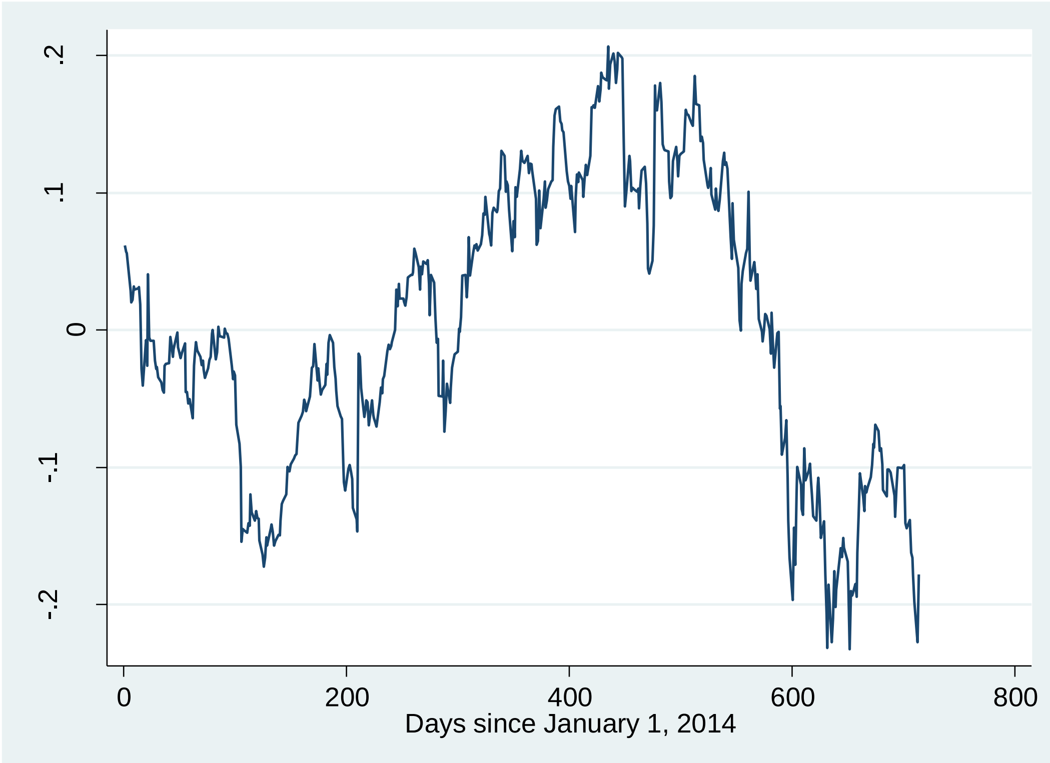
```
. graph twoway line ln_stockprice time
```




```
. reg ln_stockprice time

      Source |       SS           df       MS      Number of obs   =       510
-------------+----------------------------------   F(1, 508)       =    753.50
       Model |  8.24221675         1  8.24221675   Prob > F        =    0.0000
    Residual |  5.55681806       508  .010938618   R-squared       =    0.5973
-------------+----------------------------------   Adj R-squared   =    0.5965
       Total |  13.7990348       509  .027110088   Root MSE        =    .10459

------------------------------------------------------------------------------
ln_stockpr~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0006168   .0000225    27.45   0.000     .0005726    .0006609
       _cons |   8.442151   .0092682   910.87   0.000     8.423943     8.46036
------------------------------------------------------------------------------
```

- **Conclusion:** increase of stockprice of ASML is 0.0617 percent per day

```
. predict residu_ln_stockprice, resid
. graph twoway line residu_ln_stockprice time
```



# Seasonality

## Seasonality

*Aim: to introduce seasonality*

- If a time series has daily, monthly or quarterly observations, it may exhibit seasonality (e.g. the number of transactions in housing market, unemployment, sickness absenteeism, the stockprice) varies specific patterns repeat over the year
- Many econometric techniques are available to account for seasonality. E.g. by adding seasonal dummy variables:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} \\ &\quad + \delta_2 febr_t + \delta_3 march_t + \ldots + \delta_{12} dec_t + u_t \end{aligned}$$

- Apply an *F*-test of joint significance to test for seasonality:

$H_0 : \delta_2 = 0, \delta_3 = 0, \ldots, \delta_{12} = 0$
$H_1 : H_0$ not true

# Example: Seasonality for the stockprice?

```
. tab month, gen(dmonth)

      month |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         45        8.82        8.82
          2 |         40        7.84       16.67
          3 |         43        8.43       25.10
          4 |         44        8.63       33.73
          5 |         43        8.43       42.16
          6 |         43        8.43       50.59
          7 |         46        9.02       59.61
          8 |         42        8.24       67.84
          9 |         44        8.63       76.47
         10 |         45        8.82       85.29
         11 |         41        8.04       93.33
         12 |         34        6.67      100.00
------------+-----------------------------------
      Total |        510      100.00

. reg stockprice dmonth*
note: dmonth2 omitted because of collinearity

      Source |       SS       df       MS              Number of obs =     510
-------------+------------------------------           F( 11,   498) =    2.39
       Model |    22816973    11  2074270.28           Prob > F      =  0.0069
    Residual |   432492583   498  868459.002           R-squared     =  0.0501
-------------+------------------------------           Adj R-squared =  0.0291
       Total |   455309556   509  894517.792           Root MSE      =  931.91

------------------------------------------------------------------------------
  stockprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     dmonth1 |  -50.09094   202.5107    -0.25   0.805    -447.9715    347.7896
     dmonth2 |          0  (omitted)
     dmonth3 |   339.3381   204.7149     1.66   0.098    -62.87323    741.5495
     dmonth4 |   51.49089   203.5907     0.25   0.800    -348.5117    451.4935
     dmonth5 |   100.1954   204.7149     0.49   0.625     -302.016    502.4067
     dmonth6 |   372.8324   204.7149     1.82   0.069    -29.37901    775.0437
     dmonth7 |   164.3509   201.4721     0.82   0.415    -231.4892     560.191
     dmonth8 |  -36.32133   205.8861    -0.18   0.860    -440.8337     368.191
     dmonth9 |   74.79882   203.5907     0.37   0.713    -325.2038    474.8014
    dmonth10 |   1.359701   202.5107     0.01   0.995    -396.5209    399.2403
    dmonth11 |    527.865   207.1072     2.55   0.011     120.9534    934.7767
    dmonth12 |   621.5738   217.3807     2.86   0.004     194.4773     1048.67
       _cons |   5688.963   147.3481    38.61   0.000     5399.462    5978.464
------------------------------------------------------------------------------
```

```
. testparm dmonth*

 ( 1)   dmonth1 = 0
 ( 2)   dmonth3 = 0
 ( 3)   dmonth4 = 0
 ( 4)   dmonth5 = 0
 ( 5)   dmonth6 = 0
 ( 6)   dmonth7 = 0
 ( 7)   dmonth8 = 0
 ( 8)   dmonth9 = 0
 ( 9)   dmonth10 = 0
 (10)   dmonth11 = 0
 (11)   dmonth12 = 0

       F( 11,   498) =    2.39
            Prob > F =    0.0069
```

**Conclusion:** The *F*-statistic is statistically significant (*p*-value close to zero), so the null hypothesis that there is no seasonal effect is rejected (i.e. there is a seasonal effect).

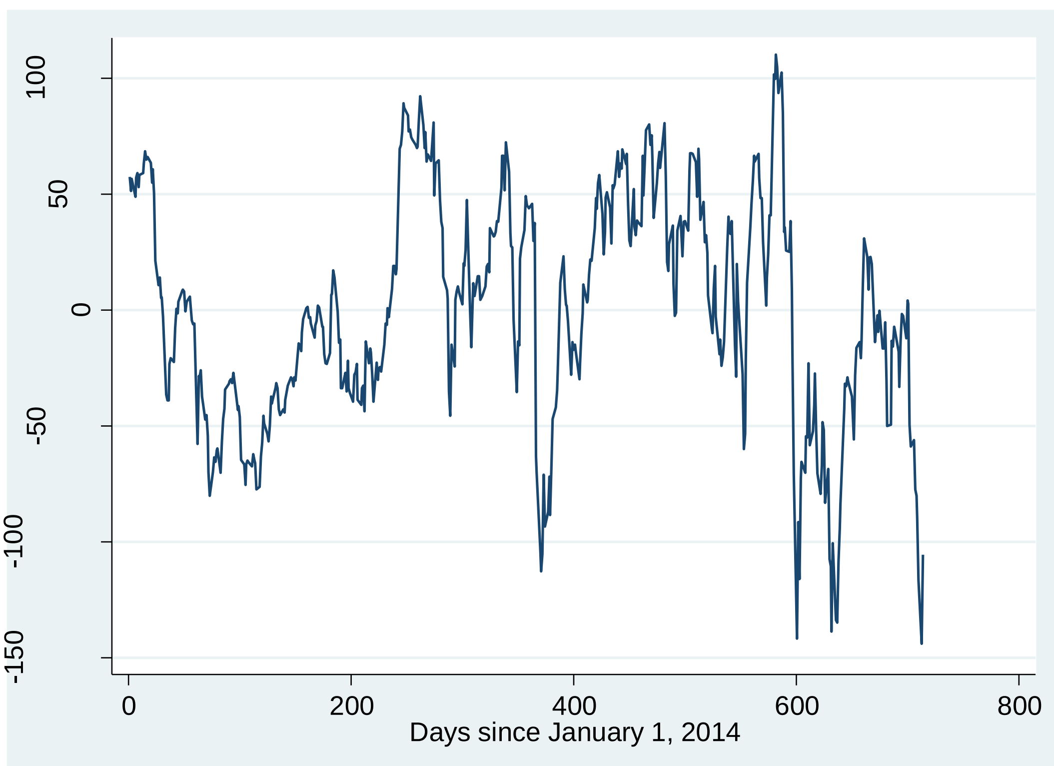
# Seasonality and time trend for the stockprice?

```
. reg stockprice dmonth* time
note: dmonth2 omitted because of collinearity

      Source |       SS       df       MS              Number of obs =     510
-------------+------------------------------           F( 12,   497) =   95.93
       Model |  318007857    12  26500654.8           Prob > F      =  0.0000
    Residual |  137301699   497  276260.964           R-squared     =  0.6984
-------------+------------------------------           Adj R-squared =  0.6912
       Total |  455309556   509  894517.792           Root MSE      =  525.61

------------------------------------------------------------------------------
  stockprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     dmonth1 |   91.70753   114.2999     0.80   0.423    -132.8631    316.2782
     dmonth2 |          0  (omitted)
     dmonth3 |   199.7107   115.5398     1.73   0.085    -27.29596    426.7173
     dmonth4 |  -197.7151   115.0796    -1.72   0.086    -423.8175    28.38725
     dmonth5 |  -258.5893   115.9813    -2.23   0.026    -486.4634   -30.71509
     dmonth6 |  -147.9337   116.5547    -1.27   0.205    -376.9345      81.067
     dmonth7 |  -468.8164   115.2709    -4.07   0.000    -695.2948   -242.3379
     dmonth8 |  -798.5986   118.4397    -6.74   0.000    -1031.303   -565.8943
     dmonth9 |  -813.6997   117.9999    -6.90   0.000     -1045.54   -581.8595
    dmonth10 |  -1000.258   118.2563    -8.46   0.000    -1232.602   -767.9138
    dmonth11 |  -635.7264   122.1135    -5.21   0.000    -875.6488    -395.804
    dmonth12 |  -369.5838   126.2982    -2.93   0.004    -617.7281   -121.4396
        time |   4.188337   .1281297    32.69   0.000     3.936594    4.440079
       _cons |   4731.928   88.11189    53.70   0.000      4558.81    4905.046
------------------------------------------------------------------------------

. testparm dmonth*

 ( 1)  dmonth1 = 0
 ( 2)  dmonth3 = 0
 ( 3)  dmonth4 = 0
 ( 4)  dmonth5 = 0
 ( 5)  dmonth6 = 0
 ( 6)  dmonth7 = 0
 ( 7)  dmonth8 = 0
 ( 8)  dmonth9 = 0
 ( 9)  dmonth10 = 0
 (10)  dmonth11 = 0
 (11)  dmonth12 = 0

       F( 11,   497) =   19.42
            Prob > F =    0.0000
```

**Conclusions:**

- Both trend (because time is individually significant) and seasonal effects (since *p*-value of joint test of parameters on dmonth* is zero) are present.
- The estimated parameter on *dmonth12*indicates that the stockprice is 370 lower in December relative to February, ceteris paribus on the time trend.
- The parameter on the time trend indicates that the stockprice increased by 4.2 each day, ceteris paribus on month.
- The residuals may be interpreted as the stockprice, after having controlled for month of the year and the time trend (see graph below).

```
. predict uhat, resid
. graph twoway line uhat time
```



# Spurious regression

## Spurious regression

*Aim: to show that it is important to include a trend in a regression equation.*

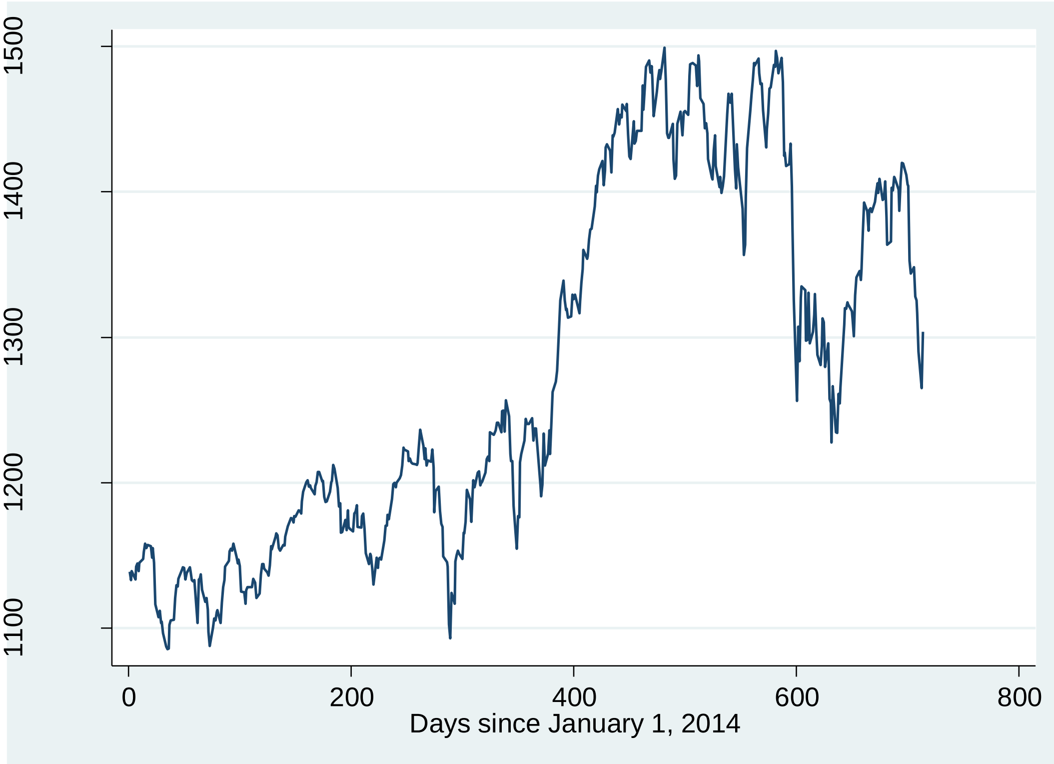
- In a regression equation:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 t + e_t \qquad t=1,2,\ldots \quad \text{with } Ee_t = 0, Var(e_t) = \sigma^2$$

- One should use a time trend in order to take account of unobserved, trending variables that affect both $x_t$ and $y_t$.
- The non-inclusion of a trend variable *t*, it may lead to **spurious** regression.
- There may appear to be an effect of *x* on *y* if the trend, *t*, is not included in the regression equation.

# Example: the AEXindex

```
. graph twoway line AEXindex time
```




```
. reg stockprice time

      Source |       SS       df       MS              Number of obs =     510
-------------+------------------------------           F(  1,   508) =  670.24
       Model |  259001715     1   259001715            Prob > F      =  0.0000
    Residual |  196307841   508  386432.758            R-squared     =  0.5688
-------------+------------------------------           Adj R-squared =  0.5680
       Total |  455309556   509  894517.792            Root MSE      =  621.64

------------------------------------------------------------------------------
  stockprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |    3.45742   .1335481    25.89   0.000     3.195045    3.719794
       _cons |   4624.423   55.08719    83.95   0.000     4516.197     4732.65
------------------------------------------------------------------------------
```

```
. reg stockprice AEXindex

      Source |       SS       df       MS              Number of obs =     510
-------------+------------------------------           F(  1,   508) = 1945.79
       Model |   361048404     1   361048404           Prob > F      =  0.0000
    Residual |  94261152.5   508   185553.45           R-squared     =  0.7930
-------------+------------------------------           Adj R-squared =  0.7926
       Total |   455309556   509  894517.792           Root MSE      =  430.76

------------------------------------------------------------------------------
  stockprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    AEXindex |   6.745643   .1529238    44.11   0.000     6.445202    7.046084
       _cons |  -2743.479   195.9661   -14.00   0.000    -3128.483   -2358.476
------------------------------------------------------------------------------

. reg stockprice AEXindex time

      Source |       SS       df       MS              Number of obs =     510
-------------+------------------------------           F(  2,   507) =  987.63
       Model |   362312913     2   181156456           Prob > F      =  0.0000
    Residual |  92996643.5   507  183425.332           R-squared     =  0.7958
-------------+------------------------------           Adj R-squared =  0.7949
       Total |   455309556   509  894517.792           Root MSE      =  428.28

------------------------------------------------------------------------------
  stockprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    AEXindex |   6.189308   .2607943    23.73   0.000     5.676938    6.701679
        time |   .4143712   .1578186     2.63   0.009     .1043122    .7244302
       _cons |  -2181.998   289.2974    -7.54   0.000    -2750.367   -1613.629
------------------------------------------------------------------------------

. reg stockprice AEXindex time dmonth*
note: dmonth3 omitted because of collinearity

      Source |       SS       df       MS              Number of obs =     510
-------------+------------------------------           F( 13,   496) =  238.65
       Model |   392550949    13  30196226.9           Prob > F      =  0.0000
    Residual |  62758606.9   496  126529.449           R-squared     =  0.8622
-------------+------------------------------           Adj R-squared =  0.8586
       Total |   455309556   509  894517.792           Root MSE      =  355.71

------------------------------------------------------------------------------
  stockprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    AEXindex |   7.818362   .3221127    24.27   0.000     7.185489    8.451236
        time |  -.5077373   .2120192    -2.39   0.017    -.9243039   -.0911707
     dmonth1 |   230.4776    77.3481     2.98   0.003     78.50732     382.448
     dmonth2 |  -37.91592   78.47657    -0.48   0.629    -192.1034    116.2716
     dmonth3 |          0  (omitted)
     dmonth4 |   -436.194   76.32767    -5.71   0.000    -586.1594   -286.2286
     dmonth5 |  -401.6394   76.88386    -5.22   0.000    -552.6975   -250.5812
     dmonth6 |  -186.5459   77.40419    -2.41   0.016    -338.6264   -34.46533
     dmonth7 |  -359.3218   77.20025    -4.65   0.000    -511.0016   -207.6419
     dmonth8 |  -319.9934   83.08023    -3.85   0.000     -483.226   -156.7608
     dmonth9 |   60.49232   89.53297     0.68   0.500    -115.4183     236.403
    dmonth10 |   43.81716   93.26672     0.47   0.639    -139.4294    227.0637
    dmonth11 |   102.5033   89.28221     1.15   0.251    -72.91465    277.9213
    dmonth12 |   552.3123   95.44704     5.79   0.000     364.7819    739.8426
       _cons |  -3855.709   366.7792   -10.51   0.000    -4576.342   -3135.077
------------------------------------------------------------------------------
```

# Properties of OLS-estimators – consistent estimator

## Chapter 11: Under which assumptions does OLS give consistent parameter estimates?

*Aim: to introduce the assumptions for consistency*

**Theorem 11.1 Estimates are consistent if the following assumptions hold true:**

- Linear model
- No perfect multicollinearity
- Contemporaneous exogeneity of the explanatory variables
  - Contemporaneous exogeneity will be explained below.

**In addition:** two other assumptions are required for consistency:

1) Stationarity of all variables of the regression equation
   - Stationarity will be explained in week 3.

2) Weak dependency of all variables of the regression equation.
   - Weak dependence will be explained below.

# Contemporaneous exogeneity

## Strict exogeneity

*Aim: to introduce the concept of strict exogeneity*

Below is a static multivariate regression equation (with $k$ explanatory variables):

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$$

- the error term $u$ in period $t$ is independent of the explanatory variables for ***all*** time periods.
- The error term $u$ in period $t$ is independent of all right-hand side variables in the past:
  - $u_t$ is independent of $x_{11}, x_{12}, ..., x_{1k}$ for $t = 1$
  - $u_t$ is independent of $x_{21}, x_{22}, ..., x_{2k}$ for $t = 2$
  - …
  - $u_t$ is independent of $x_{t-1,1}, x_{t-1,2}, ..., x_{t-1,k}$ for $t = t$-1
    - (I added a comma in the subscript to distinguish the variable number from the time period)
- $u_t$ is independent of all right-hand side variables in the current period:
  - independent of $x_{t1}, x_{t2}, ..., x_{tk}$ for $t = t$
- The error term $u$ in period $t$ is independent of all right-hand side variables in the future:
  - $u_t$ is independent of $x_{t+1,1}, x_{t+1,2}, ..., x_{t+1,k}$ for $t = t$+1
  - $u_t$ is independent of $x_{t+2,1}, x_{t+2,2}, ..., x_{t+2,k}$ for $t = t$+2
  - …
  - $u_t$ is independent of $x_{n1}, x_{n2}, ..., x_{nk}$ for $t = n$

$X$ (capital) denotes the set of *all* explanatory variables for *all* time periods.

**Assumption TS.2 (strict exogeneity)**

For each $t$, the expected value of $u_t$, given the explanatory variables for *all* time periods, is equal to zero: $E(u_t \mid X) = 0$.

## Contemporaneous exogeneity

*Aim: to introduce the concept of weak exogeneity*

**Alternative: Assumption TS.2' Contemporaneous exogeneity**

For each *t*, the error term $u_t$ is independent of all right-hand side variables in the present: independent of $x_{t1}, x_{t2}, ..., x_{tk}$ for *t* = *t*.

More formally: For each *t*, the expected value of $u_t$, given the explanatory variables in period *t*, is equal to zero:
$E(u_t \mid x_{t1}, ... x_{tk}) = 0$.

- Note that we apply here the lower case *x* (only for period *t* of the dataset; not for the entire dataset *X*).
- This assumption implies that the error term in period *t* is uncorrelated with all regressors in period *t*:
- $Cov(u_t, x_{tj}) = 0$ (*j*=1,…,*k*)
- Remember that this assumption was applied for consistency of OLS (see week 1; Chapter 5).

## Violation of the strict exogeneity assumption for a lagged dependent variable

*Aim: to show that strict exogeneity is often an incorrect assumption.*

**Example:**
Model **with lagged dependent endogenous variable**, assuming contemporaneous exogeneity ($E(u_t \mid y_{t-1}, z_t) = 0$)

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \delta_0 z_t + u_t \qquad (1)$$

- The lagged dependent variable $y_{t-1}$ is **NOT** strictly exogenous. We apply a proof by contradiction, which consists of three steps.
  1.The model (1) implies that $u_t$ and $y_t$ are dependent.

  2. We start with a proposition, which means that we suppose there is strict exogeneity. Strict exogeneity not only implies that $E(u_t \mid y_{t-1}, z_t) = 0$, but also that $u_t$ is uncorrelated with the explanatory variables in period *t*+1. $E(u_t \mid y_t, z_{t+1}) = 0$. Thus it implies that $u_t$ and $y_t$ are independent.

  3. But according to model (1), $u_t$ and $y_t$ are dependent. Thus the proposition of strict exogeneity is false. There is a contradiction. This means that the model cannot be strictly exogenous for $y_{t-1}$, since $E(u_t \mid y_t, z_{t+1}) \neq 0$.

- In other words, applying OLS on model (1) yields biased estimates because assumption TS.2 (strict exogeneity) does not hold.

## Violation of the strict exogeneity assumption for a feedback mechanism

*Aim: to show that strict exogeneity is often an incorrect assumption.*

**Example:** (see e.g. exercise 10.2 of Wooldridge). Models with a **feedback mechanism**

- General structure of feedback mechanism:
  - $y_t$ depends on $x_t$
  - $x_t$ depends on $y_{t-1}$

Consider the following model:

$$gGDP_t = \alpha_0 + \delta_0 r_t + u_t \qquad (1)$$

$gGDP_t$: GDP-growth rate.

$r_t$: interest rate; $r_t$ is contemporaneously exogenous

- Decision of Jerome Powell on FED-interest rate (feedback mechanism):

$$r_t = \gamma_0 + \gamma_1 (gGDP_{t-1} - 3) + v_t \qquad \gamma_1 > 0 \qquad (2)$$

1. Equation (1) implies that $u_t$ and $gGDP_t$ are dependent
2. Proposition of strict exogeneity: Equation (1): $u_t$ is independent of $r_{t+1}$
3. Equation (2) implies that $r_{t+1}$ depends on $gGDP_t$,
4. Thus strict exogeneity implies that $u_t$ is independent of $gGDP_t$
5. This leads to a contradiction. Thus the proposition of strict exogeneity is false. Thus, $r_t$ cannot be strictly exogenous in equation (1).

- How do we deal with time series models, for which not all explanatory variables (RHS-variables) are strictly exogenous? If strict exogeneity is violated, we rely on assumption TS.2’ (contemporaneous exogeneity).
- Consequently, OLS yields consistent estimates.
- Definition of consistency: see previous lecture (chapter 5 of Wooldridge).
- Note that consistency is a property of large samples.

# Weak dependency

## Weak dependency

*Aim: definition and application of weak dependence*

The definition of **weakly-dependent time series:**
A stationary time series $\{x_t : t = 1, 2, ...,\}$ is weakly dependent if $x_t$ and $x_{t+h}$ are "almost independent" as $h \to \infty$. Thus $Corr(x_t, x_{t+h}) \to 0$ as $h \to \infty$.

Due to the property of weak dependency, it is not necessary to make the assumption of a random sample in order to prove consistency.

## Intermezzo: Correlation and covariance

*Aim: to summarize properties of correlation, covariance and variance*

**Remember:**

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \qquad \text{(B.29)}$$

$$-1 \leq \rho_{X,Y} \leq 1$$

**Properties that are often applied:**

$$Cov(X,Y) = E(X-\mu_X)(Y-\mu_Y)$$

$$Cov(X,X) = Var(X)$$

$$Cov(X,a) = 0$$

$$Cov(aX+c,bY+d) = abCov(X,Y)$$

$$\begin{aligned} Cov(aX+bY,cW+dZ) &= \\ &= acCov(X,W) + adCov(X,Z) \\ &\quad + bcCov(Y,W) + bdCov(Y,Z) \end{aligned}$$

$$Var(aX+bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$$

## Intertemporal correlations applied to time series

Correlation between $x_t$ and $x_{t+1}$:

$$Corr(x_t, x_{t+1}) = \frac{Cov(x_t, x_{t+1})}{\sqrt{Var(x_t)}\sqrt{Var(x_{t+1})}}$$

Correlation between $x_t$ and $x_{t+2}$:

$$Corr(x_t, x_{t+2}) = \frac{Cov(x_t, x_{t+2})}{\sqrt{Var(x_t)}\sqrt{Var(x_{t+2})}}$$

..........

Correlation between $x_t$ and $x_{t+h}$:

$$Corr(x_t, x_{t+h}) = \frac{Cov(x_t, x_{t+h})}{\sqrt{Var(x_t)}\sqrt{Var(x_{t+h})}}$$

The definition of **weakly-dependent time series:**

A stationary time series $\{x_t : t = 1, 2, ...,\}$ is weakly dependent if $x_t$ and $x_{t+h}$ are "almost independent" as $h \to \infty$. Thus $Corr(x_t, x_{t+h}) \to 0$ as $h \to \infty$.

## Examples of weakly dependent time series

The following two models are weakly dependent:

1) First-order autoregressive process (AR(1)-process)

$y_t = \delta + \rho_1 y_{t-1} + e_t \qquad |\rho_1| < 1$

We assume that the error term $e_t$: i.i.d. (identically and independently distributed), with expected value zero and constant variance: $Ee_t = 0$; $Var(e_t) = \sigma_e^2$. $e_t$ is independent of $y_{t-1}$.

2) The first-order moving average process (MA(1)-process)

$y_t = e_t + \alpha_1 e_{t-1}$

The error term $e_t$ is i.i.d. (identically and independently distributed), with expected value zero and constant variance: $Ee_t = 0$ and $Var(e_t) = \sigma_e^2$

## Random walk is not weakly dependent

*Aim: to show that a random walk is not a weakly-dependent time series*

Consider the random walk model:

$$y_t = y_{t-1} + e_t$$

- The error term $e_t$ is **i.i.d.** (**i**dentically and **i**ndependently **d**istributed), with expected value zero and constant variance: $Ee_t = 0$; $Var(e_t) = \sigma_e^2$. $e_t$ is independent of $y_{t-1}$.
- It can be shown: $Corr(y_t, y_{t-h}) = \sqrt{\frac{t-h}{t}}$ which does not converge towards zero; since for given *h*, as $t \to \infty$ then $\sqrt{\frac{t-h}{t}} \to 1$
- Consequently, a random walk is not weakly dependent (thus $Corr(y_t, y_{t-h})$ should converge to zero in the case of weak dependence).

# Example of correlation

Autocorrelations can be calculated for the stockprice

```
. corrgram stockprice, lags (20)
(note: time series has 102 gaps)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.7917   0.9900   321.51  0.0000          |------           |-------
2        0.5888   0.0444   499.69  0.0000          |----             |
3        0.5882   0.0229   677.91  0.0000          |----             |
4        0.5898  -0.1324   857.41  0.0000          |----            -|
5        0.5897        .   1037.2  0.0000          |----
6        0.7763        .   1349.5  0.0000          |------
7        0.9636        .   1831.5  0.0000          |-------
8        0.7651        .     2136  0.0000          |------
9        0.5698        .   2305.2  0.0000          |----
10       0.5692        .   2474.4  0.0000          |----
11       0.5708        .   2644.9  0.0000          |----
12       0.5694        .   2814.9  0.0000          |----
13       0.7510        .   3111.2  0.0000          |------
14       0.9333        .   3569.8  0.0000          |-------
15       0.7404        .     3859  0.0000          |-----
16       0.5507        .   4019.3  0.0000          |----
17       0.5517        .   4180.5  0.0000          |----
18       0.5543        .   4343.5  0.0000          |----
19       0.5537        .   4506.5  0.0000          |----
20       0.7303        .   4790.8  0.0000          |-----
```

For residual of stockprice after correcting for trend and month:

```
. corrgram uhat, lags (20)
(note: time series has 102 gaps)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.7638   0.9730   299.32  0.0000          |------           |-------
2        0.5446   0.0493   451.78  0.0000          |----             |
3        0.5494  -0.0802   607.26  0.0000          |----             |
4        0.5566  -0.0863   767.14  0.0000          |----             |
5        0.5496        .   923.35  0.0000          |----
6        0.6964        .   1174.6  0.0000          |-----
7        0.8420        .   1542.7  0.0000          |------
8        0.6530        .   1764.5  0.0000          |-----
9        0.4733        .   1881.2  0.0000          |---
10       0.4779        .   2000.5  0.0000          |---
11       0.4832        .   2122.7  0.0000          |---
12       0.4762        .   2241.6  0.0000          |---
13       0.6136        .   2439.4  0.0000          |----
14       0.7511        .   2736.4  0.0000          |------
15       0.5866        .   2917.9  0.0000          |----
16       0.4367        .   3018.7  0.0000          |---
17       0.4477        .   3124.9  0.0000          |---
18       0.4491        .   3231.9  0.0000          |---
19       0.4453        .   3337.4  0.0000          |---
20       0.5887        .   3522.1  0.0000          |----
```

# Implications of a unit root model

## Transformations of highly-persistent time series

*Aim: to show that we need to take the first difference of a random walk to get a weakly dependent time series.*

- Weakly dependent processes (stationary processes) are said to be integrated of order zero or I(0).

- Practically, this means that nothing needs to be done to such series before using them to regression analysis.

- Unit root processes, such as the random walk process, are said to be integrated of order one, or I(1).

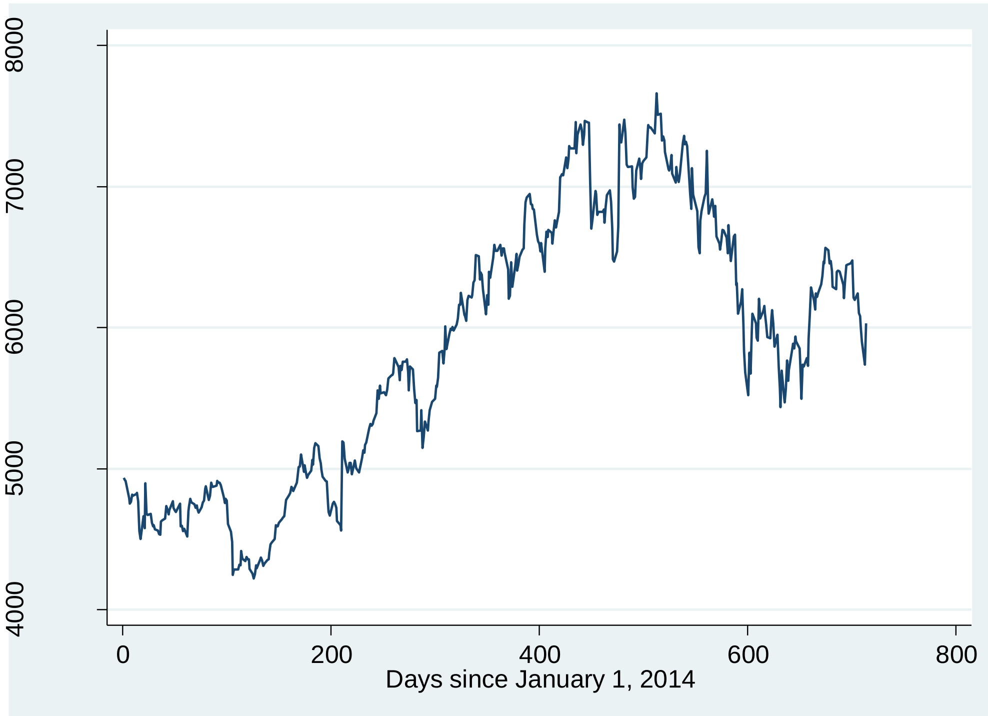
- If $\{y_t\}$ is integrated of order one:

  $y_t = y_{t-1} + e_t$
  then $\Delta y_t = y_t - y_{t-1} = e_t$

- Thus $\{\Delta y_t\}$ is integrated of order zero: I(0), because $e_t$ is i.i.d. (identically and independently distributed), with expected value zero and constant variance: $Ee_t = 0$; $Var(e_t) = \sigma_e^2$

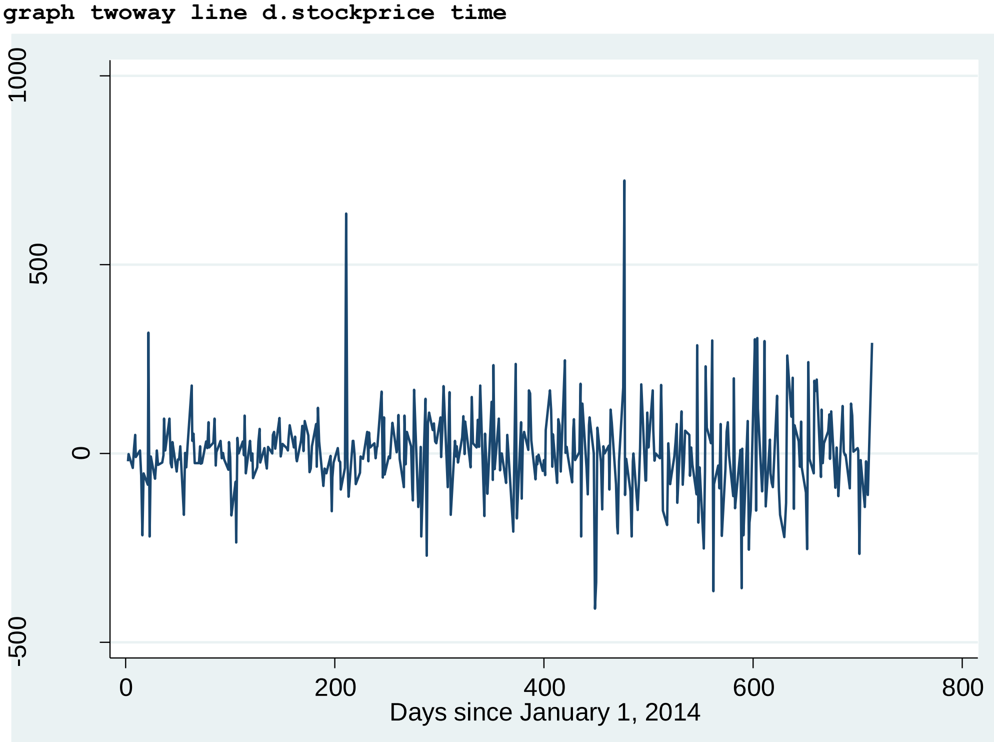
Level: *stockprice*

```
graph twoway line stockprice time
```




First difference: Δ*stockprice*

```
graph twoway line d.stockprice time
```



```
. corrgram d.stockprice, lags (20)
(note: time series has 102 gaps)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1       -0.0389  -0.0465   .62151  0.4305          |                   |
2       -0.0124  -0.0250   .68454  0.7102          |                   |
3        0.0229   0.1324   .90022  0.8254          |                   |-
4       -0.0302        .   1.2763  0.8654          |
5        0.0428        .   2.0342  0.8444          |
6       -0.0555        .   3.3119  0.7688          |
7       -0.0958        .    7.134  0.4151          |
8        0.0116        .   7.1902  0.5163          |
9        0.0147        .   7.2806  0.6079          |
10      -0.0120        .   7.3412  0.6929          |
11       0.0073        .   7.3636  0.7689          |
12      -0.0258        .   7.6438  0.8123          |
13       0.0183        .   7.7845  0.8573          |
14      -0.0262        .   8.0747  0.8854          |
15      -0.0641        .   9.8177  0.8310          |
16      -0.0010        .   9.8181  0.8760          |
17       0.0290        .   10.176  0.8961          |
18      -0.0043        .   10.184  0.9257          |
19      -0.0082        .   10.213  0.9475          |
20       0.0456        .   11.107  0.9434          |
```

# Autocorrelation

## Autocorrelation

*Aim: to introduce autocorrelation and to consider the consequences of autocorrelation in a model with strictly exogenous variables.*

- Suppose that the following condition is violated:

**Assumption TS.5 (no autocorrelation)**
Conditional on *X*, the error terms in two different time periods, $u_t$ and $u_s$, $t \neq s$, are uncorrelated: $Corr(u_t, u_s \mid X) = 0$

- **We first apply it to a model with strictly exogenous variables. Hence,**
    - There are no lagged dependent variables
    - There is no feedback mechanism

- The following model has an AR(1) error structure:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$$
$$u_t = \rho u_{t-1} + e_t \qquad |\rho| < 1$$

- Conditions TS.1 (linearity), TS.2 (strict exogeneity), and TS.3 (no perfect multicollinearity) are met. Theorem 11.1 still holds. Therefore, OLS-estimators $\hat{\beta}_j$ are consistent.
- Standard formulas of $Var(\hat{\beta}_j)$ are not valid without adjustment, *F*-tests and *t*-tests are not valid.
- Standard formulas of $Var(\hat{\beta}_j)$ underestimate the 'true' variance if there is a positive autocorrelation $0 < \rho < 1$
- In the case, the *t*-values are overestimated. Which means that one erroneously concludes too often that variables are statistically significant.

## Testing for serial correlation: Durbin-Watson statistic

*Aim: to introduce the DW-test for autocorrelation.*

- Example:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$$

Where all $x_{t1}$ are strictly exogenous. The error term follows an AR(1) process:

$$u_t = \rho u_{t-1} + e_t \qquad |\rho| < 1$$

- The error term $e_t$ is i.i.d. (identically and independently distributed), with expected value zero and constant variance: $Ee_t = 0$; $Var(e_t) = \sigma_e^2$
- $H_0 : \rho = 0$ (no autocorrelation)
  $H_1 : \rho \neq 0$ (autocorrelation)

Durbin-Watson statistic:

$$DW = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \quad \text{where } DW \approx 2(1 - \hat{\rho})$$

- Thus:
    - if $\rho = 0$ then $DW = 2$ (no autocorrelation)
    - if $\rho \approx 1$ then $DW \approx 0$ (random walk)
- However, the DW-statistic is only valid if all regressors are strictly exogenous. Also the may not be lagged dependent variables in the RHS. For this reason, we prefer to apply the Breusch-Godfrey test (see following slide)

## Breusch-Godfrey test for autocorrelation

*Aim: to introduce the BG-test for autocorrelation.*

$$u_t = \rho u_{t-1} + e_t$$

- Alternative test:
    - Estimate the regression equation using OLS. E.g. the model:
      $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$
    - Calculate the residuals $\hat{u}_t$
    - Run the regression of the residual on its lag and the explanatory variables
        - $\hat{u}_t$ on $\hat{u}_{t-1}$ and $x_{t1}, x_{t2}, ..., x_{tk}$
    - Obtain the *t*-statistic on $\hat{u}_{t-1}$ to test
      $H_0 : \rho = 0$ (no first-order autocorrelation)
      $H_1 : \rho \neq 0$ (autocorrelation)
    - There is indication of first-order autocorrelation if the *t*-statistic on $\hat{u}_{t-1}$ is statistically significant.

- If there is second-order autocorrelation:
  $u_t = \alpha + \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t$
  the procedure should be the same:
    - Run the regression of the residual on both lags of the residual and the explanatory variables
        - $\hat{u}_t$ on $\hat{u}_{t-1}, \hat{u}_{t-2}$ and $x_{t1}, x_{t2}, ..., x_{tk}$
    - Obtain the *F*-statistic to test for joint significance of $\hat{u}_{t-1}$ and $\hat{u}_{t-2}$.
    - $H_0 : \rho_1 = 0, \rho_2 = 0$ (no autocorrelation)
      $H_1 : H_0$ is not true (autocorrelation)

## Alternative procedure: Prais Winsten (FGLS)

- If the tests show evidence of autocorrelation, we should NOT use OLS to estimate the regression equation.
- Instead FGLS should be used.
- Mechanics of FGLS: consider the following model,
  $$y_t = \beta_0 + \beta_1 x_1 + u_t \qquad (1)$$
  where the error term follows an AR(1) process:
  $$u_t = \rho u_{t-1} + e_t$$
- The model is also valid in period *t*-1:
  $$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + u_{t-1} \qquad (2)$$
- and $\rho$ times equation (2) is
  $$\rho y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{t-1} + \rho u_{t-1} \qquad (3)$$
- Equation (1) – (3):
- $$y_t - \rho y_{t-1} = (1-\rho)\beta_0 + \beta_1(x_t - \rho x_{t-1}) + e_t \qquad (4)$$
  where $e_t = u_t - \rho u_{t-1}$ is uncorrelated over time (i.i.d.)
- The econometrician Prais has developed a GLS-procedure, using equation (4) to estimate the parameters, including $\rho$. It is referred to as the Prais-Winsten method or Cochrane-Orcutt method.
- This is referred to as Feasible GLS (FGLS)
- Stata command: prais (instead of reg).

## Newey-West standard errors

- Remember from the discussion of heteroskedasticity that it is possible to compute heteroskedasticity robust standard errors
- It is also possible to compute heteroskedasticity and autocorrelation robust standard errors of the estimated regression parameters (or HAC standard errors)
- Newey-West cannot always be used to solve autocorrelation. It only works if the population regression is specified correctly.

```
. tabulate broker

  Name of the broker |
     who recommends |      Freq.     Percent        Cum.
--------------------+-----------------------------------
           ABN Amro |          1        1.12        1.12
           Barclays |          1        1.12        2.25
          Citigroup |          5        5.62        7.87
      Credit Suisse |          2        2.25       10.11
      Deutsche Bank |         14       15.73       25.84
   Exane BNP Paribas |          1        1.12       26.97
      Goldman Sachs |         11       12.36       39.33
               HSBC |          1        1.12       40.45
                ING |          8        8.99       49.44
        J.P. Morgan |         11       12.36       61.80
          Jefferies |          1        1.12       62.92
   Kepler Cheuvreux |          6        6.74       69.66
     Morgan Stanley |          4        4.49       74.16
               Rabo |          4        4.49       78.65
     SNS Securities |          5        5.62       84.27
SociÃ©tÃ© GÃ©nÃ©rale |          3        3.37       87.64
                UBS |         11       12.36      100.00
--------------------+-----------------------------------
              Total |         89      100.00

. codebook recommendation

-------------------------------------------------------------------------------
------------------------------
recommendation
Brokers' recommendation
-------------------------------------------------------------------------------
------------------------------

                  type:  numeric (byte)
                 label:  recommendationl

                 range:  [1,3]                        units:  1
         unique values:  3                        missing .:  3/510

            tabulation:  Freq.   Numeric  Label
                            12         1  Sell
                           141         2  Hold
                           354         3  Buy
                             3         .

. sum dsell dhold dbuy

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       dsell |       510    .0137255    .1164633          0          1
       dhold |       510    .0431373    .2033655          0          1
        dbuy |       510    .1176471    .3225061          0          1
```

```
. reg d.stockprice time dmonth* dsell dhold dbuy
note: dmonth12 omitted because of collinearity

      Source |       SS       df       MS              Number of obs =     407
-------------+------------------------------           F( 15,   391) =    0.56
       Model |  119835.812    15   7989.0541           Prob > F      =  0.9056
    Residual |   5593565.6   391  14305.7944           R-squared     =  0.0210
-------------+------------------------------           Adj R-squared = -0.0166
       Total |  5713401.42   406  14072.4173           Root MSE      =  119.61

------------------------------------------------------------------------------
D.stockprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |  -.0187091   .0331407    -0.56   0.573    -.0838654    .0464471
     dmonth1 |  -17.22418   31.98292    -0.54   0.591    -80.10419    45.65582
     dmonth2 |   8.366207   32.44164     0.26   0.797    -55.41566    72.14808
     dmonth3 |   -6.36872   31.94242    -0.20   0.842    -69.16911    56.43167
     dmonth4 |  -15.74009   31.39928    -0.50   0.616    -77.47264    45.99246
     dmonth5 |   9.979872   31.10391     0.32   0.748    -51.17196     71.1317
     dmonth6 |  -7.017911   31.33446    -0.22   0.823    -68.62302     54.5872
     dmonth7 |  -3.690556   30.53269    -0.12   0.904    -63.71935    56.33824
     dmonth8 |  -23.17961   31.11074    -0.75   0.457    -84.34487    37.98565
     dmonth9 |   4.067236   30.76091     0.13   0.895    -56.41024    64.54472
    dmonth10 |   12.10253   30.34083     0.40   0.690    -47.54905     71.7541
    dmonth11 |   20.53099   31.38569     0.65   0.513    -41.17484    82.23682
    dmonth12 |          0  (omitted)
       dsell |  -61.79124   46.78273    -1.32   0.187    -153.7684    30.18593
       dhold |   8.054849   28.76947     0.28   0.780    -48.50737    64.61706
        dbuy |   26.74404   19.41086     1.38   0.169    -11.41869    64.90676
       _cons |   8.034494   27.81831     0.29   0.773    -46.65769    62.72668
------------------------------------------------------------------------------

. test dsell dhold dbuy

 ( 1)  dsell = 0
 ( 2)  dhold = 0
 ( 3)  dbuy = 0

       F(  3,   391) =    1.32
            Prob > F =    0.2688
```

# Test for first-order autocorrelation

**Model:**

$$\Delta \log(stockprice_t) = \beta_0 + \beta_1 \Delta \log(AEXindex_t) + \beta_2 dsell_t + \beta_3 dhold_t + \beta_4 dbuy_t + \beta_5 t + u_t$$

**Autocorrelation:** $u_t = \alpha_0 + \rho u_{t-1} + e_t$

```
. reg d.ln_stockprice d.ln_AEXindex dsell dhold dbuy time

      Source |       SS       df       MS              Number of obs =     407
-------------+------------------------------           F(  5,   401) =   65.71
       Model |  .072011984     5  .014402397           Prob > F      =  0.0000
    Residual |  .087887341   401   .00021917           R-squared     =  0.4504
-------------+------------------------------           Adj R-squared =  0.4435
       Total |  .159899325   406  .000393841           Root MSE      =   .0148

------------------------------------------------------------------------------
D.           |
ln_stockpr~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_AEXindex |
         D1. |   1.196417   .0666714    17.94   0.000     1.065348    1.327487
             |
       dsell |  -.0084334   .0056931    -1.48   0.139    -.0196254    .0027587
       dhold |  -.0023539   .0034992    -0.67   0.502    -.0092329    .0045252
        dbuy |   .0024244   .0023529     1.03   0.303    -.0022012      .00705
        time |   8.97e-07   3.60e-06     0.25   0.804    -6.18e-06    7.98e-06
       _cons |   -.000561   .0015091    -0.37   0.710    -.0035276    .0024057
------------------------------------------------------------------------------

. estat dwatson

Number of gaps in sample:  102

Durbin-Watson d-statistic(  6,   407) =  1.946667

. reg uhat l.uhat dsell dhold dbuy time

      Source |       SS       df       MS              Number of obs =     304
-------------+------------------------------           F(  5,   298) =    0.96
       Model |  .001197494     5  .000239499           Prob > F      =  0.4443
    Residual |  .074552349   298  .000250176           R-squared     =  0.0158
-------------+------------------------------           Adj R-squared = -0.0007
       Total |  .075749843   303  .000249999           Root MSE      =  .01582

------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |  -.1222667   .0587917    -2.08   0.038    -.2379663   -.0065672
             |
       dsell |  -.0047134   .0074042    -0.64   0.525    -.0192845    .0098577
       dhold |  -.0006479   .0045234    -0.14   0.886    -.0095497    .0082539
        dbuy |   .0018189   .0029198     0.62   0.534    -.0039271    .0075649
        time |  -1.72e-06   4.49e-06    -0.38   0.702    -.0000106    7.12e-06
       _cons |   .0007622   .0018799     0.41   0.685    -.0029374    .0044618
------------------------------------------------------------------------------
```

```
. estat bgodfrey

Number of gaps in sample:  102

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          4.527               1                   0.0334
---------------------------------------------------------------------------
                        H0: no serial correlation
```

- **Conclusion:** first-order autocorrelation

## Test for second-order autocorrelation

**Model:**

$$\Delta \log(stockprice_t) = \beta_0 + \beta_1 \Delta \log(AEXindex_t) + \beta_2 dsell_t + \beta_3 dhold_t + \beta_4 dbuy_t + \beta_5 t + u_t$$

**Autocorrelation:** $u_t = \alpha_0 + \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t$

```
. reg uhat l.uhat l2.uhat  dhold dbuy time

      Source |       SS       df       MS              Number of obs =     202
-------------+------------------------------           F(  5,   196) =    2.88
       Model |  .001853641     5  .000370728           Prob > F      =  0.0157
    Residual |  .025270908   196  .000128933           R-squared     =  0.0683
-------------+------------------------------           Adj R-squared =  0.0446
       Total |  .027124549   201  .000134948           Root MSE      =  .01135

------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |  -.0809593    .045317    -1.79   0.076    -.1703307    .0084122
         L2. |  -.0034182   .0463885    -0.07   0.941    -.0949028    .0880665
             |
       dhold |  -.0066112    .004407    -1.50   0.135    -.0153024      .00208
        dbuy |  -.0068313   .0026362    -2.59   0.010    -.0120302   -.0016325
        time |  -1.20e-06   3.95e-06    -0.30   0.762    -8.99e-06    6.60e-06
       _cons |   .0015463    .001665     0.93   0.354    -.0017373    .0048298
------------------------------------------------------------------------------

. test l.uhat l2.uhat

 ( 1)  L.uhat = 0
 ( 2)  L2.uhat = 0

       F(  2,   196) =    1.60
            Prob > F =    0.2048
```

- **Conclusion:** no second-order autocorrelation

## Compare OLS and FGLS estimates

```
. reg d.ln_stockprice d.ln_AEXindex dsell dhold dbuy time
      Source |       SS       df       MS              Number of obs =     407
-------------+------------------------------           F(  5,   401) =   65.71
       Model |  .072011984     5  .014402397           Prob > F      =  0.0000
    Residual |  .087887341   401   .00021917           R-squared     =  0.4504
-------------+------------------------------           Adj R-squared =  0.4435
       Total |  .159899325   406  .000393841           Root MSE      =   .0148
------------------------------------------------------------------------------
D.           |
ln_stockpr~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_AEXindex |
         D1. |   1.196417   .0666714    17.94   0.000     1.065348    1.327487
             |
       dsell |  -.0084334   .0056931    -1.48   0.139    -.0196254    .0027587
       dhold |  -.0023539   .0034992    -0.67   0.502    -.0092329    .0045252
        dbuy |   .0024244   .0023529     1.03   0.303    -.0022012      .00705
        time |   8.97e-07   3.60e-06     0.25   0.804    -6.18e-06    7.98e-06
       _cons |   -.000561   .0015091    -0.37   0.710    -.0035276    .0024057
------------------------------------------------------------------------------
. prais d.ln_stockprice d.ln_AEXindex dsell dhold dbuy time
Number of gaps in sample:  102
(note: computations for rho restarted at each gap)

Iteration 0:  rho = 0.0000
Iteration 1:  rho = -0.1086
Iteration 2:  rho = -0.1130
Iteration 3:  rho = -0.1132
Iteration 4:  rho = -0.1132
Iteration 5:  rho = -0.1132

Prais-Winsten AR(1) regression -- iterated estimates
      Source |       SS       df       MS              Number of obs =     407
-------------+------------------------------           F(  5,   401) =   67.20
       Model |   .07270838     5  .014541676           Prob > F      =  0.0000
    Residual |  .086769876   401  .000216384           R-squared     =  0.4559
-------------+------------------------------           Adj R-squared =  0.4491
       Total |  .159478256   406  .000392804           Root MSE      =  .01471
------------------------------------------------------------------------------
D.           |
ln_stockpr~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_AEXindex |
         D1. |    1.18964    .065905    18.05   0.000     1.060077    1.319202
             |
       dsell |  -.0104733   .0056516    -1.85   0.065    -.0215837    .0006372
       dhold |  -.0024235   .0034547    -0.70   0.483    -.0092152    .0043681
        dbuy |   .0028566   .0023112     1.24   0.217    -.0016869    .0074002
        time |   7.91e-07   3.31e-06     0.24   0.811    -5.71e-06    7.29e-06
       _cons |  -.0005168   .0013898    -0.37   0.710     -.003249    .0022154
-------------+----------------------------------------------------------------
         rho |  -.1131603
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    1.946667
Durbin-Watson statistic (transformed) 1.754336
```

## Conclusions:

- OLS and FGLS gives same parameters  Correlation is -0.113

**Thus to wind up: line of reasoning of this empirical application**

Step 1- Start with model with both variables in levels. The parameters are estimated with OLS.

$$\Delta \log(stockprice_t) = \beta_0 + \beta_1 \Delta \log(AEXindex_t) + \beta_2 dsell_t + \beta_3 dhold_t + \beta_4 dbuy_t + \beta_5 t + u_t$$

$$u_t = \alpha_0 + \rho u_{t-1} + e_t$$

Step 2 - DWatson: $\hat{\rho}$ of the lagged residual is small
Step 3 - Breusch Godfrey: test for autocorrelation: $\hat{\rho}$ is -0.113 and statistically significant
Step 4 - Prais Winsten (FGLS): $\hat{\rho}$ is small (-0.113)

**Conclusion:**

- autocorrelation does not lead to a major misspecification for this model!
- *t*-values and *F*-values can be interpreted in the usual way.