Utrecht University

Tutorials
Week 2

Utrecht University

Regression Analysis: with time-series data

Pdf file on Blackboard	Dataset on Blackboard		Description
C 10.2: Note: variable "pet" in eq 10.22 is called "gas" in data	barium.dta	Krupp, C. M., & Pollard, P. S. (1996). Market Responses to Antidumping Laws: Some Evidence from the U.S. Chemical Industry. *The Canadian Journal of Economics / Revue Canadienne d'Economique*, *29*(1), 199–227. https://doi.org/10.2307/136159	This exercise is related to example 10.5 from the book. Understanding the model of an economic situation, time series, adding time trends to the equation to analyze the change in the regression, use of dummies, meaning and test for heteroskedasticity, using of F-test to recognize the joint dummy variables, seasonal dummies, seasonal effects, test for seasonality, check for multicollinearity Extra exercises: serial correlation and FGLS vs. OLS
C.12.1	fertil3.dta	Whittington, L. A., Alm, J., & Peters, H. E. (1990). Fertility and the Personal Exemption: Implicit Pronatalist Policy in the United States. *The American Economic Review*, *80*(3), 545–556. http://www.jstor.org/stable/2006683	Weak assumption, strict exogeneity, serial correlation
C. 12 X	wageprc.dta		Distributed lag model (DL), testing serial Correlation, correcting standard errors for AR(1) and AR(2), and differences between effects in the long run and short run.

Utrecht University

Relevant concept for Time Series?

- 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
- Stationary vs. Dynamic models
- Trend
- Seasonality
- The assumption of the error term independently and identically distributed (iid)
- Spurious regression:
- Exogeneity:
- Autoregressive Models (AR)
- Autocorrelation (serial correlation)

Spurious regression

- If we don't correct for time, what will happen? It can be interpreted as an omitted variable if it is not added.
- That may lead to **spurious regressions** = the parameters become inconsistent; the parameters are biased if we don't control for time.
- In time series analyses, you should include t
- There might be an effect of x on y if the trend, t, is not included
- *If y and x are time-dependent*: and it is not controlled for time, then the predictions will be biased.
- More about spurious regression in week 3

Contemporaneous exogeneity vs. strict exogeneity

- Assumption of strict exogeneity:

$$\boxed{E(u_t|X)=0}$$

u_t independent of all x: error term is independent in the past, present, future

- Contemporaneous exogeneity:

$$\boxed{E(u_t|x_{t1,\ldots,}x_{tk})=0}$$

all x are unrelated to the error term in period t (a specific time)

Weak Dependency

- A statististionary time series $\{x_t: t = 1, 2, \ldots, \}$ *is weakly dependent if* x_t *and* x_{t+h} *are "almost independet" as* $h \rightarrow \infty$. Thus,

$$Corr(x_t, x_{t+h}) \rightarrow 0 \text{ as } h \rightarrow \infty$$

C. 10.2 Use the data in BARIUM.RAW for this exercise.

i) Add a linear time trend to equation 10.22 (below). Are any variables, other than the trend, statistically significant?

EXAMPLE 10.5 ANTIDUMPING FILINGS AND CHEMICAL IMPORTS

Krupp and Pollard (1996) analyzed the effects of antidumping filings by U.S. chemical industries on imports of various chemicals. We focus here on one industrial chemical, barium chloride, a cleaning agent used in various chemical processes and in gasoline production. The data are contained in the file BARIUM.RAW. In the early 1980s, U.S. barium chloride producers believed that China was offering its U.S. imports at an unfairly low price (an action known as *dumping*), and the barium chloride industry filed a complaint with the U.S. International Trade Commission (ITC) in October 1983. The ITC ruled in favor of the U.S. barium chloride industry in October 1984. There are several questions of interest in this case, but we will touch on only a few of them. First, were imports unusually high in the period immediately preceding the initial filing? Second, did imports change noticeably after an antidumping filing? Finally, what was the reduction in imports after a decision in favor of the U.S. industry?

To answer these questions, we follow Krupp and Pollard by defining three dummy variables: *befile6* is equal to 1 during the six months before filing, *affile6* indicates the six months after filing, and *afdec6* denotes the six months after the positive decision. The dependent variable is the volume of imports of barium chloride from China, *chnimp*, which we use in logarithmic form. We include as explanatory variables, all in logarithmic form, an index of chemical production, *chempi* (to control for overall demand for barium chloride), the volume of gasoline production, *gas* (another demand variable), and an exchange rate index, *rtwex*, which measures the strength of the dollar against several other currencies. The chemical production index was defined to be 100 in June 1977. The analysis here differs somewhat from Krupp and Pollard in that we use natural logarithms of all variables (except the dummy variables, of course), and we include all three dummy variables in the same regression.

Using monthly data from February 1978 through December 1988 gives the following:

$$\begin{aligned}\widehat{\log(chnimp)} = &-17.80 + 3.12\,\log(chempi) + .196\,\log(gas)\\ &\ \ (21.05)\quad\ (.48)\qquad\qquad\quad (.907)\\ &+ .983\,\log(rtwex) + .060\,befile6 - .032\,affile6 - .565\,afdec6 \qquad [10.22]\\ &\ \ (.400)\qquad\qquad\quad (.261)\qquad\quad (.264)\qquad\quad (.286)\\ n = 131,\ &R^2 = .305,\ \bar{R}^2 = .271.\end{aligned}$$

The equation shows that *befile6* is statistically insignificant, so there is no evidence that Chinese imports were unusually high during the six months before the suit was filed. Further, although the estimate on *affile6* is negative, the coefficient is small (indicating about a 3.2% fall in Chinese imports), and it is statistically very insignificant. The coefficient on *afdec6* shows a substantial fall in Chinese imports of barium chloride after the decision in favor of the U.S. industry, which is not surprising. Since the effect is so large, we compute the exact percentage change: $100[\exp(-.565) - 1] \approx -43.2\%$. The coefficient is statistically significant at the 5% level against a two-sided alternative.

The coefficient signs on the control variables are what we expect: an increase in overall chemical production increases the demand for the cleaning agent. Gasoline production does not affect Chinese imports significantly. The coefficient on log(*rtwex*) shows that an increase in the value of the dollar relative to other currencies increases the demand for Chinese imports, as is predicted by economic theory. (In fact, the elasticity is not statistically different from 1. Why?)

Utrecht University

Solution:

First we estimate the model:

```
. reg lchnimp lchempi lgas lrtwex befile6 affile6 afdec6

      Source |       SS           df       MS      Number of obs   =       131
-------------+----------------------------------   F(6, 124)       =      9.06
       Model |  19.4051607         6  3.23419346   Prob > F        =    0.0000
    Residual |  44.2470875       124  .356831351   R-squared       =    0.3049
-------------+----------------------------------   Adj R-squared   =    0.2712
       Total |  63.6522483       130  .489632679   Root MSE        =    .59735

------------------------------------------------------------------------------
     lchnimp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lchempi |   3.117193   .4792021     6.50   0.000     2.168718    4.065668
        lgas |   .1963504   .9066172     0.22   0.829    -1.598099      1.9908
      lrtwex |   .9830183   .4001537     2.46   0.015     .1910022    1.775034
     befile6 |   .0595739   .2609699     0.23   0.820    -.4569585    .5761064
     affile6 |  -.0324064   .2642973    -0.12   0.903    -.5555249     .490712
      afdec6 |   -.565245   .2858352    -1.98   0.050    -1.130993    .0005028
       _cons |    -17.803   21.04537    -0.85   0.399    -59.45769    23.85169
------------------------------------------------------------------------------
```

- lchnimp: log(chnimp)
- lgas: log(gas)
- lrtwex: log(rtwex)
- befile6: =1 for all 6 mos before filing
- affile6: =1 for all 6 mos after filing
- afdec6: =1 for all 6 mos after decision

Now the assignment asks us to include a linear trend in the specification:

```
. reg lchnimp lchempi lgas lrtwex befile6 affile6 afdec6 t

      Source |       SS       df       MS              Number of obs =     131
-------------+------------------------------           F(  7,   123) =    9.95
       Model |  23.0142898     7  3.28775569           Prob > F      =  0.0000
    Residual |  40.6379584   123  .330389906           R-squared     =  0.3616
-------------+------------------------------           Adj R-squared =  0.3252
       Total |  63.6522483   130  .489632679           Root MSE      =   .5748

------------------------------------------------------------------------------
     lchnimp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lchempi |  -.6862364   1.239711    -0.55   0.581    -3.140169    1.767696
        lgas |   .4656786   .8761779     0.53   0.596    -1.268662    2.200019
      lrtwex |   .0782237     .47244     0.17   0.869    -.8569423     1.01339
     befile6 |     .09047   .2512887     0.36   0.719    -.4069406    .5878805
     affile6 |   .0970062   .2573131     0.38   0.707    -.4123294    .6063417
      afdec6 |  -.3515018   .2825417    -1.24   0.216    -.9107758    .2077722
           t |   .0127058   .0038443     3.31   0.001     .0050963    .0203153
       _cons |  -2.367526   20.78216    -0.11   0.909    -43.50455     38.7695
------------------------------------------------------------------------------
```

- Except for the time trend, we find that none of the explanatory variables yields statistically significant coefficients at 10%.
- A possible explanation is that the results in the original specification were driven by a common trend in our variables.
- Once we explicitly include the trend in the model, there is no statistical evidence for a structural relationship anymore between our explanatory variables and the dependent variable.

ii) In the equation estimated in part i), test for joint significance of all variables except the time trend. What do you conclude?

Solution:

- The model is:

$$\log(chinimp_t) = \beta_0 + \beta_1 \log(chempi_t) + \beta_2 \log(gas_t) + \beta_3 \log(rtwex_t) + \beta_4 befile6_t + \beta_5 affile6_t + \beta_6 afdec6_t + \beta_7 t + u_t$$

- Hypotheses:

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ (not jointly significant)

H_1: H_o *is not true*. This occurs if at least one of the coefficients listed in Ho differs from zero.

- If the Fstat > Fcv, then reject Ho.
- 0.54 < 2.17, we can not reject Ho at the 5% significance level, and the variables are not jointly significant.

```
. test lchempi lgas lrtwex befile6 affile6 afdec6

 ( 1)  lchempi = 0
 ( 2)  lgas = 0
 ( 3)  lrtwex = 0
 ( 4)  befile6 = 0
 ( 5)  affile6 = 0
 ( 6)  afdec6 = 0

       F(  6,   123) =    0.54
            Prob > F =    0.7767
```

ii) In the equation estimated in part i), test for joint significance of all variables except the time trend. What do you conclude?

Solution:

- Conclusion: the coefficients of the explanatory variables (except the linear trend) do not improve the model in a statistically significant way. The do not have explanatory power.

- The degrees of freedom of the numerator is 6, because we have 6 parameter restrictions in Ho and the degrees of freedom of the denominator is 131-8=123, because in the unrestricted model we estimated 8 parameters with sample size 131.

iii) Add monthly dummy variables to the equation and test for seasonality. Does including the monthly dummies change any other estimated or their standard error in important ways?

We add 11-month dummies. Remember, one of the dummies must be omitted to avoid the dummy variable trap (perfect multicollinearity). In this case, the month January is excluded.

```
. reg  lchnimp lchempi lgas lrtwex befile6 affile6 afdec6 t feb mar apr may jun jul aug sep oct nov dec
```

Source	SS	df	MS
Model	26.1336807	18	1.45187115
Residual	37.5185675	112	.33498721
Total	63.6522483	130	.489632679

Number of obs =	131
F(18, 112) =	4.33
Prob > F =	0.0000
R-squared =	0.4106
Adj R-squared =	0.3158
Root MSE =	.57878

lchnimp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lchempi	-.4516555	1.271528	-0.36	0.723	-2.971026	2.067715
lgas	-.820624	1.345056	-0.61	0.543	-3.485679	1.844431
lrtwex	-.1971415	.5295314	-0.37	0.710	-1.24634	.852057
befile6	.1648509	.2569789	0.64	0.523	-.3443198	.6740216
affile6	.1534004	.2719856	0.56	0.574	-.3855043	.692305
afdec6	-.2950163	.2994276	-0.99	0.327	-.8882937	.298261
t	.0123389	.0039163	3.15	0.002	.0045793	.0200985
feb	-.3554148	.293754	-1.21	0.229	-.9374507	.2266211
mar	.062566	.254858	0.25	0.807	-.4424025	.5675345
apr	-.4406149	.258398	-1.71	0.091	-.9525974	.0713676
may	.031299	.2591998	0.12	0.904	-.4822721	.5448702
jun	-.20095	.2592134	-0.78	0.440	-.7145481	.312648
jul	.0111115	.2683777	0.04	0.967	-.5206446	.5428675
aug	-.1271137	.2677917	-0.47	0.636	-.6577086	.4034812
sep	-.0751929	.2583502	-0.29	0.772	-.5870807	.4366949
oct	.0797627	.2570514	0.31	0.757	-.4295517	.5890771
nov	-.2603032	.2530623	-1.03	0.306	-.7617136	.2411073
dec	.0965326	.2615525	0.37	0.713	-.4217002	.6147654
_cons	27.30007	31.39707	0.87	0.386	-34.90919	89.50934

F(11, 112) = 0,85
Prob > F = 0,5943

$H_{o:}\ \delta_2 = 0, \delta_3 = 0, \dots, \delta_{12} = 0$ (no seasonality)
$H_{1:}\ H_0\ not\ true$

$If\ Fstat > Fcv, reject\ Ho$
0,85 < 1.91 at 5% significance level.
We fail to reject the Ho, there is no seasonal effect.

Utrecht University

These exercises are not part of the book: Now check for homoskedasticity and serial correlation , considering strict exogeneity

```
. predict uhat, residuals

. gen uhat2=uhat^2

. reg uhat2 lchempi lgas lrtwex befile6 affile6 afdec6
```

Source	SS	df	MS
Model	2.55224567	6	.425374279
Residual	41.2386262	124	.332569566
Total	43.7908719	130	.33685286

Number of obs	=	131
F(6, 124)	=	1.28
Prob > F	=	0.2717
R-squared	=	0.0583
Adj R-squared	=	0.0127
Root MSE	=	.57669

uhat2	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lchempi	-1.141681	.4626243	-2.47	0.015	-2.057344	-.2260176
lgas	-.2023653	.8752532	-0.23	0.818	-1.934737	1.530006
lrtwex	-.3284217	.3863105	-0.85	0.397	-1.093038	.4361951
befile6	.0064	.2519417	0.03	0.980	-.4922632	.5050633
affile6	.0158846	.2551541	0.06	0.950	-.4891368	.520906
afdec6	.1584464	.2759468	0.57	0.567	-.3877297	.7046224
_cons	11.95703	20.31732	0.59	0.557	-28.25663	52.17069

```
. hettest, rhs fstat

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: i.i.d. error terms
Variables: All independent variables

H0: Constant variance

F(6, 124) =   0.80
 Prob > F = 0.5717
```

Ho: $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = 0$

H1: Ho is not true (heteroskedasticity)

If Ftest > Fcv at 5%, then reject Ho

0.80 < 2.10 we fail to reject Ho.

Conclusion: homoskedasticity holds.

These exercises are not part of the book:
Now check for homoskedasticity and serial correlation, considering strict exogeneity

OLS no longer provides an unbiased estimator in the presence of **serial correlation**, leading to invalid t-statistics.

```
. tsset t

Time variable: t, 1 to 131
        Delta: 1 unit

. reg uhat l.uhat lchempi lgas lrtwex befile6 affile6 afdec6

      Source |       SS           df       MS      Number of obs   =       130
-------------+----------------------------------   F(7, 122)       =      1.41
       Model |  3.31972295         7  .474246136   Prob > F        =    0.2058
    Residual |  40.9269683       122  .335466954   R-squared       =    0.0750
-------------+----------------------------------   Adj R-squared   =    0.0220
       Total |  44.2466913       129  .342997607   Root MSE        =     .5792

------------------------------------------------------------------------------
        uhat | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |   .2772056   .0881206     3.15   0.002     .1027622    .4516491
             |
     lchempi |  -.0147775   .4680184    -0.03   0.975    -.9412666    .9117116
        lgas |   .3318712   .8927836     0.37   0.711    -1.435483    2.099225
      lrtwex |   .1046564   .3919403     0.27   0.790    -.6712285    .8805414
     befile6 |  -.0268293    .253184    -0.11   0.916    -.5280324    .4743737
     affile6 |  -.0654893   .2571084    -0.25   0.799     -.574461    .4434824
      afdec6 |  -.0398831   .2774596    -0.14   0.886     -.589142    .5093758
       _cons |  -7.993576   20.79624    -0.38   0.701    -49.16181    33.17466
------------------------------------------------------------------------------
```

Ho: $\rho = 0$ (no 1st order serial correlation)
H_1: $\rho \neq 0$ (serial correlation)
If |t|> t_cv, reject Ho.
3.15 > 1.980 at 5%. Therefore, we reject Ho.
There is evidence of serial correlation.

These exercises are not part of the book:

Apply the Prais-Winsten Test and compare OLS vs. FGLS coefficients:

Can we argue - based on the OLS estimators - that the effect of the Chinese imports (afdec6) after the International Trade Commission's decision is less statistically significant? (t-statistics : -1.98)

Serial Correlation

- Serial correlation is also called autocorrelation and means that, conditional on X, the error term is correlated over subsequent time periods.

$$\boxed{Corr\ (\varepsilon_t, \varepsilon_{t-1}|X) \neq 0}$$

- Positive serial correlation: $Corr\ (\varepsilon_t, \varepsilon_{t-1}) > 0$, the error in period t is likely to be positive if the error in period t-1 is positive.
- Negative serial correlation: $Corr\ (\varepsilon_t, \varepsilon_{t-1}) < 0$, the error in period t is likely to be negative if the error in period t-1 is positive.

Serial Correlation

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$$

- Consider the model:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

- If the error term of this model is serially correlated, we can write:
- ρ shows **the correlation between** ε_t **and** ε_{t-1}
- Thus $|\rho| \leq 1$
- The close $|\rho|$ is to 1, the stronger the serial correlation.
- If ρ =0, there is no serial correlation.
- ε_t depends on ε_t-1 ; u_t is a serially correlated error term with a mean of zero and a constant variance.
- This is first order serial correlation because the error in t depends on the error in t-1

$$\boxed{Hypothesis\ for\ non\ serial\ correlation:} \quad \boxed{\begin{array}{l} H_0 = \rho = 0 \\ H_1 = \rho \neq 0 \end{array}}$$

What are the consequences of serial correlation

- If serial correlation exists, then $\widehat{\sigma^2}$ and therefore s.e. $\widehat{\beta_k}$ are incorrect.
- Hence, t-test and F-tests are invalid: cannot perform hypothesis tests.
- Import to know:
- Serial correlation in static or finite distributed lag models cause biased standard errors
- However, serial correlation does not cause bias in the estimated coefficients – with one important exception:

⇒ Serial correlation in autoregressive models (i.e. models with a lagged dependent variable) cause:

⇒ Biased standard errors

⇒ Biased coefficient estimates

How to test for serial correlation?

Two commonly used tests for serial correlation:

- Durbin-Watson test
- Breusch-Godfrey test

What to do if there is serial correlation?

- Add more lags
- In terms of economics, the t statistics are wrong with autocorrelation. Therefore, reconsidering the dynamic structure can help.

Breusch-Godfrey test (BG-test)

We want to test for first-order serial correlation in the following time series model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \qquad \text{with } |\rho| < 1$$

Stata commands:

1. Estimate the equation above.
2. Compute the residuals e_t from the estimated equation
3. Regress e_t on the 1-period lagged residual e_{t-1} and all independent variables from the equation above.
4. Use the t-value on e_{t-1} to test:

```
. reg consumption price income temp
. predict uhat, resid
. reg uhat l.uhat price income temp
```

$$H_0 \ : \ \rho = 0 \qquad \text{no 1st-order serial correlation}$$
$$H_A \ : \ \rho \neq 0 \qquad \text{1st-order serial correlation}$$

Utrecht University

C. 12.1 In example 11.6, we estimated a finite DL model in first differences:

$$\Delta gfr_t = \gamma_0 + \delta_0 \Delta pe_t + \delta_1 \Delta pe_{t-1} + \delta_2 \Delta pe_{t-2} + u_t$$

Use the data FERTILE3.RAW to test whether there is AR(1) serial correlation in the errors

First run the regression and retrieve the residuals "uhat"

```
. tset t
        time variable:  t, 1 to 72
                delta:  1 unit

. reg D. gfr D.pe L1.D.pe L2.D.pe

      Source |       SS       df       MS              Number of obs =      69
-------------+------------------------------           F(  3,    65) =    6.56
       Model |  293.259859     3  97.7532864           Prob > F      =  0.0006
    Residual |  968.199959    65   14.895384           R-squared     =  0.2325
-------------+------------------------------           Adj R-squared =  0.1971
       Total |  1261.45982    68  18.5508797           Root MSE      =  3.8595

------------------------------------------------------------------------------
       D.gfr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pe |
         D1. |  -.0362021   .0267737    -1.35   0.181     -.089673    .0172687
         LD. |  -.0139706   .0275539    -0.51   0.614    -.0689997    .0410584
        L2D. |   .1099896   .0268797     4.09   0.000     .0563071    .1636721
             |
       _cons |  -.9636787   .4677599    -2.06   0.043     -1.89786   -.0294976
------------------------------------------------------------------------------

. predict uhat, residual
(3 missing values generated)
```

11.6 FERTILITY EQUATION

In Example 10.4, we explained the general fertility rate, *gfr*, in terms of the value of the personal exemption, *pe*. The first order autocorrelations for these series are very large: $\hat{\rho}_1 = .977$ for *gfr* and $\hat{\rho}_1 = .964$ for *pe*. These autocorrelations are highly suggestive of unit root behavior, and they raise serious questions about our use of the usual OLS *t* statistics for this example back in Chapter 10. Remember, the *t* statistics only have exact *t* distributions under the full set of classical linear model assumptions. To relax those assumptions in any way and apply asymptotics, we generally need the underlying series to be I(0) processes.

We now estimate the equation using first differences (and drop the dummy variable, for simplicity):

$$\Delta\widehat{gfr} = -.785 - .043\,\Delta pe \tag{11.26}$$
$$(.502)\quad(.028)$$
$$n = 71,\ R^2 = .032,\ \bar{R}^2 = .018.$$

Now, an increase in *pe* is estimated to lower *gfr* contemporaneously, although the estimate is not statistically different from zero at the 5% level. This gives very different results than when we estimated the model in levels, and it casts doubt on our earlier analysis.

If we add two lags of Δpe, things improve:

$$\Delta\widehat{gfr} = -.964 - .036\,\Delta pe - .014\,\Delta pe_{-1} + .110\,\Delta pe_{-2} \tag{11.27}$$
$$(.468)\quad(.027)\qquad(.028)\qquad(.027)$$
$$n = 69,\ R^2 = .233,\ \bar{R}^2 = .197.$$

Even though Δpe and Δpe_{-1} have negative coefficients, their coefficients are small and jointly insignificant (*p*-value = .28). The second lag is very significant and indicates

Perform the **Breusch-Godfrey** test, for first-order autocorrelation:

First, under the assumption of strict exogeneity:

$$E(u_t \mid X) = 0 .$$

u_t independent of all x: error term is independent in the past, present, future

```
. reg uhat L.uhat

      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =      6.12
       Model |  82.0925482         1  82.0925482   Prob > F        =    0.0160
    Residual |  885.696544        66  13.4196446   R-squared       =    0.0848
-------------+----------------------------------   Adj R-squared   =    0.0710
       Total |  967.789092        67  14.4446133   Root MSE        =    3.6633

------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |   .2918091   .1179825     2.47   0.016     .0562495    .5273687
             |
       _cons |   .0180209   .4442522     0.04   0.968    -.8689571     .904999
------------------------------------------------------------------------------
```

Second, under the assumption of weak exogeneity:

$$E(u_t \mid x_{t1}, \ldots x_{tk}) = 0 .$$

all x are unrelated to the error term in period t (a specific time)

Utrecht University

```
. reg uhat L.uhat D.pe L1.D.pe L2.D.pe

      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(4, 63)        =      1.48
       Model |  83.1528991         4  20.7882248   Prob > F        =    0.2188
    Residual |  884.636193        63  14.0418443   R-squared       =    0.0859
-------------+----------------------------------   Adj R-squared   =    0.0279
       Total |  967.789092        67  14.4446133   Root MSE        =    3.7472

------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |     .29556   .1214561     2.43   0.018     .0528494    .5382706
             |
          pe |
         D1. |  -.0071763   .0261613    -0.27   0.785    -.0594554    .0451028
         LD. |   .0015739   .0267612     0.06   0.953    -.0519041     .055052
        L2D. |  -.0009125   .0261017    -0.03   0.972    -.0530725    .0512476
             |
       _cons |   .0262151   .4575848     0.06   0.954    -.8881952    .9406254
------------------------------------------------------------------------------
```

Ho: $\rho = 0$ (no 1st order serial correlation)
H_1: $\rho \neq 0$ (serial correlation)
If|t|> tcv, then reject Ho
2.43 > 2.00, reject Ho.
There is evidence for serial correlation.
Also with the p-value:
Pvalue: 0.018<0.05, reject Ho

We reject the Ho of no first-order autocorrelation at 5% level of significance because lagged uhat is significant at 5% level (p-value = 0,018)

Conclusion: there is empirical evidence in favor of first-order autocorrelation. We need to correct standard errors using e.g. the Newey estimator.

Utrecht University

After that, apply the Newey Correction / estimator (OLS + adjusted standard errors):

```
. newey D. gfr D.pe L1.D.pe L2.D.pe, lag(1)

Regression with Newey-West standard errors          Number of obs  =        69
maximum lag: 1                                      F(  3,     65)  =      8.19
                                                    Prob > F        =    0.0001
```

D.gfr	Coef.	Newey-West Std. Err.	t	P>|t|	[95% Conf.	Interval]
pe						
D1.	-.0362021	.0371648	-0.97	0.334	-.1104253	.0380211
LD.	-.0139706	.0345874	-0.40	0.688	-.0830465	.0551052
L2D.	.1099896	.0276013	3.98	0.000	.0548661	.1651131
_cons	-.9636787	.5219555	-1.85	0.069	-2.006096	.0787384

C12.X
i) Using the data in WAGEPRC.DTA, estimate the distributed lag-model from problem 11.5 . Using regressions (12.14) to test for AR(1) serial correlation.

11.5 For the U.S. economy, let *gprice* denote the monthly growth in the overall price level and let *gwage* be the monthly growth in hourly wages. [These are both obtained as differences of logarithms: $gprice = \Delta\log(price)$ and $gwage = \Delta\log(wage)$.] Using the monthly data in WAGEPRC.RAW, we estimate the following distributed lag model:

$$\begin{aligned}\widehat{gprice} = &-\underset{(.00057)}{.00093} + \underset{(.052)}{.119}\, gwage + \underset{(.039)}{.097}\, gwage_{-1} + \underset{(.039)}{.040}\, gwage_{-2}\\ &+ \underset{(.039)}{.038}\, gwage_{-3} + \underset{(.039)}{.081}\, gwage_{-4} + \underset{(.039)}{.107}\, gwage_{-5} + \underset{(.039)}{.095}\, gwage_{-6}\\ &+ \underset{(.039)}{.104}\, gwage_{-7} + \underset{(.039)}{.103}\, gwage_{-8} + \underset{(.039)}{.159}\, gwage_{-9} + \underset{(.039)}{.110}\, gwage_{-10}\\ &+ \underset{(.039)}{.103}\, gwage_{-11} + \underset{(.052)}{.016}\, gwage_{-12}\end{aligned}$$

$$n = 273,\ R^2 = .317,\ \bar{R}^2 = .283.$$

We first estimate the distributed lagged (DL) model and next, we use the Breusch-Godfrey test for autocorrelation (to test for AR(1) serial correlation)

Just to remember: lags help us to mitigate the statistical problems of autocorrelation. What we will see are dynamic regression models.

```
tsset t
        time variable:  t, 1 to 286
                delta:  1 unit

. reg gprice gwage gwage_1-gwage_12

      Source |       SS           df       MS      Number of obs   =       273
-------------+----------------------------------   F(13, 259)      =      9.25
       Model |  .000981458        13  .000075497   Prob > F        =    0.0000
    Residual |  .002113658       259  8.1608e-06   R-squared       =    0.3171
-------------+----------------------------------   Adj R-squared   =    0.2828
       Total |  .003095116       272  .000011379   Root MSE        =    .00286

------------------------------------------------------------------------------
      gprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gwage |   .1190416   .0517725     2.30   0.022     .0170929    .2209903
     gwage_1 |   .0972174   .0390409     2.49   0.013     .0203393    .1740954
     gwage_2 |   .0399518   .0390717     1.02   0.307    -.0369869    .1168905
     gwage_3 |   .0382652   .0391513     0.98   0.329    -.0388301    .1153605
     gwage_4 |   .0813362   .0393483     2.07   0.040     .0038528    .1588195
     gwage_5 |    .106852   .0391937     2.73   0.007     .0296731    .1840308
     gwage_6 |   .0949731   .0392186     2.42   0.016     .0177451    .1722011
     gwage_7 |   .1037922   .0393788     2.64   0.009     .0262488    .1813355
     gwage_8 |   .1025629   .0394884     2.60   0.010     .0248037     .180322
     gwage_9 |   .1585079   .0393341     4.03   0.000     .0810526    .2359632
    gwage_10 |   .1104412   .0392229     2.82   0.005     .0332049    .1876776
    gwage_11 |   .1033206   .0394388     2.62   0.009     .0256591     .180982
    gwage_12 |   .0156575   .0518343     0.30   0.763    -.0864128    .1177278
       _cons |  -.0009296   .0005662    -1.64   0.102    -.0020445    .0001853
------------------------------------------------------------------------------
```

Breush-Godfrey test for first-order autocorrelation, assuming weak/contemporaneous exogeneity (and not strict exogeneity):

```
. predict uhat, residual
(13 missing values generated)

. reg  uhat L.uhat gwage L. gwage L2. gwage L3. gwage L4. gwage L5. gwage L6. gwage L7. gwage L8. gwage

      Source |       SS       df       MS              Number of obs =     272
-------------+------------------------------           F( 14,   257) =    6.38
       Model |  .00054188     14  .000038706           Prob > F      =  0.0000
    Residual |  .00155816    257  6.0629e-06           R-squared     =  0.2580
-------------+------------------------------           Adj R-squared =  0.2176
       Total | .002100039    271  7.7492e-06           Root MSE      =  .00246

------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |   .5106527   .0540557     9.45   0.000     .4042042    .6171013
             |
       gwage |
         --. |   -.049533   .0453685    -1.09   0.276    -.1388743    .0398084
         L1. |  -.0102465   .0337572    -0.30   0.762    -.0767223    .0562294
         L2. |  -.0045952   .0338253    -0.14   0.892    -.0712052    .0620148
         L3. |    .002263   .0337474     0.07   0.947    -.0641936    .0687196
         L4. |   .0003814   .0340404     0.01   0.991    -.0666522    .0674151
         L5. |   .0070267   .0339435     0.21   0.836    -.0598162    .0738696
         L6. |   -.000841   .0338047    -0.02   0.980    -.0674105    .0657285
         L7. |  -.0051064   .0339603    -0.15   0.881    -.0719824    .0617696
         L8. |   .0041929   .0340396     0.12   0.902    -.0628392    .0712249
         L9. |   .0008098   .0339042     0.02   0.981    -.0659557    .0675752
        L10. |   .0027298   .0338628     0.08   0.936    -.0639541    .0694136
        L11. |   .0010463   .0340397     0.03   0.976    -.0659859    .0680784
        L12. |   .0310227   .0451507     0.69   0.493    -.0578898    .1199351
             |
       _cons |   .0001144   .0004896     0.23   0.815    -.0008497    .0010786
------------------------------------------------------------------------------
```

Ho = $\rho = 0, \textit{no serial correlation}$
H1 = $\rho \neq 0,\ \textit{serial correlation}$
|t-value| > t-critical value, reject Ho.
9.45 >1.960 (5% significance level)
We reject Ho, there is significant evidence for serial correlation among the residuals.

Conclude: the empirical evidence is in favor of first order autocorrelation (test based on the results corresponding the uhat L.1)

Also applying the Stata command for the Breush-Godfrey test for first-order autocorrelation

```
. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	12.021	1	0.0005

```
H0: no serial correlation
```

Ho = $no\ serial\ correlation$
H1 = $serial\ correlation$
If chi2 > chi2_cv, at 5% significance level, reject Ho
12.021 > 3.84, reject Ho.
There is enough evidence in favor of serial correlation

ii) Reestimate the model using standard errors that are corrected for the presence of first and/or second-order autocorrelation. You may use the “newey”stata-command for this. Does it matter if you correct for AR(1) or AR(2) error terms?

If there is serial correlation, we need them to correct standard error. How? To correct autocorrelation, we need use the Newey estimator of the Prais-Winston. The Newey correction/estimator (OLS + adjust standard errors)

```
newey gprice gwage gwage_1-gwage_12, lag(1)
est store model1
```

```
Regression with Newey-West standard errors          Number of obs     =        273
maximum lag: 1                                      F( 13,        259) =       7.45
                                                    Prob > F          =     0.0000
```

gprice	Coef.	Newey-West Std. Err.	t	P>|t|	[95% Conf.	Interval]
gwage	.1190416	.0613043	1.94	0.053	-.0016767	.2397598
gwage_1	.0972174	.0346645	2.80	0.005	.0289572	.1654775
gwage_2	.0399518	.0385938	1.04	0.302	-.0360458	.1159493
gwage_3	.0382652	.0372509	1.03	0.305	-.0350879	.1116184
gwage_4	.0813362	.040698	2.00	0.047	.001195	.1614773
gwage_5	.106852	.041456	2.58	0.011	.0252183	.1884856
gwage_6	.0949731	.0452216	2.10	0.037	.0059243	.1840219
gwage_7	.1037922	.0390687	2.66	0.008	.0268595	.1807248
gwage_8	.1025629	.0403154	2.54	0.012	.0231751	.1819506
gwage_9	.1585079	.042514	3.73	0.000	.0747908	.2422249
gwage_10	.1104412	.0430844	2.56	0.011	.0256009	.1952816
gwage_11	.1033206	.0471255	2.19	0.029	.0105226	.1961186
gwage_12	.0156575	.0538974	0.29	0.772	-.0904754	.1217904
_cons	-.0009296	.0006177	-1.50	0.134	-.002146	.0002868

```
newey gprice gwage gwage_1-gwage_12, lag(2)
est store model2
```

```
Regression with Newey-West standard errors          Number of obs     =        273
maximum lag: 2                                      F( 13,        259) =       5.96
                                                    Prob > F          =     0.0000
```

gprice	Coef.	Newey-West Std. Err.	t	P>|t|	[95% Conf.	Interval]
gwage	.1190416	.0596428	2.00	0.047	.001595	.2364881
gwage_1	.0972174	.0360511	2.70	0.007	.0262268	.1682079
gwage_2	.0399518	.0390969	1.02	0.308	-.0370364	.11694
gwage_3	.0382652	.0381432	1.00	0.317	-.0368451	.1133755
gwage_4	.0813362	.0409207	1.99	0.048	.0007566	.1619157
gwage_5	.106852	.0412433	2.59	0.010	.0256371	.1880668
gwage_6	.0949731	.0455674	2.08	0.038	.0052435	.1847028
gwage_7	.1037922	.0399055	2.60	0.010	.0252117	.1823727
gwage_8	.1025629	.0424289	2.42	0.016	.0190133	.1861124
gwage_9	.1585079	.0442851	3.58	0.000	.0713033	.2457125
gwage_10	.1104412	.0438436	2.52	0.012	.0241059	.1967765
gwage_11	.1033206	.0461922	2.24	0.026	.0123604	.1942807
gwage_12	.0156575	.050763	0.31	0.758	-.0843033	.1156183
_cons	-.0009296	.0006932	-1.34	0.181	-.0022946	.0004355

```
est table model1 model2, se


. reg uhat l.uhat l2.uhat gwage gwage_1-gwage_12

      Source |       SS           df       MS      Number of obs   =       271
-------------+----------------------------------   F(15, 255)      =      6.88
       Model |  .000603529        15  .000040235   Prob > F        =    0.0000
    Residual |   .00149082       255  5.8464e-06   R-squared       =    0.2882
-------------+----------------------------------   Adj R-squared   =    0.2463
       Total |  .002094349       270  7.7568e-06   Root MSE        =    .00242

------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |   .4038757   .0617549     6.54   0.000     .2822611    .5254902
         L2. |   .2082125   .0614613     3.39   0.001      .087176     .329249
             |
       gwage |  -.0449009    .044572    -1.01   0.315    -.1326771    .0428753
     gwage_1 |  -.0218605   .0334011    -0.65   0.513    -.0876376    .0439167
     gwage_2 |  -.0065564   .0333377    -0.20   0.844    -.0722087    .0590959
     gwage_3 |   .0011153   .0332849     0.03   0.973     -.064433    .0666635
     gwage_4 |   .0007893    .033429     0.02   0.981    -.0650429    .0666214
     gwage_5 |   .0080186   .0334868     0.24   0.811    -.0579273    .0739645
     gwage_6 |   .0000823   .0333798     0.00   0.998    -.0656529    .0658175
     gwage_7 |  -.0060688   .0333518    -0.18   0.856    -.0717489    .0596112
     gwage_8 |   .0025501   .0334564     0.08   0.939     -.063336    .0684362
     gwage_9 |   .0032809   .0333054     0.10   0.922    -.0623077    .0688695
    gwage_10 |   .0035026   .0332565     0.11   0.916    -.0619897     .068995
    gwage_11 |   .0005685   .0334864     0.02   0.986    -.0653766    .0665135
    gwage_12 |   .0271583   .0443647     0.61   0.541    -.0602096    .1145262
       _cons |    .000176   .0004821     0.36   0.715    -.0007735    .0011254
------------------------------------------------------------------------------
```

- By comparing this table AR(2) with the AR(1) we recognize that the S.E. did change a few, but is not really significant.
- We can conclude that AR(2) shows more empirical evidence in favor of AR(2)
- Does it matter if you correct the error term for AR(1) or AR(2)? No, almost not.

iii) Given the results of ii), what are the long run effects?
These can be obtained using “nlcom”command, but it could be done manually too.

Short-run effect: The contemporaneous effect (or short-run effect) is the parameter that registers the effect of the independent variable on the dependent one in the same period. It refers to the same period t. Short run effect here is δ_1=0.12

Long-run effect includes the change in the dependent variable in all periods due to changes in the independent variables. effect of the x on y. use the coefficient of the AR(2)

```
In stata: nlcom (_b[xt])+ _b([l.x])+ _b([l2.x]).....
but first reg of AR(2) error terms


. nlcom _b[gwage] + _b[gwage_1] + _b[gwage_2] + _b[gwage_3] + _b[gwage_4] + _b[gwage_5] +
 _b[gwage_6] + _b[gwage_7] + _b[
> gwage_8] + _b[gwage_9] + _b[gwage_10] + _b[gwage_11] + _b[gwage_12]

       _nl_1:  _b[gwage] + _b[gwage_1] + _b[gwage_2] + _b[gwage_3] + _b[gwage_4] + _b[
gwage_5] + _b[gwage_6] + _b[gwage_7
> ] + _b[gwage_8] + _b[gwage_9] + _b[gwage_10] + _b[gwage_11] + _b[gwage_12]

------------------------------------------------------------------------------
      gprice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   1.171919   .1479148     7.92   0.000     .8820116    1.461827
```

Utrecht University

Sharing science,
shaping tomorrow