Problem Set Three

Josh Virene

February 2024

1 Question One: Forecasts and Non-stationary time series.

Forecasts and Non-stationary time series. Let time series Y_t satisfy $y_t = \mu + \rho_{y_(t-1)} + \epsilon_t$ where ϵ_t is an i.i.d. Gaussian process with zero mean and variance σ_ϵ^2.

(a) Assume that the process Y_t is stationary, i.e., $|\rho| < 1$. Compute unconditional mean EY_t

We know $E(\epsilon_t) = 0$ as it is i.i.d, $E(\mu) = \mu$, $E(\rho Y_t) = \rho(E(y_{t-1}))$ for all $E(Y_t)$
$E(Y_t) = E(\mu) + E(\rho_{y_(t-1)}) + E(\epsilon_t)$
$E(Y_t) = \mu + \rho(EY_t)$
$E(Y_t) - \rho(EY_t) = \mu$
$E(Y_t)(1 - \rho) = \mu$
$E(Y_t) = \mu/(1 - \rho)$

(b) Compute conditional expectation $E(y_{t+h}|y_t)$. For integer h ¿ 0
$E(Y_{t+1}) = E(\mu|Y_t) + \rho(E(Y_t|Y_t) + E(\epsilon_i|Y_t)$
$= \mu * 1 + \rho EY_t$
$E(Y_{t+2}) = E(\mu|Y_t) + \rho(E(Y_{t+1}|Y_t) + E(\epsilon_i|Y_t)$
$= \mu(1 + \rho) + \rho^2 Y_t$
$E(Y_{t+3}) = E(\mu|Y_t) + \rho(E(Y_{t+2}|Y_t) + E(\epsilon_i|Y_t)$
$= \mu(1 + \rho + \rho^2) + \rho^3 Y_t$
These establish the pattern needed for our final result.
$E(Y_{t+h}) = E(\mu|Y_t) + \rho(E(Y_{t+h}|Y_t)$
$= \mu(1 + \rho^{h-1} + \rho^h) + \rho^h * Y_t$

Finally, apply the equation for a geometric sequence, our answer becomes $E(Y_{t+h}) = \mu(1 - \rho^h)/(1 - \rho) + (p^h * Y_t)$
(c) Compute $lim_h - > \infty E(y_{t+h}|y_t)$ Does it depend on y_t?
In the previous expression, $E(Y_{t+h}) = \mu(1 - \rho^h)/(1 - \rho) + (p^h * Y_t)$
We know that both the terms ρ^h will become zero as $h- > \infty$, because $|\rho| < 1$.
The expression then becomes $\mu/1 - \rho$, which does not depend on Y_t

(d) Now assume that $\rho = 1$. Compute $lim_h - > \infty E(y_{t+h}|y_t)$. Does it depend on y_t?
$E(Y_{t+h}) = \mu(1-\rho^h)/(1-\rho) + (p^h * Y_t)$
The expression simplifies to $1^h Y_t$, where $\rho = 1, lim_h - > \infty 1^h Y_t$ and we see that the limit does in fact depend on Y_t

2 Question Two: Seasonal Patterns

In this problem, you use the cement dataset (cement and its description cement Description are on Canvas). Note that you have to use tsset first to indicate to stata that your data consists of a time series. Write down the regression equation for each regression you perform. (a) **Seasonal Dummies** Generate new variable seasonal that is equal to 1 for t corresponding to January and 0 otherwise. Compute regression of gcem on L(0/10)seasonal. Which months has the largest and the smallest average values for gcem? Report the values.

Regression model: $gcem = \beta_0 + \beta_1 * I(month) + \epsilon_i$
Lagged regression: $\hat{\epsilon_t} = \beta_0 + \beta_1 * (\epsilon_{t-1}) + \hat{\beta_2} * (grres)$

```
Linear regression                               Number of obs     =        299
                                                F(12, 287)        =     210.04
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8572
                                                Root MSE          =      .0726
```

gcem	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
seasonal						
--.	-.2466031	.0289752	-8.51	0.000	-.303634	-.1895722
L1.	.1491136	.02525	5.91	0.000	.0994149	.1988124
L2.	.2396536	.0195931	12.23	0.000	.2010892	.278218
L3.	.2135449	.0115148	18.55	0.000	.1908808	.2362091
L4.	.0913427	.0089842	10.17	0.000	.0736594	.109026
L5.	.1108637	.0054218	20.45	0.000	.1001922	.1215353
L6.	-.0450786	.0077029	-5.85	0.000	-.0602401	-.0299172
L7.	.0664325	.005373	12.36	0.000	.055857	.0770079
L8.	-.0444787	.0065029	-6.84	0.000	-.0572782	-.0316792
L9.	.0159906	.0062976	2.54	0.012	.0035952	.028386
L10.	-.2346745	.011197	-20.96	0.000	-.2567131	-.2126358
L11.	-.3081064	.011494	-26.81	0.000	-.3307297	-.2854831

From this regression output, we see that the month with the largest average value for gcem is March, and the month with the smallest average value for gcem is December.

(b) **Breusch-Godfrey Test of Serial Correlation.** Perform the test for serial correlation in the error term for the regression of gcem on grres for serial correlation of order 1, 3 and 5. Compute the nR2 test using the procedure outlined below, with the statistic from the ready-made command for the Breusch-Godfrey test, which is estat bgodfrey lags(p), following regress gcem grres, where p is the order of serial correlation. Do both methods deliver the same test? What is the asymptotic distribution of the Breusch- Godfrey statistic for serial correlation of order 1, 3, and 5? The following steps outline how to perform the Breusch-Godfrey test for first-order serial correlation in the error term:

i. Perform the OLS regression gcem on grres.

ii. Obtain residuals from that regression.

iii. Generate the lagged residual.

iv. Perform the auxiliary regression of the residual on its own lag and the regressor grres.

v. Compute the Breusch-Godfrey statistic using nR2 from the above regression.

Regression model for gcem on grres: $gcem = \beta_0 + \beta_1(grres) + \epsilon_i$

Regression output:

Source	SS	df	MS			
				Number of obs	=	309
				F(1, 307)	=	2.03
Model	.071508688	1	.071508688	Prob > F	=	0.1548
Residual	10.7902283	307	.035147323	R-squared	=	0.0066
				Adj R-squared	=	0.0033
Total	10.8617369	308	.03526538	Root MSE	=	.18748

gcem	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
grres	.5119081	.3588878	1.43	0.155	-.1942831	1.218099
_cons	.0026237	.0106727	0.25	0.806	-.0183773	.0236246

First set of figures, Breusch-Godfrey test using the Stata command:

lags(*p*)	chi2	df	Prob > chi2
1	69.237	1	0.0000

H0: no serial correlation

```
. estat bgodfrey, lags(3)
```

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
3	103.322	3	0.0000

H0: no serial correlation

```
. estat bgodfrey, lags(5)
```

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
5	117.311	5	0.0000

Reported figures for the manual method of conducting the Breusch-Godfrey Test: (See appendix for regression output given by Stata)

Serial correlation of order one. $nR^2 = 0.224(308) = 68.99$

Serial correlation of order three. $nR^2 = 0.333(306) = 101.898$

Serial correlation of order five. $nR^2 = 0.378(304) = 114.912$

The asymptotic distribution of the Breusch-Godfrey statistic for serial correlation of order 1,3, and 5 is a chi-squared distribution; the degrees of freedom are given by the number of lags.

(c) **Correcting for Serial Correlation.** In order to correct for serial correlation, you can use Newey-West standard errors, using the newey command in stata. Choose the number of lags that is appropriate according to your results in (b). Report the Newey-West standard errors and compare them to the standard errors you get from the regress command. What is the difference between the Newey-West standard errors and the robust standard errors you obtain from regress. Based on the results in part (b), the appropriate number of lags is five. The reason for this is that five lags yields the highest value for the chi-squared statistic. This means that five lags is the most statistically robust. The Newey-West standard errors are reported in the following figure:

Regression with Newey-West standard errors
Maximum lag = 5

Number of obs	=	309
F(1, 307)	=	1.57
Prob > F	=	0.2110

gcem	Coefficient	Newey-West std. err.	t	P>|t|	[95% conf.	interval]
grres	.5119081	.4083955	1.25	0.211	-.2917004	1.315517
_cons	.0026237	.0122009	0.22	0.830	-.0213843	.0266317

In the above model specification, 5 lags were chosen. This comes from the formula: $p = 0.75 * T^(1/3)$

Both the non-robust, as well as the heteroscedasticity-robust standard errors are reported in the following figure:

Linear regression

Number of obs	=	309
F(1, 307)	=	1.85
Prob > F	=	0.1745
R-squared	=	0.0066
Root MSE	=	.18748

gcem	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
grres	.5119081	.3761517	1.36	0.175	-.2282536	1.25207
_cons	.0026237	.010706	0.25	0.807	-.0184428	.0236901

. regress gcem grres // regular regression

Source	SS	df	MS
Model	.071508688	1	.071508688
Residual	10.7902283	307	.035147323
Total	10.8617369	308	.03526538

Number of obs	=	309
F(1, 307)	=	2.03
Prob > F	=	0.1548
R-squared	=	0.0066
Adj R-squared	=	0.0033
Root MSE	=	.18748

gcem	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
grres	.5119081	.3588878	1.43	0.155	-.1942831	1.218099
_cons	.0026237	.0106727	0.25	0.806	-.0183773	.0236246

Comparing the standard errors from the regress command to the Newey-West standard errors, and those obtained from the robust command, we see that the non-robust regression yields the smallest standard errors, the robust regression yields slightly higher standard errors, and the Newey-West standard errors are the largest.

(d) **Testing for a structural break** Formulate a statistical hypothesis to test for significant difference in mean between January and February in notation from (a). Test this statistical hypothesis using t-test.

Number of obs = 51   R-squared = 0.6907
Root MSE = .134572   Adj R-squared = 0.6844

Source	Partial SS	df	MS	F	Prob>F
Model	1.9819689	1	1.9819689	109.44	0.0000
month	1.9819689	1	1.9819689	109.44	0.0000
Residual	.88737572	49	.01810971		
Total	2.8693446	50	.05738689		

The hypothesis we are testing is that there is no significant difference in the mean between January and February. The statistical decision is as follows, H_0 : There not is a difference in means between these two months p for $\beta_1 < 0.05$ (the coefficient on month is statistically significant)

Based on the result from this table $P < 0.05$, so we reject the null hypothesis in favor of the alternative, meaning there is a statistically significant difference in means between these months.

3 Question Three: Time Trend

Demonstrate that the OLS estimator of δ in the model $Y_t = \beta_1 + \delta(t) + \epsilon_t t = 1, ..., T$ is superconsistent (here ϵ_t is an i.i.d zero mean rv). Show also that it is unbiased in finite samples, despite the fact that Y_t is nonstationary.

Equation: $Y_t = \beta_1 + \delta_t + \epsilon_t t = 1, ..., T$

Showing that the OLS estimator of $\hat{\delta}$ is unbiased:

$$Y_t = \beta_1 + \delta_t + \epsilon_t$$

$$\overline{y} = \beta_1 + \delta_t(\overline{t}) + (1/T)\Sigma_{t=1}^{T}(t)$$
$$\overline{t} = (1/T)\Sigma_{t=1}^{T}(t)$$
$$Y_t - \overline{Y} = (\beta_1 + \delta(t) + \epsilon_t) - (\beta_1 + \delta(\overline{t} + (1/T)\Sigma_{t=1}^{T}(t))$$
$$= \delta(t - \overline{t}) + \epsilon_t - (1/T)\Sigma_{t=1}^{T}(t)$$

Plugging into the OLS formula for $del\hat{t}a$

$$\hat{\delta} = \Sigma_{t=1}^{T}((Y_t - \overline{Y})(t - \overline{t}))/\Sigma_{t=1}^{T}(t - \overline{t})^2$$

$$\hat{\delta} = \Sigma_{t=1}^{T}(\delta(t - \overline{t}) + \epsilon_t)(t - \overline{t})/\Sigma_{t=1}^{T}(t - \overline{t})^2$$

$$\hat{\delta} = \delta\Sigma_{t=1}^{T}(t - \overline{t})^2/\Sigma_{t=1}^{T}(t - \overline{t})^2 + \Sigma_{t=1}^{T}\epsilon_t(t - \overline{t})/\Sigma_{t=1}^{T}(t - \overline{t})^2$$

This final expression shows that $\hat{\delta} = \delta$ because the second term, $\Sigma_{t=1}^{T}\epsilon_t(t - \overline{t})/\Sigma_{t=1}^{T}(t - \overline{t})^2$ is equal to zero by the property $E(\epsilon_t) = 0$

Now to show that the OLS estimator of $\hat{\delta}$ is superconsistent. Reminder, the criteria for this is $\sqrt{t}(\hat{\delta}-\delta)->0$

$$\hat{\delta}-\delta = \Sigma_{t=1}^{T}\epsilon_t(t-\bar{t})/\Sigma_{t=1}^{T}(t-\bar{t})^2$$
$$= 1/\Sigma_{t=1}^{T}(t-\bar{t})^2 + (\epsilon_1(1-\bar{t}+...+\epsilon_T(T-\bar{t})$$
$$var(\hat{\delta}-\delta) = (1/\Sigma_{t=1}^{T}(t-\bar{t})^2)^2((1-\bar{t})var(\epsilon_1)+...+(T-\bar{t}var(\epsilon_T)))$$

This simplifies to: $1/\Sigma_{t=1}^{T}(t-\bar{t})^2$ since ϵ_t is i.i.d
At this point, simplification is algebraic:

$$1/\Sigma_{t=1}^{T}(t^2-2t\bar{t}+\bar{t}^2)$$

$$1/\Sigma_{t=1}^{T}(t^2) = T(T+1)(T+6)/6$$
$$= 2T^3+2T^2+T/6$$

$$1/\Sigma_{t=1}^{T}(t) = T(T+1)/2$$
$$1/\Sigma_{t=1}^{T}(\bar{t}) = (T+1)/2$$
$$2t\bar{t} = 2(T(T+1)/T)((T+1)/2)$$
$$= 3T^3+6T^2+3T/6$$

$$1/\Sigma_{t=1}^{T}(\bar{t})^2 = T((T+1)/2)^2$$

1 / $\Sigma_{t=1}^{T}(\bar{t})^2 = T^3+2T^2+T/4$

$$1/\Sigma_{t=1}^{T}(t^2-2t\bar{t}+\bar{t}^2) = 2T^3+2T^2+T/6-3T^3+6T^2+3T/6+(T+1)^2/4$$

$$1/\Sigma_{t=1}^{T}(\bar{t})^2 = T^3-T/12$$
$$var(\hat{\delta}-\delta) = 12\sigma_\epsilon^2/(T^3-T)$$
$$var(\sqrt{T}\hat{\delta}-\delta) = 12\sigma_\epsilon^2/(T^3-T)$$

Finally we can see that as t-¿ ∞, the above term -¿ 0
As $t->\infty$ the above expression $->0$

Appendix:

Second set of figures: Breusch-Godfrey test using the outlined process.

Source	SS	df	MS			
				Number of obs	=	308
				F(2, 305)	=	44.11
Model	2.41780816	2	1.20890408	Prob > F	=	0.0000
Residual	8.35981439	305	.027409228	R-squared	=	0.2243
				Adj R-squared	=	0.2192
Total	10.7776225	307	.035106262	Root MSE	=	.16556

ehat	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ehatlag_1	.4736578	.0504315	9.39	0.000	.37442	.5728956
grres	.0907785	.3170783	0.29	0.775	-.5331594	.7147165
_cons	-.0003738	.0094403	-0.04	0.968	-.0189503	.0182026

```
. esttab, se r2
```

	(1) ehat
ehatlag_1	0.474***
	(0.0504)
grres	0.0908
	(0.317)
_cons	-0.000374
	(0.00944)
N	308
R-sq	0.224

Figure 1: Serial correlation of order one. $nR^2 = 0.224(308) = 68.99$

Source	SS	df	MS
Model	2.41780816	2	1.20890408
Residual	8.35981439	305	.027409228
Total	10.7776225	307	.035106262

Number of obs	=	308
F(2, 305)	=	44.11
Prob > F	=	0.0000
R-squared	=	0.2243
Adj R-squared	=	0.2192
Root MSE	=	.16556

ehat	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
grres	.0907785	.3170783	0.29	0.775	-.5331594	.7147165
ehat						
L1.	.4736578	.0504315	9.39	0.000	.37442	.5728956
_cons	-.0003738	.0094403	-0.04	0.968	-.0189503	.0182026

```
. esttab, se r2
```

	(1) ehat
grres	0.0908
	(0.317)
L.ehat	0.474***
	(0.0504)
_cons	-0.000374
	(0.00944)
N	308
R-sq	0.224

Figure 2: Serial correlation of order three. $nR^2 = 0.333(306) = 101.898$

ehat	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
grres	.0595326	.2953768	0.20	0.840	-.521764	.6408292
ehat						
L1.	.4989758	.0560789	8.90	0.000	.3886134	.6093382
L2.	-.0848685	.0626184	-1.36	0.176	-.2081004	.0383634
L3.	-.344268	.0596829	-5.77	0.000	-.461723	-.2268131
L4.	.1490127	.0627468	2.37	0.018	.025528	.2724973
L5.	-.2644178	.0559814	-4.72	0.000	-.3745883	-.1542472
_cons	-.002121	.0085615	-0.25	0.805	-.0189698	.0147278

```
. esttab, se r2
```

	(1) ehat
grres	0.0595
	(0.295)
L.ehat	0.499***
	(0.0561)
L2.ehat	-0.0849
	(0.0626)
L3.ehat	-0.344***
	(0.0597)
L4.ehat	0.149*
	(0.0627)
L5.ehat	-0.264***
	(0.0560)
_cons	-0.00212
	(0.00856)
N	304
R-sq	0.378

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Figure 3: Serial correlation of order five. $nR^2 = 0.378(304) = 114.912$

```
*------------------------------------------------
*Title: Problem Set Three
*VirenePS3.do
*date: 2.20.2024
* Name: Josh Virene*
* ARE256B- Applied Econometrics II
* Purpose: The purpose of this script is to accomplish the tasks that
are outlined in the problem set three assignment
* ------------------------------------------------

*------------------------------------------------
*Program Setup
*------------------------------------------------
clear all
set more off

* set working directory:
global path = "C:\Users\jwvirene\Desktop\Virene PS3"
cd "C:\Users\jwvirene\Desktop\Virene PS3"
pwd // to verify directory was changed

* import the dataset upon which we run the analysis:
use "C:\Users\jwvirene\Desktop\Virene PS3\cement.dta", clear


*------------------------------------------------
*Program Setup
*------------------------------------------------
version 14             // Set Version number for backward
compatibility
set more off           // Disable partitioned output
set linesize 80        // Line size limit to make output more
readable
capture log close      // Close existing log files
log using ps3, replace // Open log file
*------------------------------------------------

* preview dataset (if needed)
browse
* we want to know the variable names:
ds

// commands so that stata knows this is time series data:
gen date = ym(year, month)
tsset date

*------------------------------------------------
* Question Two: Seasonal Patterns
*------------------------------------------------
```

```
// (a): Seasonal dummies
gen seasonal = .
replace seasonal = 1 if month == 1
replace seasonal = 0 if month != 1
// running the regression
reg gcem L(0/11).seasonal, nocons robust
// alternative way to show means
regress gcem i.month, robust

// (b): Breusch-Godfrey Test of Serial Correlation.
// set x as white noise


// perform ols regression gcem on grres
regress gcem grres // regular regression
// regress gcem grres, robust // heteroscedasticity robust regression
// method one, using the commands
estat bgodfrey, lags(1)
estat bgodfrey, lags(3)
estat bgodfrey, lags(5)
// method two, running through the steps:


// obtain residuals
predict ehat, residuals
ac ehat // checking for autocorrelation

// lagged residuals
gen ehatlag_1 = ehat[_n-1]

// Perform the auxiliary regression of the residual on its own lag and
the regressor grres
regress ehat ehatlag_1 grres
esttab, se r2

// Compute the Breusch-Godfrey statistic using nR2 from the above
regression.

// p = 1:
reg ehat grres L(1/1).ehat
esttab, se r2
di e(N)


reg ehat grres L(1/3).ehat
esttab, se r2
di e(N)


reg ehat grres L(1/5).ehat
```

```
esttab, se r2
di e(N)

// From the output, we can see that the results via the manual process
are the same as those calculated by running the Breusch-Godfrey
command estat bgodfrey, lags(n) above. The product n*R^2 gives the
same values

// (c): Correcting for Serial Correlation
// need to use the Newey command to compute the Newey-West standard
errors
newey gcem grres, lag(5) force

// (d): Testing for a structural break.

// subset the data so there are only two months- January and February
keep if month == 1 | month == 2

// re-run the regression
reg gcem month
anova gcem month

// from the ANOVA output, it is clear that we may reject the null
hypothesis that the mean value for gcem in January is the same as that
in February.


*-----------------------------------------------------

log close // Close the log, end the file

global path "C:\Users\jwvirene\Desktop\Virene PS3"

translate "$path\ps3.smcl" ///
          "$path\ps3.pdf", translator(smcl2pdf)
```

Statistics/Data analysis

```
      name:  <unnamed>
       log:  C:\Users\jwvirene\Desktop\Virene PS3\ps3.smcl
  log type:  smcl
 opened on:  22 Feb 2024, 12:51:42

1 . *---------------------------------------------------
2 .
3 . * preview dataset (if needed)
4 . browse

5 . * we want to know the variable names:
6 . ds
  year      prcpet    rdefs     grprcpet  grdefs    mar       jul       nov
  month     rresc     milemp    grres     gmilemp   apr       aug       dec
  prccem    rnonc     gprc      grnon     jan       may       sep
  ipcem     ip        gcem      gip       feb       jun       oct

7 .
8 . // commands so that stata knows this is time series data:
9 . gen date = ym(year, month)

10. tsset date

  Time variable: date, 48 to 359
          Delta: 1 unit

11.
12. *---------------------------------------------------
13. * Question Two: Seasonal Patterns
14. *---------------------------------------------------
15.
16. // (a): Seasonal dummies
17. gen seasonal = .
  (312 missing values generated)

18. replace seasonal = 1 if month == 1
  (26 real changes made)

19. replace seasonal = 0 if month != 1
  (286 real changes made)

20. // running the regression
21. reg gcem L(0/11).seasonal, nocons robust

  Linear regression                               Number of obs     =        299
                                                  F(12, 287)        =     210.04
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.8572
                                                  Root MSE          =      .0726

  ------------------------------------------------------------------------------
               |               Robust
          gcem | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
  -------------+----------------------------------------------------------------
      seasonal |
           --. |  -.2466031   .0289752    -8.51   0.000     -.303634   -.1895722
           L1. |   .1491136     .02525     5.91   0.000     .0994149    .1988124
           L2. |   .2396536   .0195931    12.23   0.000     .2010892     .278218
           L3. |   .2135449   .0115148    18.55   0.000     .1908808    .2362091
           L4. |   .0913427   .0089842    10.17   0.000     .0736594     .109026
           L5. |   .1108637   .0054218    20.45   0.000     .1001922    .1215353
           L6. |  -.0450786   .0077029    -5.85   0.000    -.0602401   -.0299172
           L7. |   .0664325    .005373    12.36   0.000      .055857    .0770079
           L8. |  -.0444787   .0065029    -6.84   0.000    -.0572782   -.0316792
           L9. |   .0159906   .0062976     2.54   0.012     .0035952     .028386
          L10. |  -.2346745    .011197   -20.96   0.000    -.2567131   -.2126358
          L11. |  -.3081064    .011494   -26.81   0.000    -.3307297   -.2854831
  ------------------------------------------------------------------------------
```

```
22. // alternative way to show means
23. regress gcem i.month, robust

  Linear regression                               Number of obs     =        309
                                                  F(11, 297)        =     224.02
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.8576
                                                  Root MSE          =     .07217

  ------------------------------------------------------------------------------
               |               Robust
          gcem | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
  -------------+----------------------------------------------------------------
         month |
            2  |   .3943454   .0378018    10.43   0.000     .3199521    .4687388
            3  |   .4843566   .0345903    14.00   0.000     .4162834    .5524297
            4  |   .4615874    .031031    14.88   0.000     .4005189    .5226559
            5  |   .3428241   .0306066    11.20   0.000     .2825909    .4030573
            6  |   .3557875   .0294685    12.07   0.000     .2977939     .413781
            7  |   .2035547   .0299556     6.80   0.000     .1446026    .2625068
            8  |   .3117213   .0294417    10.59   0.000     .2537806     .369662
            9  |   .2015412   .0296279     6.80   0.000     .1432339    .2598485
           10  |   .2639279   .0296113     8.91   0.000     .2056534    .3222025
           11  |   .0086097   .0310615     0.28   0.782    -.0525189    .0697383
           12  |  -.0615033   .0311507    -1.97   0.049    -.1228073   -.0001993
               |
         _cons |  -.2466031   .0289556    -8.52   0.000    -.3035873   -.1896188
  ------------------------------------------------------------------------------

24.
25. // (b): Breusch-Godfrey Test of Serial Correlation.
26. // set x as white noise
27.
28.
29. // perform ols regression gcem on grres
30. regress gcem grres // regular regression

        Source |       SS           df       MS      Number of obs   =       309
  -------------+----------------------------------   F(1, 307)       =      2.03
         Model |  .071508688         1  .071508688   Prob > F        =    0.1548
      Residual |  10.7902283       307  .035147323   R-squared       =    0.0066
  -------------+----------------------------------   Adj R-squared   =    0.0033
         Total |  10.8617369       308  .03526538    Root MSE        =    .18748

  ------------------------------------------------------------------------------
          gcem | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
  -------------+----------------------------------------------------------------
         grres |   .5119081   .3588878     1.43   0.155    -.1942831    1.218099
         _cons |   .0026237   .0106727     0.25   0.806    -.0183773    .0236246
  ------------------------------------------------------------------------------

31. // regress gcem grres, robust // heteroscedasticity robust regression
32. // method one, using the commands
33. estat bgodfrey, lags(1)

  Breusch-Godfrey LM test for autocorrelation
  ---------------------------------------------------------------------------
      lags(p)  |          chi2               df                 Prob > chi2
  -------------+-------------------------------------------------------------
         1     |         69.237               1                   0.0000
  ---------------------------------------------------------------------------
                          H0: no serial correlation
```

```
34. estat bgodfrey, lags(3)

  Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
3	103.322	3	0.0000

```
                    H0: no serial correlation

35. estat bgodfrey, lags(5)

  Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
5	117.311	5	0.0000

```
                    H0: no serial correlation

36. // method two, running through the steps:
37.
38.
39. // obtain residuals
40. predict ehat, residuals
  (3 missing values generated)

41. ac ehat // checking for autocorrelation

42.
43. // lagged residuals
44. gen ehatlag_1 = ehat[_n-1]
  (3 missing values generated)

45.
46. // Perform the auxiliary regression of the residual on its own lag and the regressor
  >  grres
47. regress ehat ehatlag_1 grres
```

Source	SS	df	MS
Model	2.41780816	2	1.20890408
Residual	8.35981439	305	.027409228
Total	10.7776225	307	.035106262

Number of obs	=	308
F(2, 305)	=	44.11
Prob > F	=	0.0000
R-squared	=	0.2243
Adj R-squared	=	0.2192
Root MSE	=	.16556

ehat	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ehatlag_1	.4736578	.0504315	9.39	0.000	.37442	.5728956
grres	.0907785	.3170783	0.29	0.775	-.5331594	.7147165
_cons	-.0003738	.0094403	-0.04	0.968	-.0189503	.0182026

```
48. esttab, se r2
```

	(1) ehat
ehatlag_1	0.474***
	(0.0504)
grres	0.0908
	(0.317)

```
  _cons                 -0.000374
                        (0.00944)
  ----------------------------------
  N                           308
  R-sq                      0.224
  ----------------------------------
  Standard errors in parentheses
  * p<0.05, ** p<0.01, *** p<0.001

49.
50. // Compute the Breusch-Godfrey statistic using nR2 from the above regression.
51.
52. // p = 1:
53. reg ehat grres L(1/1).ehat

      Source |       SS           df       MS      Number of obs   =       308
-------------+----------------------------------   F(2, 305)       =     44.11
       Model |  2.41780816         2  1.20890408   Prob > F        =    0.0000
    Residual |  8.35981439       305  .027409228   R-squared       =    0.2243
-------------+----------------------------------   Adj R-squared   =    0.2192
       Total |  10.7776225       307  .035106262   Root MSE        =    .16556

------------------------------------------------------------------------------
        ehat | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       grres |   .0907785   .3170783     0.29   0.775    -.5331594    .7147165
             |
        ehat |
         L1. |   .4736578   .0504315     9.39   0.000       .37442    .5728956
             |
       _cons |  -.0003738   .0094403    -0.04   0.968    -.0189503    .0182026
------------------------------------------------------------------------------

54. esttab, se r2

  ----------------------------------
                          (1)
                         ehat
  ----------------------------------
  grres                0.0908
                      (0.317)

  L.ehat                0.474***
                     (0.0504)

  _cons             -0.000374
                    (0.00944)
  ----------------------------------
  N                       308
  R-sq                  0.224
  ----------------------------------
  Standard errors in parentheses
  * p<0.05, ** p<0.01, *** p<0.001

55. di e(N)
  308

56.
57.
58. reg ehat grres L(1/3).ehat

      Source |       SS           df       MS      Number of obs   =       306
-------------+----------------------------------   F(4, 301)       =     37.59
       Model |  3.55654624         4  .88913656    Prob > F        =    0.0000
    Residual |  7.11878784       301  .023650458   R-squared       =    0.3332
-------------+----------------------------------   Adj R-squared   =    0.3243
       Total |  10.6753341       305  .035001095   Root MSE        =    .15379
```

ehat	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
grres	-.0493155	.3017448	-0.16	0.870	-.6431119	.544481
ehat						
L1.	.4948265	.054205	9.13	0.000	.3881578	.6014951
L2.	.002565	.0615926	0.04	0.967	-.1186415	.1237716
L3.	-.3350572	.0542174	-6.18	0.000	-.4417503	-.2283641
_cons	-.0013079	.0087979	-0.15	0.882	-.0186212	.0160054

59. esttab, se r2

	(1) ehat
grres	-0.0493
	(0.302)
L.ehat	0.495***
	(0.0542)
L2.ehat	0.00257
	(0.0616)
L3.ehat	-0.335***
	(0.0542)
_cons	-0.00131
	(0.00880)
N	306
R-sq	0.333

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

60. di e(N)
306

61.
62.
63. reg ehat grres L(1/5).ehat

Source	SS	df	MS
Model	4.00940058	6	.668233429
Residual	6.60210857	297	.022229322
Total	10.6115091	303	.035021482

Number of obs	=	304
F(6, 297)	=	30.06
Prob > F	=	0.0000
R-squared	=	0.3778
Adj R-squared	=	0.3653
Root MSE	=	.1491

ehat	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
grres	.0595326	.2953768	0.20	0.840	-.521764	.6408292
ehat						
L1.	.4989758	.0560789	8.90	0.000	.3886134	.6093382
L2.	-.0848685	.0626184	-1.36	0.176	-.2081004	.0383634
L3.	-.344268	.0596829	-5.77	0.000	-.461723	-.2268131
L4.	.1490127	.0627468	2.37	0.018	.025528	.2724973
L5.	-.2644178	.0559814	-4.72	0.000	-.3745883	-.1542472
_cons	-.002121	.0085615	-0.25	0.805	-.0189698	.0147278

64. esttab, se r2

	(1) ehat
grres	0.0595
	(0.295)
L.ehat	0.499***
	(0.0561)
L2.ehat	-0.0849
	(0.0626)
L3.ehat	-0.344***
	(0.0597)
L4.ehat	0.149*
	(0.0627)
L5.ehat	-0.264***
	(0.0560)
_cons	-0.00212
	(0.00856)
N	304
R-sq	0.378

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

65. di e(N)
304

66.
67. // From the output, we can see that the results via the manual process are the same
> as those calculated by running the Breusch-Godfrey command estat bgodfrey, lags(n) a
> bove. The product n*R^2 gives the same values
68.
69. // (c): Correcting for Serial Correlation
70. // need to use the Newey command to compute the Newey-West standard errors
71. newey gcem grres, lag(5) force

Regression with Newey-West standard errors Number of obs = 309
Maximum lag = 5 F(1, 307) = 1.57
Prob > F = 0.2110

gcem	Coefficient	Newey-West std. err.	t	P>|t|	[95% conf. interval]	
grres	.5119081	.4083955	1.25	0.211	-.2917004	1.315517
_cons	.0026237	.0122009	0.22	0.830	-.0213843	.0266317

72.
73. // (d): Testing for a structural break.
74.

```
75. // subset the data so there are only two months- January and February
76. keep if month == 1 | month == 2
  (260 observations deleted)

77. 
78. // re-run the regression
79. reg gcem month

      Source |       SS           df       MS      Number of obs   =        51
-------------+----------------------------------   F(1, 49)        =    109.44
       Model |  1.98196887         1  1.98196887   Prob > F        =    0.0000
    Residual |  .887375717        49  .018109709   R-squared       =    0.6907
-------------+----------------------------------   Adj R-squared   =    0.6844
       Total |  2.86934459        50  .057386892   Root MSE        =    .13457

------------------------------------------------------------------------------
        gcem | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       month |   .3943454    .037695    10.46   0.000     .3185944    .4700964
       _cons |  -.6409485   .0599507   -10.69   0.000    -.7614239   -.5204732
------------------------------------------------------------------------------

80. anova gcem month

                         Number of obs =         51    R-squared     =  0.6907
                         Root MSE      =    .134572    Adj R-squared =  0.6844

                  Source | Partial SS         df         MS        F    Prob>F
              -----------+----------------------------------------------------
                   Model |  1.9819689          1   1.9819689    109.44  0.0000
                         |
                   month |  1.9819689          1   1.9819689    109.44  0.0000
                         |
                Residual |  .88737572         49   .01810971  
              -----------+----------------------------------------------------
                   Total |  2.8693446         50   .05738689  

81. 
82. // from the ANOVA output, it is clear that we may reject the null hypothesis that th
> e mean value for gcem in January is the same as that in February.
83. 
84. 
85. *---------------------------------------------------
86. 
87. log close // Close the log, end the file
      name:  <unnamed>
       log:  C:\Users\jwvirene\Desktop\Virene PS3\ps3.smcl
  log type:  smcl
 closed on:  22 Feb 2024, 12:51:49
-------------------------------------------------------------------------------------
```