Multi-Step Forecast Variance

- Can use plug-in, iterated, or direct method
- Easiest method is direct
- Forecast variance can be computed from direct regression

$$y_t = \hat{\alpha}^* + \hat{\beta}^* y_{t-h} + \hat{u}_t$$

$$\hat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2$$

Forecast Variance and Intervals

- Standard deviation of forecast
  - **predict sf, stdf**
- Forecast Intervals (90% normal)

  gen y1L=y1-1.645*sf

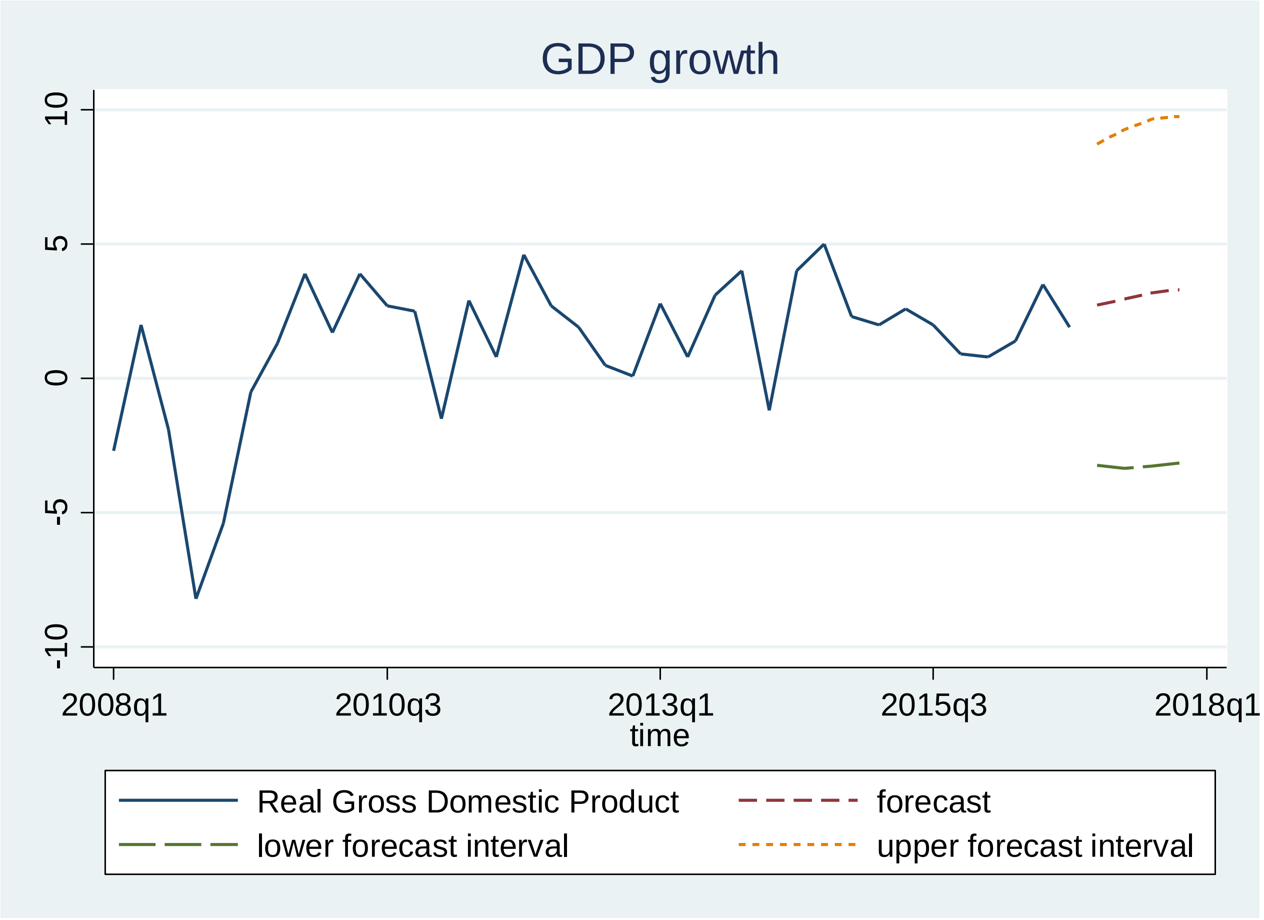
  gen y1U=y1+1.645*sf

4-step Direct Interval Forecasts

```
tsappend, add(4)
reg gdp L.gdp
predict y1
predict sf1,stdf
gen y1L=y1-1.645*sf1
gen y1U=y1+1.645*sf1
reg gdp L2.gdp
predict y2
predict sf2,stdf
gen y2L=y2-1.645*sf2
gen y2U=y2+1.645*sf2
reg gdp L3.gdp
predict y3
predict sf3,stdf
gen y3L=y3-1.645*sf3
gen y3U=y3+1.645*sf3
```

```
reg gdp L4.gdp
predict y4
predict sf4,stdf
gen y4L=y4-1.645*sf4
gen y4U=y4+1.645*sf4
egen p=rowfirst(y1 y2 y3 y4) if t>=tq(2017q1)
egen pL=rowfirst(y1L y2L y3L y4L) if t>=tq(2017q1)
egen pU=rowfirst(y1U y2U y3U y4U) if t>=tq(2017q1)
label variable p "forecast"
label variable pL "lower forecast interval"
label variable pU "upper forecast interval"
tsline gdp p pL pU if t>=tq(2008q1), title(GDP growth)
        lpattern (solid dash longdash shortdash)
```

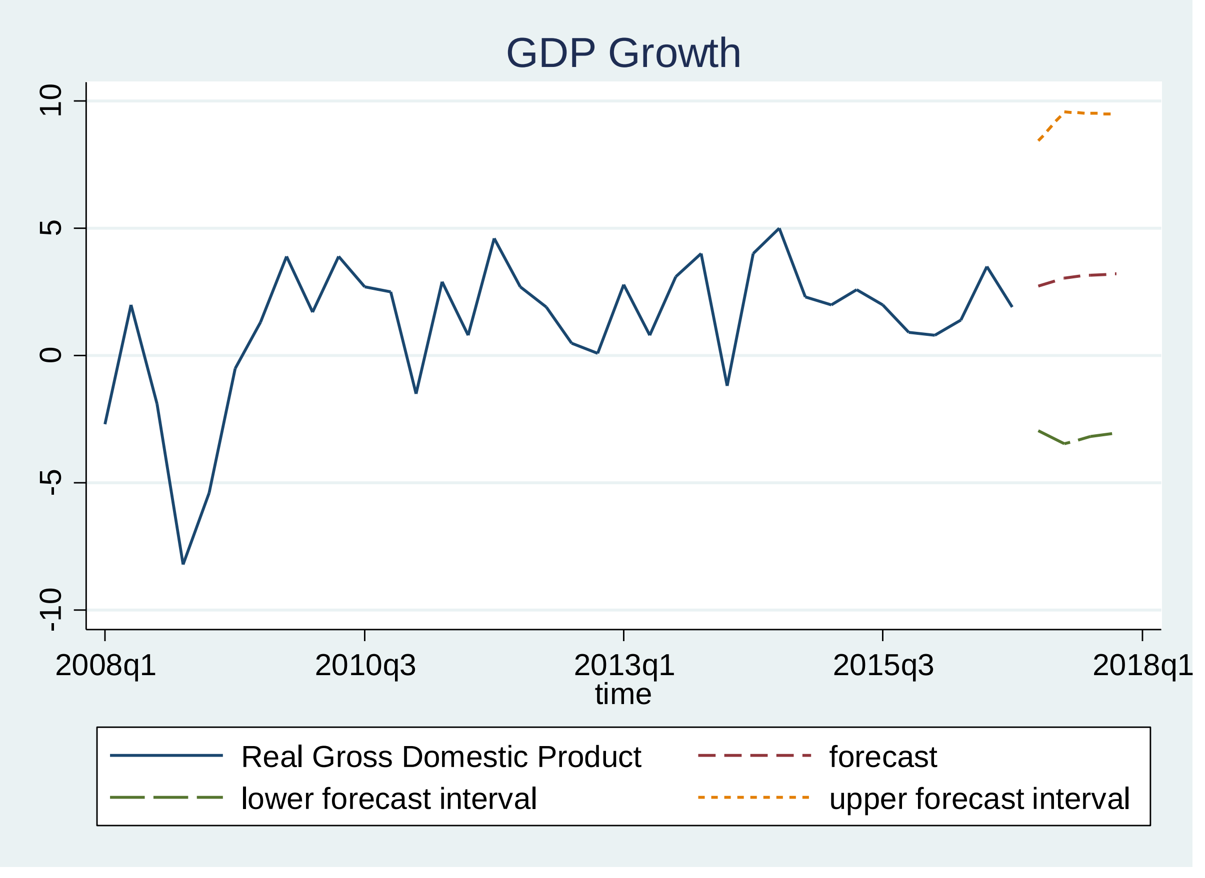
GDP Growth Forecast (Direct)



4-Step Iterated Interval Forecast

```
use realgdpgrowth.dta
tsappend, add(4)
reg gdp L.gdp
forecast create ar1
estimate store model1
forecast estimates model1
forecast solve, simulate(errors,statistic(stddev,prefix(sd_)) reps(1000) )
gen pf = f_gdp if t>=tq(2017q1)
gen pfL = pf - 1.645*sd_gdp
gen pfU = pf + 1.645*sd_gdp
label variable pf “forecast”
label variable pfL "lower forecast interval”
label variable pfU "upper forecast interval“
tsline gdp pf pfL pfU if t>=tq(2008q1), title(GDP growth) lpattern (solid dash
    longdash shortdash)
```

GDP Growth Forecast (Iterated)



AR(2) Process

- An autoregressive process of order 2, or AR(2) is

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + e_t$$

  where e_t is WN(0, σ^2)
- Using the lag operator

$$\left(1 - \beta_1 L - \beta_2 L^2\right) y_t = e_t$$

AR(2) Process with Intercept

- An autoregressive process of order 2, or AR(2) is

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + e_t$$

or

$$\left(1 - \beta_1 L - \beta_2 L^2\right) y_t = \alpha + e_t$$

Multiplier-Accelerator Model

- Due to Paul Samuelson
- Output *Y*, Consumption *C* and Investment *I*
- Aggregate Income $Y_t = C_t + I_t$
- Consumption Multiplier $C_t = a_0 + a_1 Y_{t-1}$
- Investment Accelerator $I_t = b\left(C_t - C_{t-1}\right) + e_t$
- Combine to find process for output

$$Y_t = a_0 + a_1\left(1 + b\right)Y_{t-1} - a_1 b Y_{t-2} + e_t$$

an AR(2) Process

Multiplier-Accelerator

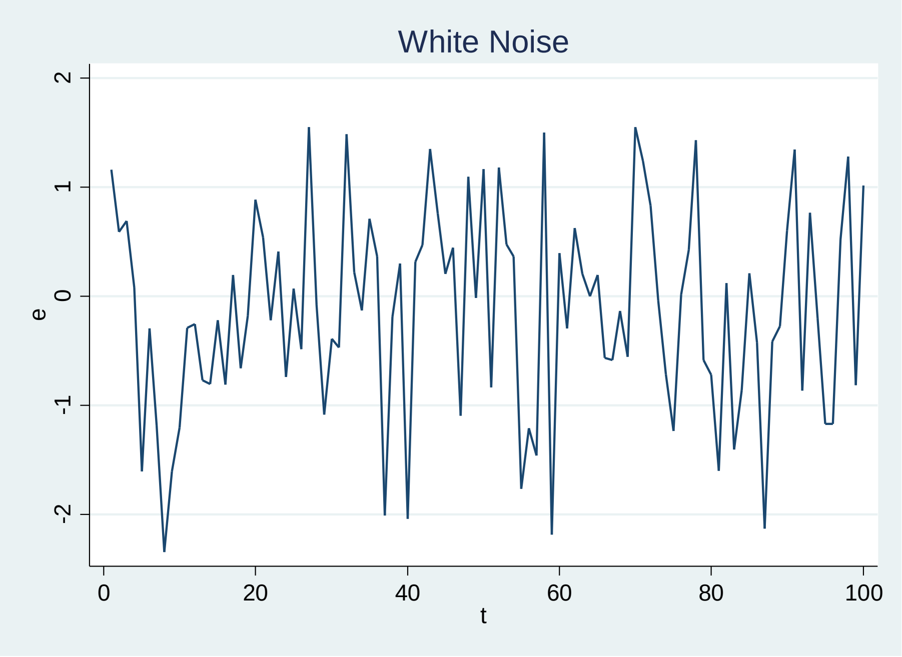
- Example

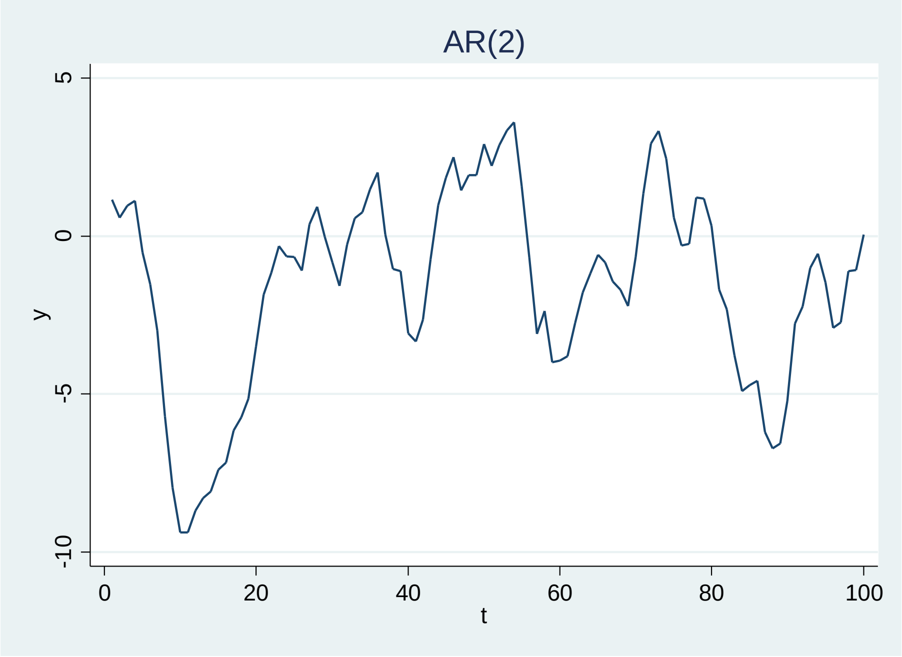
$$C_t = 0.9Y_{t-1}$$

$$I_t = .5\left(C_t - C_{t-1}\right) + e_t$$

$$Y_t = 1.35Y_{t-1} - 0.45Y_{t-2} + e_t$$

Simulated Example





Simulating AR processes in STATA

```
. set obs 100
obs was 0, now 100

. gen t=_n

. tsset t
        time variable:  t, 1 to 100
                delta:  1 unit

. gen e=rnormal()

. gen y=e

. replace y=1.35*L.y-.45*L2.y+e if t>2
(98 real changes made)
```

Random Numbers in Stata

- “Random numbers” not really random, they are pseudo-random high-dimensional chaotic sequences
- They start from a seed. The default in Stata is “123456789”, when you start the program.
- You can set the seed to any 9-digit integer
  - `set seed 2232017` (today’s date)
- numbergenerator.org

Stationarity of AR(2)

- The AR(2) process is stationary if we can invert the lag polynomial to write it as a general linear process, if

$$\left(1 - \beta_1 L - \beta_2 L^2\right) y_t = e_t$$

$$y_t = \left(1 - \beta_1 L - \beta_2 L^2\right)^{-1} e_t$$

- When is this valid?

Factors

- Write the polynomial $(1-\beta_1 L-\beta_2 L^2)$ as factors

$$1-\beta_1 L-\beta_2 L^2=(1-\lambda_1 L)(1-\lambda_2 L)$$

- Then

$$y_t=\left(1-\beta_1 L-\beta_2 L^2\right)^{-1}e_t$$
$$=\left(1-\lambda_1 L\right)^{-1}\left(1-\lambda_2 L\right)^{-1}e_t$$

- This is valid if both $(1-\lambda_1 L)$ and $(1-\lambda_2 L)$ are invertible
- This is when $|\lambda_1|<1$ and $|\lambda_2|<1$

Stationary Roots

- λ_1 and λ_2 are the inverses of the roots of the polynomial $(1-\beta_1 L-\beta_2 L^2)$
- They can be real or complex
- If $|\lambda_1|<1$ and $|\lambda_2|<1$ we say they “are within the unit circle”
- The AR(2) is stationary if the inverse roots are within the unit circle (are less than one in absolute value)

Necessary Condition

- The polynomial $(1-\beta_1 L-\beta_2 L^2)$ has a unit root if it equals 0 when L=1
- In this case, $1-\beta_1-\beta_2=0$
- Or $\beta_1+\beta_2=1$
- This means that the AR(2) is nonstationary if the sum of the AR coefficients equals 1

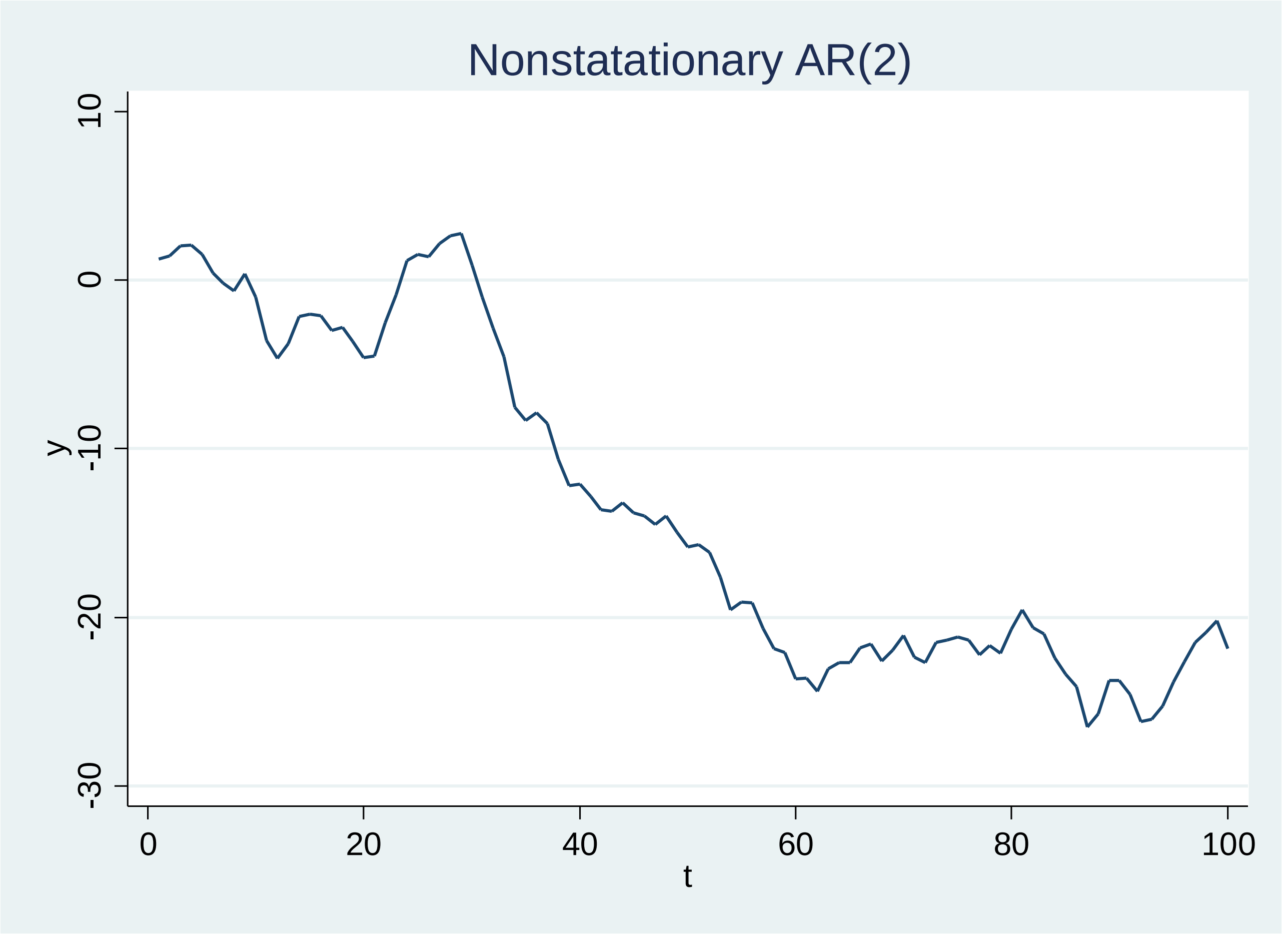
Multiplier Example

- In the model

$$Y_t = a_0 + a_1(1+b)Y_{t-1} - a_1 b Y_{t-2} + e_t$$

  the sum of the coefficients is a_1, the consumption coefficient.
- In this model, if a_1=1, then the output process has a unit root, it is nonstationary

Simulated example with a_1=1



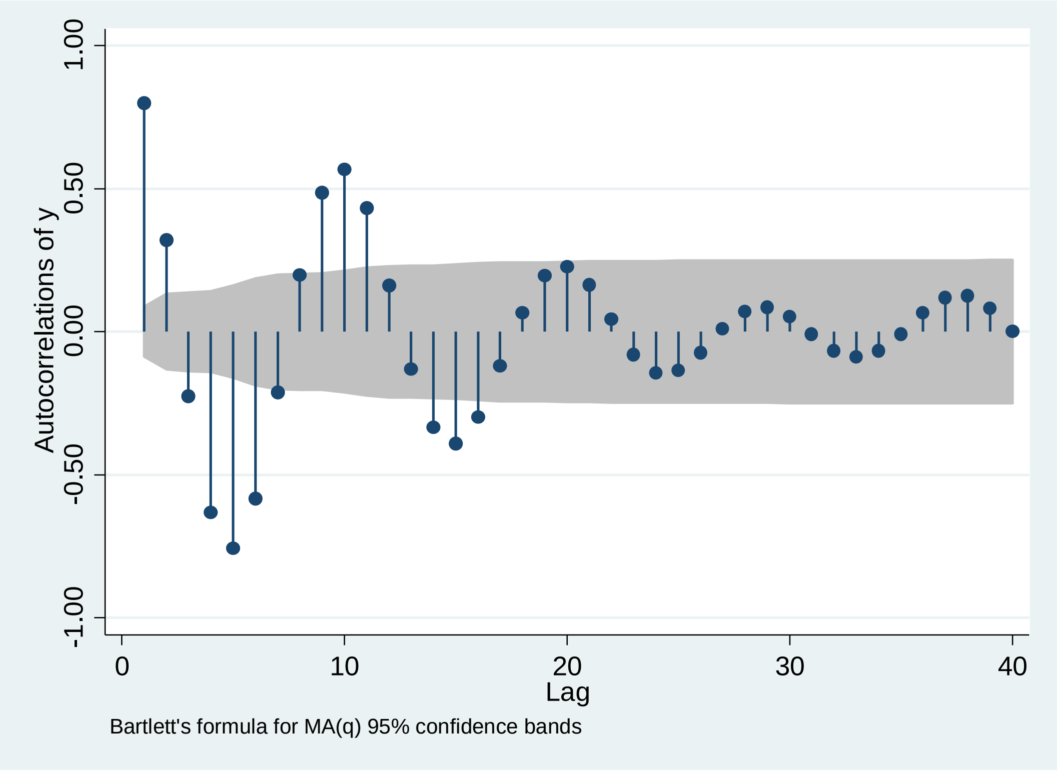
Autocorrelation of AR(2)

- The autocorrelations of an AR(2) can be much more complicated than that of an AR(1)
- Take the example

$$Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + e_t$$

- Simulated a time series with T=500 observations from this example
- Calculated sample autocorrelation function

Autocorrelation Function of AR(2)



Alternative expression

$$\begin{aligned} y_t &= \beta_1 y_{t-1} + \beta_2 y_{t-2} + e_t \\ &= (\beta_1 + \beta_2) y_{t-1} - \beta_2 (y_{t-1} - y_{t-2}) + e_t \\ &= (\beta_1 + \beta_2) y_{t-1} - \beta_2 \Delta y_{t-1} + e_t \end{aligned}$$

- The AR(2) can be written as a function of the lagged **value** and the lagged **change**
- These are equivalent expressions

Estimation of AR(2)

- Least Squares Regression

$$y_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2} + \hat{e}_t$$

Example: Unemployment Rate

```
. reg ur L.ur L2.ur

      Source |       SS           df       MS      Number of obs   =       827
-------------+----------------------------------   F(2, 824)       =   9407.23
       Model |  5339.48297         2  2669.74148   Prob > F        =    0.0000
    Residual |  233.848593       824  .283796836   R-squared       =    0.9580
-------------+----------------------------------   Adj R-squared   =    0.9579
       Total |  5573.33156       826  6.74737477   Root MSE        =    .53273

------------------------------------------------------------------------------
          ur |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ur |
         L1. |   .7638794   .0339869    22.48   0.000     .6971683    .8305905
         L2. |   .2177432   .0339522     6.41   0.000     .1511003     .284386
             |
       _cons |    .175516   .0697542     2.52   0.012     .0385992    .3124329
------------------------------------------------------------------------------
```

Example: GDP Growth

```
. reg gdp L.gdp L2.gdp
```

Source	SS	df	MS
Model	610.816291	2	305.408146
Residual	3605.07548	274	13.1572098
Total	4215.89177	276	15.2749702

Number of obs	=	277
F(2, 274)	=	23.21
Prob > F	=	0.0000
R-squared	=	0.1449
Adj R-squared	=	0.1386
Root MSE	=	3.6273

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp						
L1.	.3337403	.0601059	5.55	0.000	.2154122	.4520684
L2.	.0959558	.0600128	1.60	0.111	-.022189	.2141006
_cons	1.85748	.3079687	6.03	0.000	1.251195	2.463766

Alternative Command

```
. reg gdp L(1/2).gdp
```

Source	SS	df	MS
Model	610.816291	2	305.408146
Residual	3605.07548	274	13.1572098
Total	4215.89177	276	15.2749702

Number of obs	=	277
F(2, 274)	=	23.21
Prob > F	=	0.0000
R-squared	=	0.1449
Adj R-squared	=	0.1386
Root MSE	=	3.6273

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp						
L1.	.3337403	.0601059	5.55	0.000	.2154122	.4520684
L2.	.0959558	.0600128	1.60	0.111	-.022189	.2141006
_cons	1.85748	.3079687	6.03	0.000	1.251195	2.463766

- "L(1/2).gdp" means
  - "regress on lags 1 through 2 of gdp"

One-Step-Ahead Forecast

- The optimal forecast for T+1 given T is

$$\hat{y}_{T+1|T} = \alpha + \beta_1 y_T + \beta_2 y_{T-1}$$

- The forecast using the estimates is

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1}$$

Two-Step-Ahead Forecast

- The optimal two-step forecast is a linear function of two lags, with a MA(1) forecast error

$$
\begin{aligned}
y_t &= \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + e_t \\
&= \alpha + \beta_1 \left(\alpha + \beta_1 y_{t-2} + \beta_2 y_{t-3} + e_{t-1}\right) + \beta_2 y_{t-2} + e_t \\
&= \left(1 + \beta_1\right)\alpha + \left(\beta_1^2 + \beta_2\right) y_{t-2} + \beta_1 \beta_2 y_{t-3} + e_t + \beta_1 e_{t-1}
\end{aligned}
$$

Three-Step-Ahead Forecast

$$
\begin{aligned}
y_t &= (1+\beta_1)\alpha + \left(\alpha\beta_1 + \beta_1^2\right)y_{t-2} + \beta_1\beta_2 y_{t-3} + e_t + \beta_1 e_{t-1} \\
&= (1+\beta_1)\alpha + \left(\beta_1^2 + \beta_2\right)(\alpha + \beta_1 y_{t-3} + \beta_2 y_{t-4} + e_{t-2}) \\
&\quad + \beta_1\beta_2 y_{t-3} + e_t + \beta_1 e_{t-1} \\
&= \left(1 + \beta_1 + \beta_1^2 + \beta_2\right)\alpha + \left(\beta_1^3 + 2\beta_1\beta_2\right)y_{t-3} + \left(\beta_1^2\beta_2 + \beta_2^2\right)y_{t-4} \\
&\quad + e_t + \beta_1 e_{t-1} + \left(\beta_1^2 + \beta_2\right)e_{t-2}
\end{aligned}
$$

Iterated Rule

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1}$$

$$\hat{y}_{T+2|T} = \hat{\alpha} + \hat{\beta}_1 \hat{y}_{T+1|T} + \hat{\beta}_2 y_T$$

$$\hat{y}_{T+3|T} = \hat{\alpha} + \hat{\beta}_1 \hat{y}_{T+2|T} + \hat{\beta}_2 \hat{y}_{T+1|T}$$

$$\hat{y}_{T+h|T} = \hat{\alpha} + \hat{\beta}_1 \hat{y}_{T+h-1|T} + \hat{\beta}_2 \hat{y}_{T+h-1|T}$$

Direct Forecast

- Estimation is by least squares on two lags, h periods in past
- Forecast is least square prediction using final two observations

$$y_t = \hat{\alpha} + \hat{\beta}_1 y_{t-h} + \hat{\beta}_2 y_{t-h-1} + \hat{u}_t$$

$$\hat{y}_{T+h|T} = \hat{\alpha} + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1}$$

AR(p) Process

- An autoregressive process of order p, or AR(p) is

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + e_t$$

or

$$\left(1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p\right) y_t = \alpha + e_t$$

Stationarity

- The process is stationary if the inverses of the roots of the polynomial

$$\left(1-\beta_1 L-\beta_2 L^2-\cdots-\beta_p L^p\right)$$

are less than one (in absolute value)

- A necessary condition is that

$$\beta_1+\beta_2+\cdots+\beta_p<1$$

Alternative Representation

- We can write it as

$$y_t = \alpha + \gamma_1 y_{t-1} + \gamma_2 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p+1} + e_t$$

or

$$\Delta y_t = \alpha + (\gamma_1 - 1) y_{t-1} + \gamma_2 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p+1} + e_t$$

- These are equivalent forecasting models

Estimation of AR(p)

- Least Squares

$$y_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \hat{\beta}_1 y_{t-2} + \cdots + \hat{\beta}_p y_{t-p} + \hat{e}_t$$

Example: Unemployment Rate

```
. reg ur L(1/12).ur

      Source |       SS           df       MS      Number of obs   =       817
-------------+----------------------------------   F(12, 804)      =   1616.36
       Model |  5244.32031        12  437.026692   Prob > F        =    0.0000
    Residual |  217.382655       804  .270376436   R-squared       =    0.9602
-------------+----------------------------------   Adj R-squared   =    0.9596
       Total |  5461.70296       816  6.69326343   Root MSE        =    .51998

------------------------------------------------------------------------------
          ur |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ur |
         L1. |   .7328002   .0351296    20.86   0.000     .6638438    .8017567
         L2. |   .2756444   .0435137     6.33   0.000     .1902306    .3610583
         L3. |   .0772033   .0445546     1.73   0.084    -.0102539    .1646604
         L4. |   .0144794   .0445995     0.32   0.746    -.0730658    .1020247
         L5. |    .013264   .0445714     0.30   0.766     -.074226     .100754
         L6. |  -.0747289   .0445549    -1.68   0.094    -.1621866    .0127289
         L7. |  -.0530322   .0445815    -1.19   0.235    -.1405421    .0344776
         L8. |   .0422577   .0446145     0.95   0.344    -.0453169    .1298324
         L9. |  -.0671736   .0446553    -1.50   0.133    -.1548283    .0204811
        L10. |  -.0145669   .0444334    -0.33   0.743     -.101786    .0726523
        L11. |   .0714319    .043413     1.65   0.100    -.0137843    .1566482
        L12. |  -.0466789   .0348682    -1.34   0.181    -.1151223    .0217646
             |
       _cons |   .2756586   .0731781     3.77   0.000      .132016    .4193013
```

Iterated Forecasts

$$\hat{y}_{T+h|T} = \hat{\alpha} + \hat{\beta}_1 \hat{y}_{T+h-1|T} + \hat{\beta}_2 \hat{y}_{T+h-1|T} + \cdots + \hat{\beta}_p \hat{y}_{T-p+1|T}$$

Direct Forecasts

$$y_t = \hat{\alpha} + \hat{\beta}_1 y_{t-h} + \hat{\beta}_1 y_{t-h-1} + \cdots + \hat{\beta}_p y_{t-h-p+1} + \hat{u}_t$$

$$\hat{y}_{T+h|T} = \hat{\alpha} + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1} + \cdots + \hat{\beta}_p y_{T-p}$$

Example: Unemployment Rate

```
. reg ur L(12/23).ur
```

Source	SS	df	MS
Model	2992.38611	12	249.365509
Residual	2445.46892	793	3.08381958
Total	5437.85504	805	6.75509943

Number of obs	=	806
F(12, 793)	=	80.86
Prob > F	=	0.0000
R-squared	=	0.5503
Adj R-squared	=	0.5435
Root MSE	=	1.7561

ur	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ur						
L12.	.9785485	.1189435	8.23	0.000	.7450671	1.21203
L13.	.2807417	.1478383	1.90	0.058	-.009459	.5709423
L14.	-.0987124	.1514863	-0.65	0.515	-.396074	.1986491
L15.	-.1797846	.1517748	-1.18	0.237	-.4777125	.1181433
L16.	-.151464	.1516577	-1.00	0.318	-.449162	.1462339
L17.	-.1432067	.1514992	-0.95	0.345	-.4405936	.1541802
L18.	-.1146858	.1515847	-0.76	0.450	-.4122406	.182869
L19.	-.0124424	.151767	-0.08	0.935	-.310355	.2854701
L20.	.0052901	.1518535	0.03	0.972	-.2927923	.3033725
L21.	.0703658	.1511106	0.47	0.642	-.2262582	.3669899
L22.	.0993137	.1478976	0.67	0.502	-.1910034	.3896307
L23.	-.0168215	.1186637	-0.14	0.887	-.2497535	.2161106
_cons	2.654149	.2474935	10.72	0.000	2.168329	3.139969

- 12 periods ahead regression
- Predicted value (Jan 2018)=8.6% (current=8.3%)

12-month-ahead extrapolative forecast with AR(12)

Forecast Intervals at horizon *h*

- Residuals from the direct forecast estimates

$$y_t = \hat{\alpha} + \hat{\beta}_1 y_{t-h} + \hat{\beta}_1 y_{t-h-1} + \cdots + \hat{\beta}_p y_{t-h-p+1} + \hat{u}_t$$

- Forecast error variance

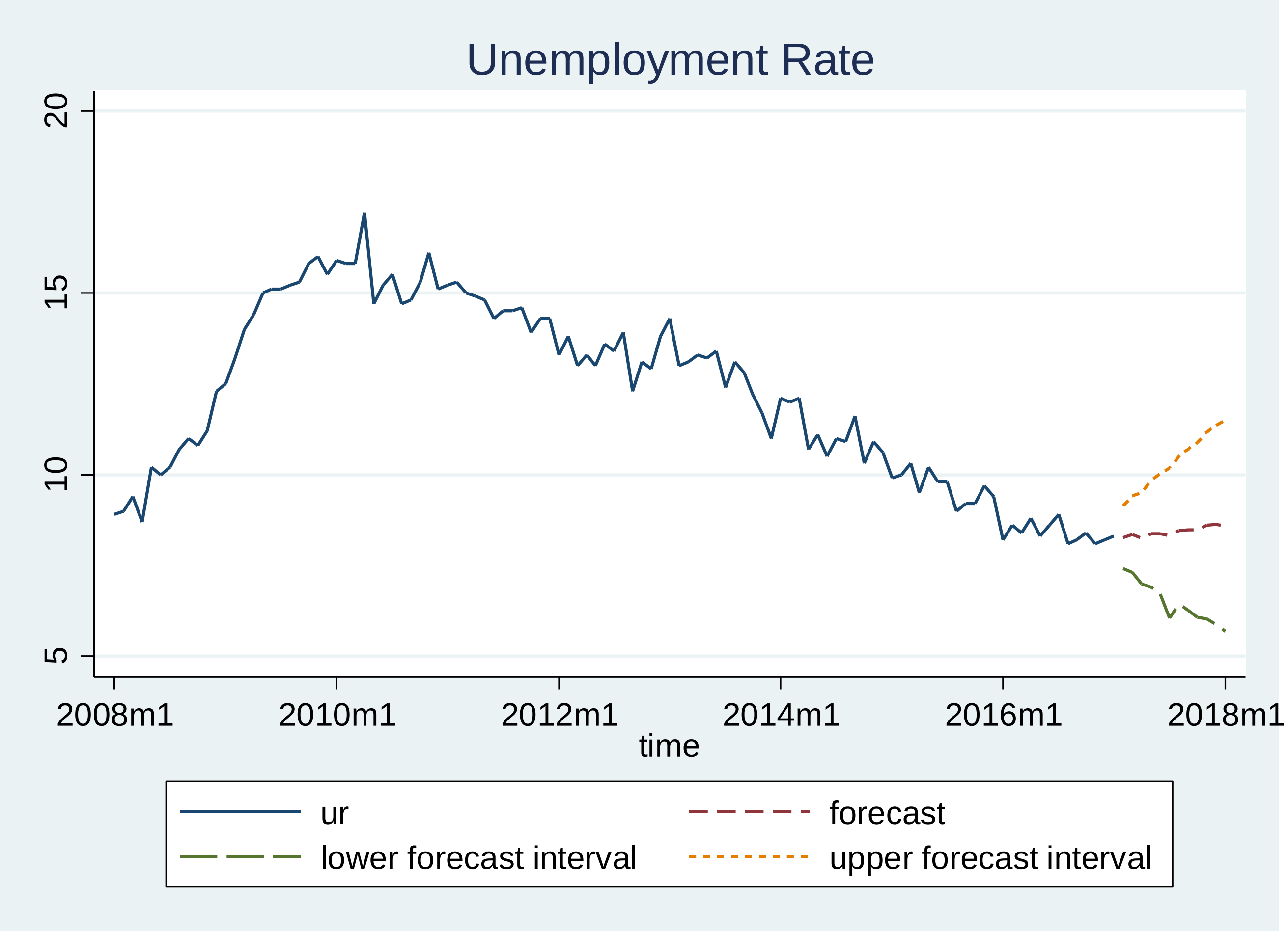
$$\hat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2$$

- (1-α)% Forecast interval

$$\hat{y}_{T+h|T} \pm \hat{\sigma}_u \cdot z_{\alpha/2}$$

- Identical to AR(1) model

90% Direct forecast intervals

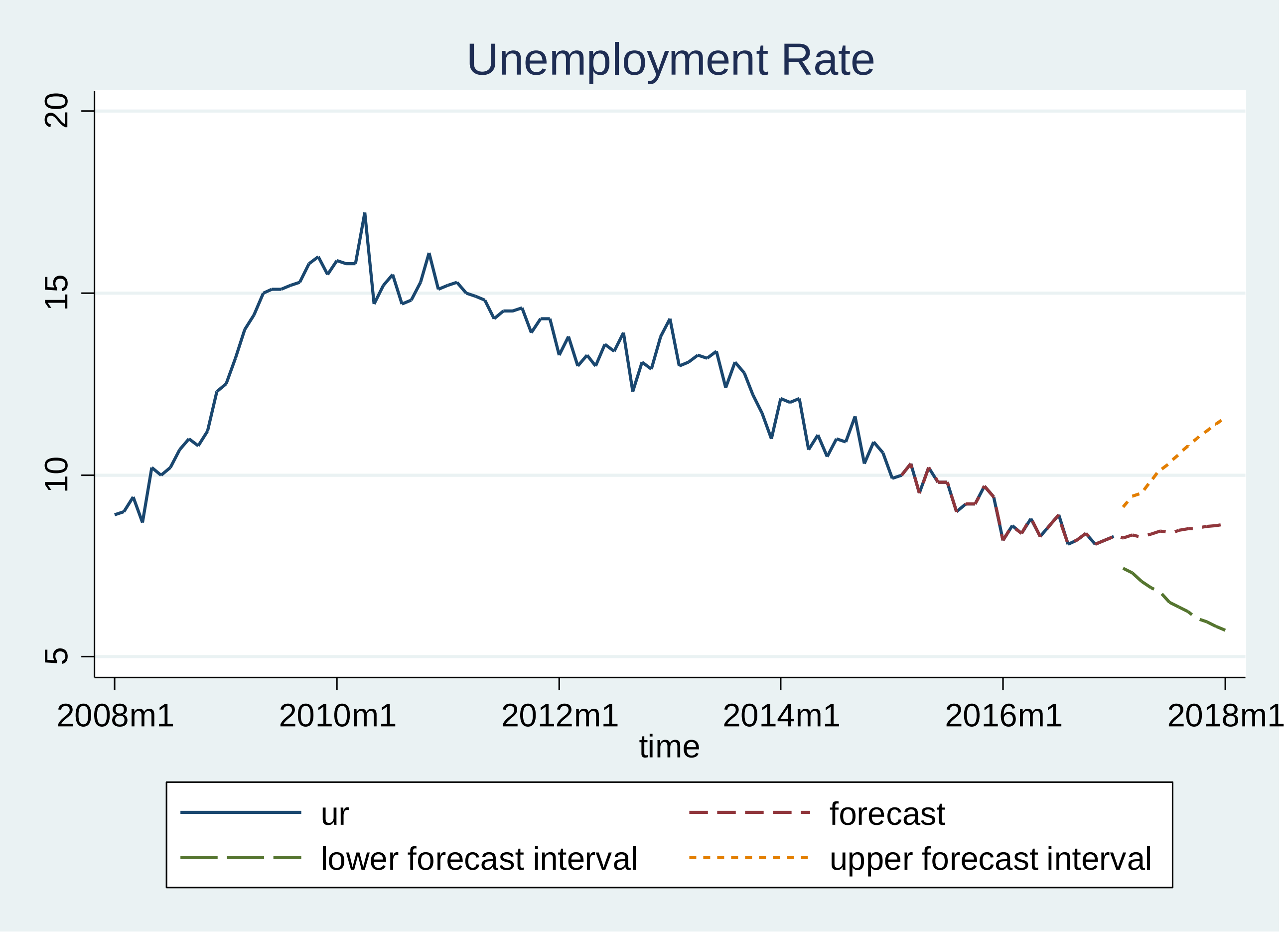


12-step Direct Interval Forecasts

```
use ur.dta, clear
tsappend, add(12)
reg ur L(1/12).ur
predict y1
predict sf1,stdf
gen y1L=y1-1.645*sf1
gen y1U=y1+1.645*sf1
reg ur L(2/13).ur
predict y2
predict sf2,stdf
gen y2L=y2-1.645*sf2
gen y2U=y2+1.645*sf2
reg ur L(3/14).ur
predict y3
predict sf3,stdf
gen y3L=y3-1.645*sf3
gen y3U=y3+1.645*sf3
…
```

```
…
egen p=rowfirst(y1 y2 y3 y4 y5 y6 y7 y8 y9 y10 y11
y12) if t>=tm(2017m2)
egen pL=rowfirst(y1L y2L y3L y4L y5L y6L y7L y8L y9L
y10L y11L y12L) if t>=tm(2017m2)
egen pU=rowfirst(y1U y2U y3U y4U y5U y6U y7U y8U
y9U y10U y11U y12U) if t>=tm(2017m2)
label variable p "forecast"
label variable pL "lower forecast interval"
label variable pU "upper forecast interval"
tsline ur p pL pU if t>=tm(2008m1),
title(Unemployment Rate) lpattern (solid dash
longdash shortdash)
tsline ur p if t>=tm(2008m1), title(Unemployment
Rate) lpattern (solid dash)
```

90% Iterated forecast intervals



12-step Iterated Interval Forecasts

```
use ur.dta, clear
tsappend, add(12)
reg ur L(1/12).ur
forecast create ar12
estimate store model1
forecast estimates model1
forecast solve, simulate(errors,statistic(stddev,prefix(sd_)) reps(1000))
gen p = f_ur if t>=tm(2017m2)
gen pL = f_ur-1.645*sd_ur if t>=tm(2017m2)
gen pU = f_ur+1.645*sd_ur if t>=tm(2017m2)
label variable p "forecast"
label variable pL "lower forecast interval"
label variable pU "upper forecast interval"
tsline ur p pL pU if t>=tm(2008m1), title(Unemployment Rate) lpattern (solid dash
longdash shortdash)
```

Assignments

- Read Diebold Chapter 9
- Problem Set # 6
  - Due Tuesday (2/28)
- Read Chapter 6 from *The Signal and the Noise*
  - Reading Reflection
  - Due Thursday (3/2)