Forecast Standard Errors

- Wooldridge, Chapter 6.4
- Multiple Regression

$$y_{t+h} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + e_t$$

- Includes intercept, trend, and autoregressive models (x can be lagged y)
- OLS estimate

$$y_{t+h} = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \cdots + \hat{\beta}_k x_{kt} + \hat{e}_t$$

Prediction Variance

- Point prediction

$$\hat{y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 x_{1T} + \hat{\beta}_2 x_{2T} + \cdots + \hat{\beta}_k x_{kT}$$

- This is also an estimate of the regression function at these values of the x's
- Variance of point prediction

$$\operatorname{var}(\hat{y}_{T+h}) = \operatorname{var}\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1T} + \hat{\beta}_2 x_{2T} + \cdots + \hat{\beta}_k x_{kT}\right)$$

- This is a function of the variances of the OLS estimates, weighted by the x's

Prediction Standard Errors

- Standard error of point prediction

$$se(\hat{y}_{T+h}) = \sqrt{\mathrm{var}(\hat{y}_{T+h})}$$

- This is the standard error of a linear combination (the x's) of the coefficients.
- Computed in STATA using stdp option for predict command
  - .predict s, stdp
- Important: This is very different than stdf

Forecast Error

- Forecast error

$$\hat{e}_{T+h} = y_{T+h} - \hat{y}_{T+h}$$

- Variance of forecast error

$$\begin{aligned}\operatorname{var}\left(\hat{e}_{T+h}\right) &= \operatorname{var}\left(y_{T+h}\right) + \operatorname{var}\left(\hat{y}_{T+h}\right) \\ &= \sigma^2 + \operatorname{var}\left(\hat{y}_{T+h}\right)\end{aligned}$$

- Two components:
  - Equation variance σ^2
  - Estimation variance $\operatorname{var}\left(\hat{y}_{T+h}\right)$

Forecast Error Variance

- Variance of forecast error

$$\operatorname{var}\left(\hat{e}_{T+h}\right) = \sigma^2 + \operatorname{var}\left(\hat{y}_{T+h}\right)$$
$$\approx \sigma^2$$

- Model variance tends to be much larger than estimation variance
- Estimation variance decreases with sample size T

Forecast standard error

$$se\left(\hat{e}_{T+h}\right) = \sqrt{\hat{\sigma}^2 + \mathrm{var}\left(\hat{y}_{T+h}\right)}$$
$$= \sqrt{\hat{\sigma}^2 + se\left(\hat{y}_{T+h}\right)^2}$$

- Computed in STATA using stdf option
  - .predict s, stdf
- Typically will be close to (just a little larger than) $\hat{\sigma}$

GDP Example

```
. reg gdp L.gdp

      Source |       SS       df       MS              Number of obs =     266
-------------+------------------------------           F(  1,   264) =   42.85
       Model |  584.759539     1  584.759539           Prob > F      =  0.0000
    Residual |  3602.38588   264  13.6454011           R-squared     =  0.1397
-------------+------------------------------           Adj R-squared =  0.1364
       Total |  4187.14542   265  15.8005487           Root MSE      =   3.694

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |
         L1. |   .3730975   .0569937     6.55   0.000     .2608774    .4853175
             |
       _cons |    2.07642   .2940597     7.06   0.000      1.49742    2.655421
------------------------------------------------------------------------------

. tsappend,add(1)

. predict sp, stdp
(1 missing value generated)

. predict sf,stdf
(1 missing value generated)

. gen s=e(rmse)
```

GDP Example

- From the Data Editor

time	sp	sf	s
2014q1	.2265	3.700	3.694

- Notice
  - s equals "Root MSE" from regression output
  - The estimates satisfy the relationship

$$sf^2 = \sqrt{sp^2 + s^2}$$

  - sf and s are very close
  - sf (standard error of forecast) is better
    - But s (error standard deviation) is often sufficient

Two-Step-Ahead Point Forecasting

- Three methods
  - Plug-in
    - Calculates optimal forecast as function of AR model
    - Replaces unknowns with estimates
  - Iterated
    - Calculates one-step forecast, and then iterates to get second-step forecast
  - Direct
    - Estimates 2-step regression function, and uses this for forecast
- We start with point forecasts, and then discuss interval forecasts

Plug-In Method

- By back-substitution

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

- Thus

$$y_{T+2} = (1+\beta)\alpha + \beta^2 y_T + e_{T+2} + \beta e_{T+1}$$

$$E\left(y_{T+2} \mid \Omega_T\right) = (1+\beta)\alpha + \beta^2 y_T$$

Point Forecast

- The optimal forecast is

$$\hat{y}_{T+2|T} = (1+\beta)\alpha + \beta^2 y_T$$

- This is a function of the AR(1) parameters
- Plug-in (replace unknowns with estimates) to obtain a feasible forecast

$$\hat{y}_{T+2|T} = (1+\hat{\beta})\hat{\alpha} + \hat{\beta}^2 y_T$$

- This method is feasible but cumbersome for multi-step forecasts and complicated models

Iterated Method

- Take conditional expectations at time T

$$y_{T+2} = \alpha + \beta y_{T+1} + e_{T+2}$$

$$E(y_{T+2} \mid \Omega_T) = \alpha + \beta E(y_{T+1} \mid \Omega_T) + E(e_{T+2} \mid \Omega_T)$$

$$= \alpha + \beta E(y_{T+1} \mid \Omega_T)$$

- The left-side is the 2-step forecast, the right-side is linear in the 1-step forecast. Thus:

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

Iteration

- We already know how to compute the one-step point forecast

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

- The second step iterates on the one-step

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

- This method is convenient in linear models (our main focus)
- It does not work in nonlinear models
- It is less useful in regression contexts (later sections)

Direct Method

- We showed that

$$y_t = (1+\beta)\alpha + \beta^2 y_{t-2} + e_t + \beta e_{t-1}$$
$$= \alpha^* + \beta^* y_{t-2} + u_t$$

where

$$\alpha^* = (1+\beta)\alpha$$
$$\beta^* = \beta^2$$
$$u_t = e_t + \beta e_{t-1}$$

Estimation of Direct Method

- This is a regression

$$y_t = \alpha^* + \beta^* y_{t-2} + u_t$$

- The error is the two-step forecast error
- It can be estimated **directly** by least-squares
- This is actually different than the iterated estimator.
- The error *u* is not white noise, but is uncorrelated with the regressor

Example – GDP Growth

- α=2.08, β=0.373, y_T =3.2, $y_{T+1|T}$ =3.3
- Plug-in:

$$\hat{y}_{T+2|T} = \left(1+\hat{\beta}\right)\times\hat{\alpha}+\hat{\beta}^2 y_T$$
$$= \left(1+.37\right)\times 2.08 + .37^2 \times 3.2$$
$$= 3.3\%$$

- Iterated:

$$\hat{y}_{T+2|T} = \hat{\alpha}+\hat{\beta}\hat{y}_{T+1|T}$$
$$= 2.08 + .37 \times 3.3$$
$$= 3.3\%$$

Example – GDP Growth

- The equality of Plug-in and Iterated 2-step forecast is typical
- The equality of the 1-step and 2-step forecast is not typical. It is an accident of the fact that last quarter's GDP growth (3.3%) is the model average: 2.08/(1 - 0.373)=3.3

STATA Forecast Command

```
. tsappend, add(2)

. forecast create ar1
  Forecast model ar1 started.

. estimates store model1

. forecast estimates model1
  Added estimation results from regress.
  Forecast model ar1 now contains 1 endogenous variable.

. forecast solve
```

- “forecast create [name1]”
- “estimates store [name2]” (after a regression)
- “forecast estimates [name2]” tells STATA to forecast using the estimates from name2
- “forecast solve” creates the forecasts, and stores then in the dataset

STATA Forecast output

time	f_gdp
2014q1	3.27033
2014q2	3.29657

- These are the one-step and two-step iterated point forecasts from the AR(1) model

GDP Growth, Direct 2-step

```
. reg gdp L2.gdp

      Source |       SS       df       MS              Number of obs =     265
-------------+------------------------------           F(  1,   263) =   13.37
       Model |  201.880328     1  201.880328           Prob > F      =  0.0003
    Residual |  3971.49269   263  15.1007327           R-squared     =  0.0484
-------------+------------------------------           Adj R-squared =  0.0448
       Total |  4173.37302   264  15.8082311           Root MSE      =   3.886

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |
         L2. |   .2192374   .0599607     3.66   0.000     .1011733    .3373015
             |
       _cons |    2.59736    .309583     8.39   0.000     1.987783    3.206936
------------------------------------------------------------------------------
```

- Estimate

$$y_t = 2.60 + 0.22 y_{t-2} + \hat{u}_t$$

- Notice .22>.14=.37^2 from iterated

Direct 2-step-ahead Forecast

- 2-step forecast

$$\begin{aligned}\hat{y}_{T+2|T} &= \hat{\alpha}^* + \hat{\beta}^* y_T \\ &= 2.60 + 0.22 \times 3.2 \\ &= 3.3\%\end{aligned}$$

- It happens to be the same as from the iterated method, but this is not typical.

2-Step Forecast Error

- Recall $y_t = \alpha^* + \beta^* y_{t-2} + u_t$

  where $u_t = e_t + \beta e_{t-1}$
- The equation error is *u*, not *e*
- It has variance

$$\begin{aligned} \mathrm{var}(u_t) &= \sigma_u^2 \\ &= \mathrm{var}(e_t + \beta e_{t-1}) \\ &= \left(1 + \beta^2\right)\sigma^2 \end{aligned}$$

- This is different than the one-step variance

Forecast variance estimation

- For forecast intervals, we need an estimate of

$$\mathrm{var}(u_t) = \sigma_u^2$$

- Not

$$\mathrm{var}(e_t) = \sigma^2$$

Plug-in Forecast variance estimation

- Use formula, and replace by estimates

$$\hat{\sigma}_u^2 = \left(1 + \hat{\beta}^2\right)\hat{\sigma}^2$$

$$\hat{\sigma}_u = \sqrt{\hat{\sigma}_u^2}$$

- This formula is hard to generalize beyond AR(1)

Example: GDP Growth Plug-in Estimate

- β=.37, σ=3.69

$$\hat{\sigma}_u = \sqrt{\left(1+\hat{\beta}^2\right)\hat{\sigma}^2}$$

$$= \sqrt{\left(1+.37^2\right)3.69^2}$$

$$= 3.9$$

Direct Forecast variance estimation

$$\hat{u}_t = y_t - \hat{\alpha}^* - \hat{\beta}^* y_{t-2}$$

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

Direct Estimate

```
. reg gdp L2.gdp

      Source |       SS       df       MS              Number of obs =     265
-------------+------------------------------           F(  1,   263) =   13.37
       Model |  201.880328     1  201.880328           Prob > F      =  0.0003
    Residual |  3971.49269   263  15.1007327           R-squared     =  0.0484
-------------+------------------------------           Adj R-squared =  0.0448
       Total |  4173.37302   264  15.8082311           Root MSE      =   3.886

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |
         L2. |   .2192374   .0599607     3.66   0.000     .1011733    .3373015
             |
       _cons |    2.59736    .309583     8.39   0.000     1.987783    3.206936
------------------------------------------------------------------------------
```

- Estimate $\hat{\sigma} = 3.886$
- Stdf $se(\hat{e}) = 3.893$

Iterated Forecast Variance Estimation

- Not easy to calculate directly
- The forecast errors *u* not a direct output
- Instead, it is typical to use *simulation* to calculate forecast variance
- This can be more flexible than the formulae
- Can be done in STATA using forecast command

```
. forecast solve, simulate(errors,statistic(stddev,prefix(sd_)) reps(1000))
```

Iterated Forecast Variance Estimation

```
. forecast solve, simulate(errors,statistic(stddev,prefix(sd_)) reps(1000))
```

- The *simulate* option creates simulated out-of-sample series from the model
- The *statistic* option tells STATA what to save (standard deviations)
- The *prefix* option tells STATA to save the standard deviations in the format sd_name, where “name” was the variable you are forecasting.
- The *reps* option tells STATA to use 1000 simulations (otherwise 50 is the default)
- This command creates the point forecasts f_gdp and standard derivations sd_gdp

GDP example

- This shows the 1-step and 2-step point forecasts (3.27 and 3.29), and the 1-step and 2-step forecast standard errors (3.7 and 3.9)

time	f_gdp	_est_model1	sd_gdp
2014q1	3.27033	0	3.70659
2014q2	3.29657	0	3.88856

- These are the same as from other methods

Two-Step-Ahead Intervals

- Normal Method
  - Forecast interval is point estimate, plus and minus the estimated standard deviation multiplied by a normal quantile
  - For a 95% interval:

$$\hat{y}_{T+2|T} \pm \hat{\sigma}_u \cdot z_{.025} = \hat{y}_{T+2|T} \pm \hat{\sigma}_u \cdot 1.96$$

  - For a 90% interval

$$\hat{y}_{T+2|T} \pm \hat{\sigma}_u \cdot z_{.05} = \hat{y}_{T+2|T} \pm \hat{\sigma}_u \cdot 1.645$$

GDP Growth Example

- In this example, the Plug-In, Iterated and Direct estimates are the same
  - $y_{T+2|T}$ =3.3%, σ_u=3.9
  - 3.3% ± 1.645*3.9=[-3.1%, 9.7%]

h-Step-Ahead Forecasting

$$\hat{y}_{T+h|T}$$

h-Step-Ahead back substitution

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

h-Step-Ahead Point Forecast

- Optimal

$$E(y_{T+h} \mid \Omega_T) = (1 + \beta + \beta^2 + \cdots + \beta^h)\alpha + \beta^h y_T$$

- Plug-In

$$\hat{y}_{T+h|T} = (1 + \hat{\beta} + \hat{\beta}^2 + \cdots + \hat{\beta}^h)\hat{\alpha} + \hat{\beta}^h y_T$$

- Iterated

$$y_{T+h} = \alpha + \beta y_{T+h-1} + e_{T+h}$$

$$E(y_{T+h} \mid \Omega_T) = \alpha + \beta E(y_{T+h-1} \mid \Omega_T)$$

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

Direct Method

- Best Linear predictor

$$y_t = \alpha^* + \beta^* y_{t-h} + u_t$$

- Least-Squares estimator

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

- h-step forecast

$$\hat{y}_{T+h|T} = \hat{\alpha}^* + \hat{\beta}^* y_T$$

Direct Estimates

- Least Squares

$$y_t = 2.07 + 0.37 y_{t-1} + \hat{e}_t$$

$$y_t = 2.60 + 0.22 y_{t-2} + \hat{u}_t$$

$$y_t = 3.23 + 0.02 y_{t-3} + \hat{u}_t$$

$$y_t = 3.51 - 0.06 y_{t-4} + \hat{u}_t$$

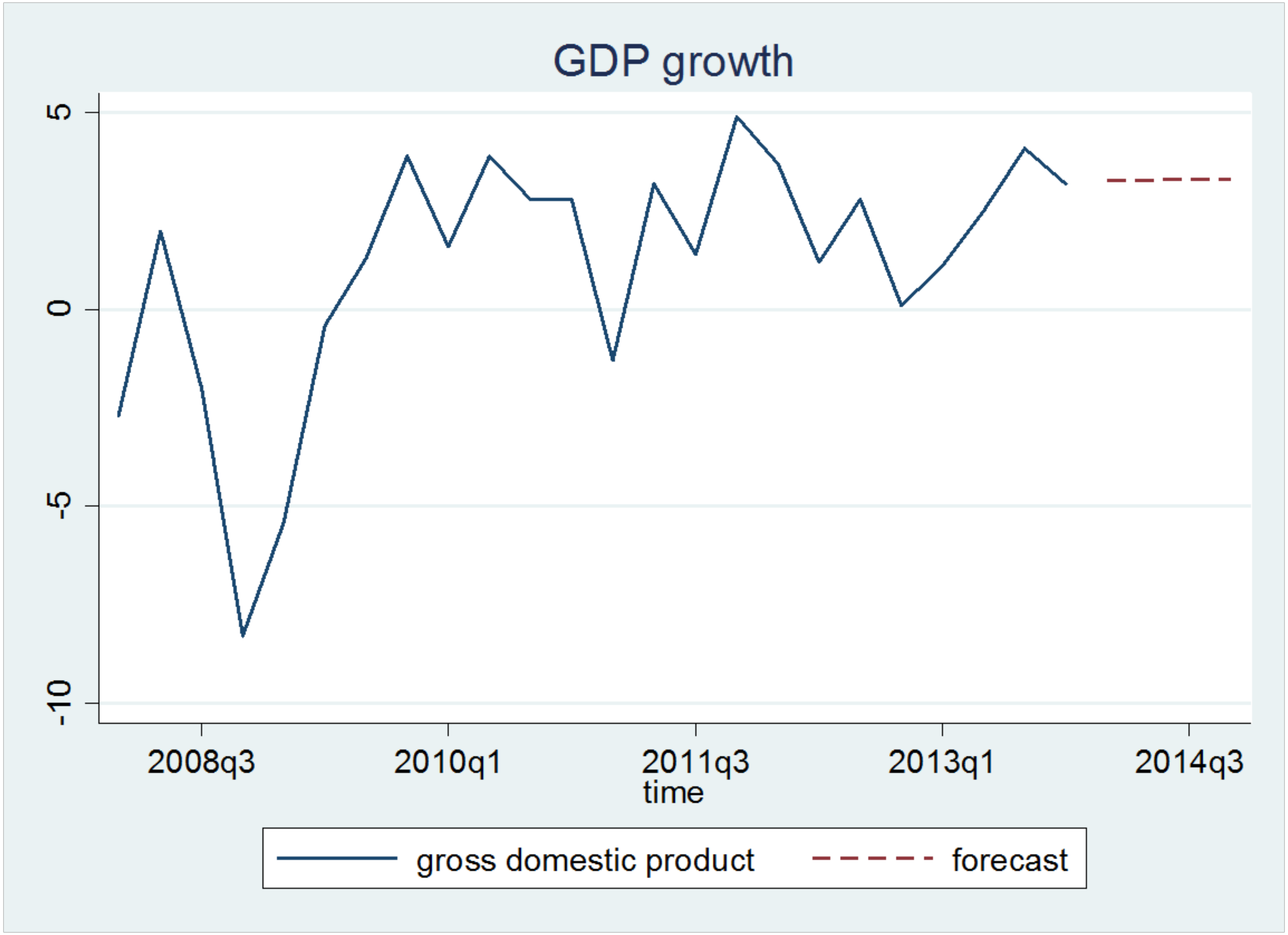
Iterated and Direct Point Estimates

	Iterated	Direct
2014Q1	3.3	3.3
2014Q2	3.3	3.3
2014Q3	3.3	3.3
2014Q4	3.3	3.3

4-Step Direct Point Forecast

```
use gdp2013.dta
tsappend, add(4)
reg gdp L.gdp
predict y1
reg gdp L2.gdp
predict y2
reg gdp L3.gdp
predict y3
reg gdp L4.gdp
predict y4
egen p=rowfirst(y1 y2 y3 y4) if t>=tq(2014q1)
label variable p “forecast”
tsline gdp p if t>=tq(2008q1), title(GDP growth) lpattern (solid dash)
```

Point Forecast (Direct)



- There are 4 periods out-of-sample
- The **predict** command computes fitted values for observations which have the needed variables.
- For the regression on the first lag (L.gdp), this works only for the first out-of-sample observation, the remainder are coded as missing.
- For the regression on the second lag (L2.gdp), this works for the fist two out-of-sample observations
- The egen command is used in STATA for more complicated versions of “generate”
- egen p=rowfirst(y1 y2 y3 y4) takes the first variable in the list which is not missing

Forecasts

t	y1	y2	y3	y4	p
2013q4	3.61	3.15	3.25	3.50	
2014q1	**3.27**	3.50	3.29	3.44	**3.27**
2014q2		**3.30**	3.33	3.35	**3.30**
2014q3			**3.30**	3.24	**3.30**
2014q4				**3.30**	**3.30**

4-Step Iterated Point Forecast

```
use gdp2013.dta
tsappend, add(4)
reg gdp L.gdp
forecast create ar1
estimate store model1
forecast estimates model1
forecast solve
gen p=f_gdp if t>=tq(2014q1)
label variable p “forecast”
tsline gdp p if t>=tq(2008q1), title(GDP growth) lpattern
   (solid dash)
```