Forecasting
ECON 492

Viktoria Hnatkovska

August 15, 2018

Models vs Reality

Many macro models have the following **autoregressive** form:

$$x_t = \rho_0 + \rho_1 x_{t-1}$$

x and ρ_0 can be vectors and ρ_1 can be a matrix

Example: neo-classical growth model

- capital accumulation equation:

$$k_{t+1} = Ak_t^\alpha \quad \Rightarrow \quad \underbrace{\ln k_{t+1}}_{x_t} = \underbrace{\ln A}_{\rho_0} + \underbrace{\alpha \ln k_t}_{\rho_1 x_{t-1}}$$

Models vs Reality

Many macro models have the following autoregressive form:

$$x_t = \rho_0 + \rho_1 x_{t-1}$$

x and ρ_0 can be vectors and ρ_1 can be a matrix

Example: two-state labor market model

- employment/unemployment dynamics:

$$\begin{bmatrix} e_t \\ u_t \end{bmatrix} = \begin{bmatrix} 1-s & f \\ 1-f & s \end{bmatrix} \begin{bmatrix} e_{t-1} \\ u_{t-1} \end{bmatrix}$$

Models vs Reality

Many macro *models* have the following autoregressive form:

$$x_t = \rho_0 + \rho_1 x_{t-1}$$

Reality is more accurately described as a vector autoregression (VAR):

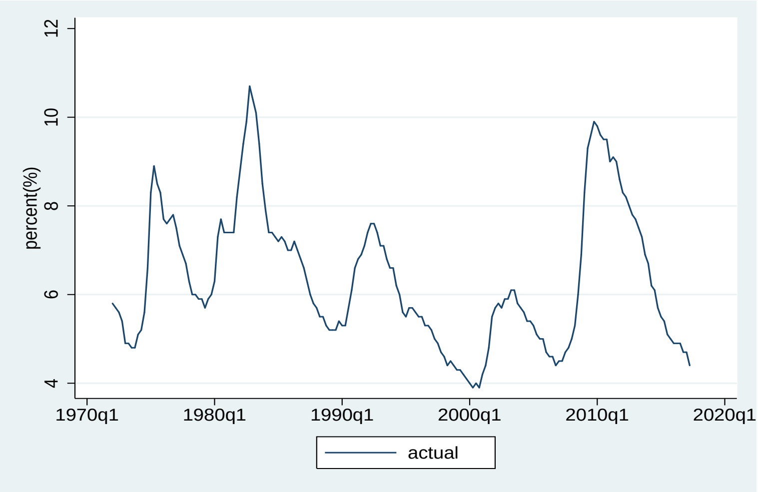
$$x_t = \rho_0 + \rho_1 x_{t-1} + \rho_2 x_{t-2} + \rho_3 x_{t-3} + \ldots + \epsilon_t$$

- system may be higher than just first-order autoregressive
- system is hit with shocks
- we don't know the values of ρ_0, ρ_1, ...
- if x_t is **stationary**, we can estimate the VAR using OLS!

Data transformations

- Seasonally adjusted?
- Transform quantities in logs?
- Stationary?

Scalar Example: Unemployment Rate



Civilian Unemployment Rate, 1972q1-2017q2. Source: fred.stlouisfed.org.

Scalar Example: Unemployment Rate

STATA commands:

```
insheet using "data20170923.csv", clear
```

- loads the data from data20170923.csv of the Excel file

```
generate qtr = tq(1967q1) + time - 1
format qtr %tq
```

- uses the variable "time" $(= 1, 2, \ldots)$ and creates new variable "qtr" $(= 1967q1, 1967q2, \ldots)$

```
tsset qtr
```

- tells STATA that the data is time series data, and "qtr" defines the time dimension

Unit root tests

Formal tests are developed to test for non-stationarity

Dickey-Fuller (DF) and augmented DF are examples of such tests

- DF runs: $y_t = \alpha + \rho y_{t-1} + \delta t + \epsilon_t$
  - H_0 is $\rho = 1$
- Augmented DF controls for possible serial correlation: $\Delta y_t = \alpha + \beta y_{t-1} + \delta t + \psi_1 \Delta y_{t-1} + \psi_2 \Delta y_{t-2} + ... + \epsilon_t$
  - H_0 is $\beta = 0 \Leftrightarrow \rho = 1$

Scalar Example: Unemployment Rate

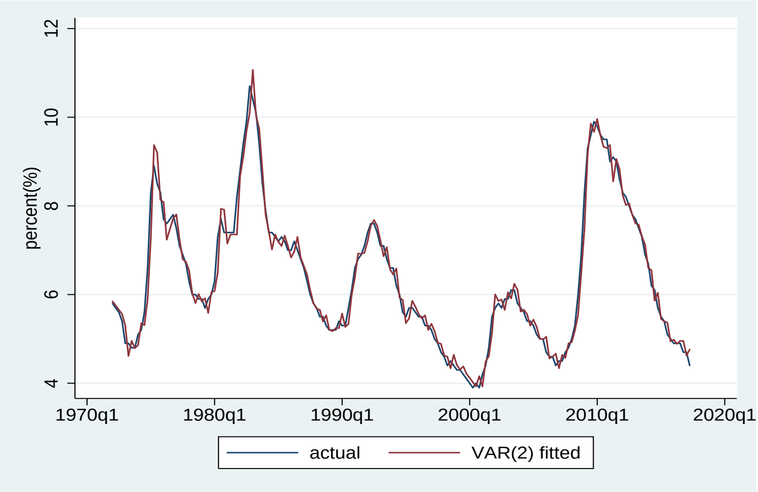
`var unemp if qtr>=tq(1972q1), lags(1 2)`

- runs an AR(2) on "unemp" using data starting in 1972q2

```
. var unemp if qtr>=tq(1972q1), lags(1/2)

Vector autoregression

Sample:  1972q1 - 2017q2                        No. of obs      =       182
Log likelihood = -9.110109                      AIC             =  .1330781
FPE            =   .066884                      HQIC            =  .1544879
Det(Sigma_ml)  =  .0647148                      SBIC            =  .1858914

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
unemp                 3     .256514   0.9731   6582.07   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
        unemp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
unemp        |
        unemp |
          L1. |   1.636853   .0549505    29.79   0.000     1.529152    1.744554
          L2. |   -.676608   .0551195   -12.28   0.000    -.7846401   -.5685758
              |
        _cons |   .2497777   .0806643     3.10   0.002     .0916786    .4078768
------------------------------------------------------------------------------
```

- So: $u_t = 0.250 + 1.637u_{t-1} - 0.677u_{t-2}, \quad R^2 = 0.973$

Scalar Example: Unemployment Rate



In forecasting we care about "fit" (i.e. R^2), not precision (std err)

Scalar Example: Unemployment Rate

We have unemployment data that ends in 2017q2

- **forecast** for 2017q3?

$$\hat{u}_{2017q3} = \hat{\rho}_0 + \hat{\rho}_1 u_{2017q2} + \hat{\rho}_2 u_{2017q1}$$

Scalar Example: Unemployment Rate

We have unemployment data that ends in 2017q2

- **forecast** for 2017q3?

$$\hat{u}_{2017q3} = \hat{\rho}_0 + \hat{\rho}_1 u_{2017q2} + \hat{\rho}_2 u_{2017q1}$$

- **forecast** for 2017q4?

Scalar Example: Unemployment Rate

We have unemployment data that ends in 2017q2

- **forecast** for 2017q3?

$$\hat{u}_{2017q3} = \hat{\rho}_0 + \hat{\rho}_1 u_{2017q2} + \hat{\rho}_2 u_{2017q1}$$

- **forecast** for 2017q4?

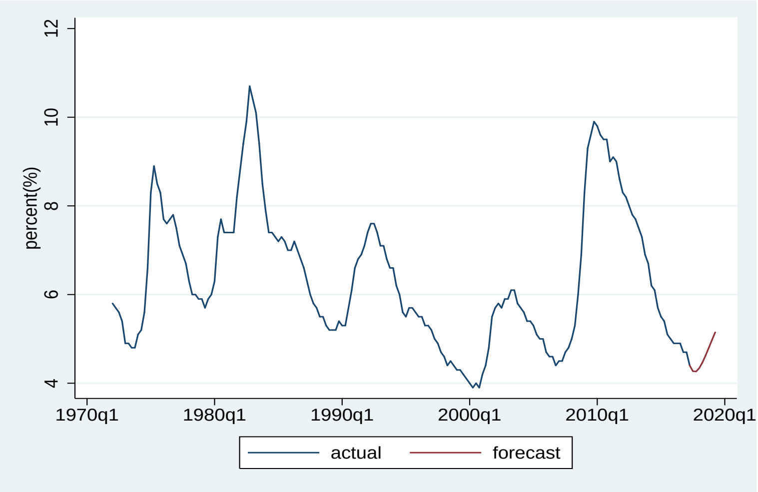
$$\hat{u}_{2017q4} = \hat{\rho}_0 + \hat{\rho}_1 \hat{u}_{2017q3} + \hat{\rho}_2 u_{2017q2}$$

- **forecast** for 2017q5?

$$\hat{u}_{2017q5} = \hat{\rho}_0 + \hat{\rho}_1 \hat{u}_{2017q4} + \hat{\rho}_2 \hat{u}_{2017q3}$$

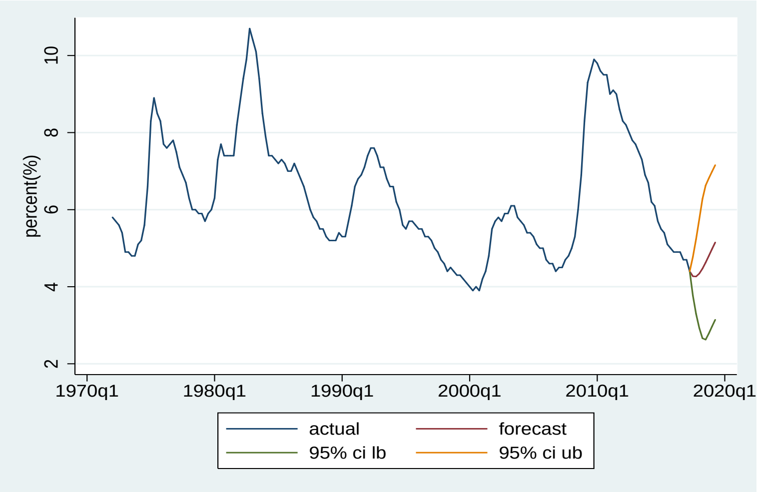
and so on ...

Scalar Example: Unemployment Rate



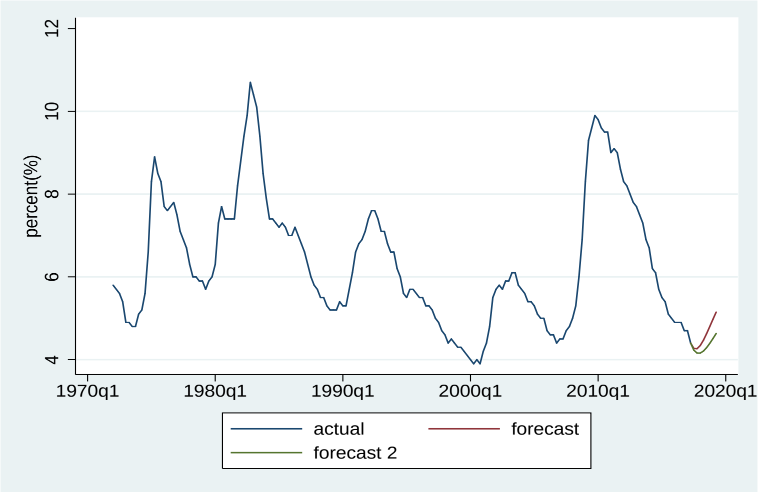
Forecast is for a return to sample average

Scalar Example: Unemployment Rate



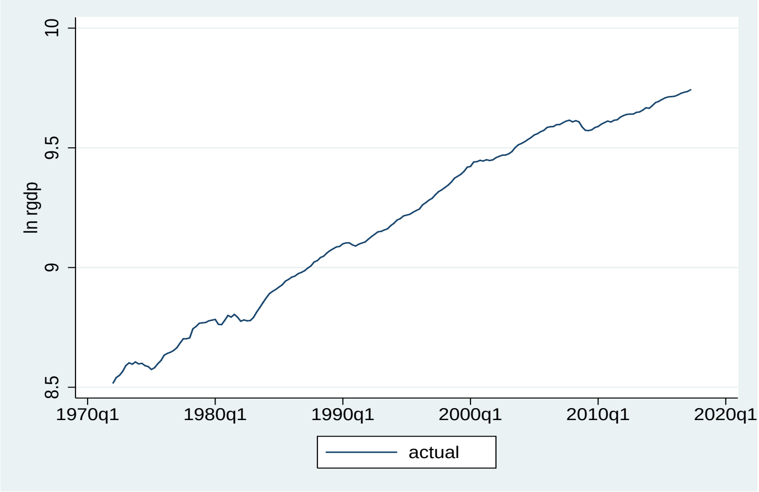
Forecasts get less accurate the farther you go

Scalar Example: Unemployment Rate



New "forecast 2" based on 1989q1-2017q2 data

Scalar Example: (natural log of) RGDP



Real GDP (2009$s), 1972q1-2017q2. Source: fred.stlouisfed.org.

Scalar Example: RGDP

Autoregressive time series models do **not** behave well for variables that are growing (i.e. non-stationary)

In addition, they're not very informative for forecasting:

$$
\begin{aligned}
y_t &= (1 + g_t) y_{t-1} \\
\ln y_t &= \underbrace{\ln(1 + g_t)}_{\approx g_t} + \ln y_{t-1}
\end{aligned}
$$

What if I run regression: $\ln y_t = \rho_0 + \rho_1 \ln y_{t-1} + \epsilon_t$?

Scalar Example: RGDP

`var lrgdp if qtr>=tq(1972q1), lags(1)`

- I should get $\rho_1 \approx 1$

```
. var lrgdp if qtr>=tq(1972q1), lags(1)

Vector autoregression

Sample:  1972q1 - 2017q2                        No. of obs      =        182
Log likelihood =  623.8009                      AIC             =  -6.832977
FPE            =  .0000631                      HQIC            =  -6.818704
Det(Sigma_ml)  =  .0000617                      SBIC            =  -6.797768

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
lrgdp                2      .0079   0.9996   411020.2   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
       lrgdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lrgdp        |
       lrgdp |
         L1. |   .9963956   .0015542   641.11   0.000     .9933495    .9994418
             |
       _cons |   .0399435   .0142906     2.80   0.005     .0119343    .0679526
------------------------------------------------------------------------------
```

Scalar Example: RGDP

If y_t is growing:

$$\Delta y_t \equiv \ln y_t - \ln y_{t-1} \approx g_t$$

is **stationary**!

I can run regression:

$$\Delta y_t = \rho_0 + \rho_1 \Delta y_{t-1} + \ldots + \epsilon_t$$

- ρ_1: if Δy_{t-1} was lower than usual last quarter, is Δy_t likely to be higher/lower than usual today?
  - in other words, are output growth fluctuations *persistent*?

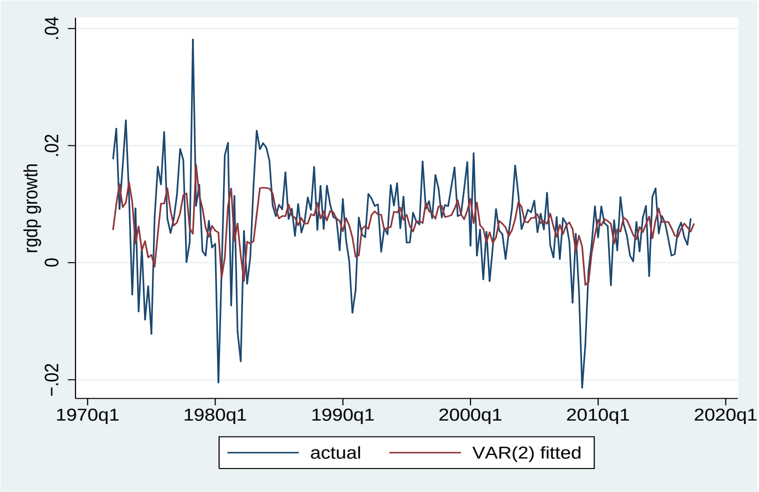
Scalar Example: RGDP

```
var growth if qtr>=tq(1972q1), lags(1 2)
```

```
. var growth if qtr>=tq(1972q1), lags(1/2)

Vector autoregression

Sample:  1972q1 - 2017q2                        No. of obs      =        182
Log likelihood = -202.1218                      AIC             =   2.254086
FPE            =  .5577813                      HQIC            =   2.275496
Det(Sigma_ml)  =  .5396911                      SBIC            =   2.306899

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
growth               3     .740767   0.1508   32.30694   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
growth       |
      growth |
         L1. |   .3286373   .0735593     4.47   0.000     .1844638    .4728109
         L2. |   .1175306   .0736008     1.60   0.110    -.0267244    .2617856
             |
       _cons |   .3790087   .0782479     4.84   0.000     .2256456    .5323717
```

- RGDP growth hard to explain: $R^2 = 0.151 \quad \Rightarrow \quad$ be cautious forecasting!

Scalar Example: RGDP



Can use growth rates to get RGDP in (log) levels, plot that instead

Vector Autoregressions

$$x_t = \rho_0 + \rho_1 x_{t-1} + \epsilon_t$$

$$\begin{bmatrix} u_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \rho_0^u \\ \rho_0^y \end{bmatrix} + \begin{bmatrix} \rho_1^{uu} & \rho_1^{uy} \\ \rho_1^{yu} & \rho_1^{yy} \end{bmatrix} \begin{bmatrix} u_{t-1} \\ \Delta y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \end{bmatrix}$$

- u_t depends not just on u_{t-1} but **also** Δy_{t-1}
- forecast for how quickly unemployment returns to "usual" depends on whether output growth is above/below "usual"
- same goes for Δy_t
- obvously, you can consider more lags

Vector Autoregressions

A more interesting example (with only one lag):

$$\begin{bmatrix} r_t \\ u_t \\ \Delta y_t \\ \Delta p_t \end{bmatrix} = \rho_0 + \rho_1 \begin{bmatrix} r_{t-1} \\ u_{t-1} \\ \Delta y_{t-1} \\ \Delta p_{t-1} \end{bmatrix} + \epsilon_t$$

- suppose r_t is the Central Bank's policy rate (overnight rate, federal funds rate)
- now u_t depends on interest rate policy
- VAR allows us to estimate the Central Bank's **"policy rule"**
  - how r_t responds to past values of Δy_t, Δp_t, etc. and today's unexpected shocks

Vector Autoregressions

```
var ffr unemp growth infl if qtr>=tq(1983q1), lags(1)
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ffr						
ffr L1.	.956098	.0133266	71.74	0.000	.9299783	.9822178
unemp L1.	-.015216	.0231139	-0.66	0.510	-.0605184	.0300864
growth L1.	.3741846	.0622191	6.01	0.000	.2522374	.4961318
infl L1.	-.005795	.0818168	-0.07	0.944	-.1661529	.154563
_cons	-.0402746	.1673516	-0.24	0.810	-.3682777	.2877285

- Signs of coefficients make sense

Vector Autoregressions

Data ends in 2017q2, so forecasting proceeds in analogous manner to scalar case

- **forecast** for 2017q3?
  - use coeff estimates and **actual** data on r, u, Δy, Δp from 2017q2 and 2017q1
- **forecast** for 2017q4?
  - use coeff estimates and **forecasted** data on r, u, Δy, Δp from 2017q3 and 2017q2
- and so on ...