HG ECON 5101

Fourth lecture – 5 Feb. 2014

1. The ARIMA command in Stata

The ARIMA command includes the possibility of covariates. It estimates a model involving both $\{y_t\}$ and a set of covariates $\{x_t\}$, assuming

$y_t - \beta' x_t \sim \text{ARMA}(p,q)$ where β is a vector of parameters, or

$$(1) \qquad y_t - \beta' x_t = \sum_{j=1}^{p} \varphi_j \left(y_{t-1} - \beta' x_{t-1} \right) + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

- This includes the possibility that $\{y_t\}$ and $\{x_t\}$ are *cointegrated* in the sense that (i) both $\{y_t\}$ and $\{x_t\}$ are non-stationary [I(1)], and (ii) that that the linear combination, $\{y_t - \beta' x_t\}$ is stationary [I(0)].
- A potential application could be a situation where some other method (e.g., Søren Johansen's method) has established that $\{y_t\}$ and $\{x_t\}$ are cointegrated, and an ARIMA analysis could be used to explore/confirm the nature of the stationarity of $\{y_t - \beta' x_t\}$.
- (1) is a special case of an ARMAX model with covariates. However, Stata has not implemented a general ARMAX- routine. General ARMAX can be handled by Kalman filtering (the `sspace` –command).

We will look at the Example 4 in the arima-documentation (pdf) in the Stata manual: The data concern two US series

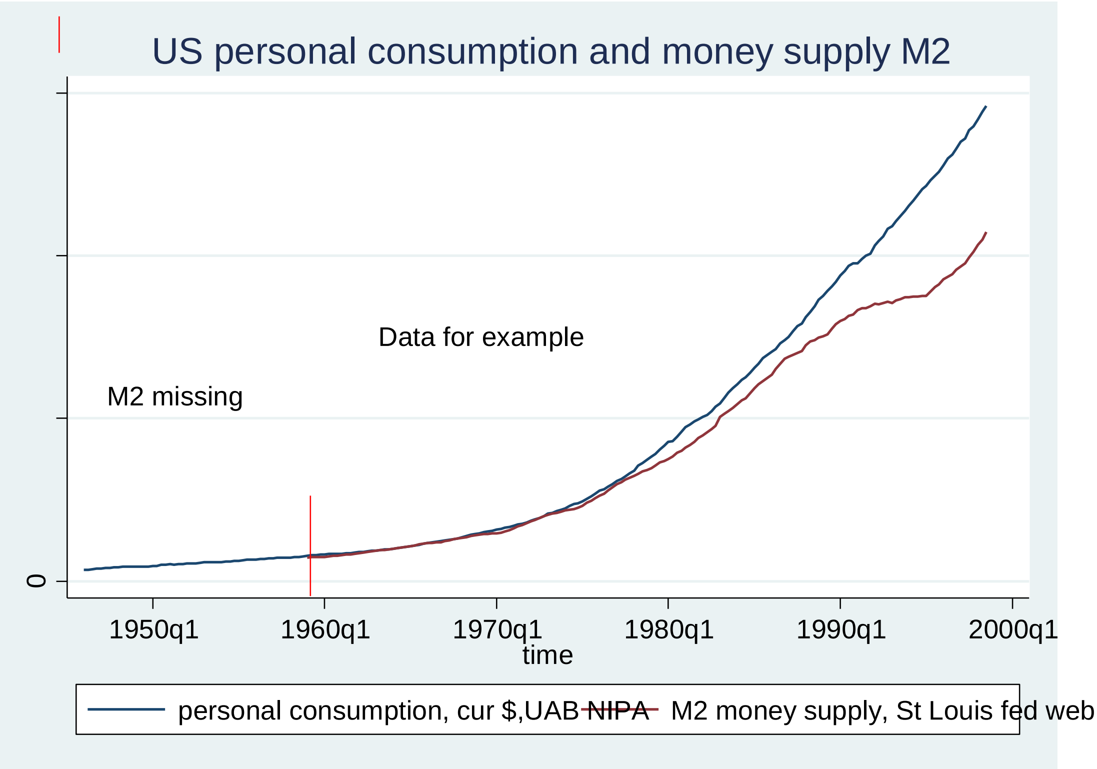
personal-consumption - $\texttt{consump}_t$
money supply - $\texttt{m2}_t$

The analysis is an update inspired by an earlier analysis by Friedman and Meiselman (1963) who postulated the simple relationship,

$$\text{consump}_t = \beta_0 + \beta_1 \text{m2}_t + u_t$$

Load data with

```
use http://www.stata-press.com/data/r13/friedman2, clear
```



Stata cuts the data in 1982q1 because of an intervention by the reserve bank in 1982 to control the inflation. I.e., a structural change in 1982.

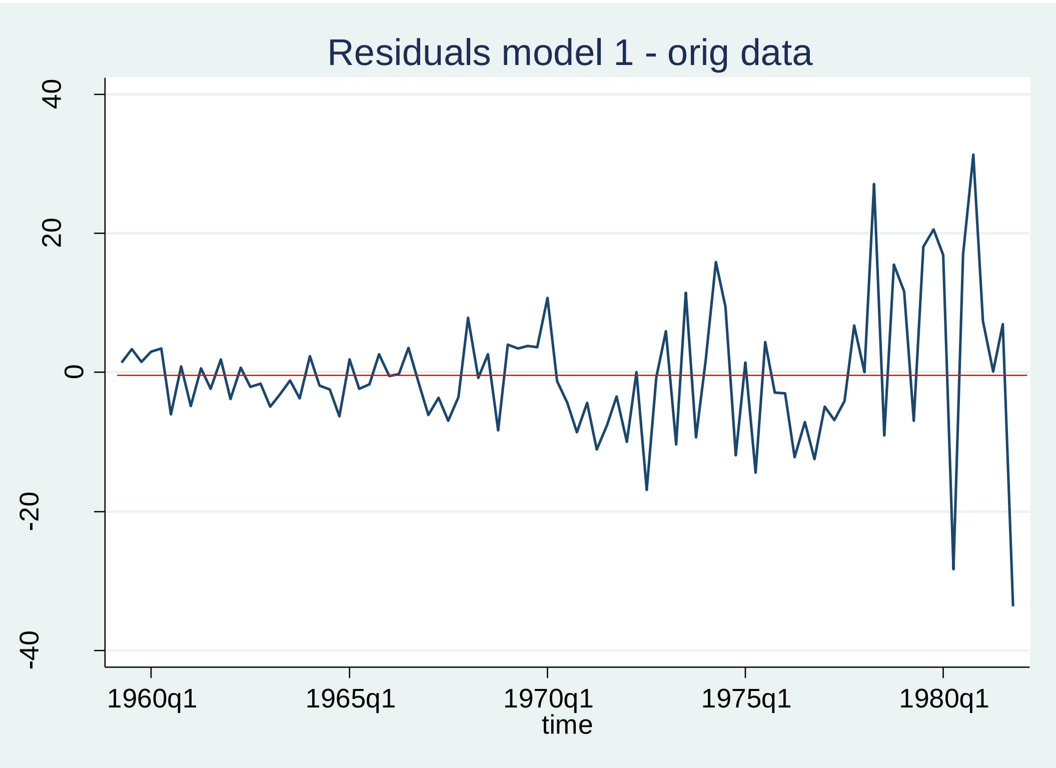
Apparent non-stationary but cointegrated series in period 1959 – 1981.

Stata's solution – postulate ARMA(1,1) for $\{y_t - \beta' x_t\}$:

```
. arima consump m2 if tin(, 1981q4), ar(1) ma(1)


ARIMA regression

Sample:  1959q1 - 1981q4                        Number of obs      =        92
                                                Wald chi2(3)       =   4394.80
Log likelihood = -340.5077                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
     consump |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
consump      |
          m2 |   1.122029   .0363563    30.86   0.000     1.050772    1.193286
       _cons |  -36.09872   56.56703    -0.64   0.523    -146.9681    74.77062
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .9348486   .0411323    22.73   0.000     .8542308    1.015467
             |
          ma |
         L1. |   .3090592   .0885883     3.49   0.000     .1354293    .4826891
-------------+----------------------------------------------------------------
      /sigma |   9.655308   .5635157    17.13   0.000     8.550837    10.75978
------------------------------------------------------------------------------
```

Now, look at the residuals.

```
predict res1,resid

tsline res1 if tin(1959q2,1981q4)
```

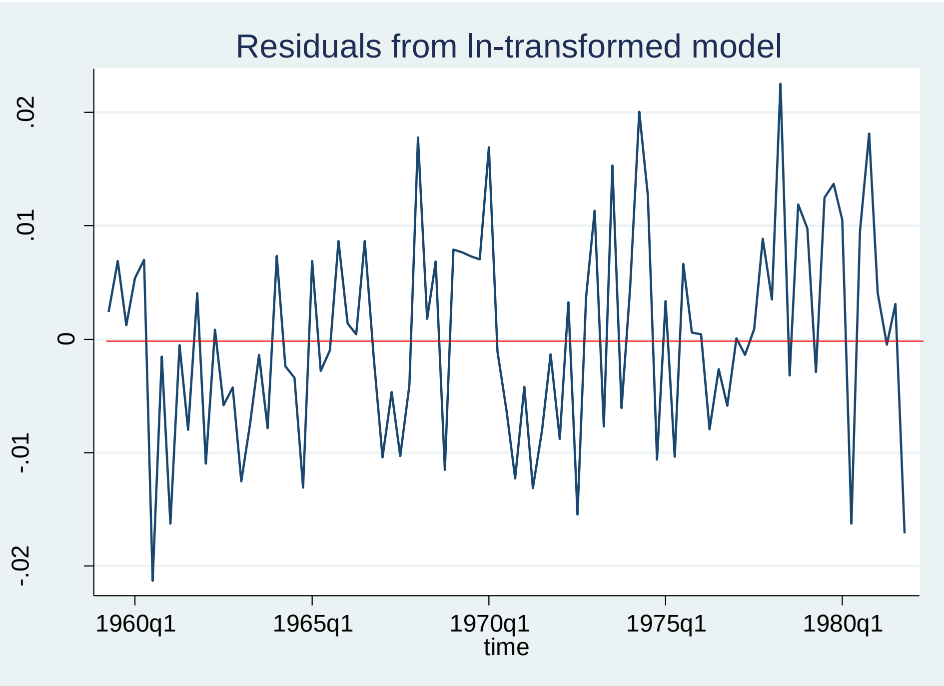


Clearly not white noise (increasing variance). So we try to stabilize the variance by taking the logs:

```
gen lm2=ln(m2)

gen lcons=ln( consump)

arima lcons lm2 if tin(, 1981q4), ar(1) ma(1)
```

```
ARIMA regression

Sample:  1959q1 - 1981q4                        Number of obs      =        92
                                                Wald chi2(3)       =   2600.44
Log likelihood =  299.9357                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
       lcons |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lcons        |
         lm2 |   .9822586   .0607853    16.16   0.000     .8631217    1.101396
       _cons |   .1832399   .4434234     0.41   0.679    -.6858541    1.052334
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .9731574   .0247959    39.25   0.000     .9245583    1.021757
             |
          ma |
         L1. |   .2496818   .1356231     1.84   0.066    -.0161347    .5154983
-------------+----------------------------------------------------------------
      /sigma |   .0091153   .0007558    12.06   0.000      .007634    .0105967
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
confidence interval is truncated at zero.
```

Residuals:



Normality test:

```
swilk resln if tin(1959q2,1981q4)

                   Shapiro-Wilk W test for normal data

    Variable |    Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
       resln |     91    0.99384      0.470    -1.665    0.95207
```

2. Forecasting (prediction).

(A good reference is Hamilton chap. 4)

Suppose we wish to predict a random variable Y from a set of predictor rv's, $X_1, X_2, \ldots, X_k$. A predictor is a function, $\hat{Y} = h(X_1, X_2, \ldots, X_k)$.

The best predictor in the minimum mean squared error (MSE) sense, i.e., minimizing $E\left[\left(Y-\hat{Y}\right)^2\right]$, is

(2) $\hat{Y} = E\left(Y \mid X_1, X_2, \ldots, X_k\right)$

(see Hamilton chap. 4 or a lecture note on prediction on the web-page for Econ 4130 (stat 2) 2013)

In time series the function in (2) often turns out to be complicated , so it is common to use a next best solution, the best linear approximation to (2) (also called the "*linear projection predictor*" from Hilbert space terminology that can be shown to exist always),

(3) $\hat{Y} = a_0 + a_1 X_1 + \cdots + a_k X_k$

where the constants $a_0, a_1, \ldots, a_k$ are determined to minimize the MSE $E\left[\left(Y-\hat{Y}\right)^2\right]$.

Note: If $(Y, X_1, X_2, \ldots, X_k)$ are jointly normal (Gaussian), the (2) and (3) are equivalent !

Solution:

Let *Y* have expectation and variance, μ_Y and σ_Y^2.

Let $\underline{X}' = (X_1, X_2, \ldots, X_k)$ have expectation $\underline{\mu}' = (\mu_1, \mu_2, \ldots, \mu_k)$ and covariance matrix $\Sigma = \left(\mathrm{cov}(X_i, X_j)\right)_{i,j=1}^{k}$.

Write $\underline{a}' = (a_1, a_2, \ldots, a_k)$, and the covariance between *Y* and $\underline{X}$ as the vector $\sigma'_{XY} = \left(\mathrm{cov}(Y, X_1), \ldots, \mathrm{cov}(Y, X_k)\right)$

Theorem 3 The MSE minimizing constants, are given as any solution to **i.** and **ii.** below**:** The solution is obtained by solving the equations

$$E\left(Y - \hat{Y}\right) = 0$$

$$E\left[(Y - \hat{Y})X_j\right] = 0 \text{ for } j = 1, 2, \ldots, k$$

leading to

i. $\mu_Y = a_0 + a_1\mu_1 + \cdots + a_k\mu_k$

ii. $\Sigma\underline{a} = \sigma_{XY} (= \mathrm{cov}(Y, \underline{X}))$

iii. If Σ is non-singular, $\underline{a} = \Sigma^{-1}\sigma_{XY}$.

iv. The MSE of the solution, becomes,

$$MSE = E\left[(Y - \hat{Y})^2\right] = \sigma_Y^2 - \sigma'_{XY}\Sigma^{-1}\sigma_{XY}$$

Proof: Exactly the same proof as for the OLS estimators in a multiple regression problem. Only replace all sample quantities, means and sample covariances, by corresponding population quantities. Or see Hamilton chap. 4. (**End of proof.**)

In a time series context, we have, e.g., observed, $Y_1, Y_2, \ldots, Y_t$ of a **causal stationary** series with expectation $\mu = E(Y_t)$ and autocovariance $\{\gamma(h)\}$, $h = 0, \pm 1, \pm 2, \ldots$.

Let $D_t = \{Y_t, Y_{t-1}, Y_{t-2}, \ldots, Y_1\}$ be all that we know at time *t*.

We want to forecast Y_{t+j} at a future time $t + j$.

Notation: We write the best linear predictor as $\hat{Y}_{t+j|t} = E(Y_{t+j} \mid D_t)$ (if feasible) or the best linear approximation by theorem 3.

With regard to Theorem 3, we have in general $Y = Y_{t+j}$, $\underline{X} = (Y_t, Y_{t-1}, \ldots, Y_0)$, $\underline{a} = (a_t, a_{t-1}, \ldots, a_1)$, and $\Sigma = \left(\gamma(i-j)\right)_{i,j=1}^{t}$ (a $t \times t$ matrix), with solution

(4) $\quad \underline{a} = \Sigma^{-1}\sigma_{t,t+j}$ where $\sigma'_{t,t+j} = (\gamma(j), \gamma(j+1), \ldots, \gamma(j+t-1))$.

- For large *t* (4) becomes unpractical and is replaced by various recursion formulas developed that do not involve matrix inversions (see Hamilton (chap. 4) for some approaches).
- The forecast problem for AR(*p*) is simple. We then only need the *p* values, $Y_t, Y_{t-1}, \ldots, Y_{t-p+1}$ to predict Y_{t+j}.
- The forecast problem for ARMA(*p*,*q*) for $q \geq 1$ is more complicated. We then need the whole history, $Y_t, Y_{t-1}, \ldots, Y_1$ to predict Y_{t+j}.

To simplify the argument below I assume a slightly stronger assumption for the white noise series $\{\varepsilon_t\}$, namely that

(5) $\quad E(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = 0$ for all *t* (in the causal case we look at here).

Note that (5) implies the weaker assumption that the ε_t's are uncorrelated, but uncorrelated ε_t's does not imply (5) except in the Gaussian case (!).

The (causal) AR(*p*) case:

(6) $\quad Y_{t+j} = \varphi_0 + \varphi_1 Y_{t+j-1} + \cdots + \varphi_p Y_{t+j-p} + \varepsilon_{t+j}$

Let $D_t = \{Y_t, Y_{t-1}, \ldots\}$. From (6) we have

(7)
$$\hat{Y}_{t+j|t} = E(Y_{t+j} \mid D_t) = E\left(\varphi_0 + \varphi_1 Y_{t+j-1} + \cdots + \varphi_p Y_{t+j-p} + \varepsilon_{t+j} \mid D_t\right) =$$
$$= \varphi_0 + \varphi_1 E\left(Y_{t+j-1} \mid D_t\right) + \cdots + \varphi_p E\left(Y_{t+j-p} \mid D_t\right) + E\left(\varepsilon_{t+j} \mid D_t\right)$$

The solution of (6) (for all *j*) is (where $\mu = E(Y_s)$)

(8) $$Y_{t+j} = \mu + \varepsilon_{t+j} + \psi_1 \varepsilon_{t+j-1} + \psi_2 \varepsilon_{t+j-2} + \cdots$$

showing that D_t only depends on $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$. Hence (using (5)) the last term in (7) must be 0, $E\left(\varepsilon_{t+j} \mid D_t\right) = 0$ for $j \geq 1$. Then, from (7)

$$\hat{Y}_{t+j|t} = \varphi_0 + \varphi_1 \hat{Y}_{t+j-1|t} + \cdots + \varphi_p \hat{Y}_{t+j-p|t}$$

In particular we get (noting that $E\left(Y_s \mid D_t\right) = Y_s$ for $s \leq t$):

$$\hat{Y}_{t+1|t} = \varphi_0 + \varphi_1 Y_t + \cdots + \varphi_p Y_{t-p+1}$$
$$\hat{Y}_{t+2|t} = \varphi_0 + \varphi_1 \hat{Y}_{t+1|t} + \varphi_2 Y_t + \cdots + \varphi_p Y_{t-p+2}$$

$$\hat{Y}_{t+j|t} = \varphi_0 + \varphi_1 \hat{Y}_{t+j-1|t} + \cdots + \varphi_p \hat{Y}_{t+j-p|t} \quad \text{for} \quad j > p$$

The (causal) ARMA(*p,q*) case:
Consider the simplest ARMA(0,1) =MA(1) case (see Hamilton chap. 4 for the general case):

(9) $$Y_{t+j} = \varphi_0 + \varepsilon_{t+j} + \theta \varepsilon_{t+j-1}$$

where we assume that the MA error term is invertible (i.e., $|\theta| < 1$).
Here the problem arises for the first forecast,

$$\hat{Y}_{t+1|t} = E(Y_{t+1} \mid D_t) = \varphi_0 + E(\varepsilon_{t+1} \mid D_t) + \theta E(\varepsilon_t \mid D_t)$$

Since in the causal case (as for the AR(p) case), D_t only depends on $\varepsilon_t, \varepsilon_{t-1}, \ldots$, we must have $E(\varepsilon_{t+1} \mid D_t) = 0$, so we get

(10) $$\hat{Y}_{t+1|t} = E(Y_{t+1} \mid D_t) = \varphi_0 + \theta E(\varepsilon_t \mid D_t)$$

Using the invertibility of $\{\varepsilon_t\}$, we have from (9)

$$\varepsilon_t = -\frac{\varphi_0}{1+\theta} + Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} + \cdots + (-\theta)^r Y_{t-r} + \cdots$$

showing that

$$\hat{Y}_{t+1|t} = E(Y_{t+1} \mid D_t) = \varphi_0 + \theta E(\varepsilon_t \mid D_t) =$$
$$= \varphi_0 - \frac{\theta\varphi_0}{1+\theta} + \theta Y_t - \theta^2 Y_{t-1} + \theta^3 Y_{t-2} + \cdots + \theta(-\theta)^r Y_{t-r} + \cdots$$

An approximation to this is then obtained by truncating the *Y*-series after $r = t-1$, giving

$$\hat{Y}_{t+1|t} \approx \varphi_0 - \frac{\theta\varphi_0}{1+\theta} + \theta Y_t - \theta^2 Y_{t-1} + \theta^3 Y_{t-2} + \cdots + \theta(-\theta)^{t-1} Y_1$$

This approximation will obviously improve as *t* increases since $\theta^t \underset{t\to\infty}{\longrightarrow} 0$.
For $j \geq 2$, however, the prediction gets easy since then
$E(\varepsilon_{t+j} \mid D_t) = E(\varepsilon_{t+j-1} \mid D_t) = 0$, implying simply

$$\hat{Y}_{t+j|t} = \varphi_0$$

In the general ARMA(p,q) case we get similarly for $1 \leq j \leq q$

$$\hat{Y}_{t+j|t} = \varphi_0 + \varphi_1 \hat{Y}_{t+j-1|t} + \cdots + \varphi_p \hat{Y}_{t+j-p|t} + \theta_j \hat{\varepsilon}_t + \theta_{j+1} \hat{\varepsilon}_{t-1} + \cdots + \theta_q \hat{\varepsilon}_{t+j-q}$$

where the $\hat{\varepsilon}_s$'s must be predicted by $Y_s, Y_{s-1}, \ldots,$
but for $j > q$, we get as before

$$\hat{Y}_{t+j|t} = \varphi_0 + \varphi_1 \hat{Y}_{t+j-1|t} + \cdots + \varphi_p \hat{Y}_{t+j-p|t}$$

Exact predictors can be derived as well (see Hamilton chap. 4) but the formulas are slightly complicated…

3. The property of "mean reversion" for causal stationary processes.

The mean $\mu = E(Y_t)$ in a stationary time series $\{Y_t\}$ seems to play the role as an *attractor* in the series – in the sense that, if an observation of Y_t is far away from μ, the next observation has a tendency to be closer to μ. This tendency some economists tend to call "*mean reversion*".

In prediction the tendency becomes evident:

The general solution of a causal ARMA time series has the form

$$Y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots$$

and at time point $t + j$ in the future

$$(11) \quad Y_{t+j} = \mu + \varepsilon_{t+j} + \psi_1 \varepsilon_{t+j-1} + \psi_2 \varepsilon_{t+j-2} + \cdots + \psi_j \varepsilon_t + \cdots$$

Using (5), all $E(\varepsilon_{t+j} \mid D_t) = 0$ for $j > 0$. Hence, the forecast becomes

$$(12) \quad \hat{Y}_{t+j|t} = E(Y_{t+j} \mid D_t) = \mu + \psi_j \varepsilon_t + \psi_{j+1} \varepsilon_{t-1} + \cdots$$

Since $\psi_j \to 0$ as $j \to \infty$, we see that the error term will approach 0 as *j* increases (it is not hard to prove this formally using the Hilbert-space concept "mean-square" convergence).

Hence $\hat{Y}_{t+j|t} \to \mu$ as *j* increases showing the attractor property of μ.

As a contrast, it is easy to see that random walks do not share the mean reversion property. Consider the simple RW,

$$Y_t = \mu + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$$

We get

$$Y_{t+j} = \mu + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t + \varepsilon_{t+1} + \cdots + \varepsilon_{t+j} = Y_t + \varepsilon_{t+1} + \cdots + \varepsilon_{t+j}$$

implying

$$\hat{Y}_{t+j|t} = E(Y_{t+j} \mid D_t) = Y_t \text{ for all } j > 0.$$

Hence, no mean reversion in random walks.

Final notes to forecasting.

- See the example of forecasting the Norwegian lnGDP in lecture notes 3 (LN3). There we estimated the model based on a reduced data set leaving the last observations to be predicted. To achieve this we could use the Stata arima post-estimation command `predict`. The stata manual did the same thing for the example we started with (see the pdf-manual).
- However, suppose we have a covariate time series $\{x_t\}$ as in the example, we have used the whole series and want to forecast future values beyond the data. Then the `predict` command cannot be used. We have to use the `forecast` command instead (as illustrated in the exercises for seminar 1). In order to utilize the covariate $\{x_t\}$ in the forecasting, we need to forecast x_{t+j} first (using `forecast`) an use the predicted values of x_{t+j} in the equation as basis for forecasting $\{y_{t+j}\}$, again using `forecast`.
- Suppose we have estimated an Arima(p,1,q) for $\{y_t\}$ which is the same as estimating an Arma(p,q) for $\{\Delta y_t\}$. Having predicted values for $\{\Delta\hat{y}_t\}$, we can calculate predicted values for the original series $\{y_t\}$ by cumulative sums (called "*integration*" in the cointegration literature): $\hat{y}_2 = y_1 + \Delta\hat{y}_2, \quad \hat{y}_3 = y_1 + \Delta\hat{y}_2 + \Delta\hat{y}_3, \ \ldots, \hat{y}_t = y_1 + \Delta\hat{y}_2 + \Delta\hat{y}_3 + \cdots + \Delta\hat{y}_t.$ This is done automatically in Stata by the option `y` for `predict`.

4. Introduction to VAR(p)

(An excellent reference for multivariate time series is H. Lütkepohl, "*New Introduction to Multiple Time Series Analysis*", Springer Verlag, 2005).

I use Lütkepohl's notation in the following.

Let $y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Kt} \end{pmatrix}$ be a *K*-dimensional vector of time series.

DEF. We say that $\{y_t\}$ is a *vector autoregressive time series* of order *p* ($y_t \sim \text{VAR}(p)$) if

(13) $y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$

where $\nu' = (\nu_1, \nu_2, \ldots, \nu_K)$ are constants, $A_1, A_2, \ldots, A_p$ are square $K \times K$ coefficient matrices, and $u_t' = (u_{1t}, u_{2t}, \ldots, u_{Kt})$ is a *K*-dimensional *white noise* vector satisfying,

$$E(u_t) = 0, \quad \text{cov}(u_t) = E\left(u_t u_t'\right) = \Sigma_u, \quad E\left(u_t u_s'\right) = \underline{0} \text{ for } t \neq s$$

(i.e., $\{u_t\}$ is a special case of a covariance stationary vector process).

DEF. We say that $\{y_t\}$ is a *structural vector autoregressive time series* of order p ($y_t \sim \text{SVAR}(p)$) if there is a non-singular $K \times K$ matrix *B* such that

(14) $By_t = \nu^* + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \varepsilon_t$ where $\varepsilon_t \sim WN(0, \Sigma_\varepsilon)$

Multiplying (14) by B^{-1} we get the *reduced form* (13) with
$\nu = B^{-1}\nu^*, \quad A_j = B^{-1}A_j^*$ for $j = 1, \ldots, p$, and $u_t = B^{-1}\varepsilon_t \sim WN(0, \Sigma_u)$ where

$\Sigma_u = B^{-1}\Sigma_\varepsilon (B^{-1})'$. Although important in dynamic modeling, we will not discuss SVAR models further in this part of the course.

The VAR(1) case: We will look at first the VAR(1) model

(15) $y_t = \nu + A_1 y_{t-1} + u_t$

Note. It turns out (see below) that the more general VAR(*p*) can be considered a special case of VAR(1)!

Successive substitution in (15) gives (where I_K is the K-dimensional identity matrix)

(16) $y_t = (I_K + A_1 + \cdots + A_1^{t-1})\nu + A_1^t y_0 + (u_t + A_1 u_{t-1} + \cdots + A_1^{t-1} u_1)$

For this to stabilize to something stationary, we must have that A_1^t must converge to a 0-matrix (written $\underline{0}$, i.e., a $K \times K$ matrix of zeroes). This happens if and only if all eigenvalues of A_1 have modulus strictly less than 1:

> [**Review of eigenvalues and eigenvectors.** Let A be a square $K \times K$ matrix. If *b* is a vector $\neq 0$, and λ a scalar such that $Ab = \lambda b$ (which is the same as $\left(A - \lambda I_K\right)b = 0$), we say that *x* is an *eigenvector* and λ a corresponding *eigenvalue*.
>
> For the equation $\left(A - \lambda I_K\right)b = 0$ to be possible for an $x \neq 0$, the matrix $A - \lambda I_K$ must be singular with determinant = 0, i.e.,

$$\det(A-\lambda I_K)=\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{11}-\lambda \end{vmatrix} = p(\lambda)$$, i.e., a polynomial of order *K*. $p(\lambda)$ must have *K* roots, $\lambda_1, \lambda_2, \ldots, \lambda_K$, making $p(\lambda_i) = 0$, some of which may be complex.

Assume for simplicity that all $\lambda_1, \lambda_2, \ldots, \lambda_K$ are different and let $b_1, b_2, \ldots, b_K$ be the corresponding eigenvectors. Then

(17) $Ab_i = \lambda_i b_i$ for $i = 1, 2, \ldots, K$

Collecting all b_i in a matrix, $B = (b_1, b_2, \ldots, b_K) \sim K \times K$, it can be shown that *B* must be non-singular, and (17) becomes

(18) $AB = (Ab_1, Ab_2, \ldots, Ab_K) = (\lambda_1 b_1, \lambda_2 b_2, \ldots, \lambda_K b_K) = B\Lambda$,

where $\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{pmatrix}$ is a diagonal matrix.

Multiplying (18) by B^{-1} from the right gives

(19) $A = B\Lambda B^{-1}$

Using (19), we get $A^2 = B\Lambda B^{-1} B\Lambda B^{-1} = B\Lambda I_K \Lambda B^{-1} = B\Lambda^2 B^{-1}$, $A^3 = B\Lambda^3 B^{-1}$ … etc, and in general

(20) $A^j = B\Lambda^j B^{-1}$

On the other hand, regular matrix multiplication gives

$$\Lambda^2 = \begin{pmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K^2 \end{pmatrix}, \text{ and in general, } \Lambda^j = \begin{pmatrix} \lambda_1^j & 0 & \cdots & 0 \\ 0 & \lambda_2^j & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K^j \end{pmatrix}$$

So, if and only if all $\left|\lambda_j\right| < 1$, $\Lambda^j \underset{j\to\infty}{\rightarrow} \underline{0}$ (i.e., the 0-matrix).

From (20) we then have

$$\lim_{j\to\infty} A^j = \lim_{j\to\infty} B\Lambda^j B^{-1} = B\left(\lim_{j\to\infty}\Lambda^j\right)B^{-1} = \underline{0}$$

(21**) Basic stability condition for VAR(1)**

In other words, the condition that $A^j \underset{j\to\infty}{\rightarrow} \underline{0}$ is that all eigenvalues of A have modulus strictly less than 1.

Note. If some of the λ_i's are equal, (20) is still valid, but with a slightly more complicated Λ ("Jordan form") **.** Λ^j can still be calculated and the conclusion is the same for stability.

End of review.]

We need another matrix formula

Lemma 1 Let A be $K \times K$ satisfying the stability condition (21). Then
i. $(I_K + A + \cdots + A^{t-1}) = (I_K - A)^{-1}(I_K - A^t)$
ii. The infinite matrix series converges as $t \to \infty$ and is equal to $I_K + A + A^2 + \cdots = \sum_{j=0}^{\infty} A^j = (I_K - A)^{-1}$

[**Proof.** Usual matrix multiplication gives

$$(I_K - A)(I_K + A + \cdots + A^{t-1}) =$$
$$= I_K + A + A^2 + \cdots + A^{t-1}$$
$$- A - A^2 - \cdots - A^{t-1} - A^t = I_K - A^t$$

We must have $(I_K - A)$ is non-singular (i.e., $\det(I_K - A) \neq 0$ since 1 is not an eigenvalue of *A*). Then, multiplying both sides of the equality by $(I_K - A)^{-1}$ gives **i.**

ii.: By definition we have

$$I_K + A + A^2 + \cdots = \sum_{j=0}^{\infty} A^j \overset{Def}{=} \lim_{t\to\infty}(I_K + A + \cdots + A^{t-1}) =$$
$$= \lim_{t\to\infty}(I_K - A)^{-1}(I_K - A^t) = (I_K - A)^{-1}$$

End of proof.]

Under stability (all eigenvalues of A_1 have modulus less than 1) the solution (16) (reproduced)

(16) $y_t = (I_K + A_1 + \cdots + A_1^{t-1})\nu + A_1^t y_0 + (u_t + A_1 u_{t-1} + \cdots + A_1^{t-1} u_1)$

will approach a stationary process as $t \to \infty$:

The first term, $(I_K + A_1 + \cdots + A_1^{t-1})\nu \underset{t\to\infty}{\to} (I_K - A_1)\nu$

The second term $\to 0$

The third term converges to a stationary process (when we imagine that the white noise process $\{u_t\}$ has been going on since $-\infty$:

Theorem 4. If all eigenvalues of A_1 have modulus less than 1, then

i. $z_t = u_t + A_1 u_{t-1} + A_1^2 u_{t-2} + \cdots = \sum_{j=0}^{\infty} A_1^j u_{t-j}$ is a well defined random variable and the time series $\{z_t\}$ is stationary with $E(z_t) = 0$ and autocovariance matrices

$$\Gamma_z(h) = E(z_{t+h} z_t') = \sum_{j=0}^{\infty} A_1^{j+h} \Sigma_u (A_1^j)'$$

ii. The VAR(1) equation, $y_t = \nu + A_1 y_{t-1} + u_t$, has a causal stationary solution

$$y_t = \mu + u_t + A_1 u_{t-1} + A_1^2 u_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} A_1^j u_{t-j}$$

where $\mu = E(y_t) = (I_K - A_1)^{-1} \nu$, and autocovariance matrices

$$\Gamma_y(h) = E(\mathrm{y}_{t+h} - \mu)(y_t - \mu)' = \sum_{j=0}^{\infty} A_1^{j+h} \Sigma_u (A_1^j)'$$

(The proof of this follows from elementary Hilbert space theory combined with the concept of "mean square convergence" between random variables – a convergence concept that is slightly stronger than convergence in probability and/or distribution.)

Note that the autocovariance matrices $\Gamma_y(h)$ are not depending on *t*, implying covariance stationarity of the $\{y_t\}$ series.

So, theorem 4 shows that the stability condition implies the existence of a causal stationary solution for the VAR(1) model.

5. The stability condition for general VAR(p) models.

We then need the following useful lemma.

Notation. For any square matrix *B*, write the determinant, $\det(B) = |B|$.

Lemma 2. The stability condition for VAR(1) (i.e., that all eigenvalues of A_1 have modulus less than 1), is equivalent to the following condition

(22) $|I_K - A_1 z| \neq 0$ for all *z* such that $|z| \leq 1$
where *z* is a scalar variable.

[**Proof.** Write (22) (remember that a constant taken out of a determinant must be raised to the power of *K*)

$$|I_K - A_1 z| = (-z)^K \left| A_1 - \frac{1}{z} I_K \right|$$

(We can exclude the possibility $z = 0$ since $|I_K| = 1 \neq 0$). Hence (22) is equivalent to

(23) $\left| A_1 - \frac{1}{z} I_K \right| \neq 0$ for all $|z| \leq 1$

Assume that all eigenvalues of A_1 have modulus strictly less than 1. Then we must have

$|A_1 - \lambda I_K| \neq 0$ for any λ with $|\lambda| \geq 1$, i.e., for any *z* with $\left|\frac{1}{z}\right| = \frac{1}{|z|} \geq 1$, i.e.,

for any *z* with $|z| \leq 1$, so (23) and therefore (22) must be true.

Assume now that (22) is true. Then (23) is true, implying that

$\left| A_1 - \frac{1}{z} I_K \right| = 0$ can have no solution with $\left|\frac{1}{z}\right| \geq 1$. In other words, the

equation, $|A_1 - \lambda I_K| = 0$ can have no solution with $|\lambda| \geq 1$. So all solutions (i.e., eigenvalues) must fulfill $|\lambda| < 1$ - which is the stability condition.

End of proof.]

Any VAR(p) can be formulated as a VAR(1).:

If y_t is a K-dimensional VAR(p):

(24) $y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$

we can define a Kp -dimensional VAR(1), $Y_t = \underline{\nu} + AY_{t-1} + U_t$ by putting

$$Y_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} \sim Kp \times 1 , \quad \underline{\nu} = \begin{pmatrix} \nu \\ \underline{0} \\ \vdots \\ \underline{0} \end{pmatrix} \sim Kp \times 1$$

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{pmatrix} \sim (Kp \times Kp) , \quad U_t = \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \sim (Kp \times 1)$$

or

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \nu \\ \underline{0} \\ \vdots \\ \underline{0} \end{pmatrix} + \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

The matrix A is sometimes called *the companion matrix* for the VAR(p)

Some manipulation of determinants (as done in Hamilton for the univariate AR(p) case), shows

(25) $\det\left(I_{Kp} - Az\right) = \det\left(I_K - A_1 z - \cdots - A_p z^p\right)$

From Lemma 2 we get **our stability condition for the VAR(p**) in (24)
(26)

The VAR(p), $y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$, is **stable** if and only if
$\det\left(I_K - A_1 z - \cdots - A_p z^p\right) \neq 0$ for all $\lvert z \rvert \leq 1$

The polynomial, $\det\left(I_K - A_1 z - \cdots - A_p z^p\right)$, we may *call the companion polynomial* of the VAR(p) process.

The criterion (26) then says that the VAR(p) process is **stable** if and only if all roots of the companion polynomial , $\det\left(I_K - A_1 z - \cdots - A_p z^p\right)$, are outside the unit circle (as in the univariate case).

Note also that because of Theorem 4 the stability of a VAR(p) implies (via its VAR(1) representation) that it has a causal stationary solution which has a MA(∞) form.

Knowing that $\{y_t\}$ is stationary, the expected value, $\mu = E(y_t)$, is easily found by taking expectations of $y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$ directly, giving

$$\mu = \nu + A_1 \mu + \cdots + A_p \mu \quad \text{or} \quad \left(I_K - A_1 - \cdots - A_p\right)\mu = \nu$$

giving

(27) $$\mu = E(y_t) = \left(I_K - A_1 - \cdots - A_p\right)^{-1} \nu$$

To find the MA(∞) solution for y_t, we may use the VAR(1) and theorem 4. Introduce the $(K \times Kp)$ matrix

$$J = \left(I_K : 0 : \ldots : 0\right)$$

Then $y_t = JY_t$ and from theorem 4 (putting $\underline{\mu} = E(Y_t) = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}$)

$$y_t = JY_t = J\underline{\mu} + J\sum_{j=0}^{\infty} A^j U_{t-j} = \mu + J\sum_{j=0}^{\infty} A^j U_{t-j}$$

Example 1 (taken from Lütkepohl):

Consider the bivariate VAR(2) model

$$y_t = \nu + \begin{pmatrix} .5 & .1 \\ .4 & .5 \end{pmatrix} y_{t-1} + \begin{pmatrix} 0 & 0 \\ .25 & 0 \end{pmatrix} y_{t-2} + u_t$$

Is this a stable (and a causal stationary) process?

The companion polynomial becomes

$$\det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} z^2 \right) = \det\begin{pmatrix} 1-.5z & -.1z \\ -.4z-.25z^2 & 1-.5z \end{pmatrix} =$$
$$= 1 - z + .21z^2 - .025z^3$$

with roots, $z_1 = 1.3, \quad z_2 = 3.55 + 4.26i, \quad z_3 = 3.55 - 4.26i$

Since $|z_2| = |z_3| = \sqrt{3.55^2 + 4.26^2} = 5.545$,

we see that all roots are outside the unit circle – so the process is stable and a causal stationary solution exists. (**End of example**.)

(A little bit on forecasting and estimation comes next lecture.)