ApEc 8213: Econometric Analysis III -- Lecture #10

Time Series Econometrics, Part 3
Hansen, Chapter 14, Sections 14.19 – 14.29

I. Introduction/Linear Models (14.19)

This lecture presents the most commonly used linear time series models. Lecture 11 explains how to estimate them.

We saw in the last lecture that there are **two ways of representing any** non-deterministic (i.e. $\mu_t = \mu$ for all t) covariance stationary times series variables/processes:

1. Projection representation: $Y_t = \mu + \sum_{j=0}^{\infty} b_j e_{t-j}$

2. Autoregressive representation: $Y_t = \mu + \sum_{j=1}^{\infty} a_j Y_{t-j} + e_t$

In practice, we cannot use models where j goes to ∞, but instead **we approximate Y_t by setting j to a "reasonable" number**, which could be as small as 1 but may be larger. As we will see, **these two representations have different properties**, and it is **very important to have a method**, based on the data and perhaps on theory, **to decide which representation best fits the data**.

The **projection representation** is a **moving average process**, while an **autoregressive representation** is a **autoregressive process**.

II. Moving Average Processes (14.20, 14.21)

The simplest moving average process is a **first-order moving average process**, which is denoted by **MA(1)**:

$$Y_t = \mu + e_t + \theta e_{t-1}$$

where ***e_t* is a strictly stationary and ergodic white noise process** with variance σ^2. It is called "moving average" because Y_t is a weighted average of the "shocks" (e_t and e_{t-1}).

The MA(1) has the following "moments":

$$\mathrm{E}[Y_t] = \mu$$

$$\mathrm{Var}(Y_t) = (1 + \theta^2)\sigma^2$$

$$\gamma(1) = \mathrm{Cov}(Y_t, Y_{t-1}) = \theta\sigma^2$$

$$\rho(1) = \frac{\mathrm{Cov}(Y_t, Y_{t-1})}{\mathrm{Var}(Y_t)} = \frac{\theta}{1+\theta^2}$$

$$\gamma(k) = \rho(k) = 0, \text{ for } k \geq 2$$

In words, and MA(1) process allows for **correlation of *Y_t* over time**, but **only for observations one time period apart**. There is no correlation over time for observations 2 or more observations apart. If ***θ* > 0** the **correlation is positive**, and if ***θ* < 0** the correlation is **negative**.

MA(q) Process (qth – order moving average process)

The MA(1) can be generalized to q lagged e_t terms:

$$Y_t = \mu + \theta_0 e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

with $\theta_0 = 1$. An MA(q) process has the following moments:

$$\mathrm{E}[Y_t] = \mu$$

$$\mathrm{Var}(Y_t) = (\textstyle\sum_{j=0}^{q} \theta_j^2)\sigma^2$$

$$\gamma(k) = \mathrm{Cov}(Y_t, Y_{t-k}) = (\textstyle\sum_{j=0}^{q-k} \theta_{j+k}\theta_j)\sigma^2, \text{ for } k \leq q$$

$$\rho(k) = \frac{\mathrm{Cov}(Y_t, Y_{t-k})}{\mathrm{Var}(Y_t)} = \frac{\sum_{j=0}^{q-k} \theta_{j+k}\theta_j}{\sum_{j=0}^{q} \theta_j^2}, \text{ for } k \leq q$$

$$\gamma(k) = \rho(k) = 0, \text{ for } k > q$$

In words, Y_t is **correlated over time with any observation within *q* time periods** (from Y_{t+q} to Y_{t-q}) but is uncorrelated with all observations more than q time periods away from it.

By Theorem 14.5 (see Lecture 8), any MA(q) process Y_t is **strictly stationary** and **ergodic**. You can think of this process is one that "smooths" e_t.

MA(∞) Process

An **infinite-order moving average process**, which is denoted by **MA(∞)** and is also called a **linear process**, is:

$$Y_t = \mu + \sum_{j=0}^{\infty} \theta_j e_{t-j}$$

Recall that e_t is a strictly stationary and ergodic white noise process with variance σ^2. **We also require that**:

$$\sum_{j=0}^{\infty} |\theta_j| < \infty.$$

By Theorem 14.6 (see Lecture 8), Y_t is strictly stationary and ergodic.

The MA(∞) process has the following moments:

$$\mathrm{E}[Y_t] = \mu$$

$$\mathrm{Var}(Y_t) = (\textstyle\sum_{j=0}^{\infty} \theta_j^2)\sigma^2$$

$$\gamma(k) = \mathrm{Cov}(Y_t, Y_{t-k}) = (\textstyle\sum_{j=0}^{\infty} \theta_{j+k}\theta_j)\sigma^2$$

$$\rho(k) = \frac{\mathrm{Cov}(Y_t, Y_{t-k})}{\mathrm{Var}(Y_t)} = \frac{\sum_{j=0}^{\infty} \theta_{j+k}\theta_j}{\sum_{j=0}^{\infty} \theta_j^2}$$

III. First-Order Autoregressive Processes (14.22, 14.23)

We now turn to autoregressive representations of Y_t. The simplest one is the **first-order autoregression process**, which can be denoted by **AR(1)**:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + e_t \qquad (14.25)$$

where e_t is a strictly stationary and ergodic white noise process with variance σ^2. Hansen says: "The AR(1) model is probably the single most important model in econometric time series analysis."

Hansen gives an **example** where ***Y_t* = the number of employed people**. Suppose that $1 - \alpha_1$ of employees lose there job in every time period (so that α_1 is the proportion who keep their jobs). Let u_t be the number of new jobs that open up in the same time period, which has a mean of $E[u_t] = \alpha_0$. Define e_t as $u_t - \alpha_0$, that is, the deviation from the average of u_t. Together, these imply equation (14.25).

Hanson shows two plots of 2 AR(1) processes (page 477).

Under what conditions is an AR(1) process stationary?

One "trick" to this is to **"convert" it to an MA(∞) process**. To start, replace Y_{t-1} with $\alpha_0 + \alpha_1 Y_{t-2} + e_{t-1}$:

$$Y_t = \alpha_0 + \alpha_1(\alpha_0 + \alpha_1 Y_{t-2} + e_{t-1}) + e_t$$

$$= \alpha_0 + \alpha_1\alpha_0 + \alpha_1^2 Y_{t-2} + \alpha_1 e_{t-1} + e_t$$

This is so much fun! Let's do it *t* times:

$$Y_t = \alpha_0(1 + \alpha_1 + \alpha_1^2 + \ldots + \alpha_1^{t-1}) + \alpha_1^t Y_0 + \alpha_1^{t-1} e_1 + \alpha_1^{t-2} e_2 + \ldots + e_t$$

$$= \alpha_0 \sum_{j=0}^{t-1} \alpha_1^j + \alpha_1^t Y_0 + \sum_{j=0}^{t-1} \alpha_1^j e_{t-j} \quad (14.26)$$

We could go back in time to $-\infty$. In this case, will Y_t converge as $t \to \infty$? This depends on the first and second terms since the infinite sum of the e_t terms has a mean of 0. Clearly $\sum_{j=0}^{\infty} \alpha_1{}^j = \infty$ if $\alpha_1 \geq 1$, and this also holds if $\alpha_1 \leq -1$.

A useful result for infinite sums is:

Theorem 14.20. The sum $\sum_{k=0}^{\infty} \beta^k = \frac{1}{1-\beta}$ is **absolutely convergent** if $|\beta| < 1$.

Thus if $|\alpha_1| < 1$ then $\alpha_0 \sum_{j=0}^{t-1} \alpha_1^j$ converges to $\alpha_0/(1 - \alpha_1)$ as $t \to \infty$. Thus:

Theorem 14.21. If $E[|e_t|] < \infty$ and $|\alpha_1| < 1$, then the AR(1) process in (14.25) has the MA(∞) convergent representation:

$$Y_t = \mu + \sum_{j=0}^{\infty} \alpha_1^j e_{t-j} \quad (14.27)$$

where $\mu = \alpha_0/(1 - \alpha_1)$. The AR(1) process Y_t is strictly stationary and ergodic.

The moments of an AR(1) process are easy to work out. To calculate the mean of Y_t, take expectations of both sides of (14.25), noting that stationarity implies $E[Y_t] = E[Y_{t-1}]$:

$$E[Y_t] = E[\alpha_0 + \alpha_1 Y_{t\text{-}1} + e_t]$$

$$= \alpha_0 + \alpha_1 E[Y_{t\text{-}1}]$$

$$E[Y_t] = \alpha_0/(1 - \alpha_1)$$

$$\text{Var}(Y_t) = \text{Var}(\alpha_1 Y_{t\text{-}1} + e_t)$$

$$= \alpha_1^2 \text{Var}(Y_{t\text{-}1}) + \sigma^2$$

$$\text{Var}(Y_t) = \sigma^2/(1 - \alpha_1^2)$$

$$\text{Cov}(Y_t, Y_{t\text{-}k}) = \gamma_k = E[(Y_t - E[Y_t])(Y_{t\text{-}k} - E[Y_{t\text{-}k}])]$$

$$= E[(\mu + \textstyle\sum_{j=0}^{\infty} \alpha_1^j e_{t\text{-}j} - \mu)(\mu + \sum_{j=0}^{\infty} \alpha_1^j e_{t\text{-}j\text{-}k} - \mu)]$$

$$= E[(e_t + \alpha_1 e_{t\text{-}1} + \alpha_1^2 e_{t\text{-}2} + \ldots)(e_{t\text{-}k} + \alpha_1 e_{t\text{-}k\text{-}1} + \alpha_1^2 e_{t\text{-}k\text{-}2} + \ldots)$$

$$= \alpha_1^k \sigma^2 + \alpha_1^{k+2} \sigma^2 + \alpha_1^{k+4} \sigma^2 + \ldots$$

$$= \alpha_1^k \sigma^2/(1 - \alpha_1^2)$$

$$\rho(k) = \text{Cov}(Y_t, Y_{t\text{-}k})/\text{Var}(Y_t)$$

$$= \alpha_1^k$$

In words, the **correlation of Y_t over time for an AR(1) process is a gradual "decay to zero"**. If $\alpha_1 > 0$ the autocorrelations are all positive, while if $\alpha_1 < 0$ the autocorrelations alternative between positive and negative.

AR(1) processes can be expressed using the lag operator:

$$(1 - \alpha_1 \mathrm{L})Y_t = \alpha_0 + e_t \qquad (14.28)$$

This can also be expressed as $\alpha(\mathrm{L})Y_t = \alpha_0 + e_t$, where $\alpha(\mathrm{L}) = (1 - \alpha_1\mathrm{L})$. "Generalizing" this function we can write $\alpha(z) = (1 - \alpha_1 z)$, the **autoregressive polynomial** of Y_t for an AR(1) process.

If you get good at using lag operators, you can do the following (assuming that $(1 - \alpha_1\mathrm{L})^{-1}$ is something "real"):

$$Y_t = (1 - \alpha_1\mathrm{L})^{-1}(1 - \alpha_1\mathrm{L})Y_t = (1 - \alpha_1\mathrm{L})^{-1}(\alpha_0 + e_t) \qquad (14.29)$$

But what is $(1 - \alpha_1\mathrm{L})^{-1}$? Of course, you recall Theorem 14.20 from page 6, which is that for $|x| < 1$:

$$\sum_{j=0}^{\infty} x^j = \frac{1}{1-x} = (1-x)^{-1}$$

Let $x = \alpha_1 z$. Then we have:

$$(1 - \alpha_1 z)^{-1} = \sum_{j=0}^{\infty} \alpha_1^j z^j \qquad (14.30)$$

Finally, replace z with L to get:

$$(1 - \alpha_1\mathrm{L})^{-1} = \sum_{j=0}^{\infty} \alpha_1^j \mathrm{L}^j$$

Substituting this into equation (14.29) yields:

$$Y_t = (1 - \alpha_1 \mathrm{L})^{-1}(\alpha_0 + e_t)$$

$$= \left(\sum_{j=0}^{\infty} \alpha_1^j \mathrm{L}^j\right)(\alpha_0 + e_t)$$

$$= \sum_{j=0}^{\infty} \alpha_1^j (\alpha_0 + e_{t-j})$$

$$= \frac{\alpha_0}{1-\alpha_1} + \sum_{j=0}^{\infty} \alpha_1^j e_{t-j}$$

which is equation (14.27).

This shows an **important concept**, which is that **a polynomial $\alpha(z)$ is invertible if**:

$$(\alpha(z))^{-1} = \sum_{j=0}^{\infty} \alpha_j z^j$$

is absolutely convergent. In particular, the **AR(1)** autoregressive polynomial $\alpha(z) = 1 - \alpha_1 z$ **is invertible if** $|\alpha_1| < 1$ (since $(1 - \alpha_1\mathrm{L})^{-1} = \sum_{j=0}^{\infty} \alpha_1^j L^j$). This is the **same condition** for a **stationary** AR(1) process.

Unit Root Processes and Explosive AR(1) Processes

What happens if $|\alpha_1| \geq 1$? First consider the case $\alpha_1 = 1$. Also set $\alpha_0 = 0$ (**why**?). This is called a **random walk**:

$$Y_t = Y_{t-1} + e_t$$

It is also called a **unit root process**, a **martingale**, or an **integrated process**. If we start at Y_0, then Y_t is:

$$Y_t = Y_0 + \sum_{j=1}^{t} e_j$$

Note that Y_0 does not "shrink to 0" as t increases. This is **not stationary** (**why?**). Also, the autoregressive polynomial $\alpha(z) = 1 - z$ is **not invertible**, which means that Y_t **cannot be written as a convergent function of an infinite history past e_t terms**.

Note that AR(1) processes, if hit by a "shock", will gradually return to their long run means if $|\alpha_1| < 1$. This is not the case for a random walk; it is not "mean-reverting".

Finally, **what if $\alpha_1 > 1$? Recall equation (14.26)**. The $\alpha_0 \sum_{j=0}^{t-1} \alpha_1^j$ term will not converge as t increases, instead it will "explode", since it has exponential growth. This is clearly not stationary. **What if $\alpha_1 < -1$?**

IV. Second-Order Autoregressive Processes (14.24)

The **second-order autoregressive process**, or AR(2), is:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + e_t \qquad (14.31)$$

Hansen gives an example of a very old macroeconomic model that depicts GDP as an AR(2) process (pp.481-82).

If you like lag operators, you can write (14.31) as:

$$Y_t - \alpha_1 \mathrm{L} Y_t - \alpha_2 \mathrm{L}^2 Y_t = \alpha_0 + e_t$$

or as $\alpha(\mathrm{L})Y_t = \alpha_0 + e_t$, where $\alpha(\mathrm{L}) = 1 - \alpha_1\mathrm{L} - \alpha_2\mathrm{L}^2$. We call $\boldsymbol{\alpha(z)} = 1 - \alpha_1 z - \alpha_2 z^2$ the **autoregressive polynomial** of Y_t.

As with the AR(1) model, we would like to know what values of α_1 and α_2 lead to stationary AR(2) models. It is convenient to transform it into a simple VAR process (see Chap. 15). **Define $\widetilde{Y}_t = (Y_t, Y_{t-1})'$, which is stationary if and only if Y_t is stationary**. Equation (14.31) implies:

$$\begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \end{pmatrix} + \begin{pmatrix} \alpha_0 + e_t \\ 0 \end{pmatrix}$$

which can be written as:

$$\widetilde{Y}_t = \boldsymbol{A}\widetilde{Y}_{t-1} + \tilde{e}_t$$

where $\boldsymbol{A} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix}$ and $\tilde{e}_t = (\alpha_0 + e_t, 0)'$. Theorem 15.6 shows that a VAR process is strictly stationary and ergodic if the $\mathrm{E}[\|\tilde{e}_t\|] < \infty$ and all eigenvalues of $\boldsymbol{A}$ are < 1 in absolute value. The eigenvalues (denoted by λ) of *any* square matrix $\boldsymbol{A}$ satisfy $\det(\boldsymbol{A} - \boldsymbol{I}_2\lambda) = 0$:

$$\det(\boldsymbol{A} - \boldsymbol{I}_2\lambda) = \lambda^2 - \lambda\alpha_1 - \alpha_2 = \lambda^2\alpha(1/\lambda) = 0$$

where $\alpha(z) = 1 - \alpha_1 z - \alpha_2 z^2$ is an autoregressive polynomial (so $\alpha(1/\lambda) = 1 - \alpha_1(1/\lambda) - \alpha_2(1/\lambda)^2$).

The eigenvalues λ satisfy $\det(\boldsymbol{A} - \boldsymbol{I}_2\lambda) = 0$ for the two roots of the quadratic equation $\lambda^2 - \lambda\alpha_1 - \alpha_2$, which are:

$$\lambda_j = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2} \qquad (14.34)$$

Note that these values of λ are **real if $\alpha_1^2 + 4\alpha_2 \geq 0$**, and are **complex numbers if $\alpha_1^2 + 4\alpha_2 < 0$**.

By Theorem 15.6, **the AR(2) process is stationary if** the solutions to the quadratic equation satisfy **$|\lambda_1| < 1$ and $|\lambda_2| < 1$**. "Some algebra" shows that **$|\lambda_1| < 1$ and $|\lambda_2| < 1$ if and only if**:

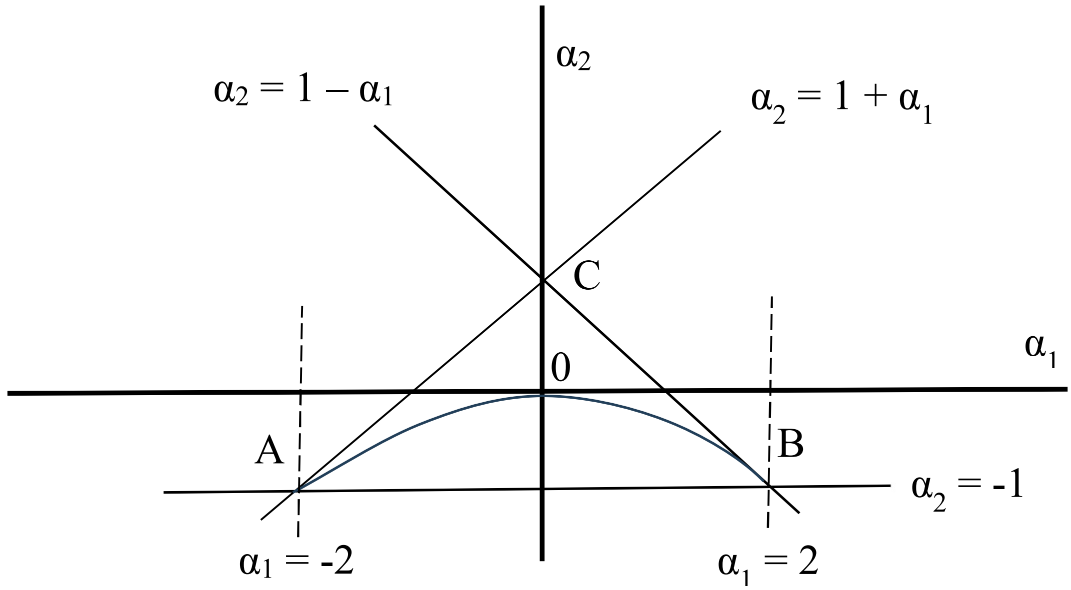
$$\alpha_1 + \alpha_2 < 1 \qquad (14.35)$$

$$\alpha_2 - \alpha_1 < 1 \qquad (14.36)$$

$$\alpha_2 > -1 \qquad (14.37)$$

The following graph shows the values of α_1 and α_2 for which these conditions hold:

Stationarity Region for AR(2) Processes



Stationary processes are those **inside the triangle ABC**. Within the triangle, values **above the curve** is the area were the ***λ*'s are real numbers**, and **below** the curve they are **complex numbers**, which can lead to "interesting" oscillating patterns. Overall, the AR(2) process is much more flexible than the AR(1) process. We conclude with:

Theorem 14.22. If $E[|e_t|] < \infty$ and $|\lambda_j| < 1$ for λ_j defined in equation (14.34), or equivalently the inequalities in equations (14.35)-(14.37) hold, then the AR(2) process in equation (14.31) is absolutely convergent, strictly stationary, and ergodic.

In Section 14.25, Hansen discusses AR(p) processes, where $p > 2$. These are relatively rarely used so this is optional material.

V. Impulse Response Function (14.26)

When Y_t is correlated over time, a "shock" (a particularly large absolute value of e_t) can have effects not only at time t but also in later time periods via the correlation. **The coefficients of the moving average representation can be expressed as**:

$$Y_t = b(\mathrm{L})e_t$$

$$= \sum_{j=0}^{\infty} b_j e_{t-j}$$

$$= b_0 e_t + b_1 e_{t-1} + b_2 e_{t-2} + \ldots$$

This moving average representation is called the **impulse response function (IRF)**. More specifically, the impulse response function is **defined as**:

$$\partial Y_{t+j}/\partial e_t = b_j$$

In words, this is the **impact of a shock at time *t* on** the value of ***Y* at time *t* + *j***. How is this calculated? Assume you have estimated $\alpha_1, \alpha_2, \ldots \alpha_p$ for an AR(p) process.

Method 1 uses a **recursive** method. Consider AR(p):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + e_t$$

Clearly, $\partial Y_t/\partial e_t = b_0 = 1$, and $\partial Y_{t+1}/\partial e_t = \partial Y_t/\partial e_{t\text{-}1} = b_1 =$ $= \partial Y_t/\partial Y_{t\text{-}1}\times\partial Y_{t\text{-}1}/\partial e_{t\text{-}1} = \alpha_1 b_0 = \alpha_1$, and $\partial Y_{t+2}/\partial e_t = \partial Y_t/\partial e_{t\text{-}2}$ $= \alpha_2 + \partial Y_t/\partial Y_{t\text{-}1}\times\partial Y_{t\text{-}1}/\partial e_{t\text{-}2} = b_2 = \alpha_2 + \alpha_1^2 = \alpha_2 b_0 + \alpha_1 b_1$. Continuing this way gives:

$$b_0 = 1$$

$$b_1 = \alpha_1 = \alpha_1 b_0$$

$$b_2 = \alpha_2 + \alpha_1^2 = \alpha_1 b_1 + \alpha_2 b_0$$

$$\vdots$$

$$b_j = \alpha_1 b_{j\text{-}1} + \alpha_2 b_{j\text{-}2} + \ldots \alpha_p b_{j\text{-}p}$$

Question: Do the effects b_j (of e_t on Y_{t+j}) exist when $j > p$?

Method 2 (see Hansen for the derivation). Define $\boldsymbol{A}$ as

$$\boldsymbol{A} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

Then:

$$b_j = S'\boldsymbol{A}^j S \qquad (14.42)$$

where S is a $p\times1$ vector with 1 in first row and 0 elsewhere.

Sometimes the IRF is scaled by σ, where σ^2 = Var(e_t). We can define $\varepsilon_t = e_t/\sigma$, then $\text{IRF}_j = \partial Y_{t+j}/\partial \varepsilon_t = \sigma b_j$.

VI. ARMA and ARIMA Processes (14.27)

Sometimes researchers **combine AR and MA processes**, which can be denoted by ARMA(p, q):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

where again e_t is a strictly stationary and ergodic white noise process. We can **also express this as**:

$$\alpha(\text{L})Y_t = \alpha_0 + \theta(\text{L})e_t$$

Theorem 14.25. The ARMA(p, q) process is strictly stationary and ergodic if (autoregressive) roots of α(L) lie outside the unit circle. (Note: Roots are inverses of the eigenvalues (λ's).) In this case, Y_t can be written as:

$$Y_t = \mu + b(\text{L})e_t$$

where $b_j = O(j^p\beta^j)$ (i.e. it is bounded), and $\sum_{j=0}^{\infty} |b_j| < \infty$.

Finally, suppose Y_t is not, but $Y_t - Y_{t-1}$ is, ARMA(p,q). Then Y_t is an **autoregressive-integrated moving average process**, denoted by ARIMA(p, 1, q). More generally, if $\Delta^d Y_t$ is ARMA(p, q) then Y_t is ARIMA(p, d, q).

In Section 14.28 Hansen briefly discusses whether any Y_t process that can be depicted as a function of an infinite sum of e_t terms will satisfy the mixing conditions. In fact, Andrews (1984) showed that **even a simple AR(1) process may not satisfy the mixing conditions**, by presenting a peculiar e_t that takes only two possible values, so it is a discrete, as opposed to a continuous, variable. Thus a **necessary (but probably not sufficient) condition for an autoregressive process to satisfy mixing is that e_t does not have a discrete distribution**.

VII. Identification (14.29)

Strictly speaking, the parameters of a model are identified if they are uniquely determined by the probability distribution of the observations. For our purposes, we focus on the mean and the second moments (variances and covariances), so we say that the **coefficients of a stationary MA, AR or ARMA** process are **identified if they are uniquely determined by the autocorrelation function**. Given an estimate of $\rho(k)$, are the model coefficients unique?

In general, **MA and ARMA models are not identified**. To see why, consider a simple MA(1) model:

$$Y_t = e_t + \theta e_{t-1}$$

Recall that $\rho(1) = \frac{\theta}{1+\theta^2}$. Next consider a model $Y_t = e_t + \omega e_{t-1}$, and suppose that $\omega = 1/\theta$. For this model we have:

$$\rho(1) = \frac{\omega}{1+\omega^2} = \frac{1/\theta}{1+(1/\theta)^2} = \frac{\theta}{\theta^2+1}$$

Thus we cannot use $\rho(1)$ to distinguish between these two models. Hansen says that the "standard solution" is to choose the model that produces an invertible MA(1) model, and only one model can do this. He does not explain this, but what it means is: Choose the θ that is < 1.

Hansen then discusses the **MA(2) model**:

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

He shows that there is **a similar problem**. That is if θ_1 is replaced by $\omega_1 = 1/\theta_1$ the $\rho(1)$ and $\rho(2)$ autocorrelations are the same. This also happens when θ_2 is replaced by $\omega_2 = 1/\theta_2$ (and θ_1 does not change) and when both θ_1 and θ_2 are replaced by ω_1 and ω_2. Thus in this case there are four different models that have the same $\rho(1)$ and $\rho(2)$.

This problem extends to higher order MA(q) models. Again, he says that the "standard solution" is to choose the one model that has invertible roots.

For **ARMA models**, **identification is even harder**. To see the problem, **consider a simple ARMA(1,1)** model:

$$Y_t = \alpha Y_{t-1} + e_t + \theta e_{t-1}$$

This can be written using lag operators as:

$$(1 - \alpha\text{L})Y_t = (1 + \theta\text{L})e_t$$

If $\alpha = -\theta$ then this model simplifies to $Y_t = e_t$. Another way to see this, without lag operators, is that when $\alpha = -\theta$ we have $Y_t = \theta Y_{t-1} + e_t - \theta e_{t-1}$. But this is equivalent to $Y_t = e_t$, because this simpler model implies that $Y_{t-1} = e_{t-1}$ and thus $\theta Y_{t-1} = \theta e_{t-1}$, and subtracting this from $Y_t = e_t$ gives the original ARMA(1, 1) model. **Thus for the ARMA(1, 1) model there is an infinite set of models** (the set with $\alpha = -\theta$ for any value of α) **that are identical and not identified**. This is called the "cancelling roots" problem.

This **extends to higher order ARMA(p, q) models**. Hansen says that the **"standard solution"** is to **assume that this never happens**. That is assume that $\alpha \neq -\theta$ (and for higher order models assume $\alpha_1 \neq -\theta_1$, $\alpha_2 \neq -\theta_2$, etc.). Hansen also says to "be careful" with ARMA(1, 1) models.

Next, consider **AR(p) models**. Here there is **good news**. They are **always unique and identified**. This works because we can always regress Y_t and its lagged values.

Hansen shows this by defining $X_t = (1, Y_{t-1}, \ldots Y_{t-p})'$ and defining $\alpha = (\alpha_0, \alpha_1, \ldots \alpha_p)'$. Thus the AR(p) model is:

$$Y_t = X_t'\alpha + e_t \qquad (14.45)$$

$E[e_t X_t] = 0$ since e_t is assumed to be white noise for any AR(p) process. Then α can be expressed as:

$$\alpha = (E[X_t X_t'])^{-1} E[X_t Y_t] \qquad (14.46)$$

This is unique, and so α is identified, if $\boldsymbol{Q} = E[X_t X_t']$ is positive definite, which Hansen says is "generally true":

Theorem 14.27. In the AR(p) model given in equation (14.38), if $0 < \sigma^2 < \infty$ then $\boldsymbol{Q} > 0$ and α is unique and identified.

Hansen extends this result to "approximating" AR(p) models. That is, assume that Y_t is a non-deterministic (i.e. $\sigma^2 > 0$) stationary process. Define α as the "best linear predictor", that is the α in equation (14.46). Then e_t is defined by equation (14.45); it still has $E[e_t X_t] = 0$, but it may not be white noise or an MDS. Then the "approximating" AR(p) is still identified:

Theorem 14.28. If Y_t is strictly stationary, not purely deterministic, and $E[Y_t^2] < \infty$, then for any p, $\boldsymbol{Q} = E[X_t X_t']$ > 0 (pos. def.) and thus α in equation (14.46) is identified.

Optional Material: AR(p) Processes (Section 14.25)

While rarely used in practice, one can generalize to **p^th^-order autoregressive processes**, denoted by **AR(p)**:

$$Y_t = \alpha_0 + \alpha_1 Y_{t\text{-}1} + \alpha_2 Y_{t\text{-}2} + \ldots + \alpha_p Y_{t\text{-}p} + e_t \qquad (14.38)$$

where e_t is a strictly stationary and ergodic white noise process. This can be written using lag operators:

$$Y_t - \alpha_1 \mathrm{L} Y_{t\text{-}1} - \alpha_2 \mathrm{L}^2 Y_t - \ldots - \alpha_p \mathrm{L}^p Y_{t\text{-}p} = \alpha_0 + e_t$$

or as $\alpha(\mathrm{L})Y_t = \alpha_0 + e_t$, where:

$$\alpha(\mathrm{L}) = 1 - \alpha_1 \mathrm{L} - \alpha_2 \mathrm{L}^2 - \ldots - \alpha_p L^p \qquad (14.39)$$

We can denote $\boldsymbol{\alpha(z)} = 1 - \alpha_1 \mathrm{L} - \alpha_2 z^2 - \ldots - \alpha_2 z^p$ as the **autoregressive polynomial of $\boldsymbol{Y_t}$**.

The following matrix is analogous to the matrix $\boldsymbol{A}$ for the AR(2) case:

$$
\boldsymbol{A} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \quad (14.40)
$$

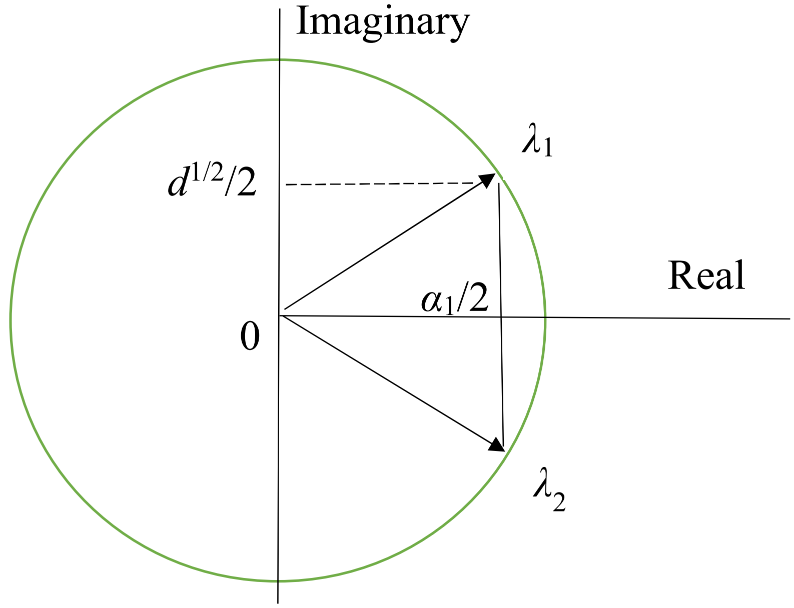
The eigenvalues λ_j of this $\boldsymbol{A}$ are the reciprocals of the roots, r_j, of the autoregressive polynomial in (14.39). Those roots are the solutions for $\alpha(r) = 0$.

By Theorem 15.6, Y_t is stationary if either of the following two equivalent conditions hold:

1. $|\lambda_j| < 1$ for all $j = 1, 2, \ldots p$.

2. All roots r_j of $\alpha(z)$ satisfy $|r_j| > 1$.

For a complex number z, $|z| = 1$ defines the "unit circle" in the complex plane, and any $|z| > 1$ is outside that circle. Let $\lambda_1 = (\alpha_1 + \sqrt{d}\,)/2$ and $\lambda_2 = (\alpha_1 - \sqrt{d}\,)/2$, $d = \alpha_1^2 + 4\alpha_2 < 0$, be the two roots, this shows them on the unit circle:



The **main conclusion** here is:

Theorem 14.23. If $E[e_t] < \infty$ and all roots r_j of $\alpha(z)$ lie outside the unit circle, then the AR(p) process in equation (14.38) is absolutely convergent, strictly stationary and ergodic.

In practice, you can take the estimates of $\alpha_1, \ldots \alpha_p$, solve for the λ_j factors, and check that $|\lambda_j| < 1$, or that $|r_j| > 1$, for all j, where $r_j = \lambda_j^{-1}$.

When all the roots of $\alpha(z)$ lie outside the unit circle then the polynomial $\alpha(z)$ is invertible. Inverting $\alpha(\mathrm{L})Y_t = \alpha_0 + e_t$ allows one to express Y_t in moving average form:

$$Y_t = \mu + b(\mathrm{L})e_t$$

where

$$b(L) = (\alpha(z))^{-1} = \sum_{j=0}^{\infty} b_j z^j \qquad (14.41)$$

and $\mu = \alpha(1)^{-1}\alpha_0$.

This implies that $Y_t = \mu + b_0 e_t + b_1 e_{t-1} + b_2 e_{t-2}$ … Then …

Theorem 14.24. If all the roots of the autoregressive polynomial $\alpha(z)$ lie outside of the unit circle, then (14.41) holds with $|b_j| \le (j + 1)^p \lambda^j$ and $\sum_{j=0}^{\infty} |\lambda_j| < \infty$, where $\lambda = \max_j |r_j^{-1}| < 1$.

In theory, once you solve for all the b_j terms, you can estimate the moving average equation for Y_t with a finite number of lagged errors and check these conditions.