ApEc 8213: Econometric Analysis III -- Lecture #9

Time Series Econometrics, Part 2
Hansen, Chapter 14, Sections 14.10 – 14.18

This lecture presents more theoretical results, along with some more practical/applied concepts. The next two lectures will be much more applied.

I. Martingale Difference Sequences (14.10, 14.11)

Some time series variables/processes are **unforecastable**. For such variables, **the *conditional* expectation** for the variable at time t based on ***any*** **information** at time t-1 or earlier **equals the *unconditional* expectation**. This simply means that none of the past information is useful to determine its current expectation. **An unforecastable process is called a martingale difference sequence (MDS).**

As the name suggests, an MDS is the difference, from time t-1 to time t, of a times series variable/process called a **martingale**. A martingale is the (summed) process:

$$S_t = \sum_{j=1}^{t} e_j$$

where $E[e_t] = 0$ for all t, and $Cov(e_t, e_{t-k}) = 0$ for all $k \geq 1$.

Question: Is this e_t covariance (weakly) stationary?

Going back to **MDS**, it is useful to **be specific about the information set**, which we can **denote by** $\mathscr{F}_t$, that turns out to be useless for forecasting the MDS. The **most common information set** used is the **natural filtration**, which is defined as $\mathscr{F}_t = \sigma(e_t, e_{t-1}, \ldots)$, but other information sets are possible (e.g. including information on other variables). Yet, **unless otherwise indicated, we will assume that the information set is the natural filtration**: $\mathscr{F}_t$ is the σ-field of e_t, that is the set of the **current and all past values of** $\boldsymbol{e_t}$.

We can now define the MDS:

Definition 14.4. The process $(e_t, \mathscr{F}_t)$ is a **Martingale Difference Sequence** if:

1. e_t is adapted to $\mathscr{F}_t$ (which means $E[e_t|\, \mathscr{F}_t] = e_t$)
2. $E[|e_t|] < \infty$ (we will see later that $E[e_t] = 0$)
3. $E[e_t|\, \mathscr{F}_{t-1}] = 0$ **[Write on board]**

Comment: If $\mathscr{F}_t$ is the natural filtration, e_t is adapted to $\mathscr{F}_t$ almost by definition.

Question: Is it possible for $E[e_t] = 0$ if $E[|e_t|] \neq 0$?

This definition means that any e_t that is **an MDS is unforecastable** (at least the mean is unforecastable).

Applying the law of iterated expectations to e_t, we have $E[e_t] = E[E[e_t| \mathscr{F}_{t-1}]] = E[0] = 0$.

Question: What is the difference between e_t being an MDS and e_t being an i.i.d. variable with mean zero?

Answer: An e_t that is i.i.d. with mean zero is an MDS, but an MDS is not necessarily i.i.d with mean zero.

To see why, start with an e_t that is i.i.d. with mean zero. Because it is independently distribution over time, e_t is independent of $\mathscr{F}_{t-1}$, and so $E[e_t| \mathscr{F}_{t-1}] = E[e_t] = 0$. So this e_t is an MDS.

However, an MDS is not necessarily i.i.d with mean zero. We can show this by constructing a **counterexample**. Let the variable u_t be i.i.d. and distributed as N(0, 1). Define:

$$e_t = u_t u_{t-1} \qquad (14.11)$$

By the conditioning theorem (see page 21 of Hansen):

$$E[e_t| \mathscr{F}_{t-1}] = E[u_t u_{t-1}| \mathscr{F}_{t-1}] = u_{t-1}E[u_t| \mathscr{F}_{t-1}] = 0$$

Thus e_t is an MDS. Yet the e_t in (14.11) is not i.i.d. This is shown by calculating the (auto)covariance of e_t^2 (this is ***not*** the correlation of the variances of e_t and e_{t-1}, which are constants and so they have no covariance):

$$\text{Cov}(e_t^2, e_{t-1}^2) = \text{E}[e_t^2 e_{t-1}^2] - \text{E}[e_t^2]\text{E}[e_{t-1}^2]$$

$$= \text{E}[u_t^2 u_{t-1}^4 u_{t-2}^2] - 1$$

$$= \text{E}[u_t^2][\text{E}u_{t-1}^4]\text{E}[u_{t-2}^2] - 1$$

$$= 1\times3\times1 - 1$$

$$= 2$$

Although e_t is not correlated over time, its square (e_t^2) is correlated over time, so this e_t is not i.i.d. (because its square is correlated over time, i.e. not independent).

(Note: Hansen mentions "square integrable" here. This property simply means that the expectation of the square of e_t is $< \infty$.)

An **important property** of an **MDS** is that it is **serially uncorrelated**. This can be shown as follows:

$$\text{Cov}(e_t, e_{t-k}) = \text{E}[e_t\, e_{t-k}]$$

$$= \text{E}[\text{E}[e_t\, e_{t-k} | \mathscr{F}_{t-1}]] \text{ (by the law of iterated expectations)}$$

$$= \text{E}[e_{t-k} \times \text{E}[e_t \mid \mathscr{F}_{t-1}]] \text{ (by the conditioning theorem)}$$

$$\text{E}[e_{t-k} \times 0]$$

$$= 0$$

More generally, an MDS has $\text{Cov}(e_t, e_{t-k}) = 0$ for all $k \neq 0$. This is summarized in the following theorem:

Theorem 14.10. If $(e_t, \mathscr{F}_t)$ is an MDS and $\text{E}[e_t^2] < \infty$, then e_t is serially uncorrelated.

However, it is ***not*** **true that a serially uncorrelated variable/process is an MDS**. This can also be shown by a **counterexample**. Consider $e_t = u_t + u_{t-1}u_{t-2}$, where u_t is i.i.d. and distributed as N(0, 1). **This is not an MDS** because $\text{E}[e_t \mid \mathscr{F}_{t-1}] = u_{t-1}u_{t-2} \neq 0$. But it *is* serially uncorrelated:

$$\text{Cov}(e_t, e_{t-1}) = \text{E}[e_t\, e_{t-1}]$$

$$= \text{E}[(u_t + u_{t-1}u_{t-2})(u_{t-1} + u_{t-2}u_{t-3})]$$

$$= \text{E}[u_t u_{t-1} + (u_{t-1})^2 u_{t-2} + u_t u_{t-2} u_{t-3} + u_{t-1}(u_{t-2})^2 u_{t-3}]$$

$$= \text{E}[u_t]\text{E}[u_{t-1}] + \text{E}[(u_{t-1})^2]\text{E}[u_{t-2}] + \text{E}[u_t]\text{E}[u_{t-2}]\text{E}[u_{t-3}]$$

$$+ \text{E}[u_{t-1}]\text{E}[(u_{t-2})^2]\text{E}[u_{t-3}]$$

$$= 0$$

You can also show that $\text{Cov}(e_t, e_{t-k}) = 0$ for any $k \neq 0$.

Finally and important MDS is the **special case** of a homoscedastic martingale difference sequence:

Definition 14.5. The MDS (e_t, $\mathscr{F}_t$) is a **Homoscedastic Martingale Difference Sequence** if $E[e_t^2 | \mathscr{F}_{t-1}] = \sigma^2$. Strictly speaking this should be called a conditional homoscedastic MDS, because any strictly stationary MDS has a constant variance, $E[e_t^2]$, but only a homoscedastic MDS has a constant ***conditional*** variance: $E[e_t^2 | \mathscr{F}_{t-1}] = \sigma^2$.

Any i.i.d. sequence with a mean of zero is a homoscedastic MDS, and (by definition) every homoscedastic MDS is an MDS. But **not every MDS is (conditionally) homoscedastic**. For example, $e_t = u_t u_{t-1}$ is an MDS (see above), but it is not conditionally homoscedastic, because:

$$E[e_t^2 | \mathscr{F}_{t-1}] = (u_{t-1})^2 E[u_t^2 | \mathscr{F}_{t-1}] = (u_{t-1})^2$$

Also, a **homoscedastic MDS is not necessarily i.i.d.** For example, let $e_t = \sqrt{1-(2/\eta_{t-1})} \times T_t$, where T_t is distributed as a student t distribution with degrees of freedom $\eta_{t-1} = 2 + (e_{t-1})^2$. This has $E[e_t | \mathscr{F}_{t-1}] = 0$ and $E[e_t^2 | \mathscr{F}_{t-1}] = 1$, and thus it is a homoscedastic MDS. However, it is not i.i.d. because the conditional distribution of e_t depends on e_{t-1} via η_{t-1}.

In terms of forecastability, **an i.i.d. time series variable is completely unforecastable**, while the **mean of an MDS process is not forecastable but other moments** (e.g. the variance) **could be forecastable**.

A Central Limit Theorem for Martingale Differences

For some estimation purposes we will want to work out the asymptotic distribution of a sample mean of an MDS. More specifically, we are interested in the following sample mean:

$$S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \qquad (14.12) \quad \textbf{[Write on board]}$$

where u_t, a vector, is an MDS with E[u_t] = 0 and E[$u_t u_t'$] = $\boldsymbol{\Sigma} < \infty$. Here is the theorem:

Theorem 14.11. MDS Central Limit Theorem. If u_t is a strictly stationary and ergodic martingale difference sequence and E[$u_t u_t'$] = $\boldsymbol{\Sigma} < \infty$, then as $n \rightarrow \infty$,

$$S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \underset{d}{\rightarrow} \text{N}(0, \boldsymbol{\Sigma})$$

II. Mixing & CLT for Correlated Observations (14.12, 14.13)

The Central Limit Theorem in Theorem 14.11 is for **MDS variables**, but these are **not correlated** (in terms of their mean) **over time**. So **we still need a CLT for variables that are correlated over time**. It turns out that this is **hard to do for ergodic variables**, but it is possible if we "tighten up" the correlation of such variables over time. This leads to the (rather abstract) concept of **mixing**.

Consider a variable Y_t for which there are (at least) two events, A and B, that can happen at any point in time. We want a **concept that measures the correlation** (or lack of it) **between these two events at different points in time** for Y_t. The following **measure of dependence** is very useful:

$$\alpha(A, B) = |\text{Prob}[A \cap B] - \text{Prob}[A]\times\text{Prob}[B]|$$

This **equals 0 when A and B are independent**; otherwise it is > 0. Note: **$\alpha(A, B)$ is called the discrepancy** of A and B.

Question: Suppose that A and B are events that are "negatively correlated". Shouldn't this lead to a situation where $\alpha(A, B)$ is < 0?

Intuition: A = sunny morning, B = cloudy afternoon
Let Prob[A] = Prob[B] = 0.5 and Prob[$A \cap B$] = 0.10.

Next, consider **two events that are separated in time by ℓ time periods**. More specifically, let event A be an event that takes place ℓ or more time periods in the past, while B is an event that takes place in the present or the future:

$$A \in \mathcal{F}_{-\infty}^{t-\ell} = \sigma(\ldots, Y_{t-\ell-1}, Y_{t-\ell})$$

$$B \in \mathcal{F}_{t}^{\infty} = \sigma(Y_t, Y_{t+1}, \ldots,)$$

We can measure discrepancy, $\alpha(A, B)$, for any pair of events from these two sets. **To measure dependence**

over time for these two sets of Y_t that are ℓ time periods apart, we **find the largest discrepancy**, $\alpha(A, B)$, **for any pair of events** from these two sets, which we denote by $\alpha(\ell)$:

$$\alpha(\ell) = \sup_{A \in \mathcal{F}_{-\infty}^{t-\ell},\ B \in \mathcal{F}_{t}^{\infty}} \alpha(A, B)$$

The closer this is to 0, the less dependence over time for Y_t.

The constants $\boldsymbol{\alpha(\ell)}$ are called **strong mixing coefficients**. $\boldsymbol{Y_t}$ **has strong mixing if** $\boldsymbol{\alpha(\ell) \to 0}$ **as** $\boldsymbol{\ell \to \infty}$. That is, as the gap in time between A and B increases, the degree of dependence falls, and eventually reaches independence as $\ell \to \infty$. Note than **any** $\boldsymbol{Y_t}$ **that has strong mixing is also ergodic**.

Hansen briefly describes another type of mixing, $\boldsymbol{\beta}$**-mixing**, in Section 14.12. This is optional.

As you might expect, the **mixing property is preserved for transformations** of variables:

Theorem 14.12. If Y_t has **strong** mixing coefficients $\alpha_Y(\ell)$ and $X_t = \phi(Y_t, Y_{t\text{-}1}, Y_{t\text{-}2}, \ldots, Y_{t\text{-}q})$, then X_t has mixing coefficients $\alpha_X(\ell) \leq \alpha_Y(\ell - q)$, for $\ell \geq q$. The coefficients

$\alpha_X(\ell)$ have the same summation and rate properties as $\alpha_Y(\ell)$ (see bottom of page 468 of Hansen).

Hansen says that mixing is a "useful tool" because of the following inequalities. Note that they all refer to limits on correlation or on autocorrelation over time.

Theorem 14.13. Let $\mathcal{F}_{-\infty}^{t}$ and $\mathcal{F}_{t}^{\infty}$ (information sets) be constructed from the random pair (X_t, Z_t), where, for the definition of $\alpha(\ell)$ on p.9, $\mathcal{F}_{-\infty}^{t-\ell} = \sigma(\ \dots X_{t-\ell-1}, Z_{t-\ell-1}, X_{t-\ell}, Z_{t-\ell})$ and $\mathcal{F}_{t}^{\infty} = \sigma(X_t, Z_t, X_{t+1}, Z_{t+1}\ \dots)$.

1. If $|X_t| \leq C_1$ and $|Z_t| \leq C_2$, then

$$|\text{Cov}(X_{t-\ell}, Z_t)| \leq 4C_1C_2\alpha(\ell).$$

2. If $\text{E}[|X_t|^r] < \infty$ and $\text{E}[|Z_t|^q] < \infty$ for $1/r + 1/q < 1$, then

$$\text{Cov}(X_{t-\ell}, Z_t)| \leq 8(\text{E}[|X_t|^r])^{1/r}(\text{E}[|Z_t|^q])^{1/q}\alpha(\ell)^{1 - 1/r - 1/q}.$$

3. If $\text{E}[Z_t] = 0$ and $\text{E}[|Z_t|^r] < \infty$ for $r \geq 1$, then

$$\text{E}[|\text{E}[Z_t|\mathcal{F}_{-\infty}^{t-\ell}]|] \leq 6(\text{E}[|Z_t|^r])^{1/r}\alpha(\ell)^{1 - 1/r}$$

A final theorem is:

Theorem 14.14. If Y_t is i.i.d. then it is strong mixing and ergodic.

Central Limit Theorem for Correlated Observations

At last, consider a CLT for a normalized mean S_n as defined in equation (14.12) on page 7 above that allows the variables u_t to be serially correlated.

To start, **when u_t is a single variable**, the variance of S_n can be written as follows:

$$\mathrm{Var}(S_n) = \sigma^2 + 2\sum_{\ell=1}^{n}\left(1 - \frac{\ell}{n}\right)\gamma(\ell)$$

where $\sigma^2 = \mathrm{Var}(u_t)$ and $\gamma(\ell) = \mathrm{Cov}(u_t, u_{t-\ell})$. Since $\gamma(\ell) = \gamma(-\ell)$, this can be written as:

$$\mathrm{Var}(S_n) = \sum_{\ell=-n}^{n}\left(1 - \frac{|\ell|}{n}\right)\gamma(\ell) \qquad (14.14)$$

When u_t is a vector, define the (co)variance matrix as:

$$\boldsymbol{\Sigma} = \mathrm{E}[u_t u_t']$$

and the serial covariance matrix as:

$$\boldsymbol{\Gamma}(\ell) = \mathrm{E}[u_t u_{t-\ell}'].$$

Note that $\boldsymbol{\Gamma}(\ell) = \boldsymbol{\Gamma}(-\ell)'$. $\mathrm{Var}(S_n)$ for the vector case is then:

$$\text{Var}(S_n) = \mathbf{\Sigma} + \sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)(\mathbf{\Gamma}(\ell) + \mathbf{\Gamma}(\ell)')$$

$$= \sum_{\ell=-n}^{n}\left(1-\frac{|\ell|}{n}\right)\mathbf{\Gamma}(\ell) \qquad (14.15)$$

Question: What happened to $\mathbf{\Sigma}$?

A **necessary** (but not sufficient) **condition for S_n to converge to a normal distribution** is that the variance in (14.15) converges to a limit. **This variance will converge if the following sum converges** as $n \to \infty$:

$$\sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)\mathbf{\Gamma}(\ell) = \frac{1}{n}\sum_{\ell=1}^{n-1}\sum_{j=1}^{\ell}\mathbf{\Gamma}(j) \to \sum_{\ell=0}^{\infty}\mathbf{\Gamma}(\ell) \quad (14.16)$$

A **necessary condition** for equation (14.16) to hold is that the **covariances $\mathbf{\Gamma}(\ell)$ fall to 0** as $\ell \to \infty$. A **sufficient condition** is that the **covariances are "absolutely summable"**, which can be checked using a "mixing inequality" (see middle of page 470 of Hansen).

The **bottom line** is that equation (14.15) converges if $\text{E}[||u_t||^r] < \infty$ and $\sum_{\ell=0}^{\infty}\alpha(\ell)^{1-2/r} < \infty$, for some $r > 2$.

Under these assumptions we have:

$$\text{Var}(S_n) \to \sum_{\ell=-\infty}^{\infty}\mathbf{\Gamma}(\ell) \equiv \mathbf{\Omega} \qquad (14.17)$$

The matrix $\boldsymbol{\Omega}$ is often called the **long-run variance of u_t.**

This can be **summarized** in the following CLT theorem:

Theorem 14.15. If u_t is strictly stationary with mixing coefficients $\alpha(\ell)$, E[u_t] = 0, and for some $r > 2$ both E[$\|u_t\|^r$] < ∞ and $\sum_{\ell=0}^{\infty} \alpha(\ell)^{1-2/r} < \infty$, then equation (14.17) is convergent and

$$S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \underset{d}{\rightarrow} \mathrm{N}(0, \boldsymbol{\Omega})$$

Some comments about this:

1. This requires at least $r > 2$ finite moments, which is stronger than the requirement for the MDS CLT (which required only that the second moment, that is the variance, was finite).

2. The summability condition on the mixing coefficients is "considerably stronger" than ergodicity.

3. There is a trade-off regarding the choice of r. A larger r requires more moments to be finite, but a smaller r requires faster "decay" in the mixing coefficients.

From now on, the material will be less theoretical.

III. Linear Projection and White Noise (14.14, 14.15)

You have learned about linear projections last semester, where some variable Y was projected onto a vector of X variables. For **time series** analysis **one often projects Y onto its lagged values**. In theory, Y can be projected onto an infinite number of lagged values. **We assume throughout that Y is weakly (covariance) stationary**. Recall that, when Y and the X variables have bounded variances, the linear projection of Y on the X variables is the β that minimizes the sum of the squared residuals, i.e. it minimizes $S(\beta) = \mathrm{E}[(Y - \beta'X)^2]$. This solution for this β is $(\mathrm{E}[X'X])^{-1}\mathrm{E}[X'Y]$, and the **projection of Y on X, denoted by $\mathscr{P}[Y \mid X]$**, is the prediction based on this solution:

$$\mathscr{P}[Y \mid X] = X'(\mathrm{E}[X'X])^{-1}\mathrm{E}[X'Y]$$

where the unique **projection error** is $e = Y - \mathscr{P}[Y \mid X]$.

In theory, we can **project Y onto an infinite number of its lagged values**, which is denoted by $\mathscr{P}_{t-1}[Y_t] = \mathscr{P}[Y_t \mid \tilde{Y}_{t-1}]$, where $\tilde{Y}_t = (\ldots\ Y_{t-2}, Y_{t-1})$. This projection is unique, and it has a **unique projection error**:

$$e_t = Y_t - \mathscr{P}_{t-1}[Y_t] \qquad (14.18)$$

This projection error has a **mean of 0** and a **finite variance**:

$$\sigma^2 = \mathrm{E}[e_t^2] \leq \mathrm{E}[Y_t^2] < \infty$$

It is **also serially uncorrelated**, even if Y_t is serially correlated. Also, by Theorem 14.2, if Y_t is strictly stationary then $\mathscr{P}_{t-1}[Y_t]$ and e_t are strictly stationary. More formally:

Theorem 14.16. If the variable Y_t is weakly (covariance) stationary, it has projection equation

$$Y_t = \mathscr{P}_{t-1}[Y_t] + e_t.$$

The projection error e_t satisfies

$$\mathrm{E}[e_t] = 0$$

$$\mathrm{E}[e_{t-j}\, e_t] = 0 \quad \text{for } j \geq 1$$

$$\sigma^2 = \mathrm{E}[e_t^2] \leq \mathrm{E}[Y_t^2] < \infty \qquad (14.19)$$

Thus, e_t is also weakly stationary. Finally, if Y_t is strictly stationary then e_t is strictly stationary.

Note: Unlike e_t, Y_t could be (auto)correlated over time.

White Noise

The projection error has a mean of 0, a finite variance, and is serially uncorrelated. A variable with these properties is called a "white noise" process.

Definition 14.6. The variable/process e_t is **white noise** if $E[e_t] = 0$, $E[e_t^2] = \sigma^2 < \infty$, and $Cov(e_t\, e_{t-k}) = 0$ for $k \neq 0$.

An MDS is white noise (Theorem 14.10), but the reverse is not the case, as shown by the example $e_t = u_t + u_{t-1}u_{t-2}$. The **relationship between these variables/processes is**:

i.i.d. with $E[e_t] = 0 \subset$ homoscedastic MDS $\subset$ MDS $\subset$ white noise

Note that **the *conditional* variance of an MDS**, **and of a white noise process can, be heteroscedastic**.

IV. The Wold Decomposition (14.16, 14.17 and 14.18)

It Y_t is weakly stationary it has a useful linear representation:

Theorem 14.17: The Wold Decomposition. If Y_t is weakly (covariance) stationary and $\sigma^2 > 0$, where σ^2 is the projection error variance in equation (14.19), then Y_t has the following linear representation:

$$Y_t = \mu_t + \sum_{j=0}^{\infty} b_j e_{t-j} \qquad (14.20)$$

where e_t are the white noise projection errors in (14.18), $b_0 = 1$, and

$$\sum_{j=0}^{\infty} b_j^2 < \infty \qquad (14.21)$$

and

$$\mu_t = \lim_{m\to\infty} \mathscr{P}_{t-m}[Y_t] \qquad (14.22)$$

The infinite sum in equation (14.20) is called a **linear process**. Hansen says that the Wold Decomposition is "a foundational result for linear time series analysis." In particular, since any weakly stationary process can be written in this way, this justifies linear models as approximations (like a Taylor's expansion of some function).

The series μ_t is the part of Y_t that is "perfectly predictable" from past values, and it is called the **deterministic component**. **In most cases, $\mu_t = \mu$ (a constant)**, and then Y_t is called a **non-deterministic** stationary time series. In this case, the **Wold Decomposition simplifies to**:

Theorem 14.18. If Y_t is weakly (covariance) stationary and non-deterministic, then is has the following linear representation:

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

where b_j satisfies equation (14.21) and e_t is a set a white noise projection errors as in equation (14.18).

Lag Operator

A **convenient notation** for time series models is the lag operator, which is defined as:

Definition 14.7. The **lag operator** L satisfies:

$$\mathrm{L}Y_t = Y_{t-1}$$

The lag operator is multiplicative in the sense that:

$$\mathrm{L}^2 Y_t = \mathrm{L}\mathrm{L}Y_t = \mathrm{L}Y_{t-1} = Y_{t-2}$$

More generally:

$$\mathrm{L}^k Y_t = Y_{t-k}$$

A useful "function" is the following notation:

$$b(\mathrm{L}) = b_0 + b_1\mathrm{L} + b_2\mathrm{L}^2 + b_3\mathrm{L}^3 + \ldots$$

This allows us to write the Wold decomposition as:

$$Y_t = \mu + b_0 e_t + b_1 \mathrm{L} e_t + b_2 \mathrm{L}^2 e_t + \ldots$$

$$= \mu + (b_0 + b_1\mathrm{L} + b_2\mathrm{L}^2 + \ldots)e_t$$

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

Note: This notation is a special case of the following **general notation** $\boldsymbol{b(z)} = b_0 + b_1z + b_2z^2 + \ldots$ which concerns some variable z that may have nothing to do with lagged variables.

The Autoregressive Wold Decomposition

The projection error e_t is a function of the current and past values of Y_t. The Wold Decomposition allows one to express Y_t as a function of current and past values of e_t, so one can substitute out the lagged e_t terms with past values of Y_t, and thus **express $\boldsymbol{Y_t}$ as a function of past values of $\boldsymbol{Y_t}$**. The assumptions necessary for this are given by:

Theorem 14.19. If Y_t is covariance stationary, non-deterministic, with Wold representation $Y_t = b(\mathrm{L})e_t$, such that $|b(z)| \geq \delta > 0$ for all complex $|z| \leq 1$, and for some integer $s \geq 0$ the Wold coefficients satisfy the condition $\sum_{j=0}^{\infty}(\sum_{k=0}^{\infty} k^s b_{j+k})^2 < \infty$, then Y_t has the representation:

$$Y_t = \mu + \sum_{j=1}^{\infty} a_j Y_{t-j} + e_t \qquad (14.23)$$

for some coefficients μ and a_j. These coefficients satisfy $\sum_{k=0}^{\infty} k^s |a_k| < \infty$, so (14.23) is convergent.

Equation (14.23) is called an **infinite-order autoregressive representation** with autoregressive coefficients a_j.

So what is $\boldsymbol{b(z)}$? It is a polynomial in z, and thus a solution to the equation $b(z) = 0$ is a **root** of $b(z)$. The assumption $|b(z)| > 0$ for $|z| \leq 1$ means that the roots of $b(z)$ lie *outside* the unit circle $|z| = 1$ (the circle can be a complex plane with radius one).

Theorem 14.19 makes the stronger restriction that **$\boldsymbol{|b(z)|}$ is bounded away from 0** for z on or within the unit circle. This **excludes** the possibility of an **infinite number of roots** outside of but **arbitrarily close to the unit circle**. Finally, the summability assumption for the b coefficients on the Wold Decomposition ensures the convergence of the α_j autoregressive coefficients.

To understand the restriction on the roots of $b(z)$, consider the **simple case of $\boldsymbol{b(z) = 1 - b_1 z}$** (we will see in the next lecture that this is a first-order moving average model). The requirement that $|b(z)| \geq \delta$ for $|z| \leq 1$ means that $|b_1| \leq 1 - \delta$ (see footnote on p.474 of Hansen). Thus $|b_1|$ must be < 1. For the infinite polynomial case it requires $\sup_j |b_j| < 1$.