ApEc 8213: Econometric Analysis III -- Lecture #8

Time Series Econometrics, Part 1
Hansen, Chapter 14, Sections 14.1 – 14.9

I. Introduction and Examples (14.1, 14.2)

A variable Y_t is a **time series** process if its observations are ordered in time. These lectures will focus on a single variable, a "univariate" model. There are also "multivariate" models (e.g. VAR models), which are covered extensively in Chapter 15 of Hansen's book.

Some basic notation:

1. For any given sample, the overall sample size is denoted by n (other authors use T), so the subscripts for the sample are $t = 1, 2, \ldots n$.

2. Hansen defines the **frequency** of a time series as the number of observations per year, which is denoted by s. Thus for quarterly data $s = 4$, and for monthly data $s = 12$.

Perhaps the **most important difference** is that Y_t usually is **not independent over time**. For example, Y_t is usually correlated with Y_{t-1}. Thus **asymptotic theory needs to be modified** to account for this.

Hansen provides some examples, including graphs, in Section 14.2, focusing on U.S. macroeconomic data. In many cases it is quite clear that Y_t is closely related to Y_{t-1}.

Here is a graph of U.S. money supply (M1) from a time series book by Walter Enders:
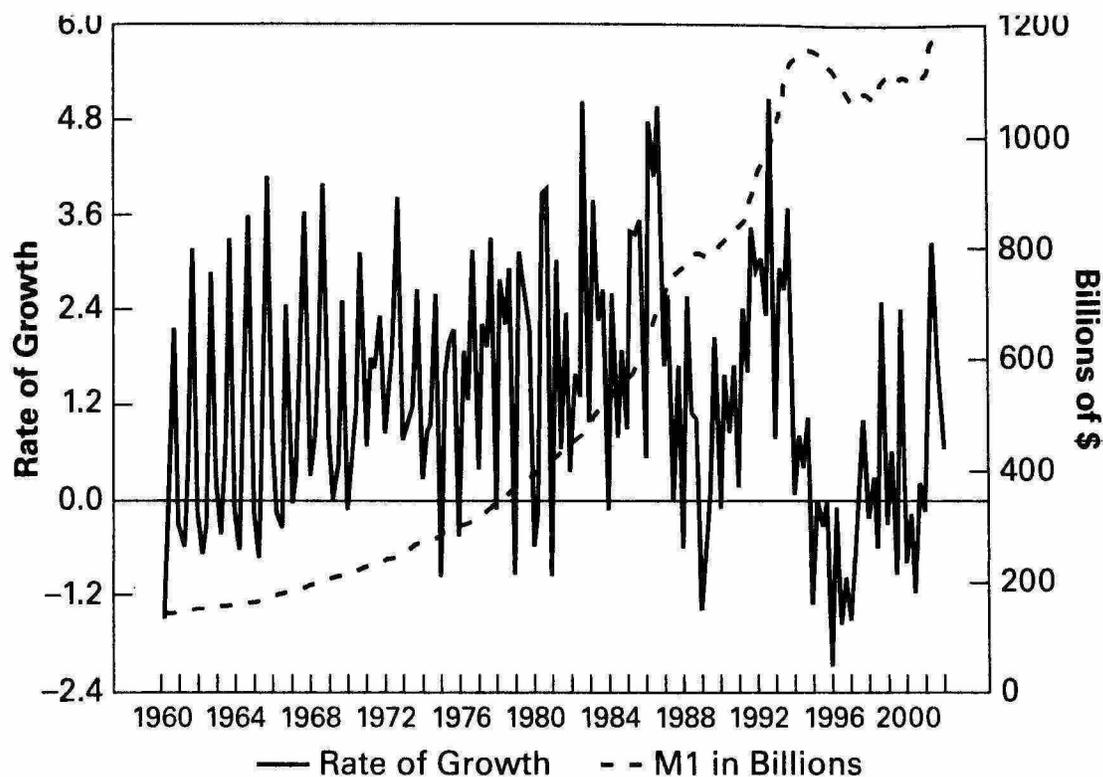

FIGURE 2.7   The Level and Growth Rate of M1

Note that actual money supply (the dashed line) is not very stable, but the rate of growth of the money supply is somewhat more stable.

II. Differences and Growth Rates (14.3)

It is often useful to transform Y_t so that it has certain convenient properties, such as stationarity (discussed below).

Differences

One common way to do this is to take differences over time. The **first difference** of a time series process Y_t is:

$$\Delta Y_t = Y_t - Y_{t-1}$$

The **second difference** is:

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

$$= Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2})$$

$$= Y_t - 2Y_{t-1} + Y_{t-2}$$

An **annual** (or **year-on-year**) change in a time series process with frequency s is defined as:

$$\Delta_s Y_t = Y_t - Y_{t-s}$$

Growth Rates

Closely related are growth rates. The **simplest type** is a **one-period growth rate**, which is denoted by Q_t defined and defined as:

$$Q_t = 100\left(\frac{\Delta Y_t}{Y_{t-1}}\right) = 100\left(\frac{Y_t}{Y_{t-1}} - 1\right) \quad (14.1)$$

which is multiplied by 100 to express it in percent terms.

I think Hansen had in mind quarterly (3-month) data, which is what he uses the letter Q.

It is common to report "**annualized**" **growth rates for variables that have a higher frequency ($s > 1$)**. Such rates are denoted by A_t and are defined as follows:

$$A_t = 100\left(\left(\frac{Y_t}{Y_{t-1}}\right)^s - 1\right) \quad (14.2)$$

Intuitively, the growth rate from time $t - 1$ to time t, a period of less than one year, is expressed in terms of **how much growth would occur in one year if that growth rate from $t - 1$ to t continued for one year**.

Another type of growth rate is a **year-on-year growth rate**, which can be denoted by G_t and are defined as:

$$G_t = 100\left(\frac{\Delta_s Y_t}{Y_{t-s}}\right) = 100\left(\frac{Y_t}{Y_{t-s}} - 1\right)$$

This simply measures the overall change in Y_t for a period of one year (from time $t - s$ to time t) and **does not use information about Y_t between those two time periods**.

Logarithmic Transformations

Growth rates can be closely (but not exactly) expressed by using logarithmic transformations, **if the growth rate is "small"** (e.g. 10% or less). More specifically:

$$Q_t \approx 100\Delta\log(Y_t)$$
$$A_t \approx s{\times}100\Delta\log(Y_t)$$
$$G_t \approx 100\Delta_s\log(Y_t)$$

To see how this works, consider Q_t.

$$\Delta\log(Y_t) = \log(Y_t) - \log(Y_{t-1})$$

$$= \log\left(\frac{Y_t}{Y_{t-1}}\right)$$

$$= \log\left(\frac{Y_{t-1}(1+g_t)}{Y_{t-1}}\right), \quad \text{where } g_t = \frac{Y_t}{Y_{t-1}} - 1$$

$$= \log(1) + \log(1 + g_t)$$

$$\approx g_t$$

$$= \Delta Y_t / Y_{t-1} = Q_t / 100$$

So $g_t = Q_t/100 \approx \Delta\log(Y_t)$, or equivalently $Q_t \approx 100\, \Delta\log(Y_t)$.

Log transformations are convenient for statistical analysis because they use simple differences, which are easier to use for such analysis compared to, for example, A_t, which is a nonlinear transformation of a ratio.

III. Stationary Processes (14.4, 14.5 and 14.6)

A time series variable Y_t **has some kind of underlying distribution *at time t*,** but strictly speaking that underlying distribution is **not some population from which Y_t has been sampled**. In addition, a series of Y_t, such as the set **(Y_t, Y_{t+1}, … Y_{t+t}) has a *joint* distribution**. For this set of variables, it is **useful to calculate means, variances, covariances**, and so forth. But this works only if Y_t has some kind of "constancy" or "stability". **For example, GDP per capita** usually increases over time, and so it **does not have a stable mean**. However, **its growth rate may have a stable mean**.

Stationarity

This need for constancy or stability leads to the concept of **stationarity**. There is a "weak" form (which is usually all one needs for linear models) and a "strict" form. To be as general as possible, Hansen defines it for the general case where Y_t **can be a vector**, but **for most purposes** you can **think of it as a scalar**.

Definition 14.1. The vector $\{Y_t\}$ is **covariance stationary, or weakly stationary,** if the **mean**

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

and **covariance matrix**

$$\Sigma = \mathrm{Var}(Y_t) = \mathrm{E}[(Y_t - \mu)(Y_t - \mu)']$$

are **finite** and **do not depend on** t, and the **autocovariances**

$$\Gamma(k) = \mathrm{Cov}(Y_t, Y_{t-k}) = \mathrm{E}[(Y_t - \mu)(Y_{t-k} - \mu)']$$

are independent of t (**do not depend on** t) for all k. $\Gamma(k)$ is finite, since Σ is finite (by Cauchy-Schwartz inequality).

Strictly speaking, these **means and variances of Y_t are characteristics of Y_t at time t**, which means that Y_t **at any time t is a variable** that has a mean and a variance

just for that time, and the **observed Y_t is a random draw from that distribution**. **Weak stationarity** means that these **means** and **variances** (and the **covariance** across any given interval of time) **are the same for all** t.

The **autocovariances** summarize the linear dependence of Y_t and its lags. **When Y_t is a single variable** (not a vector) it is **useful to express Var(Y_t) as σ^2 and express autocovariances as correlations**:

$$\rho(k) = \text{Corr}(Y_t,\ Y_{t\text{-}k}) = \frac{\text{Cov}(Y_t, Y_{t-k})}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t-k})}} = \frac{\gamma(k)}{\sigma^2} = \frac{\gamma(k)}{\gamma(0)}$$

Note that, by definition $\rho(k) = \rho(-k)$ for all k.

The **second definition of stationarity includes the entire (joint) distribution of Y_t:**

Definition 14.2. $\{Y_t\}$ is **strictly stationary** if the joint distribution of $(Y_t,\ Y_{t+1},\ \dots\ Y_{t+\ell})$ does not depend on t for all ℓ.

If the variances are bounded, then a process that is strictly stationary is also weakly (covariance) stationary.

From now on, "stationary" means "strictly stationary". Strict stationarity is useful for proving various theorems, as will be seen. Here is the first theorem that Hansen presents:

Theorem 14.1. If Y_t is i.i.d. (independently and identically distributed) then it is strictly stationary.

Note that **many, if not most, time series are *not* i.i.d.** (but some are).

Hansen gives some examples of processes that are strictly stationary, and some that are not strictly stationary. Here are a few of them. Note that e_t is always i.i.d and $E[e_t] = 0$.

Examples of Strictly Stationary Processes

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

$Y_t = Z$ for some random variable Z

$Y_t = (-1)^t Z$ for some random variable Z that is
         symmetrically distributed around 0

Examples of Processes that are NOT Strictly Stationary

$Y_t = t$

$Y_t = (-1)^t$

$Y_t = \sqrt{t} \times e_t$

$Y_t = Y_{t-1} + e_t$ with $Y_0 = 0$

Transformations of Stationary Processes

An important property of (strictly) **stationary processes** is that **any transformation** of such a process **is also** (strictly) **stationary**:

Theorem 14.2. If Y_t, which could be a vector, is strictly stationary, and $X_t = \phi(Y_t, Y_{t-1}, Y_{t-2}, \ldots)$ is a random vector of dimension q, then X_t is strictly stationary.

Convergent Series

An **example of a transformation** that uses the full past history is the **"infinite-order moving average" transformation**. This is defined, with Y_t **and** a_j both **as scalars**, as:

$$X_t = \sum_{j=0}^{\infty} a_j Y_{t-j} \qquad (14.3)$$

Many time-series models use this form. Of course, an **infinite sum can go off to infinity**. If this is the case, we say that X_t does not exist (does not have a finite limit). So it is **useful to clarify when this would not happen**. This leads to the concept of **convergence** of an infinite series:

Definition 14.3. **Convergence** The **infinite series** $X_t = \sum_{j=0}^{\infty} a_j Y_{t-j}$ in equation (14.3) **converges almost surely** if it has a finite limit with probability one as $n \to \infty$. When this is the case, we say that X_t is **convergent**.

Theorem 14.3. If Y_t is strictly stationary, $E[Y] < \infty$, and $\sum_{j=0}^{\infty} |a_j| < \infty$, then $X_t = \sum_{j=0}^{\infty} a_j Y_{t-j}$ in equation (14.3) converges almost surely, and X_t is strictly stationary.

IV. Ergodicity and the Ergodic Theorem (14.7, 14.8)

Eventually we will want to show that estimated parameters for time series models are consistent, and that their variances have well-defined limiting distributions. This requires using laws of large numbers (LLNs) and central limit theorems (CLTs). However, **strict stationarity is not sufficient to make use of LLNs and CLTs**. What we will need is a property called **ergodicity**.

To see why strict stationarity is insufficient. Consider a **very simple example: $Y_t = Z$** for all t for some random variable Z. Note that Z does not have a t subscript. For a given sample of Y_t, we **draw a value for Z** and we **use this same value for each Y_t**, so in our sample Y_t **is a constant** for all values of t. Y_t **is strictly stationary**, but as we will see it is **not ergodic**.

One way to see the problem is that, by definition, $E[Y_t] = E[Z]$, but our sample values for Y_t will give an inconsistent estimate of $E[Y_t]$ because the **sample mean of Y_t** will be "stuck" at the draw of Z and **will not converge to E[Z] as n** $\rightarrow \infty$.

11

Ergodicity

Strictly speaking, a time series variable Y_t is **ergodic** if **all "invariant" events are "trivial"**. This is explained more rigorously below, but for now note that an **"invariant event"** is an **event that is not affected by a "shift" in time** (i.e. replace Y_t with $Y_{t+\ell}$ for some ℓ), and a **"trivial" event** is an event that **has a probability of zero or one**.

Intuitively, a times series variable is **ergodic if its variation over time** passes through all parts of the "sample space" of its distribution and **does not get "stuck" in a subspace** of that sample space.

Before defining ergodicity rigorously, here are some **basic properties of ergodic variables**.

First, any i.i.d (independently and identically distributed) **sequence is ergodic**:

Theorem 14.4. If any **vector** Y_t is i.i.d., then it is strictly stationary and ergodic.

Second, **ergodicity**, like stationarity, is **preserved under transformations**.

Theorem 14.5. If any **vector** Y_t is strictly stationary and ergodic, and X_t, which is defined as $X_t = \phi(Y_t, Y_{t-1}, Y_{t-2}, \ldots)$, is a random vector, then X_t is strictly stationary and ergodic.

An **example of this** is the infinite-order moving average transformation in **equation (14.3)** if Y_t is strictly stationary and ergodic:

Theorem 14.6. If a **variable** Y_t is strictly stationary, ergodic, and $E[|Y|] < \infty$, and $\sum_{j=0}^{\infty} |a_j| < \infty$, then $X_t = \sum_{j=0}^{\infty} a_j Y_{t-j}$ is also strictly stationary and ergodic.

Another useful property is that the **Cesàro sum** (maybe it should be called Cesàro mean of a sum), which is the sum of the autocovariances, **of a strictly stationary ergodic variable** has a limit of 0 as $n \to \infty$:

Theorem 14.7. If a **variable** Y_t is strictly stationary, ergodic, and $E[Y^2] < \infty$, then its Cesàro sum has a limit of 0:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{\ell=1}^{n} \mathrm{Cov}(Y_t, Y_{t+\ell}) = 0 \qquad (14.4)$$

This means that the covariances are finite, and that **for large enough ℓ** it is always the case that **$\mathrm{Cov}(Y_t, Y_{t+\ell}) = 0$**. This implies that an infinite sum of these covariance terms will be finite, so dividing by n implies that the "average" covariance $\to 0$ as $n \to \infty$.

Definition of Ergodicity

This is optional, but see if you think it is interesting.

Consider Y_t, a **vector** of m time series variables.

An **event** for Y_t is a specific "infinite history" of Y_t, which can be denoted by $\tilde{Y}_t = (\dots Y_{t-1}, Y_t, Y_{t+1}, \dots)$. Any $\tilde{Y}_t$ is a sample drawn from the underlying joint distribution. Let G be the subspace of R^m that contains all possible events of Y_t. We can **denote a specific event** by $A = \{\tilde{Y}_t \in G\}$ with $G \subset R^{m\infty}$.

The ℓth **time-shift** of $\tilde{Y}_t$ is defined as $\tilde{Y}_{t+\ell} = (\dots Y_{t-1+\ell}, Y_{t+\ell}, Y_{t+1+\ell}, \dots)$. Thus $\tilde{Y}_{t+\ell}$ replaces each observation Y_t in $\tilde{Y}_t$ with its ℓth shifted value $Y_{t+\ell}$.

An event A is **invariant** if it is not affected by a time shift, which can be expressed as $A_\ell = A$. An **example** of this is $A = \{\max_{-\infty < t < \infty} Y_t \leq 0\}$, which is the "event" that the maximum value of Y_t for all t from $-\infty$ to ∞ is ≤ 0. This is **invariant** because shifting Y_t forward or backward in time does not change the maximum value of Y_t from $-\infty$ to ∞.

An event A is **trivial** if either $\text{Prob}[A] = 0$ or $\text{Prob}[A] = 1$. That is, A **has no uncertainty or randomness**.

Recall the definition of ergodic: Y_t **is ergodic** if all invariant events are trivial. That is, **any event that is not affected by a time shift is not random**.

Consider the event $A = \{ \max\limits_{-\infty < t < \infty} Y_t \leq 0 \}$. **Recall the example** $Y_t = Z$, which in effect means that Y_t is a constant for all t. **In this case A is certainly invariant** to a time shift.

Suppose that $Z \sim N(0, 1)$. Then $\text{Prob}[A] = 0.5$, so A **is not a trivial event**, which means that Y_t **is not ergodic**.

In contrast, suppose that Y_t is i.i.d. as $N(0, 1)$. In this case there is a **new "draw" of Y_t for each time period**. A is still invariant to a time shift, but the **probability that $Y_t \leq 0$ for every time period is zero**, so this **is ergodic**.

Theorem 14.8. A strictly stationary time series for a vector Y_t is ergodic if and only if for all events A and B:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{\ell=1}^{n} \text{Prob}[A_\ell \cap B] = \text{Prob}[A] \times \text{Prob}[B] \qquad (14.5)$$

This is very abstract. The limit is the Cesàro sum of $\text{Prob}[A_\ell \cap B]$. A closely related to property of time series variables is called **mixing**, which states that $\text{Prob}[A_\ell \cap B]$ converges to $\text{Prob}[A] \times \text{Prob}[B]$. In fact, mixing implies ergodicity. We will cover mixing more in the next lecture.

Hayashi (2000) has a less abstract definition:

A (strictly) stationary process for the variable Y_t is ergodic if, for any two bounded functions f and g, we have:

$$\lim_{n \to \infty} \left| E[f(Y_t, \ldots Y_{t+k})g(Y_{t+n}, \ldots Y_{t+n+\ell})] \right|$$

$$= \left| E[f(Y_t, \ldots Y_{t+k})] \right| \times \left| E[g(Y_{t+n}, \ldots Y_{t+n+\ell})] \right|$$

Intuitively, variables that are far apart in time are essentially statistically independent.

The Ergodic Theorem

Hansen says that this is "one of the most famous results of time series theory". He also says that it can be expressed in different ways, but "for simplicity" he states the theorem in terms of convergence in probability.

Theorem 14.9: **Ergodic Theorem**.

If the **vector** Y_t is strictly stationary, ergodic, and $\mathrm{E}[\|Y\|] < \infty$, then as $n \to \infty$

$$\mathrm{E}[\|\bar{Y} - \mu\|] \to 0 \qquad (14.6)$$

and

$$\bar{Y} \underset{p}{\to} \mu \qquad (14.7)$$

where $\bar{Y}$ **is the sample mean** of Y, $\mu = \mathrm{E}[Y]$, and $\|\ \|$ is the "distance" between two vectors (for $\|Y\|$ the other vector is all zeros), which for any two vectors Y_a and Y_b of size m can be expressed as $((Y_{a1} - Y_{b1})^2 + (Y_{a2} - Y_{b2})^2 + \ldots + (Y_{am} - Y_{bm})^2)^{1/2}$.

The important point is that **the mean of a vector of ergodic time series variables**, or transformations of them, **can be consistently estimated** by the sample mean.

This is analogous to a (Weak) Law of Large Numbers.

Hansen provides a proof of the ergodic theorem for a single variable (scalar case) at the end of Section 14.8, but this is optional.

V. Conditioning on Information Sets (14.9)

In theory, we can think of time series variables as having an "infinite history" that goes back in time to -∞. **Many uses of time series analysis** involve taking **conditional expectations** based on past values of some variable Y. For **example**, **forecasting** a variable could involve taking the conditional expectation of Y_{t+1} conditional on Y_t, Y_{t-1}, etc.

We can also, in theory, define conditional expectations in a way that conditions on "infinite" histories.

To start, recall from probability theory that an **outcome** is **an element of a sample space** (a given value of all the possible values that a variable can take). An **event** is a **set of outcomes**. Finally, a **probability law** is a "rule" that assigns non-negative real numbers to events. When the events are collections of infinite histories, a probability law can assign such numbers (probabilities) to collections of infinite histories.

Suppose we want to define a **conditional expectation given an infinite past history**. We can write this as:

$$E_{t-1}[Y_t] = E_t[Y_t | Y_{t-1}, Y_{t-2}, \dots] \qquad (14.10)$$

This conditional expectation exists if $E[Y_t] < \infty$ and the probabilities of the history of Y_t (which can be denoted by $\tilde{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \ldots))$ are well defined.

Hansen says that he usually avoids using measure-theoretic terminology in his textbook, but he makes an exception here. He **restates (14.10) as**:

$$E_{t-1}[Y_t] = E_t[Y_t | \mathscr{F}_{t-1}] \qquad (14.10')$$

where $\mathscr{F}_{t-1} = \sigma(\tilde{Y}_{t-1})$ is the "σ-field" generated by the history $\tilde{Y}_{t-1}$. **A σ-field is a collection of sets that satisfies certain regularity conditions**. The **intuition** is that $\mathscr{F}_{t-1}$ **contains all the "information" available in the history $\tilde{Y}_{t-1}$.** In fact, economists call $\mathscr{F}_{t-1}$ an **information set** instead of a σ-field. Hansen says that he will use the statistical notation, that is $\mathscr{F}_{t-1}$, but he will use the economists' terminology and call $\mathscr{F}_{t-1}$ an "information set".

Here are some **properties of information sets**:

1. They are **nested** $\mathscr{F}_{t-1} \subset \mathscr{F}_t$. In words, information accumulates over time and is not "lost".

2. One must **always specify which variables are included** as part of the information set.

3. The conditional expectations in (14.10′) **"obey" the law of iterated expectations** and the conditioning theorem, which means that:

$$E[E[Y_t | \mathscr{F}_{t\text{-}1}] | \mathscr{F}_{t\text{-}2}] = E[Y_t | \mathscr{F}_{t\text{-}2}]$$

$$E[E[Y_t | \mathscr{F}_{t\text{-}1}]] = E[Y_t]$$

$$E[Y_{t\text{-}1} \times Y_t | \mathscr{F}_{t\text{-}1}] = Y_{t\text{-}1} \times E[Y_t | \mathscr{F}_{t\text{-}1}]$$

This lecture was very abstract, to "set the stage" for more applied time series analysis, so please be patient. The next lecture is also very abstract, but the following two lectures are more applied.