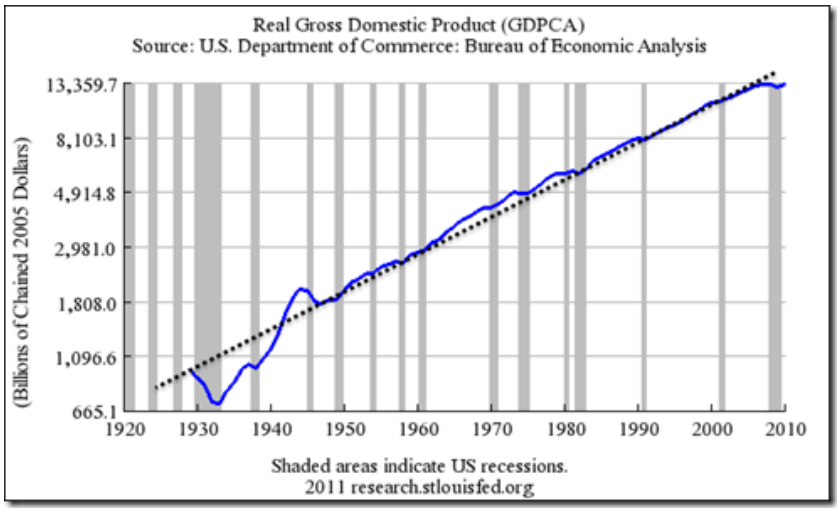
Notes on Time Series Econometrics

Diego Vilán *

Fall 2012

These notes cover topics in time series/macroeconometrics. Some of the topics included are: stochastic processes, with a focus on autoregressive and moving average models. Also covered are multivariate autoregressions (VARs), impulse response analysis and state-space modeling. As usual, it all begins with the data. Consider the following plot of US GDP:



The time series above can be regarded as a particular realization of a stochastic process. We will argue that, if the data generating process (DGP) satisfies certain conditions, then a single realization of the time series will be sufficient for us to understand the underlying nature of the existing relationship. This temporal homogeneity in the behavior of a series is usually referred to as stationarity.

1 Stationarity

Stationarity is a probabilistically meaningful measure of regularity. This regularity may be exploited to estimate unknown parameters and characterize the dependence between observations across time. Broadly speaking, a process is said to be stationary if its probability distribution remains unchanged over time. The invariability of the structural relation (i.e.: the conditional distribution, conditional

*__DISCLAIMER__: I wrote these notes as a study aid for myself. They are work in progress and could be incomplete, inaccurate and even somewhat incorrect. Keep that in mind should you decide to use them. Comments and suggestions welcomed!

mean and variance) makes it possible to use historical data to estimate the underlying coefficients governing the DGP. It also allows us to forecast the future based on past observations.

In short, why do we care about stationary processes?

- Fundamentally because stationarity implies that the underlying relationships between variables described by the DGP is not changing over time. If the DGP would change frequently and in an unpredictable way, constructing a meaningful model of it would be almost impossible.
- Moreover, given the nature of time series observations, studying one realization of a stationary stochastic process will afford us an understanding of the existing relationship.
- Since we often think of economic equilibria as being characterized by stable, long-term relations, stationarity processes fit well that description.
- On a technical side, many standard results in econometrics (e.g.: CLT) often require stationarity. Nonstationarity will introduce several complications: estimators and test statistics will often have nonstandard asymptotic distributions.

Formal definitions:

Definition 1 *(Strict stationary) A stochastic process* $\{x_t\}$ *is strictly stationary if the joint distribution of* $\{x_t, x_{t+1}, ..., x_{t+h}\}$ *depends only on h and not on t.*

Strict stationarity requires that the joint distribution of a stochastic process does not depend on time and so the only factor affecting the relationship between two observations is the gap between them.

Strict stationarity is weaker than i.i.d. since a stationary process may be dependent. Nonetheless it is still often too strong of an assumption for most financial and macroeconomic time series. Consequently, we do not usually require strict stationarity but rather covariance stationarity.

Definition 2 *(Weak or covariance stationary) A stochastic process* $\{x_t\}$ *is covariance stationary if:*

$$\begin{aligned} E[x_t] &= \mu \quad \forall t \\ V[x_t] &= \sigma^2 < \infty \quad \forall t \\ E[(x_t - \mu)(x_{t-s} - \mu)] &\quad \forall t \end{aligned}$$

Covariance stationarity requires that both the unconditional mean and variance are finite and do not change over time. Note that covariance stationarity only applies to unconditional moments, so a covariance process may still have varying conditional mean.

These two types of stationarity are related although neither nests the other. If a process is strictly stationary and has finite second moments, then it is covariance stationary. If a process is covariance stationary and the joint distribution of the studentized residuals does not depend on time, then the process is strictly stationary.

For purposes of these notes, stationarity will usually imply weak/ covariance stationarity.

Remark 1 *Most economic time series in their original form are usually not stationary. It is widely accepted that time series are often governed by four main components that can be associated to different types of temporal variations*[1]*:*

1. *A trend of long term movement*
2. *A seasonal effect*
3. *A cyclical component*
4. *A residual, irregular or random effect*

Nonstationary time series often undergo some transformation to become stationary (for e.g.: differencing or detrending). These notes will only briefly mention some of these topics in the appendix and focus almost exclusive on stationary processes.

2 Stochastic Processes

Definition 3 *(Stochastic process) Any process that generates a sequence of random variables* $\{x\}$ *that are ordered in an immutable fashion. One such way of ordering is by time, usually noted as* $\{x_t\}$*.*

Alternative definition

Definition 4 *(Stochastic process) Any process that is ordered (i.e.: is a sequence that is immutable) is a stochastic process. A time series is an example of a stochastic process.*

Note that a stochastic process is a sequence of random variables, while a realization of such a process is a sequence of real numbers.

Associated with this stochastic process is its history:

$$H_t = \{x_t, x_{t-1}, x_{t-2}, ...\}$$

The expected value at time t of a particular random variable is:

$$E_t(x_s) = E(x_s|H_t)$$

or its expected value conditional on all information available at time t.

There are two special classes of stochastic processes that are particularly useful for Macroeconomics: (i) Markov Processes and (ii) Linear Stochastic Processes.

[1] These four components are usually combined together either using an additive or multiplicative model.

2.1 Markov Processes

Definition 5 *(Markov process) A Markov process is a stochastic process which satisfies the following property:*

$$P(x_{t+1}|H_t) = P(x_{t+1}|x_t)$$

where $P(x)$ is the probability distribution of x.

Definition 6 *(Markov chain) A Markov chain is a Markov process with a countable space. A Markov chain can be defined using three elements: (i) an n-dimensional vector $\omega \in \mathrm{R^n}$ defining the state space, (ii) an n-dimensional vector of initial probabilities π_0*

$$\pi_{0i} = P(x_0 = \omega_i)$$

and (iii) an n-by-n transition matrix:

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{21} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1N} & 0 & \cdots & p_{NN} \end{bmatrix}$$

where $p_{ij} = P(x_{t+1} = \omega_j | x_t = \omega_i)$

In order for everything to be well defined, we also require:

$$\sum_{i=0}^{n} \pi_{0i} = 1$$
$$\sum_{j=0}^{n} p_{i,j} = 1$$

Also, let the n-vector π_t be defined as:

$$\pi_{ti} = P(x_t = \omega_i)$$

Note that π_t is not a random object.

Why might we find Markov chains convenient for modeling? Two main reasons:

1. Dynamic Programming problems are often easier to solve for a discrete state space and there is usually a Markov chain which is "close enough" to another stochastic process to serve as a more convenient approximation.
2. Several of the properties of a given Markov chain can be easily derived from its transition matrix.

(Provide Example)

We can prove that there will always be at least one stationary distribution for any transition matrix.

When will there be exactly one?

Proposition 1 *Let P be a Markov transition matrix such that there exists $n \geq 1$ such that $(P^n)_{ij} > 0 \; \forall i, j$. Then P is asymptotically stationary and has a unique stationary distribution.*

In other words, if every element of P is strictly positive, then for any initial distribution π_0 :

$$\lim_{t \to \infty} \pi_t = \pi$$

2.2 Linear (stationary) Stochastic Processes

2.2.1 White Noise

The simplest form of a covariance (or weakly) stationary process is the *white noise*[2] process.

Definition 7 *(White noise) A stochastic process $\{e_t\}$ is a white noise process if the following holds:*

$$\begin{aligned} E(e_t) &= 0 \\ E(e_t^2) &= \sigma^2 \\ E(e_t e_\tau) &= 0 : t \neq \tau \end{aligned}$$

In short:

i) zero mean

ii) constant variance

iii) no serial correlation

It should be noted that although $\{e_t\}$ are serially uncorrelated, they are not **necessarily** serially independent, since they are not **necessarily** normally distributed[3].

Definition 8 *(Gaussian white noise) A stochastic process $\{e_t\}$ is a white noise process if the following holds:*

$$\begin{aligned} E(e_t) &= 0 \\ E(e_t^2) &= \sigma^2 \\ E(e_t e_\tau) &= E(e_t)E(e_\tau) : t \neq \tau \\ e_t &\sim N(0, \sigma^2) \end{aligned}$$

The white noise process is important because it will be the foundation for most of the other stochastic process we are interested in. For example, we can use a white noise process to construct a wide range of ARMA processes.

[2]Why white noise? This name comes from the engineering literature where certain instruments are able to analyze the light waves into components of various frequencies. The white light has the property that all frequencies enter equally. It turns out that with truly iid realizations of a random variable normally distributed as inputs, one obtains a flat spectrum for the light waves which is similar to that of white light. Just as all rainbow colors can be obtained from white color, many stationary processes can be written as a linear combination of a white noise process.

[3]Recall that zero correlation only implies independence for normally distributed random variables

2.2.2 ARMA Models

Autoregressive moving average (ARMA) processes will form the core of time-series analysis. The ARMA class can be decomposed into two smaller classes: autoregressive (AR) and moving average (MA) processes.

Definition 9 *(Moving Average MA(1)) The stochastic process* $\{x_t\}$ *is a moving average process of order* 1 *if:*

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

where $\{\varepsilon_t\}$ is a white noise process with the additional property that $E_{t-1}[\varepsilon_t] = 0$. This assumption follows from the fact that the innovation is unpredictable using the time $t-1$ information set. The process is called a moving average since it is built as a weighted average of a white noise process.

The unconditional mean is:

$$\begin{aligned} E(y_t) &= E(\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t) \\ &= \mu + \theta_1 E(\varepsilon_{t-1}) + E(\varepsilon_t) \\ &= \mu + \theta_1 0 + 0 \\ &= \mu \end{aligned}$$

The conditional mean is:

$$\begin{aligned} E_{t-1}(y_t) &= E_{t-1}(\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t) \\ &= \mu + \theta_1 E_{t-1}(\varepsilon_{t-1}) + E_{t-1}(\varepsilon_t) \\ &= \mu + \theta_1\varepsilon_{t-1} + 0 \\ &= \mu + \theta_1\varepsilon_{t-1} \end{aligned}$$

The differences in the means reflect the persistence of the previous shocks in the current period. The variances can be similarly derived.

The unconditional variance is:

$$\begin{aligned} V(y_t) &= E[(\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t - E[\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t])^2)] \\ &= E[(\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t - \mu)^2] \\ &= [\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t - \mu)^2] \\ &= \theta_1^2 E[\varepsilon_{t-1}^2] + E[\varepsilon_t^2] + 2\theta_1 E(\varepsilon_{t-1}\varepsilon_t)] \\ &= \sigma^2\theta_1^2 + \sigma^2 + 0 \\ &= \sigma^2(1 + \theta_1^2) \end{aligned}$$

where $E(\varepsilon_{t-1}\varepsilon_t = 0)$ follows from the white noise assumption.

The conditional variance is:

$$\begin{aligned} V_{t-1}(y_t) &= E_{t-1}[(\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t - E_{t-1}[\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t])^2)] \\ &= E_{t-1}[(\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t - \mu - \theta_1\varepsilon_{t-1})^2] \\ &= E_{t-1}[\varepsilon_t^2] \\ &= \sigma_t^2 \end{aligned}$$

where σ_t^2 is the conditional variance of $\{\varepsilon_t\}$. White noise processes do not necessarily have to be homoskedastic, although if $\{\varepsilon_t\}$ is, then $V_{t-1}(y_t) = \sigma^2$. Like the mean, the unconditional variance and the conditional variance are different. The unconditional variance is unambiguously larger than the average conditional variance due to the extra variability introduced by the moving average term.

The autocovariances can be derived as:

$$\begin{aligned} \gamma_{1t} &= E[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))] \\ &= E[(\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t - \mu)(\mu + \theta_1\varepsilon_{t-2} + \varepsilon_{t-1} - \mu)] \\ &= E[\theta_1\varepsilon_{t-1}^2 + \theta_1\varepsilon_t\varepsilon_{t-2} + \varepsilon_t\varepsilon_{t-1} + \theta_1^2\varepsilon_{t-1}\varepsilon_{t-2}] \\ &= \theta_1\sigma^2 + 0 + 0 + 0 \\ &= \theta_1\sigma^2 \end{aligned}$$

$$\begin{aligned} \gamma_{2t} &= E[(y_t - E(y_t))(y_{t-2} - E(y_{t-2}))] \\ &= E[(\mu + \theta_1\varepsilon_{t-1} + \varepsilon_t - \mu)(\mu + \theta_1\varepsilon_{t-3} + \varepsilon_{t-2} - \mu)] \\ &= E[(\theta_1\varepsilon_{t-1} + \varepsilon_t)(\theta_1\varepsilon_{t-3} + \varepsilon_{t-2})] \\ &= E[\theta_1\varepsilon_{t-1}\varepsilon_{t-2} + \theta_1\varepsilon_{t-3}\varepsilon_t + \varepsilon_t\varepsilon_{t-2} + \theta_1^2\varepsilon_{t-1}\varepsilon_{t-3})] \\ &= \theta_1 E[\varepsilon_{t-1}\varepsilon_{t-2}] + \theta_1 E[\varepsilon_{t-3}\varepsilon_t] + E[\varepsilon_t\varepsilon_{t-2}] + \theta_1^2 E[\varepsilon_{t-1}\varepsilon_{t-3})] \\ &= 0 + 0 + 0 + 0 = 0 \end{aligned}$$

The MA(1) can be generalized into the class of MA(q) processes by including additional lagged errors.

Definition 10 *(Moving Average MA(q)) The stochastic process $\{y_t\}$ is a moving average process of order q if:*

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q} \\ &= \varepsilon_t + \sum_{q=1}^{Q} \theta_q\varepsilon_{t-q} \end{aligned}$$

where $\{\theta_1, \theta_2, ..., \theta_q\}$ is a sequence of real numbers and $\{\varepsilon_t\}$ is a white noise process with the additional property that $E_{t-1}(\varepsilon_t) = 0$.

Key things to remember:

1. $E(y_t) = \mu$
2. $V(y_t) = (1 + \sum_{q=1}^{Q} \theta_q^2)\sigma^2$
3. $$\gamma_{jt} = \begin{cases} \sigma^2 \sum_{i=0}^{q-j} \theta_i \theta_{i+j} & \text{if } j \leq q \\ 0 & \text{if } j > q \end{cases}$$

   where $\theta_0 = 1$.

2.2.3 Autoregressive Processes

Another class of ARMA processes is the autoregressive process.

Definition 11 *(First Order Autoregressive Process) A first order autoregressive process (AR(1)) has dynamics which follow:*

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$$

where $\{\epsilon_t\}$ is a white noise process with the additional property that $E_{t-1}(\epsilon_t) = 0$.

Unlike the MA(1) process, y appears on both sides of the equation. However, this is just a convenience and the process can be re-written to provide an expression that depends only on the disturbances ϵ_t and an initial condition.

$$\begin{aligned} y_t &= \phi_0 + \phi_1 y_{t-1} + \varepsilon_t \\ &= \phi_0 + \phi_1(\phi_0 + \phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_0 + \phi_1\phi_0 + \phi_1^2 y_{t-2} + \varepsilon_t + \phi_1\varepsilon_{t-1} \\ &= \phi_0 + \phi_1\phi_0 + \phi_1^2(\phi_0 + \phi_1 y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \phi_1\varepsilon_{t-1} \\ &\vdots \\ &= \sum_{i=0}^{t-1} \phi_1^i \phi_0 + \sum_{i=0}^{t-1} \phi_1^i \varepsilon_{t-i} + \phi_1^t y_0 \end{aligned}$$

Using backwards substitution, an AR(1) can be expressed as an MA(t). In many cases the initial condition is unimportant and the AR process can be assumed to have begun long ago in the past. As long as $|\phi_1| < 1$, then $\lim_{t\to\infty} \phi^t y_0 \to 0$ and the effect of the initial condition will be negligible.

Using the infinite history version of an AR(1) and assuming that $|\phi_1| < 1$, then the above solution simplifies to:

$$\begin{aligned} y_t &= \phi_0 + \phi_1 y_{t-1} + \varepsilon_t \\ &= \sum_{i=0}^{\infty} \phi_1^i \phi_0 + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} + 0 \\ &= \frac{\phi_0}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \end{aligned}$$

where the identity $\sum_{i=0}^{\infty} = \frac{1}{1-\phi_1}$ is used.

This expression of an AR process is known as an MA(∞) representation and it is useful for deriving properties.

The unconditional mean:

$$\begin{aligned} E(y_t) &= E\left[\frac{\phi_0}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}\right] \\ &= \frac{\phi_0}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i E(\varepsilon_{t-i}) \\ &= \frac{\phi_0}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i 0 \\ &= \frac{\phi_0}{1-\phi_1} \end{aligned}$$

As long as $\{y_t\}$ is covariance stationary, so that $E(y_t) = E(y_{t-1}) = \mu$, the unconditional mean may also be derived as:

$$\begin{aligned} E(y_t) &= E(\phi_0 + \phi_1 y_{t-1} + \varepsilon_t) \\ &= \phi_0 + \phi_1 E(y_{t-1}) + E(\varepsilon_{t-1}) \\ \mu &= \phi_0 + \phi_1 \mu + 0 \\ \mu(1-\phi_1) &= \phi_0 \\ E(y_t) &= \frac{\phi_0}{1-\phi_1} \end{aligned}$$

The unconditional variance:

$$\begin{aligned} V(y_t) &= V(\phi_0 + \phi_1 y_{t-1} + \varepsilon_t) \\ &= V(\phi_0) + V(\phi_1 y_{t-1}) + V(\varepsilon_t) + 2Cov(\phi_1 y_{t-1}, \varepsilon_t) \\ &= 0 + \phi_1^2 V(y_{t-1}) + \sigma^2 + 2 * 0 \\ &= \phi_1^2 V(y_t) + \sigma^2 \\ V(y_t) - \phi_1^2 (y_t) &= \sigma^2 \\ V(y_t) &= \frac{\sigma^2}{1-\phi_1^2} \end{aligned}$$

where $Cov(y_{t-1}, \varepsilon_t) = 0$ follows from the white noise assumption given that y_{t-1} is a function of $\varepsilon_{t-1}, \varepsilon_{t-2}, ...$

The conditional variance:

$$\begin{aligned} V_{t-1}(y_t) &= E_{t-1}[(\phi_0 + \phi_1 y_{t-1} + \varepsilon_t)^2] \\ &= E_{t-1}(\varepsilon_t^2) \\ &= \sigma_t^2 \end{aligned}$$

The unconditional variance is again larger than the average conditional variance and the variance explodes as $|\phi_1|$ approaches 1 or -1.

The autocovariance:

$$\gamma_{jt} = \phi_1^j \frac{\sigma^2}{1 - \phi_1^2}$$

The AR(1) can be extended to the AR(p) class by including additional lags of y_t.

3 Analysis of Multiple Time Series

Multivariate time-series analysis extends many of the ideas of univariate time-series to systems of equations. The primary model is the vector autoregression (VAR), a direct and natural extension of the univariate autoregression. Most results that apply to univariate time-series can be directly ported to multivariate time-series with a minimal change of notation and the use of linear algebra.

3.1 Vector Autoregressions

Definition 12 *(Vector Autoregressive process) Insert text ...*

Most of the ideas associated with scalar processes carry over in vector language. For example, the scalar AR(1) process will be stationary iff $|a| < 1$, while the VAR(1) will be stationary iff the eigenvalues of XX are all less than 1.

3.1.1 Simple/Canonical VAR

Companion Form.

3.1.2 Strcutural VAR

Hablar sobre simultaneous eq models and stuff ... Structural VARS usually take the form of:

$$X_t = B_0 X_t + B_1 X_{t-1} + C u_t$$

where u_t is a vector of uncorrelated white noise processes, each with unit variance and the diagonals of B_0 are zero.

The idea is that shocks may **directly** affect only each variable, but that variables may affect each other contemporaneously. With this in mind is that these shocks are often called *structural* shocks.

A structural VAR is different from a simple VAR in at least two ways:

1. Variables do not affect each other contemporaneously
2. Shocks are not correlated with each other

Problem of estimation

3.1.3 Reduced Form VAR

4 Appendix

4.1 Probability and Statistical Preliminaries

Let x and y be random variables characterized by a probability distribution and a be a constant term.

Definition 13 *1st central moment (Mean):*

$$E(x) = \mu$$

Definition 14 *2nd central moment (Covariance):*

$$E\left[(x - E(x)(y - E(y)\right] = E(xy) - E(x)E(y)$$

The variance is just a special case of the covariance when the two variables are identical. That is to say the Variance is the Covariance of the random variable. with itself:

$$\begin{aligned} Cov(x,x) &= E\left[(x - E(x)(x - E(x)\right] \\ &= E[(x - E(x))^2] \\ &= E(x^2) - [E(x)]^2 \\ &= Var(x) \end{aligned}$$

Properties:

$$\begin{aligned} Var(a) &= 0 \\ Var(x) &= \sigma^2 \\ Var(a+x) &= Var(x) + Var(a) = \sigma^2 \\ Var(ax) &= a^2 Var(x) = a^2\sigma^2 \\ Var(x+y) &= Var(x) + Var(y) + 2Cov(x,y) \end{aligned}$$

Additional properties:

$$\begin{aligned} Cov(x,a) &= 0 \\ Cov(x,y) &= Cov(y,x) \\ Cov(ax+by) &= abCov(x,y) \\ Cov(a+x,b+y) &= Cov(x,y) \end{aligned}$$

Definition 15 *(Statistical Independence)*
Two random variables x and y are statistical independent iff:

$$\begin{aligned} E(xy) &= E(x)E(y) \\ &\Rightarrow Cov(x,y) = 0 \end{aligned}$$

Note: Statistical independence implies no correlation between x and y (i.e.: $Cov(x,y) = 0$). However $Cov(x,y) = 0$ does not imply statistical independence unless x and y are normally distributed.

Definition 16 *(Autocovariance)*
The variance of a random variable at difference points in time.

$$\begin{aligned} \gamma_{jt} &= E[(x_t - E(x_t))(x_{t-j} - E(x_{t-j})] \\ &= E(x_t x_{t-j}) - E(x_t)E(x_{t-j}) \end{aligned}$$

If $\{x_t\}_{t=0}^{\infty}$ *has zero mean then:*

$$\gamma_{jt} = E(x_t x_{t-j})$$

Why do we care about autocovariances? Two reasons:

1. Forecasting: helps us understand a variable's evolution over time.
2. Modeling: helps us understand how persistent a process is. Models should reflect this persistence.

Just as it is useful to normalize covariances by dividing by the respective variables' standard deviations, it is also useful to normalize autocovariances.

Definition 17 *(Autocorrelation)*
Insert text

4.2 Difference Equations

Definition 18 *(Difference equation) A difference equation expresses the value of a variable as a function of its own lagged values, time and other variables.*

Consider the following difference equation given by:

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + x_t \tag{1}$$

Note:

1. The order of the difference equation is given by the number of lags (i.e.: the value of n)
2. The term x_t is called the driving or forcing process
3. From an appropriate choice of the forcing process, we can obtain a wide variety of important macroeconomic models (see stochastic processes).

4. This n-th order linear difference equation is a special case where the coefficients a_i are constant. Economic theory may indicate how the various a_i are function of variables in the economy. As long as they do not depend on any values of y_t or x_t, we can regard them as parameters.

Remark 2 *Classification of difference equations*

A difference equation can be classified according to:

1. Order
2. Linearity/nonlinear
3. Autonomonous/ nonautonomous
4. Stochastic/deterministic

Order. The order of a difference equation is determined by the highest order of difference contained in the equation. An *nth* order difference equation contains variables at most n periods apart.

Autonomous. A difference equation is said to be autonomous if it does not depend on time explicitly; otherwise it is non autonomous. For example:

$$y_{t+1} = 2y_t + 3t$$

is nonautonomous because it depends explicitly on the variable t. On the other hand:

$$y_{t+1} = 2y_t + 3$$

is autonomous because it does not depend explicitly on t. Autonomous difference equations are most prevalent in economics.

Linear or nonlinear. A difference equation is nonlinear if it involves any nonlinear terms in y_t, y_{t+1} and so on. For example:

$$y_{t+1} = 2y_t^2 + 3$$

is a nonlinear, autonomous, first order difference equation.

Stochastic or deterministic. A difference equation is stochastic if the forcing process is itself a random variable which follows a stochastic process. Alternatively, for a deterministic difference equation, the driving process is either a constant value, or a known sequence of values. One could think of this deterministic component as being captured by the intercept of the equation.

4.2.1 Solution Concept

The concept of a solution to a difference equation is different from the concept of a solution to an algebraic equation where the solution is a variable or a number.

A solution to a linear difference equation expresses the value of y_t as a function of the forcing process $\{x_t\}$, t and possibly some known initial condition (y_0) for the $\{y_t\}$ sequence. A solution to a difference equation is itself a function that makes the difference equation true. In other words, given the difference equation (1), we seek to find the primitive function $f(t)$ that satisfies the difference equation for all permissible values of t and $\{x_t\}$. In a time series context, the solution will be a sequence of values that satisfy the difference equation **at all points in time**.

A corollary of the above, is that the substitution of a solution into the difference equation must results in an identity. Consider the following examples:

Example 1 *Deterministic difference equation*

Consider the difference equation

$$y_t = y_{t-1} + 2 \tag{2}$$

A potential solution to this difference equation would be:

$$y_t = 2t + c \tag{3}$$

where c is an arbitrary constant. If (3) is a solution to (2) it must hold for all values of t. As such, for t and t_{-1} we would have:

$$\begin{aligned} y_t &= 2t + c \\ y_{t-1} &= 2(t-1) + c \end{aligned}$$

Substitute the above into (2) yields:

$$2t + c = 2(t-1) + c + 2 \tag{4}$$

which verifies that (4) is an identity and hence (3) a solution to (2).

Example 2 *Stochastic difference equation*

Consider the following law of motion for investment

$$i_t = 0.7i_{t-1} + \varepsilon_t \tag{5}$$

The proposed solution in this case is:

$$i_t = \sum_{i=0}^{\infty} (0.7)^i \varepsilon_{t-i} \tag{6}$$

To check the validity of this solution, we iterate (5) one period backwards:

$$i_{t-1} = \sum_{i=0}^{\infty} (0.7)^i \varepsilon_{t-1-i} \tag{7}$$

Substitute (6) and (7) into (5) yields:

$$\varepsilon_t + 0.7\varepsilon_{t-1} + (0.7)^2\varepsilon_{t-2} + (0.7)^3\varepsilon_{t-3} + ... = 0.7[\varepsilon_{t-1} + 0.7\varepsilon_{t-2} + (0.7)^2\varepsilon_{t-3} + ...] + \varepsilon_t \tag{8}$$

The two sides of (8) are identical, which proves that (6) is a solution to (5).

4.2.2 Iterative Solutions

A solution to a linear difference equation expresses the linear difference equation as a function of $\{x_t\}$, time and possibly some initial condition y_0. How to find the particular solution will depend on whether this initial condition is actually known. The two approaches are:

1. Iterate Forward (with an initial condition)
2. Iterate Backwards (without an initial condition)

(a) Solution by iteration with an initial condition:

If the value of y in some specific period is known, a direct method of solution is to iterate **forward** from that period onwards to obtain the subsequent time path of the entire $\{y\}$. Refer to this known value of y as the initial condition y_0. To illustrate this technique we'll use a first order difference equation:

(i) Deterministic Case: consider the following difference equation

$$y_t = a_1 y_{t-1} + b \tag{9}$$

Lets say a value y_0 is known at $t = 0$. Then at $t = 1$ equation (9) implies that:

$$y_1 = a_1 y_0 + b \tag{10}$$

At $t = 2$:

$$\begin{aligned} y_2 &= a_1 y_1 + b \\ &= a_1(b + a_1 y_0) + b \\ &= a_1^2 y_0 + b(a_1 + 1) \end{aligned}$$

At $t = 3$:

$$\begin{aligned} y_3 &= a_1 y_2 + b \\ &= a_1[a_1^2 y_0 + b(a_1 + 1)] + b \\ &= a_1^3 y_0 + b(a_1^2 + a_1 + 1) \end{aligned}$$

Keep iterating the until we obtain:

$$y_t = a_1^t y_0 + b(a_1^{t-1} + a_1^{t-2} + ... + a_1 + 1)$$

Note that the above expression between brackets can be also written as:

$$1 + a_1 + a_1^2 + ... + a_1^{t-1} = \begin{cases} \sum_{j=0}^{t-1} a_1^j & \text{if } a_1 \neq 1 \\ t & \text{if } a_1 = 1 \end{cases}$$

Making the solution to the difference equation:

$$y_t = \begin{cases} a_1^t y_0 + b \sum_{j=0}^{t-1} a_1^j & \text{if } a_1 \neq 1 \\ y_0 + bt & \text{if } a_1 = 1 \end{cases}$$

Remark 3 *There is only one solution that satisfies both the difference equation and an initial condition. However, there is, in general, an infinite number of solutions to a linear, first-order difference equation.*

The general solution would be given by:

$$y_t = \begin{cases} C a_1^t y_0 + a_0 \left(\frac{1-a_1^t}{1-a_1} \right) & \text{if } a_1 \neq 1 \\ C + a_0 t & \text{if } a_1 = 1 \end{cases}$$

where C stands for an arbitrary constant. In other words, the presence of an initial condition eliminates the arbitrariness of C.

(ii) Stochastic Case: consider the following difference equation

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t \tag{11}$$

where $\{\varepsilon_t\}$ follows some kind of stochastic process, reflecting the uncertainty about the value of y_t.

Given a known value y_0 of $\{y_t\}$, it follows that y_1 will be given by:

$$y_1 = a_0 + a_1 y_0 + \varepsilon_1 \tag{12}$$

Similarly, y_2 must be:

$$\begin{aligned} y_2 &= a_0 + a_1 y_1 + \varepsilon_2 \\ &= a_0 + a_1[a_0 + a_1 y_0 + \varepsilon_1] + \varepsilon_2 \\ &= a_0 + a_0 a_1 + (a_1)^2 y_0 + a_1 \varepsilon_1 + \varepsilon_2 \end{aligned}$$

Repeated iterations yield:

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} \tag{13}$$

Assuming $|a_1| < 1$, $\lim_{t\to\infty} a_1^t y_0 = 0$ then (13) converges to:

$$y_t = \frac{a_0}{(1 - a_1)} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \tag{14}$$

Remark 4 *Should the difference equation miss a constant term a_0, then the sequence would converge to:*

$$y_t = \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \tag{15}$$

(b) Solution by iteration without an initial condition:
When no initial condition is given (or the series is assumed to be infinite), the solution can be found by iterating backwards.

To be completed ...

4.2.3 Steady State and Convergence

An important property of **autonomous** difference equations is that they **often** have a steady state. A steady state is the value of y at which the dynamic system becomes stationary. That is to say, y_{t+1} takes the same value as y_t for all values of t.

Definition 19 *(Steady State) The steady state or stationary value in a difference equation is defined as the value of y at which the system comes to rest. This implies that $y_{t+1} = y_t = \bar{y}\ \forall t$.*

Theorem 1 *(Existence of steady state)*
In a linear, autonomous, first order difference equation there always exist a steady state as long as $a \neq 1$.

Example 3 *Finding the steady state:*
Consider the deterministic difference equation (9). Its steady state would be:

$$\begin{aligned} \bar{y} &= a_1\bar{y} + b \\ \Rightarrow \bar{y} &= \frac{b}{1-a_1} : a_1 \neq 1 \end{aligned}$$

4.2.4 Convergence to steady state

If y ever becomes equal to its steady-state value, it will remain at that value for all successive time periods. The key question is: if y begins at any arbitrary value different from its steady state, will it always tend to converge towards this point? To answer this we need to study the solution to the difference equation.

Theorem 2 *(Convergence of first-order difference equations)*
A linear, first order difference equation will converge to its steady state value (ss) if and only if $|a_1| < 1$.

While convergence is guaranteed if $|a_1| < 1$, the path that y_t will take over time is vary depending on the sign of a_1. In turn, we'll have that:

1. If $0 < a_1 < 1$, y_t converges monotonically towards the ss
2. If $-1 < a_1 < 1$, y_t converges in an oscillating path towards the ss

For all other cases the divergence will be:

1. If $a_1 > 1$, y_t diverges exponentially away from the ss
2. If $-a_1 < -1$, y_t diverges in an oscillating path away from the ss

Key things to remember

(1) How to find an iterative solution to a difference equation:

- If y_0 is known you iterate forwards. If y_0 is unknown, you iterate backwards.

(2) Does a steady state exist? If so, how to find it.

- A linear, autonomous, first-order difference equation always has a steady state so long as $a_1 \neq 1$.

(3) Does the difference equation converge to its steady state? If so, in which way?

- For a first-order difference equation, convergence to the steady state will depend on the sign and magnitude of the coefficient on the lagged variable (a_1).

4.3 Nonstationary Time Series

Nonstationary time series present some particular difficulties and standard inference often fails when a process depends explicitly on t. Nonstationarities can be classified into one of four categories:

- Seasonalities
- Deterministic Trends (also known as time trends)
- Unit Roots (also known as stochastic trends)
- Structural breaks

Each type has unique features. Seasonalities are technically a form of deterministic trend, although their analysis is sufficiently similar to stationary time series that little is lost in treating a seasonal time series as if it were stationarity. Process with deterministic trends have unconditional means which depend on time and unit roots processes have unconditional variances that grow over time. Structural breaks are an encompassing class which may result in either or both mean and variance exhibiting time dependance.

4.4 Achieving Stationarity

Economic Time Series often violate our assumption of covariance stationarity. In particular, their mean is typically changing over time. For example, the average value of U.S. GDP in the 2000s is much higher than the average value that at the dawn of the 20th century.

One way of dealing with this type of nonstationarity is by using stationary-inducing transformations of the data. A preliminary transformation often used is to take logs of the time series. This is useful because most macro data typically grow in an exponential fashion and if a series grows with an exponential trend, taking logs will make the trend linear. Moreover, taking logs also deals with certain types of heteroskedasticity.

4.4.1 First-differencing

A first alternative is to take first differences after taking logs of the data. In turn, we define a new variable Δy_t:

$$\Delta y_t = y_t - y_{t-1}$$

This transformation almost always induces stationarity for processes that have means (in log levels) that change over time in a systematic way.

Note that because the data has been logged, Δy_t measures the growth rate of the variable. This is because the log-difference transformation of a variable represents that variable in terms of its growth rates. Most growth rates of economic variables are stationary.

4.4.2 Removing deterministic trend

An alternative approach to induce stationarity for processes that grow over time is to remove a deterministic trend from their logged values. For example, removing a linear trend means taking the residuals from the following regression:

$$u_t = y_t - \hat{\mu} - \alpha t$$

In other words, u_t would be the resulting stationary series. One could also add quadratic, cubic, etc. terms to this regression to remove non-linear trends. In practice, removing these higher order terms is not commonly done.

4.4.3 Filtering

A third alternative is to use some sort of filter to remove the nonstationarities. Examples of such filters are the Hodrick-Prescott (HP) filter or some band-pass filer. A thorough discussion of these methods requires getting into the frequency domain (rather than time domain). Intuitively, however, we would take the stationary component as:

$$u_t = y_t - y_t^{HP}$$

where y_t^{HP} represents the trend component of the time series as identified by the HP filter.

References:

- J. H. Stock and M.W. Watson, 2003. Introduction to Econometrics.
- J. H. Stock and M.W. Watson, 2001. Vector Autoregression. Journal of Economic Perspectives 15, 101-115.
- W. H. Greene, 2000. Econometric Analysis.
- J.D., Hamilton 1994. Time Series Analysis.
- W. Enders, 2004. Applied Econometric Time Series.