ECONOMICS 7344, Spring 2023

Bent E. Sørensen

March 21, 2023

1 Short Introduction to Time Series

A time series is a sequence of stochastic variables observed over time. Most of what we do in economic modeling involves time series. If you have any dynamic model, you will likely have agents maximizing expected utility or profits. Time series models are designed to provide convenient, yet flexible, models that allow for analytical or numerical expectations to be calculated. Time series models also allow for convenient, parsimonious models of observed data to be estimated. The modeling and estimation angles are highly complementary and uses the same models, but modeling and estimation still involves different methods. This note is about the basic models and, especially in the homeworks and the application to the PIH model, with a focus on finding expected values of future variables. (We will focus on time series estimation in the econometrics classes.)

A time series is a collection of stochastic variables $x_1, .., x_t, .., x_T$ indexed by an integer value t. The interpretation is that the series represent a vector of stochastic variables observed at equal-spaced time intervals. The series is also some times called a stochastic process.

The distinguishing feature of time series is that of temporal dependence: the distribution of x_t *conditional* on previous value of the series depends on the outcome of those previous observations, i.e., the outcomes are not independent. For the purpose of analyzing a time series we will usually model the time series over all the

non-negative integers: $x_t \; ; t = \{0, 1, .., \infty\}$ or $x_t \; ; t = \{-\infty, .., 0, 1, .., \infty\}$. Time 1 *or* time 0 will be the first period that you observe the series. In a specific model, you will have to be explicit about the initial value, as will be clear from the following.

1.1 Stationarity

Definition A time series is called *stationary* (more precisely *covariance stationary*) if

$$E(x_t) = \mu \; ,$$
$$E[(x_t - \mu)^2] \; = \; \gamma(0) \; ,$$
$$E[(x_t - \mu)(x_{t+k} - \mu)] \; = \; \gamma(k) \; = \; \gamma(-k) \; = \; E[(x_t - \mu)(x_{t-k} - \mu)]; \; k = 1, 2, ... \; ,$$

where $\gamma(k) \; ; \; k = 0, 1..$ are independent of t and finite.

There is a quite long tradition in time series to focus on only the first two moments of the process, rather than on the actual distribution of x_t. If the process is normally distributed all information is contained in the first two moments and most of the statistical theory of time series estimators is asymptotic and more often than not dependent on only the first two moments of the process. Consider a stationary process: the $\gamma(k)$'s for $k \neq 0$ are called **autocovariances** and if we divide by the variance we obtain the **autocorrelations** $\rho(k) = \gamma(k)/\gamma(0)$ (note that this is a special case of a correlation-coefficient $corr(x, y) = cov(x, y)/(\sqrt{var(x)} * \sqrt{var(y)})$ because the variances of x_t and x_{t-k} are identical by stationarity). These are the correlation of x_t with it own lagged values.

Note that if Σ_T is the matrix of variances and covariance of $x_1, ..., x_T$ then

$$\Sigma_T = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(T-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(T-2) & \dots & \dots & \dots & \gamma(1) \\ \gamma(T-1) & \gamma(T-2) & \dots & \gamma(1) & \gamma(0) \end{pmatrix}.$$

So if we let Ω_T be the matrix of autocorrelations, i.e. $\Sigma_T = \gamma(0)\Omega_T$ we will have

$$\Omega_T = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(T-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(T-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho(T-1) & \rho(T-2) & \dots & \rho(1) & 1 \end{pmatrix}.$$

Time series models are simple models for the (auto-) correlation of the x_t's that allow us to parameterize the $T*(T+1)/2$ variances and autocoveriances as functions of a much lower number of parameters.

Definition: A stationary process e_t with mean 0 is called *white noise* if $\gamma(k) = 0$ for $k \neq 0$. Of course this implies that the autocorrelation matrix is just an identity matrix, so the standard OLS assumptions on the error term can also be formulated as "the error term is assumed to be white noise."

1.1.1 The lag-operator, lag-polynomials, and their inverses

The lag operator L is defined by $Lx_t = x_{t-1}$. We will also define the symbol L^k as $L^k x_t = x_{t-k}$. Lag operators can be given a very precise mathematical definition, which we will not do here (we would need much more background math), but for us they first of all serve as a very compact notation, so don't feel mystified.

You should think of the lag-operator as moving the whole process $\{x_t \; ; t = -\infty, ..., \infty\}$. Notice that it is here practical to assume that the series is defined for

all integer t, rather than starting at some period 0—in applications this may not always make sense and in these cases you may have to worry about your starting values. We will define **lag polynomials** as polynomials in the lag operator as follows. Let a(L) be the lag polynomial

$$a(L) \;=\; a_0 \;+\; a_1 L \;+\; ... \;+\; a_p L^p$$

which is defined as an operator such that

$$a(L)x_t \;=\; a_0 x_t \;+\; a_1 x_{t-1} \;+\; ... \;+\; a_p x_{t-p} \;.$$

This simply means that the equation above defines a(L) by the way if operates on x_t.

A key observation is that we can add and multiply lag-polynomials in exactly the same way as we can add and multiply polynomials in complex variables. For example, if $a(L) = (1 - a\,L)$ and $b(L) = (1 - b\,L)$ then

$$a(L)b(L)x_t \;=\; (1-a\,L)(1-b\,L)x_t \;=\; (1-a\,L)(x_t - b\,x_{t-1}) \;=\; x_t - b\,x_{t-1} - a\,L(x_t - b\,x_{t-1})$$

which, after simplifying, gives you

$$a(L)b(L)x_t \;=\; x_t - (a+b)\,x_{t-1} + ab\,x_{t-2} \;=\; (1 - (a+b)\,L + ab\,L^2)x_t \;.$$

Notice, that if we denote $c(L) = 1 - (a+b)\,L + ab\,L^2$, the coefficients of the lag polynomial $c(L)$ are equal to to the coefficients of the polynomial $c(z) = 1 - (a + b)z + abz^2 = (1 - a\,z)(1 - b\,z)$ where z is a real or complex variable. For a given lag-polynomial $a(L)$ we therefore define the corresponding *z-transform* (also a label from the engineering literature) $a(z)$ where z is a complex number. The point is that if we define $a(z)$ and $b(z)$ as the complex polynomials we get from substituting the complex number z for L then we know that $a(z) * b(z)$ is a complex polynomial which we can denote $c(z)$. Because the operations are similar, it is also the case

that $c(L) = a(L) * b(L)$. Conceptually, $c(L)$ is a function that works on a time series and $c(z)$ is totally different animal—we just use $c(z)$ as a quick way to find the coefficients of $c(L)$.

One can invert lag-polynomials (although it is not always the case that an inverse exist, as in the case of matrices). We use the notation

$$a(L)^{-1} \text{ or, occasionally (esp. among engineers), but don't do it, } \frac{1}{a(L)} .$$

$a(L)^{-1}$ is defined as the operator such that $a(L)^{-1}a(L)x_t = a(L)a(L)^{-1}x_t = x_t$ for any series x.

All problems concerning inversion of lag polynomials can be reduced to inversion of the first order polynomial $1-az$—we will show the details for the first order polynomial. You all know the formula (valid for $|a| < 1$)

$$\frac{1}{1-a} \;=\; 1 \;+\; a \;+\; a^2 \;+\; a^3 \;+\; ....$$

(If you are not fully familiar with this equation, then you should take a long look at it, prove it for yourself, and play around with it. It is very important for the following and in other contexts.)

For $|a| < 1$ we get

$$\frac{1}{1-az} \;=\; 1 \;+\; az \;+\; (az)^2 \;+\; (az)^3 \;+\; ....$$

which converges if $|z| = 1$. In order for infinite polynomials like this to converge (when a is numerically smaller than unity), the z-transform $a(z)$ is defined only for complex numbers of length 1, i.e., $|z| = 1$. We can therefore find the inverse of the lag-polynomial $1 - a\,L$ as

$$\frac{1}{1-aL} \;=\; I \;+\; aL \;+\; (aL)^2 \;+\; (aL)^3 \;+\; ....$$

so

$$(1-aL)^{-1}x_t \;=\; x_t \;+\; ax_{t-1} \;+\; a^2x_{t-2} \;+\; a^2x_{t-3} \;+\; ....$$

(Here is where one would need the math to be really rigorous to make precise the conditions under which this infinite sum converges. In almost all case we work with in macro, this is not a problem (for example, if x is stationary, it is clear that the variance and mean converges for a numerically smaller than unity).

Inversion of a general lag polynomial.

It is well known that if $a(z) = 1+a_1z+....+a_kz^k$ is a kth order complex polynomial then

$$a(z) \;=\; (1-\alpha_1 z)(1-\alpha_2 z)...(1-\alpha_k z)$$

where $\frac{1}{\alpha_1}, ..., \frac{1}{\alpha_k}$ are the roots of the polynomial. Recall that the roots will be complex conjugates (although sometimes real) if the coefficients of $a(z)$ are real numbers. Of course, the lag-polynomial $a(L)$ factors the same way. This means that

$$a(L) \;=\; (1-\alpha_1 L)(1-\alpha_2 L)...(1-\alpha_k L)\,.$$

The inverse of the general scalar lag polynomial is now simply defined by inverting the "component" first order lag polynomials one-by-one. Consider the AR-model (to be defined soon) of form

$$a(L)x_t \;= \mu + u_t\,.$$

Inverting the lag-polynomial, we have

$$x_t \;=\; a(L)^{-1}\mu + a(L)^{-1}u_t \;=\; a(L)^{-1}\mu + (1-\alpha_k L)^{-1}....(1-\alpha_k L)^{-1}u_t\;;$$

because we can invert the stable first order lag polynomials one by one. What about the first term? Because μ is a constant, $L\mu = \mu$ so instead if L, you put a number 1 in the lag-polynomial, so you get

$$a(L)^{-1}\mu = \frac{\mu}{a(1)}\,.$$

Let us verify how this works for the stable lag polynomial with one lag:

$$(1-aL)^{-1}\mu = (1+aL+a^2L^2...)\mu = \mu + a\mu + a^2\mu... = (1+a+a^2...)\mu = \frac{1}{1-a}\mu\,.$$

Example: Consider the polynomial $a(L) = (1+.3L-.1L^2)$. This can be factored as $a(L) = (1+.5L)(1-.2L)$. Since $(1+.5L)^{-1} = 1-.5L+.25L^2....$ and $(1-.2L)^{-1} = 1+.2L+.04L^2....$, we have $1+.3L-.1L^2)^{-1} = (1-.5L+.25L^2....)(1+.2L+.04L^2....) = 1-.3L+.28L^2.....$ This means that we can write the stationary model

$$x_t = -.3x_{t-1} + .1x_{t-2} + u_t = u_t - .3u_{t-1} + .28u_{t-2} + ...$$

Finally, notice that to invert the stable *model*, I need to assume that it is stationary because this is equivalent to the model having started at $-\infty$.

The **difference operator** Δ is defined as $\Delta = I - L$ giving $\Delta x_t = x_t - x_{t-1}$.

1.2 AR models:

The most commonly used type of time series models are the auto regressive (AR) models.

$$x_t = \mu + a_1x_{t-1} + ... + a_kx_{t-k} + u_t\,,$$

where the innovation u_t is white noise with constant variance σ^2. Here k is a positive integer called the order of the AR-process. (The terms AR-model and AR-process are used interchangeably.) Such a process is usually referred to as an AR(k) process. Note that μ here is a constant and not the mean of x_t (we try to use somewhat consistent notation, but we do not reserve any particular symbols for one single use.)

In empirical work, it is almost always (implicitly) assumed that x_t and the lagged x-variables are observed, while the innovations u_t are not. The a coefficients can

be estimated by regression (or maximum likelihood).

An AR model is a model of correlations between observations, but in macro-modeling, we usually think of time as progressing and we are interested in conditional means (or variances) at t conditional on where we were at $t-1$. (For a statistician, it does not matter if time was going backwards, and the model were $x_t = \mu + ax_{t+1} + u_t$ where u_t is uncorrelated with all *future* observations. The mean and standard deviations that is used for estimation are the same as for the model $x_t = \mu + ax_{t-1} + e_t$, where e_t is uncorrelated with all *past* observations. Can you see that? It follows by logic, but if you are not comfortable with that, verify it using the formula for the conditional mean in the normal distribution.) The AR(k) model directly give this conditional expectation:

$$E_{t-1}x_t \;=\; \mu \;+\; a_1x_{t-1} \;+\; ... \;+\; a_kx_{t-k}\;.$$

When we maximize, say, utility, over the future expected outcomes, we need to do some work. We will illustrate that in the case of the PIH-model later.
Most of the intuition for AR processes can be gained from looking at the AR(1) process, which is also by far the most commonly applied model. Consider the process:

$$x_t \;=\; \mu \;+\; ax_{t-1} \;+\; u_t \;\;(*)\;.$$

If $|a| \;<\; 1$ the process is called *stable*. For instance, assume that x_0 is a fixed number. Then $x_1 = \mu + ax_0 + u_1$ using (*). You can use (*) again and find $x_2 = (1+a)\mu + a^2x_0 + au_1 + u_2$. Continuing this (often called "iterating" equation (*) or "recursive iteration" or (best) "recursive substitution") you get

$$x_t = \frac{1-a^t}{1-a}\mu + x_0a^t \;+\; a^{t-1}u_1 \;+\; ... \;+\; au_{t-1} + u_t\;.$$

If the u_t error terms are i.i.d. with variance σ^2 then

$$var(x_t) \;=\; \sigma^2[1 + a^2 + ... + (a^2)^{(t-1)}] \;=\; \frac{\sigma^2\,(1 - a^{2t})}{1 - a^2} \;\to\; \frac{\sigma^2}{1 - a^2}\;,$$

for $t \to \infty$.

Notice that this process is not stationary if x_0 is fixed (as the variance varies with t). However, for large t the process is approximately stationary (the variance is approximately $\sigma^2/(1 - a^2)$ independently of t), and if the process was started at $-\infty$ then x_0 would be stochastic with variance $\frac{\sigma^2}{1-a^2}$. So the stationary AR(1) is a process that we imagine started in the infinite past. Using the lag operator methodology, we have (for $\mu = 0$ for simplicity)

$$(1 - aL)x_t = u_t \;\Rightarrow\; x_t = (1 - aL)^{-1}u_t \;\Rightarrow\; x_t = u_t + au_{t-1} + a^2u_{t-2} + ...$$

Compare this to what we obtained from iterating the process, and you can see that lag-operator manipulations in essence is nothing but a convenient way of iterating the process all the way back to the infinite past.

Notice that when we simulate complicated dynamic models and want a stationary distribution (assuming the model has one), we often start from a fixed value and simulated future outcomes. Then to consider the stationary distribution, we discard the first many observations (where "many" depends on the actual model, the closer it is to being non-stationary, the more you need). However, for simple AR models, we can directly draw the first observation as a random variable with the stationary mean and variance and as you iterate future values, they will have this same mean and variance.

Notice what happens if $a \;\to\; 1$: the variance tends to infinity and the stationary initial condition can not be defined - at least in typical case as where $u_t \sim N(0, \sigma^2)$

since a normal distribution with infinite variance is not well defined.
The AR(1) process

$$x_t \;=\; x_{t-1} \;+\; u_t \;,$$

with u_t iid white noise, is called a *random walk*. Random walk models (and other models with infinite variance) used to be considered somewhat pathological (they also behave "strangely" when used in regression models). Hall's 1998 demonstration that consumption is a random walk (more precisely a martingale) in a not-too-farfetched version of the PIH model, was therefore a major surprise for the profession. (The difference between a random walk and a martingale is mainly that in a random walk the variance of the u_t term is constant, which it need not be for a martingale.)

It is commonly assumed in modeling (based on empirical work) that log wages can be described as a deterministic component (a function of age and education, in particular) and a random component which is the sum of a random walk and a white noise component. (The random walk component may capture a regular job with employers setting annual wage increases, rather than levels, and the white noise component capturing one-off opportunities or misfortunes.)

The AR(k) model can also be written

$$a(L)x_t \;=\; \mu \;+\; u_t \;,$$

where the log-polynomial $a(L)$ associated with the AR(k) model is

$$a(L) = 1 - a_1 L - ... - a_k L^k \;.$$

(Note, I here us the term "associated with," but you may encounter different ways of denoting the lag-polynomial "of" the AR(k) model. This usually doesn't create

confusion.)

We say that the AR(k) model is *stable* if all roots of the lag-polynomial $a(L)$ are outside the unit circle.

Example: If you are given for example an AR(2) process, like

$$x_t \;=\; 1.5x_{t-1} \;+\; x_{t-2} \;+\; u_t \;,$$

you should be able to tell if the process is stable. In the example we find the roots of the polynomial $1 - 1.5z - z^2$ to be

$$r_i \;=\; .5 * (-1.5 \pm \sqrt{2.25 + 4}) \;,$$

so the roots are –2 and .5. Since .5 is less than one in absolute value the process is not stable.

Stability versus stationarity:

Theorem: A stationary model is stable.

The same theorem said differently: If a process is not stable it is not stationary.

The proof (for the AR(1) case) follows from the derivation above. For a non-stable model the variance goes to infinity so it cannot be constant and finite.
The "opposite" statement is not true. It is possible for a model to be stable, but not stationary. E.g., this will be the case for a stable model that starts from some (usually fixed) x_0 in period 0, unless x_0 is a random variable with exactly the right

mean and variance.

1.3 MA models:

The simplest time series for modeling are the moving average (MA) models:

$$x_t \;=\; \mu \;+\; u_t \;+\; b_1 u_{t-1} \;+\; ... \;+\; b_p u_{t-p} \;=\; \mu \;+\; b(L)u_t,$$

where the innovation u_t is white noise and the lag-polynomial is defined by the equation. Observe that for the MA-model, μ is the mean of x_t (which is not the case for the AR-model). The positive integer p is called the **order** of the MA-process. MA processes are quite easy to *analyze* because they are given as a sum of independent (or uncorrelated) variables. In computational economics, you will typically draw the innovations from a random number generator.

MA-models are not so easy to *estimate* econometrically: since it is only the x_t's that are observed, the u_t's are unobserved, i.e., latent variables, that one cannot regress on.

Consider the simple scalar MA(1)-model (I leave out the mean for simplicity)

$$(*)\; x_t \;=\; u_t \;+\; b u_{t-1}\;.$$

If u_t is an independent series of $N(0, \sigma_u^2)$ variables, then this model implies that x_t has mean zero and autocovariances: $\gamma(0) = (1+b^2)\sigma_u^2$; $\gamma(1) = \gamma(-1) = b\sigma_u^2$; $\gamma(k) = 0$; $k \neq -1, 0, 1$. (Notice that I here figured out what the model says about the distribution of the observed x's. In economic models, the u_t terms may have an economic interpretation, such a wage raise or increment in productivity, but in many empirical applications of time series the MA- (or AR-) model simply serves

as a very convenient way of modeling the autocovariances of the $x-$ variables.) The autocorrelations between the observations in the MA(1) model are trivially 0, except for $\rho(-1) = \rho(1) = \frac{b}{1+b^2}$.

Consider equation (*) again. In lag-operator notation it reads

$$x_t \ = \ (1 \ + \ bL)u_t \ ,$$

which can be inverted to

$$u_t \ = \ (1 \ + \ bL)^{-1}x_t \ = \ x_t \ - \ bx_{t-1} \ + \ b^2x_{t-2} \ - \ ...$$

It is quite obvious that this expression is not meaningful if $|b| \geq 1$ since the power term blows up (think of the stationary case, where it is clear that the variance of $b^k x_{t-k}$ goes to infinity). In the case where $|b| < 1$, the right hand side converges to a well defined random variable (under standard assumptions).

Definition: The scalar MA(q) model is called *invertible* if all the roots of the lag-polynomial b(L) (strictly speaking the corresponding z-transform b(z)) are outside the unit circle.

An ARMA model is a model that has both AR- and MA-components. For example, the model

$$x_t = .3x_{t-1} + .1x_{t-2} + u_t + .4u_{t-1}$$

is called an ARMA(2,1) model. The ARMA model inherits the properties of AR and MA models. An ARMA-model, using lag-operators, takes the form

$$a(L)x_t = \mu + b(L)u_t \ ,$$

where $a(L)$ and $b(L)$ are lag-polynomials, which almost always are of finite order. If the AR-component of the ARMA-model is stable, we say the ARMA model is

stable and if it is stationary and defined for all t, we can write the model as on infinite MA model:

$$x_t = \mu' + a^{-1}(L)b(L)u_t \ ,$$

where the constant $\mu' = a^{-1}(1)\mu$. This is a result we will need when covering the PIH-model in details. In economics, we often start from an AR or ARMA model which describes behavior and find the corresponding MA model:

$$x_t \ = \ \mu \ + \ u_t \ + \ b_1 u_{t-1} \ + \ ... \ + \ b_k u_{t-k} + ... \, .$$

If the u terms in an economic model are fundamental innovations (shocks), it is very common to describe the behavior of the model via "impulse response functions," which is a terms from engineering where the u's are called impulses and the interest is in how future x's are affected. We can directly read that impact on x_t of a one unit shock at $t-1$ is b_1, which implies that a unity shock at period t impacts x_{t+1} by an amount b_1. Similarly, a one unit shock at t impacts x_{t+k} by an amount b_k, etc. The b-coefficients as a function of the lag is called an impulse response function. There are thousands and thousands of papers showing impulse response functions, typically for vector valued time series, but these are derived using the same logic and tools are we here discuss in the univariate case.

If the MA component of the ARMA model is invertible, we can also write x_t as an infinite AR-model:

$$a(L)b^{-1}(L)x_t = \mu'' + u_t \ .$$

In empirical work, x_t is our observed data, and this formula shows how the u_t terms could be found from the x_ts, if those were observed since "way back."

More on recursive substitution. Consider a stationary AR(2) model:

$$x_t = \mu + a_1 x_{t-1} + a_2 x_{t-2} + u_t \, .$$

We can directly observe $E_{t-1}x_t$ by setting the innovation term to zero, but what about $E_{t-2}x_t$? We can use the law of iterated expectation $E_{t-2}x_t = E_{t-2}\{E_{t-1}x_t\}$: we know what $E_{t-1}x_t$ is as a function of x_{t-1} and further lags:

$$E_{t-1}x_t = \mu + a_1 x_{t-1} + a_2 x_{t-2} + u_t \,. \quad (**)$$

Then we take

$$E_{t-2}E_{t-1}x_t = \mu + a_1 E_{t-2}x_{t-1} + a_2 x_{t-2} + u_t \,.$$

Here, $E_{t-2}x_{t-1}$ involves taking an expectation "one step ahead," so it has the same form as (**) with the t-index moved on period back, so we have

$$E_{t-2}E_{t-1}x_t = \mu + a_1(\mu + a_1 x_{t-2} + a_2 x_{t-3}) + a_2 x_{t-2} \,,$$

where the term in parenthesis is the right-hand side of (**) with the innovation set to 0 and the time index "moved on period back." Then we simplify and get

$$E_{t-2}x_t = (1 + a_1)\mu + (a_1^2 + a_2)x_{t-2} + a_2 x_{t-3} \,,$$

where all terms on the right-hand side now are deterministic or have time indices of $t - 2$ or earlier.

Solving economic models, we may be more interested in expected future values. You can "iterate forward" or you can "iterate backwards," as we just did but then move the time index forward which is OK because relations between the x's only depend on the time distance between them, so moving the t index two periods ahead *everywhere*, we get

$$E_t x_{t+2} = (1 + a_1)\mu + (a_1^2 + a_2)x_t + a_2 x_{t-1} \,.$$

It is a super important to be able to find the expectations of future values conditional on present and/or past values as an agent's behavior (consumption, say) depends on present and future expected values (of income, say).