

Week 1 - January 28, 2008
POL 606 Time Series Analysis
Prof. Matthew Lebo
Introduction to Time Series and Difference Equations

Introduction

A time series is a collection of observations made sequentially in time.

Examples:

Economic time series. Time series arise in economics such as share prices on successive days, export totals in successive months, average income per month, unemployment, CPI, consumer confidence ratings, GDP per year, and thousands more.

Famous example is the Beveridge wheat price index which consists of the average wheat price in nearly 50 places in various countries measured in successive years from 1500 to 1869.

Marketing time series. Sales figures in successive weeks or months. Studying these is an important aspect of understanding business – what are the effects of advertising, seasonal trends, etc.

Binary Processes. A special kind of time series when observations can take only one of two values, usually 0 and 1. Occur in communication theory – whether a switch is on or off. There are some of these in political science – especially at the micro-level, e.g. Presidential approval.

Point Processes. A different type of time series where we consider a series of events occurring randomly in time.

Event Counts. For each time point the number of times an event occurs is counted. For example, the number of conflicts that a country is engaged in on a yearly basis. Where event counts end and traditional time series begin can be a fuzzy line – for example, we could count the weekly number of casualties in Iraq and it would look and perhaps behave just like a time series. The more rare the event, the more distinctive is event count analysis from traditional time series analysis.

Time series are also present in the physical sciences, process control, and demographics.

The longest yearly recorded time series is the yearly high water mark of the Nile, dating back about 3000 years.

Terminology

Continuous – a time series is said to be continuous when observations are made continuously in time. A continuous time series may take on only a discrete set of values.

Discrete – a time series is said to be discrete when observations are taken only at specific times, usually equally spaced. We call a series discrete even when the possible values it takes on are continuous.

Most of what we study in political science are discrete time series – though studies in international relations rely more on continuous time series and often *event counts*.

Discrete time series can arise in several ways.

Given a continuous time series we could read off the values at equal intervals (such as approval measures). This may be called a *sampled* series.

Or, a variable may not have an instantaneous value but we can aggregate the values over equal intervals of time. Examples are exports measured monthly or daily rainfall.

Or, some variables are inherently discrete like the daily closing price of the Dow Jones Industrial Average.

The Key Problem:

Much statistical theory is concerned with random samples of independent observations.

The special feature of time series analysis is the fact that successive observations are usually *not* independent and that the analysis must take into account the time *order* of the observations.

When successive observations are dependent, future values may be predicted from past observations.

If a time series can be predicted exactly, it is said to be *deterministic*.

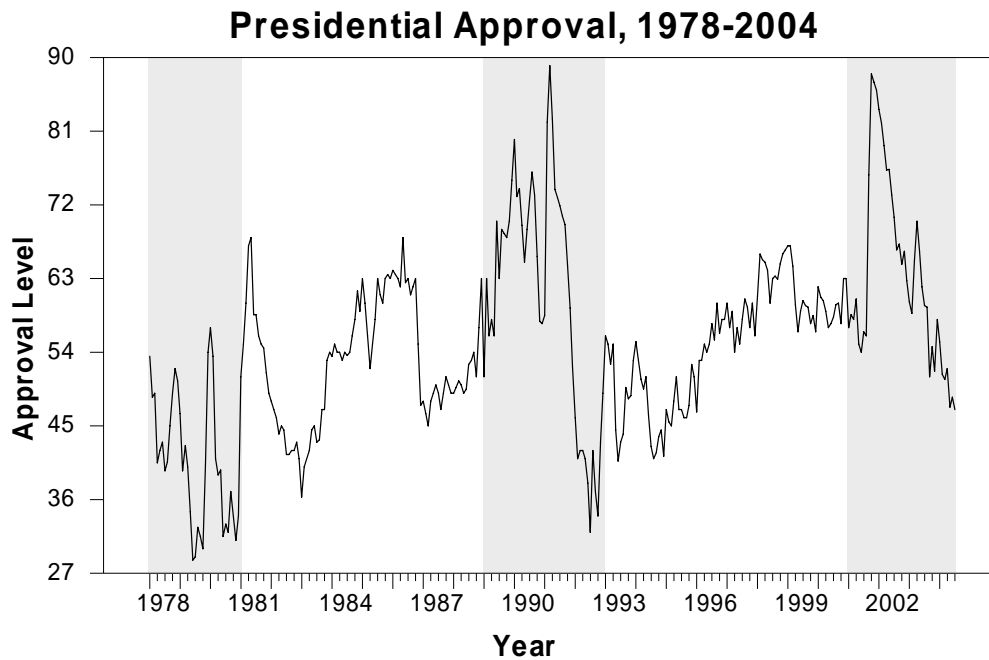
But most time series are *stochastic* in that the future is only partly determined by past values.

For stochastic series exact predictions are impossible and must be replaced by the idea that future values have a probability distribution which is conditioned by a knowledge of past values.

Objectives of Time Series Analysis

Several possible objectives: description, explanation, prediction, and control.

Description. The first step in any analysis is usually to plot the data and obtain simple descriptive measures of the main properties of the series.



Be sure to check for missing data.

Missing data pose a particular problem to time series analysis.

For the rare missing data we can impute using prediction or averages.

For more frequent missing data we can aggregate up to a longer time period.

What are the highs and lows? Average values? Outlying values?

Outlying values in particular are important to note – they may indicate a particularly interesting period.

There is a lot more to this than in cross-sectional data.

We may look for seasonal trends or autoregressive trends.

Is the series *mean stationary*, that is, does it return to a mean that is consistent over time?

Is the series *variance stationary*, that is, is the variance constant over time?

Are there major turning points in the series where suddenly the data structure look to be quite different?

In time series we will learn to model a variable in and of itself.

Explanation. When observations are taken on two or more variables, we can begin the familiar process of building multivariate models.

Prediction or forecasting. Given an observed time series we may want to predict the future values of the series.

This is very useful in economics and – for political scientists – in electoral research.

Control. Not very relevant to political or social scientists. If we can understand what factors affect a process and make predictions about its future values, we can affect those future values by manipulating independent variables in the present. Politicians and economists try this with the macro economy. Worth mentioning here because it a focus of Box and Jenkins (1970).

Data

Data are observed over a discrete period of time such as a month or a year.

This should be consistent over the length of the series – though with public opinion series you take what you can get. E.g. Gallup doesn't conduct its polls on the 15th of every month. But every month there is some Gallup poll that can be used to construct a measure of presidential approval.

Data is constructed into a matrix with the time points as the rows and the variables as the columns. So time points are our cases.

See Week 1 – Introduction to RATS

See also, Week 1 Presidential Approval RATS example.

Difference Equations

Time Series Data and Lag Operators

A time series is a collection of observations indexed by the date of each observation.

We want to study the dynamics of events over time.

Study a variable whose value at date, t , is denoted y_t .

A series begins with a particular date such as $t = 1$ and ends at another such at $t = T$.

Our series can be expressed as:

$$(y_1, y_2, \dots, y_T)$$

And we may find it useful to think about earlier ($y_0, y_{-1}, y_{-2}, \dots$) or later (y_{T+1}, y_{T+2}, \dots) observations than we have.

So we can think of our observed sample as being a finite segment of a “doubly infinite” sequence.

$$\{y_t\}_{t=-\infty}^{\infty} = \{y_{-\infty}, \dots, y_{-1}, y_0, y_1, y_2, \dots, y_T, y_{T+1}, \dots, y_{\infty}\}$$

Only part of this is our observed sample.

*This is one argument against Bayesian criticism.

We can express y_t in terms of y at different periods, the effects of error terms in the present period or at other periods, the effects of independent variables in the present or other periods.

One important time series is a *Gaussian white noise process*,

$$y_t = \varepsilon_t$$

Where $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is a sequence of independent random variables each of which is $N(0, \sigma^2)$.

Time Series Operators

There are several short-hands in time series.

The multiplication operator:

$y_t = \beta x_t$ is a shorthand for an infinite sequence of multiplications, one for each date of t . The operator multiplies the value x takes on at any date t by some constant β to generate the value of y for that date.

Just as if we were discussing y_t the same conventions apply as with scalar multiplication.

Likewise for addition, subtraction, and division.

An operator that is unique to time series is the lag operator, L . (Sometimes called the backshift operator, B).

L means “lag back”.

$$Ly_t = y_{t-1}$$

We can use it to lag back more than one period:

$$L(Ly_t) = y_{t-2}$$

In general, for any integer k ,

$$L^k y_t = y_{t-k}$$

The lag operator and the multiplication operator are commutative:

$$L(\beta y_t) = \beta L y_t.$$

One particularly useful way to use the lag operator is to difference a series:

apply $(1-L)$ to the series y_t .

This gives us:

$$\begin{aligned} y_t - y_{t-1} \\ = \Delta y_t \end{aligned}$$

So, $(1-L)$ is equivalent to Δ , the *difference operator*.

We can describe $(1-L)$ as a filter.

We use it to make a non-stationary series stationary.

Difference Equations

We can relate the value of y at date t to another variable, W_t and the value y took on in the previous period:

$$y_t = \phi y_{t-1} + w_t \quad (1)$$

[$\phi = phi$]

This is a *linear first-order difference equation*.

A difference equation is an expression relating a variable y_t to its previous values, time and other variables.

It's called a *first-order* difference equation because only the first lag of the variable (y_{t-1}) appears in the equation.

An example is Goldfeld's estimated money demand function for the United States.

It related the log of the real money holdings of the public (m_t) to the log of aggregate real income (I_t), the log of the interest rate on bank accounts (r_{bt}), and the log of interest rate on commercial paper (r_{ct}):

$$m_t = 0.27 + 0.72m_{t-1} + 0.91I_t - 0.045r_{bt} - 0.019r_{ct}.$$

This is a special example of (1) where $y_t = m_t$, $\phi = 0.72$ and

$$w_t = 0.27 + 0.91I_t - 0.045r_{bt} - 0.019r_{ct}.$$

Ender's examples of difference equations:

$$T_t = 1 + 0.1t$$

This is a trend. We can see that it accumulates over time. As t gets larger, so does T_t .

$$S_t = 1.6 \sin t \left(\frac{\pi}{6} \right)$$

Here there is seasonality.

$$I_t = 0.7I_{t-1} + \varepsilon_t$$

Is *irregular* in that it contains a stochastic component, ε_t .

Some popular difference equations include the *random walk* which consists of a variable whose changes from one period to the next are stochastic but expected to be zero.

$$y_{t+1} = y_t + \varepsilon_{t+1}$$

We can think of this as a particular version of:

$$y_{t+1} = \phi y_t + \varepsilon_{t+1} \text{ where } \phi = 1.$$

The error term is random "noise" with an expected value of zero.

We can change the expression of this model around.

$$\Delta y_{t+1} = \varepsilon_{t+1}$$

By substituting over and over we can express y_t as a function of its original value plus the sum of all subsequent shocks.

As an accumulation of shocks, it is non-stationary.

We use differencing here to get a stationary series with "*white noise residuals*,"

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

$$y_{t-1} - y_{t-2} = \varepsilon_{t-1}$$

Expected value of ε_t and ε_{t-1} is zero.

What are examples of random walks?

The value of a stock, lots of economic variables such as the money supply, and the national debt.

Try the coin example.

We begin with a score of zero.
In each time period we flip a coin.
If it is heads, we add one.
If it is tails we subtract one.

$$y_0 = 0$$

The coin flips are our random noise component.

So, $y_1 = y_0 + \varepsilon_1$, $y_2 = y_1 + \varepsilon_2$, $y_3 = y_2 + \varepsilon_3$ and so on.

We can plot this over time and see the “random walk.” We begin at each time point where we were at the end of the previous time point.

We meander with no greater reason to expect a return to zero than there is a move further away from zero.

Such a series is non-stationary in the mean.

A lot fewer political variables are random walks than was once thought.

Also, there is the random walk with drift model:

$$y_t = a + 1.0y_{t-1} + \varepsilon_t$$
$$y_{t-1} = a + 1.0y_{t-2} + \varepsilon_{t-1}$$

This is identical to the random walk except for the added constant.

a allows some trending over time.

Why? Note that substituting for y_{t-1} in the first equation gives us:

$$y_t = a + 1.0(a + 1.0y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$
$$y_t = a + a + y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$
$$y_t = 2a + y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

Try the coin flip experiment again.

At time 0 we have a value of zero. Let $a = 0.5$.

We have the same results as in the random walk experiment but we are adding a constant with each time point – thus we expect a trend upwards even when the disturbances have an expected value of zero.

For the random walk with drift model, substitute in and we will get:

$$y_t = a * t + y_0 + \sum_{j=1}^t \varepsilon_j$$

That is, y_t = some initial value plus the accumulation of shocks.

These models have “perfect memory” in that the shocks do not dissipate over time.

No “discounting.”

Random walk models are also called “stochastic trend models.”

Coefficient of 1.0 on lagged y is “unit root” coefficient.

Long-run behavior are very different between deterministic and stochastic.

For short-term forecasts the differences are minimal but for long-term the differences are substantial.

If these models are non-stationary, where are the stationary series?

When we difference a random walk series and look at the changes from one time point to the next.

Instead of using as our dependent variable, y_t accumulating over time, we will use as our dependent variable the changes in y_t from one time point to the next.

The mean of the differenced series is constant – 0. Its expected value doesn’t change over the course of the series.

A real world example is the Dow Jones.

It accumulates from day to day but the differences are mean stationary.

Reduced-Form and Structural Equations

System of equations:

$$\begin{aligned} y_t &= c_t + i_t & 0 < \alpha < 1 \\ c_t &= \alpha y_t - 1 + \varepsilon_{ct} & \beta > 0 \\ i_t &= \beta(c_t - c_{t-1}) + \varepsilon_{it} \end{aligned}$$

y_t, c_t, i_t are real GDP, consumption, and investment in time period t .

The endogenous variable for each equation is on the left-hand side of the equation.

y_{t-1} & c_{t-1} are lagged endogenous variables.

ε_{ct} & ε_{it} are error terms with means = 0.

α & β are parameters to be estimated.

The first equation explains aggregate GDP as the sum of consumption and investment. There is no error here – this is an *identity* equation – that is, it is true by definition.

The second equation tells us that consumption in the present period depends on income in the previous period plus some random disturbance.

The third equation is the “accelerator principle.” Investment is proportional to the change in consumption – as consumption grows investment will grow.

The third (and the first) equation is a **structural equation** because it expresses an endogenous variable, i_t , as a function of the current realization of another endogenous variable, c_t .

A **reduced-form** equation is one that expresses a variable in terms of its own lags, lags of other endogenous variables, current and past values of exogenous variables, and error terms.

The consumption function is already in reduced form – current consumption depends only on lagged income and the current value of the stochastic disturbance term ε_{ct} .

We can turn the investment function into a reduced form equation by substituting the second equation into the third to get:

$$\begin{aligned} i_t &= \beta[\alpha y_{t-1} + \varepsilon_{ct} - c_{t-1}] + \varepsilon_{it} \\ &= \alpha\beta y_{t-1} - \beta c_{t-1} + \beta\varepsilon_{ct} + \varepsilon_{it} \end{aligned}$$

Also, and more importantly, we can get a reduced form equation for the first equation, GDP:

$$y_t = c_t + i_t$$

becomes:

$$\begin{aligned} y_t &= \alpha y_{t-1} + \varepsilon_{ct} + \alpha\beta(y_{t-1} - y_{t-2}) + \beta(\varepsilon_{ct} - \varepsilon_{ct-1}) + \varepsilon_{it} \\ &= \alpha(1 + \beta)y_{t-1} - \alpha\beta y_{t-2} + (1 + \beta)\varepsilon_{ct} + \varepsilon_{it} - \beta\varepsilon_{ct-1} \end{aligned}$$

This last equation is important because it is a *univariate* reduced-form equation.

y_t is expressed only as a function of its own lags and disturbance terms.

This is a very important type of model to us because it allows us to model a series in terms of its own past history and to make predictions without the use of independent variables. What Hibbs called “models of ignorance.”

Difference Equations and their Solutions

Begin with the function $y = f(t)$.

When t takes on a specific value t^* we get a specific value for the dependent variable, y_{t^*} .

So y_{t^*+h} is the value of y when $t = t^*+h$.

The first difference of y is the value of the function when evaluated at $t = t^*+h$ minus the value of the function evaluated at t^* :

$$\begin{aligned}\Delta y_{t^*+h} &\equiv f(t^*+h) - f(t^*) \\ &\equiv y_{t^*+h} - y_{t^*}\end{aligned}$$

We are usually interested in the case where $h=1$.

So that,

$$\begin{aligned}\Delta y_t &= f(t) - f(t-1) \equiv y_t - y_{t-1} \\ \Delta y_{t+1} &= f(t+1) - f(t) \equiv y_{t+1} - y_t\end{aligned}$$

We can also use second differences, where the second differences are the changes between the first differences:

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

Second differences aren't used in political science – its hard to think of an example where they would be necessary.

To need them, the first differences would have to be non-stationary.

For example, the rate of change would be a random walk.

Likewise, knowing how to compute the n^{th} difference isn't very useful to us.

Forcing Process

The forcing process is the collection of terms that are neither constants nor autoregressive parameters.

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

In the above, $\sum_{i=1}^n a_i y_{t-i}$ are the autoregressive processes determining y_t .

This portion of the equation will tell us how y_t is determined by its own past values.

x_t is the forcing process.

It can be any function of time (like a trend), current and lagged values of other variables, and/or stochastic disturbances.

Solving Difference Equations

Means expressing the current realization of a variable in terms of independent variables, t , error terms, and any given initial conditions.

For example, using the first-order difference equation:

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

If we have a value for y_0 , we can express y_1 as:

$$y_1 = a_0 + a_1 y_0 + \varepsilon_1$$

And then express y_2 as:

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

Which eventually becomes:

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

This is a solution of the first equation because it expresses y_t as a function of t , the forcing process $x_t = \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}$, and the known value of y_0 .

One thing to pay attention to here is the auto-regressive effects over time.

Page 13 Figure 1.2 to see how the effect of y in period t continues to affect a series over time.