STATISTICAL HORIZONS

Time Series Analysis

Jon Pevehouse, Ph.D.

Upcoming Seminar:

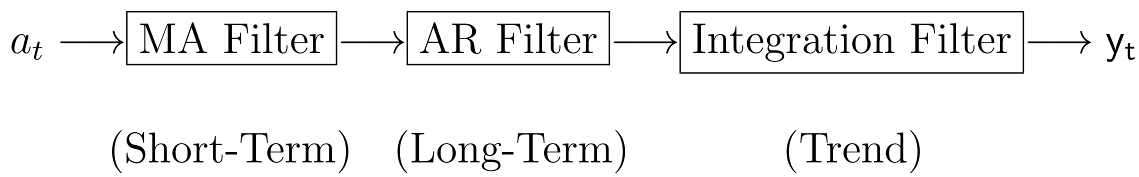
August 3-6, 2021, Remote Seminar

ARIMA models

Foundations and Estimation

Jon C.W. Pevehouse, University of Wisconsin-Madison

White noise drives ARIMA models



$$E(a_t) = E(a_{t-1}) = \cdots = 0 \qquad (2.4)$$

$$E(a_t^2) = E(a_{t-1}^2) = \cdots = \sigma^2 \qquad (2.5)$$

$$E(a_t a_{t-j}) = E(a_{t-k} a_{t-k-j}) = 0 \text{ for all } j \text{ and } k.$$

ARIMA models: notations

1. White noise is the driving force of all ARIMA (p, d, q) models
2. There exists a fundamental stability to social/mechanical processes and recent inputs are more important than previous observations
   - **Integrated processes**: realized as the sum of all past shocks, represented by $(0, d, 0)$ models. (The middle term in AR**I**MA)
   - **Auto-regressive (AR) processes**: realized as exponentially weighted sum of past shocks, represented by $(p, 0, 0)$ models. (The first term in **AR**IMA)
   - **Moving average (MA) processes**: finite persistence – a random shock enters the system and then persists for no more than q observations before vanishing entirely. (The last term in ARI**MA**)

ARIMA models: notations

3. Partitioning:

$$a_t \rightarrow \boxed{\text{MA filter}} \rightarrow \boxed{\text{AR filter}} \rightarrow \boxed{\text{Integration filter}} \rightarrow y_t$$

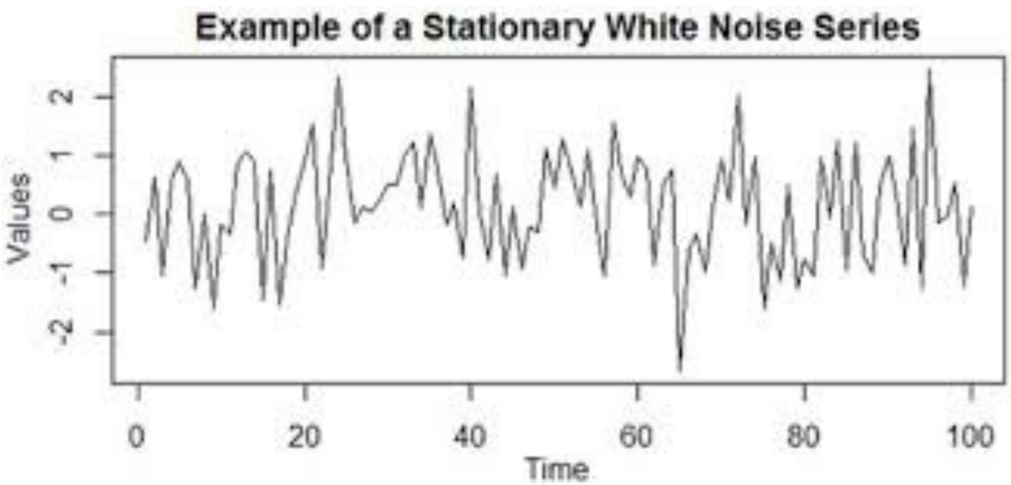
(short-term) (long-term) (trend)

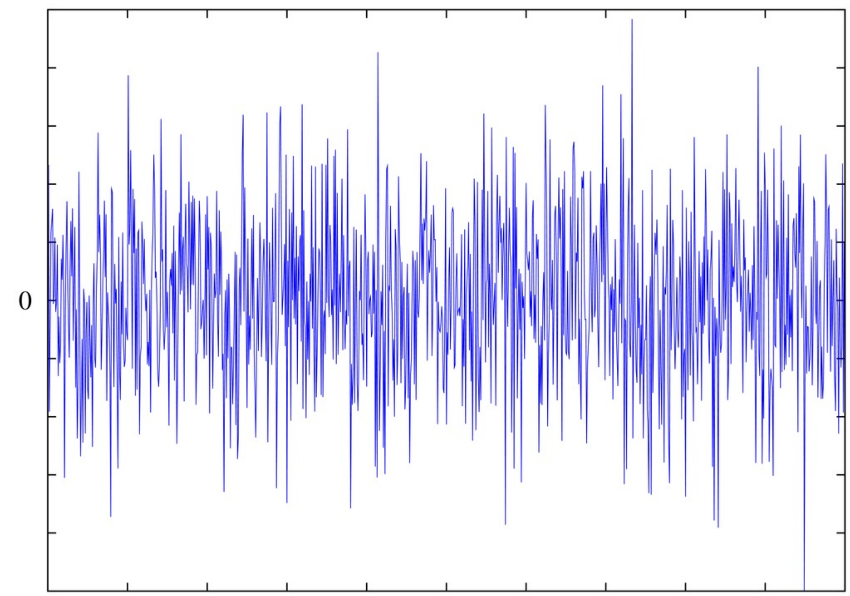
4. Formally:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \ldots + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots$$

where $a_t \sim NID(0, \sigma_a^2)$, i.e. "white noise"

What does white noise look like?





Basic motivation

The basic motivation of this approach is to do better than models of simple or sophisticated smoothing and extrapolation.

- Describe important features of the time series pattern
- Explain how past realizations explain the future, or how series interact
- Produce better forecasts

Some additional concepts to help motivate…

Trend is a nuisance that we want to get rid of (a very long-term pattern of movement in the same direction, e.g. population growth, prices always increasing, etc.).

Drift is a shorter-term movement that we want to model.

OLS model vs. time-series approach:

OLS Trend Equation:

$$\hat{y} = \hat{b_0} + \hat{b_1} t$$

Difference Equation of Trend:

$$\hat{y} = y_{t-1} + \hat{\theta}_0$$

These equations describe exactly the same deterministic trend (when $\hat{b_1} = \hat{\theta}_0$) but there is a major practical difference between them:

- No satisfactory way to estimate b_1 in the OLS equation
- The analogous parameter in the Difference Equation can be estimated as the estimated mean of the differenced series:

$$\hat{\theta}_0 = \bar{z} = \frac{1}{N} \sum_{t-1}^{N} z_t$$

Example...

Difference Equation of Trend:

$$\hat{y} = y_{t-1} + \hat{\theta}_0$$

Note that this is essentially a difference equation where

$$y_t = 1 * y_{t-1} + \theta_0$$
$$y_t - y_{t-1} = \theta_0$$

And notice that without the trend, the original form of the difference equation of trend becomes

$$y_t = y_{t-1} + a_t$$

Example...

So we have:

$$
\begin{aligned}
y_0 &= y_0 \\
y_1 &= y_0 + a_1 \\
y_2 &= y_1 + a_2 = y_0 + a_1 + a_2 \\
y_3 &= y_2 + a_3 = y_0 + a_1 + a_2 + a_3 \\
&\vdots \\
y_t &= y_{t-1} - a_t = y_0 + a_1 + a_2 + a_3 + \ldots + a_t
\end{aligned}
$$

The big picture point is that this model will outperform simple extrapolation **if** it is correctly specified.

Problem estimating OLS trend equation:

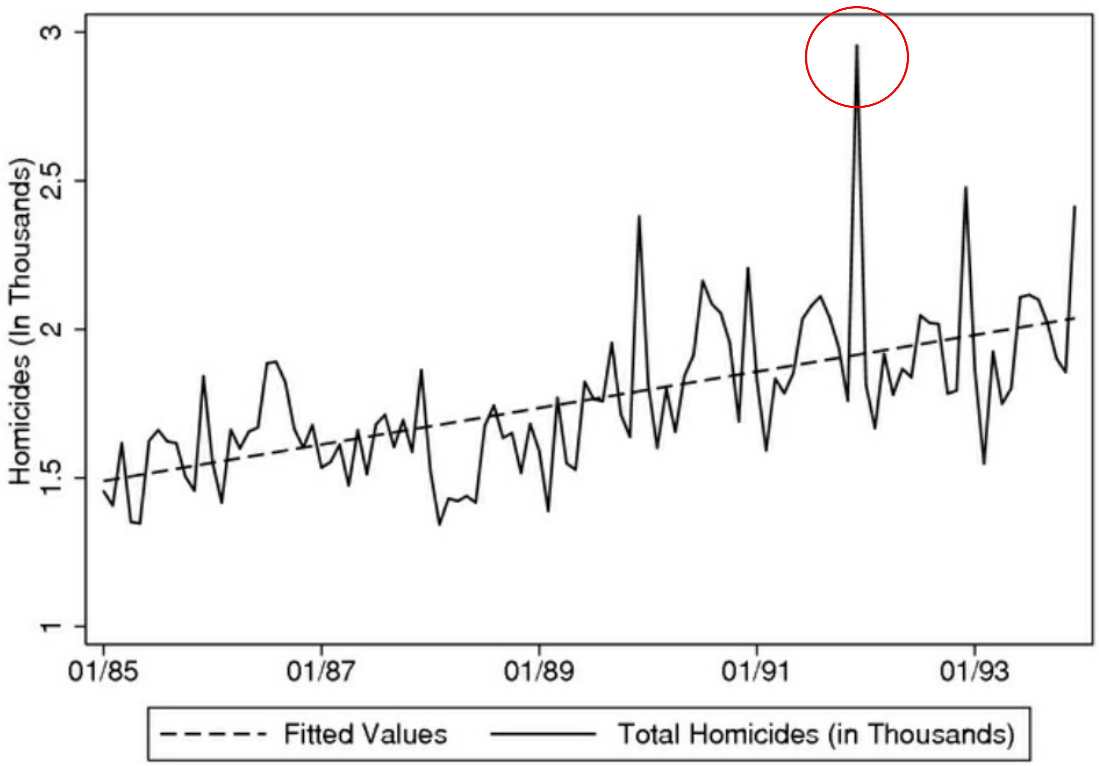
The OLS Trend Equation ($\hat{y} = \hat{b_0} + \hat{b_1} t$) is problematic for a number of reasons:

- Cannot be estimated with much accuracy
  - Sensitive to outliers
  - Sensitive to y_1 and y_N (the sum of squares overestimates the first and last observations)

    Generally speaking, we would like to have *dynamic*, not static, estimates of the trend

    * Each observation in OLS is given equal weight, whereas we want more recent observations to matter more
    * The trend line should fit the beginning, middle, and end of the series (rather than just the two end points)
- This approach is subject to ocular pitfalls – given an observed realization of finite length, it's hard to distinguish between drift (stochastic) and trend (deterministic)

Monthly U.S. Homicide Statistics



Box-Jenkins approach:

1. Study generic forms and properties
2. Armed with general, theoretical knowledge, examine your data to see which possibility applies
   - Identify, then estimate
3. Assess your guess
   - Diagnosis (may iterate again)
4. Meta-analysis
   - R^2, RMS, over-modeling, under-modeling

General auto-regression (AR) process

Two ways of describing an AR(p) process, where p is the order of autoregression:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + a_t \qquad \text{Lag}$$

Equivalently:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) y_t = a_t \qquad \text{Backward Shift}$$

AR(1) Process

Consider an AR(1) or ARIMA$(1, 0, 0)$:

$$y_t = \phi_1 y_{t-1} + a_t, \quad \text{or equivalently}$$
$$(1 - \phi B) y_t = a_t$$

This is similar to an OLS regression, with the additional stipulation that $-1 < \phi < 1$.

Why do we have these bounds on ϕ?

AR(1) Process

Consider an AR(1) or ARIMA$(1,0,0)$:

$$y_t = \phi_1 y_{t-1} + a_t, \quad \text{or equivalently}$$
$$(1-\phi B)y_t = a_t$$

This is similar to an OLS regression, with the additional stipulation that $-1 < \phi < 1$.

Why do we have these bounds on ϕ?

Because we want ϕ to get smaller, not larger, when it's exponentiated. Without these bounds, past values won't decay in significance.

When $\phi_1 = 1$, the impact of distant shocks doesn't diminish over time – a property of a random walk.

When $\phi_1 > 1$, past shocks get larger and larger. This violates stationarity – leads to nonconstant variance and explosive behavior of the series as past shocks get amplified more and more over time.

AR(1) Process

Notice that...

$$
\begin{aligned}
y_0 &= a_0 \\
y_1 &= \phi_1 y_0 + a_1 \\
&= \phi_1 a_0 + a_1 \\
y_2 &= \phi_1 y_1 + a_2 \\
&= \phi_1(\phi_1 a_0 + a_1) + a_2 \\
&= \phi_1^2 a_0 + \phi_1 a_1 + a_2 \\
y_3 &= \phi_1 y_2 + a_3 \\
&= \phi_1(\phi_1^2 a_0 + \phi_1 a_1 + a_2) + a_3 \\
&= \phi_1^3 a_0 + \phi_1^2 a_1 + \phi_1 a_2 + a_3 \\
&\vdots \\
y_t &= \phi_1 y_{t-1} + a_t \\
&= \phi_1^t a_0 + \phi_1^{t-1} a_1 + \phi_1^{t-2} a_2 + \ldots + a_t
\end{aligned}
$$

An ARIMA$(1, 0, 0)$ process is essentially an **infinite sum of exponentially weighted random shocks**:
$y_t = \sum_{i=0}^{\infty} \phi_1^i a_{t-1}$

Parameter bounds on Φ

Given that an ARIMA$(1, 0, 0)$ is an infinite sum of exponentially weighted random shocks:

$$y_t = \sum_{i=0}^{\infty} \phi_1^i a_{t-1}$$

Consider again our bounds on ϕ_1, $-1 < \phi_1 < 1$:

$$y_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \ldots + \phi_1^k a_{t-k} + \ldots$$

(where $\phi_1^k a_{t-k} + \ldots$ is truncated because ϕ_1^k goes to zero rapidly.)

This lets us see clearly how past values only diminish in effect when ϕ_1 becomes smaller when exponentiated.

Relationship to stationarity…

$$y_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \ldots + \phi_1^k a_{t-k} + \ldots$$

Let $\phi_1 = 1$, then $(1 - B)y = a_t$, and we get:

$$y_t = a_t + (1)a_{t-1} + (1)^2 a_{t-2} + \ldots$$

$\rightarrow$ Note that the impact of distant shocks does **not** diminish over time.

Now let $\phi_1 = 1.5$, then $(1 - 1.5B)y_t = a_t$, and we get:

$$y_t = a_t + 1.5a_{t-1} + 2.25a_{t-2} + \ldots + (1.5)^k a_{t-k} + \ldots$$

$\rightarrow$ Now past shocks become more and more important.

The moving average (MA) process

Moving average (MA) processes – e.g. MA(1) / ARIMA$(0, 0, 1)$ – capture the idea of **finite persistence**.

$$y_t = a_t - \theta_1 a_{t-1}, \quad \text{or equivalently}$$
$$y_t = (1 - \theta_1 B) a_t$$

McCleary and Hay demonstrate that this implies:

$$y = a_t - \sum_{i=1}^{\infty} \theta_1^i y_{t-1}$$

Consequently, an ARIMA$(0, 0, 1)$ process can be expressed as an **infinite sum of exponentially weighted past observations**.

MA (1) process

$$y_t = a_t - \theta_1 a_{t-1}$$

Now apply the backshift operator…

$$y_{t-1} = a_{t-1} - \theta_1 a_{t-2}$$

Solve for a_{t-1} and substitute back into previous equation….

MA(1) process

$$y_t = a_t - \theta_1(y_{t-1} + \theta_1 a_{t-2})$$

$$= a_t - \theta_1 y_{t-1} - \theta_1^2 a_{t-2}.$$

Express y_{t-2} in terms of a_{t-2} and substitute back into the equation for y_t yields:

$$y_t = a_t - \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \theta_1^3 a_{t-3}.$$

Continuing this process infinitely yields:

$$y_t = a_t - \sum_{i=1}^{\infty} \theta_i^i y_{t-i}.$$

MA Process stationarity

Like with the ϕ parameter in an AR process, we again require that $-1 < \theta_1 < 1$. This is done to prevent "explosiveness." Values outside this range may indicate a need for differencing (or alternatively that there has been too much differencing).

Equivalence of MA and AR process

To this point, AR processes have been expressed in terms of the sum of shocks, while MA processes have been expressed in terms of sums of past observations.

Invertibility: If conditions are met, a *finite* order MA process has an equivalent AR process of *infinite* order. Similarly, for a stationary AR process of *any* order, there exists an equivalent MA process of *infinite* order.

Equivalence of MA and AR process

Consider an AR(1):

$$y_t = \phi_1 y_{t-1} + a_t$$
$$y_t = \phi_1 B y_t + a_t$$
$$a_t = y_t - \phi_1 B y_t$$
$$a_t = (1 - \phi_1 B) y_t$$
$$y_t = \left(\frac{1}{(1 - \phi_1 B)} \right) a_t$$

In this instance the AR process is written in terms of shocks, but the y_{t-1} term has been eliminated. This implies that **any AR(1) process is an infinite MA process**.

Consider an MA(1):

$$y_t = a_t - \theta_1 a_{t-1}$$
$$y_t = (1 - \theta_1 B) a_t$$
$$a_t = \left(\frac{1}{(1 - \theta_1 B)} \right) y_t$$

Thus, **any MA(1) process is an infinite AR process**.

For $-1 < r < 1$, the sum converges as $n \to \infty$,in which case

$$S \equiv S_\infty = \sum_{k=0}^{\infty} r^k = \frac{1}{1 - r}$$

Taken together, these results illustrate the equivalence of AR and MA processes.

Identifying an AR(1) or MA(1) process…

How do we determine if the data were generated by an AR(1) or an MA(1) process?

Generally speaking, we have two tools:

- Autocorrelation Function (ACF)
- Partial Autocorrelation Function (PACF)