ARIMA Models

Jamie Monogan

University of Georgia

January 16, 2018

Objectives

By the end of this meeting, participants should be able to:

- Argue why standard regressions on trending series are inappropriate.
- Describe the logic of the Box-Jenkins modeling strategy.
- Define stationarity and describe the processes of a stationary series.
- Explain why Maximum Likelihood Estimation is essential for ARIMA models and how MLE is tailored to time series' needs.
- Identify, estimate, and diagnose ARIMA models.
- Identify and estimate seasonal elements of an ARIMA model.

Trend on Trend Spuriousness

- We also start with error specification because it is essential to accurately testing our theories.
- Imagine that we have a simple hypothesis, say
  - Population$\longrightarrow$GDP
- Our measures of each trend (upward, but it doesn't matter if they are opposite).
- Granger and Newbold thesis: any two variables which trend will produce spurious evidence of causal connection.
- In the negative: It is impossible to observe the absence of a statistical relationship between any two trending series!
- Source: Granger and Newbold. 1974. Spurious Regressions in Econometrics. *Journal of Econometrics*. 2: 111-120.

What's the bias with trending series?

- The definition of unbiased is: $E(\hat{\beta}) = \beta$
- For the case of two trending series we estimate two components:
  1. β itself, and
  2. the ratio of the the two trends.
- Thus $\hat{\beta} = \beta +$ Trend-Ratio
- Thus $\hat{\beta}$ is unbiased if and only if Trend-Ratio $= 0.0$
- It cannot be zero if both series trend.

First Differencing Trending Series?

- First differencing is performing the operation (1-B)z = a
- or simply: $w_t = z_t - z_{t-1}$
- In R: use the "diff" function from the "timeSeries" library
- In Stata:
  - tsset month
  - d.approval
- So, under what conditions is it allowable to regress one trending time series on another?
  - **Never!**

Modeling Form: End Game for Inference & Forecasting

Regression

$$\begin{aligned} y_t &= \{\beta_0 + \beta_1 x_1 + \beta_2 X_2 + \cdots\} + u_t \\ &= \{structure\} + error \end{aligned}$$

The two components are structure and error: Y = structure + error

Box-Jenkins

- y_t = [transfer function] + ARIMA Model
- y_t =f(x) + N_t
- So we start out working on ARIMA models of error aggregation, the N_t, and then later develop transfer functions as tests of theories we care about. (Opposite what we usually do.)
- The causal flow of the transfer function cannot be observed until we successfully model the error processes.

How a Time Series is Produced

- We assume the data generating process is:
- $a_t \longrightarrow$ Linear filter $\longrightarrow z_t$
- A white noise input is systematically filtered into the observed time series.

Getting Back to White Noise

- $a_t \longrightarrow$ Linear filter $\longrightarrow z_t$
- implies
- $z_t \longrightarrow$ Inverse of Linear filter $\longrightarrow a_t$
- So, if we can solve for $z = f(a)$, then we can invert f and produce a (which is white noise).

The Box-Jenkins Procedure

Identification What class of models probably produced z_t?

Estimation What are the model parameters?

Diagnosis Are the residuals, a_t, from the estimated model white noise?

Empirically Identifying the Error Process

- We can infer the data generating process because knowing the mathematics, we know the empirical "signature."
- We will develop the signatures of a family of error aggregation models that are autoregressive (AR), integrated (I), moving average (MA)—and all combinations ARIMA(P,D,Q)

AR(1): A very important special case

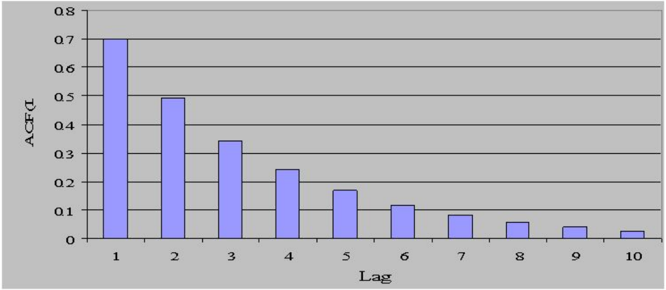
- Notation: AR(1) means autoregressive, 1st order
  - Only the first lag of z appears in the equation
- $z_t = \theta_0 + \phi_1 z_{t-1} + a_t$
- What is its signature? To answer that question it is useful to transform the equation into "shock" form, where z is a function of all previous a's.
- $z_t = \theta_0 + a_t + \phi_1 a_{t-1} + \phi_2 a_{t-2} + \phi_3 a_{t-3} \cdots + \phi_{t-1} a_1$

Simplify

- That's an ugly equation that has T terms, but it has useful information about the expected association of each of the a's with z_t.
- Lag 1: ϕ, that is ϕ^1
- Lag 2: ϕ^2
- Lag 3: ϕ^3
- Lag k: ϕ^k

Autocorrelation Function for AR(1)

- Since ϕ is constrained to be <1.0, that means that each exponential power of ϕ is a progressively smaller number, looking like:



Next Steps

- If we observe an empirical series that shows this (very common) pattern of autocorrelation (and a couple other details), we judge it to be AR(1)
- This is the IDENTIFICATION stage: we're using empirical evidence such as correlograms to determine the error process.
- Once we've tentatively judged the class of model, we estimate the parameters of such a model using MLE (ARIMA ESTIMATION, more later).
- After we estimate the model, calculate the residuals, a_t.
- Now the really neat part: If our judgment was correct, a_t, the estimated residuals must be white noise and we know how to test for that property! (DIAGNOSIS)
- If it is white noise, then we can use this filtered series for our analysis.

Stationarity and Integration

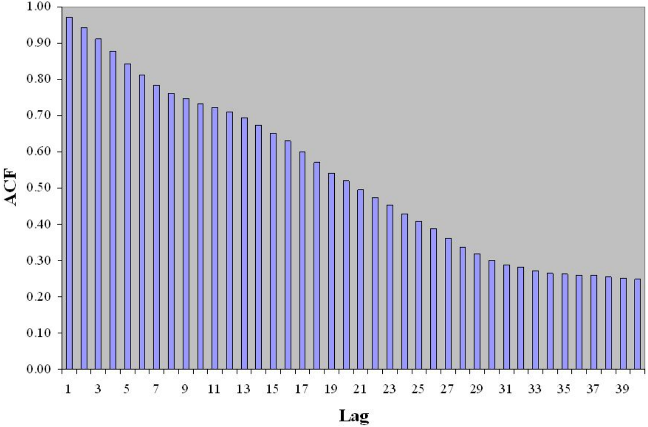
- A stationary series is one that tends to return to some equilibrium level after being disburbed.
- A nonstationary series, or an **integrated** series, has no equilibrium. The most common integrated series is the random walk (i.e., DJIA).
- Box-Jenkins models are defined only for stationary time series.
- The good news: Integrated series can be made stationary by differencing them. (Which you know how to do.)

How to Know Stationarity

- The ACF of a stationary series tends to approach zero after just a few lags
  - And stay there.
- Integrated series show systematic behavior over very long lag lengths.
- In the regression tradition, we will develop the Dickey-Fuller test for unit roots.

Macropartisanship: A Non-stationary Series

Slowly Decaying Autocorrelation Function



Three Models and Their Notation

The AR(1) Model

- Functional: $z_t = \phi z_{t-1} + a_t$, where -1 $< \phi <$ 1
- Backshift: (1 - ϕB)$z_t = a_t$
- Shock form: $z_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3 a_{t-3} + \ldots + \phi^t a_0$

The MA(1) Model

- Functional: $z_t = \theta_0 + a_t - \theta_1 a_{t-1}$
- Backshift: $z_t = \theta_0 + (1 - \theta_1 B)a_t$
- Shock form **is** MA, so there is no expansion

The I(1) Model

- Functional: $z_t = z_{t-1} + a_t$
- Shock form: $z_t = a_t + a_{t-1} + a_{t-2} + a_{t-3} \ldots + a_0$
- Different from AR(1): No decay terms on shocks, *permanent memory*.

The General ARMA(P,Q) Model

$$z_t = \theta_0 + \sum_{i=1}^{p} \phi_i z_{t-i} + \sum_{i=1}^{q} \theta_i a_{t-i} + a_t$$

In econometric notation:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \epsilon_t$$

Some Low Order Models

White noise $z = a$

Random walk $z_t - z_{t-1} = a_t$

- $(1 - B)z = a$
- $\Delta z = a$ [i.e., cumulated white noise]

AR(1) $z_t = \phi z_{t-1} + a_t$

- $(1 - \phi B)z = a$
- Note that $(1 - \phi B) = a$ for ϕ=1.0 is then exactly a random walk. $\phi = 1.0$ is called a unit root.

MA(1) $z_t = a_t - \theta a_{t-1}$

- $z = (1 - \theta B)a$

IMA(1,1) $(1 - B)z = (1 - \theta B)a$

Thus an IMA(1,1) is simply a MA(1) operating on first differences.

ARMA(1,1) $(1 - \phi B)z = (1 - \theta B)a$

Autocorrelation Function Expectations

AR(1)

- $z_t = \phi z_{t-1} + a_t$
- $z_{t-1} = \phi z_{t-2} + a_{t-1}$ (from stationarity)
- $z_t = \phi(\phi z_{t-2} + a_{t-1}) + a_t$ substituting
- $z_t = \phi a_{t-1} + \phi^2 z_{t-2} + a_t$
- $z_t = \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3 z_{t-3} + a_t$
- ...
- $z_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3 a_{t-3} + \ldots \phi^\ell a_{t-\ell}$
- We expect exponential decay in ACF: 1, ϕ , ϕ^2 , ϕ^3 , ...

MA(1)

- $z_t = \theta_0 + a_t - \theta_1 a_{t-1}$
- ACF: $E(\rho_1) = -\theta_1/(1 + \theta_1^2)$
- Crude empirical rule of thumb: ACF(1)= $-\theta_1/2$.
- This implies ACF(1) negative and less than .5 in absolute value.

Cookbook Rules for Identification

Fuller List: Table 6.1 in Box, Jenkins, & Reinsel (2008, 199)

AR(P) Exponential decay in the ACF, P significant spikes in the PACF

I(1) Slow decay in the ACF, 1 significant spike in the PACF

MA(Q) Q significant spikes in the ACF, exponential decay in the PACF

The Partial Autocorrelation Function

- The PACF shows the autocorrelation at lag k *controlling for all previous lags*.
- Thus it shows the effects at lag k which could not have been predicted from lower lags.
- In effect then, it shows the independent effects of processes at lag k.
- PACF(1) = ACF(1)

Least Squares and MLE

- Some ARIMA models, e.g., AR(1), are essentially linear and could be estimated by least squares.
- For example $z_t = \phi_1 z_{t-1} + a_t$ can be estimated by least squares regression if you just drop the first case.
  - R:
    - `z <- ts(data$z1)`
    - `l.z <- lag(z, -1)`
    - `data2 <- ts.union(z, l.z)`
    - `reg.1 <- lm(z~l.z, data=data2)`
  - Stata: `tsset month`, then `reg z l.z`
  - The coefficient on the lagged dependent variable is a LS estimate of ϕ_1
- In practice ARIMA software uses a generalized maximum likelihood algorithm for all ARIMA models.
- The ϕ's estimated by LS and ML are not identical, but the difference is nearly always trivial.
- This is **not** a case like OLS, where LS and ML solutions are proven identical when OLS assumptions hold.

Maximum Likelihood Unmasked

- Maximum Likelihood Estimation is really nothing more than efficient trial and error.
- It has three components:
  1. A function to be maximized, the log of likelihood for the equation
     - Why log instead of likelihood itself?
  2. An algorithm for generating efficient guesses of parameter values
  3. Starting values for the parameters.

Log of Likelihood for ARIMA Estimation

$$LL(\theta) = -\frac{T}{2} log(2\pi) - \frac{T}{2} log(\sigma^2) - \sum_{t=1}^{T} \frac{a_t^2}{2\sigma^2}$$

- This applies to any ARIMA(P,D,Q) model.
- When the likelihood is known, as here, the problem reduces to finding out how to estimate θ and a_t.

MA(1) illustration

- z_t = -$\theta_1 a_{t-1}$ + a_t + θ_0
- Drop θ_0 for simplicity
- Note inherent nonlinearity of -$\theta_1 a_{t-1}$
  - Both θ and a_{t-1} are unobserved quantities to be estimated

Step by step

- Presume for the moment that we somehow know θ
- How do we estimate a_t?
- *Except for the first case*; just solve one case at a time: z_t is given

Solve the MA(1) Equation for a_t

Just Algebra

1. $z_t = -\theta a_{t-1} + a_t$
2. $z_t + \theta a_{t-1} = + a_t$ (adding θa_{t-1} to both sides)
3. $a_t = z_t + \theta a_{t-1}$ (reversing)

- So, beginning at time zero, if we know a_0, we can solve for a_1, if we know a_1, we can solve for a_2, if we know a_2, we can solve for a_3, recursive all the way to a_T
- So assuming or computing a value for a_0 is the key to everything.

Conditional Maximum Likelihood

- $E(a_t)=0.0$;
  - Therefore initialize $a_0 = 0.0$
- Then maximize log-likelihood *conditional* on that starting value
- The assumption will be false, but its effect is *transient*
  - That is guaranteed by the stationarity condition

Unconditional Maximum Likelihood

- Backforecast a_0
- Because the backforecast is a product of known z and maximum likelihood estimates of θ and a, it will be optimum.
- Hence full ML is preferred to Conditional ML

Backforecasting for MA(1)

- Given: $z_t = -\theta_1 a_{t-1} + a_t$
- $z_t - a_t = -\theta_1 a_{t-1}$
- $(z_t - a_t)/-\theta_1 = a_{t-1}$
- Inititalize $a_{T+1} = 0$. Solve each previous a_{t-1} recursively, even a_0.

Why is this optimal?

- Backforecasting's assumption about unobservables is at a series' *end*,
- and the error in that assumption is transient, so
- the transient error will not affect the a_0 forecast at the series' origin.

Cyclical Data

- So far, difference equations have focused on trend and error processes in relation to recent values.
- Perhaps the data cycle across quarters or months.

One Approach: Linear Seasonal Terms

- Perhaps just add another autoregressive or moving-average term into the difference equation:
  - ARMA(1,4): $y_t = a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_4 \epsilon_{t-4}$
  - ARMA(4,1): $y_t = a_1 y_{t-1} + a_4 y_{t-4} + \epsilon_t + \beta_1 \epsilon_{t-1}$
- The rub: nonseasonal patterns will interact with the seasonal.
- This influences the empirical signature.

Multiplicative Seasonality

- The multiplicative model accounts for interaction among terms:
  - ARMA(1,1), MA seasonal: $(1 - a_1 L)y_t = (1 - \beta_1 L)(1 - \beta_4 L^4)\epsilon_t$
  - ARMA(1,1), AR seasonal: $(1 - a_1 L)(1 - a_4 L^4)y_t = (1 - \beta_1 L)\epsilon_t$
- MA seasonal, in functional terms:
  - $y_t = a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_4 \epsilon_{t-4} + \beta_1 \beta_4 \epsilon_{t-5}$
- Now our task will be spotting the empirical signature of such a series.

Notation

- The seasonal ARIMA model is now: ARIMA(p,d,q)(P,D,Q)$_s$
- When differencing series, subscripts refer to the seasonal period, and superscripts refer to number of differences.
- Using the above terms, our differencing notation is: $\Delta^d \Delta_s^D$

For Next Time

- Write down the research question for your term paper. What is the status of the project? Brand new? Data gathered? What?
- Complete questions #1 and #2.a-2.c from page 185 of *Political Analysis Using R*.
- **Reading:** *Time Series Analysis for the Social Sciences*, Chapter 2 (pp. 58-67) and Section 7.3 (pp. 187-205).