Gov 2002 - Section 8 - Time Series Analysis, With Applications to Dynamic Ideal Point Estimation

Connor Jerzak

Harvard University, Department of Government

November 2, 2016

Introduction

- **Setup**.
  - **"Nothing is so difficult as not deceiving oneself."** – Ludwig Wittgenstein, *Culture and Value*, p. 34.
  - This principle is especially true of time series analysis.
    1. Why? There are often more researcher degrees of freedom in time series analysis. This is because researchers not only can specify how the current value the covariates relate to the outcome, but they also model the relationship of the past covariates to current outcome.
    2. More researcher degrees of freedom means it is unacceptably publish "statistically significant" evidence consistent with any hypothesis.

Our goal

- Goals of time series analysis:
  1. Reduce omitted variable bias. The current value of X may depend on the past value, so we must condition on the past.
  2. Control rate of false positives. Without time series method, the rate of false positives is virtually 1 for non-stationary data (esp. data with a trend). We'll return to this.
  3. Get better predictions.
  4. Understand the temporal dynamics of a political phenomenon.

Time series notation

- Y_t is the value of Y at time t.
- y_t denotes a specific value of Y_t at t.
- y_{t-i} denotes the specific value of Y_t at the time point i time periods previous to t. Called the "i^{th} lag."
- $\Delta y_t = y_t - y_{t-1}$. Δ is the difference operator.
- $\Delta^2 y_t = \Delta(\Delta y_t)$. This is called the "second difference."
- Time series models take the following form:

$$Y_t = \beta_0 + f(Y_{t-1}, Y_{t-2}, ..., Y_{t-i}) + \epsilon_t, \tag{1}$$

where f is arbitrary and $\mathbb{E}[\epsilon_t | Y_{t-1}, Y_{i-2}, ..., Y_{i-p}] = 0$

Stationarity

- Most of the models in time series analysis assume stationarity.
- **Definition**.

$$F_X(x_{t_1}, ..., x_{t_k}) = F_X(x_{t_1+\tau}, ..., x_{t_k+\tau})$$

  where $F_X(\cdot)$ denotes the CDF. In other words, the parameters generating the data such not change over time. The mean/variance should be independent of t. The joint distribution of the future is the same as the joint distribution of the past.
- Why assume this?
  1. Spurious regression is highly likely if two or more series are non-stationary.
  2. Statistical results which hold for iid random variables (i.e. law of large numbers, etc.) also hold for stationary sequences.
  3. Stationary series have a finite variance; for nonstationary series, $\sigma^2 \to \infty$ as $t \to \infty$.

Enforcing stationarity

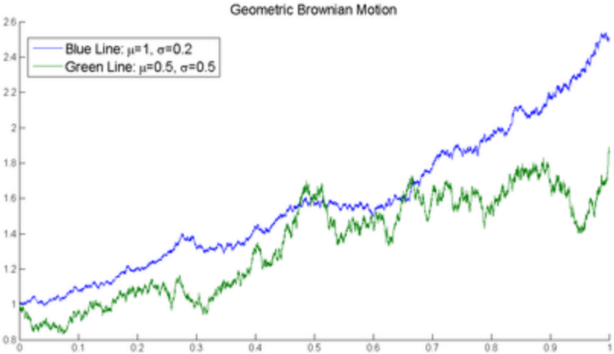
- Make data stationary by analyzing ΔY_t or $\Delta^2 Y_t$ instead of Y_t.
- Take out time or stationary trends by looking at residuals after regressing on the time or season variable.
- Dickey-Fuller Test. Tests for null hypothesis of non-stationary data. `adf.test` from `tseries` in `R`.
  1. Intuition: If a series non-stationary, then positive changes and negative changes will occur with probabilities that do not depend on the current level of the series; in a random walk, where you are now does not affect which way you will go next.

     $$y_t = \rho y_{t-1} + \epsilon_t.$$

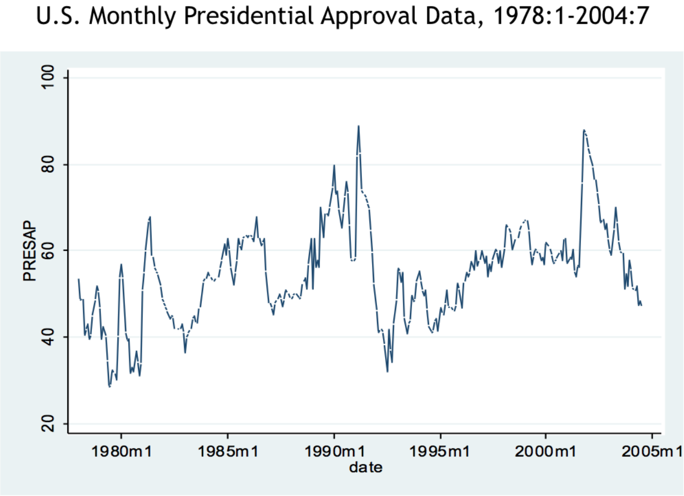
     The Dickey-Fuller tests whether $|\rho| > 1$. If $|\rho| > 1$, then stochastic process is on an explosive path. For further intuition, consider:

     $$\begin{aligned} Y_t &= \rho Y_{t-1} + \epsilon_t = \rho(\rho Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \ldots \to \infty \text{ if } \rho > 1 \end{aligned}$$

Non-stationary data



Stationary data



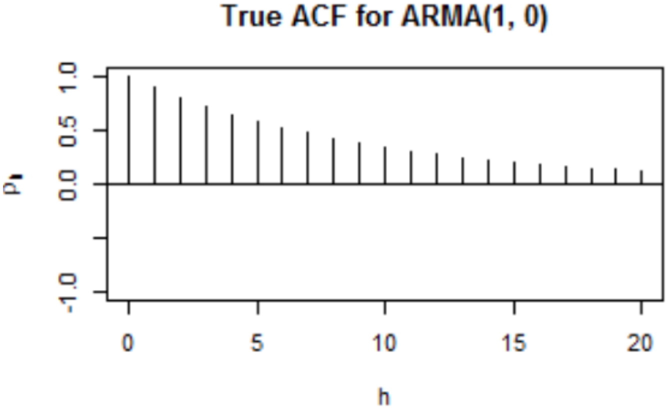
ARMA model

- Main building block of time series analysis: combines the autoregressive (AR) and moving average (MA) framework.
  1. AR(p) models assume that the value of X_t depends on $X_{t-1}, ..., X_{t-p}$.

$$X_t = \beta_0 + \sum_{i=1}^{p} \beta_i X_{t-i} + \epsilon_t, \quad \epsilon_t \sim D(0, \sigma_t^2), \beta_i < 1.$$

  Example of AR(1) process: changes in government budgets. The budget might increase in year t due to an increase in a high frequency trading tax, and then might increase again (by a smaller amount) in the next year. The budget might increase again because more people every year enter the financial markets, but it might increase less than before since people adapt to the tax and trade less.

Visualizing ARMA models



ARMA model

2. MA(q) models assume that the value of X_t depends on the value of past shocks and starting value. (Example: hurricane and changes in oil price; lemonade example).

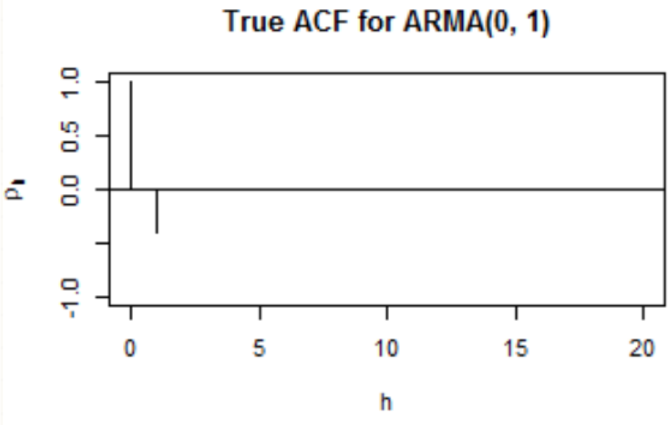
$$X_t = \mu + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i} + \epsilon_t,$$

where ϵ_{t-i} is white noise. Note that AR(p) can be written as an MA(∞):

$$y_t = \phi_1 y_{t-1} + \epsilon_t = \phi_1(\phi_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2^2 \epsilon_{t-2} + \ldots$$

Fitting the MA estimates is more complicated than with autoregressive models (AR models) because the lagged error terms are not observable. Iterative non-linear fitting procedures need to be used in place of linear least squares.

Visualizing ARMA models



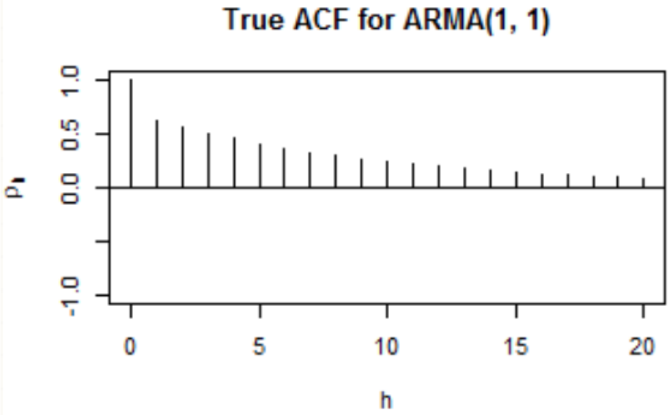
ARMA model

An ARMA(p, q) model combines features from an AR(p) and an MA(q):

$$X_t = \left\{ \sum_{i=1}^{p} \beta_i X_{t-i} \right\} + \left\{ \sum_{i=1}^{q} \gamma_i \epsilon_{t-i} \right\} + \sum_{i=1}^{d} \eta_i d_{t-i} + \epsilon_t,$$

where d_t is an exogenous input.

Visualizing ARMA models



Code

```
#test for stationarity
adf.test(my_data$approve,
         alternative = "stationary",
         k = 1)

#perform arima model
approval_vec_diff <- diff( my_data$approve )
my_arima <- arima(approval_vec_diff,
                  order = c(1, 0, 1))
my_arima
```

Extensions of the ARMA model

- **When to use ARMA?** Use an ARMA model if a process is characterized by unobserved shocks, as well as the past level of the process.
- **GARCH Extension.** If a ARMA model is assumed for the error variance, the model is called a generalized autoregressive conditional heteroscedasticity (GARCH) model. This generalization allows the volatility of a process to vary.

Dynamic Ideal Point Estimation

- To what extent does the preferences of political actors change over time? Martin & Quinn attempt to answer this question in the context of the Supreme Court in Martin & Quinn (2002).
- Significance.
  - Get better predictions of decisions (although not huge, 63% $\rightarrow$ 74% $\rightarrow$ 76%).
  - Weaken assumptions about invariant preferences.
  - Study the influence of shocks on political preferences.

Random utility model

Let $u^{(a)}_{t,k,j}$ be the utility to justice j of voting to affirm case k in term t, and $u^{(a)}_{t,k,j}$ be the utility from voting to reverse the case.

$$\begin{aligned}
z_{t,k,j} &= u^{(r)}_{t,k,j} - u^{(a)}_{t,k,j};\\
&= \left[-(\theta_{t,j} - x^{(r)}_k)^2 + \xi^{(r)}_{t,k,j} \right] - \left[-(\theta_{t,j} - x^{(a)}_k)^2 + \xi^{(a)}_{t,k,j} \right];\\
&= \left[x^{(a)}_k x^{a}_k - x^{(r)}_k x^{(r)}_k \right] + 2[x^{(r)}_k - x^{(a)}_k] + [\xi^{(r)}_{t,k,j} - \xi^{(a)}_{t,k,j}];\\
&= \alpha_k + \beta_k \theta_{t,j} + \epsilon_{t,k,j}.
\end{aligned}$$

Random utility model, continued

$$v_{t,k,j} = \begin{cases} 1 & \text{if} \quad z_{t,k,j} > 0 \\ 0 & \text{if} \quad z_{t,k,j} \leq 0. \end{cases}$$

Assumption: Justice j will vote to reverse case k if $z_{t,k,j} > 0$.

- $v_{t,k,j}$ is 1 if the justice decides to reverse the case.
- $\theta_{t,j}$ is justice j's ideal point in a unidimensional space for term t.
- $\epsilon_{t,k,j} \sim N(0,1)$.
- $x_k^{(r)}$ is the location of a policy in the uni-dimensional space under a reversal; $x_k^{(a)}$ is the location under an affirmative vote.

The Bayesian Model

- The target posterior (MCMC used for estimation):

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta} | \boldsymbol{V}) \propto f(\boldsymbol{V} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$$

- The (Bernoulli) likelihood. This represents the density of observing the vote choice, given the parameters:

$$f(\boldsymbol{V} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \propto \prod_{t=1}^{T} \prod_{k \in K_t} \prod_{j \in J_k} \left[\Phi(\alpha_k + \beta_k \theta_{t,j}) \right]^{v_{t,k,j}} \cdot \left[1 - \Phi(\alpha_k + \beta_k \theta_{t,j}) \right]^{1 - v_{t,k,j}}$$

The Bayesian Model

- The prior for $\alpha_k, beta_k$:

$$\begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} \sim N_2(\mathbf{b}_0, \ \ \mathbf{B}_0), \ \ k \in \{1, ..., K\}$$

- Random walk prior on $\theta_{t,j}$:

$$\theta_{t,j} \sim N(\theta_{t-1,j} \Delta_{\theta_{t,j}}),$$

for $t = \underline{T}_j$ (justice j's first term) to $t = \bar{T}_j$ (j's last term). In addition:

$$\theta_{0,j} \sim N(m_{0,j}, \ \ C_{0,j}).$$

Visualizing the Ideal Point Results

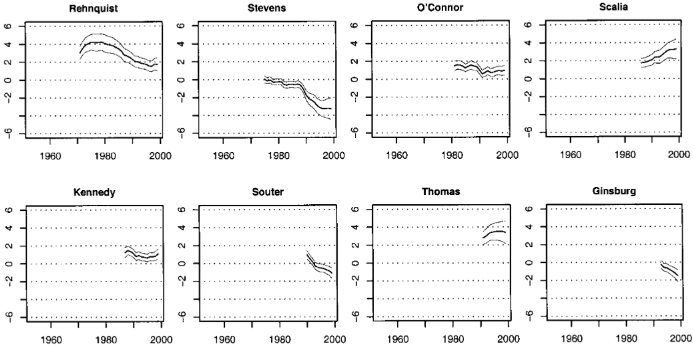


Fig. 1 Posterior density summary of the ideal points of selected justices for the terms in which they served for the dynamic ideal point model.

The upshot

- Time series analysis is a challenge. Many researcher degrees of freedom. "True" model could be arbitrarily complex.
- Be sure to make your data stationary before proceeding with regression analysis.
- Temporal dynamics can be substantively interesting - think carefully about how to model them!

Acknowledgements

- These slides build from material presented by Pearl, Rubin, Jordan, Jackman, and others.