PLSC 504 – Fall 2024
Time Series Analysis:
An Introduction

October 2, 2024

Primitives & Terminology

Longitudinal Data = Data reflecting temporal variation.

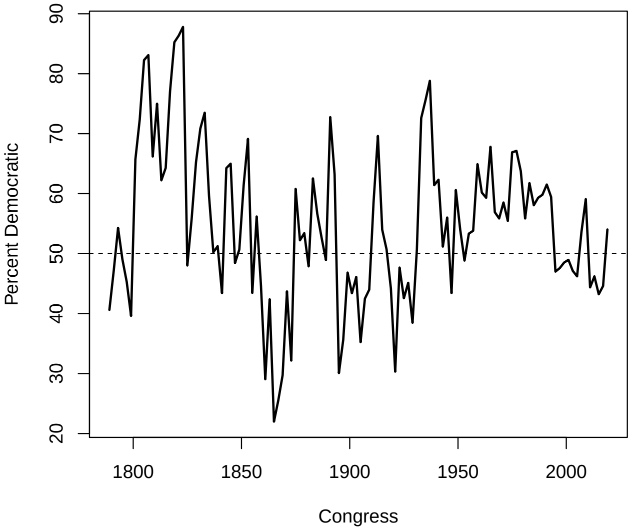
Typically:

- $i \in \{1, 2, ...N\}$ indexes cross-sectional units
- $t \in \{1, 2, ...T\}$ indexes temporal observations

Terminology:

- Cross-Sectional: N is large, $T = 1$
- Time Series: T is large, $N = 1$
- Panel: N is large, T is small(ish)
- TSCS: $N \approx T$, or $N < T$ (but $N > 1$)

Democratic House Membership, 1789-2019



Time Series Flavors

- **Univariate**: Properties of series, ARIMA, etc.
- **Bivariate**: Pairs of series / regression
  - "Distributed Lag" models
  - Structural breaks / intervention analysis
  - Cointegration / Error Correction
  - Others...
- **Multivariate**: Systems of equations
  - "Seemingly unrelated" regressions
  - Vector Autoregressions ("VARs") and their variants
- **Higher-moment properties**: ARCH, GARCH, etc.

Quick Review of Error Correlation

OLS requires:

$$E(u_i, u_j) = 0 \;\forall\; i \neq j$$

or...

$$E(\mathbf{uu}') = \sigma^2\mathbf{I}$$

$$\begin{aligned} \text{Correlated residuals} \quad & \rightarrow \underline{\text{no}} \text{ bias in } \hat{\boldsymbol{\beta}}\text{, but} \\ & \rightarrow \text{biased, inconsistent estimates of } \widehat{\mathbf{V}(\hat{\boldsymbol{\beta}})} \end{aligned}$$

Time Series: Autocorrelation

Example:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + u_t \\ u_t &= \rho u_{t-1} + e_t \end{aligned}$$

with $e_t \sim i.i.d.\ N(0, \sigma_u^2)$ and $\rho \in [-1, 1]$ (typically).

→ "First-order autoregressive" ("AR(1)") errors.

Serially Correlated Errors and OLS

Detection

- Plot of residuals vs. lagged residuals
- Runs test (Geary test)
- Durbin-Watson *d*
  - Calculated as:

$$d = \frac{\sum_{t=2}^{N}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{N}\hat{u}_t^2}$$

  - Non-standard distribution
  - Only detects first-order autocorrelation

Serially Correlated Errors and OLS

What to do about it?

- GLS, incorporating ρ / $\hat{\rho}$ into the equation
- First-difference equations (regressing changes of Y on changes of $\mathbf{X}$)
- Cochrane-Orcutt / Prais-Winsten:
  1. Estimate the basic equation via OLS, and obtain residuals
  2. Use the residuals to consistently estimate $\hat{\rho}$ (i.e. the empirical correlation between u_t and u_{t-1})
  3. Use this estimate of $\hat{\rho}$ to estimate the difference equation:

$$(Y_t - \rho Y_{t-1}) = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$$

  4. Save the residuals, and use them to estimate $\hat{\rho}$ again
  5. Repeat this process until successive estimates of $\hat{\rho}$ differ by a very small amount

Time Series, In General

Consider a random series **Y**:

$$\mathbf{Y} = \{Y_0, Y_1, ...Y_T\}$$

with moments:

$$\begin{aligned}
\mu_t &= E(Y_t) \\
\sigma_t^2 \equiv \text{Var}(Y_t) &= E[(Y_t - \mu_t)^2] \\
\gamma_{t,t-s} \equiv \text{Cov}(Y_t, Y_{t-s}) &= E[(Y_t - \mu_t)(Y_{t-s} - \mu_{t-s})]
\end{aligned}$$

Two typical assumptions: stationarity and ergodicity.

Time Series: Stationarity

Stationarity: A constant d.g.p. over time.[1]

Mean stationarity:

$$E(Y_t) = \mu \; \forall \; t$$

Variance stationarity:

$$Var(Y_t) = E[(Y_t - \mu)^2] \equiv \sigma_Y^2 \; \forall \; t$$

Covariance stationarity:

$$Cov(Y_t, Y_{t-s}) = E[(Y_t - \mu)(Y_{t-s} - \mu)] = \gamma_s \; \forall \; s$$

---

[1] A stricter form of stationarity requires that the joint probability distribution (in other words, *all* the moments) of series of observations $\{Y_1, Y_2, ... Y_t\}$ is the same as that for $\{Y_{1+s}, Y_{2+s}, ... Y_{t+s}\}$ for all t and s.

Time Series: Ergodicity

Ergodicity: Asymptotic (in T) independence.

Means that:

$$Cov(Y_t, Y_{t-s}) = \gamma_s \to 0 \;\; as \;\; s \to \infty$$

Stationary ergodic processes:

- are measure-preserving,
- can have their statistical properties inferred/estimated from a sufficiently long observation of the process.

Autocovariance and Autocorrelation

For a stationary ergodic series:

$$\begin{aligned}
\hat{\mu} = \bar{Y} &= T^{-1} \sum_{t=1}^{T} Y_t \\
\hat{\sigma}^2 &= T^{-1} \sum_{t=1}^{T} (Y_t - \bar{Y})^2 \\
\hat{\gamma}_s &= T^{-1} \sum_{t=s+1}^{T} (Y_t - \bar{Y})(Y_{t-s} - \bar{Y}), \ s = 1, 2, 3, ...
\end{aligned}$$

Autocovariance → Autocorrelation:

$$\hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\sigma}^2}, \ s = 0, \pm 1, \pm 2, ...$$

The "ARIMA" Approach

"ARIMA" = Autoregressive Integrated Moving Average...

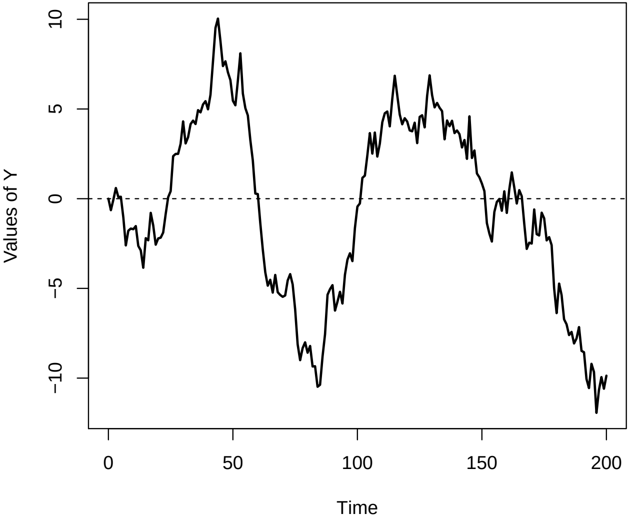
A (first-order) integrated series ("random walk") is:

$$Y_t = Y_{t-1} + u_t,\ u_t \sim i.i.d.(0, \sigma_u^2)$$

...a/k/a a "random walk":

$$\begin{aligned} Y_t &= Y_{t-2} + u_{t-1} + u_t \\ &= Y_{t-3} + u_{t-2} + u_{t-1} + u_t \\ &= \sum_{t=0}^{T} u_t \end{aligned}$$

I(1) series with $u_t \sim N(0, 1)$, $T = 200$



I(1) Series Properties

I(1) series are not stationary.

Variance:

$$\mathrm{Var}(Y_t) \equiv E(Y_t)^2 \quad = \quad t\sigma^2$$

Autocovariance:

$$\mathrm{Cov}(Y_t, Y_{t-s}) = |t-s|\sigma^2.$$

Both depend on t...

I(1) Series: Differencing

For an I(1) series:

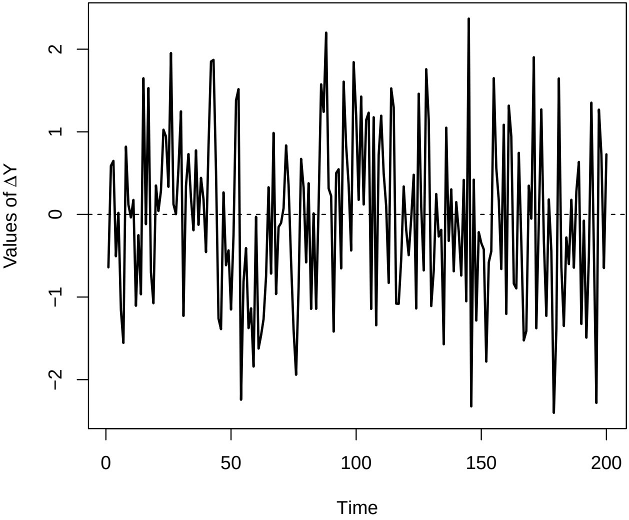
$$Y_t - Y_{t-1} = u_t$$

which we often write in terms of the difference operator Δ (or sometimes ∇):

$$\Delta Y_t = u_t$$

The differenced series is just the (stationary, ergoditic) white-noise process u_t.

A Differenced I(1) Series



Higher-order [I(d)] Integrated Series

An I(d) series is one which, differenced d times, becomes a stationary series. E.g., the I(2) series:

$$Y_t = u_t + 2u_{t-1} + 3u_{t-2} + ...$$

Differencing:

$$\begin{aligned} \Delta Y_t &= [u_t + 2u_{t-1} + 3u_{t-2} + ...] - [u_{t-1} + 2u_{t-2} + 3u_{t-3} + ...] \\ &= \sum_{j=0}^{T} u_{t-j} \end{aligned}$$

Differencing again:

$$\begin{aligned} \Delta^2 Y_t \equiv (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) &= [u_t + u_{t-1} + ...] - [u_{t-1} + u_{t-2} + ...] \\ &= u_t \end{aligned}$$

which is stationary / ergoditic.

AR Processes

An AR(1) process:

$$Y_t = \phi Y_{t-1} + u_t$$

with $u_t \sim i.i.d.(0, \sigma_u^2)$.

Equivalently:

$$Y_t - \phi Y_{t-1} = u_t$$

or:

$$\begin{aligned} Y_t &= \phi[\phi Y_{t-2} + u_{t-1}] + u_t \\ &= \phi^2 Y_{t-2} + \phi u_{t-1} + u_t \\ &= \phi^2[\phi Y_{t-3} + u_{t-2}] + \phi u_{t-1} + u_t \\ &= \phi^3 Y_{t-3} + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \\ &= \ldots \\ &= \sum_{j=0}^{T-1} \phi^j u_{t-j} + \phi^T Y_0 \end{aligned}$$

AR(1) Series: Properties

Mean:

$$\begin{aligned} E(Y_t) &= E\left(\sum_{j=0}^{T-1} \phi^j u_{t-j}\right) + E(\phi^T Y_{t-T}) \\ &= \phi^T Y_0 \end{aligned}$$

That means that:

- For $|\phi| > 1$, $|E(Y_t)|$ is increasing in t
- For $|\phi| = 1$, $E(Y_t) = Y_0$
- For $|\phi| < 1$:
  - the importance of Y_0 decreases over time
  - asymptotically, $E(Y_t) = 0$

AR(1) and Stationarity

For an AR(1) series with $|\phi| < 1$:

$$Y_t = \sum_{j=0}^{\infty} \phi^j u_{t-j}$$

Means that:

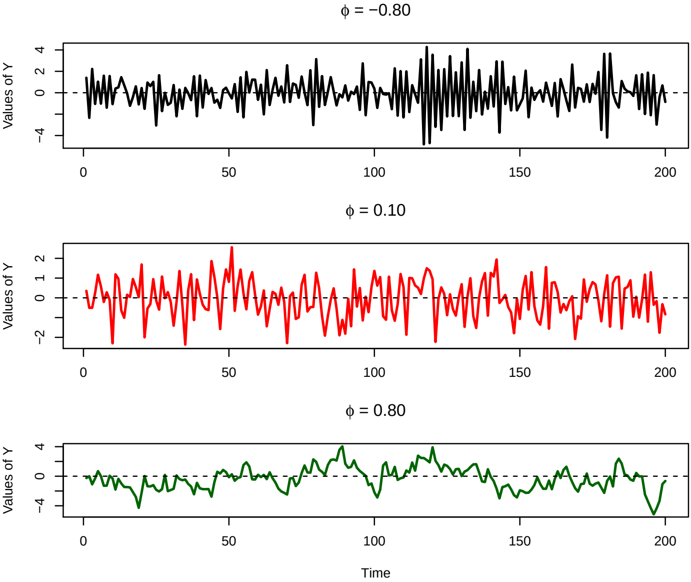
- the effects of a shock *persist* over time; however,
- the *size* of those effects *decay* over time
- In the limit, the effect of any one shock very much later is zero

Similarly:

$$\begin{aligned} \text{Var}(Y_t) \equiv E(Y_t^2) &= E(\sum_{j=0}^{\infty} \phi^j u_{t-j})^2 \\ &= \sum_{j=0}^{\infty} \phi^{2j} E(u_{t-j}^2) \\ &= \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} \\ &= \frac{\sigma^2}{1 - \phi^2} \end{aligned}$$

which is stationary iff $|\phi| < 1$.

AR(1) Series, Illustrated



Notation Digression: The Lag Operator

"Lag operator" L:

$$\begin{aligned} LY_t &= Y_{t-1} \\ L^2 Y_t &= Y_{t-2} \\ L^s Y_t &= Y_{t-s} \\ L^0 Y_t &= Y_t \end{aligned}$$

Means we can write:

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} \\ &= (1 - L)Y_t \end{aligned}$$

So an AR(1) equation is:

$$(1 - \phi L)Y_t = u_t$$

Higher-Order AR(p) Series

An AR(p) series:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + u_t$$

Equivalently:

$$\begin{aligned} Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - ... - \phi_p Y_{t-p} &= u_t \\ (1 - \phi L - \phi_2 L^2 - ... - \phi_p L^p) Y_t &= u_t \end{aligned}$$

Moving Average (MA) Series

MA(1) series:

$$Y_t = u_t + \theta u_{t-1}, \; t = 1, 2, ... T$$

with $u_t \sim \text{i.i.d.}(0, \sigma_u^2)$.

Mean:

$$\begin{aligned} \mu &= E(Y_t) \\ &= E(u_t) + \theta E(u_{t-1}) \\ &= 0 \end{aligned}$$

Variance:

$$\begin{aligned} \sigma_Y^2 &= E[(u_t + \theta u_{t-1})(u_t + \theta u_{t-1})] \\ &= E(u_t^2) + \theta^2 E(u_{t-1}^2) + 2\theta E(u_t u_{t-1}) \\ &= (1 + \theta^2)\sigma_u^2 \end{aligned}$$

MA(1), continued

Autocovariance at one lag:

$$\begin{aligned}
\gamma_1 &= E[(u_t + \theta u_{t-1})(u_{t-1} + \theta u_{t-2})] \\
&= E(u_t u_{t-1}) + \theta E(u_{t-1}^2) + \theta E(u_t u_{t-2}) + \theta^2 E(u_{t-1} u_{t-2}) \\
&= \theta E(u_{t-1}^2) \\
&= \theta \sigma_u^2
\end{aligned}$$

Autocovariance at $s > 1$ lags:

$$\begin{aligned}
\gamma_s &= E[(u_t + \theta u_{t-s})(u_{t-1} + \theta u_{t-(s+1)})] \\
&= E(u_t u_{t-s}) + \theta E(u_{t-s}^2) + \theta E(u_t u_{t-(s+1)}) + \theta^2 E(u_{t-s} u_{t-(s+1)}) \\
&= 0
\end{aligned}$$

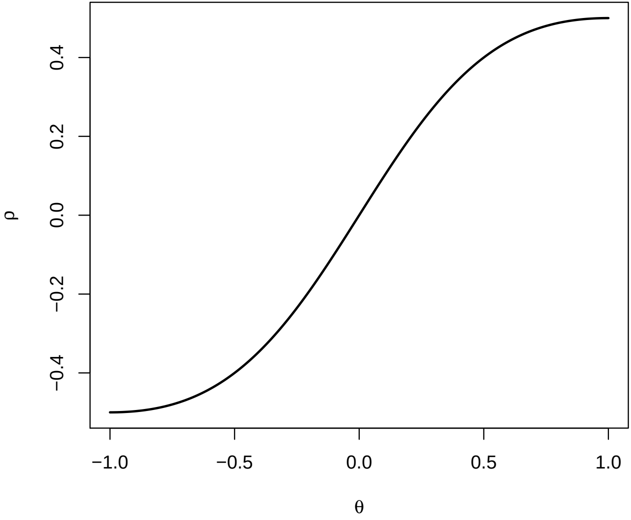
MA(1), continued

This means that:

- the means, variances and autocovariances are all independent of t,
- the ACF is zero for all lags greater than one,
- the ACF at one lag is dependent on the degree of dependence in the moving average process. In particular, since $\rho = \frac{\gamma}{\sigma^2}$:

$$\begin{aligned} \rho &= \frac{\theta \sigma_u^2}{(1+\theta^2)\sigma_u^2} \\ &= \frac{\theta}{1+\theta^2} \end{aligned}$$

Relationship between θ and ρ for an MA(1) series

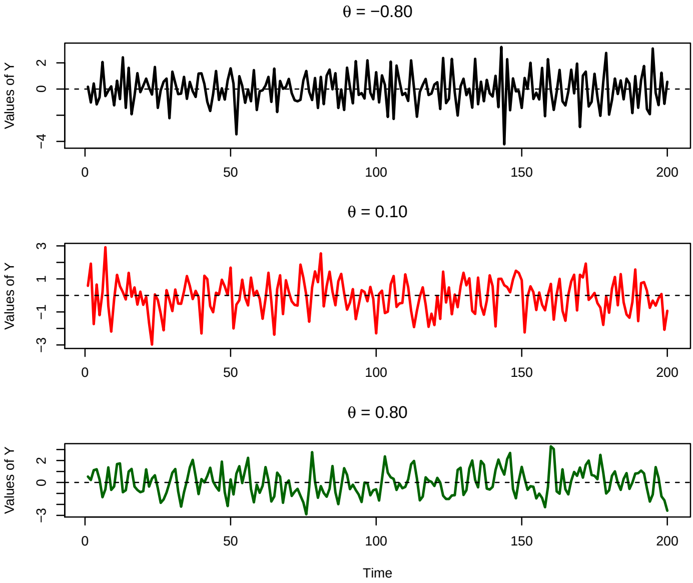


MA(1) Series Properties

For an MA(1) series:

- $\theta = 0 \leftrightarrow \rho_1 = 0$, a "white noise" process (no temporal dependence in Y_t).
- For $\theta > 0$:
  - $\rho_1 > 0$
  - successive values of Y_t will be *positively* related, and
  - the series will be "smoother" than a white noise series.
- For $\theta < 0$:
  - $\rho_1 < 0$
  - successive values of Y_t will be *negatively* correlated, and
  - the series will be less "smooth" than a white noise sequence.

MA(1) Series Illustrated



MA = AR...

Rewriting our MA(1) series:

$$\begin{aligned}
Y_t &= u_t + \theta u_{t-1} \\
&= u_t + \theta(Y_{t-1} - \theta u_{t-2}) \\
&= u_t + \theta Y_{t-1} - \theta^2 u_{t-2} \\
&= u_t + \theta Y_{t-1} - \theta^2(Y_{t-2} - \theta u_{t-3}) \\
&= u_t + \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 u_{t-3} \\
&= \ldots \\
&= u_t + \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 Y_{t-3} - \ldots - (-\theta)^{(T-1)} Y_{t-(T-1)} - (-\theta)^T u_0
\end{aligned}$$

As $T \rightarrow \infty$:

$$Y_t = -\sum_{j=1}^{\infty} (-\theta)^j Y_{t-j} + u_t$$

The Wold Decomposition Theorem

Any weakly stationary, purely nondeterministic stochastic process can be written as a linear combination of a sequence of uncorrelated random variables.

that is, as:

$$Y_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$$

with $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

MA(q) Models

MA(q) model:

$$Y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \theta_q u_{t-q}$$

MA(q) models are:

- always stationary,
- invertable if the AR(p) form of the model satisfies the stationarity conditions above.

ARMA / ARIMA Models

A general ARMA(p, q) model is:

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + u_t + \theta_1 u_{t-1} + ... + \theta_q u_{t-q}$$

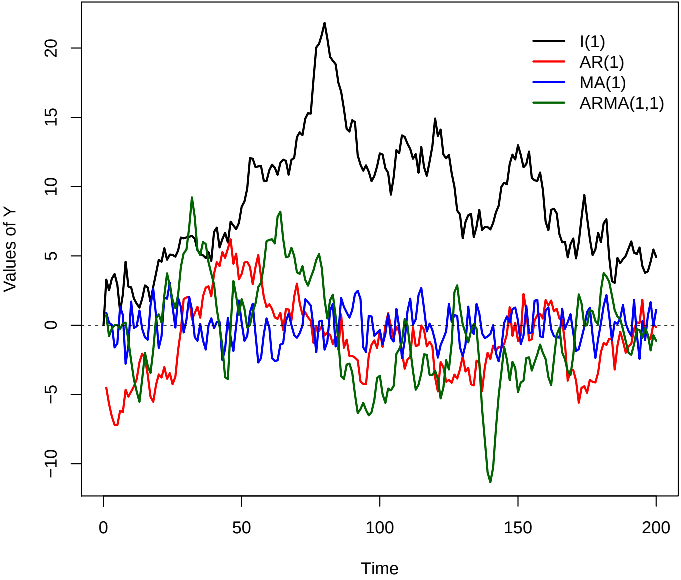
Equivalently:

$$(1 - \phi L - \phi_2 L^2 - ... - \phi_p L^p) Y_t = (1 + \theta L + \theta_2 L^2 + ... + \theta_q L^q) u_t$$

If the model also requires dth-order differencing in order to achieve stationarity, we call it an ARIMA(p, d, q) model; e.g., the ARIMA($p, 1, q$) model:

$$\Delta Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + u_t + \theta_1 u_{t-1} + ... + \theta_q u_{t-q}$$

ARIMA Series: A Comparison ($\phi = \theta = 0.9$)



Series Characteristics: ACFs and PACFs

Autocorrelation function (ACF):

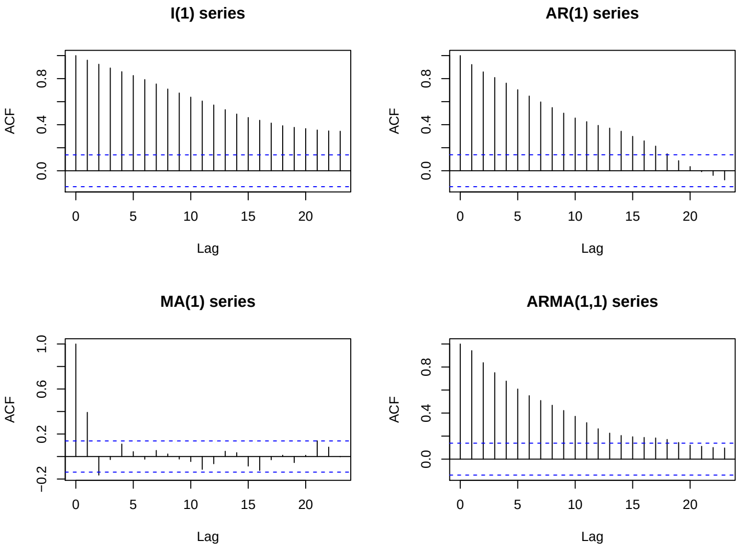
$$\hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\sigma}^2},\ s = 0, \pm 1, \pm 2, ...$$

The partial autocorrelation function (PACF) is the correlation between Y_t and Y_{t-s} after controlling for ("partialling out") the the common linear effects of the intermediate lags.[2]

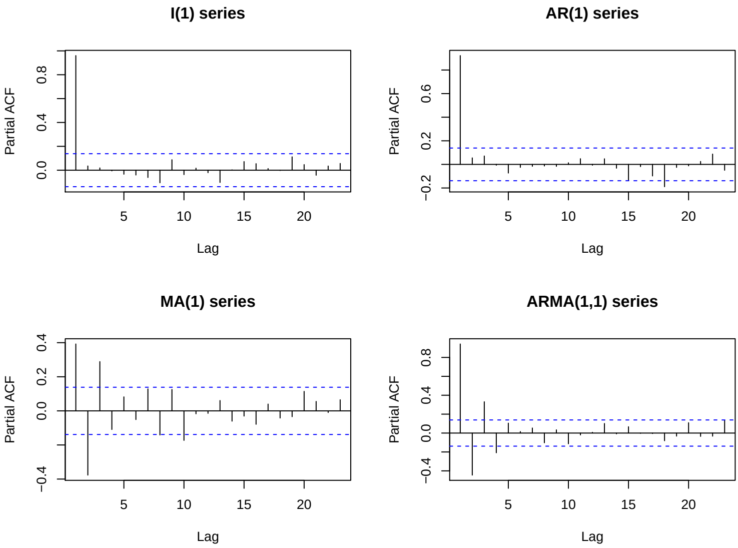
---

[2] The PACF is estimated from a solution to the Yule-Walker system of equations. A good mathematical treatment of this is in Box et al. 1994, pp. 64-69

ARIMA Series: ACFs



ARIMA Series: PACFs



ARIMA Modeling: Intuition

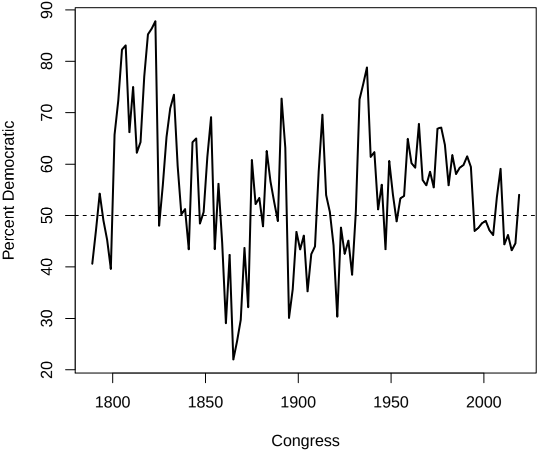
Identification $\rightarrow$ *Estimation* $\rightarrow$ *Diagnostics*

In general:

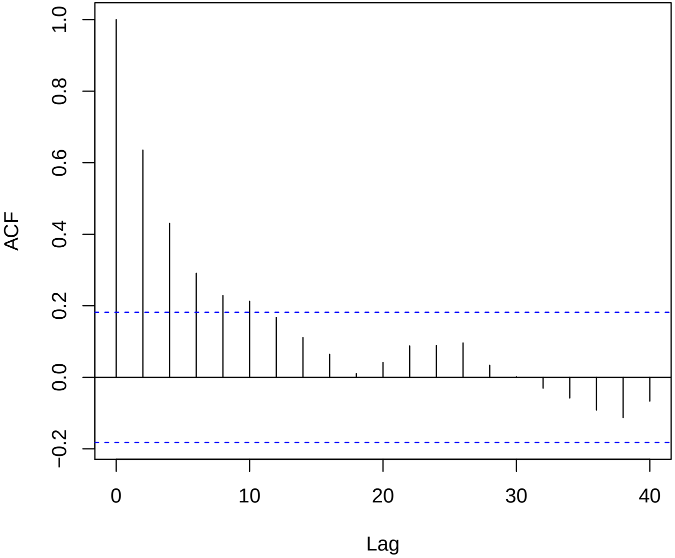
1. Determine of the series is stationary. If so, proceed to (3).
2. If not, difference the series until it is.
3. Examine the ACFs and PACFs of the stationary series, in order to determine the data generating process (AR, MA, or some combination thereof).
4. Fit a model – starting simple – using ARIMA/MLE.
5. Examine and test the residuals to determine if they are white noise. If they are, proceed to (7).
6. If they are not, go back to (3) and try again.
7. Proceed with inference, hypothesis testing, forecasting, and the like.

Example: Democratic House Membership

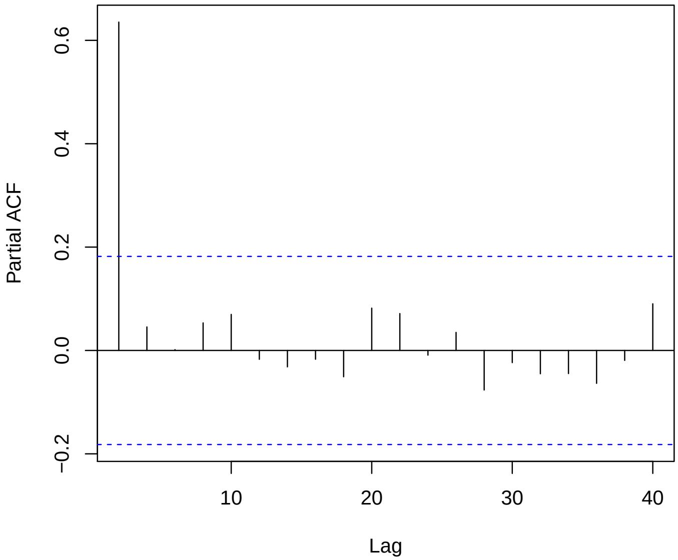
$Y =$ percentage of U.S. House members identifying with the Democratic party, by Congress, 1789-2019.



Democratic House Membership: ACF



Democratic House Membership: PACF



ARIMA Model Fitting

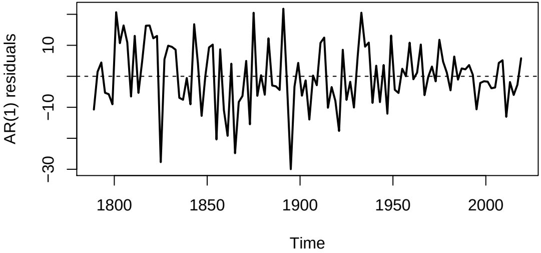
```
> DH.AR1 <- arima(DH.TS,order=c(1,0,0),method="ML") # AR(1)
> summary(DH.AR1)

Coefficients:
        ar1  intercept
      0.636      54.53
s.e.  0.071       2.53

sigma^2 estimated as 101:  log likelihood = -432.6,  aic = 871.2

Training set error measures:
                 ME  RMSE   MAE    MPE  MAPE   MASE     ACF1
Training set 0.1038 10.06 7.951 -3.858 16.03 0.9476 -0.03006


> DH.ARMA11 <- arima(DH.TS,order=c(1,0,1),method="ML") # ARMA(1,1)
> summary(DH.ARMA11)

Coefficients:
        ar1     ma1  intercept
      0.679  -0.072     54.495
s.e.  0.107   0.148      2.651

sigma^2 estimated as 101:  log likelihood = -432.5,  aic = 873

Training set error measures:
                 ME  RMSE   MAE    MPE  MAPE   MASE      ACF1
Training set 0.1123 10.04 7.915 -3.824 15.94 0.9434 -0.001164
```

ARIMA Model Selection

```
> # Model selection via LR test:
>
> lrtest(DH.AR1,DH.ARMA11)

Likelihood ratio test

Model 1: arima(x = DH.TS, order = c(1, 0, 0), method = "ML")
Model 2: arima(x = DH.TS, order = c(1, 0, 1), method = "ML")
  #Df LogLik Df Chisq Pr(>Chisq)
1   3   -433
2   4   -432  1  0.24       0.62
```

```
> # Automated version:
>
> DH.robot <- auto.arima(DH.TS)
> summary(DH.robot)
Series: DH.TS
ARIMA(1,0,0) with non-zero mean

Coefficients:
        ar1   mean
      0.636  54.53
s.e.  0.071   2.53

sigma^2 estimated as 103:  log likelihood=-432.6
AIC=871.2   AICc=871.4   BIC=879.5

Training set error measures:
                 ME  RMSE   MAE    MPE  MAPE   MASE     ACF1
Training set 0.1037 10.06 7.951 -3.858 16.03 0.1452 -0.03005
```

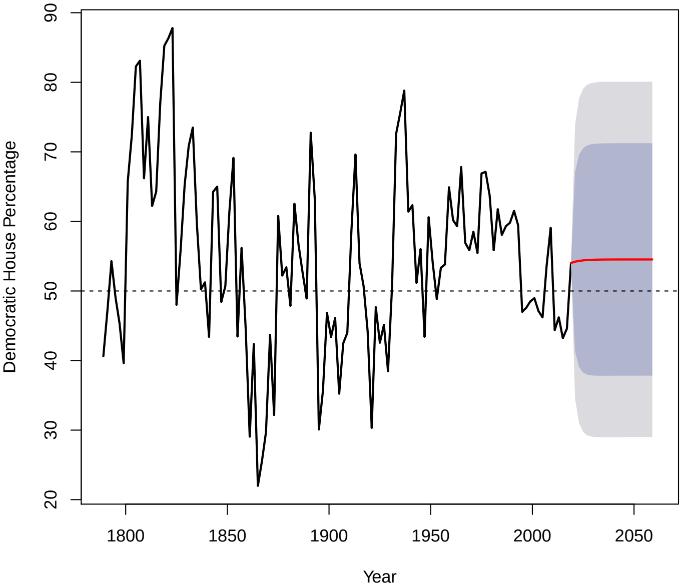
ARIMA: Residuals Analysis



Box-Pierce and Ljung-Box tests:

```
> Box.test(DH.AR1$residuals)

 Box-Pierce test

data:  DH.AR1$residuals
X-squared = 0.1, df = 1, p-value = 0.7

> Box.test(DH.AR1$residuals,type="Ljung")

 Box-Ljung test

data:  DH.AR1$residuals
X-squared = 0.11, df = 1, p-value = 0.7
```

ARIMA: Forecasting (40-Year Window)



Unit Roots, (Co)Integration, and Error Correction

Unit Roots...

A basic I(1) series is:

$$Y_t = Y_{t-1} + u_t$$

Special case of AR(1):

$$Y_t = \rho Y_{t-1} + u_t$$

where $\rho = 1$.

Recall:

- $|\rho| < 1.0 \leftrightarrow$ stationary series
- $|\rho| = 1.0 \leftrightarrow$ non-stationary series

Other Non-Stationary Series

Trending series:

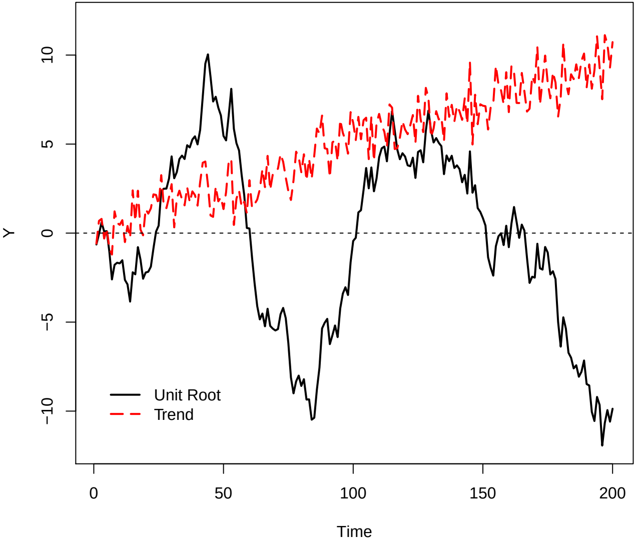
$$Y_t = \beta t + u_t$$

Means:

$$\begin{aligned}
\Delta Y_t \equiv Y_t - Y_{t-1} &= \beta t + u_t - Y_{t-1} \\
&= \beta(t) + u_t - \beta(t-1) - u_{t-1} \\
&= \beta(t - t + 1) + u_t - u_{t-1} \\
&= \beta + \Delta u_t
\end{aligned}$$

which is also stationary.

Unit Root vs. Trend



Unit Root? Trend?

Key: Examine

$$Y_t = \rho Y_{t-1} + \beta t + u_t$$

and test for $H_0 : \hat{\beta} = 0$:

- Cannot reject $\hat{\beta} = 0 \leftrightarrow$ “random walk” without a trend;
- Reject $\hat{\beta} = 0 \leftrightarrow$ series has a deterministic trend.

ρ, Generally

- $|\rho| > 1$
  - Series is nonstationary / *explosive*
  - Past shocks have a greater impact than current ones
  - Uncommon
- $|\rho| < 1$
  - *Stationary* series
  - Effects of shocks die out exponentially according to ρ
  - Is mean-reverting
- $|\rho| = 1$
  - Nonstationary series
  - Shocks persist at full force
  - Not mean-reverting; variance increases with t

Dickey-Fuller...

Basic idea:

$$Y_t = \rho Y_{t-1} + u_t$$

and test for $\hat{\rho} = 1$.

Issues:

- $\hat{\rho} = \frac{\sum Y_t Y_{t-1}}{\sum Y_{t-1}^2}$ is consistent for ρ, but
- the distribution of $\hat{\rho}$ is not t...
  - $T(\hat{\rho} - 1) \rightarrow \frac{\int_0^1 W(r) dW(r)}{\int_0^1 W(r)^2 dr}$, so
  - $t \equiv \frac{\hat{\rho}}{s.e.(\hat{\rho})} = \frac{\frac{1}{2}[W(1)^2 - 1]}{[\int_0^1 W(r)^2 dr]^{\frac{1}{2}}}$

This is:

- Right-skewed (so t-statistics will tend to be large and negative),
- → tend to *overreject* the null hypothesis if we use the standard t-distribution.

The Dickey-Fuller Test

1. Estimate $\hat{\rho}$,
2. Test $H_0 : \hat{\rho} = 1$ using a non-standard set of critical values,
3. Ensure that $\hat{u}$ are white noise.

Equivalent:

$$\begin{aligned} \Delta Y_t &= (\rho - 1) Y_{t-1} + u_t \\ &= \delta Y_{t-1} + u_t \end{aligned}$$

and test for $\hat{\delta} = 0$.

Dickey-Fuller Variants

“Drift”:

$$Y_t = \alpha + \rho Y_{t-1} + u_t$$

which is also:

$$Y_t = Y_0 + \alpha t + \sum_{t=1}^{T} u_t$$

This requires testing both $\hat{\rho} = 1$ and $\hat{\alpha} = 0$.

Dickey-Fuller Variants

"Trend":

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + u_t$$

- α is now a "constant".
- Requires a (slightly) different set of critical values,
- Also: F-tests on the joint nulls: $\rho = 1$ and $\beta = 0$

In general, it's worse to *omit* a drift / trend from a model when the data generating process has one, than it is to *include* one where the DGP is driftless / trendless.

More Variants

Suppose we have:

$$\Delta Y_t = \sum_{i=1}^{p} d_i \Delta Y_{t-i} + u_t$$

This yields:

$$Y_t = Y_{t-1} + \sum_{i=1}^{p} d_i \Delta Y_{t-i} + u_t$$

If this is true, a standard D.F. test:

$$Y_t = \hat{\rho} Y_{t-1} + u_t$$

will have AR(1) errors / residuals...

“Augmented” D-F Tests

Estimate:

$$\Delta Y_t = \rho Y_{t-1} + \sum_{i=1}^{p} d_i \Delta Y_{t-i} + u_t$$

or:

$$\Delta Y_t = \alpha + \beta t + \rho Y_{t-1} + \sum_{i=1}^{p} d_i \Delta Y_{t-i} + u_t$$

Choosing p via AIC/BIC, etc. until we are certain that the residuals $\hat{u}$ are uncorrelated / white noise.

Phillips-Perron Test

For:

$$\Delta Y_t = \alpha + \rho Y_{t-1} + u_t$$

The distribution of $\hat{\rho}$ depends on $\frac{\sigma^2}{\sigma_e^2}$, where

- $\sigma^2 = \mathrm{Var}(u)$ and
- $\sigma_e^2 = \lim_{T\to\infty} T^{-1} \sum_{j=1}^{T} E[(\sum_{i=1}^{t} u_i)_j^2]$.

"Phillips-Perron":

1. Estimate $\hat{\rho}$, then
2. use an empirical estimate of σ^2 and σ_e^2 to adjust the statistic ("Z_ρ" and "Z_τ").

Other Unit Root Tests: KPSS

Kwiatkowski et al. ("KPSS"):

$$LM = \sum_{t=1}^{T} \frac{S_t^2}{\hat{\sigma}_\epsilon^2}$$

where $S_t^2 = \sum_{i=1}^{t} \hat{u}_i$ and $\hat{\sigma}_\epsilon^2$ is $\widehat{\mathrm{Var}(\epsilon_t)}$ from:

$$Y_t = \alpha + \epsilon_t$$

or:

$$Y_t = \alpha + \beta t + \epsilon_t$$

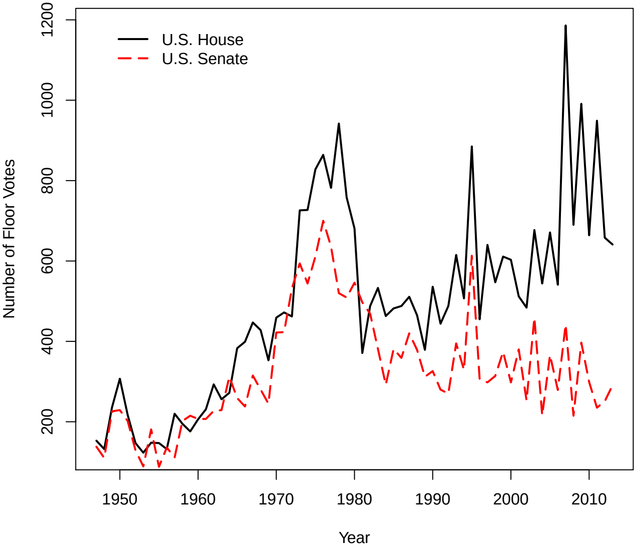
Variance-Ratio Tests

- Idea: if a series is stationary, the variance of the series is not increasing over time.
- Intuition: Compare the variance of a subset of the data "early" in the series with a similarly-sized subset "later" in the process. E.g.,

$$\hat{\rho} = \frac{1}{T} \sum \frac{[\sum (Y)^2]}{\sum (Y^2)}$$

- References: Hamilton (p. 531-32), Cochrane (1988), Lo and McKinlay (1988), Cecchetti and Lam (1991), etc.

House and Senate Votes, 1946-2013



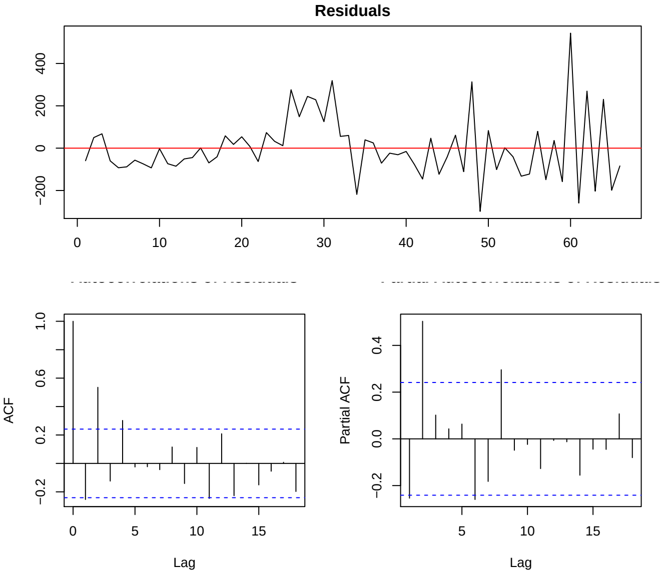
Dickey-Fuller Tests

```
> HDF<-ur.df(HVotes.TS,type="none",lags=0)
> summary(HDF)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ˜ z.lag.1 - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-449.76  -48.99   25.39   91.19  666.09

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
z.lag.1 -0.03899    0.03846  -1.014    0.315

Residual standard error: 170.5 on 65 degrees of freedom
Multiple R-squared:  0.01556, Adjusted R-squared:  0.0004131
F-statistic: 1.027 on 1 and 65 DF,  p-value: 0.3146


Value of test-statistic is: -1.0135

Critical values for test statistics:
     1pct  5pct 10pct
tau1 -2.6 -1.95 -1.61
```

Dickey-Fuller Tests (continued)

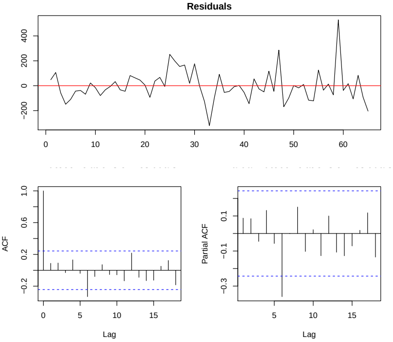
```
> HDF.D<-ur.df(HVotes.TS,type="drift",lags=0)
> summary(HDF.D)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Value of test-statistic is: -3.2294 5.2847

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86


> HDF.T<-ur.df(HVotes.TS,type="trend",lags=0)
> summary(HDF.T)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Value of test-statistic is: -4.7424 7.576 11.2834

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.04 -3.45 -3.15
phi2  6.50  4.88  4.16
phi3  8.73  6.49  5.47
```

Plotting...


Residuals
400
200
0
−200
0
10
20
30
40
50
60
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag
Partial ACF
0.4
0.2
0.0
−0.2
5
10
15
Lag

Augmented D-F Tests

```
> HADF.T1<-ur.df(HVotes.TS,type="trend",lags=1)
> summary(HADF.T1)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Value of test-statistic is: -2.112 1.8218 2.3959

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.04 -3.45 -3.15
phi2  6.50  4.88  4.16
phi3  8.73  6.49  5.47
```



ADF Tests, continued...

```
> HADF.BIC<-ur.df(HVotes.TS,type="trend",selectlags="BIC")
> summary(HADF.BIC)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Call:
lm(formula = z.diff ˜ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-320.30  -68.08  -15.52   48.23  528.75

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  72.1790    38.4963   1.875   0.0656 .
z.lag.1      -0.2368     0.1121  -2.112   0.0388 *
tt            1.6859     1.3295   1.268   0.2096
z.diff.lag   -0.5465     0.1081  -5.055 4.19e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 127.5 on 61 degrees of freedom
Multiple R-squared:  0.4823, Adjusted R-squared:  0.4568
F-statistic: 18.94 on 3 and 61 DF,  p-value: 8.484e-09


Value of test-statistic is: -2.112 1.8218 2.3959

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.04 -3.45 -3.15
phi2  6.50  4.88  4.16
phi3  8.73  6.49  5.47
```

Phillips-Perron Test

```
> HPP <- ur.pp(HVotes.TS, type="Z-tau",model="trend",lags="short")
> summary(HPP)

##################################
# Phillips-Perron Unit Root Test #
##################################

Call:
lm(formula = y ~ y.l1 + trend)

Residuals:
    Min      1Q  Median      3Q     Max
-298.55  -84.95  -27.39   57.55  543.23

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 265.9933    57.5321   4.623 1.92e-05 ***
y.l1          0.4682     0.1121   4.176 9.30e-05 ***
trend         4.5726     1.4060   3.252  0.00184 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 149.6 on 63 degrees of freedom
Multiple R-squared:  0.6157, Adjusted R-squared:  0.6035
F-statistic: 50.47 on 2 and 63 DF,  p-value: 8.267e-14

Value of test-statistic, type: Z-tau  is: -4.8552

           aux. Z statistics
Z-tau-mu               4.122
Z-tau-beta             3.356

Critical values for Z statistics:
                     1pct     5pct     10pct
critical values -4.101251 -3.47789 -3.166276
```

Plotting...

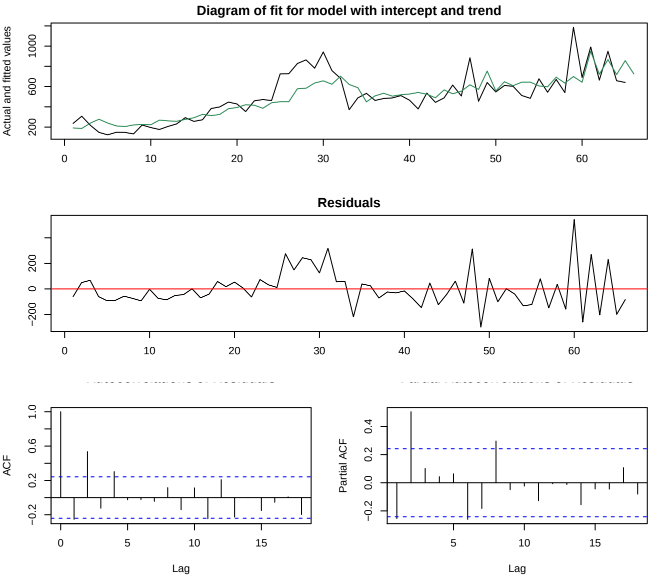
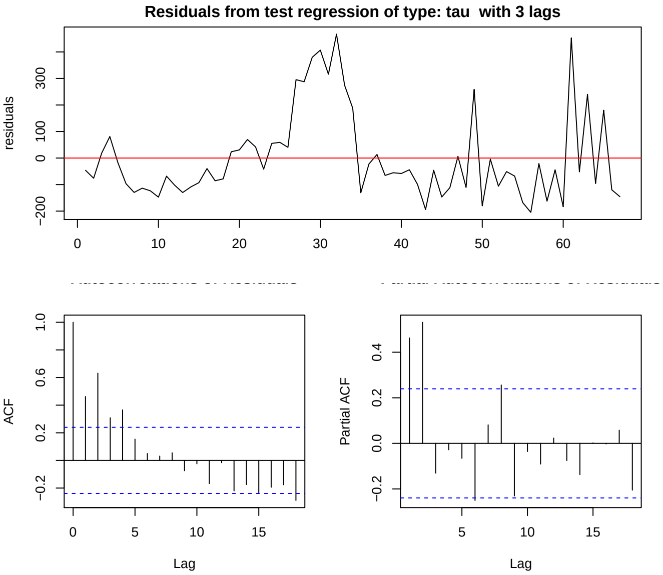


Diagram of fit for model with intercept and trend
Actual and fitted values
200
600
1000
0
10
20
30
40
50
60
Residuals
−200
0
200
0
10
20
30
40
50
60
ACF
−0.2
0.2
0.6
1.0
0
5
10
15
Lag
Partial ACF
−0.2
0.0
0.2
0.4
5
10
15
Lag

KPSS Test (null = stationarity)

```
> HKPSS <- ur.kpss(HVotes.TS,type="tau",lags="short")
> summary(HKPSS)

#######################
# KPSS Unit Root Test #
#######################

Test is of type: tau with 3 lags.

Value of test-statistic is: 0.1568

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

Plotting...


Residuals from test regression of type: tau with 3 lags
residuals
−200
0
100
300
0
10
20
30
40
50
60
ACF
−0.2
0.2
0.6
1.0
0
5
10
15
Lag
Partial ACF
−0.2
0.0
0.2
0.4
5
10
15
Lag

Variance-Ratio Test

```
> library(egcm)

> H.VRatio <- bvr.test(HVotes.TS, detrend=TRUE)

> H.VRatio

 Breitung Variance Ratio Test for a Unit Root

data:  HVotes.TS
rho = 0.0058012, p-value = 0.1872
alternative hypothesis: stationary
```

General Issues

- Choosing Lag Lengths...
- “Knife-Edge” Tests and Bayesians vs. Frequentists...
- Why bother? Differencing...

Unit Roots in R

Packages:

- `tseries`
- `uroot` (including seasonality)
- `urca`
- `egcm`
- `CADFtest`

Fundamentals

Consider:

$$Y_t = \rho Y_{t-1} + u_t$$

Properties:

1. Linear transformations of Y_t preserve the level of integration.
2. For two I(0) series X_t and Y_t, $aX_t + bY_t + c$ is also I(0).
3. For two I(1) series X_t and Y_t, $aX_t + bY_t + c$ is *generally* I(1) as well.
4. If X_t is I(0) and Y_t is I(1), then $aX_t + bY_t + c$ is I(1) (*integration dominates stationarity*).

More Fundamentals

Now consider:

$$\begin{aligned} X_t &= W_t + u_{Xt} \\ Y_t &= AW_t + u_{Yt}, \\ W_t &\sim I(1), \\ u_{Xt}, u_{Yt} &\sim I(0) \end{aligned}$$

Here, both X_t and Y_t will be I(1), but:

$$\begin{aligned} Z_t &= Y_t - AX_t \\ &= [AW_t + u_{Yt}] - A[W_t + u_{Xt}] \\ &= AW_t + u_{Yt} - AW_t - Au_{Xt} \\ &= u_{Yt} - Au_{Xt} \end{aligned}$$

is I(0).

Cointegration

If there exist two I(1) series X_t, Y_t such that $Z_t = \mu + aX_t + bY_t$ is I(0), then X_t and Y_t are said to be cointegrated.

Intuition:

- The nonstationarity in X and Y arises from a *common component* (W_t) which is I(1).
- Combining X and Y cancels out the common part, leaving I(0) "residuals."

Cointegration as an Attractor

Above:

$$X_t = AY_t$$

means that:

$$Z_t = X_t - AY_t \;\sim\; \text{i.i.d.}(0, \sigma_Z^2)$$

Examples:

1. Murray (1994): Drunks and puppies...
2. Regional commodity prices...

Cointegration in Practice

First: Assess whether X and Y are I(1).

Then rewrite:

$$Z_t = X_t - \alpha Y_t$$

as the cointegrating regression:

$$X_t = \alpha Y_t + Z_t.$$

Cointegration implies:

$$\begin{aligned} Z_t &\sim I(0), \text{ and} \\ \text{Var}(Z_t) &< \infty. \end{aligned}$$

Cointegration: A Roadmap

1. Determine that the two series are I(1),
2. Estimate $X_t = \hat{\alpha} Y_t + Z_t$ using OLS,
3. Examine $\hat{Z}_t$ for stationarity, using
    - a Durbin-Watson test ("CRDW"), and/or
    - standard unit root tests (DF, ADF, KPSS, etc.)

Wait, What? OLS???

Recall that for cointegrated X_t and Y_t, $\text{Var}(Z_t) < \infty$.

Suppose we (incorrectly) estimate δ:

$$\tilde{Z}_t = X_t - \delta Y_t$$

Then $\tilde{Z}_t$ is I(1), and $\lim_{T \to \infty} [\text{Var}(\tilde{Z}_t)] = \infty$, while in finite samples:

$$\text{Var}(\tilde{Z}_t) > \text{Var}(\hat{Z}_t),$$

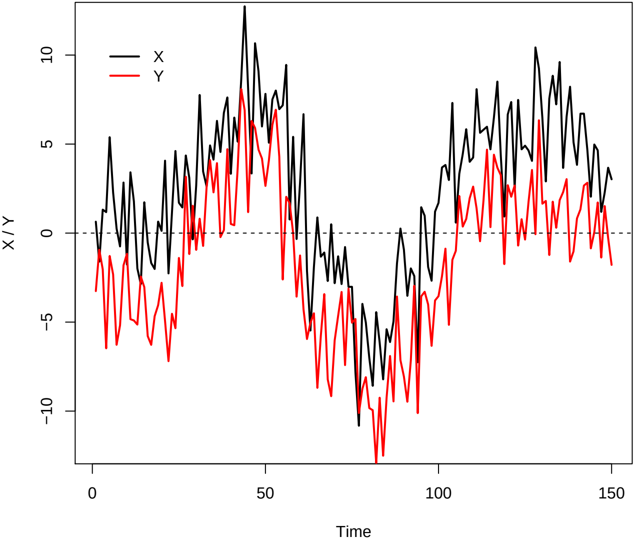
Thus:

- OLS minimizes residual variance →
- OLS makes $\hat{Z}_t$ stationary [and makes $\text{Var}(Z_t)$ finite] →
- $\hat{\alpha}_{OLS} \to \alpha$ as $T \to \infty$ ("superconsistent").

A Simulated Example

```
> T <- 150
>
> set.seed(7222009)
>
> W <- cumsum(rnorm(T)) # I(1)
> X <- 2 + 0.8*W + 2*rnorm(T)
> Y <- -2 + 0.8*W + 2*rnorm(T)
>
> W.TS <- ts(W,start=1,end=T) # time series objects
> X.TS <- ts(X,start=1,end=T)
> Y.TS <- ts(Y,start=1,end=T)
```

Our Series



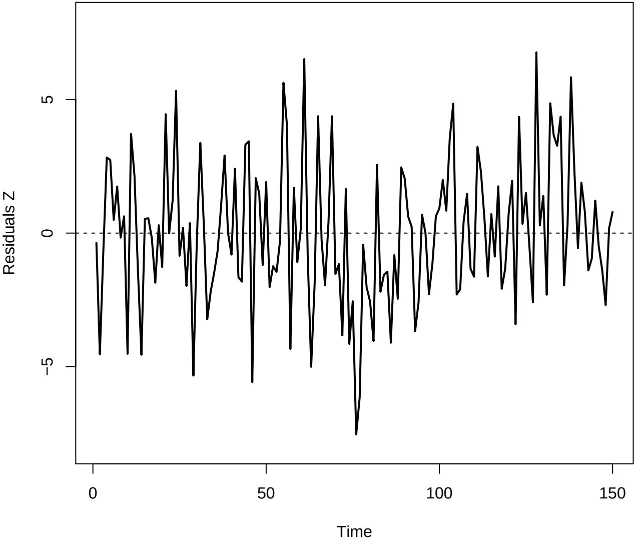
Unit Root Tests

```
> summary(ur.df(X.TS,type="trend",lags=1))  ### ADF

Value of test-statistic is: -3.2619 3.5655 5.3352

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47

> summary(ur.df(Y.TS,type="trend",lags=1))

Value of test-statistic is: -2.6832 2.4 3.5999

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47


> summary(ur.kpss(X.TS,type="tau",lags="short"))  ### KPSS

Value of test-statistic is: 0.3143

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216

> summary(ur.kpss(Y.TS,type="tau",lags="short"))

Value of test-statistic is: 0.2913

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

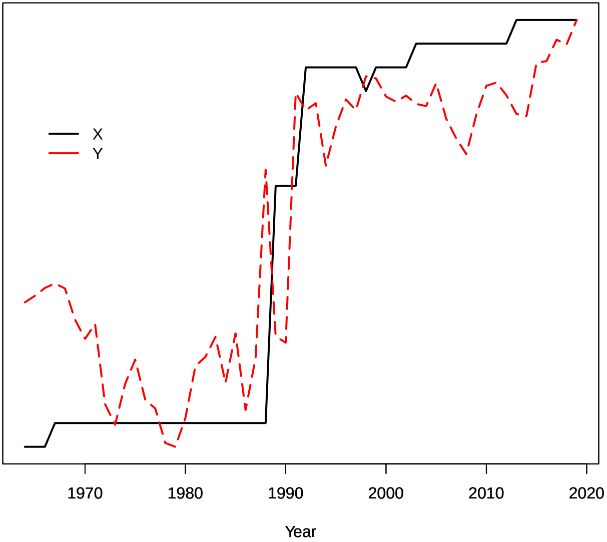
CI Regression + Residual Check

```
> CI.reg <- lm(X~Y)
> # summary(CI.reg)
>
> Zhats.TS <- ts(CI.reg$residuals,start=1,end=T)
>
> summary(ur.df(Zhats.TS,type="trend",lags=1)) ### ADF

Value of test-statistic is: -9.0599 27.3846 41.0621

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47

> summary(ur.kpss(Zhats.TS,type="tau",lags="short"))  ### KPSS

Value of test-statistic is: 0.1625

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

Residuals from Cointegrating Regression



Two (More) Series, 1964-2019



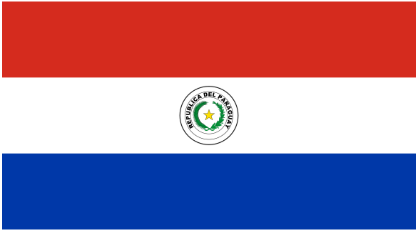
A Regression

```
> summary(lm(Y.TS~X.TS))

Call:
lm(formula = Y.TS ~ X.TS)

Residuals:
    Min      1Q  Median      3Q     Max
-3751.1  -806.2   108.1   597.6  4657.5

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 21177.78     194.42  108.93   <2e-16 ***
X.TS          403.65      24.97   16.17   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1453 on 54 degrees of freedom
Multiple R-squared:  0.8288, Adjusted R-squared:  0.8256
F-statistic: 261.4 on 1 and 54 DF,  p-value: < 2.2e-16
```

Another Regression (Lagged X)

```
> summary(dyn$lm(Y.TS~lag(X.TS,-1)))  ### Lagged X

Call:
lm(formula = dyn(Y.TS ~ lag(X.TS, -1)))

Residuals:
    Min      1Q  Median      3Q     Max
-3864.3  -787.9   138.0   594.6  4613.1

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   21277.22     188.35  112.96   <2e-16 ***
lag(X.TS, -1)   410.54      24.26   16.92   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1396 on 53 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.8438, Adjusted R-squared:  0.8409
F-statistic: 286.3 on 1 and 53 DF,  p-value: < 2.2e-16
```

X = Ford Mustang prices, in constant $US



Y = Paraguay's POLITY IV (democracy / autocracy) score

Time Series Regression

Lagged Xs:

$$Y_t = \alpha + \beta X_{t-1} + u_t$$

often yield autocorrelated errors / spurious regressions.

vs. "difference equations":

$$\Delta Y_t = \alpha + \beta \Delta X_{t-1} + u_t$$

where all effects are "short term."

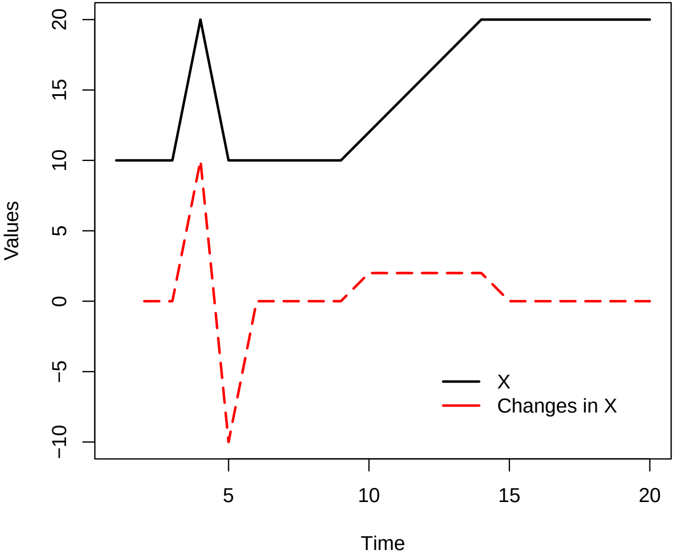
Error-Correction Models

$$\Delta Y_t = \beta \Delta X_{t-1} + \rho(Y_{t-1} - \alpha - \gamma X_{t-1})$$

Terms:

- $\hat{\beta} \rightarrow$ short-term relationship between X and Y.
- $\hat{\alpha}$ and $\hat{\gamma} \rightarrow$ long-term relationship between X and Y (the "attractor" – the equilibrium distance between X and Y).
- $\hat{\rho} \rightarrow$ the rate at which the model "reequilibrates" (formally,the proportion of the disequilibrium which is corrected with each period)

ECM: Toy Example



Two Cointegrated Series

Two examples:

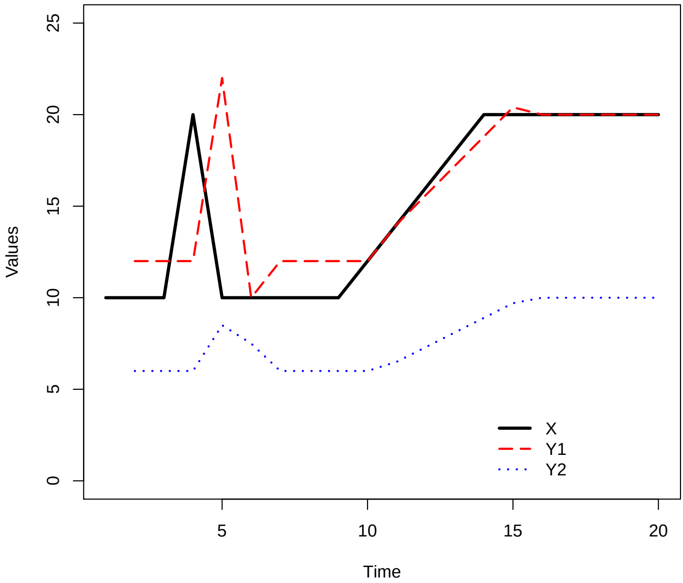
$$\Delta Y_{1t} = 1.0\Delta X_{t-1} - 0.8(Y_{1t-1} - 5.0 - 1.0X_{t-1})$$

$$\Delta Y_{2t} = 0.25\Delta X_{t-1} - 0.2(Y_{2t-1} - 10.0 - 2.0X_{t-1})$$

Note:

- ΔY_1 has a larger degree of short-term dependence on X.
- The "equilibrium distance" for ΔY_1 is $Y_t = 5 + X_t$; that in ΔY_2 is $Y_t = 10 + 2X_t$.
- Y_1 reequilibrates at a much faster rate than does Y_2.

Pictures!



Estimation and Testing: Two-Steps...

If both X_t and Y_t are cointegrated, then a general ECM is:

$$\begin{aligned}\Delta Y_t &= \beta_0 + \beta_1 \Delta X_{t-1} + \beta_2 \Delta X_{t-2} + ... + \beta_k \Delta X_{t-k} + \rho Z_{t-1} + u_t \\ &= \beta_0 + \beta_1 \Delta X_{t-1} + \beta_2 \Delta X_{t-2} + ... + \beta_k \Delta X_{t-k} + \rho (X_t - \alpha Y_t) + u_t\end{aligned}$$

where the ΔXs capture the short-term dependence between X and Y.

The Engle-Granger Two-Step:

1. Estimate the cointegrating regression $Y_t = \alpha + \gamma X_t + e_t$,
2. From these estimates, generate $\hat{Z}_t = Y_t - \hat{\alpha} - \hat{\gamma} X_t$,
3. Include $\hat{Z}_{t-1}$ for Z_{t-1} in the model above.

The One-Step Approach

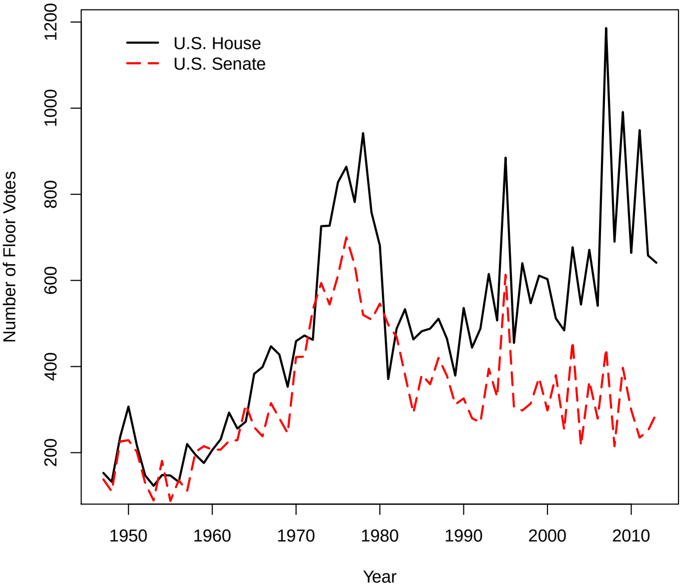
Substituting $Z_t = X_t - \alpha Y_t$ into the equation above, we get:

$$\begin{aligned}
\Delta Y_t &= \beta_0 + \beta_1 \Delta X_{t-1} + \rho(Y_{t-1} - \alpha - \gamma X_{t-1}) + u_t \\
&= (\beta_0 - \rho\alpha) + \beta_1 \Delta X_{t-1} + \rho Y_{t-1} - \rho\gamma X_{t-1} + u_t
\end{aligned}$$

Suggests a single equation, where changes in Y_t are a function of:

- Lagged ΔXs,
- X_{t-1}, and
- Y_{t-1}.

Example: House and Senate Votes, 1946-2013



Unit Roots? (A: Yes)

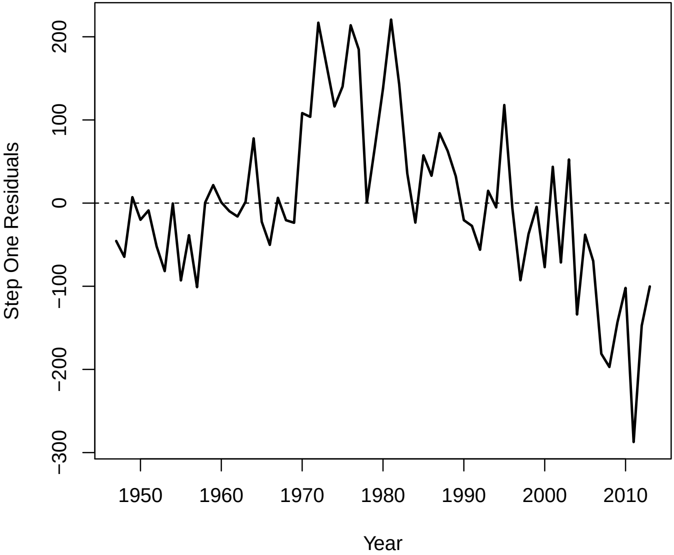
```
> summary(ur.df(HVotes.TS,type="trend",lags=1))

Value of test-statistic is: -2.112 1.8218 2.3959

> summary(ur.df(SVotes.TS,type="trend",lags=1))

Value of test-statistic is: -1.6975 1.3998 2.033

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.04 -3.45 -3.15
phi2  6.50  4.88  4.16
phi3  8.73  6.49  5.47
```

```
> summary(ur.kpss(HVotes.TS,type="tau",lags="short"))

Value of test-statistic is: 0.1568

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216

> summary(ur.kpss(SVotes.TS,type="tau",lags="short"))

Value of test-statistic is: 0.3122

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

ECM, Step One

```
> StepOne <- lm(SVotes.TS˜HVotes.TS) # CI regression
> summary(StepOne)

Call:
lm(formula = SVotes.TS ˜ HVotes.TS)

Residuals:
     Min       1Q   Median       3Q      Max
-287.302  -54.055   -6.071   54.946  220.677

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.43749   28.20367   4.199 8.31e-05 ***
HVotes.TS     0.42557    0.05154   8.257 1.02e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 100.3 on 65 degrees of freedom
Multiple R-squared:  0.5119, Adjusted R-squared:  0.5044
F-statistic: 68.18 on 1 and 65 DF,  p-value: 1.021e-11
```

Step One Residuals...☹



ECM: Step Two

```
> DSV.TS <- diff(SVotes.TS)  # difference Y
> DHVLag.TS <- lag(diff(HVotes.TS),k=-1) # Lag differenced X
> Ztminus1.TS <- lag(Zt.TS,k=-1) # Lag residuals
> df <- ts.intersect(DSV.TS,DHVLag.TS,Ztminus1.TS)
```

```
> StepTwo <- lm(DSV.TS ˜ DHVLag.TS + Ztminus1.TS,
+               data = df)
> summary(StepTwo)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.56516   10.93399   0.509   0.6126
DHVLag.TS   -0.27917    0.06362  -4.388 4.53e-05 ***
Ztminus1.TS -0.27269    0.10992  -2.481   0.0158 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 88.04 on 62 degrees of freedom
Multiple R-squared:  0.2875, Adjusted R-squared:  0.2645
F-statistic: 12.51 on 2 and 62 DF,  p-value: 2.734e-05
```

What Does That Mean?

The fitted ECM is:

$$\begin{aligned}\Delta\text{Senate Votes}_t &= 5.56 - 0.28\Delta\text{House Votes}_{t-1} \\ &\quad -0.27[\text{Senate Votes}_{t-1} - 118.4 - 0.43(\text{House Votes}_{t-1})]\end{aligned}$$

This means that:

- The "short term" relationship between Senate and House votes has about 0.28.
- Following a "shock," the relationship between the two returns to its equilibrium level at a rate of about 27 percent per year.

One-Step ECM

```
> SVLag.TS <- lag(SVotes.TS,k=-1) # Lag Y
> HVLag.TS <- lag(HVotes.TS,k=-1) # Lag X
> df2 <- ts.intersect(DSV.TS,DHVLag.TS,SVLag.TS,HVLag.TS)
> OneStep <- lm(DSV.TS~DHVLag.TS+SVLag.TS+HVLag.TS,
+               data=df2)

> summary(OneStep)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 67.79858   29.46161   2.301 0.024808 *
DHVLag.TS   -0.25004    0.06750  -3.704 0.000458 ***
SVLag.TS    -0.27186    0.10943  -2.484 0.015744 *
HVLag.TS     0.05466    0.06773   0.807 0.422782
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 87.64 on 61 degrees of freedom
Multiple R-squared:  0.3052, Adjusted R-squared:  0.2711
F-statistic: 8.933 on 3 and 61 DF,  p-value: 5.411e-05
```

Other Topics...

- Distributed lag models
- ("Granger") Causality
- Intervention analysis, "structural breaks," and threshold models
- Vector autoregression (VAR) models
- Autoregressive Conditional Heteroscedasticity (ARCH and GARCH) models
- Fractional integration / long-memoried series
- Spectral domain analysis
- Time series models for binary, ordered, event count, etc. outcomes
- Models for spatio-temporal data

Resources

Some favorite references:

- Box-Steffensmeier, Janet M., John R. Freeman, Matthew P. Hitt, and Jon C. Pevehouse. 2015. *Time Series Analysis for the Social Sciences*. New York: Cambridge University Press.
- Fuller, W.A. 1996. *Introduction to Statistical Time Series*, 2nd ed. New York: Wiley.
- Pickup, Mark. 2014. *Introduction to Time Series Analysis*. New York: SAGE Publications. Quantitative Applications in the Social Sciences.
- Shumway, Robert H., and David S. Stoffer. 2016. *Time Series Analysis and Its Applications, With R Examples*. New York: Springer.

Syllabi (examples; some also include panel data models):

- Adolph (2024)
- Enns (2023)
- Mitchell (2020)
- Philips (2022)