CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

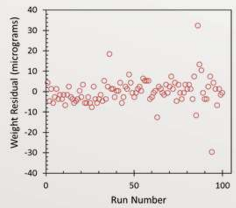
Lecture 39
Autocorrelation in Time Series

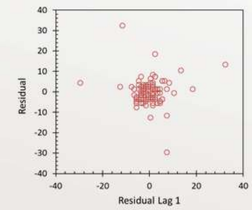
Chris A. Mack
Adjunct Associate Professor

http://www.lithoguru.com/scientist/statistics/

Lag Plots

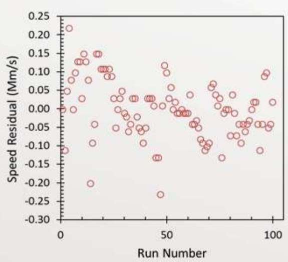
- A plot of ε_i versus ε_{i-1} (when residuals are ordered in time or other natural sequence) helps to discover correlations between a residual and its preceding residual
  – A lag of n plots ε_i versus ε_{i-n}

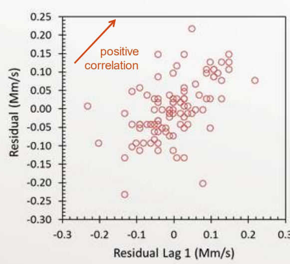




NBS measurements for a standard weight (Data Set 1)

Lag Plots





Michelson speed of light measurements (Data Set 1)

Modeling Autocorrelation

- The past is sometimes the best predictor of the future
- Consider the first-order autoregressive model AR(1):
  – $y_i = f(x_i) + \varepsilon_i$
  – $\varepsilon_i = \rho\varepsilon_{i-1} + u_i,\ u_i \sim N(0, \sigma^2)$ iid
  – ρ = autocorrelation parameter, $|\rho| < 1$
    - $\rho = 0$ means no autocorrelation
    - $\rho > 0$ means positive autocorrelation ("persistence")

AR(1) Properties

- Properties of the error terms:
  – $E[\varepsilon_i] = 0$
  – $var[\varepsilon_i] = \sigma^2/(1-\rho^2)$
  – $cov[\varepsilon_i, \varepsilon_{i-1}] = \rho\sigma^2/(1-\rho^2) = \rho var[\varepsilon_i]$
  – $cov[\varepsilon_i, \varepsilon_{i-s}] = \rho^s\sigma^2/(1-\rho^2) = \rho^s var[\varepsilon_i]$, for $s > 0$
- Two questions:
  – Can we test for autocorrelation?
  – How can we estimate ρ?

Durbin-Watson Test

- We wish to perform a hypothesis test:
  – $H_0: \rho = 0$
  – $H_A: \rho > 0$ (negative autocorrelation is rare)
- Define the Durbin-Watson statistic

$$D = \frac{\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2}$$

Based on the raw residuals

- If D is smaller than a critical value, we reject the null hypothesis

Durbin-Watson Test

- We can't obtain an exact critical value for D. Instead, we define upper (d_U) and lower (d_L) critical values so that
  - $D > d_U$, can't reject H_0
  - $D < d_L$, conclude H_A
  - $d_L < D < d_U$, test is inconclusive
- Critical values are from
  - N.E. Savin and K.J. White, "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors," *Econometrica*, **45**, 1989-1996 (1977).

Durbin-Watson Test

- This test only checks for autocorrelation with a lag of 1; longer lags may not be detected
- Note that $D \approx 2(1-\hat{\rho})$
  - When $D \approx 2$ we have no autocorrelation
  - As $D \to 0$ we approach perfect autocorrelation
- While negative autocorrelation is rare, we can test for it
  - Test statistic is $4 - D$
  - Use the Durbin-Watson critical tables as before

Consequences of Autocorrelation

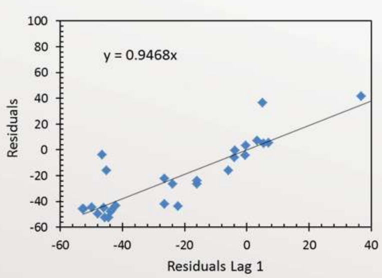
- Regression coefficients remain unbiased but are no longer minimum variance estimates
- For positive autocorrelation, our MSE estimate of the residual variance will underestimate the true variance (sometimes seriously)
  - Estimated standard errors for regression coefficients will be too small
  - Statistical tests (such as F or t) on the model will not be appropriate

How Do We Estimate ρ?

- If the Durbin-Watson test detects autocorrelated residuals, we have three common ways to estimate ρ:
  - Slope of residual lag plot (through the origin)
  - Correlation(e_i,e_{i-1}): $r = s_{xy}/(s_x s_y)$ (preferred)
  - Use $r' = \frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2}$

$$slope \approx r\left(1 + \frac{e_n^2 - e_1^2}{2SSE}\right), \qquad r' \approx r\left(1 - \frac{e_n^2 + e_1^2}{2SSE}\right)$$

Estimating the Autocorrelation Coefficient



Correlation of e_i,e_{i-1} = 0.923

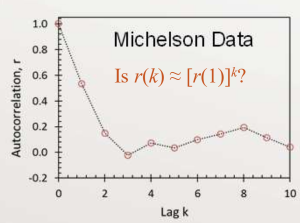
$$\frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2} = 0.846$$

Flow Meter Accumulated Liquid vs. Temperature (Data Set 2)

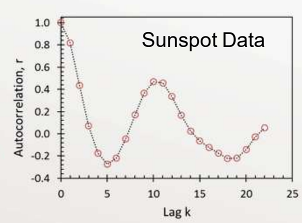
Is AR(1) Appropriate?

Called the ACF, the autocorrelation function

- Determine $r(k)$, r = correlation coefficient, k = lag



AR(1) is appropriate



AR(1) is not appropriate

How to Fix Autocorrelated Residuals

- First, look for an improved model
  - Missing predictor variables often show up as autocorrelated residuals (especially if experiments are not randomized): an artificial time dependence
- If autocorrelation is a function of drift, aging, or other such behavior, then
  - Improve the measurement/experiment to remove the time/space/order dependence
  - Use an autoregressive model (last resort)

First-Order Autoregressive Correction

- For a linear AR(1) model,
  - $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
  - $\varepsilon_i = \rho\varepsilon_{i-1} + u_i$, $u_i \sim N(0, \sigma^2)$ iid
- Define transformed variables
  - $y'_i = y_i - \rho y_{i-1}$, $x'_i = x_i - \rho x_{i-1}$
  - $\beta'_0 = \beta_0(1-\rho)$, $\beta'_1 = \beta_1$
- Then, the transformed regression model is
  - $y'_i = \beta'_0 + \beta'_1 x'_i + u_i$

First-Order Autoregressive Correction

- Perform a standard (OLS) regression on the transformed data, then
  - $b_0 = b'_0/(1-r)$, $SE[b_0] = SE[b'_0]/(1-r)$
  - $b_1 = b'_1$, $SE[b_1] = SE[b'_1]$
  - r = our estimate of ρ

The final model:

$$y_i = r\,y_{i-1} + (1-r)b_0 + b_1(x_i - r\,x_{i-1}) + u_i$$

More on Autoregressive Correction

- The Cochrane-Orcutt procedure: test the residuals after the autoregressive transformation and linear regression
  - Does the Durbin-Watson test show it is clear of autocorrelated residual behavior?
  - If not, iterate the whole procedure
- Alternate approaches
  - Just assume $\rho \approx 1$, see what happens
  - Find an MLE of ρ while fitting model (nonlinear regression)

Lecture 39: What have we learned?

- What is an AR(1) model?
- What is the Durbin-Watson test and what can it tell you?
- How can we estimate the correlation parameter ρ of an AR(1) model?
- How can we correct for autocorrelation when doing a regression?