Chapter 9

Serially Correlated Errors

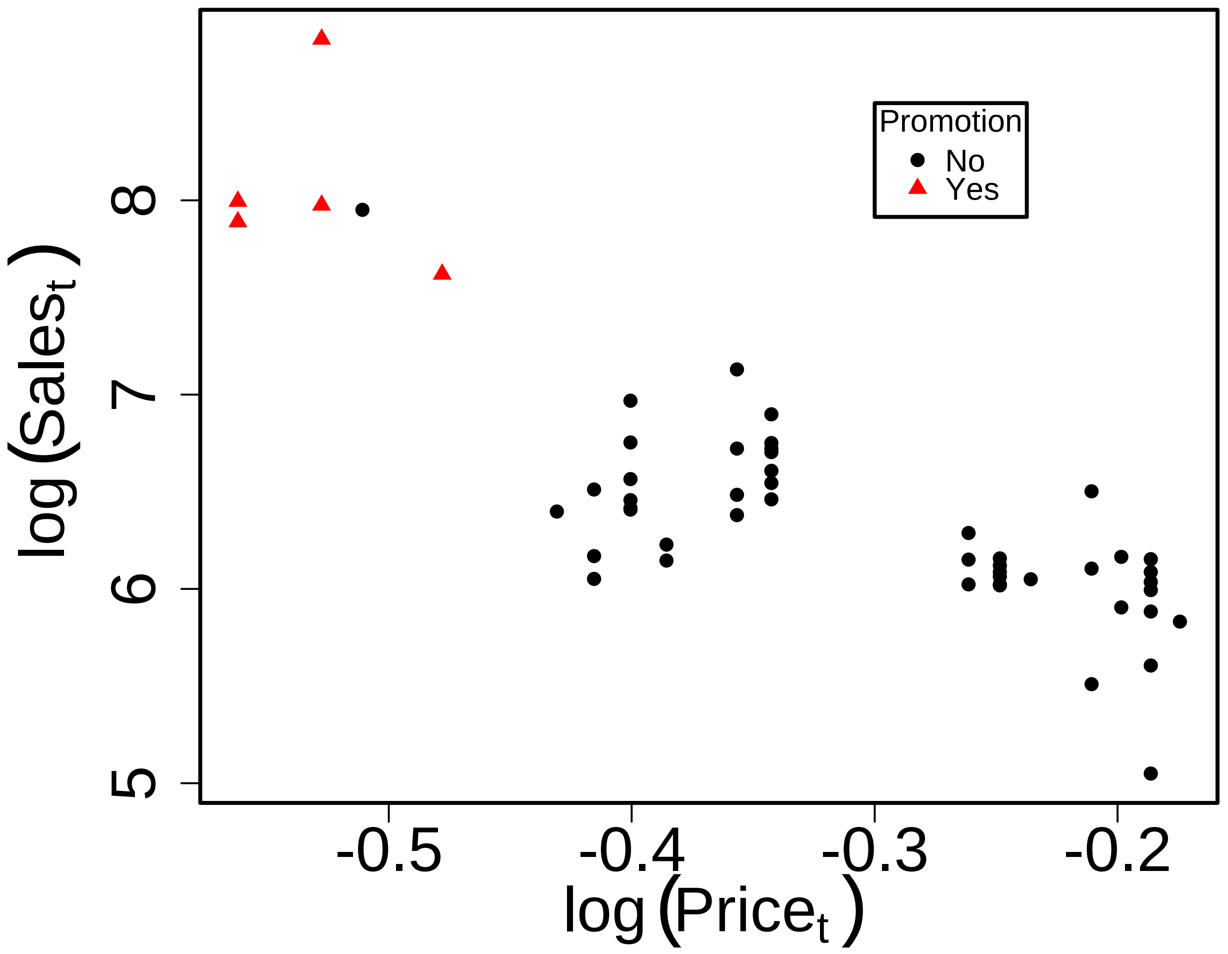
Serially Correlated Errors

- Data are often collected over time.
- Our assumption so far has been 0 correlation among the errors.
- Now we use Generalized Least Squares to fit models with autocorrelated errors.

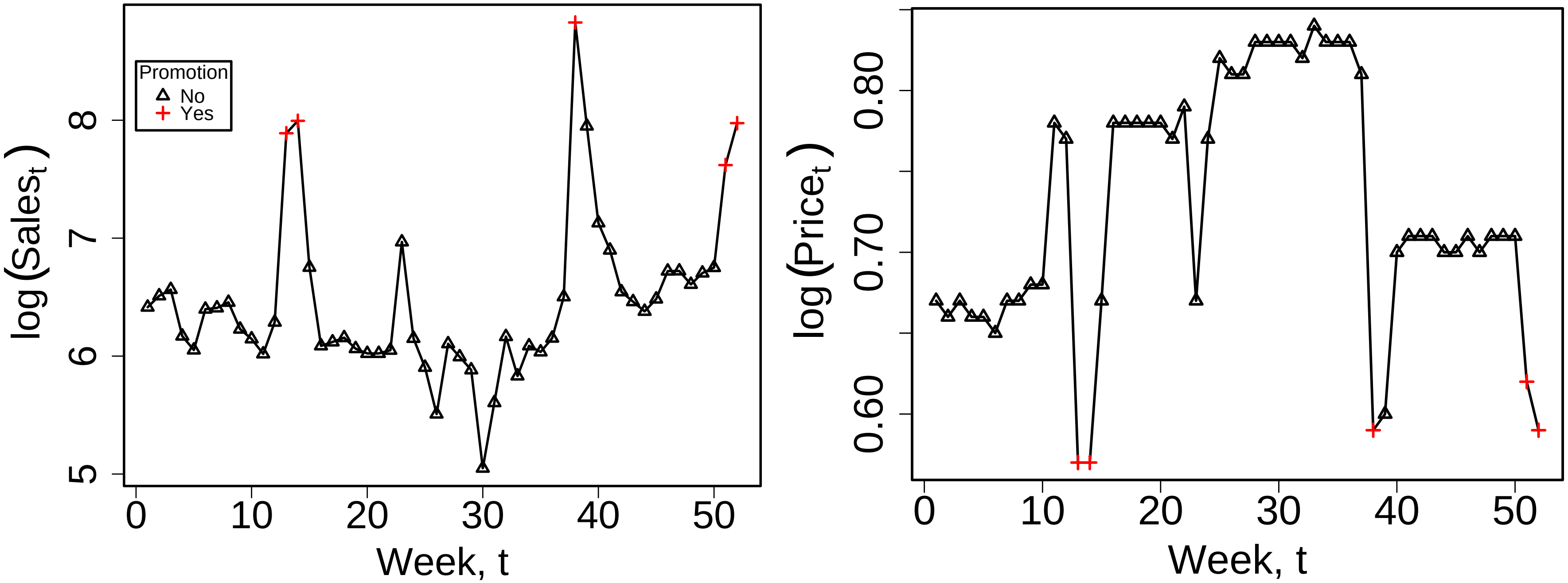
What is Statistics?

- Separating noise from pattern.
- Two main areas
  - Estimation: More difficult with serial correlation
  - Prediction: Easier with serial correlation

Example: Food Sales at the Grocery Store

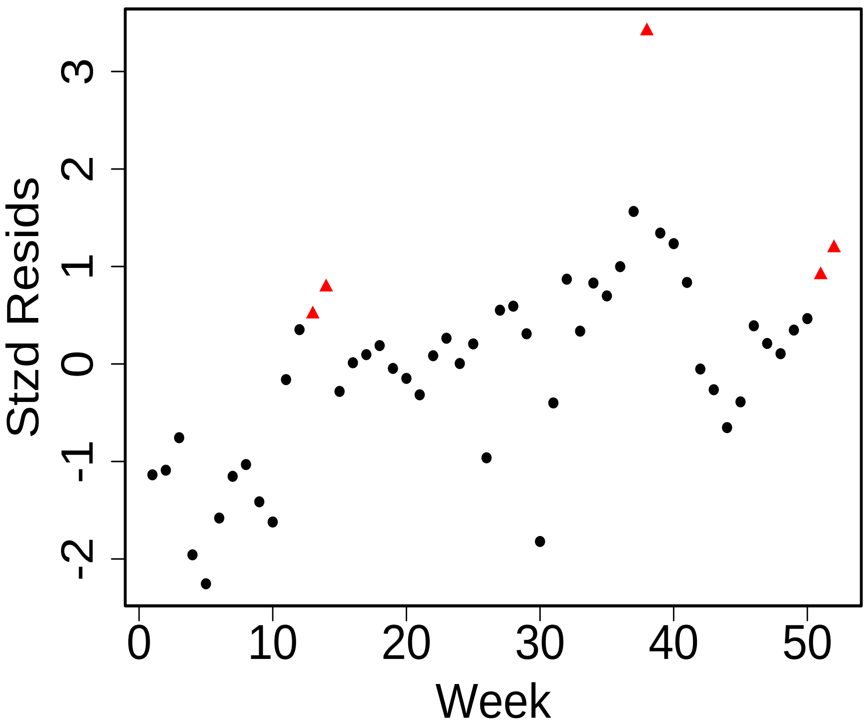


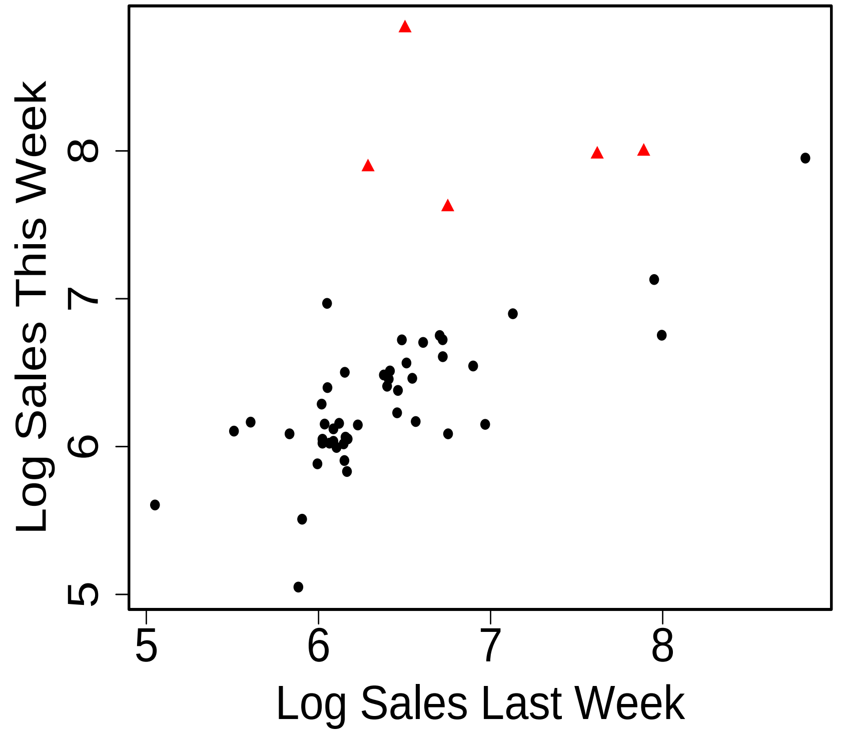
Example: Food Sales



Example: Food Sales

Model: log(Sales) = $\beta_0 + \beta_1$ log(Price) + e





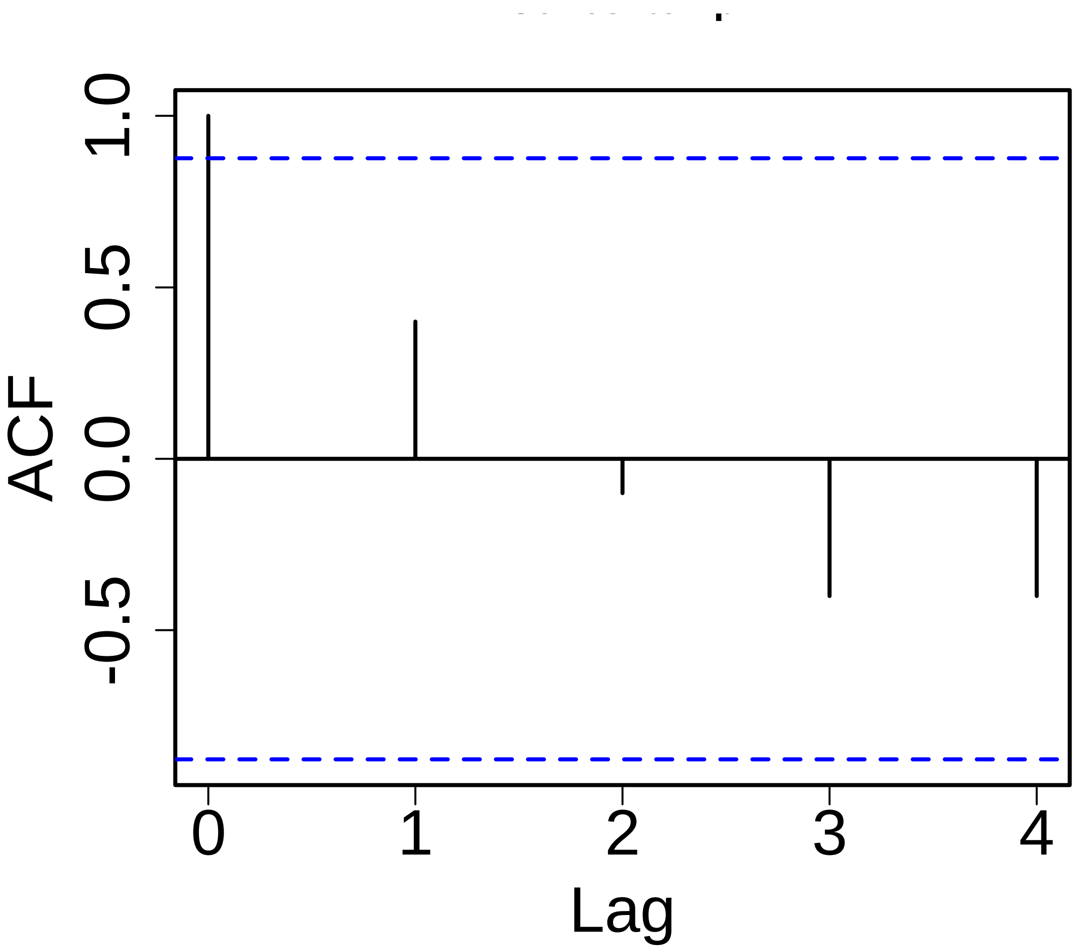
Autocorrelation Example

- Weather in College Station: 85, 86, 87, 88, 89 degrees Fahrenheit.

$$\text{Autocorrleation}(l) = \frac{\sum_{t=l+1}^{n} (y_t - \bar{y})(y_{t-l} - \bar{y})}{\sum_{t=n}^{n} (y_t - \bar{y})^2}$$

- We use subscript t to denote time.

Autocorrelation Example

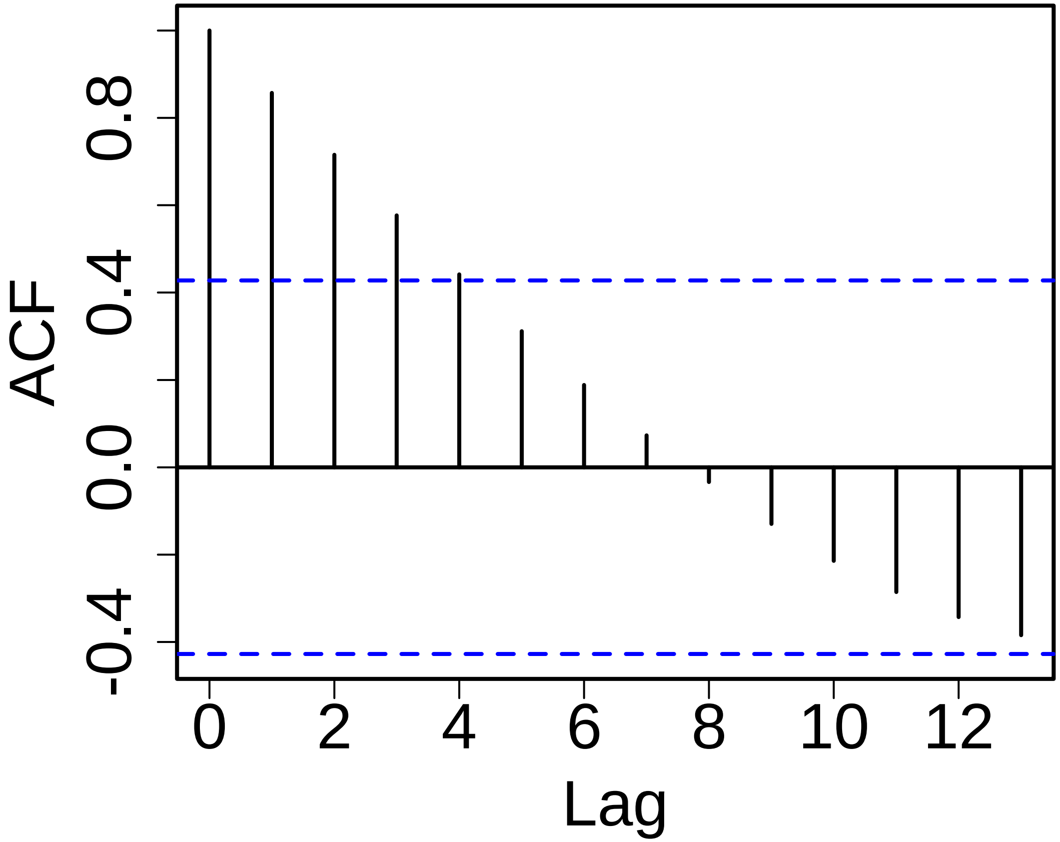


Dashed lines correspond to -2/√n and + 2/√n.

Another option: the Durbin-Watson statistic to test for significance.

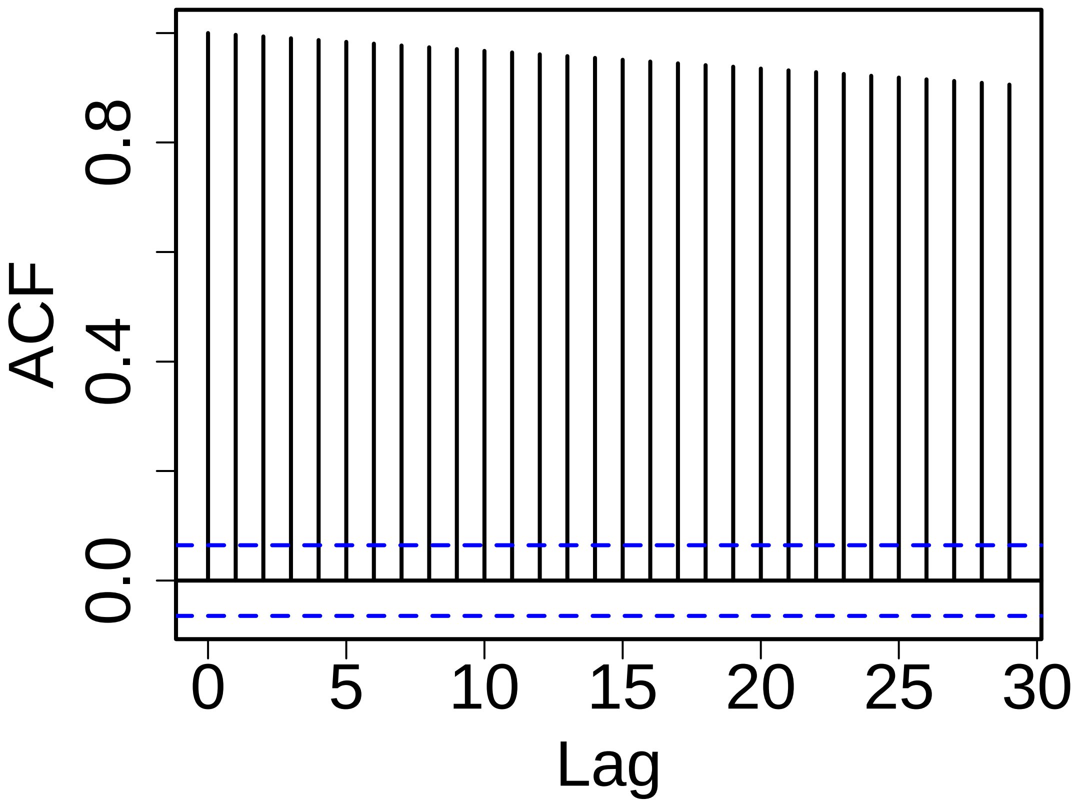
Autocorrelation Example

- Weather now 85, 86, ..., 105

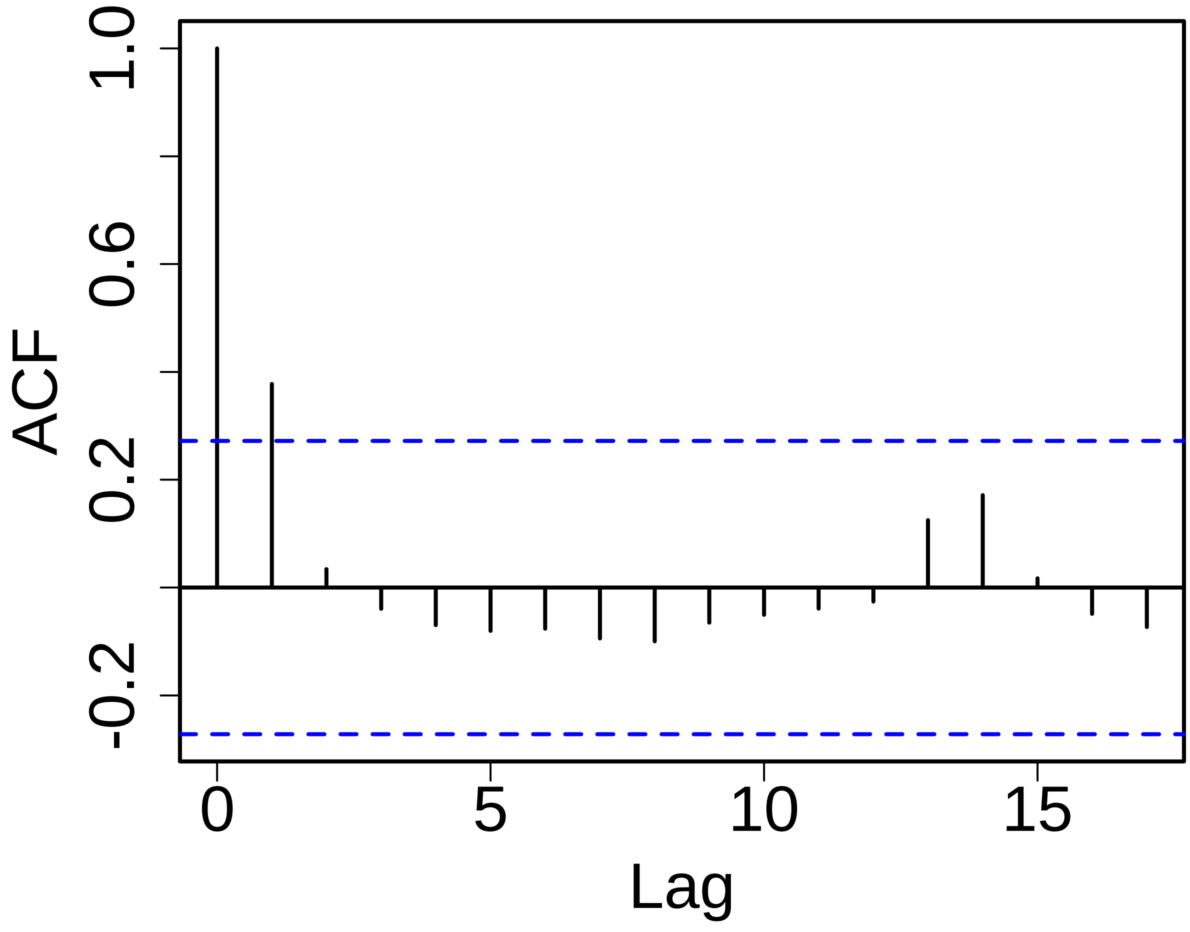


Autocorrelation Example

- Weather now 85, 86, ..., 1005



Autocorrelation: Food Sales



AR(1) Model

- What is an AR(1)?
- First, white noise term: $\nu_t \sim N(0, \sigma^2)$

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

- What values could ρ take?

Generalized Least Squares Models

- Examples:
  - Y_t = # games won by Astros in year t; X_t = Population of Houston / Population of U.S.
  - Y_t = Fidelity Real Estate Fund price on day t; X_t = S&P 500

- Errors are AR(1)

$$Y_t = \beta_0 + \beta_1 X_t + e_t \quad e_t = \rho\, e_{t-1} + \nu_t \quad \nu_t \sim N(0, \sigma_\nu^2)$$

$$E[e_t] =$$

We assume that the time series e_t is **stationary**:

- $E[e_t] = \mu$ for all t
- $E[e_t^2] < \infty$ for all t
- $Cov(e_r, e_s) = cov(e_{r+t}, e_{s+t})$

Generalized Least Squares Models

$$Y_t = \beta_0 + \beta_1 X_t + e_t \qquad e_t = \rho\, e_{t-1} + \nu_t \qquad \nu_t \sim N(0, \sigma_\nu^2)$$

$$\sigma_e^2 = Var(e_t) =$$

Generalized Least Squares Models

$$Y_t = \beta_0 + \beta_1 X_t + e_t \qquad e_t = \rho\, e_{t-1} + \nu_t \qquad \nu_t \sim N(0, \sigma^2_\nu)$$

$$\mathrm{Cov}(e_t, e_{t-1}) =$$

$$\mathrm{Corr}(e_t, e_{t-1}) =$$

Generalized Least Squares Models

$$Y_t = \beta_0 + \beta_1 X_t + e_t \qquad e_t = \rho\, e_{t-1} + \nu_t \qquad \nu_t \sim N(0, \sigma_\nu^2)$$

$$\mathrm{Cov}(e_t, e_{t-2}) =$$

Generalized Least Squares Models

- What is the covariance matrix now?

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}Y)$$

Durbin-Watson Test Statistic

- [Table here]

$$D_l = \frac{\sum_{t=l+1}^{n} (e_t - e_{t-l})^2}{\sum_{i=1}^{n} e_t^2}$$

- Always between 0 and 4.
- Smaller values of the statistic indicate positive autocorrelation of the residuals; larger values indicate negative autocorrelation.
- Not appropriate when lagged values of response variable are included as predictors.

+

Durbin-Watson Test Statistic

Testing for positive correlation:
H_o: The error terms are not correlated.
H_a: The error terms are positively correlated

The Important One!

If $D <$ lower bound, there is evidence that the error terms are positively correlated.

Testing for negative correlation:
H_o: The error terms are not correlated.
H_a: The error terms are negatively correlated

If $(4 – D) <$ lower bound, there is evidence that the error terms are negatively correlated.

Durbin-Watson Test Statistic

$$D_l = \frac{\sum_{t=l+1}^{n} (e_t - e_{t-l})^2}{\sum_{t=1}^{n} e_t^2}$$

$$= \frac{\sum_{t=l+1}^{n} (e_t^2 - 2e_t e_{t-l} + e_{t-l}^2)}{\sum_{t=1}^{n} e_t^2}$$

$$= \frac{\sum_{t=l+1}^{n} e_t^2}{\sum_{t=1}^{n} e_t^2} + \frac{\sum_{t=l+1}^{n} e_{t-l}^2}{\sum_{t=1}^{n} e_t^2} - 2\frac{\sum_{t=l+1}^{n} e_t e_{t-l}}{\sum_{t=1}^{n} e_t^2}$$

- As the sample size increases, the first two terms tend to 1. The third is an estimator of ρ, so this tends toward 1 + 1 – 2ρ.
- If ρ increases from 0 to 1, D moves from 2 to 0; as ρ moves from 0 to -1, D moves from 2 to 4.

Transforming to iid Errors: t > 1

- Goal: transform a regression model with AR(1) errors into a related model with uncorrelated errors so we can use all of our usual diagnostics.

$$Y_t = \beta_0 + \beta_1 x_t + e_t = \beta_0 + \beta_1 x_t + \rho e_{t-1} + \nu_t$$

$$Y_{t-1} = \beta_0 + \beta_1 x_{t-1} + e_{t-1}$$

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{t-1} + \rho e_{t-1}$$

- Subtract:

$$Y_t - \rho Y_{t-1} = (\beta_0 + \beta_1 x_t + e_t) - (\rho\beta_0 + \rho\beta_1 x_{t-1} + \rho e_{t-1})$$

Transforming to iid Errors: t > 1

$$Y_t - \rho Y_{t-1} = (\beta_0 + \beta_1 x_t + e_t) - (\rho\beta_0 + \rho\beta_1 x_{t-1} + \rho e_{t-1})$$

- Substitute: $e_t = \rho e_{t-1} + \nu_t$

$$\begin{aligned} Y_t - \rho Y_{t-1} &= \beta_0 + \beta_1 x_t + \rho e_{t-1} + \nu_t - (\rho\beta_0 + \rho\beta_1 x_{t-1} + \rho e_{t-1}) \\ &= (1-\rho)\beta_0 + \beta_1(x_t - \rho x_{t-1}) + \nu_t \end{aligned}$$

- What is the cool thing about this new model?

- Define new variables (Cochrane-Orcutt transformation):

$$\begin{aligned} Y_t^* &= Y_t - \rho Y_{t-1} \\ x_{t1}^* &= 1 - \rho \\ x_{t2}^* &= x_t - \rho x_{t-1}, t = 2\ldots, n \end{aligned}$$

Transforming to iid Errors: t = 1

- Then we can rewrite the model as:

$$Y_t^* = \beta_0 x_{t1}^* + \beta_1 x_{t2}^* + \nu_t, t = 2, \ldots, n$$

- And we still have to deal with the first observation.
  - Remember the first observation was: $Y_1 = \beta_0 + \beta_1 x_1 + e_1$
  - The variance of the first error was: $\sigma_\nu^2/(1-\rho^2)$
  - So if we multiply, we get the same variance in the error as the other observations in the transformed model:

$$\sqrt{1-\rho^2}Y_1 = \sqrt{1-\rho^2}\beta_0 + \sqrt{1-\rho^2}\beta_1 x_1 + \sqrt{1-\rho^2}e_1$$

$$\text{Var}\left(\sqrt{1-\rho^2}e_1\right) =$$

Transforming to iid Errors: t = 1

- So we define the Prais-Winsten transformation

$$
\begin{aligned}
Y_1^* &= \sqrt{1-\rho^2}Y_1 \\
x_{11}^* &= \sqrt{1-\rho^2} \\
x_{12}^* &= \sqrt{1-\rho^2}x_1 \\
e_1^* &= \sqrt{1-\rho^2}e_1
\end{aligned}
$$

- And write the model equation for Y_1 as:

$$Y_1^* = \beta_0 x_{11}^* + \beta_1 x_{12}^* + e_1^*$$

- Pro: the error variances match now.
- Con: (x_{12}^*, Y_1^*) is generally a point of high leverage.

Transforming to iid Errors

- We could use the formulas we just developed and multiply all of them by $\sqrt{1-\rho^2}$ and equivalently define:

$$Y_1^* = Y_1$$

$$Y_t^* = (Y_t - \rho Y_{t-1})\sqrt{1-\rho^2}, t = 2, \ldots, n$$

Transforming: Matrices

- Now consider the general matrix form:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$$

where the errors have mean 0 and covariance matrix Σ.

- Earlier we found the estimator of β is given by:

$$\hat{\beta}_{GLS} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Sigma^{-1}\mathbf{Y})$$

Let's break up Σ into pieces; since it is a symmetric positive-definite matrix it can be written as:

$$\Sigma = SS'$$

where S is a lower triangular matrix with positive diagonal entries. This is the Cholesky decomposition of Σ.

Transforming: Matrices

- Then we can rewrite our estimator as:

$$\hat{\beta}_{GLS} = (\mathbf{X}'(SS')^{-1}\mathbf{X})^{-1}(\mathbf{X}'(SS')^{-1}\mathbf{Y})$$

Transforming: Matrices

- So use the transformation:

$$\mathbf{Y}^* = S^{-1}\mathbf{Y}$$

$$\mathbf{X}^* = S^{-1}\mathbf{X}$$

$$\mathbf{e}^* = S^{-1}\mathbf{e}$$

- And then we produce a model with uncorrelated errors:

$$\mathbf{Y}^* = \mathbf{X}^*\beta + \mathbf{e}^*$$

$$\text{Var}(\mathbf{e}^*) =$$

- Con: numerically unstable.

“Feasible” GLS

- Goal: estimate the parameter vector β.
- Problem: errors are correlated.
- Solution: Figure out correlation structure and use that:

$$\hat{\beta}_{GLS} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Sigma^{-1}\mathbf{Y})$$

- …But I don’t know ρ.

Reminder: Variance Matrix

$$\Sigma = \frac{\sigma_\nu^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix}$$

- Feasible because one extra parameter ρ, not n extra parameters.

“Feasible” GLS

- Statisticians: People who run around and put hats on things:

$$\tilde{\beta}_{GLS} = (\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{Y})$$

- Translation: Estimate ρ from the data and call it a day.

Feasible GLS

- That works if we only have variance to estimate:

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{SXX}} \qquad T = \frac{\hat{\beta}_1 - \beta_1}{s / \sqrt{SXX}}$$

- But if we have to estimate ρ as well, what is the distribution of $\tilde{\beta}$?
- Large-sample: consistent and asymptotically normal in the case where errors are AR(1) (estimator of ρ consistent).