Review Topics

INSR 260, Spring 2009
Bob Stine
Regression

Models

- Simple regression model (SRM) - marginal slope
  - Scatterplot of y on x, e on x
- Multiple regression model (MRM) - partial slope
  - Scatterplot matrix of y with \( x_1, \ldots, x_k \)
  - Collinearity, VIF
- Fitted values, residuals, errors

Assumptions

- Linear equation
- Independence
- Constant variance
- Normality
- Transformations (esp. logs)
- Durbin–Watson test, residuals
- Plot of e on \( \hat{y} \)
- Normal quantile plot

Categorical variables

- Dummy variables
- Use in seasonal models
- Interactions
Inference in Regression

**Coefficients**
- One slope \( H_0: \beta_j = 0 \) \( t \)-statistic, p-value, conf int
- All \( H_0: \beta_1=...=\beta_k=0 \) F-statistic (Anova table)
  - Test size of \( R^2 \)
- Some \( H_0: \text{some } \beta_j=0 \) Partial F
  - Test collection representing categorical variable
  - Test using Effect Test or by comparison of \( R^2 \) in models
  - Importance of avoiding multiple t-tests, multiplicity

**Prediction**
- Components of standard error
  - Random, unexplained variation (RMSE, \( \sigma_\varepsilon \))
  - Extrapolation ("distance value")
- Intervals
  - Confidence interval for mean \( \hat{\gamma} = E(Y|X_1,\ldots,X_k) \)
  - Prediction interval for individual future \( Y \) value
Exponential Smoothing

- Simple exponential smoothing
  - Geometrically weighted average of past values
  - Recursive form, updating equation: \( l_t = l_{t-1} + \alpha(y_t - l_{t-1}) \)

- Model with underlying state
  - Evolving underlying state: \( L_t = L_{t-1} + \alpha \varepsilon_t \)
  - Observation has mean \( L_{t-1} \): \( y_t = L_{t-1} + \varepsilon_t \)

- Prediction
  - Predict \( y_{n+f} \) using estimated state: \( \mathbb{E} y_{n+f} = L_n \) (same for all \( f \))
  - Error grows as extrapolate: \( \mathbb{E}(y_{n+f} - \hat{y}_{n+f})^2 = \sigma^2(1+(f-1)\alpha^2) \)

- Equivalent to IMA(1,1)
  - Exponential smoothing is special case of a non-stationary
    ARIMA model
Exponential Smoothing

IMA(1,1)
- Equation: \[ y_t - y_{t-1} = -\theta a_{t-1} + a_t \]
- Predictions: \[ \hat{y}_{n+1} = y_n - \theta a_n, \quad \hat{y}_{n+f} = \hat{y}_{n+1} \]

Same predictors
- Equation: \[ l_t = l_{t-1} + \alpha(y_t - l_{t-1}) \]
- Predictor of \( y_t \) in exponential smooth is \( l_{t-1} \)
- Relabel \( l_{t-1} = \hat{y}_t \)
- Equation becomes: \[ \hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t) \]
  
  at \( t=n \) \[ \hat{y}_{n+1} = \hat{y}_n + \alpha(y_n - \hat{y}_n) \]
  
  use observations \[ \hat{y}_{n+1} = y_n + \alpha a_n \]
- Prediction equations agree with \( \alpha = -\theta \)

Implication
- If an IMA(1,1) looks like good fit, use expo smoothing
- If not, some other type of smoothing is appropriate
ARIMA Models

Stationarity
- Use differencing to produce stationary series

Correlation functions
- Autocorrelation function
- Partial autocorrelation function

Uses
- detect non-stationarity
- identify model
- evaluate/check residuals

Different types of dependence
- Autoregression
  - Geometric weighting of past errors, finite past observations
  - TAC decays geometrically, TPAC cuts off
- Moving Average
  - Geometric weighting of past observations, finite past errors
  - TAC cuts off, TPAC decays geometrically
ARIMA Models

Model identification
- Plots of data
- Correlation functions
  - Do residuals appear uncorrelated?
    - Box-Pierce, Box-Ljung statistics (avoid multiplicity)
    - Accumulate squared residual autocorrelations
- Selection criteria (AIC, SBC/BIC)
  - Penalized complicated models, reward for parsimonious specification

Prediction, prediction intervals
- Extrapolate form of model, recursively using predictions
  - Fill in $y_{n+f}$ with $\hat{y}_{n+f}$, $a_{n+f}$ with 0
- Predictions revert to mean
- Prediction standard errors grow toward $SD(y_t)$
General Comments

- Read questions carefully before answering

- Open textbook (no other books)
  - Buy a calculator if you want to use one.
  - No telephones, laptops, other electronics allowed.
    - Shut off ahead of time to avoid issues.

- Exam structure
  - Brief description of context of data
  - JMP output
  - Short answer
  - Multiple choice

- Get plenty of sleep the night before!