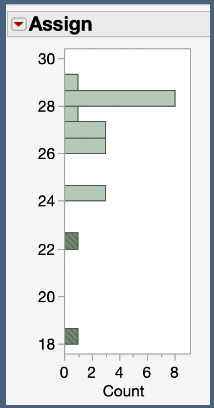
Statistics 5350/7110
Forecasting

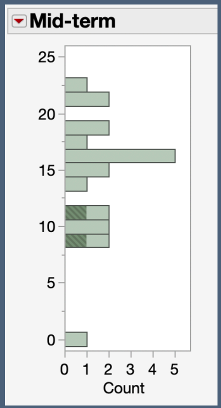
Lecture 16
Forecasting ARMA Models

Professor Robert Stine

Preliminaries

- Questions?





- Hackathon
  - Wharton AI & Analytics Datathon team registration!
  - Predict "clearing prices" for Wharton's CourseMatch algorithm.
  - Event Information (sign up form here: https://forms.gle/9NNo3EaFt8pU75hY6)
- Quick review
  - Identifying models (while ago)
  - Estimating ARMA models
  - Diagnostics for ARMA models as in regression

Today's Topics

Text, §4.3

- Forecasting ARMA processes

- Roles of different representations of the model
  - Difference equation reveals forecast
  - MA representation reveals uncertainty of the forecast
  - AR representation describes how data impact current value (least used)

- Assumptions
  - We pretend that we know more than we actually do
  - Typically these assumptions are benign (e.g. not a really short time series)

- Example

Optimal Prediction

- Conditional mean
  - Information in a set of k random variables $Z_1, \ldots, Z\_k$
  - Predictor that minimizes the expected squared error given this information
$$\widehat{X} = \min_{f(Z)} E\left((X - f(Z))^2 \mid Z_1, Z_2, \ldots, Z_k\right)$$
  - The best predictor is the conditional mean
$$\widehat{X} = E\left(X \mid Z_1, Z_2, \ldots, Z_k\right)$$
  - The mean squared prediction error (MSPE) is thus the conditional variance
$$E\left((X - \widehat{X})^2 \mid Z_1, Z_2, \ldots, Z_k\right)$$
  - Note the resemblance to least squares regression
- Specialized to time series analysis
  - Observe series up to time n, defining the information set $(Z = X_1, \ldots, X_n)$
  - Value to predict is m periods into the "future", say X_{n+m}
  - Examples clarify these concepts…

This information set defines a "sigma field" in more advanced courses.

Forecasting ARMA Models

- Three representations of an ARMA model
  - Difference equation

$$\phi(B)\,X_t = \theta(B)\,w_t$$

  - As a moving average

$$X_t = \left(\psi(B) = \frac{\theta(B)}{\pi(B)}\right) w_t \quad \Rightarrow \quad X_t = w_t + \sum_{j=1}^{\infty} \psi_j\, w_{t-j}$$

  - As an autoregression

$$\left(\pi(B) = \frac{\phi(B)}{\theta(B)}\right) X_t = w_t \quad \Rightarrow \quad X_t = w_t + \sum_{j=1}^{\infty} \pi_j\, X_{t-j}$$

- Relevance for forecasting
  - ARMA specification reveals the optimal predictor
  - Weights from MA form reveal the variance of the prediction error (MSPE)
  - Weights from AR form suggest how to construct the predictor from data

AR(1) Example

- Start with most familiar example

$$X_t = \phi\, X_{t-1} + w_t$$

- Best predictor and squared error, one-step ahead

$X_{1:n} = \{X_1, X_2, \ldots, X_n\}$

  - Predictor

$$\hat{X}_{n+1|n} = E(X_{n+1} \,|\, X_{1:n}) = E(\phi X_n + w_{n+1} \,|\, X_{1:n}) = \phi X_n$$

book notation is x^n_{n+1}

  - Expected squared error

$$P_{n+1|n} = E((X_{n+1} - \hat{X}_{n+1|n})^2 \,|\, X_{1:n}) = E(w^2_{n+1} \,|\, X_{1:n}) = \sigma^2_w$$

book notation P^n_{n+1}

- Best predictor and squared error, two steps ahead
  - Predictor

$$\hat{X}_{n+2|n} = E(X_{n+2} \,|\, X_{1:n}) = E(\phi^2 X_n + w_{n+2} + \phi w_{n+1} \,|\, X_{1:n}) = \phi^2\, X_n$$

  - Expected squared error

$$P_{n+2|n} = E((X_{n+2} - \hat{X}_{n+2|n})^2 \,|\, X_{1:n}) = E((w_{n+2} + \phi w_{n+1})^2) \,|\, X_{1:n}) = \sigma^2_w(1 + \phi^2)$$

AR(1) Example, Second Approach

- Process $$X_t = \phi X_{t-1} + w_t$$
- Convenient notation
  - Use square brackets to denote conditional expectation given $X_1, X_2, \ldots, X_n$
  $$E(X_{n+k} \mid X_{1:n}) = [X_{n+k}], \quad 0 < k$$
  - Rules presume we know p and q, parameters ($\hat{\phi} = \phi$), and observable errors
  $$[X_t] = X_t, \quad 1 \le t \le n \qquad\qquad [X_{n+m}] = \hat{X}_{n+m|n}, \quad 0 < m$$
  $$[w_t] = w_t, \quad 1 \le t \le n \qquad\qquad [w_{n+m}] = 0, \quad 0 < m$$
- Best predictor
  - One step ahead
  $$E(X_{n+1} \mid X_{1:n}) = [\phi X_n + w_{n+1}] = \phi[X_n] + [w_{n+1}] = \phi X_n$$
  - Two steps ahead
  $$E(X_{n+2} \mid X_{1:n}) = [\phi X_{n+1} + w_{n+2}] = \phi[X_{n+1}] + [w_{n+2}] = \phi^2 X_n$$
  - m steps ahead
  $$E(X_{n+m} \mid X_{1:n}) = [\phi X_{n+m-1} + w_{n+m}] = \phi[X_{n+m-1}] = \phi^m X_n$$

AR(1) Example, Prediction Errors

- Model
  - Use difference equation to find predictor

$$X_t = \phi\, X_{t-1} + w_t$$

  - Use moving average form to find expected squared prediction error

$$X_t = w_t + \psi_1\, w_{t-1} + \psi_2\, w_{t-2} + \cdots = w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \cdots$$

- Squared error of predictor
  - One step ahead

$$[X_{n+1}] = [w_{n+1} + \psi_1 w_n + \psi_2 w_{n-1} + \cdots] = [\psi_1 w_n + \psi_2 w_{n-1} + \cdots]$$
$$\Rightarrow P_{n+1|n} = E(w_{n+1}^2) = \sigma_w^2$$

  - Two steps ahead

$$[X_{n+2}] = [w_{n+2} + \psi_1 w_{n+1} + \psi_2 w_n + \cdots] = [\psi_2 w_n + \psi_3 w_{n-1} + \cdots]$$
$$\Rightarrow P_{n+2|n} = E(w_{n+2} + \psi_1 w_{n+1})^2 = \sigma_w^2(1 + \phi^2)$$

  - m steps ahead

$$P_{n+m|n} = E(w_{n+m} + \psi_1 w_{n+m-1} + \cdots + \psi_{m-1} w_{n+1})^2 = \sigma_w^2(1 + \phi^2 + \phi^4 + \cdots + \phi^{2(m-1)})$$

Harder Example: ARMA(2,1)

- Process
  - Difference equation $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + w_t + \theta_1 w_{t-1}$
- One step ahead
  - Predictor

$$
\begin{aligned}
[X_{n+1}] &= [\phi_1 X_n + \phi_2 X_{n-1} + w_{n+1} + \theta_1 w_n] \\
&= \phi_1 [X_n] + \phi_2 [X_{n-1}] + [w_{n+1}] + \theta_1 [w_n] \\
&= \boxed{\phi_1 X_n + \phi_2 X_{n-1} + \theta_1 w_n}
\end{aligned}
$$

  - Mean squared prediction error (MSPE)

$$P_{n+1|n} = \sigma_w^2$$

Always like this for MSPE at lead 1

- Two steps ahead
  - Predictor $[X_{n+2}] = \phi_1 [X_{n+1}] + \phi_2 [X_n] + [w_{n+2}] + \theta_1 [w_{n+1}] = \phi_1 [X_{n+1}] + \phi_2 X_n$
  - MSPE

$$P_{n+2|n} = E(\phi_1 w_{n+1} + w_{n+2} + \theta_1 w_{n+1})^2 = \sigma_s^2 (1 + \underbrace{(\phi_1 + \theta_1)}_{\psi_1}{}^2)$$

Aside: Difference Equations

- How do you find the MA weights, the ψ_j ?
- Easy: Use R
  - That's what the R function ARMAtoMA does.
- How's it work?
  - It's all about polynomials
  - Plug in MA representation for X_t

$$\phi(B)X_t = \theta(B)w_t \quad \Rightarrow \quad \phi(B)\,\psi(B)w_t = \theta(B)w_t$$

  - Hence must have equivalent polynomials

$$\phi(B)\,\psi(B) = \theta(B)$$

  - Insert Φ(B) and θ(B) for ARMA(2,1)

$$(1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1 + \theta_1 B$$

  - Equate coefficients of powers of B

$$1 + \underbrace{(\psi_1 - \phi_1)}_{=\theta_1}B + \underbrace{(\psi_2 - \phi_1\psi_1 - \phi_2)}_{=0}B^2 + \cdots = 1 + \theta_1 B$$

Forecast Errors and MA Representation

- Moving average representation
  - Key to understanding forecast errors
  - Partition future value into unpredictable future and "known" past (uncorrelated with future)

$$X_{n+m} = \underbrace{w_{n+m} + \psi_1 w_{n+m-1} + \psi_2 w_{n+m-2} + \cdots + \psi_{m-1} w_{n+1}}_{\text{future}} \mid \underbrace{\psi_m w_n + \psi_{m+1} w_{n-1} + \cdots}_{\text{past}}$$

- Implications
  - Forecast error at lead m composed of m uncorrelated random variables

$$X_{n+m} - \hat{X}_{n+m|n} = x_{n+m} + \psi_1 w_{n+m-1} + \psi_2 w_{n+m-2} + \cdots + \psi_{m-1} w_{n+1}$$

  - MSPE grows with m, increasing monotonically toward process variance
  - Forecast errors are correlated over different lead times

Prediction error 1 step ahead $X_{n+1} - \hat{X}_{n+1|n} = w_{n+1}$

Prediction error 2 steps ahead $X_{n+2} - \hat{X}_{n+2|n} = w_{n+2} + \psi_1 w_{n+1}$

Prediction error 2 steps ahead $X_{n+3} - \hat{X}_{n+3|n} = w_{n+3} + \psi_1 w_{n+2} + \psi_2 w_{n+1}$

Discussion of Procedure

- Back-substitution
  - For any AR(p) we can recursively decompose as sum of future errors plus weights on last p values

$$X_{n+m} = \underbrace{w_{n+m} + \psi_1 w_{n+m-1} + \cdots + \psi_{n+1} w_{n+1}}_{\text{MA}} + \underbrace{\xi_{1,m} X_n + \xi_{2,m} X_{n+1} + \cdots + \xi_{p,m} X_{n-p-1}}_{\text{AR}}$$

  - MA weights describe the unpredictable noise
  - AR weights show how data impact prediction (weights constructed like π_j in the AR representation)
  - If there's an MA component, the AR piece is an infinite sum!
- Approximations
  - Pretend we know both p and q as well as the coefficients $\hat{\phi}_j = \phi_j,\ \hat{\theta}_j = \theta_j,\ \hat{\mu} = \mu,\ s_w^2 = \sigma_w^2$
  - Presume we know $w_1, \ldots, w_n$
    Intuitive if the process is AR(p), but requires infinite sum if MA component
    For example, compute w_n in the ARMA(2,1) example.
  - Ignore errors when truncate infinite sums, as if we know the infinite past
    Not such a big effect since we presume stationarity and invertibility

Implications of Procedure

- What happens as we predict farther into the future
  - Predictions will revert to the mean
  - Mean squared prediction error (MSPE) will monotonically grow to the variance of the process
- Why must forecasts mean revert?
  - AR(∞) weights show diminishing role of data in forecast
  - These weights decay exponentially fast to zero for a stationary process
  - Eventually prediction puts essentially "no weight" on values far in the past
- Why does the MSPE grow monotonically to process variance?
  - As we extrapolate farther out,

    Errors accumulate. We know less and less.

    Observed data becomes less relevant. AR weights approach 0.
  - Eventually, as the lead time m increases,

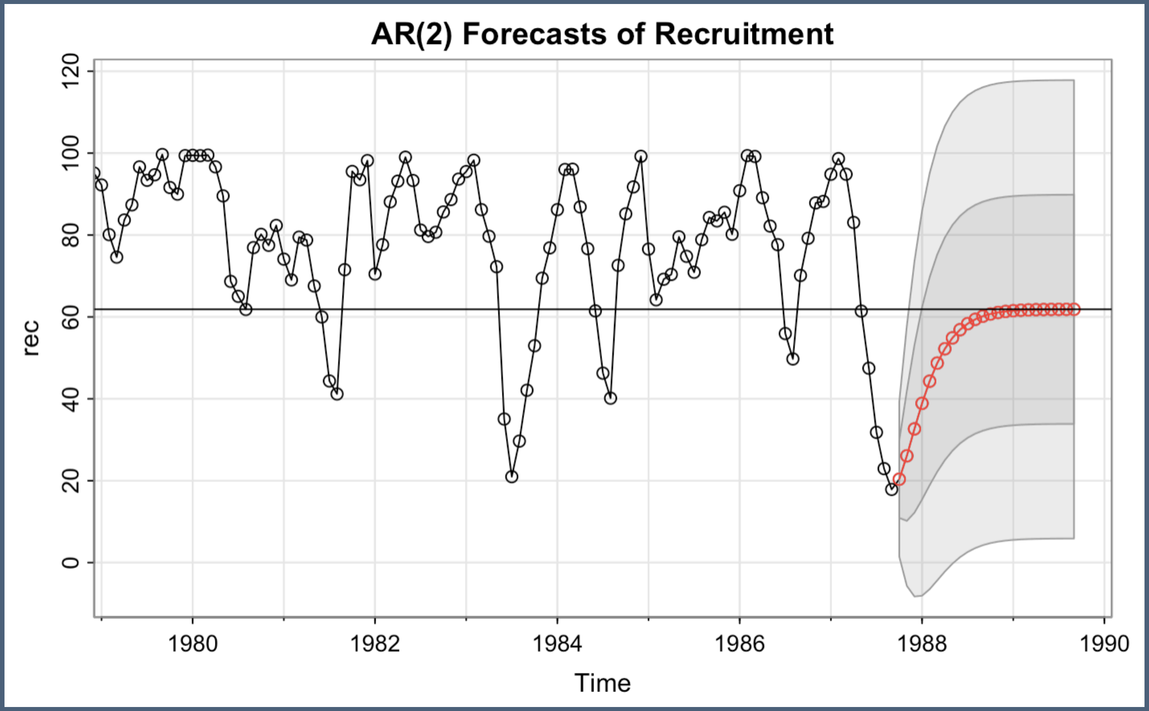
$$E\left(X_{n+m} - \hat{X}_{n+m|n}\right)^2 = \sigma_w^2(1 + \psi_1^2 + \cdots + \psi_{n+m-1}^2) \to \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \text{Var}(X_t)$$

Example

Example 4.31

- Estimated model for fish recruitment
  - AR(2) model
  - Predictions damp to estimated mean
  - Standard errors grow

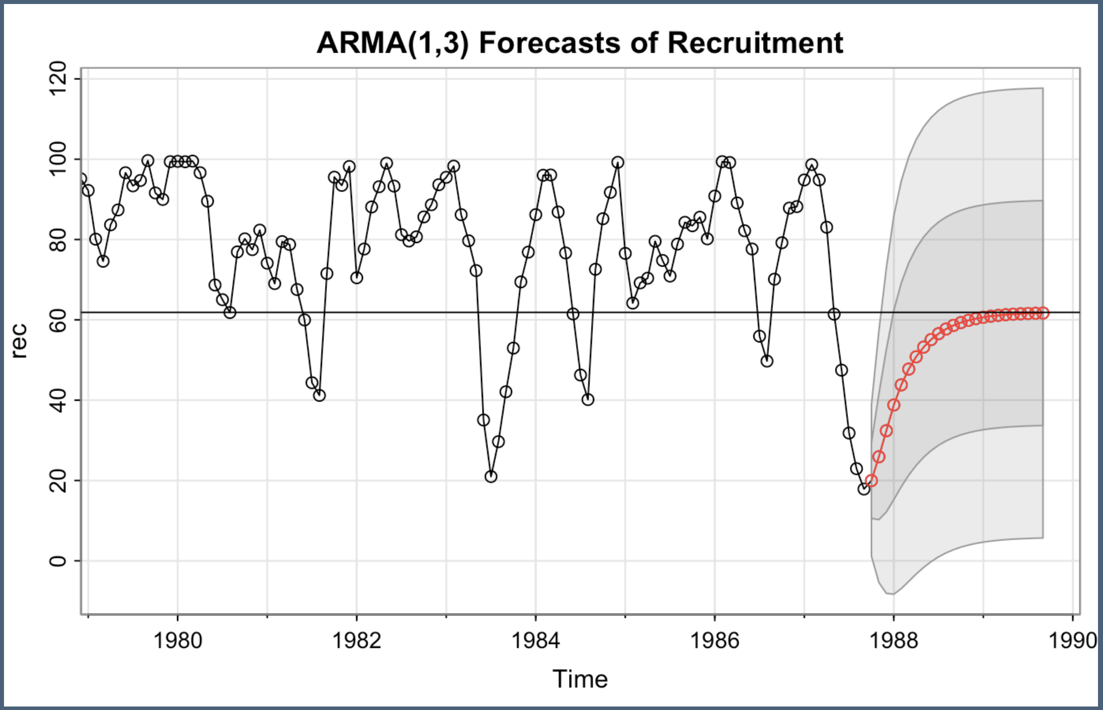
- Comments
  - Pointwise intervals
  - Process does not recognize limit at 100
  - Other models give very similar predictions



Example

- Estimated model for fish recruitment
  - ARMA(1,3) model is similar
  - Predictions damp to estimated mean
  - Standard errors grow

Example 4.31



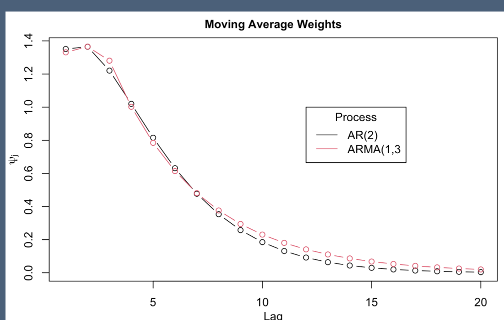
- Comments
  - How different are these?

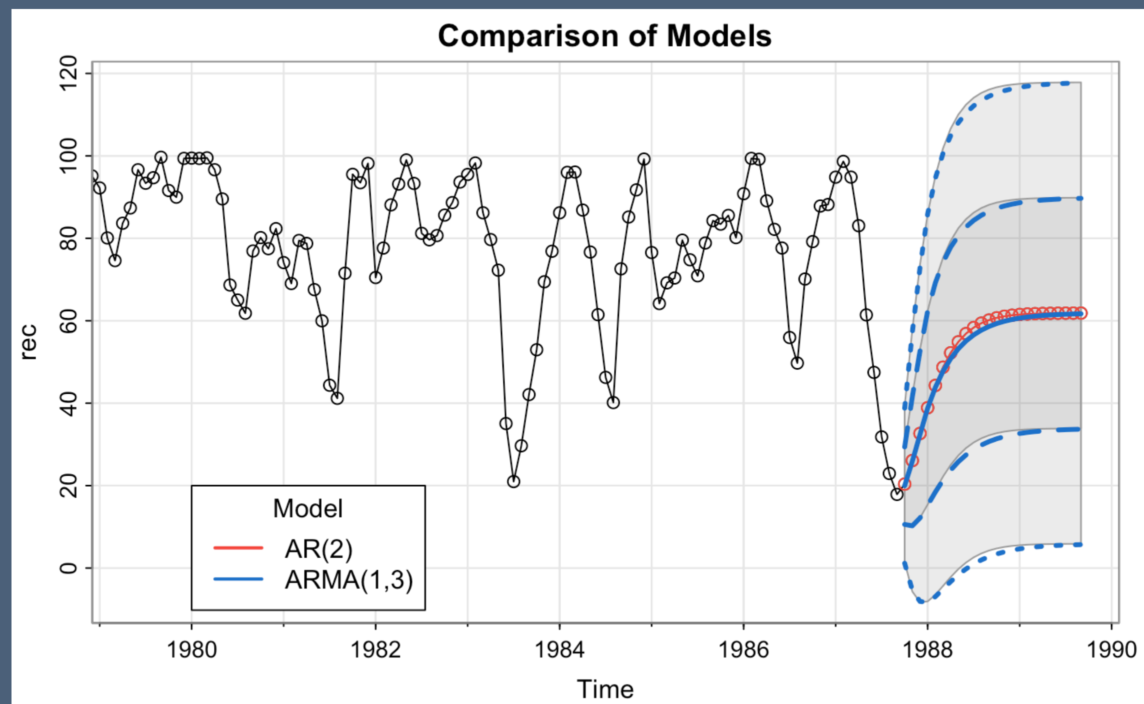
Example

Example 4.31

- Estimated model for fish recruitment
  - ARMA(1,3) model is similar
  - Predictions damp to estimated mean
  - Standard errors grow

- Comments
  - How different are these?
  - MA weights show similarity





Forecasting in Practice

- Prior analysis makes several large assumptions
  - Predictor minimizes squared error rather than some other loss function (implicitly Gaussian)
  - We know that the process is ARMA(p,q) — stationary and invertible
  - We not only know both p and q, we also know the coefficients of that process
- Consider the last of these...
  - AR(1) model, prediction of X_{n+2} is
$$\hat{X}_{n+2|n} = \phi \hat{X}_{n+1} = \phi^2 X_n$$
  - Estimated conditional LS coefficient comes from regression of X_{t+1} on X_t.
  - Common practice is to square the estimated coefficient
$$\hat{X}_{n+2|n} \approx \hat{\phi}^2 X_n$$
  - That's not the same as the regression of X_{t+2} on X_t. Which is better?
  - Fortunately, only an issue if we have short time series.
- Issue becomes more relevant for p > 1

What's next?

Textbook Chapter 5

- ARIMA models
- Recognizing “hidden” ARIMA models: exponential smoothing
- Forecasting non-stationary ARIMA models