Forecasting ARMA Models

INSR 260, Spring 2009
Bob Stine

Overview

- Review
  - Model selection criteria
  - Residual diagnostics
- Prediction
  - Normality
  - Stationary vs non-stationary models
  - Calculations
- Case study

Review

- Autoregressive, moving average models
  - AR(p)
    - $Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + a_t$
  - MA(q)
    - $Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$ Watch negative signs
  - ARMA(p,q)
    - $Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + a_t - \theta_1 a_{t-1}$ ARMA(2,1)
- Common feature
  - Every stationary ARMA model specifies Yt as a weighted sum of past error terms

    $$Y_t = a_t + w_1 a_{t-1} + w_2 a_{t-2} + w_3 a_{t-3} + \ldots$$

    - e.g., AR(1) sets $w_j = \varphi^j$
  - ARMA models for non-stationary data
    - Differencing produces a stationary series.
    - These differences are a weighted average of prior errors.

Modeling Process

- Initial steps
  - Before you work with data: think about context
    - What do you expect to find in a model?
    - What do you need to get from a model? ARIMA = short-term forecasts
    - Set a baseline: What results have been obtained by other models?
  - Plot time series
  - Inspect SAC, SPAC
- Estimation
  - Fit initial model, explore simpler & more complex models
  - Check residuals for problems
    - Ljung-Box test of residual autocorrelations
    - Residual plots show outliers, other anomalies
- Forecasting
  - Check for normality
  - Extrapolate pattern implied by dependence
  - Compare to baseline estimates

Forecasting ARMA

- Characteristics
  - Forecasts from stationary models revert to mean
    - Integrated models revert to trend (usually a line)
  - Accuracy deteriorates as extrapolate farther
    - Variance of prediction error grows
    - Prediction intervals at fixed coverage (e.g. 95%) get wider
- Calculations
  - Fill in unknown values with predictions
  - Pretend estimated model is the true model
- Example: ARMA (2,1) $Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + a_t - \theta_1 a_{t-1}$
  - One-step ahead: $\hat{Y}_{n+1} = \delta + \varphi_1 Y_n + \varphi_2 Y_{n-1} + (\hat{a}_{n+1}=0) - \theta_1 \hat{a}_n$
  - Two $\hat{Y}_{n+2} = \delta + \varphi_1 \hat{Y}_{n+1} + \varphi_2 Y_n + (\hat{a}_{n+2}=0) - \theta_1 (\hat{a}_{n+1}=0)$
  - Three $\hat{Y}_{n+3} = \delta + \varphi_1 \hat{Y}_{n+2} + \varphi_2 \hat{Y}_{n+1} + 0 + 0$
  - AR gradually damp out, MA terms disappear (as in autocorrelations)

Accuracy of Forecasts

- Assume
  - Estimated model is true model
- Key fact
  - ARMA models represent Y_t as weighted sum of past errors
- Theory: Forecasts omit unknown error terms
  - $Y_{n+1} = \mu + a_{n+1} + w_1\, a_n + w_2\, a_{n-1} + w_3\, a_{n-2} + \ldots$

    ("known": $w_1\, a_n + w_2\, a_{n-1} + w_3\, a_{n-2} + \ldots$)

    $\hat{Y}_{n+1} = \mu + w_1\, a_n + w_2\, a_{n-1} + \ldots$

    $\Rightarrow Y_{n+1} - \hat{Y}_{n+1} = a_{n+1}$
  - $Y_{n+2} = \mu + a_{n+2} + w_1\, a_{n+1} + w_2\, a_n + w_3\, a_{n-1} + \ldots$

    $\hat{Y}_{n+2} = \mu + w_2\, a_n + w_3\, a_{n-1} + \ldots$

    $\Rightarrow Y_{n+2} - \hat{Y}_{n+2} = a_{n+2} + w_1\, a_{n+1}$
  - Variance of forecast error grows as (a_t are iid)

    $$\sigma^2(1 + w_1^2 + w_2^2 + \ldots)$$

Example: ARMA(1,1)

- Simulated data
  - Know that we're fitting the right model
  - Compare forecasts to actual future values

δ is not the mean of the series

- Estimated model

Parameter Estimates

Term	Estimate	Constant Estimate
AR1	0.7467	0.07988923
MA1	-0.7954	
Intercept	0.3154	

intercept = mean

δ

Reverse sign on moving average estimates

- Forecasts

	Std Err Pred ARMA(1,1)	Actual ARMA(1,1)	Residual ARMA(1,1)	Lower CL (0.95) ARMA(1,1)	Predicted ARMA(1,1)	Upper CL (0.95) ARMA(1,1)
196	0.99	4.71	0.98	1.78	3.73	5.67
197	0.99	4.81	0.44	2.43	4.38	6.32
198	0.99	4.31	0.28	2.08	4.02	5.97
199	0.99	2.74	-0.77	1.57	3.52	5.47
200	0.99	1.22	-0.29	-0.43	1.51	3.46
201	0.99	•	•	-1.18	0.76	2.71
202	1.82	•	•	-2.93	0.65	4.23
203	2.15	•	•	-3.65	0.57	4.79
204	2.32	•	•	-4.04	0.50	5.04
205	2.40	•	•	-4.25	0.45	5.16
206	2.45	•	•	-4.38	0.42	5.22
207	2.47	•	•	-4.46	0.39	5.24

SD(y) = 2.50

0.315

n=200

$\hat{y}_{201}=0.080+0.75(1.22)+0.80(-0.29)$

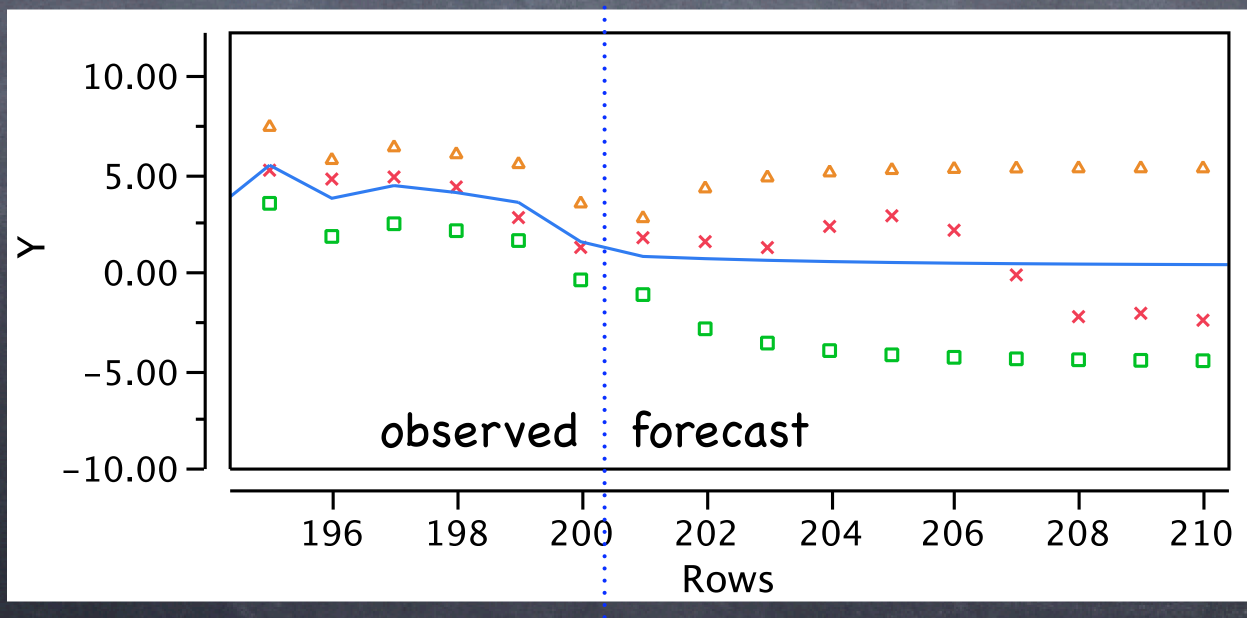
$\hat{y}_{202}=0.080+0.75(0.76)+0.80(0)$

$\hat{y}_{203}=0.080+0.75(0.65)$

$\hat{y}_{n+f}= \delta + \varphi y_{n+f-1} - \theta a_{n+f-1}$

Forecasts

- Forecasts revert quickly to series mean
  - Unless model is non-stationary or has very strong autocorrelations
- Prediction intervals open as extrapolate
  - Variance of prediction errors rapidly approaches series variance



Detailed Calculations

- MA(2) $Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
- Forecasting
  - 1-step
    $$Y_{n+1} = \mu + a_{n+1} - \theta_1 a_n - \theta_2 a_{n-1}$$
    $$\hat{Y}_{n+1} = \mu - \theta_1 a_n - \theta_2 a_{n-1}$$
    - $Y_{n+1} - \hat{Y}_{n+1} = a_{n+1}$
    - $Var(Y_{n+1} - \hat{Y}_{n+1}) = \sigma^2$
  - 2-steps
    $$Y_{n+2} = \mu + a_{n+2} - \theta_1 a_{n+1} - \theta_2 a_n$$
    $$\hat{Y}_{n+2} = \mu - \theta_2 a_n$$
    - $Y_{n+2} - \hat{Y}_{n+2} = a_{n+2} - \theta_1 a_{n+1}$
    - $Var(Y_{n+2} - \hat{Y}_{n+2}) = \sigma^2(1+\theta_1^2)$
  - 3-steps
    $$Y_{n+3} = \mu + a_{n+3} - \theta_1 a_{n+2} - \theta_2 a_{n+1}$$
    $$\hat{Y}_{n+3} = \mu$$
    - $Y_{n+3} - \hat{Y}_{n+3} = a_{n+3} + \theta_1 a_{n+2} + \theta_2 a_{n+1}$
    - $Var(Y_{n+3} - \hat{Y}_{n+3}) = \sigma^2(1+\theta_1^2+\theta_2^2)$

Example: MA(2) w/Numbers

Model Summary

DF	177.0000	Stable	Yes
Sum of Squared Errors	188.3783	Invertible	Yes
Variance Estimate	1.0643		
Standard Deviation	1.0316		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
MA1	1	-0.9962	0.0647622	-15.38	<.0001*	0.16199858
MA2	2	-0.3803	0.0590225	-6.44	<.0001*	
Intercept	0	0.1620	0.1802599	0.90	0.3700	

	Actual MA(2)	Residual MA(2)	Predicted MA(2)	Std Err Pred MA(2)	Lower CL (0.95)	Upper CL (0.95)
197	3.0455	1.0554	1.9902	1.0316	-0.0318	4.0122
198	2.4779	0.5533	1.9246	1.0316	-0.0974	3.9465
199	1.2582	0.1436	1.1146	1.0316	-0.9074	3.1366
200	0.4691	-0.0464	0.5155	1.0316	-1.5065	2.5375
201	•	•	0.1704	1.0316	-1.8516	2.1924
202	•	•	0.1443	1.4562	-2.7098	2.9984
203	•	•	0.1620	1.5081	-2.7939	3.1179
204	•	•	0.1620	1.5081	-2.7939	3.1179
205	•	•	0.1620	1.5081	-2.7939	3.1179
206	•	•	0.1620	1.5081	-2.7939	3.1179

- Predictions
  - $\hat{Y}_{n+1} = \mu - \theta_1 a_n - \theta_2 a_{n-1}$
    $= 0.162 + 0.996(-.0464) + 0.380(.144$
    $= 0.1705$
  - $\hat{Y}_{n+2} = \mu - \theta_1 \hat{a}_{n+1} - \theta_2 a_n$
    $= 0.162 + 0.996(0) + 0.380(-.0464)$
    $= 0.1443$
  - $\hat{Y}_{n+3} = \mu - \theta_1 \hat{a}_{n+2} - \theta_2 \hat{a}_{n+1}$
    $= 0.046 + 0.996(0) + 0.380(0)$
    $= 0.162$
- SD of prediction error
  - 1-step:
    $\sigma = 1.0316$
  - 2-steps
    $\sigma \text{ sqrt}(1 + \theta_1^2) = 1.0316(1+.996^2)^{1/2}$
    $= 1.456$
  - 3-steps
  - $\sigma \text{ sqrt}(1+\theta_1^2+\theta_2^2) = 1.0316(1+.996^2+.380^2)^{1/2}$
    $= 1.508$

Detailed Calculations

- AR(2) $Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + a_t$
- Forecasting
  - 1-step $Y_{n+1} = \delta + \varphi_1 Y_n + \varphi_2 Y_{n-1} + a_{n+1}$
    $\hat{Y}_{n+1} = \delta + \varphi_1 Y_n + \varphi_2 Y_{n-1}$
    - $Y_{n+1} - \hat{Y}_{n+1} = a_{n+1}$
    - $Var(Y_{n+1} - \hat{Y}_{n+1}) = \sigma^2$
  - 2-steps $Y_{n+2} = \delta + \varphi_1 Y_{n+1} + \varphi_2 Y_n + a_{n+2}$
    $\hat{Y}_{n+2} = \delta + \varphi_1 \hat{Y}_{n+1} + \varphi_2 Y_n$
    - $Y_{n+2} - \hat{Y}_{n+2} = a_{n+2} + \varphi_1(Y_{n+1} - \hat{Y}_{n+1}) = a_{n+2} + \varphi_1 a_{n+1}$
    - $Var(Y_{n+2} - \hat{Y}_{n+2}) = \sigma^2(1+\varphi_1^2)$
  - 3-steps $Y_{n+3} = \delta + \varphi_1 Y_{n+2} + \varphi_2 Y_{n+1} + a_{n+3}$
    $\hat{Y}_{n+3} = \delta + \varphi_1 \hat{Y}_{n+2} + \varphi_2 \hat{Y}_{n+1}$
    - $Y_{n+3} - \hat{Y}_{n+3} = a_{n+3} + \varphi_1(Y_{n+2} - \hat{Y}_{n+2}) + \varphi_2(Y_{n+1} - \hat{Y}_{n+1})$
      $= a_{n+3} + \varphi_1(a_{n+2} + \varphi_1 a_{n+1}) + \varphi_2 a_{n+1}$
      $= a_{n+3} + \varphi_1 a_{n+2} + (\varphi_1^2 + \varphi_2) a_{n+1}$
    - $Var(Y_{n+3} - \hat{Y}_{n+3}) = \sigma^2(1+\varphi_1^2+(\varphi_1^2 + \varphi_2)^2)$

Example: AR(2) w/Numbers

Mean	0.1555
Std	1.6440

Model Summary

DF	177.0000	Stable	Yes
Sum of Squared Errors	175.7592	Invertible	Yes
Variance Estimate	0.9930		
Standard Deviation	0.9965		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	0.9745	0.0722429	13.49	<.0001*	0.04617102
AR2	2	-0.2449	0.0722063	-3.39	0.0009*	
Intercept	0	0.1707	0.2699548	0.63	0.5280	

	Actual AR(1)	Residual AR(1)	Predicted AR(1)	Std Err Pred AR(1)	Lower CL (0.95) AR(1)	Upper CL (0.95) AR(1)
196	2.947	0.946	2.001	0.996	0.048	3.954
197	3.046	0.847	2.198	0.996	0.245	4.151
198	2.478	0.186	2.292	0.996	0.339	4.245
199	1.258	-0.457	1.715	0.996	-0.238	3.668
200	0.469	-0.196	0.665	0.996	-1.288	2.618
201	•	•	0.195	0.996	-1.758	2.148
202	•	•	0.121	1.391	-2.606	2.848
203	•	•	0.117	1.559	-2.938	3.171
204	•	•	0.130	1.621	-3.047	3.308
205	•	•	0.144	1.642	-3.075	3.363

0.1707 1.644

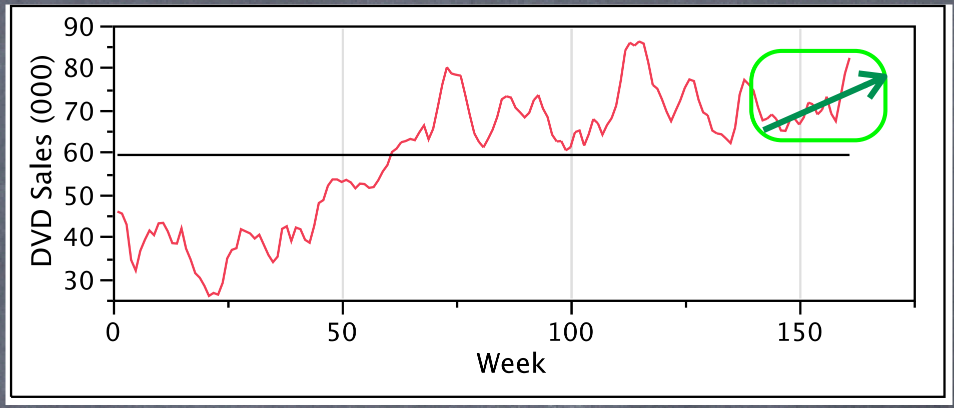
- Predictions
  - $\hat{Y}_{n+1} = \delta + \varphi_1 Y_n + \varphi_2 Y_{n-1}$
    $= 0.046 + 0.975\ (.469) - 0.245(1.258)$
    $= 0.195$
  - $\hat{Y}_{n+2} = \delta + \varphi_1 \hat{Y}_{n+1} + \varphi_2 Y_n$
    $= 0.046 + 0.975\ (.195) - 0.245(.469)$
    $= 0.121$
  - $\hat{Y}_{n+3} = \delta + \varphi_1 \hat{Y}_{n+2} + \varphi_2 \hat{Y}_{n+1}$
    $= 0.046 + 0.975\ (.121) - 0.245(.195)$
    $= 0.117$
- SD of prediction error
  - 1-step:
    $\sigma = 0.9965$
  - 2-steps
    $\sigma\ \mathrm{sqrt}(1 + \varphi_1^2) = 0.9965(1+.9745^2)^{1/2}$
    $= 1.391$
  - 3-steps
  - $\sigma\ \mathrm{sqrt}(1+\varphi_1^2+(\varphi_1^2+\varphi_2)^2) = 0.9965(1+.9745^2+ (.9745^2-.2449)^2)^{1/2}$
    $= 1.559$

Integrated Forecasts

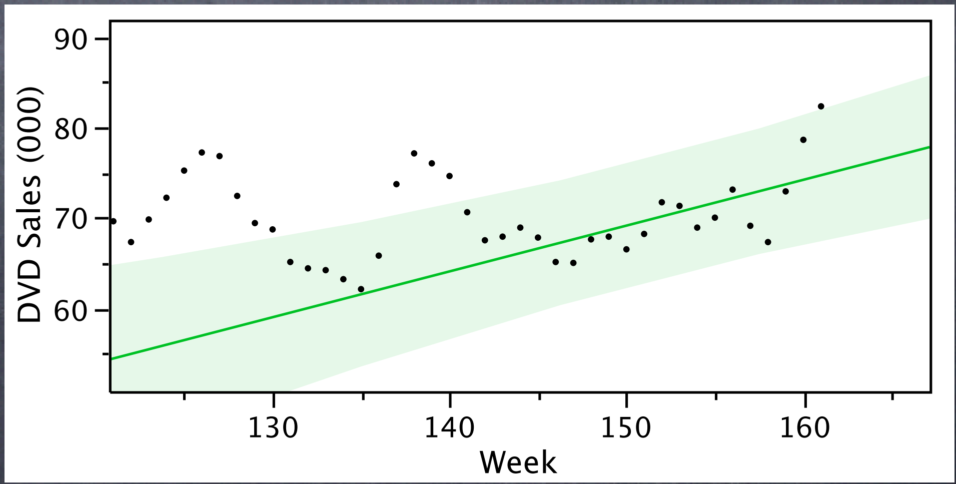
- Forecasts of stationary ARMA processes damp down to mean, with widening prediction intervals
- Integrated forecasts
  - After differencing (usually once) the model predicts the changes in the process.
  - Forecasts of changes behave like forecasts of a stationary ARMA process
    - Hence, predicted changes revert to mean change
    - Accuracy of predicted changes diminishes
  - Software "integrates" (accumulates) predicted changes back to the level of the observations
    - Sums the estimated future changes
    - Combines the standard errors of the forecasts
    - Takes into account that the forecasts are correlated

Case Study: DVDs

- Objective: Forecast DVD unit sales 6 weeks out
- Simple baseline model: the "ruler"
  - Fit ruler to the end of the data
  - Only use last 20 weeks of data to fit model



  - Pretend used linear regression to get prediction intervals



Forecasts

Baseline Model

	Prediction	Lower	Upper
1	75.19	67.91	82.48
2	75.70	68.32	83.08
3	76.21	68.72	83.69
4	76.71	69.11	84.31
5	77.22	69.50	84.94
6	77.73	69.88	85.57

???

Arima Forecasts

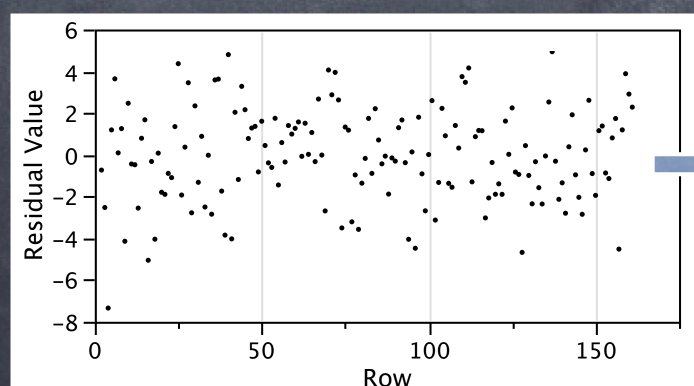
ARIMA Model for DVDs

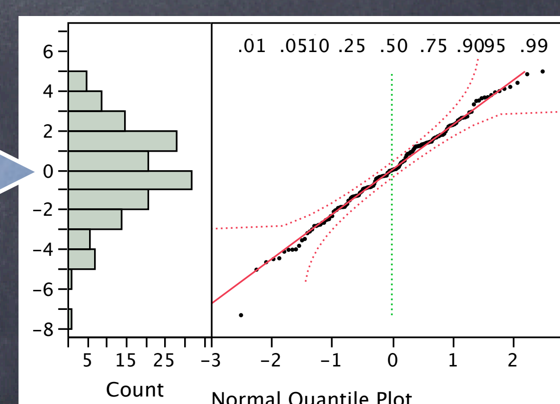
- Time series modeling
  - Use full time series, all 161 weeks
  - Differenced data to obtain stationary process
  - Settled upon IMA(1,6) model (previous class)
    - Compared variety of ARIMA models
    - Used model selection criteria to decide which to use

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
MA1	1	−0.6168536	0.0768915	−8.02	<.0001*	0.23371712
MA2	2	0.0410673	0.0857010	0.48	0.6325	
MA3	3	−0.0121524	0.0846485	−0.14	0.8860	
MA4	4	0.0539297	0.0871212	0.62	0.5368	
MA5	5	0.1817214	0.1070533	1.70	0.0916	
MA6	6	0.4577184	0.0793308	5.77	<.0001*	
Intercept	0	0.2337171	0.1592325	1.47	0.1442	

- Residuals
  - Normal
  - Initial outlier





Forecasts

Baseline Model

Estimate	Lower	Upper	Length
75.19	67.91	82.48	14.57
75.70	68.32	83.08	14.76
76.21	68.72	83.69	14.97
76.71	69.11	84.31	15.2
77.22	69.50	84.94	15.44
77.73	69.88	85.57	15.69

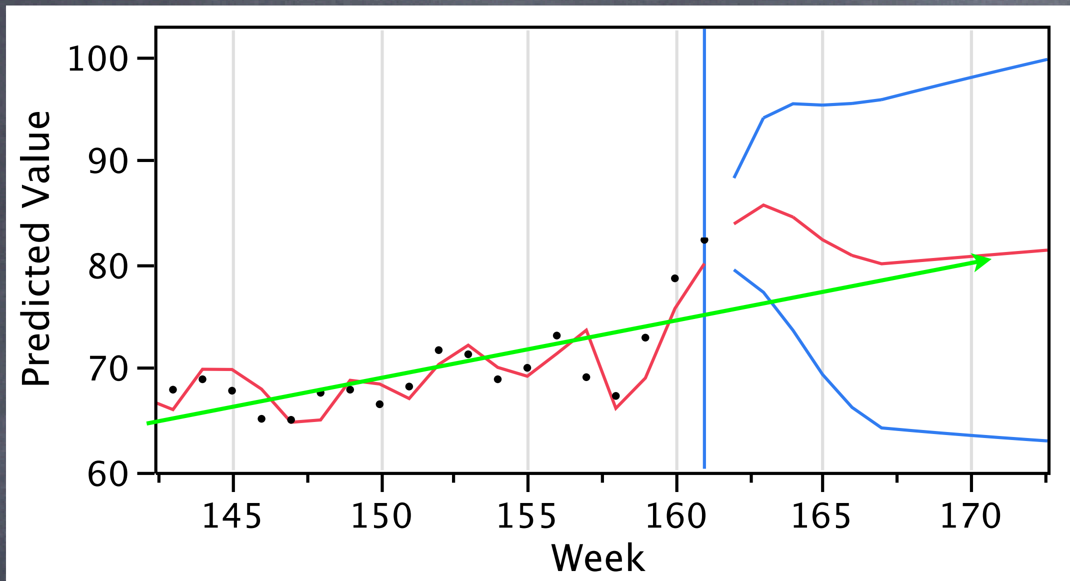
IMA(6)

	Estimate	Lower	Upper	Length
1	83.77	79.35	88.20	8.85
2	85.60	77.16	94.04	16.88
3	84.54	73.55	95.53	21.98
4	82.30	69.23	95.37	26.14
5	80.70	65.96	95.44	29.48
6	79.82	63.89	95.75	31.86

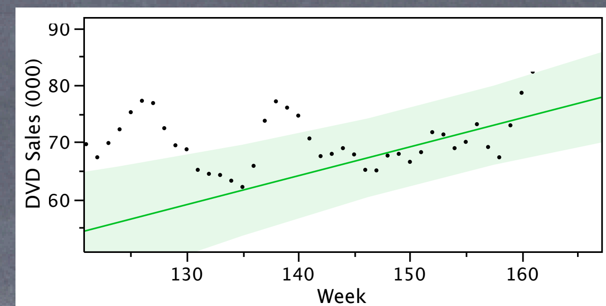
- Discussion
  - IMA forecasts are initially much higher, then shrink
  - Regression forecasts ≈ constant, with equal "accuracy"*
  - IMA model claims to be more accurate for one period, then offers very wide prediction intervals
  - Which is better? (Text fits ARIMA(2,1,6))

*Accuracy? These are claims of accuracy. Who knows if either is the "true" model.

View of Forecasts



Arima estimates appear more aligned with short-term future of data series.



- Eventual linear trend in predictions is characteristic of an integrated model
- Rapid widening of prediction intervals typical for an integrated ARIMA model

Summary

- Forecasting procedure
  - Only begins once model is identified
  - Substitution of estimates for needed values
  - Treat estimated model as if "true" model
- Case study shows
  - Designed for short-term forecasting
  - ARIMA forecasts revert to long-run form quickly
    - Mean if stationary, trend if integrated
  - Prediction intervals rapidly widen as extrapolate
- Long-term forecasts?
  - Need leading indicators or a crystal ball!