Overview of Time Series and Forecasting:

Data taken over time (usually equally spaced)

Y_t = data at time t μ = mean (constant over time)

Models:

"**Autoregressive**"

$$(Y_t - \mu) = \alpha_1 (Y_{t-1} - \mu) + \alpha_2 (Y_{t-2} - \mu) + \cdots + \alpha_p (Y_{t-p} - \mu) + e_t$$

e_t independent, constant variance: "**White Noise**"

How to find p? Regress Y on lags.

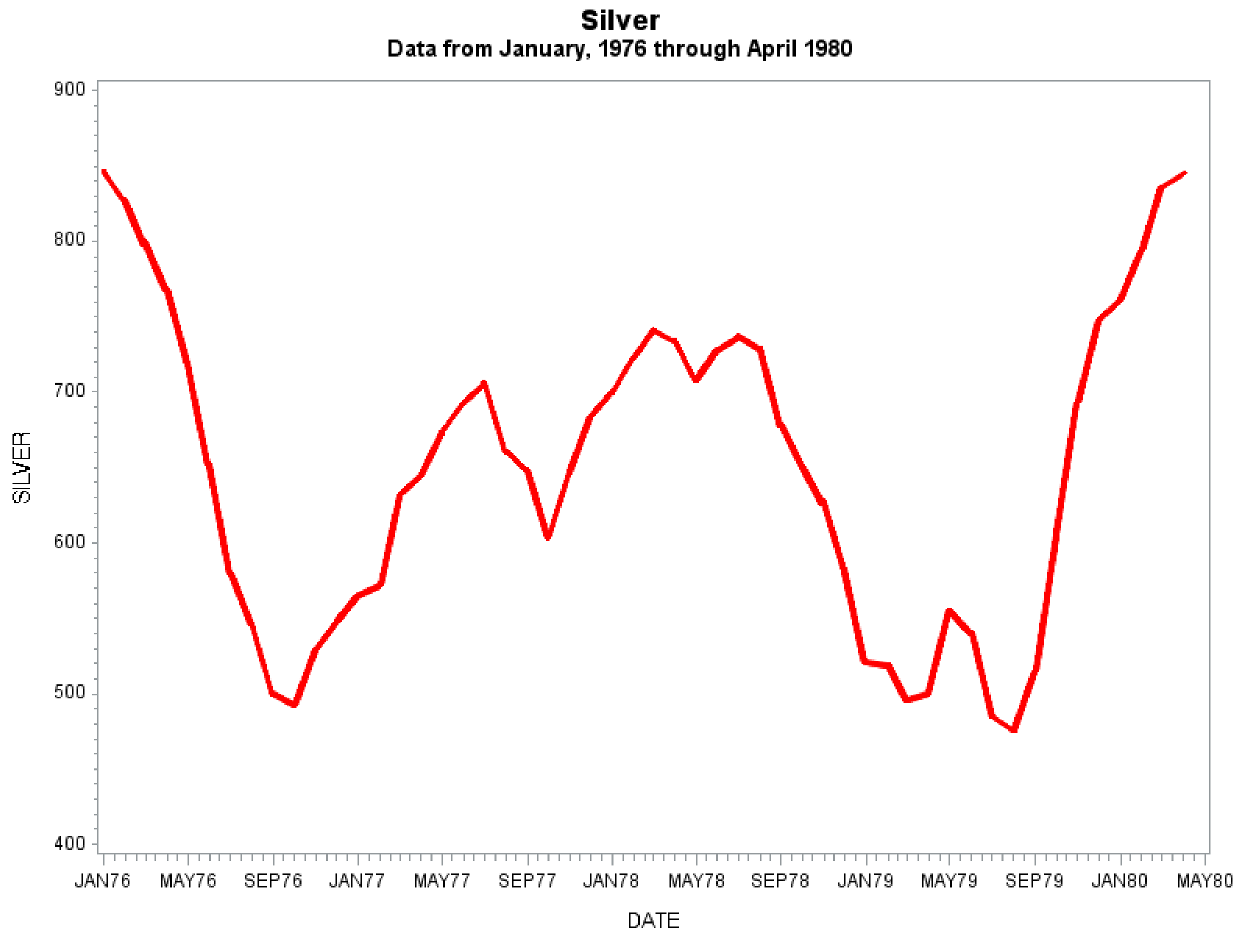
PACF Partial Autocorrelation Function

(1) Regress Y_t on Y_{t-1} then Y_t on Y_{t-1} and Y_{t-2}

then Y_t on Y_{t-1}, Y_{t-2}, Y_{t-3} etc.

(2) Plot last lag coefficients versus lags.

Example 1: Supplies of Silver in NY commodities exchange:

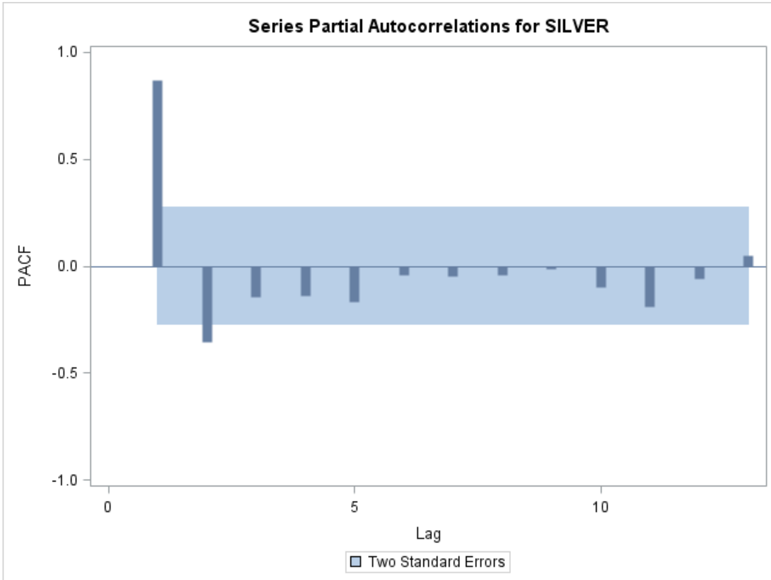


Getting PACF (and other identifying plots). SAS™ code:

```
PROC ARIMA data=silver
  plots(unpack) = all;
    identify var=silver; run;
```

™ SAS and its products are registered trademarks of SAS Institute, Cary NC.

PACF



"Spikes" outside 2 standard error bands are statistically significant

Two spikes → p=2

$$(Y_t - \mu) = \alpha_1 (Y_{t-1} - \mu) + \alpha_2 (Y_{t-2} - \mu) + e_t$$

How to estimate μ and α's ? PROC ARIMA's ESTIMATE statement.

Use maximum likelihood (ml option)

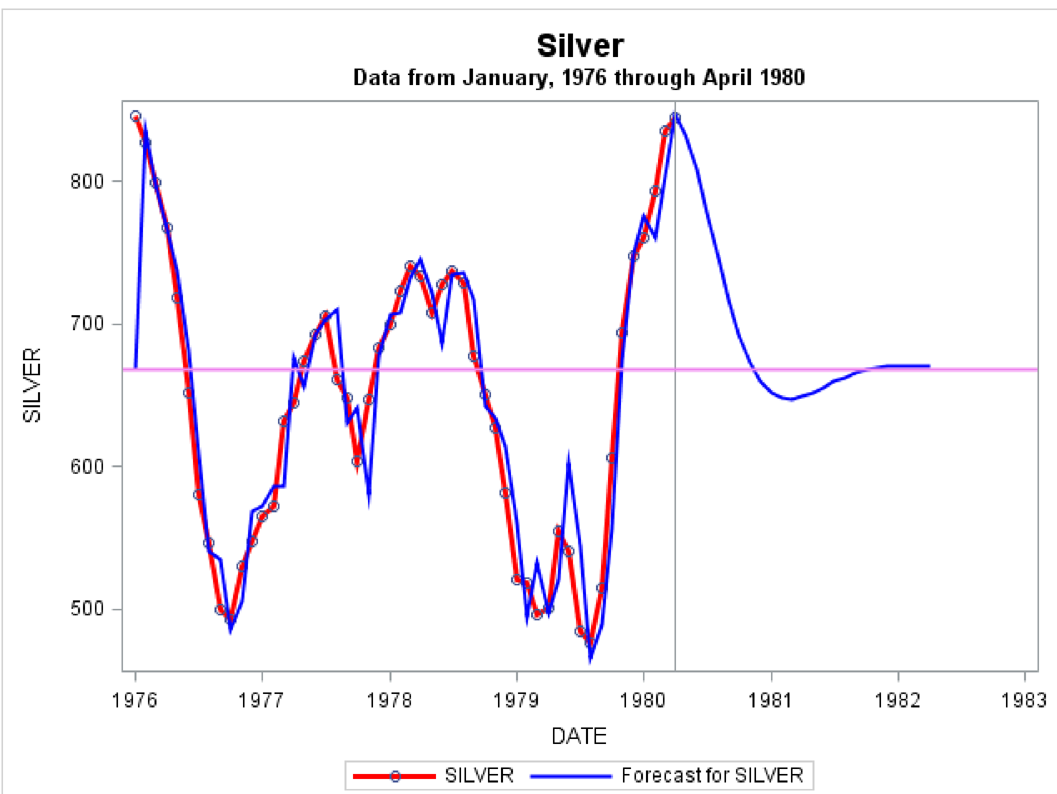
```
PROC ARIMA data=silver plots(unpack) = all;
  identify var=silver;

   estimate p=2 ml;
```

Maximum Likelihood Estimation					
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**
MU	668.29592	38.07935	17.55	<.0001	0
AR1,1	1.57436	0.10186	15.46	<.0001	1
AR1,2	-0.67483	0.10422	-6.48	<.0001	2

$$(Y_t - \mu) - \alpha_1 (Y_{t-1} - \mu) - \alpha_2 (Y_{t-2} - \mu) = e_t$$

$$(Y_t - 668) - 1.57(Y_{t-1} - 668) + 0.67(Y_{t-2} - 668) \quad = \quad e_t$$

$$(Y_t - 668) \quad = \quad 1.57(Y_{t-1} - 668) - 0.67(Y_{t-2} - 668) + e_t$$



Backshift notation: $B(Y_t)=Y_{t-1}$, $B^2(Y_t)=B(B(Y_t))=Y_{t-2}$

$$(1-1.57B+0.67B^2)(Y_t-668)=e_t$$

SAS output: (uses backshift)

Autoregressive Factors
Factor 1: 1 - 1.57436 B**(1) + 0.67483 B**(2)

Checks:

(1) Overfit (try AR(3))

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	664.88129	35.21080	18.88	<.0001	0
AR1,1	1.52382	0.13980	10.90	<.0001	1
AR1,2	-0.55575	0.24687	-2.25	0.0244	2
AR1,3	-0.07883	0.14376	-0.55	**0.5834**	**3**

(2) Residual autocorrelations

Residual r_t

Residual autocorrelation at lag j: $\mathrm{Corr}(r_t, r_{t-j}) = \rho(j)$

Estimate, square, and sum k of these multiply by sample size n. PROC ARIMA: k in sets of 6. Box-Pierce Q statistic. Limit distribution Chi-square if errors independent. Later modification: Box-Ljung statistic for H_0:residuals uncorrelated

$$n\sum_{j=1}^{k}\left(\frac{n+2}{n-j}\right)\hat{\rho}_j^2$$

SAS output:

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.49	4	**0.4794**	-0.070	-0.049	-0.080	0.100	-0.112	0.151
12	5.97	10	**0.8178**	0.026	-0.111	-0.094	-0.057	0.006	-0.110
18	10.27	16	**0.8522**	-0.037	-0.105	0.128	-0.051	0.032	-0.150
24	16.00	22	**0.8161**	-0.110	0.066	-0.039	0.057	0.200	-0.014

Residuals uncorrelated←→Residuals are White Noise
←→ Residuals are unpredictable

SAS computes Box-Ljung on original data too.

Autocorrelation Check for White Noise									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	81.84	6	**<.0001**	0.867	0.663	0.439	0.214	-0.005	-0.184
12	142.96	12	<.0001	-0.314	-0.392	-0.417	-0.413	-0.410	-0.393

Data autocorrelated ←→ predictable!

Note: All p-values are based on an assumption called "stationarity" discussed later.

How to predict?

$$(Y_t - \mu) - \alpha_1 (Y_{t-1} - \mu) - \alpha_2 (Y_{t-2} - \mu) = e_t$$

One step prediction

$$\widehat{Y}_{t+1} = \mu + \alpha_1 (Y_t - \mu) + \alpha_2 (Y_{t-1} - \mu), \quad future\ error = e_{t+1}$$

Two step prediction

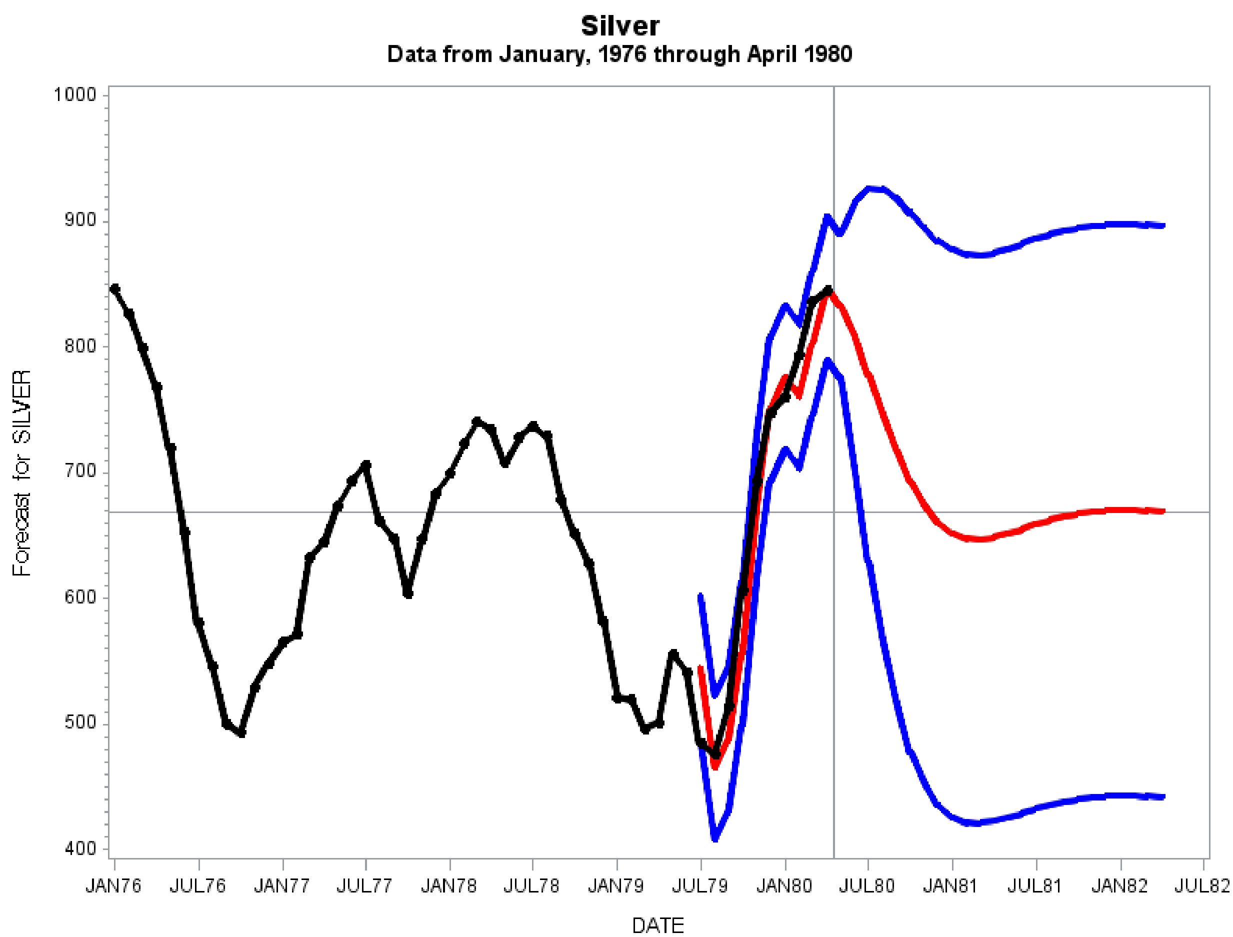
$$\widehat{Y}_{t+2} = \mu + \alpha_1(\widehat{Y}_{t+1} - \mu) + \alpha_2(Y_t - \mu), \quad error = e_{t+2} + \alpha_1 e_{t+1}$$

etc.

<u>*Prediction error*</u> variance (σ^2 = variance(e_t))

$$\sigma^2,\ (1+\alpha_1^2)\sigma^2,\ \ldots$$

From prediction error variances, get 95% prediction intervals. Can estimate variance of e_t from past data. SAS PROC ARIMA does it all for you!



Moving Average, MA(q), and ARMA(p,q) models

MA(1) $Y_t = \mu + e_t - \theta e_{t-1}$ Variance $(1+\theta^2)\sigma^2$

$Y_{t-1} = \mu + e_{t-1} - \theta e_{t-2}$ $\rho(1)=-\theta/(1+\theta^2)$

$Y_{t-2} = \mu + e_{t-2} - \theta e_{t-3}$ $\rho(2)=0/(1+\theta^2)=0$

Autocorrelation function "**ACF**" ($\rho(j)$) is 0 after lag q for MA(q). PACF is useless for identifying q in MA(q).

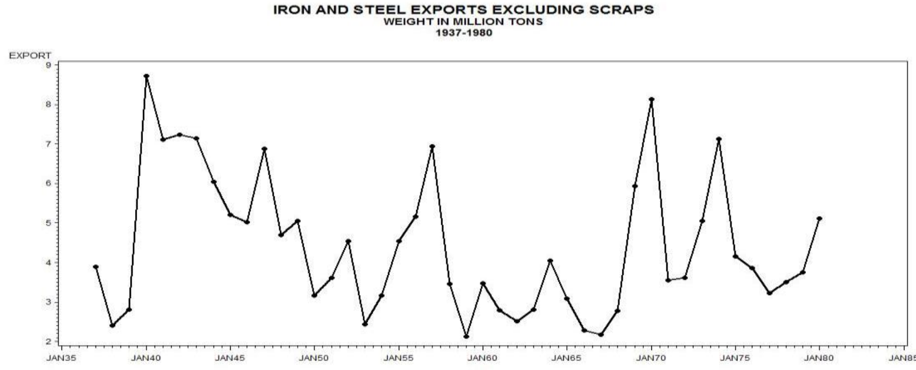
PACF drops to 0 after lag 3 → AR(3) p=3

ACF drops to 0 after lag 2 → MA(2) q=2

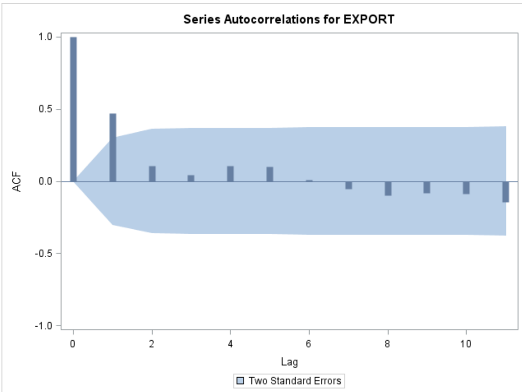
Neither drops → ARMA(p,q) p= ___ q=____

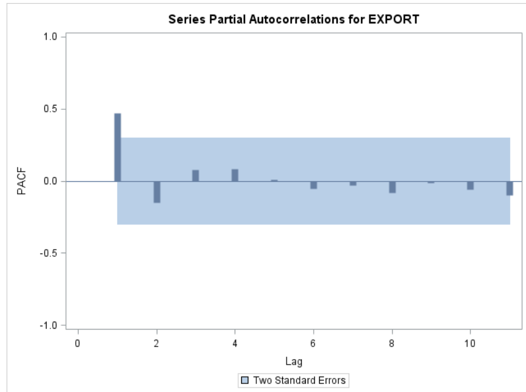
$$(Y_t - \mu) - \alpha_1 (Y_{t-1} - \mu) - ... - \alpha_p (Y_{t-p} - \mu) = e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$$

Example 2: Iron and Steel Exports.



```
PROC ARIMA plots(unpack)=all;
  Identify VAR=EXPORT;
```





ACF (could be MA(1))
Spike at lags 0, 1

PACF (could be AR(1))
No spike at lag 0

```
  Estimate P=1 ML;
  Estimate Q=2 ML;
  Estimate Q=1 ML;
```

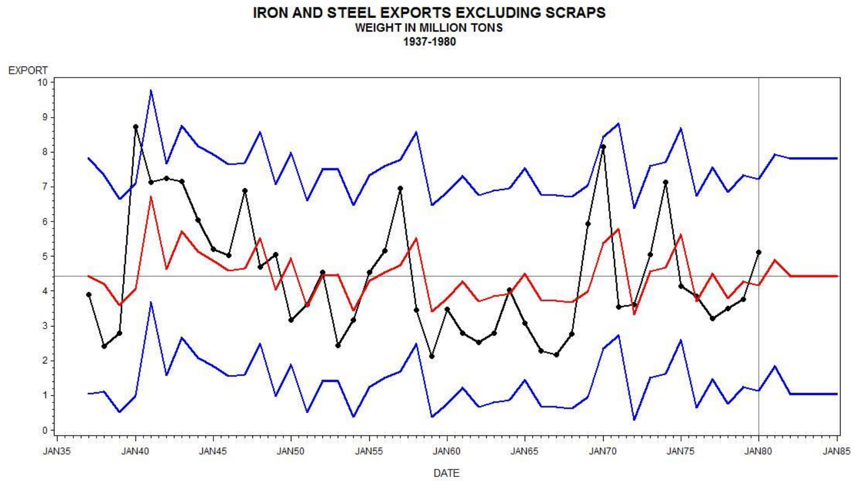
Maximum Likelihood Estimation

Parameter	Estimate	t Value	Approx Pr>|t|	Lag
MU	4.42129	10.28	<.0001	0
AR1,1	0.46415	3.42	0.0006	1
MU	4.43237	11.41	<.0001	0
MA1,1	-0.54780	-3.53	0.0004	1
MA1,2	-0.12663	-0.82	0.4142	2
MU	**4.42489**	**12.81**	**<.0001**	**0**
MA1,1	**-0.49072**	**-3.59**	**0.0003**	**1**

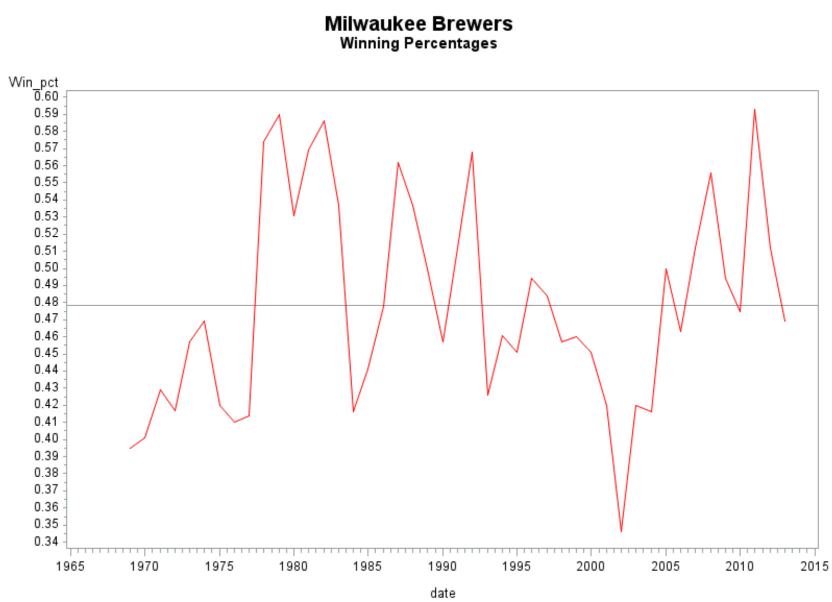
How to choose? AIC - smaller is better

```
AIC                    165.8342   (MA(1))
AIC                    166.3711   (AR(1))
AIC                    167.1906   (MA(2))
```

```
Forecast lead=5 out=out1 id=date interval=year;
```



Example 3: Brewers' Proportion Won



```
                 Mean of Working Series       0.478444
                    Standard Deviation        0.059934
                    Number of Observations          45

                                 Autocorrelations
Lag               -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  Correlation                                                    Std Error
0    1.00000     |                    |********************|         0
1    0.52076     |              .     |**********          |  0.149071
2    0.18663     |             .      |****    .           |  0.185136
3    0.11132     |            .       |**       .          |  0.189271
4    0.11490     |            .       |**       .          |  0.190720
5    -.00402     |            .       |         .          |  0.192252
6    -.14938     |            .    ***|         .          |  0.192254
7    -.13351     |            .    ***|         .          |  0.194817
8    -.06019     |            .      *|         .          |  0.196840
9    -.05246     |            .      *|         .          |  0.197248
10   -.20459     |            .   ****|         .          |  0.197558
11   -.22159     |            .   ****|         .          |  0.202211
12   -.24398     |            .  *****|         .          |  0.207537
                          "." marks two standard errors
```

Could be MA(1)

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	17.27	6	0.0084	0.521	0.187	0.111	0.115	-0.004	-0.149
12	28.02	12	0.0055	-0.134	-0.060	-0.052	-0.205	-0.222	-0.244

<u>NOT</u> White Noise!

SAS Code:

```
proc arima data=brewers;
 identify var=Win_Pct nlag=12; run;
 estimate q=1 ml;
```

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.47791	0.01168	40.93	<.0001	0
MA1,1	**-0.50479**	0.13370	**-3.78**	**0.0002**	1

AIC -135.099

Autocorrelation Check of **Residuals**

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.51	5	**0.6219**	0.095	0.161	0.006	0.119	0.006	-0.140
12	11.14	11	**0.4313**	-0.061	-0.072	0.066	-0.221	-0.053	-0.242
18	13.54	17	**0.6992**	0.003	-0.037	-0.162	-0.010	-0.076	-0.011
24	17.31	23	**0.7936**	-0.045	-0.035	-0.133	-0.087	-0.114	0.015

Estimated Mean 0.477911
Moving Average Factors
Factor 1: 1 + 0.50479 B**(1)

Partial Autocorrelations

```
Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1        0.52076     |              .     |**********           |
  2       -0.11603     |              .   **|     .               |
  3        0.08801     |              .     |**   .               |
  4        0.04826     |              .     |*    .               |
  5       -0.12646     |              .  ***|     .               |
  6       -0.12989     |              .  ***|     .               |
  7        0.01803     |              .     |     .               |
  8        0.01085     |              .     |     .               |
  9       -0.02252     |              .     |     .               |
 10       -0.20351     |              . ****|     .               |
 11       -0.03129     |              .    *|     .               |
 12       -0.18464     |              . ****|     .               |
```

OR … could be AR(1)

```
estimate p=1 ml;
```

```
                        Maximum Likelihood Estimation
                          Standard                  Approx
Parameter     Estimate       Error  t Value     Pr > |t|      Lag

  MU           0.47620    0.01609    29.59        <.0001        0
  AR1,1        0.53275    0.12750     4.18        <.0001        1
              AIC  -136.286  (vs. -135.099)
```

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.57	5	0.6134	0.050	-0.133	-0.033	0.129	0.021	-0.173
12	8.66	11	0.6533	-0.089	0.030	0.117	-0.154	-0.065	-0.181
18	10.94	17	0.8594	0.074	0.027	-0.161	0.010	-0.019	0.007
24	13.42	23	0.9423	0.011	-0.012	-0.092	-0.081	-0.106	0.013

```
                 Model for variable Win_pct
                 Estimated Mean      0.476204
                   Autoregressive Factors
                 Factor 1:   1 - 0.53275 B**(1)
```

Conclusions for Brewers:

Both models have statistically significant parameters.

Both models are sufficient (no lack of fit)

Predictions from MA(1):

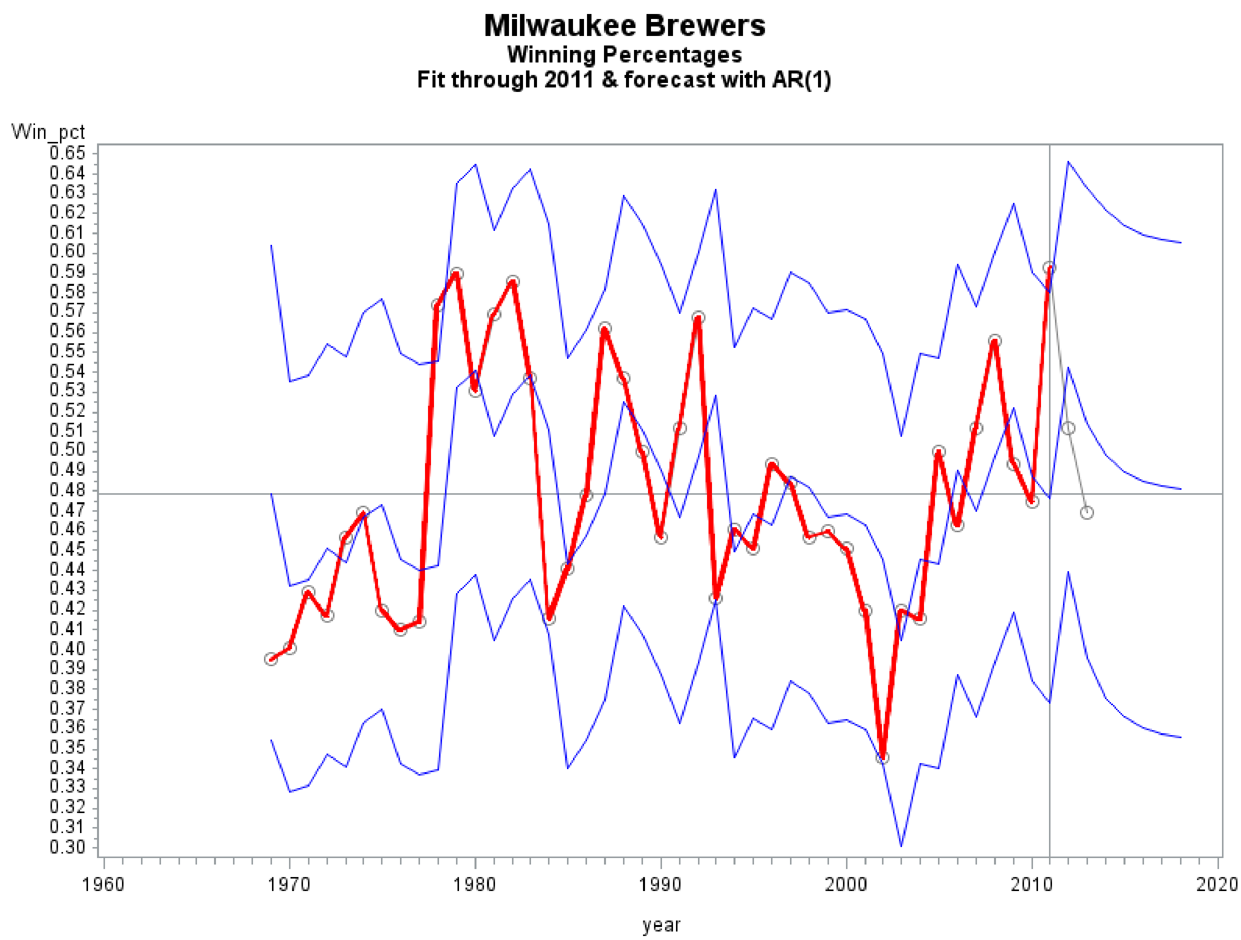
First one uses correlations

The rest are on the mean.

Predictions for AR(1):

Converge exponentially fast toward mean

Not much difference but AIC prefers AR(1)



Stationarity

(1) Mean constant (no trends)
(2) Variance constant
(3) Covariance $\gamma(j)$ and correlation

$$\rho(j) = \gamma(j)/\gamma(j)$$

between Y_t and Y_{t-j} depend only on j

ARMA(p,q) model

$$(Y_t - \mu) - \alpha_1 (Y_{t-1} - \mu) - \ldots - \alpha_p (Y_{t-p} - \mu) = e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$$

Stationarity guaranteed whenever solutions of equation (roots of polynomial)

$$X^p - \alpha_1 X^{p-1} - \alpha_2 X^{p-2} - \ldots - \alpha_p = 0$$

are all <1 in magnitude.

Examples

(1) $Y_t - \mu = .8(Y_{t-1} - \mu) + e_t$ $X - .8 = 0 \rightarrow X = .8$
stationary

(2) $Y_t - \mu = 1.00(Y_{t-1} - \mu) + e_t$ **nonstationary**

Note: $Y_t = Y_{t-1} + e_t$ Random walk

(3) $Y_t - \mu = 1.6(Y_{t-1} - \mu) - 0.6(Y_{t-1} - \mu) + e_t$

"characteristic polynomial"

$X^2 - 1.6X + 0.6 = 0 \rightarrow$ **$X=1$** or $X=0.6$

nonstationary (unit root X=1)

$(Y_t - \mu) - (Y_{t-1} - \mu) = 0.6[(Y_{t-1} - \mu) - (Y_{t-2} - \mu)] + e_t$

$(Y_t - Y_{t-1}) = 0.6(Y_{t-1} - Y_{t-2}) + e_t$

First differences form stationary AR(1) process!

No mean – no mean reversion – no gravity pulling toward the mean.

(4) $Y_t - \mu = 1.60(Y_{t-1} - \mu) - 0.63(Y_{t-1} - \mu) + e_t$

$X^2 - 1.60X + 0.63 = 0 \rightarrow X = 0.9$ or $X = 0.7$

$|\text{roots}| < 1 \rightarrow$ **stationary**

$(Y_t - \mu) - (Y_{t-1} - \mu) =$

$\mathbf{-0.03(Y_{t-1} - \mu)} + 0.63[(Y_{t-1} - \mu) - (Y_{t-2} - \mu)] + e_t$

$Y_t - Y_{t-1} = \mathbf{-0.03(Y_{t-1} - \mu)} + 0.63(Y_{t-1} - Y_{t-2}) + e_t$

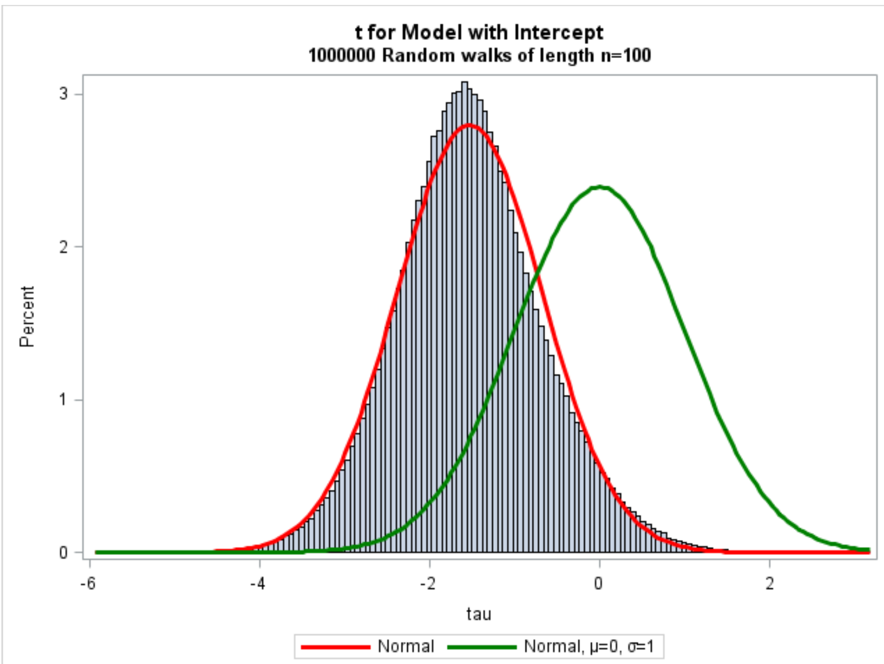
Unit Root testing (H_0:Series has a unit root)

Regress

$Y_t - Y_{t-1}$ on $\mathbf{Y_{t-1}}$ and $(Y_{t-1} - Y_{t-2})$

Look at t test for $\mathbf{Y_{t-1}}$. If it is significantly negative then stationary.

Problem: Distribution of t stat is ***not t*** ***<u>distribution</u>*** under unit root hypothesis. Distribution looks like this histogram:



Overlays: N(sample mean & variance)   N(0,1)

Correct distribution: Dickey-Fuller test in PROC ARIMA.

-2.89 is the correct (left) 5th %ile

46% of t's are less than -1.645

(the normal 5th percentile)

Example 1: Brewers

```
proc arima data=brewers;
 identify var=Win_Pct nlag=12 stationarity=(ADF=0);
```

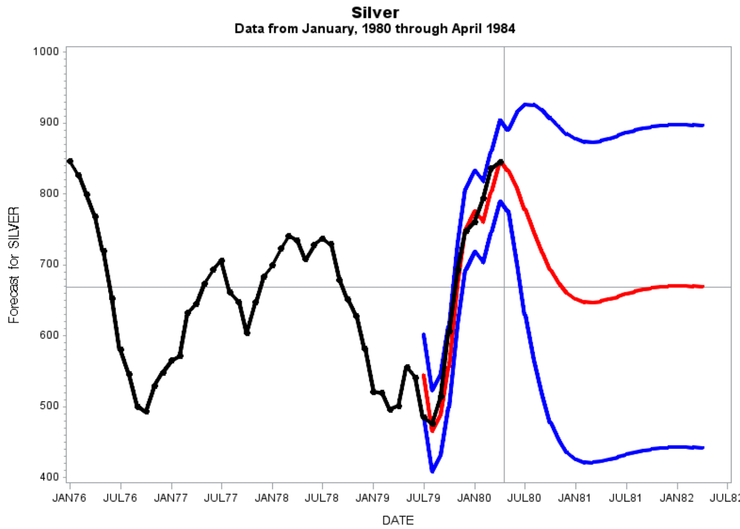
Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-0.1803	0.6376	-0.22	0.6002
Single Mean	0	-21.0783	0.0039	-3.75	0.0062
Trend	0	-21.1020	0.0287	-3.68	0.0347

Conclusion reject H_0:unit roots so Brewers series is stationary (mean reverting).

0 lags → do not need lagged differences in model (just regress Y_t-Y_{t-1} on Y_{t-1})

Example 2: Stocks of silver revisited



Needed AR(2) (2 lags) so regress

$Y_t - Y_{t-1}$ (D_Silver) on

Y_{t-1} (L_Silver) and $Y_{t-1} - Y_{t-2}$ (D_Silver_1)

PROC REG:

Parameter Estimates

Variable	DF	Parameter Estimate	t Value	Pr>|t|	
Intercept	1	75.58073	2.76	0.0082	
L_Silver	1	-0.11703	-2.78	0.0079	☹ wrong distn.
D_Silver_1	1	0.67115	6.21	<.0001	☺ OK

PROC ARIMA:

Augmented Dickey-Fuller Unit Root Tests

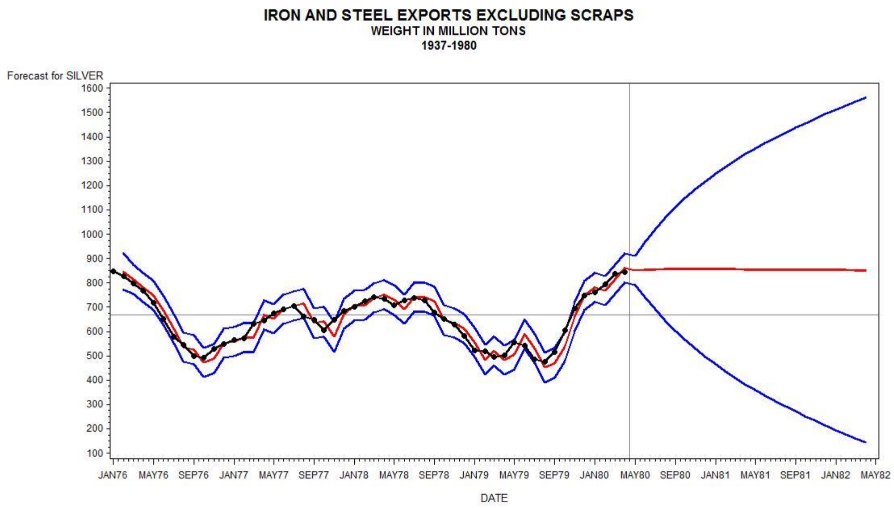
Type	Lags	Rho	Pr<Rho	Tau	Pr<Tau	
Zero Mean	1	-0.2461	0.6232	-0.28	0.5800	
Single Mean	1	-17.7945	0.0121	**-2.78**	**0.0689**	☺ OK
Trend	1	-15.1102	0.1383	-2.63	0.2697	

Same t statistic, corrected p-value!

Conclusion: Unit root → difference the series.

1 lag → need 1 lagged difference in model (regress $Y_t - Y_{t-1}$ on Y_{t-1} and $Y_{t-1} - Y_{t-2}$)

```
PROC ARIMA data=silver;
  identify var=silver(1) stationarity=(ADF=(0));
  estimate p=1 ml;
  forecast lead=24 out=outN ID=date
                         Interval=month;
```



Unit root forecast & forecast interval

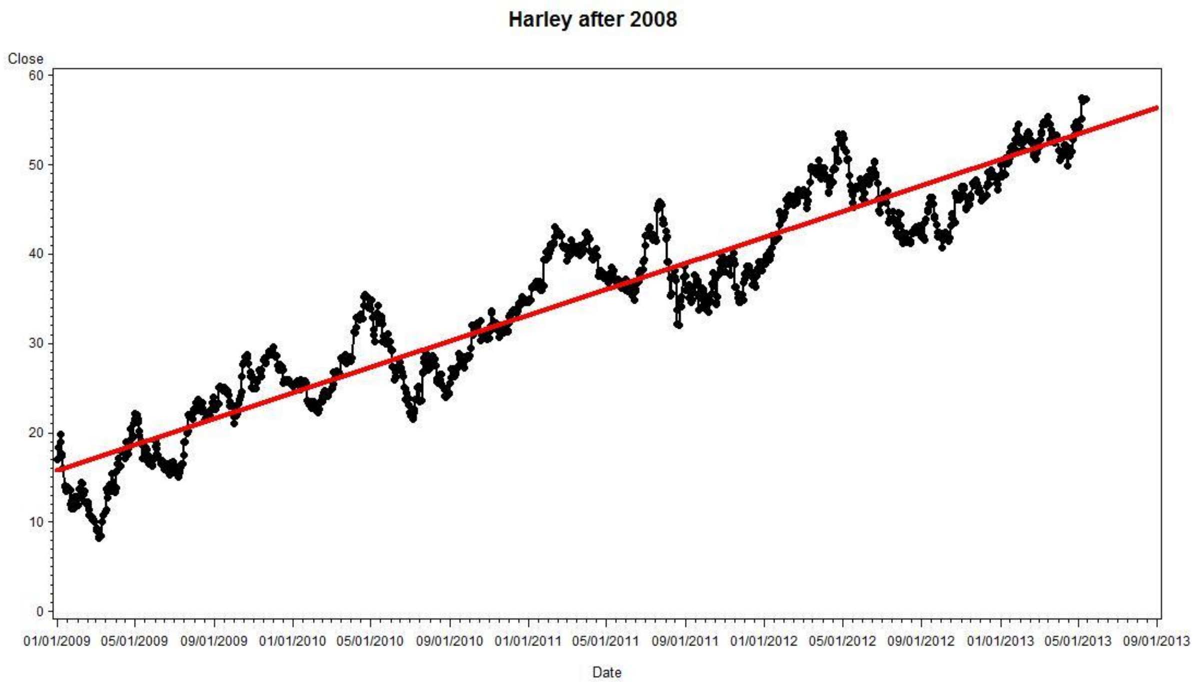
PROC AUTOREG

Fits a regression model (least squares)

Fits stationary autoregressive model to error terms

Refits accounting for autoregressive errors.

Example 3: AUTOREG Harley-Davidson closing stock prices 2009-present.




```
proc autoreg data=Harley;
model close=date/ nlag=15 backstep;
run;
```

One by one, AUTOREG eliminates insignificant lags then:

Estimates of Autoregressive Parameters			
Lag	**Coefficient**	**Standard Error**	**t Value**
1	-0.975229	0.006566	-148.53

Final model:

Parameter Estimates					
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**
Intercept	1	-412.1128	35.2646	-11.69	<.0001
Date	1	0.0239	0.001886	12.68	<.0001

Error term Z_t satisfies Z_t-0.97Z_{t-1}=e_t.

Example 3 ARIMA: Harley-Davidson closing stock prices 2009-present.

Apparent upward movement: Linear trend or nonstationary?

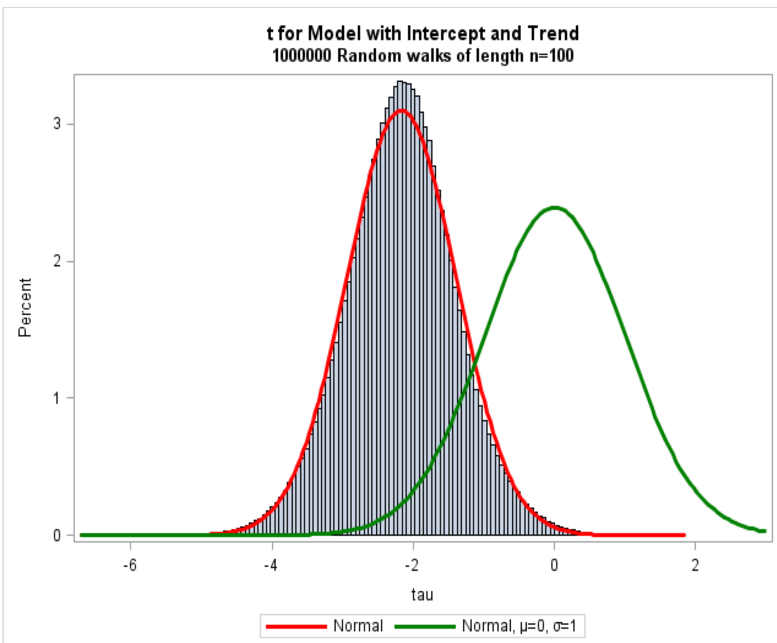
Regress

$Y_t - Y_{t-1}$ on 1, t, Y_{t-1} (& lagged differences)

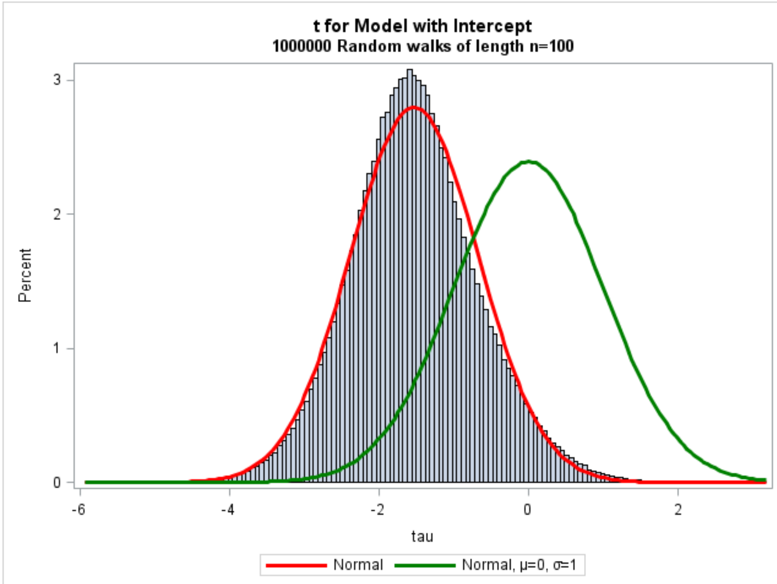
H_0: $Y_t = \beta + Y_{t-1} + e_t$ “random walk with drift”

H_1: $Y_t = \alpha + \beta t + Z_t$ with Z_t stationary

New distribution for Y_{t-1} t-test



With trend



Without trend

1million simulation runs in 7 seconds!

SAS code for Harley stock closing price

```
proc arima data=Harley;

identify var=close stationarity=(adf)
   crosscor=(date) noprint;

Estimate input=(date) p=1 ml;

forecast lead=120 id=date interval=weekday
out=out1; run;
```

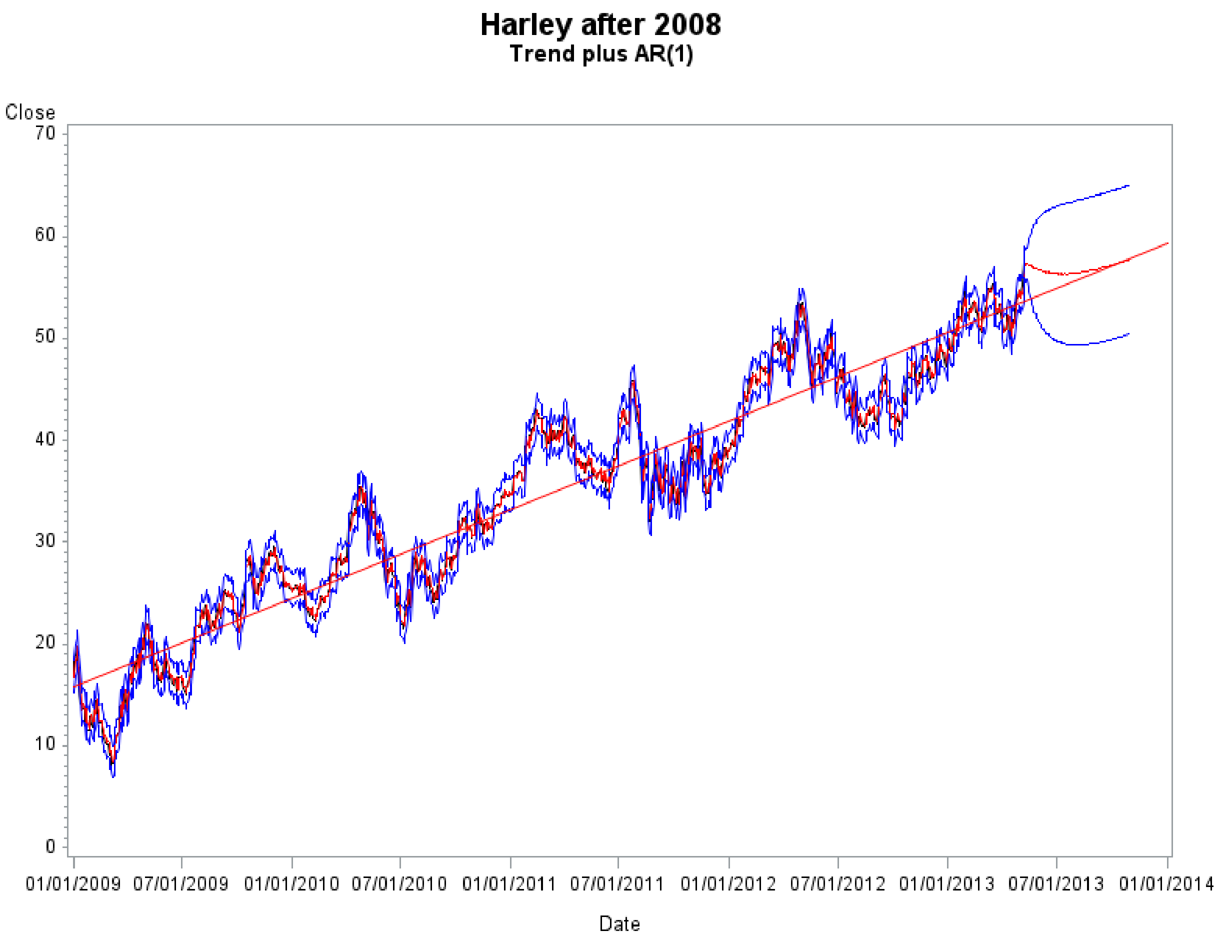
Stationarity test (0,1,2 lagged differences):

Augmented Dickey-Fuller Unit Root Tests					
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**
Zero Mean	**0**	0.8437	0.8853	1.14	0.9344
	1	0.8351	0.8836	1.14	0.9354
	2	0.8097	0.8786	1.07	0.9268
Single Mean	**0**	-2.0518	0.7726	-0.87	0.7981
	1	-1.7772	0.8048	-0.77	0.8278
	2	-1.8832	0.7925	-0.78	0.8227
Trend	**0**	-27.1559	0.0150	-3.67	**0.0248**
	1	-26.9233	0.0158	-3.64	0.0268
	2	-29.4935	0.0089	-3.80	0.0171

Conclusion: stationary around a linear trend.

Estimates: trend + AR(1)

Maximum Likelihood Estimation							
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**	**Variable**	**Shift**
MU	-412.08104	35.45718	-11.62	<.0001	0	Close	0
AR1,1	0.97528	0.0064942	150.18	<.0001	1	Close	0
NUM1	0.02391	0.0018961	12.61	<.0001	0	Date	0



Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.20	5	0.6694	-0.005	0.044	-0.023	0.000	0.017	0.005
12	6.49	11	0.8389	-0.001	0.019	0.003	-0.010	0.049	-0.003
18	10.55	17	0.8791	0.041	-0.026	-0.022	-0.023	0.007	-0.011
24	16.00	23	0.8553	0.014	-0.037	0.041	-0.020	-0.032	0.003
30	22.36	29	0.8050	0.013	-0.026	0.028	0.051	0.036	0.000
36	24.55	35	0.9065	0.037	0.016	-0.012	0.002	-0.007	0.001
42	29.53	41	0.9088	-0.007	-0.021	0.029	0.030	-0.033	0.030
48	49.78	47	0.3632	0.027	-0.009	-0.097	-0.026	-0.074	0.026

NCSU Energy Demand

Type of day

Class Days

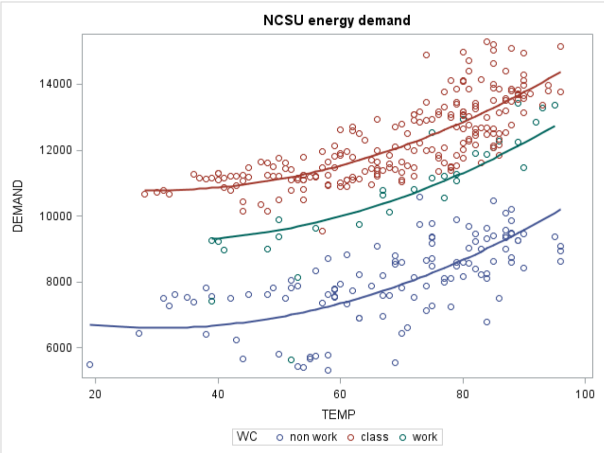
Work Days (no classes)

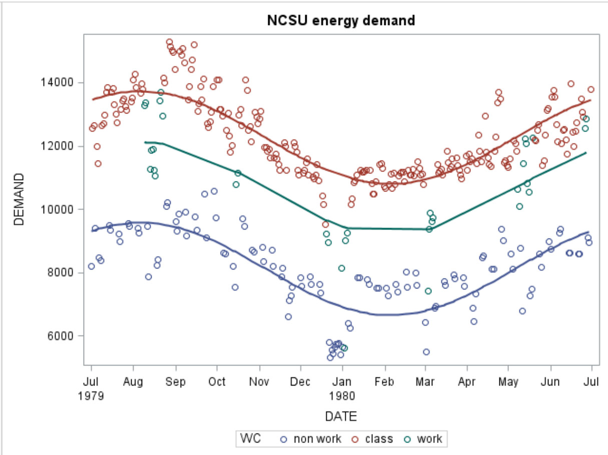
Holidays & weekends.

Temperature Season of Year

Step 1: Make some plots of energy demand vs. temperature and season. Use type of day as color.

Seasons: S = A $\sin(2\pi t/365)$, C=B $\sin(2\pi t/365)$





Temperature Season of Year

Step 2: PROC AUTOREG with all inputs:

```
PROC AUTOREG data=energy;
MODEL DEMAND = TEMP TEMPSQ CLASS WORK S C
   /NLAG=15 BACKSTEP DWPROB;
output out=out3
   predicted = p predictedm=pm
   residual=r residualm=rm;
run;
```

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.559658	0.043993	-12.72
5	-0.117824	0.045998	-2.56
7	-0.220105	0.053999	-4.08
8	0.188009	0.059577	3.16
9	-0.108031	0.051219	-2.11
12	0.110785	0.046068	2.40
14	-0.094713	0.045942	-2.06

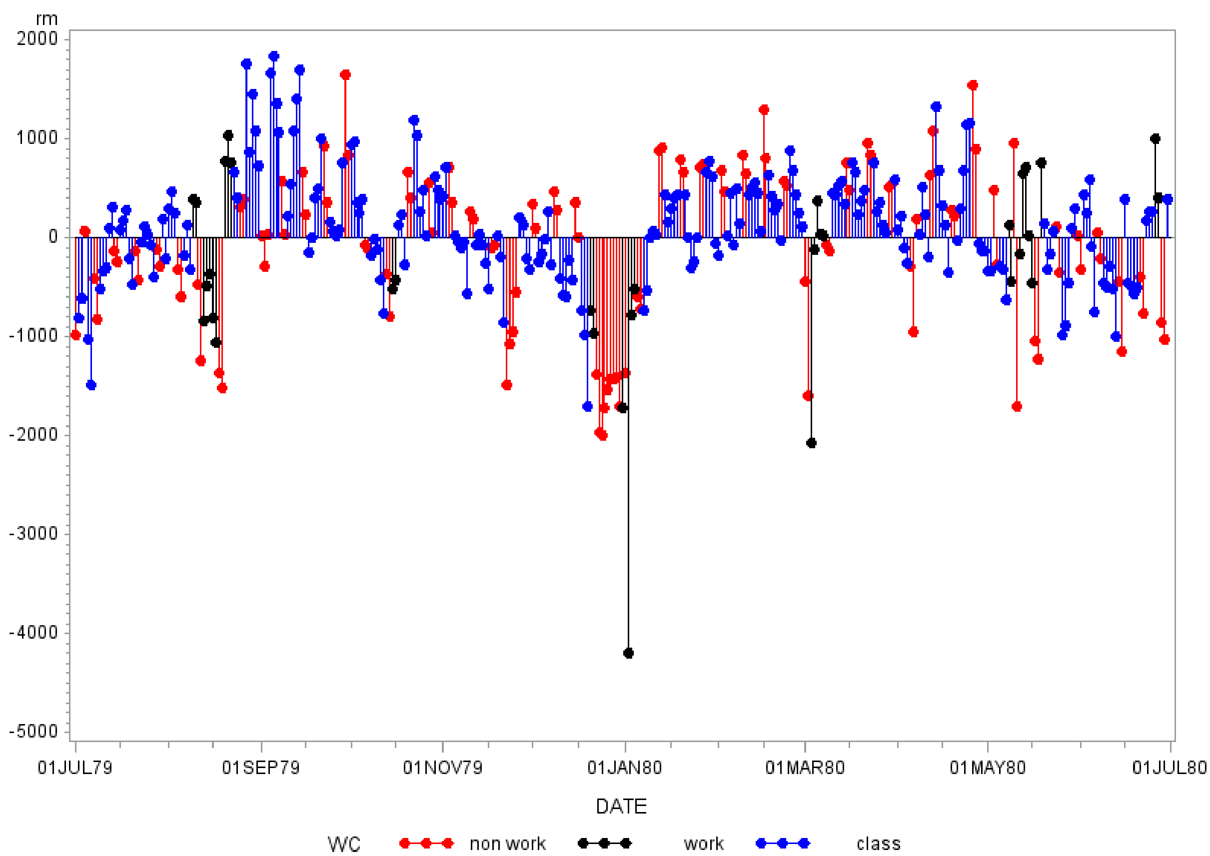
Autocorrelation at 1, 7, 14, and others.

After autocorrelation adjustments, trust t tests etc.

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	6076	296.5261	20.49	<.0001
TEMP	1	28.1581	3.6773	7.66	<.0001
TEMPSQ	1	0.6592	0.1194	5.52	<.0001
CLASS	1	1159	117.4507	9.87	<.0001
WORK	1	2769	122.5721	22.59	<.0001
S	1	-764.0316	186.0912	-4.11	<.0001
C	1	-520.8604	188.2783	-2.77	0.0060

Need better model?



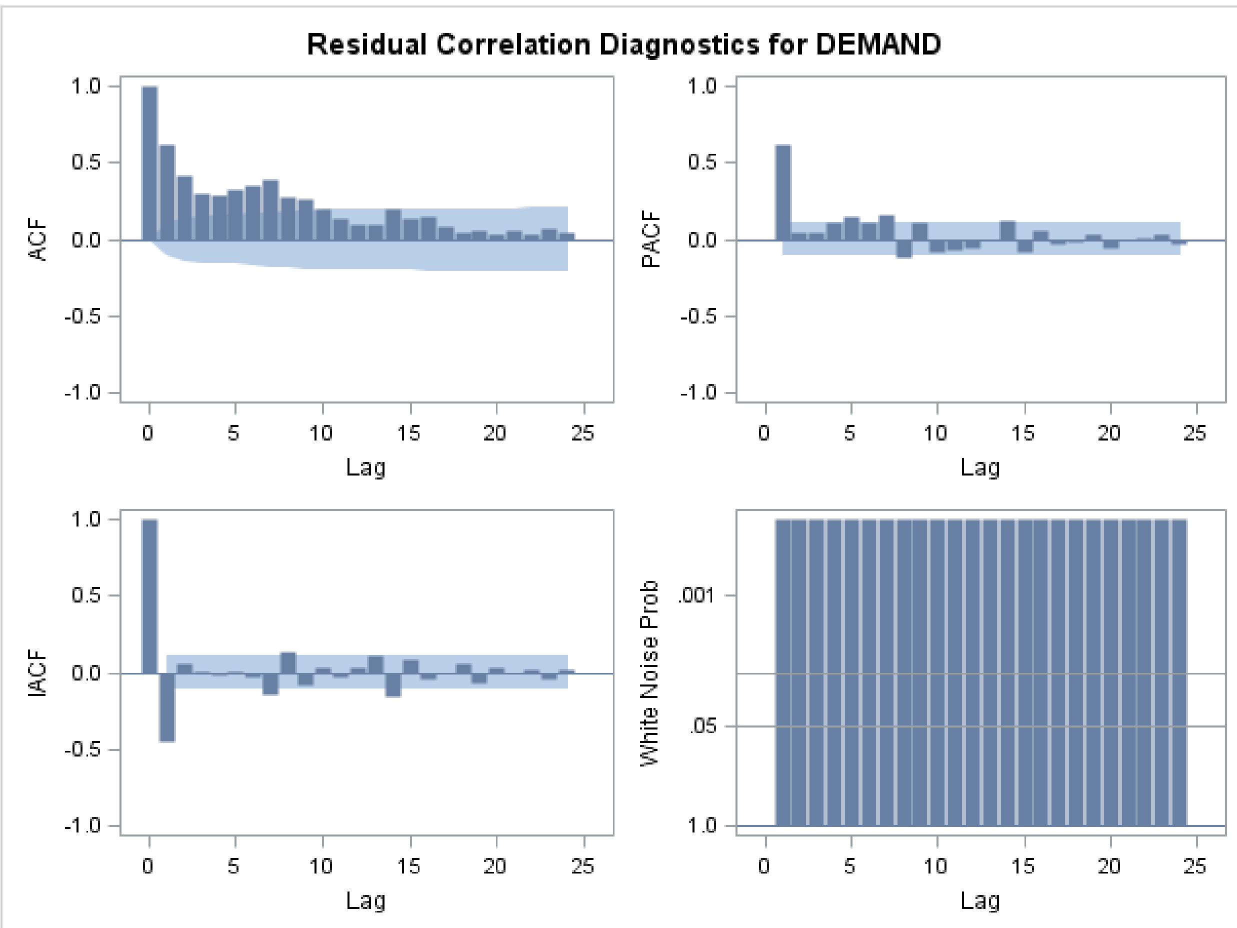
•

Residuals from regression part. Large residuals on workday near Christmas. Add dummy variable.

Same idea: PROC ARIMA

Step 1: Graphs

Step 2: Regress on inputs, diagnose residual autocorrelation:



Not white noise (bottom right)

Activity (bars) at lag 1, 7, 14

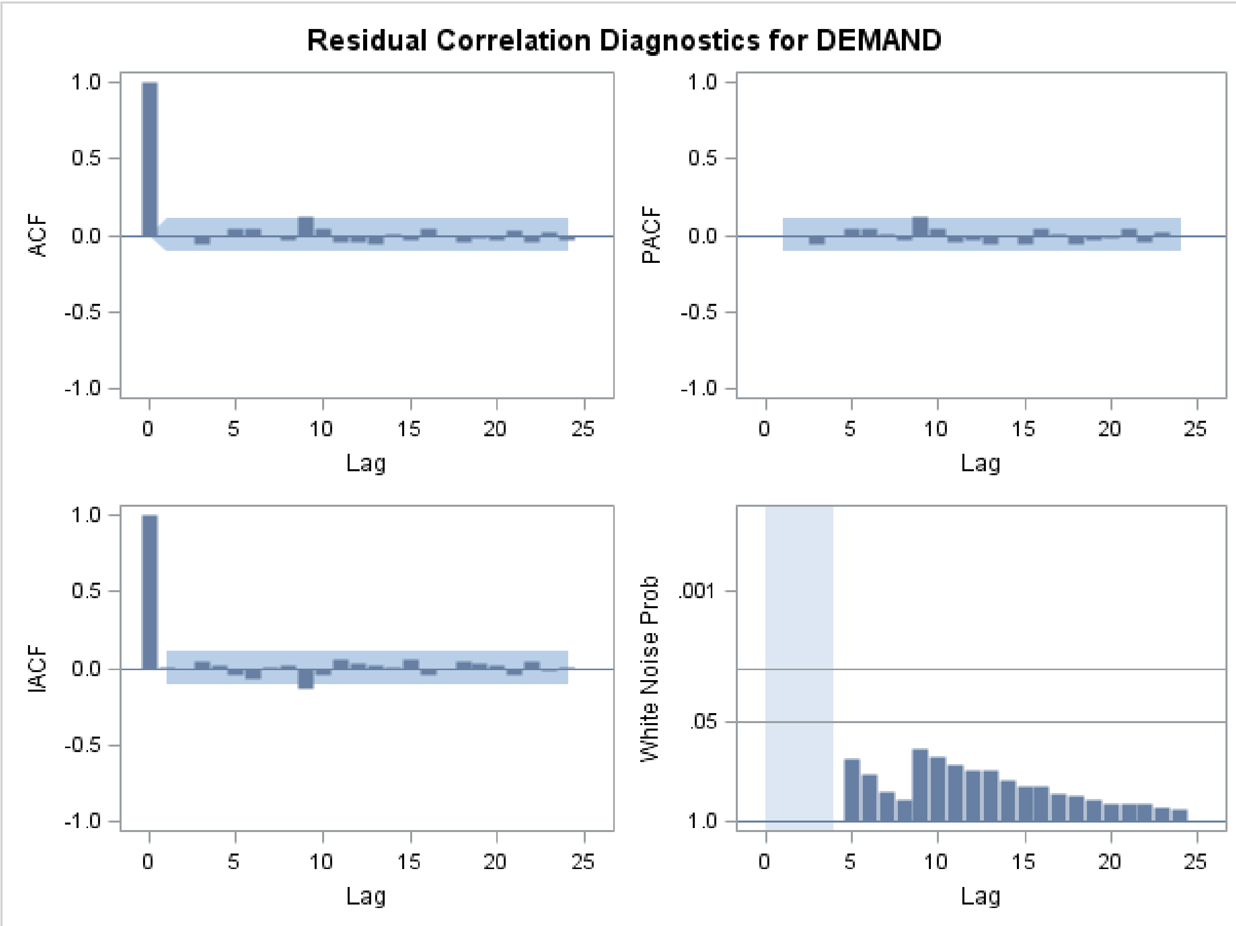
(3) Estimate resulting model from diagnostics plus trial and error:

```
  e input = (temp tempsq class work s c) p=1
q=(1,7,14) ml;
```

Maximum Likelihood Estimation							
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**	**Variable**	**Shift**
MU	6183.1	300.87297	20.55	<.0001	0	DEMAND	0
MA1,1	0.11481	0.07251	1.58	0.1133	1	DEMAND	0
MA1,2	-0.18467	0.05415	-3.41	0.0006	7	DEMAND	0
MA1,3	-0.13326	0.05358	-2.49	0.0129	14	DEMAND	0
AR1,1	0.73980	0.05090	14.53	<.0001	1	DEMAND	0
NUM1	26.89511	3.83769	7.01	<.0001	0	TEMP	0
NUM2	0.64614	0.12143	5.32	<.0001	0	TEMPSQ	0
NUM3	912.80536	122.78189	7.43	<.0001	0	CLASS	0
NUM4	2971.6	123.94067	23.98	<.0001	0	WORK	0
NUM5	-767.41131	174.59057	-4.40	<.0001	0	S	0
NUM6	-553.13620	182.66142	-3.03	0.0025	0	C	0

(Note: class days get class effect plus work effect)

(4) Check model fit (stats look OK):



Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.86	2	0.2398	-0.001	-0.009	-0.053	-0.000	0.050	0.047
12	10.71	8	0.2188	0.001	-0.034	**0.122**	0.044	-0.039	-0.037
18	13.94	14	0.4541	-0.056	0.013	-0.031	0.048	-0.006	-0.042
24	16.47	20	0.6870	-0.023	-0.028	0.039	-0.049	0.020	-0.029
30	24.29	26	0.5593	0.006	0.050	-0.098	0.077	-0.002	0.039
36	35.09	32	0.3239	-0.029	-0.075	0.057	-0.001	**0.121**	-0.047
42	39.99	38	0.3817	0.002	-0.007	0.088	0.019	-0.004	0.060
48	43.35	44	0.4995	-0.043	0.043	-0.027	-0.047	-0.019	-0.032

Looking for “outliers” *that can be explained*

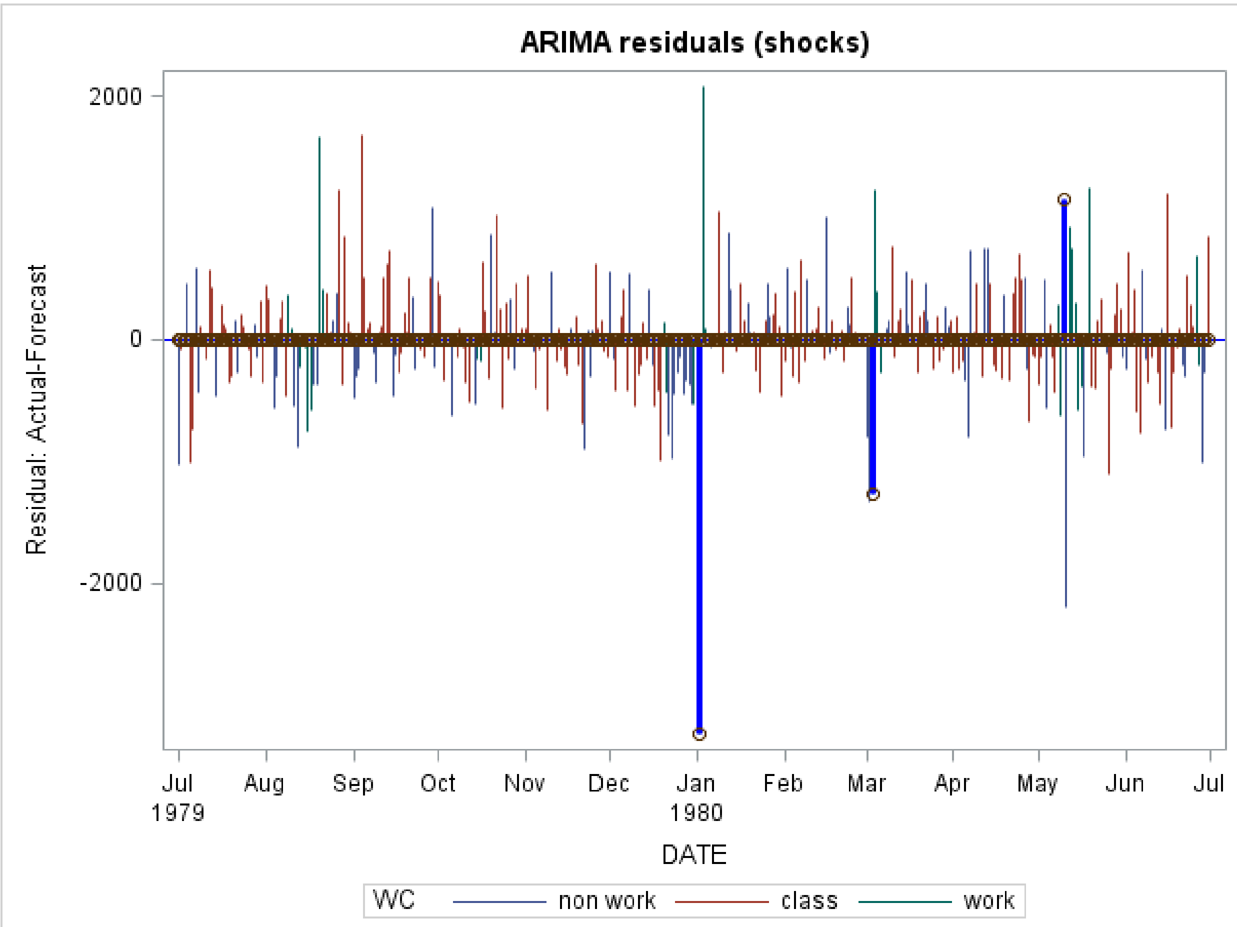
```
   * 0.05/365 = .0001369863 (Bonferroni) *;
outlier type=additive alpha=.0001369863 id=date;
format date weekdate.;
run;

/*****************************************************
January 2, 1980 Wednesday: Hangover Day :-)   .

March 3,1980 Monday:
On the afternoon and evening of March 2, 1980, North
Carolina experienced a major winter storm with heavy
snow across the entire state and near blizzard
conditions in the eastern part of the state. Widespread
snowfall totals of 12 to 18 inches were observed over
Eastern North Carolina, with localized amounts ranging
up to 22 inches at Morehead City and 25 inches at
Elizabeth City, with unofficial reports of up to 30
inches at Emerald Isle and Cherry Point (Figure 1).
This was one of the great snowstorms in Eastern North
Carolina history. What made this storm so remarkable
was the combination of snow, high winds, and very cold
temperatures.


May 10,1980  Saturday: Last day of Spring semester.
*****************************************************/;
```

Outlier Details

Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob>ChiSq
186	**Wednesday**	Additive	-3250.9	87.76	<.0001
315	**Saturday**	Additive	1798.1	28.19	<.0001
247	**Monday**	Additive	-1611.8	22.65	<.0001

Outlier Details

Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob>ChiSq
186	**02-JAN-1980**	Additive	-3250.9	87.76	<.0001
315	**10-MAY-1980**	Additive	1798.1	28.19	<.0001
247	**03-MAR-1980**	Additive	-1611.8	22.65	<.0001



Outliers: Jan 2 (hangover day!), March 3 (snowstorm), May 10 (graduation day).

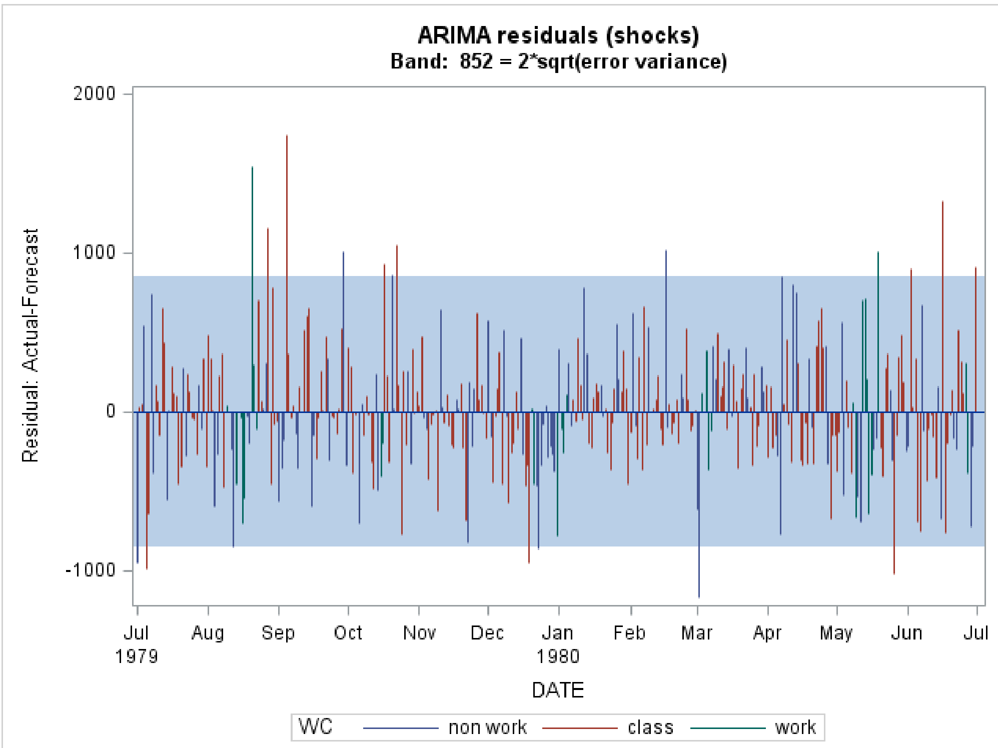
AR(1) → 'rebound' outlying residuals next day.

Add dummy variables for explainable outliers

```
data next; merge outarima energy; by date;
hangover      = (date="02Jan1980"d);
storm         = (date="03Mar1980"d);
graduation    = (date="10May1980"d);

Proc ARIMA data=next;
  identify var=demand crosscor=(temp tempsq class work
            s c hangover graduation storm) noprint;
  estimate input = (temp tempsq class work s c hangover
            graduation storm) p=1 q=(7,14) ml;
  forecast lead=0 out=outARIMA2 id=date  interval=day;
run;
```

Maximum Likelihood Estimation							
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**	**Variable**	**Shift**
MU	6127.4	259.43918	23.62	<.0001	0	DEMAND	0
MA1,1	-0.25704	0.05444	-4.72	<.0001	7	DEMAND	0
MA1,2	-0.10821	0.05420	-2.00	0.0459	14	DEMAND	0
AR1,1	0.76271	0.03535	21.57	<.0001	1	DEMAND	0
NUM1	27.89783	3.15904	8.83	<.0001	0	TEMP	0
NUM2	0.54698	0.10056	5.44	<.0001	0	TEMPSQ	0
NUM3	626.08113	104.48069	5.99	<.0001	0	CLASS	0
NUM4	3258.1	105.73971	30.81	<.0001	0	WORK	0
NUM5	-757.90108	181.28967	-4.18	<.0001	0	S	0
NUM6	-506.31892	184.50221	-2.74	0.0061	0	C	0
NUM7	-3473.8	334.16645	-10.40	<.0001	0	hangover	0
NUM8	2007.1	331.77424	6.05	<.0001	0	graduation	0
NUM9	-1702.8	333.79141	-5.10	<.0001	0	storm	0

Constant Estimate	1453.963
Variance Estimate	181450
Std Error Estimate	425.9695
AIC	5484.728
SBC	5535.462
Number of Residuals	366



Model looks fine.

AUTOREG - regression with AR(p) errors

ARIMA – regressors, differencing, ARMA(p,q) errors