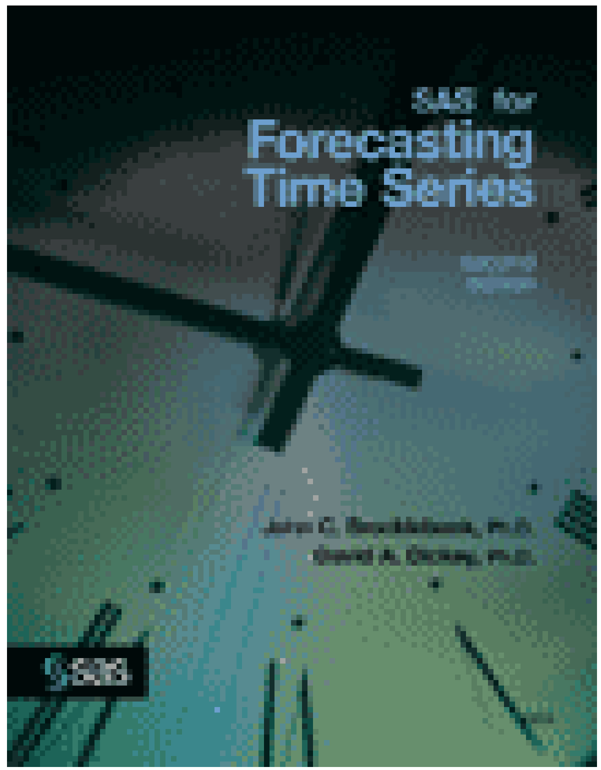
SAS for Forecasting Time Series –

Part 3: The General ARIMA Model - 2

Charlie Hallahan

January 20, 2010

SAS for Forecasting Time Series”, 2nd edition by Brocklebank & Dickey, 2003



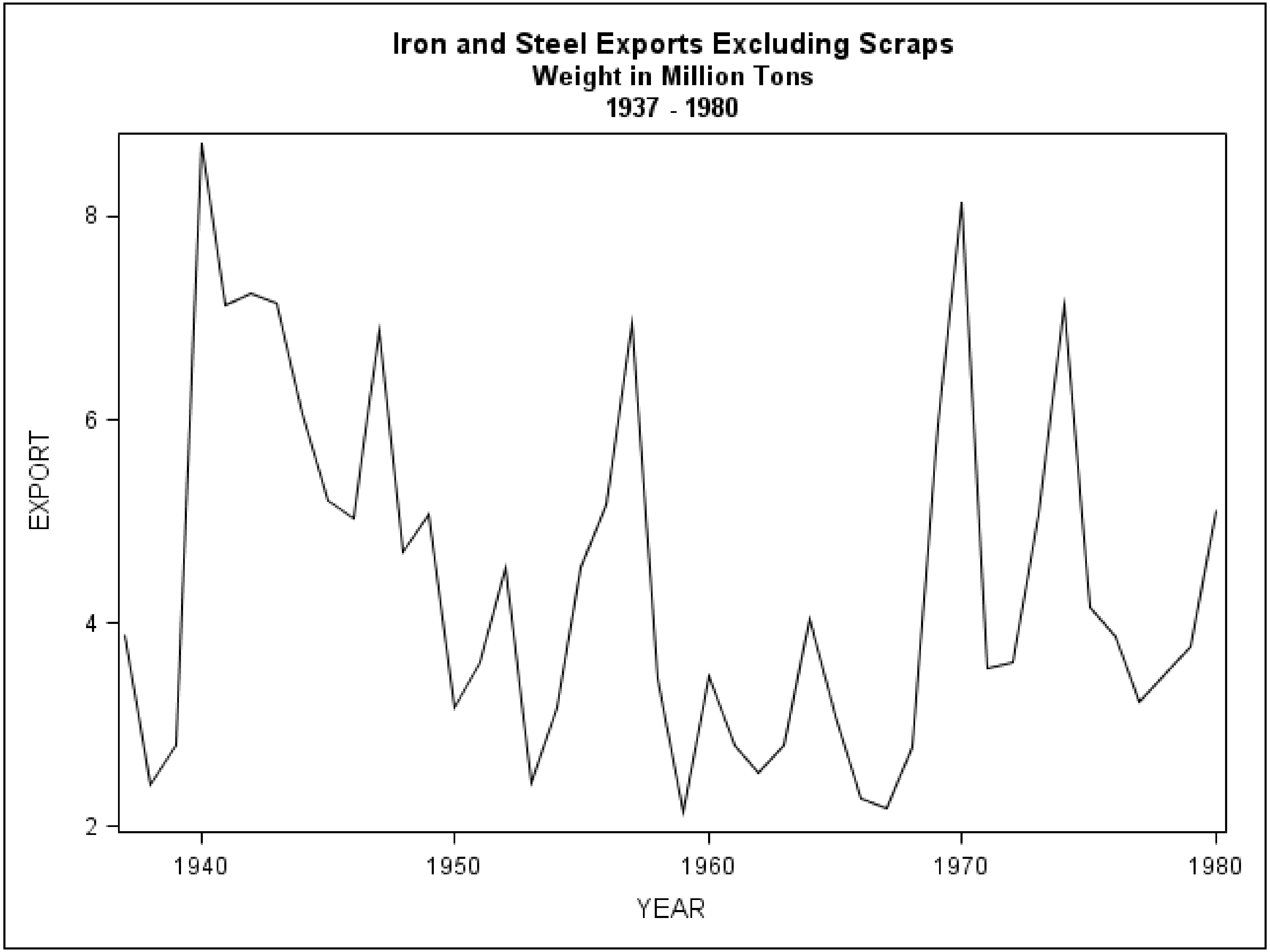
Example: Iron and Steel Export Analysis

```
DATA steel;
   input export  @@;
   retain year 1936;
   year+1;
   cards;
 3.89 2.41 2.8  8.72 7.12 7.24 7.15 6.05 5.21 5.03 6.88 4.7  5.06 3.16
 3.62 4.55 2.43 3.16 4.55 5.17 6.95 3.46 2.13 3.47 2.79 2.52 2.8  4.04
 3.08 2.28 2.17 2.78 5.94 8.14 3.55 3.61 5.06 7.13 4.15 3.86 3.22
 3.5  3.76 5.11
;
RUN;

title "Iron and Steel Exports Excluding Scraps";
title2 "Weight in Million Tons";
title3 "1937 - 1980";
proc sgplot data = steel;
            series x=year y=export;
run;

proc arima data=steel;
            identify var=export nlag=10;
run;
```

Example: Iron and Steel Export Analysis



Example: Iron and Steel Export Analysis

```
                      Name of Variable = EXPORT

                 Mean of Working Series    4.418182
                 Standard Deviation         1.73354
                 Number of Observations          44


                           Autocorrelations

Lag   Covariance   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1   Std Error

  0     3.005160       1.00000   |                    |********************|           0
  1     1.418238       0.47193   |              .     |*********           |    0.150756
  2     0.313839       0.10443   |             .      |**    .             |    0.181248
  3     0.133835       0.04453   |             .      |*     .             |    0.182611
  4     0.310097       0.10319   |             .      |**    .             |    0.182858
  5     0.296534       0.09867   |             .      |**    .             |    0.184176
  6     0.024517       0.00816   |             .      |      .             |    0.185374
  7    -0.159424       -.05305   |             .     *|      .             |    0.185382
  8    -0.299770       -.09975   |             .    **|      .             |    0.185727
  9    -0.247158       -.08224   |             .    **|      .             |    0.186940
 10    -0.256881       -.08548   |            .     **|       .            |    0.187761

                     "." marks two standard errors
```

Example: Iron and Steel Export Analysis

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        0.47193     |                .     |*********          |
  2       -0.15218     |                .  ***|     .             |
  3        0.07846     |                .     |**   .             |
  4        0.08185     |                .     |**   .             |
  5        0.01053     |                .     |     .             |
  6       -0.05594     |                .    *|     .             |
  7       -0.03333     |                .    *|     .             |
  8       -0.08310     |                .   **|     .             |
  9       -0.01156     |                .     |     .             |
 10       -0.05715     |                .    *|     .             |
```

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	**12.15**	**6**	**0.0586**	0.472	0.104	0.045	0.103	0.099	0.008

The **ACF** suggests a **MA(1)** process and the **PACF** an **AR(1)** process. The test for white noise just falls into the acceptance range, so we'll ignore it for the time being based on the other plots. We can fit both models and see which has the lowest MSE. The appropriate commands are **`estimate p=1;`** and **`estimate q=1;`**.

Example: Iron and Steel Export Analysis

MA(1) Model

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	**4.42102**	**0.34703**	**12.74**	**<.0001**	**0**
MA1,1	**-0.49827**	**0.13512**	**-3.69**	**0.0006**	**1**

Constant Estimate	4.421016
Variance Estimate	**2.412583**
Std Error Estimate	1.553249
AIC	165.5704
SBC	169.1388
Number of Residuals	44

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates

Parameter	MU	MA1,1
MU	1.000	-0.008
MA1,1	-0.008	1.000

Example: Iron and Steel Export Analysis

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	**Pr > ChiSq**	--------------------Autocorrelations--------------------					
6	1.31	5	**0.9336**	0.059	0.094	-0.028	0.085	0.075	-0.020
12	3.23	11	**0.9873**	-0.006	-0.079	-0.052	-0.013	-0.146	0.039
18	6.68	17	**0.9874**	0.063	-0.001	0.044	-0.092	0.096	-0.149
24	14.00	23	**0.9268**	-0.206	-0.135	-0.114	-0.084	0.014	-0.072

Model for variable EXPORT

Estimated Mean 4.421016

Moving Average Factors

Factor 1: 1 + 0.49827 B**(1)

The moving average parameter is highly significant and the residuals pass the white noise test out to 24 lags. The **MSE** is 2.41.

Example: Iron and Steel Export Analysis

AR(1) Model

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	**4.41217**	**0.43509**	**10.14**	**<.0001**	**0**
AR1,1	**0.47368**	**0.13622**	**3.48**	**0.0012**	**1**

Constant Estimate	2.322229
Variance Estimate	**2.444518**
Std Error Estimate	1.563495
AIC	166.149
SBC	169.7174
Number of Residuals	44

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates

Parameter	MU	AR1,1
MU	1.000	0.006
AR1,1	0.006	1.000

Example: Iron and Steel Export Analysis

AR(1) Model

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	2.19	5	0.8224	0.074	-0.151	-0.057	0.072	0.086	-0.020
12	4.32	11	0.9597	-0.020	-0.072	-0.018	-0.006	-0.165	0.046
18	7.29	17	0.9794	0.096	0.013	0.007	-0.061	0.130	-0.102
24	12.95	23	0.9530	-0.216	-0.094	-0.081	-0.039	0.042	-0.050

Model for variable EXPORT

Estimated Mean 4.412166

Autoregressive Factors

Factor 1: 1 - 0.47368 B**(1)

The autoregressive parameter is also highly significant and the residuals pass the white noise test out to 24 lags. The **MSE** is 2.44. Based on these results, the MA(1) model would be preferred (slightly) to the AR(1) model.

Estimation Methods Used in PROC ARIMA

PROC ARIMA supports three estimation techniques for **ARIMA(p,d,q)** models.

- **conditional least squares (CLS)** the default
- **unconditional least squares (ULS)**
- **maximum likelihood (ML)**

The **CLS** method minimizes the quantity: $\sum_{t=p+1}^{n} e_t^2$ where p is the order of the **AR** part of the process and e_t is a residual.

Estimation Methods Used in PROC ARIMA

The **METHOD = ULS** technique more accurately computes prediction error variances and finite sample predictions than **METHOD = CLS**.

METHOD = CLS assumes a constant variance and the same linear combination of past values as the optimum prediction.

With **METHOD = ML**, the function minimized is the negative log likelihood, assuming normality for the residuals.

While giving similar results for large samples, simulation studies show that **ML** is the most accurate.

The **Yule - Walker equations** are used as starting values for the parameters.

ESTIMATE Statement for Series 8

```
proc arima data=ffc2010.series;
        identify var = Y8 noprint;
        estimate p=1 q=1 printall grid;
        estimate p=2 q=2;
run;
```

The `printall` option shows all the iterations.

To check that a global, and not a local minimum, has been achieved, the `grid` option evaluates the sum of squares (or likelihood) on a grid surrounding the final parameter estimates.

ESTIMATE Statement for Series 8

Preliminary Estimation

Initial Autoregressive Estimates

	Estimate
1	0.51197

Initial Moving Average Estimates

	Estimate
1	-0.42193

Constant Term Estimate	**-0.28015**
White Noise Variance Est	1.10836

Conditional Least Squares Estimation

Iteration	SSE	MU	MA1,1	AR1,1	Constant	Lambda	R Crit
0	157.49	-0.57405	-0.42193	0.51197	-0.28015	0.00001	1
1	156.79	-0.56250	-0.42443	0.56575	-0.24427	1E-6	0.066581
2	156.79	-0.56301	-0.42325	0.56643	-0.24411	1E-7	0.001026
3	**156.79**	**-0.56285**	**-0.42358**	**0.56626**	**-0.24413**	**0.01**	**0.000345**

ESTIMATE Statement for Series 8

ARIMA Estimation Optimization Summary

Estimation Method	Conditional Least Squares
Parameters Estimated	3
Termination Criteria	Maximum Relative Change in Estimates
Iteration Stopping Value	0.001
Criteria Value	0.000777
Alternate Criteria	Relative Change in Objective Function
Alternate Criteria Value	8.225E-9
Maximum Absolute Value of Gradient	0.045945
R-Square Change from Last Iteration	0.000345
Objective Function	Sum of Squared Residuals
Objective Function Value	156.7896
Marquardt's Lambda Coefficient	0.01
Numerical Derivative Perturbation Delta	0.001
Iterations	3

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	**Approx Pr > |t|**	Lag
MU	**-0.56285**	**0.26840**	**-2.10**	**0.0377**	**0**
MA1,1	**-0.42358**	**0.09620**	**-4.40**	**<.0001**	**1**
AR1,1	**0.56626**	**0.08599**	**6.59**	**<.0001**	**1**

ESTIMATE Statement for Series 8

Constant Estimate	-0.24413
Variance Estimate	1.066596
Std Error Estimate	1.032761
AIC	**438.3219**
SBC	**447.3539**
Number of Residuals	150

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates

Parameter	MU	MA1,1	AR1,1
MU	1.000	-0.002	0.007
MA1,1	-0.002	1.000	0.606
AR1,1	0.007	0.606	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	2.63	4	0.6220	0.026	0.015	-0.095	-0.004	-0.079	0.026
12	4.92	10	0.8962	0.046	0.102	0.033	-0.022	0.012	-0.004
18	6.22	16	0.9855	-0.046	-0.009	0.037	0.006	0.058	0.027
24	14.40	22	0.8867	-0.071	0.014	-0.197	0.018	0.041	0.013
30	17.45	28	0.9392	0.068	-0.006	-0.007	0.016	-0.093	-0.052

ESTIMATE Statement for Series 8

SSE Surface on Grid Near
Estimates: MA1,1 (Y8)

MU (Y8)	-0.42858	**-0.42358**	-0.41858
-0.56785	156.80	156.79	156.80
-0.56285	156.80	**156.79**	156.79
-0.55785	156.80	156.79	156.80

SSE Surface on Grid Near
Estimates: AR1,1 (Y8)

MU (Y8)	0.56126	**0.56626**	0.57126
-0.56785	156.80	156.79	156.80
-0.56285	156.80	**156.79**	156.80
-0.55785	156.80	156.79	156.80

SSE Surface on Grid Near
Estimates: AR1,1 (Y8)

MA1,1 (Y8)	0.56126	**0.56626**	0.57126
-0.42858	156.79	156.80	156.81
-0.42358	156.80	156.79	156.80
-0.41858	156.81	156.79	156.79

ESTIMATE Statement for Series 8

ARMA(2,2) MODEL

Estimated Mean -0.56285

Autoregressive Factors

Factor 1: 1 - 0.56626 B(1)**

Moving Average Factors

Factor 1: 1 + 0.42358 B(1)**

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.54923	0.25741	-2.13	0.0345	0
MA1,1	-1.06041	0.46917	-2.26	0.0253	1
MA1,2	-0.36641	0.18939	-1.93	0.0550	2
AR1,1	**-0.04251**	**0.48048**	**-0.09**	0.9296	1
AR1,2	0.26889	0.31413	0.86	0.3934	2

Note: by over fitting the model, the AR(1) parameter is not significant. Both the AIC and SBC criteria favor the ARMA(1,1) model (see p. 16)

Constant Estimate	-0.42489
Variance Estimate	1.067367
Std Error Estimate	1.033134
AIC	**440.3755**
SBC	**455.4287**
Number of Residuals	150

* AIC and SBC do not include log determinant.

ESTIMATE Statement for Series 8

Correlations of Parameter Estimates

Parameter	MU	MA1,1	MA1,2	AR1,1	AR1,2
MU	1.000	-0.003	-0.001	0.001	0.005
MA1,1	-0.003	1.000	0.920	0.985	-0.928
MA1,2	-0.001	0.920	1.000	0.890	-0.749
AR1,1	0.001	0.985	0.890	1.000	-0.923
AR1,2	0.005	-0.928	-0.749	-0.923	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	0.45	2	0.7980	0.003	0.008	-0.009	-0.040	-0.032	0.009
12	2.71	8	0.9515	0.051	0.095	0.036	-0.026	0.020	-0.000
18	4.40	14	0.9925	-0.046	-0.007	0.045	-0.017	0.074	0.012
24	13.68	20	0.8466	-0.074	0.034	-0.206	0.024	0.049	-0.016
30	17.32	26	0.8991	0.088	-0.023	-0.005	0.024	-0.094	-0.044

Estimated Mean -0.54923

Autoregressive Factors

Factor 1: 1 + 0.04251 B**(1) - 0.26889 B**(2)

Moving Average Factors

Factor 1: 1 + 1.06041 B**(1) + 0.36641 B**(2)

Nonstationary Series

The theory behind **PROC ARIMA** requires that a series be **stationary**.

Given an **ARMA**(p,q) model:

$$\left(1-\alpha_1 B-\alpha_2 B^2-\ldots-\alpha_p B^p\right)\left(Y_t-\mu\right)=\left(1-\beta_1 B-\beta_2 B^2-\ldots-\beta_q B^q\right)e_t$$

$\left(1-\alpha_1 M-\alpha_2 M^2-\ldots-\alpha_p M^p\right)=0$ is called the **characteristic equation** of Y_t.

Y_t is stationary if the roots of the characteristic equation all fall outside the unit circle.

Note: for $p=1$, $1-\alpha_1 M=0 \Rightarrow \ |M|=|1/\alpha_1|>1 \Leftrightarrow |\alpha_1|<1$.

Nonstationary Series

For example, the series $\left(1-1.5B+.64B^2\right)\left(Y_t-\mu\right)=\left(1+.8B\right)e_t$ is stationary, but $\left(1-1.5B+.5B^2\right)\left(Y_t-\mu\right)=\left(1+.8B\right)e_t$ is nonstationary.

$\left(1-1.5M+.5M^2\right)=\left(1-M\right)\left(1-.5M\right) \Rightarrow$ 1 is a root, i.e., a **unit root**. The other root is 2.

Note that $\left(1-1.5B+.5B^2\right)\left(Y_t-\mu\right)=Y_t-1.5Y_{t-1}+.5Y_{t-2}-\left(1-1.5+.5\right)\mu$ and the mean μ drops out. The implication is that forecasts of Y_t won't eventually converge to the series mean μ.

On the other hand, forecasts of stationary series do eventually converge to the series mean μ.

Nonstationary Series

If we define $W_t = Y_t - Y_{t-1} = (1-B)Y_t = \Delta Y_t,$ then the nonstationary model becomes $(1-.5B)W_t = e_t + .8e_{t-1},$ a stationary series.

If the first difference of a nonstationary series is stationary, the series is said to be **integrated of order 1** or $I(1)$.

A very slowly decaying **ACF** is a clue that a series may be nonstationary. Formal unit root tests, such as the **Dickey - Fuller test**, are discussed later in the Chapter.

In **PROC ARIMA**, first differences are indicated in the **IDENTIFY** statement as follows: IDENTIFY var = Y(1). **Second differences**, although rarely required, are obtained as follows: IDENTIFY var = Y(1,1).

Effects of Differencing on Forecasts

While a model requiring differencing to become stationary is estimated in its differenced form, predictions are always given for the series in levels.

To see how this works, consider the the nonstationary model previously considerd:
$Y_t = 1.5Y_{t-1} + .5Y_{t-2} + e_t$. We saw that if $W_t = Y_t - Y_{t-1}$, then $W_t = .5W_{t-1} + e_t$.

Given data $Y_1, \ldots, Y_n$, then $Y_{n+1} = Y_n + Y_{n+1} - Y_n = Y_n + W_{n+1} = Y_n + .5W_n + e_{n+1}$.
We can then define $\hat{Y}_{n+1} = Y_n + .5W_n$ (after setting all future error terms to zero).

Since $Y_{n+2} = Y_n + Y_{n+1} - Y_n + Y_{n+2} - Y_{n+1} = Y_n + W_{n+1} + W_{n+2}$

$$= Y_n + .5W_n + e_{n+1} + .5\left(.5W_n + e_{n+1}\right) + e_{n+2} = Y_n + .5W_n + (.5)^2 W_n + 1.5e_{n+1} + e_{n+2}$$

Thus, $\hat{Y}_{n+2} = Y_n + .5W_n + (.5)^2 W_n$. In general, $\hat{Y}_{n+k} = Y_n + \sum_{i=1}^{k} .5^i W_n$ which converges to

$Y_n + (1)W_n$ since $\sum_{i=1}^{\infty} .5^i = 1$.

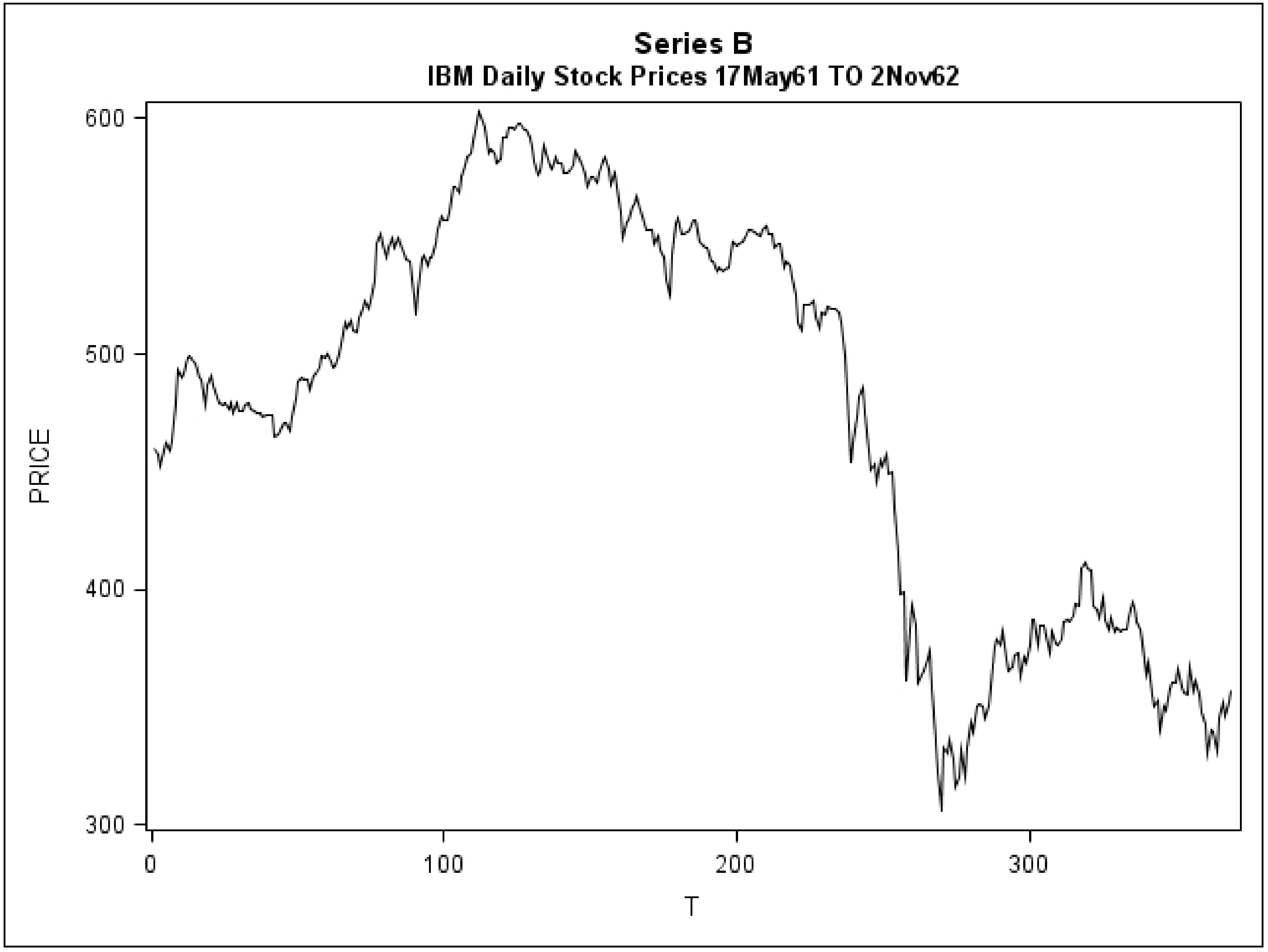
Effects of Differencing on Forecasts

We can also easily calculate the **forecast errors** and their **variances**.

For example, $Y_{n+1} - \hat{Y}_{n+1} = e_{n+1}$ with variance σ^2.

$Y_{n+2} - \hat{Y}_{n+2} = 1.5e_{n+1} + e_{n+2}$ with variance $3.25\sigma^2$.

Examples: Forecasting IBM Series and Silver Series



Examples: Forecasting IBM Series and Silver Series

Identifying the IBM price series.

```
proc arima data= ffc2010.ibm;
        identify var = price center nlag=15;
        identify var = price(1) nlag=15;
run;
```

Examples: Forecasting IBM Series and Silver Series

Very strong evidence of a nontationary series

```
                      Name of Variable = PRICE

              Mean of Working Series              0
              Standard Deviation           84.10504
              Number of Observations            369

                          Autocorrelations

Lag  Covariance  Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1    Std Error

  0    7073.658      1.00000    |                    |********************|            0
  1    7026.966      0.99340    |                  . |********************|     0.052058
  2    6973.914      0.98590    |                .   |********************|     0.089771
  3    6918.629      0.97808    |               .    |********************|     0.115443
  4    6868.433      0.97099    |               .    |******************* |     0.136059
  5    6817.810      0.96383    |              .     |******************* |     0.153695
  6    6763.587      0.95617    |             .      |******************* |     0.169285
  7    6705.771      0.94799    |             .      |******************* |     0.183337
  8    6645.401      0.93946    |            .       |******************* |     0.196172
  9    6580.448      0.93028    |            .       |******************* |     0.208008
 10    6522.985      0.92215    |           .        |******************  |     0.218993
 11    6466.010      0.91410    |           .        |******************  |     0.229274
 12    6407.497      0.90583    |          .         |******************  |     0.238947
 13    6348.092      0.89743    |          .         |******************  |     0.248078
 14    6289.664      0.88917    |          .         |******************  |     0.256726
 15    6230.941      0.88087    |         .          |******************  |     0.264940

                    "." marks two standard errors
```

Examples: Forecasting IBM Series and Silver Series

ACF looks much better for the 1st difference

```
                     Name of Variable = PRICE

Period(s) of Differencing                                  1
Mean of Working Series                              -0.27989
Standard Deviation                                  7.248345
Number of Observations                                   368
Observation(s) eliminated by differencing                  1

                          Autocorrelations

Lag   Covariance   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1     Std Error

  0    52.538509       1.00000   |                    |********************|            0
  1     4.496014       0.08558   |                  . |**                  |     0.052129
  2    -0.072894       -.00139   |                  . | .                  |     0.052509
  3    -2.853759       -.05432   |                  .*| .                  |     0.052509
  4    -1.820817       -.03466   |                  .*| .                  |     0.052662
  5    -1.261461       -.02401   |                  . | .                  |     0.052723
  6     6.350064       0.12086   |                  . |**                  |     0.052753
  7     3.585725       0.06825   |                  . |*.                  |     0.053500
  8     1.871606       0.03562   |                  . |*.                  |     0.053736
  9    -3.483286       -.06630   |                  .*| .                  |     0.053801
 10     1.149218       0.02187   |                  . | .                  |     0.054022
 11     4.043788       0.07697   |                  . |**                  |     0.054046
 12     2.816399       0.05361   |                  . |*.                  |     0.054343
 13    -2.508704       -.04775   |                  .*| .                  |     0.054487
 14     3.445101       0.06557   |                  . |*.                  |     0.054600
 15    -3.470001       -.06605   |                  .*| .                  |     0.054814

                    "." marks two standard errors
```

Examples: Forecasting IBM Series and Silver Series

In fact, the 1st difference appears to be white noise or MA(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        0.08558    |                   . |**                |
  2       -0.00877    |                   . | .                |
  3       -0.05385    |                   .*| .                |
  4       -0.02565    |                   .*| .                |
  5       -0.01940    |                   . | .                |
  6        0.12291    |                   . |**                |
  7        0.04555    |                   . |*.                |
  8        0.02375    |                   . | .                |
  9       -0.06241    |                   .*| .                |
 10        0.04501    |                   . |*.                |
 11        0.08667    |                   . |**                |
 12        0.02638    |                   . |*.                |
 13       -0.07034    |                   .*| .                |
 14        0.07191    |                   . |*.                |
 15       -0.05660    |                   .*| .                |
```

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	**9.98**	**6**	**0.1256**	0.086	-0.001	-0.054	-0.035	-0.024	0.121
12	**17.42**	**12**	**0.1344**	0.068	0.036	-0.066	0.022	0.077	0.054

Examples: Forecasting IBM Series and Silver Series

```
proc arima data= ffc2010.ibm;
        identify var = price(1) noprint;
        estimate q=1 noconstant;
run;
```

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	**-0.08658**	**0.05203**	**-1.66**	**0.0970**	**1**

Variance Estimate	52.36132
Std Error Estimate	7.236112
AIC	2501.943
SBC	2505.851
Number of Residuals	368

* AIC and SBC do not include log determinant.

Examples: Forecasting IBM Series and Silver Series

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.99	5	0.2217	0.001	0.005	-0.051	-0.026	-0.030	0.120
12	13.94	11	0.2365	0.056	0.039	-0.070	0.024	0.072	0.054
18	**31.04**	**17**	**0.0198**	-0.057	0.079	-0.081	**0.118**	**0.113**	0.040
24	39.05	23	0.0196	0.041	0.072	-0.089	-0.027	0.066	0.025
30	49.83	29	0.0094	0.028	-0.100	-0.055	0.051	0.028	0.099
36	56.47	35	0.0122	0.072	-0.074	-0.063	-0.007	0.022	0.035
42	64.42	41	0.0112	0.066	-0.085	0.059	-0.060	0.018	0.017
48	76.33	47	0.0044	-0.116	-0.037	0.073	0.005	0.069	0.057

The large autocorrelations at lags 16 and 17 are ignored since they don't occur at seasonal lags, and the more parsimonious MA(1) model is kept.

Examples: Forecasting IBM Series and Silver Series

```
proc arima data= ffc2010.ibm;
        identify var = price(1) noprint;
        estimate q=1 noconstant noprint;
        forecast lead = 15;
run;
```

```
 Model for variable PRICE
Period(s) of Differencing    1
No mean term in this model.
Moving Average Factors
Factor 1:  1 + 0.08658 B**(1)
Forecasts for variable PRICE
```

Obs	Forecast	Std Error	95% Confidence Limits	
370	357.3837	7.2361	343.2012	371.5662
371	357.3837	10.6856	336.4403	378.3270
372	357.3837	13.2666	331.3817	383.3857
373	357.3837	15.4215	327.1581	387.6093
374	357.3837	17.3102	323.4563	391.3110
375	357.3837	19.0122	320.1205	394.6469
376	357.3837	20.5738	317.0597	397.7077
377	357.3837	22.0251	314.2154	400.5520
378	357.3837	23.3864	311.5472	403.2202
379	357.3837	24.6727	309.0260	405.7414
380	357.3837	25.8953	306.6299	408.1375
381	357.3837	27.0626	304.3420	410.4254
382	357.3837	28.1816	302.1487	412.6187
383	357.3837	29.2579	300.0392	414.7282
384	357.3837	30.2960	298.0047	416.7627

Note that all the forecasts are the same.
Why this is is explained on the next page.

Examples: Forecasting IBM Series and Silver Series

Note: if $Y_t - Y_{t-1} = e_t - \beta e_{t\text{-}1}$ as in the IBM *ARIMA*(0,1,1) model, then we can write $e_t = \left(Y_t - Y_{t-1}\right) + \beta\left(Y_{t-1} - Y_{t-2}\right) + \beta^2\left(Y_{t-2} - Y_{t-3} + \cdots\right)$ by repeated back substitution or

$$Y_t = e_t + \left(1-\beta\right)\left(Y_{t-1} + \beta Y_{t-2} + \beta^2 Y_{t-3} + \cdots\right) \text{ or } \hat{Y}_t = e_t + \left(1-\beta\right)\left(Y_{t-1} + \beta Y_{t-2} + \beta^2 Y_{t-3} + \cdots\right)$$

This forecasting formula involves an infinite sum, but any finite approximation would seem to produce forecasts that changed as the lead time increases. So why are the forecasts on the previous page all the same after one-step ahead?

The reason is that PROC ARIMA estimates the residuals, $\hat{e}_t$, and uses the relationship $Y_t = Y_{t-1} + e_t - \beta e_{t\text{-}1}$. Using the estimate $\hat{e}_t$ for e_t and setting all future residuals equal to their mean value of zero, we get $\hat{Y}_t = Y_{t-1} - \hat{\beta}\hat{e}_{t\text{-}1}$, $\hat{Y}_{t+1} = \hat{Y}_t$, $\hat{Y}_{t+2} = \hat{Y}_t$, etc.

Models for Nonstationary Data

There are formal tests for unit root nonstationarity. In particular, the **Dickey - Fuller** tests (1979, 1981) will be discussed.

We'll illustrate the process with an **AR(2)** model: $Y_t - \mu = \alpha_1\left(Y_{t-1} - \mu\right) + \alpha_2\left(Y_{t-2} - \mu\right) + e_t$.
This can be written as $\left(Y_t - \mu\right) - \alpha_1\left(Y_{t-1} - \mu\right) - \alpha_2\left(Y_{t-2} - \mu\right) = e_t$
or $(1 - \alpha_1 B - \alpha_2 B^2)\left(Y_t - \mu\right) = e_t$.

With a little algebra, this can be rewritten as:

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

or

$$\Delta Y_t = \rho\left(Y_{t-1} - \mu\right) - \alpha_2 \Delta Y_{t-1} + e_t$$

Stationarity depends on the roots of the **characteristic equation**: $1 - \alpha_1 M - \alpha_2 M^2 = 0$

Models for Nonstationary Data

Y_t is **nonstationary** if 1 is a root of the characteristic equation.

But $1-\alpha_1 M - \alpha_2 M^2 = 0$ for $M = 1 \Rightarrow 1-\alpha_1-\alpha_2 = 0 \Rightarrow \rho = 0.$

Thus a test for a unit root is to estimate the model $\Delta Y_t = \rho\left(Y_{t-1}-\mu\right)+\beta\Delta Y_{t-1}+e_t$ and test $H_0 : \rho = 0$ against the alternative $H_1 : \rho < 0.$

Since the usual test statistic for this test does not have a t-distribution in this instance, **Dickey** and **Fuller** (1979, 1981) derived the distribution of this "t-test" and provided tables of critical values.

The test is called the **Dickey - Fuller test**. Since the assumption that e_t is white noise may not hold for an AR(1) model, additional lagged differences may be needed to satisfy this assumption leading to the **Augmented Dickey - Fuller** or **ADF** test.

Models for Nonstationary Data

There are three possible assumptions to make concerning Y_t.

1. Y_t has zero mean, i.e., $\mu = 0$
2. Y_t has a non-zero mean μ
3. Y_t has a non-zero mean μ and deterministic trend βt.

Each assumption leads to a different test statistic and critical values when testing for a unit root, i.e., whether $\rho = 0$ or $\rho < 0$.

The regressions involved in the unit root tests for these three situations are given on the next page.

Models for Nonstationary Data

Regress ΔY_t on these:	AR(1) in regression form
$Y_{t-1,}\Delta Y_{t-1},\ldots,\Delta Y_{t-k}$	$Y_t = \rho Y_{t-1} + e_t$
$Y_{t-1,}1,\Delta Y_{t-1},\ldots,\Delta Y_{t-k}$	$Y_t - \mu = \rho\left(Y_{t-1} - \mu\right) + e_t$
$Y_{t-1,}1,t,\Delta Y_{t-1},\ldots,\Delta Y_{t-k}$	$Y_t - \alpha - \beta t = \rho\left(Y_{t-1} - \alpha - \beta(t-1)\right) + e_t$

AR(1) in deviations form	$H_0 : \rho = 1$
$\Delta Y_t = \left(\rho - 1\right)Y_{t-1} + e_t$	$\Delta Y_t = e_t$
$\Delta Y_t = \left(\rho - 1\right)\mu + \left(\rho - 1\right)Y_{t-1} + e_t$	$\Delta Y_t = e_t$
$\Delta Y_t = \left(\rho - 1\right)(\alpha - \beta t) + \beta + \left(\rho - 1\right)Y_{t-1} + e_t$	$\Delta Y_t = \beta + e_t$

Models for Nonstationary Data

Note that in the first two cases, under the null hypothesis of $\rho = 1$, Y_t is a random walk while the third case says that Y_t is a random walk with drift β.

Only fit the first model if you're sure that Y has mean 0.

Use the third model if Y_t appears to have a deterministic trend t. In this case, under the alternative, Y_t is stationary around a deterministic trend.

If $|\rho| < 1$, then a forecast of Y_{n+k} would be $\hat{Y}_{n+k} = \alpha + \beta(n+k) + \rho^k (Y_n - \alpha - \beta k)$ with forecast error variance $\left(1 + \rho^2 + ... + \rho^{2k-2}\right)\sigma^2 \to \sigma^2 / (1 - \rho^2)$ as $k \to \infty$.

However, if $\rho = 1$, then the forecast error variance becomes infinite.

Models for Nonstationary Data

Note: while the parameter estimates for $Y_{t\text{-}1}$, 1, and t all have nonstandard distributions, the parameter estimates for all the lagged differences in the models have a limiting distribution that is normal.

Thus, standard t-tests can be used on individual coefficients and F-tests on sets of lagged differences.

Also, the distribution for the parameter estimate of $Y_{t\text{-}1}$, namely $\hat{\rho}-1$, is different for each of the three cases considered above.

On the other hand, the limiting distribution of $\hat{\rho}-1$ remains the same no matter how many lagged differences are included.

Models for Nonstationary Data

To illustrate how the wrong conclusion could be reached if the standard *t*-distribution is used for $\rho - 1$, we'll use data for the stocks of silver on the New York Commodities Exchanged that were analyzed in Chapter 2.

DEL denotes the difference, **DELi** for the i^{th} lagged difference, and **LSILVER** for the lagged level of **SILVER**.

Twelve years of monthly data begins in January, 1976.

The variable **PART** is used to distinguish observations through March, 1980.

Models for Nonstationary Data

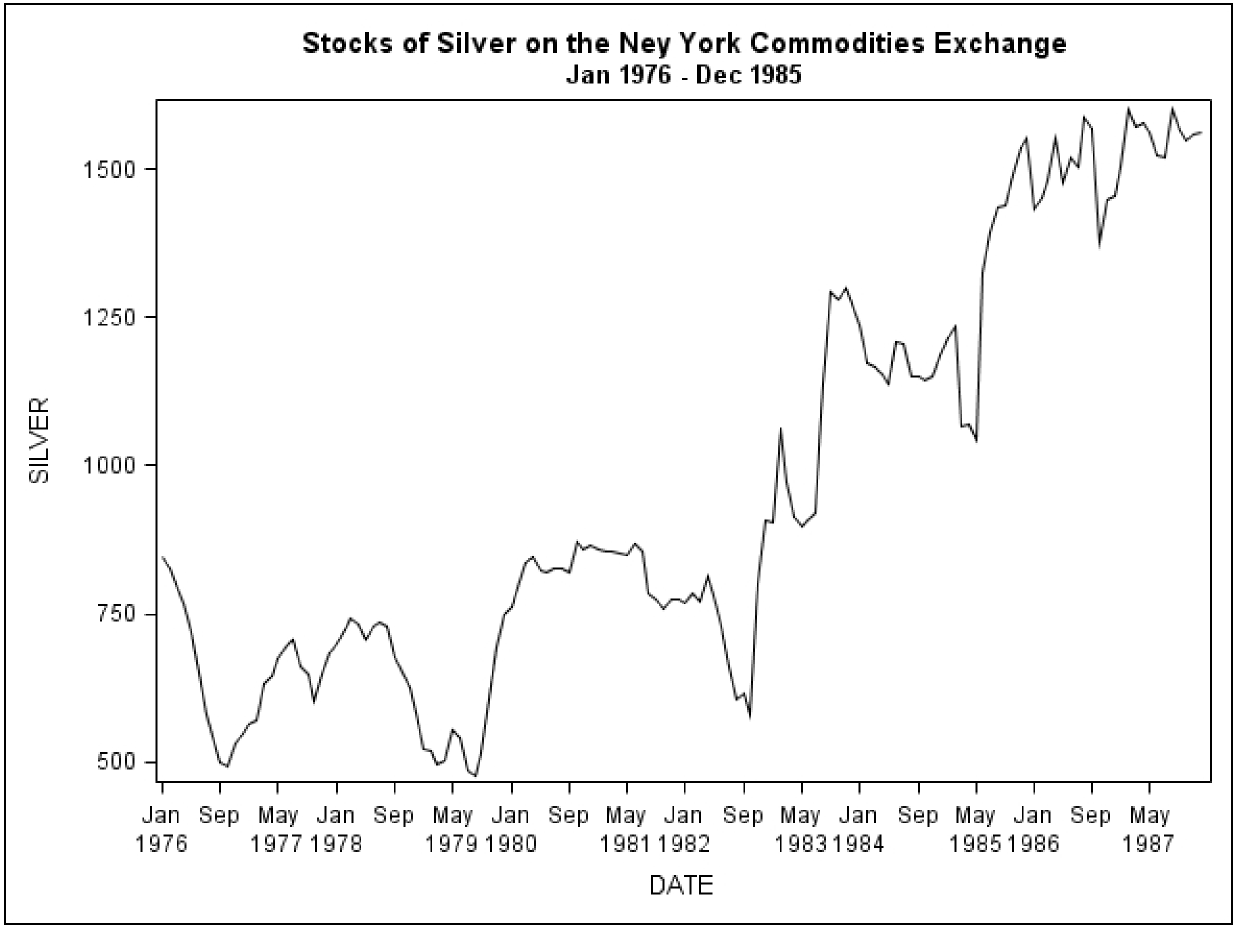
```
DATA ffc2010.SILVER;
   INPUT SILVER @@; DEL = SILVER-LAG(SILVER);
   TITLE 'MONTH END STOCKS OF SILVER';
   RETAIN DATE '01DEC75'D;
   DATE=INTNX('MONTH',DATE,1);
   FORMAT DATE MONYY.;
   PART=1;
   IF DATE > '01APR80'D THEN PART=2;
   IF PART=1 THEN SILV=SILVER;
   ELSE SILV=.;
   OUTPUT;
   RETAIN;
   DEL4=DEL3;
   DEL3=DEL2;
   DEL2=DEL1;
   DEL1=DEL;
   LSILVER=SILVER;
  *  SAS dataset up to 845.  Some errors found and
     corrected                  ;
   CARDS;
 846  827  799  768  719  652  580  546  500  493  530  548
 565  572  632  645  674  693  706  661  648  604  647  684
 700  723  741  734  708  728  737  729  678  651  627  582
 :
1502 1600 1573 1577 1561 1522 1521 1601 1564 1548 1558 1563
;
RUN;
```

Models for Nonstationary Data

First 10 observations for the SILVER dataset

DATE	SILVER	DEL	PART	SILV	DEL4	DEL3	DEL2	DEL1	LSILVER
JAN76	846	.	1	846	.	.	.	.	.
FEB76	827	-19	1	827	.	.	.	.	846
MAR76	799	-28	1	799	.	.	.	-19	827
APR76	768	-31	1	768	.	.	-19	-28	799
MAY76	719	-49	1	719	.	-19	-28	-31	768
JUN76	652	-67	1	652	-19	-28	-31	-49	719
JUL76	580	-72	1	580	-28	-31	-49	-67	652
AUG76	546	-34	1	546	-31	-49	-67	-72	580
SEP76	500	-46	1	500	-49	-67	-72	-34	546
OCT76	493	-7	1	493	-67	-72	-34	-46	500

Models for Nonstationary Data



Note that we'll only use data up to March, 1980 – before the obvious upward trend begins.

Models for Nonstationary Data

```
title "F-test on lagged differences";
proc reg data=ffc2010.silver;
        model del =     lsilver del1 del2 del3 del4 / noprint;
        test del2=0, del3=0, del4=0;
        where part = 1;
run;
```

Test 1 Results for Dependent Variable DEL

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	1152.19711	**1.32**	**0.2803**
Denominator	41	871.51780		

Conclusion: we can drop the extraneous lagged differences.

Models for Nonstationary Data

We now do the incorrect thing and use a standard *t-test* to test for a unit root.

Dependent Variable: DEL

Number of Observations Read	52
Number of Observations Used	50
Number of Observations with Missing Values	2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	34685	17342	20.66	<.0001
Error	47	39451	839.37389		
Corrected Total	49	74136			

Root MSE	28.97195	R-Square	0.4679
Dependent Mean	0.36000	Adj R-Sq	0.4452
Coeff Var	8047.76393		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	75.58073	27.36395	2.76	0.0082
LSILVER	**1**	**-0.11703**	**0.04216**	**-2.78**	**0.0079**
DEL1	1	0.67115	0.10806	6.21	<.0001

Models for Nonstationary Data

Using a standard *t*-test, we would reject the hypothesis that the coefficient of LSILVER is zero, i.e., that ρ = 1, and conclude that SILVER is stationary.

The correct 95% critical value, as determined by Dickey and Fuller, is -2.86.

proc arima data = ffc2010.silver;
identify var = silver stationarity = (adf = (**1**)) ;
run;

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	1	-0.2461	0.6232	-0.28	0.5800		
Single Mean	**1**	-17.7945	0.0121	**-2.78**	**0.0689**	3.86	0.1197
Trend	1	-15.1102	0.1383	-2.63	0.2697	4.29	0.3484

The relevant test here, using 1 lag as previously determined and assuming a non-zero mean (since all values of sliver are above 400), is for Tau = 0 (i.e., $\tau = \rho - 1$).
Note that Tau = -2.78, the same value as with **PROC REG**, however, the correct critical value yields a *p*-value of 0.0689, leading to a conclusion of **non-stationarity**.