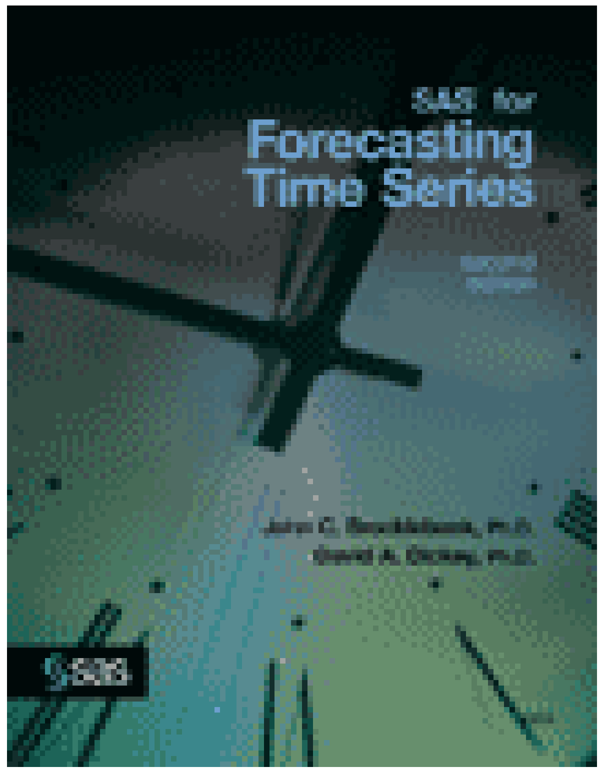
SAS for Forecasting Time Series –

Chapter 3: The General ARIMA Model - 3

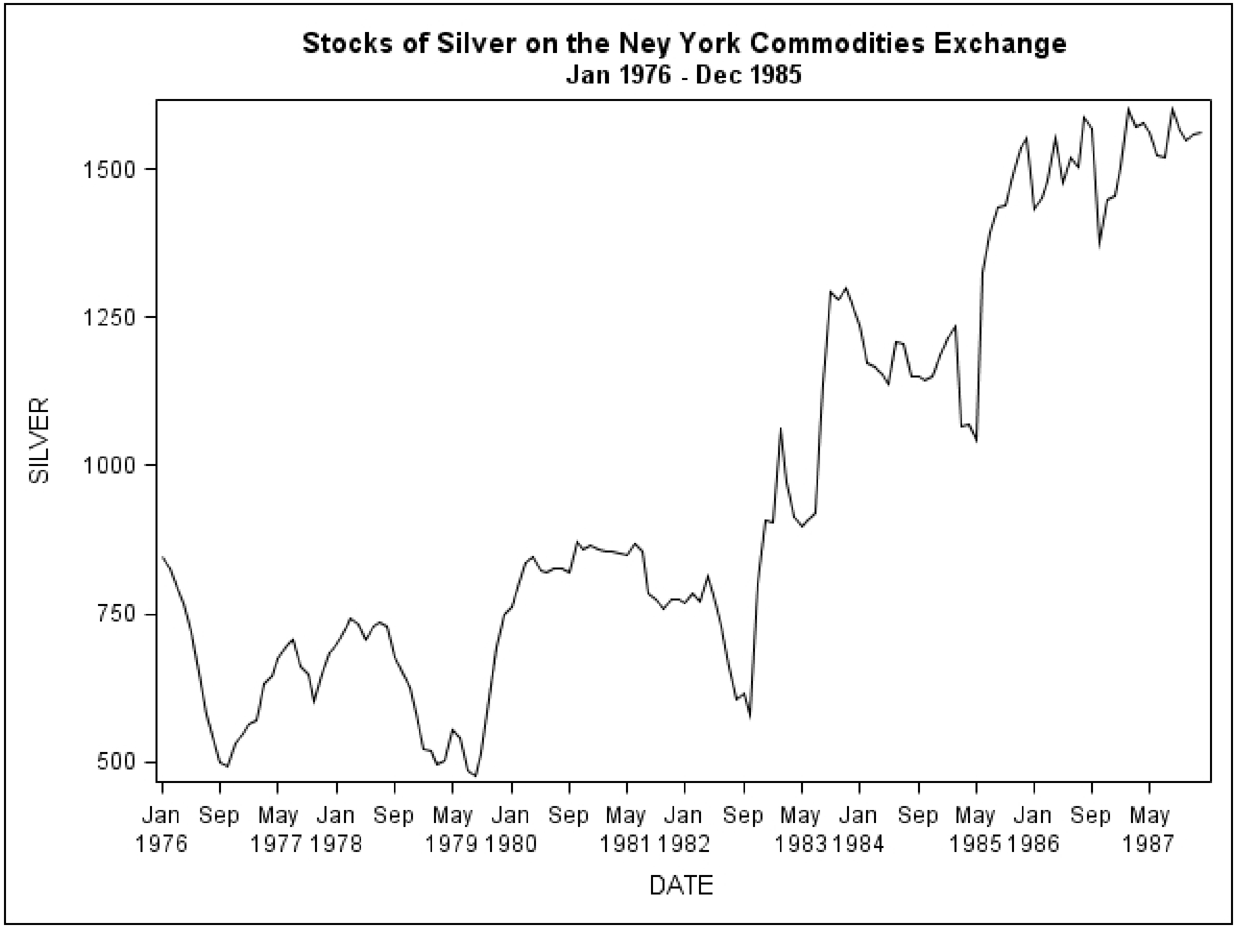
Charlie Hallahan

February 24, 2010

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



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

Recalling the graph on the previous page and the strong upward trending after the period used for testing for a unit root, the false conclusion of stationarity would lead to unduly optimistic forecast error bounds.

Treating the series as nonstationary leads to much more realistic forecast error bounds.

To illustrate the effects of trends, we'll now look at the logarithm of the closing price of Amazon.com stock.

As seen in the following graph, the series is tightly clustered around a linear trend.

Models for Nonstationary Data

Models for Nonstationary Data

Here is the **ACF** for the levels of **CLOSE**.

```
                                          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      1.507838        1.00000    |                    |********************|             0
  1      1.499991        0.99480    |                  . |********************|      0.044324
  2      1.491712        0.98931    |                 .  |********************|      0.076506
  3      1.482859        0.98343    |                .   |********************|      0.098482
  4      1.473232        0.97705    |               .    |********************|      0.116185
  5      1.463401        0.97053    |               .    |******************* |      0.131339
  6      1.453959        0.96427    |              .     |******************* |      0.144745
  7      1.444492        0.95799    |              .     |******************* |      0.156858
  8      1.435207        0.95183    |             .      |******************* |      0.167960
  9      1.425840        0.94562    |             .      |******************* |      0.178243
 10      1.416339        0.93932    |            .       |******************* |      0.187840
 11      1.406670        0.93291    |            .       |******************* |      0.196852
 12      1.396887        0.92642    |            .       |******************* |      0.205355
 13      1.387059        0.91990    |           .        |******************  |      0.213408
 14      1.377306        0.91343    |           .        |******************  |      0.221061
 15      1.367502        0.90693    |           .        |******************  |      0.228356
 16      1.357946        0.90059    |           .        |******************  |      0.235326
 17      1.347846        0.89389    |          .         |******************  |      0.242002
 18      1.337100        0.88677    |          .         |******************  |      0.248404
```


The very slow decline certainly indicates nonstationarity.

Models for Nonstationary Data

Here is the **ACF** for the first differences of **CLOSE**.

```
                                     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     0.0037493        1.00000    |                    |********************|              0
  1    0.00017619        0.04699    |                  . |*.                  |       0.044368
  2    0.00007870        0.02099    |                  . | .                  |       0.044466
  3    0.00017190        0.04585    |                  . |*.                  |       0.044485
  4    -0.0001363        -.03635    |                  .*| .                  |       0.044578
  5    -0.0000161        -.00431    |                  . | .                  |       0.044636
  6    0.00005285        0.01409    |                  . | .                  |       0.044637
  7    -0.0002315        -.06176    |                  .*| .                  |       0.044646
  8    -0.0001189        -.03171    |                  .*| .                  |       0.044814
  9    -0.0000900        -.02400    |                  . | .                  |       0.044858
 10    0.00002114        0.00564    |                  . | .                  |       0.044883
 11    0.00001634        0.00436    |                  . | .                  |       0.044885
 12    0.00007216        0.01925    |                  . | .                  |       0.044885
 13    0.00015685        0.04183    |                  . |*.                  |       0.044902
 14    0.00005506        0.01469    |                  . | .                  |       0.044978
 15    -0.0001558        -.04155    |                  .*| .                  |       0.044988
 16    0.00008757        0.02336    |                  . | .                  |       0.045063
 17    0.00007530        0.02008    |                  . | .                  |       0.045087
 18    0.00017094        0.04559    |                  . |*.                  |       0.045105
```

Looks like white noise.

Models for Nonstationary Data

Here is the **PACF** for the first differences of **CLOSE**.

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        0.04699    |                  . |*.                  |
  2        0.01882    |                  . | .                  |
  3        0.04411    |                  . |*.                  |
  4       -0.04103    |                  .*| .                  |
  5       -0.00245    |                  . | .                  |
  6        0.01389    |                  . | .                  |
  7       -0.05977    |                  .*| .                  |
  8       -0.02794    |                  .*| .                  |
  9       -0.02067    |                  . | .                  |
 10        0.01553    |                  . | .                  |
 11        0.00243    |                  . | .                  |
 12        0.01788    |                  . | .                  |
 13        0.03960    |                  . |*.                  |
 14        0.00813    |                  . | .                  |
 15       -0.04885    |                  .*| .                  |
 16        0.02114    |                  . | .                  |
 17        0.02201    |                  . | .                  |
 18        0.04841    |                  . |*.                  |
```

Models for Nonstationary Data

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	3.22	6	0.7803	0.047	0.021	0.046	-0.036	-0.004	0.014
12	6.24	12	0.9037	-0.062	-0.032	-0.024	0.006	0.004	0.019
18	9.77	18	0.9391	0.042	0.015	-0.042	0.023	0.020	0.046
24	12.28	24	0.9766	-0.010	-0.005	-0.035	-0.045	0.008	-0.035

So taking first differences transforms the series into white noise, i.e., the first differences are definitely stationary.

The question is: Is the series **stationary around a linear trend** or does it have a **unit root** (and thus require first differencing)?

Models for Nonstationary Data

So far we've been only using the **ACF** and **PACF** to help identify an **ARIMA**(p,d,q) model.

A third function used in the identification step is the **inverse autocorrelation function**, **IACF**.

It can be particulary useful in determing whether or not a series has been **overdifferenced**.

Given the ARMA(p,q) model $\phi(B)Y_t = \theta(B)e_t$, recall that Y_t is **stationary** if the roots of $\phi(B)$ are all outside the unit circle and is **invertible** if the roots of $\theta(B)$ are all outside the unit circle.

If both conditions hold, then the so-called **dual model** $\theta(B)Y_t = \phi(B)e_t$ is also stationary and invertible.

Being stationary, the **ACF** of the dual model, called the **IACF** of Y, should exponentially decline.

Models for Nonstationary Data

Note that if Y_t is a pure **AR** series, then the **IACF** of Y_t is the same as the **ACF** of a pure **MA** process and will cut off after a few lags, i.e. the **IACF** behaves like the **PACF** of Y_t.

On the other hand, if Y_t is a pure **MA** series, then the **IACF** of Y_t is the same as the **ACF** of a pure **AR** process and will decline exponentially, i.e. the **IACF** again behaves like the **PACF** of Y_t.

The **IACF** is estimated in **PROC ARIMA** as follows: 1^{st}, a high-order **AR** model is fit to the data by means of the **Yule - Walker equations**. This step results in an estimated polynomial ϕ(B). The **IACF** is then calculated as the **ACF** of a pure **MA** process defined by the polynomial ϕ(B).

Models for Nonstationary Data

Here is the **IACF** for the levels of **CLOSE**.

```
                        Inverse 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       -0.49317    |          **********| .                  |
   2        0.00012    |                  . | .                  |
   3       -0.02366    |                  . | .                  |
   4       -0.00355    |                  . | .                  |
   5        0.03411    |                  . |*.                  |
   6       -0.02025    |                  . | .                  |
   7        0.01472    |                  . | .                  |
   8       -0.00604    |                  . | .                  |
   9       -0.00172    |                  . | .                  |
  10       -0.00142    |                  . | .                  |
  11       -0.00087    |                  . | .                  |
  12       -0.00469    |                  . | .                  |
  13        0.01064    |                  . | .                  |
  14       -0.01339    |                  . | .                  |
  15        0.03416    |                  . |*.                  |
  16       -0.02223    |                  . | .                  |
  17       -0.02395    |                  . | .                  |
  18        0.03719    |                  . |*.
```

This looks like the **ACF** of an **MA(1)** model, i.e., the original series could be an **AR(1)** model.

Models for Nonstationary Data

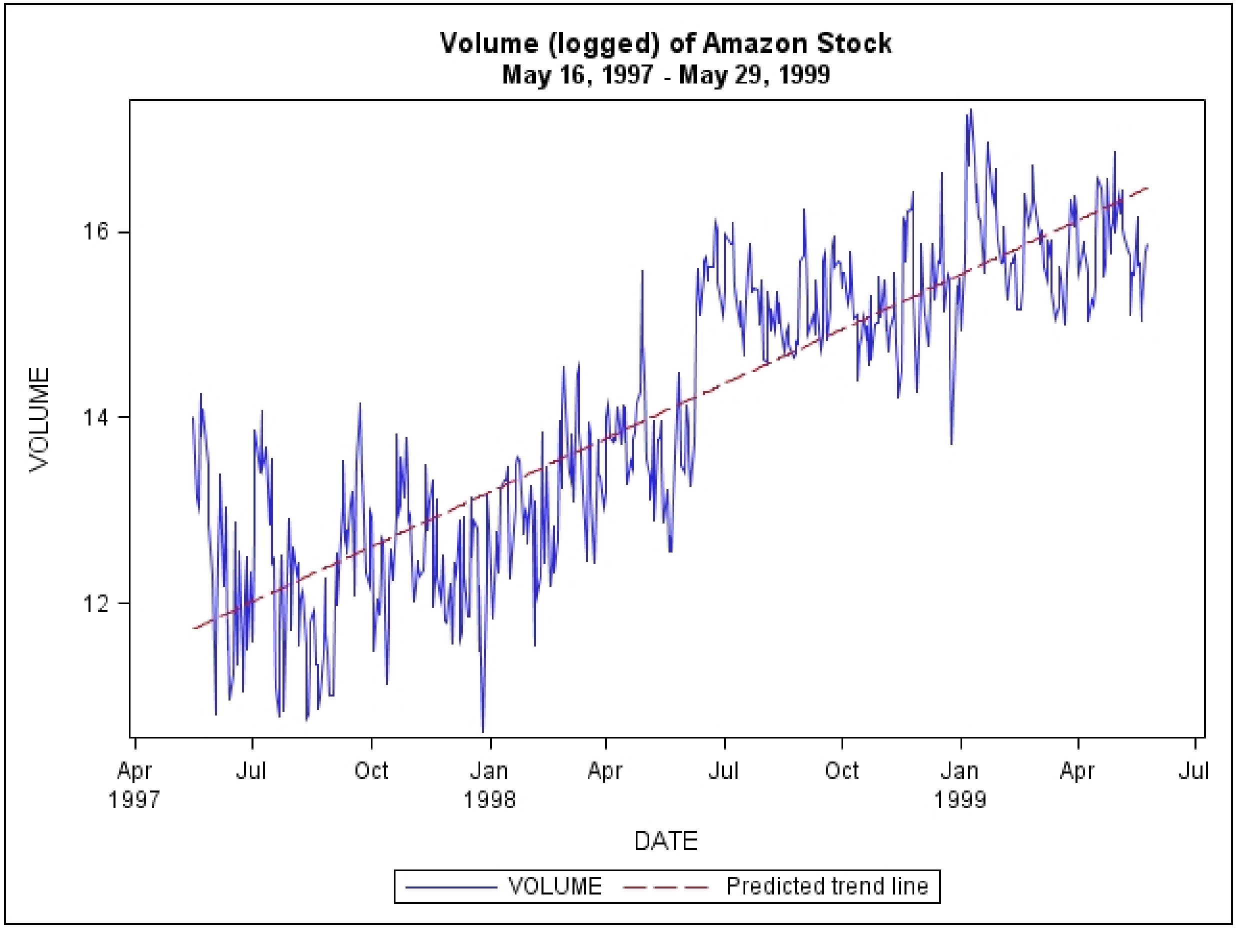
Here is the **IACF** for the first difference of **CLOSE**.

```
                    Inverse 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       -0.04365    |                   .*| .                  |
   2       -0.01691    |                   . | .                  |
   3       -0.04750    |                   .*| .                  |
   4        0.03442    |                   . |*.                  |
   5       -0.00135    |                   . | .                  |
   6       -0.02547    |                   .*| .                  |
   7        0.06211    |                   . |*.                  |
   8        0.02538    |                   . |*.                  |
   9        0.01200    |                   . | .                  |
  10       -0.01610    |                   . | .                  |
  11       -0.00202    |                   . | .                  |
  12       -0.01726    |                   . | .                  |
  13       -0.03674    |                   .*| .                  |
  14       -0.00937    |                   . | .                  |
  15        0.06007    |                   . |*.                  |
  16       -0.01998    |                   . | .                  |
```

This looks like the **ACF** of white noise.

Models for Nonstationary Data

In contrast, look at the same picture for the **volume** of Amazon stock.



Models for Nonstationary Data

If a series Y_t has a (near) unit root on the **MA** side, i.e., is not invertible, then the **IACF** of Y_t would be the **ACF** for a (near) nonstationary series and will die off slowly.

IACF of the levels of volume.

```
                          Inverse 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       -0.46787    |            *********| .                  |
  2        0.06602    |                   . |*.                  |
  3       -0.05657    |                   .*| .                  |
  4        0.00131    |                   . | .                  |
  5       -0.03400    |                   .*| .                  |
  6       -0.00470    |                   . | .                  |
  7        0.04408    |                   . |*.                  |
  8       -0.00976    |                   . | .                  |
  9       -0.09283    |                   **| .                  |
 10        0.11628    |                   . |**                  |
 11       -0.04616    |                   .*| .                  |
 12       -0.06186    |                   .*| .                  |
 13        0.08452    |                   . |**                  |
 14       -0.03373    |                   .*| .                  |
 15       -0.01124    |                   . | .                  |
 16        0.01678    |                   . | .                  |
 17        0.01158    |                   . | .                  |
 18        0.00119    |                   . | .                  |
```

Models for Nonstationary Data

IACF of the first differences of volume

```
                         Inverse 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        0.48216    |                  . |**********          |
  2        0.44816    |                  . |*********           |
  3        0.34266    |                  . |*******             |
  4        0.30682    |                  . |******              |
  5        0.25213    |                  . |*****               |
  6        0.24854    |                  . |*****               |
  7        0.23624    |                  . |*****               |
  8        0.18675    |                  . |****                |
  9        0.14088    |                  . |***                 |
 10        0.20330    |                  . |****                |
 11        0.13295    |                  . |***                 |
 12        0.11437    |                  . |**                  |
 13        0.15524    |                  . |***                 |
 14        0.11829    |                  . |**                  |
 15        0.09978    |                  . |**                  |
 16        0.10919    |                  . |**
```

What are we to make of this behavior of the **IACF** for the first differences of volume?

Models for Nonstationary Data

Chang and **Dickey (1993)** give a detailed proof of what happens to the **IACF** when a series is **overdifferenced**.

They find that an essentially linear descent in the **IACF** is consistent with overdifferencing.

It's important to distinguish between a **trend stationary series** (one that is stationary around a deterministic trend) and a **difference stationary series** (one with a unit root that must be differenced).

In the first case, **forecasts** will converge to the long run trend line, while in the second case, forecasts of levels will wander unpredictably.

For series with obvious trend, such as **CLOSE** and **VOLUME**, should use the third form of the **ADF** unit root test.

Models for Nonstationary Data

ADF test for **CLOSE**

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	1	0.8393	0.8841	1.90	0.9867		
Single Mean	1	-0.9880	0.8871	-0.83	0.8081	3.20	0.2496
Trend	**1**	**-17.0203**	**0.1231**	**-2.78**	**0.2046**	**3.92**	**0.3884**

ADF test for **VOLUME**

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	2	0.0144	0.6861	0.02	0.6909		
Single Mean	2	-14.2100	0.0474	-2.60	0.0944	3.42	0.1920
Trend	**2**	**-85.7758**	**0.0007**	**-6.35**	**<.0001**	**20.18**	**0.0010**

Conclusion: **CLOSE is difference stationary and VOLUME is trend stationary**.

Even though **CLOSE** hovers close to its trend line and **VOLUME** wanders quite a bit , pictures don’t tell all the story.

Differencing to Remove a Linear Trend

A linear trend can always be removed by taking a difference.

For example, if $Y_t = \alpha + \beta t + e_t$ where e_t is white noise, then $\Delta Y_t = \beta + e_t - e_{t-1}$.

But, in the process of removing the trend $\alpha + \beta t$, we have introduced a non-invertible **MA** term into the model, i.e., the **IACF** of ΔY_t will be the **ACF** of a unit root process and, hence, decline very slowly.

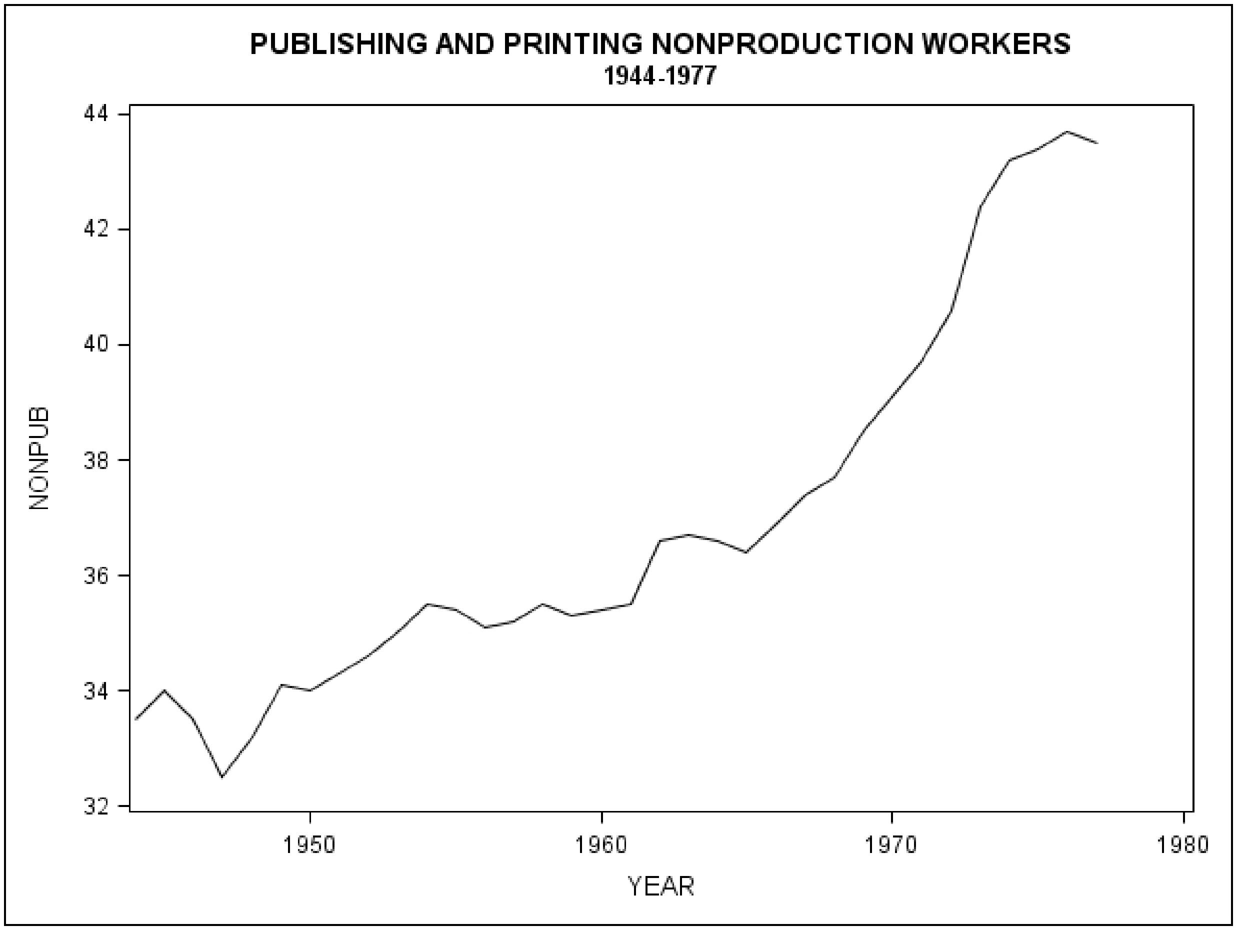
In this way, the **IACF** can detect when a series has been over-differenced.

Since the orginal series Y_t is not stationary, the correct approach in this instance to removing the trend is to regress Y_t on 1 and t and work with the residuals $\hat{e}_t$.

If e_t is not white noise, but instead is a nonstationary series whose difference Δe_t is stationary, then taking first differences of Y_t is appropriate.

Differencing to Remove a Linear Trend

Consider the following data series of nonproduction workers in the publishing and printing business (BLS, 1977).



Differencing to Remove a Linear Trend

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	1	0.2542	0.7348	1.89	0.9837		
Single Mean	1	0.9881	0.9853	0.67	0.9894	1.87	0.6055
Trend	1	-6.5467	0.6645	-1.88	0.6397	3.18	0.5589

Unit root tests for the **levels** definitely **accept** a unit root.

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	1	-17.4139	0.0016	-2.89	0.0053		
Single Mean	1	-33.1413	0.0002	-3.97	0.0044	7.89	0.0010
Trend	1	-48.9346	<.0001	-4.34	0.0088	9.81	0.0010

While unit root tests for the **first differences reject** a unit root.

Differencing to Remove a Linear Trend

```
                                   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      0.263930        1.00000    |                    |********************|             0
   1      0.082387        0.31216    |             .      |******.             |      0.174078
   2     -0.025565        -.09686    |            .     **|       .            |      0.190285
   3     0.0079422        0.03009    |            .       |*      .            |      0.191774
   4      0.034691        0.13144    |            .       |***    .            |      0.191917
   5      0.010641        0.04032    |            .       |*      .            |      0.194626
   6     -0.014492        -.05491    |            .      *|       .            |      0.194879
   7     -0.028101        -.10647    |            .     **|       .            |      0.195347
   8     -0.018074        -.06848    |            .      *|       .            |      0.197098

                                  "." marks two standard errors


                                 Inverse 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       -0.41859    |            ********|      .             |
                 2        0.23512    |             .      |***** .             |
                 3       -0.09418    |             .    **|      .             |
                 4       -0.06272    |             .     *|      .             |
                 5        0.02405    |             .      |      .             |
                 6       -0.01591    |             .      |      .             |
                 7        0.05923    |             .      |*     .             |
                 8        0.02180    |             .      |      .             |
```

Differencing to Remove a Linear Trend

```
                    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        0.31216     |             .      |******.             |
   2       -0.21528     |             .  ****|      .             |
   3        0.15573     |             .      |***   .             |
   4        0.05219     |             .      |*     .             |
   5       -0.00997     |             .      |      .             |
   6       -0.03898     |             .     *|      .             |
   7       -0.09630     |             .    **|      .             |
   8       -0.02738     |             .     *|      .             |
```

The diagnostic plots point to a **MA(1)** model.

```
proc arima  data = ffc2010.WORKERS;
        identify var = nonpub(1) noprint ;
        estimate q = 1;
        forecast lead = 10;
run;
```

Differencing to Remove a Linear Trend

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	**0.30330**	**0.12294**	**2.47**	**0.0193**	**0**
MA1,1	**-0.46626**	**0.16148**	**-2.89**	**0.0070**	**1**

Constant Estimate	0.3033
Variance Estimate	0.238422
Std Error Estimate	0.488284
AIC	48.27419
SBC	51.2672
Number of Residuals	33

* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	**Pr > ChiSq**	Autocorrelations					
6	1.01	5	0.9619	-0.033	-0.089	0.020	0.119	-0.032	-0.036
12	3.80	11	0.9754	-0.054	-0.114	0.157	-0.093	0.072	-0.057
18	7.41	17	0.9776	0.064	0.001	-0.175	-0.108	-0.027	0.085
24	10.07	23	0.9909	0.007	-0.023	0.067	-0.003	-0.123	0.057

Differencing to Remove a Linear Trend

Forecasts for variable NONPUB

Obs	Forecast	Std Error	95% Confidence Limits	
35	43.5635	0.4883	42.6065	44.5206
36	43.8668	0.8666	42.1683	45.5654
37	44.1701	1.1241	41.9669	46.3733
38	44.4734	1.3327	41.8613	47.0855
39	44.7767	1.5129	41.8116	47.7419
40	45.0800	1.6737	41.7996	48.3605
41	45.3833	1.8204	41.8154	48.9513
42	45.6866	1.9562	41.8526	49.5206
43	45.9899	2.0831	41.9072	50.0727
44	46.2932	2.2027	41.9761	50.6104

Differencing to Remove a Linear Trend

The estimated parameter **MU** represents the mean of the first difference and is called the **drift**.

Note that if $Y_t - Y_{t-1} = \mu + e_t$, then $Y_t = Y_{t-1} + \mu + e_t = Y_{t-2} + \mu + e_{t-1} + \mu + e_t$ and by repeatedly back-substituting we can write $Y_t = Y_0 + \mu t + \sum_{j=0}^{t} e_j$, so μ becomes the slope of a deterministic trend.

We also see that $Var(Y_t) = Var\left(\sum_{j=0}^{t} e_j\right) = (t+1)\sigma^2$, so that Y_t has infinite variance.

Other Identification Techniques

The book goes on to discuss three other identification tools called **ESACF**, **SCAN**, and **MINIC**. I won't cover these here.