Chapter 8
The AUTOREG Procedure

Chapter Table of Contents

OVERVIEW .......................................................... 303

GETTING STARTED ................................................. 305
  Regression with Autocorrelated Errors .................................. 305
  Forecasting Autoregressive Error Models .................................. 311
  Testing for Autocorrelation ............................................. 313
  Stepwise Autoregression .............................................. 315
  Testing for Heteroscedasticity ........................................... 317
  Heteroscedasticity and GARCH Models ................................... 320

SYNTAX .............................................................. 324
  Functional Summary .................................................. 324
  PROC AUTOREG Statement ............................................ 327
  BY Statement ....................................................... 328
  MODEL Statement .................................................... 328
  HETERO Statement ................................................... 335
  RESTRICT Statement ................................................. 337
  TEST Statement ..................................................... 338
  OUTPUT Statement ................................................... 339

DETAILS ............................................................ 342
  Missing Values ..................................................... 342
  Autoregressive Error Model ........................................... 342
  Alternative Autocorrelation Correction Methods ....................... 346
  GARCH, IGARCH, EGARCH, and GARCH-M Models .................... 347
  R^2 Statistics and Other Measures of Fit ............................. 352
  Generalized Durbin-Watson Tests .................................... 354
  Testing ............................................................ 358
  Predicted Values .................................................... 363
  OUT= Data Set ...................................................... 366
  OUTTEST= Data Set ................................................ 367
  Printed Output ...................................................... 368
  ODS Table Names .................................................... 368

EXAMPLES .......................................................... 370
  Example 8.1 Analysis of Real Output Series .......................... 370
Part 2. General Information

Example 8.2 Comparing Estimates and Models .................................. 374
Example 8.3 Lack of Fit Study ......................................................... 377
Example 8.4 Missing Values ............................................................. 380
Example 8.5 Money Demand Model .................................................. 384
Example 8.6 Estimation of ARCH(2) Process ....................................... 387

REFERENCES ................................................................................. 392
Chapter 8
The AUTOREG Procedure

Overview

The AUTOREG procedure estimates and forecasts linear regression models for time series data when the errors are autocorrelated or heteroscedastic. The autoregressive error model is used to correct for autocorrelation, and the generalized autoregressive conditional heteroscedasticity (GARCH) model and its variants are used to model and correct for heteroscedasticity.

When time series data are used in regression analysis, often the error term is not independent through time. Instead, the errors are *serially correlated* or *autocorrelated*. If the error term is autocorrelated, the efficiency of ordinary least-squares (OLS) parameter estimates is adversely affected and standard error estimates are biased.

The autoregressive error model corrects for serial correlation. The AUTOREG procedure can fit autoregressive error models of any order and can fit subset autoregressive models. You can also specify stepwise autoregression to select the autoregressive error model automatically.

To diagnose autocorrelation, the AUTOREG procedure produces generalized Durbin-Watson (DW) statistics and their marginal probabilities. Exact $p$-values are reported for generalized DW tests to any specified order. For models with lagged dependent regressors, PROC AUTOREG performs the Durbin $t$-test and the Durbin $h$-test for first-order autocorrelation and reports their marginal significance levels.

Ordinary regression analysis assumes that the error variance is the same for all observations. When the error variance is not constant, the data are said to be *heteroscedastic*, and ordinary least-squares estimates are inefficient. Heteroscedasticity also affects the accuracy of forecast confidence limits. More efficient use of the data and more accurate prediction error estimates can be made by models that take the heteroscedasticity into account.

To test for heteroscedasticity, the AUTOREG procedure uses the portmanteau test statistics and the Engle Lagrange multiplier tests. Test statistics and significance $p$-values are reported for conditional heteroscedasticity at lags 1 through 12. The Bera-Jarque normality test statistic and its significance level are also reported to test for conditional nonnormality of residuals.

The family of GARCH models provides a means of estimating and correcting for the changing variability of the data. The GARCH process assumes that the errors, although uncorrelated, are not independent and models the conditional error variance as a function of the past realizations of the series.
The AUTOREG procedure supports the following variations of the GARCH models:

- generalized ARCH (GARCH)
- integrated GARCH (IGARCH)
- exponential GARCH (EGARCH)
- GARCH-in-mean (GARCH-M)

For GARCH-type models, the AUTOREG procedure produces the conditional prediction error variances as well as parameter and covariance estimates.

The AUTOREG procedure can also analyze models that combine autoregressive errors and GARCH-type heteroscedasticity. PROC AUTOREG can output predictions of the conditional mean and variance for models with autocorrelated disturbances and changing conditional error variances over time.

Four estimation methods are supported for the autoregressive error model:

- Yule-Walker
- iterated Yule-Walker
- unconditional least squares
- exact maximum likelihood

The maximum likelihood method is used for GARCH models and for mixed AR-GARCH models.

The AUTOREG procedure produces forecasts and forecast confidence limits when future values of the independent variables are included in the input data set. PROC AUTOREG is a useful tool for forecasting because it uses the time series part of the model as well as the systematic part in generating predicted values. The autoregressive error model takes into account recent departures from the trend in producing forecasts.

The AUTOREG procedure permits embedded missing values for the independent or dependent variables. The procedure should be used only for ordered and equally spaced time series data.
Regression with Autocorrelated Errors

Ordinary regression analysis is based on several statistical assumptions. One key assumption is that the errors are independent of each other. However, with time series data, the ordinary regression residuals usually are correlated over time. It is not desirable to use ordinary regression analysis for time series data since the assumptions on which the classical linear regression model is based will usually be violated.

Violation of the independent errors assumption has three important consequences for ordinary regression. First, statistical tests of the significance of the parameters and the confidence limits for the predicted values are not correct. Second, the estimates of the regression coefficients are not as efficient as they would be if the autocorrelation were taken into account. Third, since the ordinary regression residuals are not independent, they contain information that can be used to improve the prediction of future values.

The AUTOREG procedure solves this problem by augmenting the regression model with an autoregressive model for the random error, thereby accounting for the autocorrelation of the errors. Instead of the usual regression model, the following autoregressive error model is used:

\[
\begin{align*}
    y_t &= x_t^\prime \beta + \nu_t \\
    \nu_t &= -\varphi_1 \nu_{t-1} - \varphi_2 \nu_{t-2} - \cdots - \varphi_m \nu_{t-m} + \epsilon_t \\
    \epsilon_t &\sim \text{IN}(0, \sigma^2)
\end{align*}
\]

The notation \( \epsilon_t \sim \text{IN}(0, \sigma^2) \) indicates that each \( \epsilon_t \) is normally and independently distributed with mean 0 and variance \( \sigma^2 \).

By simultaneously estimating the regression coefficients \( \beta \) and the autoregressive error model parameters \( \varphi_i \), the AUTOREG procedure corrects the regression estimates for autocorrelation. Thus, this kind of regression analysis is often called autoregressive error correction or serial correlation correction.

Example of Autocorrelated Data

A simulated time series is used to introduce the AUTOREG procedure. The following statements generate a simulated time series \( Y \) with second-order autocorrelation:

```sas
data a;
    ul = 0; ull = 0;
    do time = -10 to 36;
        u = +1.3 * ul - .5 * ull + 2*rannor(12346);
        y = 10 + .5 * time + u;
        if time > 0 then output;
        ull = ul; ul = u;
    end;
run;
```
The series Y is a time trend plus a second-order autoregressive error. The model simulated is

\[ y_t = 10 + 0.5t + \nu_t \]
\[ \nu_t = 1.3\nu_{t-1} - 0.5\nu_{t-2} + \epsilon_t \]
\[ \epsilon_t \sim \text{IN}(0,4) \]

The following statements plot the simulated time series Y. A linear regression trend line is shown for reference. (The regression line is produced by plotting the series a second time using the regression interpolation feature of the SYMBOL statement. Refer to SAS/GRAPH Software: Reference, Version 6, First Edition, Volume 1 for further explanation.)

```
title "Autocorrelated Time Series";
proc gplot data=a;
  symbol1 v=dot i=join;
  symbol2 v=none i=r;
  plot y * time = 1 y * time = 2 / overlay;
run;
```

The plot of series Y and the regression line are shown in Figure 8.1.

![Autocorrelated Time Series](image)

**Figure 8.1.** Autocorrelated Time Series

Note that when the series is above (or below) the OLS regression trend line, it tends to remain above (below) the trend for several periods. This pattern is an example of positive autocorrelation.
Time series regression usually involves independent variables other than a time-trend. However, the simple time-trend model is convenient for illustrating regression with autocorrelated errors, and the series Y shown in Figure 8.1 is used in the following introductory examples.

**Ordinary Least-Squares Regression**

To use the AUTOREG procedure, specify the input data set in the PROC AUTOREG statement and specify the regression model in a MODEL statement. Specify the model by first naming the dependent variable and then listing the regressors after an equal sign, as is done in other SAS regression procedures. The following statements regress Y on TIME using ordinary least squares:

```sas
proc autoreg data=a;
   model y = time;
run;
```

The AUTOREG procedure output is shown in Figure 8.2.

The output first shows statistics for the model residuals. The model root mean square error (Root MSE) is 2.51, and the model $R^2$ is .82. Notice that two $R^2$ statistics are shown, one for the regression model (Reg Rsq) and one for the full model (Total Rsq) that includes the autoregressive error process, if any. In this case, an autoregressive error model is not used, so the two $R^2$ statistics are the same.

Other statistics shown are the sum of square errors (SSE), mean square error (MSE), error degrees of freedom (DFE, the number of observations minus the number of parameters), the information criteria SBC and AIC, and the Durbin-Watson statistic. (Durbin-Watson statistics and SBC and AIC are discussed in the "Details" section later in this chapter.)
The output then shows a table of regression coefficients, with standard errors and $t$-tests. The estimated model is

$$y_t = 8.23 + .502t + \epsilon_t$$

$$Est. Var(\epsilon_t) = 6.32$$

The OLS parameter estimates are reasonably close to the true values, but the estimated error variance, 6.32, is much larger than the true value, 4.

**Autoregressive Error Model**

The following statements regress $Y$ on $TIME$ with the errors assumed to follow a second-order autoregressive process. The order of the autoregressive model is specified by the NLAG=2 option. The Yule-Walker estimation method is used by default. The example uses the METHOD=ML option to specify the exact maximum likelihood method instead.

```sas
proc autoreg data=a;
   model y = time / nlag=2 method=ml;
run;
```

The first part of the results are shown in Figure 8.3. The initial OLS results are produced first, followed by estimates of the autocorrelations computed from the OLS residuals. The autocorrelations are also displayed graphically.

---

**The AUTOREG Procedure**

**Dependent Variable  y**

<table>
<thead>
<tr>
<th></th>
<th>SSE</th>
<th>DFE</th>
<th>MSE</th>
<th>Root MSE</th>
<th>SBC</th>
<th>AIC</th>
<th>Regress R-Square</th>
<th>Total R-Square</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>214.953429</td>
<td>34</td>
<td>6.32216</td>
<td>2.51439</td>
<td>173.659101</td>
<td>170.492063</td>
<td>0.8200</td>
<td>0.8200</td>
<td>0.4752</td>
</tr>
</tbody>
</table>

| Variable | DF | Estimate | Error  | $t$ Value | Pr > |t|   |
|----------|----|----------|--------|-----------|------|----|
| Intercept | 1  | 8.2308   | 0.8559 | 9.62      | <.0001|
| time     | 1  | 0.5021   | 0.0403 | 12.45     | <.0001|

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
<th>-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.9709</td>
<td>1.000000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.5169</td>
<td>0.756485</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.0241</td>
<td>0.338995</td>
<td></td>
</tr>
</tbody>
</table>

| Preliminary MSE | 1.7943 |

---

**Figure 8.3.** Preliminary Estimate for AR(2) Error Model
The maximum likelihood estimates are shown in Figure 8.4. Figure 8.4 also shows the preliminary Yule-Walker estimates used as starting values for the iterative computation of the maximum likelihood estimates.

![The AUTOREG Procedure](image)

### Estimates of Autoregressive Parameters

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.169057</td>
<td>0.148172</td>
<td>-7.89</td>
</tr>
<tr>
<td>2</td>
<td>0.545379</td>
<td>0.148172</td>
<td>3.68</td>
</tr>
</tbody>
</table>

Algorithm converged.

### Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>54.7493022</td>
</tr>
<tr>
<td>MSE</td>
<td>1.71092</td>
</tr>
<tr>
<td>SBC</td>
<td>133.476508</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.7280</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>2.2761</td>
</tr>
</tbody>
</table>

| Variable     | DF | Estimate | Standard Error | t Value | Pr > |t| |
|--------------|----|----------|----------------|---------|------|---|
| Intercept    | 1  | 7.8833   | 1.1693         | 6.74    | <.0001 |
| time         | 1  | 0.5096   | 0.0551         | 9.25    | <.0001 |
| AR1          | 1  | -1.2464  | 0.1385         | -9.00   | <.0001 |
| AR2          | 1  | 0.6283   | 0.1366         | 4.60    | <.0001 |

Autoregressive parameters assumed given.

| Variable     | DF | Estimate | Standard Error | t Value | Pr > |t| |
|--------------|----|----------|----------------|---------|------|---|
| Intercept    | 1  | 7.8833   | 1.1678         | 6.75    | <.0001 |
| time         | 1  | 0.5096   | 0.0551         | 9.26    | <.0001 |

**Figure 8.4.** Maximum Likelihood Estimates of AR(2) Error Model

The diagnostic statistics and parameter estimates tables in Figure 8.4 have the same form as in the OLS output, but the values shown are for the autoregressive error model. The MSE for the autoregressive model is 1.71, which is much smaller than the true value of 4. In small samples, the autoregressive error model tends to underestimate \( \sigma^2 \), while the OLS MSE overestimates \( \sigma^2 \).

Notice that the total R\(^2\) statistic computed from the autoregressive model residuals is .954, reflecting the improved fit from the use of past residuals to help predict the next Y value. The Reg Rsq value .728 is the R\(^2\) statistic for a regression of transformed variables adjusted for the estimated autocorrelation. (This is not the R\(^2\) for the estimated trend line. For details, see "R\(^2\) Statistics and Other Measures of Fit" later in this chapter.)
The parameter estimates table shows the ML estimates of the regression coefficients and includes two additional rows for the estimates of the autoregressive parameters, labeled A(1) and A(2).

The estimated model is

\[ y_t = 7.88 + 0.5096t + \nu_t \]

\[ \nu_t = 1.25\nu_{t-1} - 0.628\nu_{t-2} + \epsilon_t \]

\[ \text{Est. Var}(\epsilon_t) = 1.71 \]

Note that the signs of the autoregressive parameters shown in this equation for \( \nu_t \) are the reverse of the estimates shown in the AUTOREG procedure output. Figure 8.4 also shows the estimates of the regression coefficients with the standard errors recomputed on the assumption that the autoregressive parameter estimates equal the true values.

**Predicted Values and Residuals**

The AUTOREG procedure can produce two kinds of predicted values and corresponding residuals and confidence limits. The first kind of predicted value is obtained from only the structural part of the model, \( x_t^T \beta \). This is an estimate of the unconditional mean of the response variable at time \( t \). For the time trend model, these predicted values trace the estimated trend. The second kind of predicted values include both the structural part of the model and the predicted values of the autoregressive error process. The full model (conditional) predictions are used to forecast future values.

Use the OUTPUT statement to store predicted values and residuals in a SAS data set and to output other values such as confidence limits and variance estimates. The P= option specifies an output variable to contain the full model predicted values. The PM= option names an output variable for the predicted mean. The R= and RM= options specify output variables for the corresponding residuals, computed as the actual value minus the predicted value.

The following statements store both kinds of predicted values in the output data set. (The printed output is the same as previously shown in Figure 8.3 and Figure 8.4.)

```sas
proc autoreg data=a;
  model y = time / nlag=2 method=ml;
  output out=p p=yhat pm=trendhat;
run;
```

The following statements plot the predicted values from the regression trend line and from the full model together with the actual values.

```sas
title "Predictions for Autocorrelation Model";
proc gplot data=p;
  symbol1 v=star i=none;
```
symbol2 v=circle i=join;
symbol3 v=none i=join;
plot y * time = 1 yhat * time = 2
trendhat * time = 3 / overlay ;
run;

The plot of predicted values is shown in Figure 8.5.

**Figure 8.5.** PROC AUTOREG Predictions

In Figure 8.5 the straight line is the autocorrelation corrected regression line, traced out by the structural predicted values TRENDHAT. The jagged line traces the full model prediction values. The actual values are marked by asterisks. This plot graphically illustrates the improvement in fit provided by the autoregressive error process for highly autocorrelated data.

### Forecasting Autoregressive Error Models

To produce forecasts for future periods, include observations for the forecast periods in the input data set. The forecast observations must provide values for the independent variables and have missing values for the response variable.

For the time trend model, the only regressor is time. The following statements add observations for time periods 37 through 46 to the data set A to produce an augmented data set B:
To produce the forecast, use the augmented data set as input to PROC AUTOREG, and specify the appropriate options in the OUTPUT statement. The following statements produce forecasts for the time trend with autoregressive error model. The output data set includes all the variables in the input data set, the forecast values (YHAT), the predicted trend (YTREND), and the upper (UCL) and lower (LCL) 95% confidence limits.

```sas
proc autoreg data=b;
  model y = time / nlag=2 method=ml;
  output out=p p=yhat pm=ytrend
       lcl=lcl ucl=ucl;
run;
```

The following statements plot the predicted values and confidence limits, and they also plot the trend line for reference. The actual observations are shown for periods 16 through 36, and a reference line is drawn at the start of the out-of-sample forecasts.

```sas
title "Forecasting Autocorrelated Time Series";
proc gplot data=p;
  plot y*time=1 yhat*time=2 ytrend*time=3
     lcl*time=3 ucl*time=3 / overlay href=36.5;
  where time >= 16;
  symbol1 v=star i=none;
  symbol2 v=circle i=join;
  symbol3 v=none i=join;
run;
```

The plot is shown in Figure 8.6. Notice that the forecasts take into account the recent departures from the trend but converge back to the trend line for longer forecast horizons.
Testing for Autocorrelation

In the preceding section, it is assumed that the order of the autoregressive process is known. In practice, you need to test for the presence of autocorrelation.

The Durbin-Watson test is a widely used method of testing for autocorrelation. The first-order Durbin-Watson statistic is printed by default. This statistic can be used to test for first-order autocorrelation. Use the DWPROB option to print the significance level (p-values) for the Durbin-Watson tests. (Since the Durbin-Watson p-values are computationally expensive, they are not reported by default.)

You can use the DW= option to request higher-order Durbin-Watson statistics. Since the ordinary Durbin-Watson statistic only tests for first-order autocorrelation, the Durbin-Watson statistics for higher-order autocorrelation are called generalized Durbin-Watson statistics.

The following statements perform the Durbin-Watson test for autocorrelation in the OLS residuals for orders 1 through 4. The DWPROB option prints the marginal significance levels (p-values) for the Durbin-Watson statistics.

```sas
proc autoreg data=a;
   model y = time / dw=4 dwprob;
run;
```

The AUTOREG procedure output is shown in Figure 8.7. In this case, the first-order Durbin-Watson test is highly significant, with p < .0001 for the hypothesis of no first-order autocorrelation. Thus, autocorrelation correction is needed.
The AUTOREG Procedure

Dependent Variable  y

Ordinary Least Squares Estimates

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>214.953429</td>
</tr>
<tr>
<td>MSE</td>
<td>6.32216</td>
</tr>
<tr>
<td>SBC</td>
<td>173.659101</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.8200</td>
</tr>
<tr>
<td>DFE</td>
<td>34</td>
</tr>
<tr>
<td>Root MSE</td>
<td>2.51439</td>
</tr>
<tr>
<td>AIC</td>
<td>170.492063</td>
</tr>
<tr>
<td>Total R-Square</td>
<td>0.8200</td>
</tr>
</tbody>
</table>

Durbin-Watson Statistics

<table>
<thead>
<tr>
<th>Order</th>
<th>DW</th>
<th>Pr &lt; DW</th>
<th>Pr &gt; DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4752</td>
<td>&lt;.0001</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>1.2935</td>
<td>0.0137</td>
<td>0.9863</td>
</tr>
<tr>
<td>3</td>
<td>2.0694</td>
<td>0.6545</td>
<td>0.3455</td>
</tr>
<tr>
<td>4</td>
<td>2.5544</td>
<td>0.9818</td>
<td>0.0182</td>
</tr>
</tbody>
</table>

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

| Variable | DF | Estimate | Standard Error | t Value | Approx Pr > |t| |
|----------|----|----------|----------------|---------|-------------|----|
| Intercept| 1  | 8.2308   | 0.8559         | 9.62    | <.0001      |
| time     | 1  | 0.5021   | 0.0403         | 12.45   | <.0001      |

Figure 8.7. Durbin-Watson Test Results for OLS Residuals

Using the Durbin-Watson test, you can decide if autocorrelation correction is needed. However, generalized Durbin-Watson tests should not be used to decide on the autoregressive order. The higher-order tests assume the absence of lower-order autocorrelation. If the ordinary Durbin-Watson test indicates no first-order autocorrelation, you can use the second-order test to check for second-order autocorrelation. Once autocorrelation is detected, further tests at higher orders are not appropriate. In Figure 8.7, since the first-order Durbin-Watson test is significant, the order 2, 3, and 4 tests can be ignored.

When using Durbin-Watson tests to check for autocorrelation, you should specify an order at least as large as the order of any potential seasonality, since seasonality produces autocorrelation at the seasonal lag. For example, for quarterly data use DW=4, and for monthly data use DW=12.

Lagged Dependent Variables

The Durbin-Watson tests are not valid when the lagged dependent variable is used in the regression model. In this case, the Durbin h-test or Durbin t-test can be used to test for first-order autocorrelation.

For the Durbin h-test, specify the name of the lagged dependent variable in the LAGDEP= option. For the Durbin t-test, specify the LAGDEP option without giving the name of the lagged dependent variable.

For example, the following statements add the variable YLAG to the data set A and regress Y on YLAG instead of TIME.
data b;
  set a;
  ylag = lag1( y );
run;

proc autoreg data=b;
  model y = ylag / lagdep=ylag;
run;

The results are shown in Figure 8.8. The Durbin h statistic 2.78 is significant with a p-value of .0027, indicating autocorrelation.

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|-------|---|
| Intercept| 1  | 1.5742   | 0.9300         | 1.69    | 0.0999|
| ylag     | 1  | 0.9376   | 0.0510         | 18.37   | <.0001|

Figure 8.8. Durbin h -Test With a Lagged Dependent Variable

Stepwise Autoregression

Once you determine that autocorrelation correction is needed, you must select the order of the autoregressive error model to use. One way to select the order of the autoregressive error model is stepwise autoregression. The stepwise autoregression method initially fits a high-order model with many autoregressive lags and then sequentially removes autoregressive parameters until all remaining autoregressive parameters have significant t-tests.

To use stepwise autoregression, specify the BACKSTEP option, and specify a large order with the NLAG= option. The following statements show the stepwise feature, using an initial order of 5:

```
proc autoreg data=a;
  model y = time / method=ml nlag=5 backstep;
run;
```

The results are shown in Figure 8.9.
The AUTOREG Procedure

Dependent Variable y

Ordinary Least Squares Estimates

| Variable   | DF | Estimate | Error | t Value | Pr > |t| |
|------------|----|----------|-------|---------|-------|----|
| Intercept  | 1  | 8.2308   | 0.8559| 9.62    | <.0001|
| time       | 1  | 0.5021   | 0.0403| 12.45   | <.0001|

Estimates of Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.9709</td>
<td>1.000000</td>
</tr>
<tr>
<td>1</td>
<td>4.5169</td>
<td>0.756485</td>
</tr>
<tr>
<td>2</td>
<td>2.0241</td>
<td>0.338995</td>
</tr>
<tr>
<td>3</td>
<td>-0.4402</td>
<td>-0.073725</td>
</tr>
<tr>
<td>4</td>
<td>-2.1175</td>
<td>-0.354632</td>
</tr>
<tr>
<td>5</td>
<td>-2.8534</td>
<td>-0.477887</td>
</tr>
</tbody>
</table>

Backward Elimination of Autoregressive Terms

| Lag | Estimate | t Value | Pr > |t| |
|-----|----------|---------|-------|----|
| 4   | -0.052908| -0.20   | 0.8442|
| 3   | 0.115986 | 0.57    | 0.5698|
| 5   | 0.131734 | 1.21    | 0.2340|

Figure 8.9. Stepwise Autoregression

The estimates of the autocorrelations are shown for 5 lags. The backward elimination of autoregressive terms report shows that the autoregressive parameters at lags 3, 4, and 5 were insignificant and eliminated, resulting in the second-order model shown previously in Figure 8.4. By default, retained autoregressive parameters must be significant at the .05 level, but you can control this with the SLSTAY= option. The remainder of the output from this example is the same as that in Figure 8.3 and Figure 8.4, and it is not repeated here.

The stepwise autoregressive process is performed using the Yule-Walker method. The maximum likelihood estimates are produced after the order of the model is determined from the significance tests of the preliminary Yule-Walker estimates.

When using stepwise autoregression, it is a good idea to specify an NLAG= option value larger than the order of any potential seasonality, since seasonality produces autocorrelation at the seasonal lag. For example, for monthly data use NLAG=13, and for quarterly data use NLAG=5.
Subset and Factored Models

In the previous example, the BACKSTEP option dropped lags 3, 4, and 5, leaving an order 2 model. However, in other cases a parameter at a longer lag may be kept while some smaller lags are dropped. For example, the stepwise autoregression method might drop lags 2, 3, and 5 but keep lags 1 and 4. This is called a subset model, since the number of estimated autoregressive parameters is smaller than the order of the model.

Subset models are common for seasonal data and often correspond to factored autoregressive models. A factored model is the product of simpler autoregressive models. For example, the best model for seasonal monthly data may be the combination of a first-order model for recent effects with a twelfth-order subset model for the seasonality, with a single parameter at lag 12. This results in an order 13 subset model with nonzero parameters at lags 1, 12, and 13. See Chapter 7, “The ARIMA Procedure,” for further discussion of subset and factored autoregressive models.

You can specify subset models with the NLAG= option. List the lags to include in the autoregressive model within parentheses. The following statements show an example of specifying the subset model resulting from the combination of a first-order process for recent effects with a fourth-order seasonal process:

```sas
proc autoreg data=a;
   model y = time / nlag=(1 4 5);
run;
```

The MODEL statement specifies the following fifth-order autoregressive error model:

\[ y_t = a + b t + \nu_t \]

\[ \nu_t = -\varphi_1 \nu_{t-1} - \varphi_4 \nu_{t-4} - \varphi_5 \nu_{t-5} + \epsilon_t \]

Testing for Heteroscedasticity

One of the key assumptions of the ordinary regression model is that the errors have the same variance throughout the sample. This is also called the homoscedasticity model. If the error variance is not constant, the data are said to be heteroscedastic.

Since ordinary least-squares regression assumes constant error variance, heteroscedasticity causes the OLS estimates to be inefficient. Models that take into account the changing variance can make more efficient use of the data. Also, heteroscedasticity can make the OLS forecast error variance inaccurate since the predicted forecast variance is based on the average variance instead of the variability at the end of the series.

To illustrate heteroscedastic time series, the following statements re-create the simulated series Y. The variable Y has an error variance that changes from 1 to 4 in the middle part of the series. The length of the series is also extended 120 observations.
Part 2. General Information

data a;
  ul = 0; ull = 0;
  do time = -10 to 120;
    s = 1 + (time >= 60 & time < 90);
    u = + 1.3 * ul - .5 * ull + s*rannor(12346);
    y = 10 + .5 * time + u;
    if time > 0 then output;
    ull = ul; ul = u;
  end;
run;

title "Heteroscedastic Autocorrelated Time Series";
proc gplot data=a;
  symbol1 v=dot i=join;
  symbol2 v=none i=r;
  plot y * time = 1 y * time = 2 / overlay;
run;

The simulated series is plotted in Figure 8.10.

Figure 8.10. Heteroscedastic and Autocorrelated Series

To test for heteroscedasticity with PROC AUTOREG, specify the ARCHTEST option. The following statements regress Y on TIME and use the ARCHTEST option to test for heteroscedastic OLS residuals. The DWPROB option is also used to test for autocorrelation.
The PROC AUTOREG output is shown in Figure 8.11. The Q statistics test for changes in variance across time using lag windows ranging from 1 through 12. (See "Heteroscedasticity and Normality Tests" for details.) The $p$-values for the test statistics are given in parentheses. These tests strongly indicate heteroscedasticity, with $p < 0.0001$ for all lag windows.

The Lagrange multiplier (LM) tests also indicate heteroscedasticity. These tests can also help determine the order of the ARCH model appropriate for modeling the heteroscedasticity, assuming that the changing variance follows an autoregressive conditional heteroscedasticity model.

```
proc autoreg data=a;
  model y = time / nlag=2 archtest dwprob;
  output out=r r=yresid;
run;
```

Figure 8.11. Heteroscedasticity Tests
Heteroscedasticity and GARCH Models

There are several approaches to dealing with heteroscedasticity. If the error variance at different times is known, weighted regression is a good method. If, as is usually the case, the error variance is unknown and must be estimated from the data, you can model the changing error variance.

The generalized autoregressive conditional heteroscedasticity (GARCH) model is one approach to modeling time series with heteroscedastic errors. The GARCH regression model with autoregressive errors is

\[
y_t = x_t \beta + \nu_t
\]

\[
\nu_t = \epsilon_t - \varphi_1 \nu_{t-1} - \ldots - \varphi_m \nu_{t-m}
\]

\[
\epsilon_t = \sqrt{h_t} \epsilon_t
\]

\[
h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j}
\]

\[
\epsilon_t \sim \text{IN}(0,1)
\]

This model combines the \(m\)th-order autoregressive error model with the GARCH\((p,q)\) variance model. It is denoted as the AR\((m)\)-GARCH\((p,q)\) regression model.

The Lagrange multiplier (LM) tests shown in Figure 8.11 can help determine the order of the ARCH model appropriate for the data. The tests are significant \((p<.0001)\) through order 12, which indicates that a very high-order ARCH model is needed to model the heteroscedasticity.

The basic ARCH\((q)\) model \((p=0)\) is a short memory process in that only the most recent \(q\) squared residuals are used to estimate the changing variance. The GARCH model \((p>0)\) allows long memory processes, which use all the past squared residuals to estimate the current variance. The LM tests in Figure 8.11 suggest the use of the GARCH model \((p>0)\) instead of the ARCH model.

The GARCH\((p,q)\) model is specified with the GARCH\(=(P=p,Q=q)\) option in the MODEL statement. The basic ARCH\((q)\) model is the same as the GARCH\((0,q)\) model and is specified with the GARCH\(=(Q=q)\) option.

The following statements fit an AR(2)-GARCH(1,1) model for the Y series regressed on TIME. The GARCH\(=(P=1,Q=1)\) option specifies the GARCH(1,1) conditional variance model. The NLAG=2 option specifies the AR(2) error process. Only the maximum likelihood method is supported for GARCH models; therefore, the METHOD= option is not needed. The CEV= option in the OUTPUT statement stores the estimated conditional error variance at each time period in the variable VHAT in an output data set named OUT.
The results for the GARCH model are shown in Figure 8.12. (The preliminary estimates are not shown.)

| Variable  | DF | Estimate | Standard Error | t Value | Pr > |t| |
|-----------|----|----------|----------------|--------|------|---|
| Intercept | 1  | 8.9301   | 0.7235         | 12.34  | <.0001 |
| time      | 1  | 0.5075   | 0.0107         | 47.30  | <.0001 |
| AR1       | 1  | -1.2301  | 0.1078         | -11.41 | <.0001 |
| AR2       | 1  | 0.5023   | 0.1057         | 4.75   | <.0001 |
| ARCH0     | 1  | 0.0850   | 0.0757         | 1.12   | 0.2614 |
| ARCH1     | 1  | 0.2103   | 0.0847         | 2.48   | 0.0130 |
| GARCH1    | 1  | 0.7376   | 0.0960         | 7.68   | <.0001 |

The normality test is not significant ($p = 0.957$), which is consistent with the hypothesis that the residuals from the GARCH model, $\epsilon_t/\sqrt{h_t}$, are normally distributed. The parameter estimates table includes rows for the GARCH parameters. ARCH0 represents the estimate for the parameter $\omega$, ARCH1 represents $\alpha_1$, and GARCH1 represents $\gamma_1$.

The following statements transform the estimated conditional error variance series VHA T to the estimated standard deviation series SHA T. Then, they plot SHA T together with the true standard deviation $S$ used to generate the simulated data.

```sas
data out;
  set out;
  shat = sqrt( vhat );
run;

title "Predicted and Actual Standard Deviations";
proc gplot data=out;
  plot s*time=1 shat*time=2 / overlay;
  symbol1 v=dot i=none;
  symbol2 v=none i = join;
run;
```

The plot is shown in Figure 8.13.
Part 2. General Information

Figure 8.13. Estimated and Actual Error Standard Deviation Series

Note that in this example the form of heteroscedasticity used in generating the simulated series $Y$ does not fit the GARCH model. The GARCH model assumes \textit{conditional} heteroscedasticity, with homoscedastic unconditional error variance. That is, the GARCH model assumes that the changes in variance are a function of the realizations of preceding errors and that these changes represent temporary and random departures from a constant unconditional variance. The data generating process used to simulate series $Y$, contrary to the GARCH model, has exogenous unconditional heteroscedasticity that is independent of past errors.

Nonetheless, as shown in Figure 8.13, the GARCH model does a reasonably good job of approximating the error variance in this example, and some improvement in the efficiency of the estimator of the regression parameters can be expected.

The GARCH model may perform better in cases where theory suggests that the data generating process produces true autoregressive conditional heteroscedasticity. This is the case in some economic theories of asset returns, and GARCH-type models are often used for analysis of financial markets data.

\textbf{Using the HETERO Statement with GARCH Models}

The HETERO statement can be combined with the GARCH= option on the MODEL statement to include input variables in the GARCH conditional variance model. For example, the GARCH(1,1) variance model with two dummy input variables $D1$ and $D2$ is

$$
\epsilon_t = \sqrt{h_t} \epsilon_t
$$

$$
h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 h_{t-1} + \eta_1 D_{1t} + \eta_2 D_{2t}
$$
The following statements estimate this GARCH model:

```sas
proc autoreg data=one;
  model y = x z / garch=(p=1, q=1);
  hetero d1 d2;
run;
```

The parameters for the variables D1 and D2 can be constrained using the COEF= option. For example, the constraints $\eta_1 = \eta_2 = 1$ are imposed by the following statements:

```sas
proc autoreg data=one;
  model y = x z / garch=(p=1, q=1);
  hetero d1 d2 / coef=unit;
run;
```

**Limitations of GARCH and Heteroscedasticity Specifications**

When you specify both the GARCH= option and the HETERO statement, the GARCH=(TYPE=EXP) option is not valid. The COVEST= option is not applicable to the EGARCH model.

**EGARCH, IGARCH, GARCH-M Models**

The AUTOREG procedure supports several variations of the generalized conditional heteroscedasticity model.

Using the TYPE= suboption of the GARCH= option, you can control the constraints placed on the estimated GARCH parameters. You can specify unconstrained, non-negativity constrained (default), stationarity constrained, or integration constrained. The integration constraint produces the integrated GARCH or IGARCH model.

You can also use the TYPE= option to specify the exponential form of the GARCH model, called the EGARCH model. The MEAN suboption of the GARCH= option specifies the GARCH-in-mean or GARCH-M model.

The following statements illustrate the use of the TYPE= option to fit an AR(2)-EGARCH(1,1) model to the series Y. (Output is not shown.)

```sas
proc autoreg data=a;
  model y = time / nlag=2 garch=(p=1, q=1, type=exp);
run;
```

See the section "GARCH, IGARCH, EGARCH, and GARCH-M Models" later in this chapter for details.
Syntax

The AUTOREG procedure is controlled by the following statements:

\[
\text{PROC AUTOREG options ;}
\text{BY variables ;}
\text{MODEL dependent = regressors / options ;}
\text{HETERO variables / options ;}
\text{RESTRICT equation , . . . , equation ;}
\text{TEST equation , . . . , equation / option ;}
\text{OUTPUT OUT = SAS data set options ;}
\]

At least one MODEL statement must be specified. One OUTPUT statement can follow each MODEL statement. One HETERO statement can follow each MODEL statement.

Functional Summary

The statements and options used with the AUTOREG procedure are summarized in the following table:

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set Options</td>
<td>AUTOREG</td>
<td>DATA=</td>
</tr>
<tr>
<td>specify the input data set</td>
<td></td>
<td>OUTTEST=</td>
</tr>
<tr>
<td>write parameter estimates to an output data set</td>
<td>AUTOREG</td>
<td>COVOUT</td>
</tr>
<tr>
<td>include covariances in the OUTTEST= data set</td>
<td>OUTPUT</td>
<td>OUT=</td>
</tr>
<tr>
<td>write predictions, residuals, and confidence limits to an output data set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Declaring the Role of Variables</td>
<td>BY</td>
<td></td>
</tr>
<tr>
<td>specify BY-group processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Printing Control Options</td>
<td>MODEL</td>
<td>ALL</td>
</tr>
<tr>
<td>request all printing options</td>
<td>MODEL</td>
<td>COEF</td>
</tr>
<tr>
<td>print transformed coefficients</td>
<td>MODEL</td>
<td>CORRB</td>
</tr>
<tr>
<td>print correlation matrix of the estimates</td>
<td>MODEL</td>
<td>COVB</td>
</tr>
<tr>
<td>print covariance matrix of the estimates</td>
<td>MODEL</td>
<td>DW=j</td>
</tr>
<tr>
<td>print DW statistics up to order j</td>
<td>MODEL</td>
<td>DWPROB</td>
</tr>
<tr>
<td>print marginal probability of the generalized Durbin-Watson test statistics for large sample sizes</td>
<td>MODEL</td>
<td>LDW</td>
</tr>
<tr>
<td>print the ( p )-values for the Durbin-Watson test be computed using a linearized approximation of the design matrix</td>
<td>MODEL</td>
<td>GINV</td>
</tr>
</tbody>
</table>
### Description

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>print the Godfrey LM serial correlation test</td>
<td>MODEL</td>
<td>GODFREY=</td>
</tr>
<tr>
<td>print details at each iteration step</td>
<td>MODEL</td>
<td>ITPRINT</td>
</tr>
<tr>
<td>print the Durbin $t$ statistic</td>
<td>MODEL</td>
<td>LAGDEP</td>
</tr>
<tr>
<td>print the Durbin $h$ statistic</td>
<td>MODEL</td>
<td>LAGDEP=</td>
</tr>
<tr>
<td>print the log likelihood value of the regression model</td>
<td>MODEL</td>
<td>LOGLIKL</td>
</tr>
<tr>
<td>print the Jarque-Bera normality test</td>
<td>MODEL</td>
<td>NORMAL</td>
</tr>
<tr>
<td>print tests for ARCH process</td>
<td>MODEL</td>
<td>ARCHTEST</td>
</tr>
<tr>
<td>print the Lagrange multiplier test</td>
<td>HETERO</td>
<td>TEST=LM</td>
</tr>
<tr>
<td>print the Chow test</td>
<td>MODEL</td>
<td>CHOW=</td>
</tr>
<tr>
<td>print the predictive Chow test</td>
<td>MODEL</td>
<td>PCHOW=</td>
</tr>
<tr>
<td>suppress printed output</td>
<td>MODEL</td>
<td>NOPRINT</td>
</tr>
<tr>
<td>print partial autocorrelations</td>
<td>MODEL</td>
<td>PARTIAL</td>
</tr>
<tr>
<td>print Ramsey’s RESET test</td>
<td>MODEL</td>
<td>RESET</td>
</tr>
<tr>
<td>print tests for stationarity or unit roots</td>
<td>MODEL</td>
<td>STATIONARITY=(PHILLIPS=)</td>
</tr>
<tr>
<td>print tests of linear hypotheses</td>
<td>TEST</td>
<td>TYPE=</td>
</tr>
<tr>
<td>specify the test statistics to use</td>
<td>MODEL</td>
<td>URSQ</td>
</tr>
<tr>
<td>prints the uncentered regression $R^2$</td>
<td>MODEL</td>
<td>URSQ</td>
</tr>
</tbody>
</table>

### Model Estimation Options

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>specify the order of autoregressive process</td>
<td>MODEL</td>
<td>NLAG=</td>
</tr>
<tr>
<td>center the dependent variable</td>
<td>MODEL</td>
<td>CENTER</td>
</tr>
<tr>
<td>suppress the intercept parameter</td>
<td>MODEL</td>
<td>NOINT</td>
</tr>
<tr>
<td>remove nonsignificant AR parameters</td>
<td>MODEL</td>
<td>BACKSTEP</td>
</tr>
<tr>
<td>specify significance level for BACKSTEP</td>
<td>MODEL</td>
<td>SLSTAY=</td>
</tr>
<tr>
<td>specify the convergence criterion</td>
<td>MODEL</td>
<td>CONVERGE=</td>
</tr>
<tr>
<td>specify the type of covariance matrix</td>
<td>MODEL</td>
<td>COVEST=</td>
</tr>
<tr>
<td>set the initial values of parameters used by the iterative optimization algorithm</td>
<td>MODEL</td>
<td>INITIAL=</td>
</tr>
<tr>
<td>specify iterative Yule-Walker method</td>
<td>MODEL</td>
<td>ITER</td>
</tr>
<tr>
<td>specify maximum number of iterations</td>
<td>MODEL</td>
<td>MAXITER=</td>
</tr>
<tr>
<td>specify the estimation method</td>
<td>MODEL</td>
<td>METHOD=</td>
</tr>
<tr>
<td>use only first sequence of nonmissing data</td>
<td>MODEL</td>
<td>NOMISS</td>
</tr>
<tr>
<td>specify the optimization technique</td>
<td>MODEL</td>
<td>OPTMETHOD=</td>
</tr>
<tr>
<td>imposes restrictions on the regression estimates</td>
<td>RESTRICT</td>
<td></td>
</tr>
<tr>
<td>estimate and test heteroscedasticity models</td>
<td>HETERO</td>
<td></td>
</tr>
</tbody>
</table>

### GARCH Related Options

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>specify order of GARCH process</td>
<td>MODEL</td>
<td>GARCH=(Q,P=)</td>
</tr>
<tr>
<td>specify type of GARCH model</td>
<td>MODEL</td>
<td>GARCH=(...TYPE=)</td>
</tr>
<tr>
<td>specify various forms of the GARCH-M model</td>
<td>MODEL</td>
<td>GARCH=(...MEAN=)</td>
</tr>
<tr>
<td>suppress GARCH intercept parameter</td>
<td>MODEL</td>
<td>GARCH=(...NOINT)</td>
</tr>
</tbody>
</table>
### Description

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>specify the trust region method</td>
<td>MODEL</td>
<td>GARCH=(.,TR)</td>
</tr>
<tr>
<td>estimate the GARCH model for the conditional t-distribution</td>
<td>MODEL</td>
<td>GARCH=(.,DIST=)</td>
</tr>
<tr>
<td>estimates the start-up values for the conditional variance equation</td>
<td>MODEL</td>
<td>GARCH=(.,STARTUP=)</td>
</tr>
<tr>
<td>specify the functional form of the heteroscedasticity model</td>
<td>HETERO</td>
<td>LINK=</td>
</tr>
<tr>
<td>specify that the heteroscedasticity model does not include the unit term</td>
<td>HETERO</td>
<td>NOCONST</td>
</tr>
<tr>
<td>impose constraints on the estimated parameters the heteroscedasticity model</td>
<td>HETERO</td>
<td>COEF=</td>
</tr>
<tr>
<td>impose constraints on the estimated standard deviation of the heteroscedasticity model</td>
<td>HETERO</td>
<td>STD=</td>
</tr>
<tr>
<td>output conditional error variance</td>
<td>OUTPUT</td>
<td>CEV=</td>
</tr>
<tr>
<td>output conditional prediction error variance</td>
<td>OUTPUT</td>
<td>CPEV=</td>
</tr>
<tr>
<td>specify the flexible conditional variance form of the GARCH model</td>
<td>HETERO</td>
<td></td>
</tr>
</tbody>
</table>

### Output Control Options

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>specify confidence limit size</td>
<td>OUTPUT</td>
<td>ALPHACL=</td>
</tr>
<tr>
<td>specify confidence limit size for structural predicted values</td>
<td>OUTPUT</td>
<td>ALPHACL=</td>
</tr>
<tr>
<td>specify the significance level for the upper and lower bounds of the CUSUM and CUSUMSQ statistics</td>
<td>OUTPUT</td>
<td>ALPHAC=</td>
</tr>
<tr>
<td>specify the name of a variable to contain the values of the Theil’s BLUS residuals</td>
<td>OUTPUT</td>
<td>BLUS=</td>
</tr>
<tr>
<td>output the value of the error variance $\sigma_t^2$</td>
<td>OUTPUT</td>
<td>CEV=</td>
</tr>
<tr>
<td>output transformed intercept variable</td>
<td>OUTPUT</td>
<td>CONSTANT=</td>
</tr>
<tr>
<td>specify the name of a variable to contain the CUSUM statistics</td>
<td>OUTPUT</td>
<td>CUSUM=</td>
</tr>
<tr>
<td>specify the name of a variable to contain the CUSUMSQ statistics</td>
<td>OUTPUT</td>
<td>CUSUMSQ=</td>
</tr>
<tr>
<td>specify the name of a variable to contain the upper confidence bound for the CUSUM statistic</td>
<td>OUTPUT</td>
<td>CUSUMUB=</td>
</tr>
<tr>
<td>specify the name of a variable to contain the lower confidence bound for the CUSUM statistic</td>
<td>OUTPUT</td>
<td>CUSUMLB=</td>
</tr>
<tr>
<td>specify the name of a variable to contain the upper confidence bound for the CUSUMSQ statistic</td>
<td>OUTPUT</td>
<td>CUSUMSQUB=</td>
</tr>
</tbody>
</table>
Chapter 8. Syntax

### Description Statement Option

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>option specify the name of a variable to contain the lower confidence bound for the CUSUMSQ statistic</td>
<td>OUTPUT</td>
<td>CUSUMSQLB=</td>
</tr>
<tr>
<td>output lower confidence limit</td>
<td>OUTPUT</td>
<td>LCL=</td>
</tr>
<tr>
<td>output lower confidence limit for structural predicted values</td>
<td>OUTPUT</td>
<td>LCLM=</td>
</tr>
<tr>
<td>output predicted values</td>
<td>OUTPUT</td>
<td>P=</td>
</tr>
<tr>
<td>output predicted values of structural part</td>
<td>OUTPUT</td>
<td>PM=</td>
</tr>
<tr>
<td>output residuals</td>
<td>OUTPUT</td>
<td>R=</td>
</tr>
<tr>
<td>output residuals from structural predictions</td>
<td>OUTPUT</td>
<td>RM=</td>
</tr>
<tr>
<td>specify the name of a variable to contain the part of the predictive error variance (( v_t ))</td>
<td>OUTPUT</td>
<td>RECPEV=</td>
</tr>
<tr>
<td>specify the name of a variable to contain recursive residuals</td>
<td>OUTPUT</td>
<td>RECRES=</td>
</tr>
<tr>
<td>output transformed variables</td>
<td>OUTPUT</td>
<td>comp TRANSFORM=</td>
</tr>
<tr>
<td>output upper confidence limit</td>
<td>OUTPUT</td>
<td>UCL=</td>
</tr>
<tr>
<td>output upper confidence limit for structural predicted values</td>
<td>OUTPUT</td>
<td>UCLM=</td>
</tr>
</tbody>
</table>

### PROC AUTOREG Statement

```
PROC AUTOREG options;
```

The following options can be used in the PROC AUTOREG statement:

**DATA= SAS-data-set**

specifies the input SAS data set. If the DATA= option is not specified, PROC AUTOREG uses the most recently created SAS data set.

**OUTEST= SAS-data-set**

writes the parameter estimates to an output data set. See "OUTEST= Data Set" later in this chapter for information on the contents of this data set.

**COVOUT**

writes the covariance matrix for the parameter estimates to the OUTTEST= data set. This option is valid only if the OUTEST= option is specified.

In addition, any of the following MODEL statement options can be specified in the PROC AUTOREG statement, which is equivalent to specifying the option for every MODEL statement: ALL, ARCHTEST, BACKSTEP, CENTER, COEF, CONVERGE=, CORRB, COVB, DW=, DWPROB, GINV, ITER, ITPRINT, MAXITER=, METHOD=, NOINT, NOMISS, NOPRINT, and PARTIAL.
Part 2. General Information

BY Statement

BY variables;

A BY statement can be used with PROC AUTOREG to obtain separate analyses on observations in groups defined by the BY variables.

MODEL Statement

MODEL dependent = regressors / options ;

The MODEL statement specifies the dependent variable and independent regressor variables for the regression model. If no independent variables are specified in the MODEL statement, only the mean is fitted. (This is a way to obtain autocorrelations of a series.)

Models can be given labels of up to eight characters. Model labels are used in the printed output to identify the results for different models. The model label is specified as follows:

label : MODEL ... ;

The following options can be used in the MODEL statement after a slash (/).

CENTER
centers the dependent variable by subtracting its mean and suppresses the intercept parameter from the model. This option is only valid when the model does not have regressors (explanatory variables).

NOINT
suppresses the intercept parameter.

Autoregressive Error Options

NLAG= number
NLAG= ( number-list )
specifies the order of the autoregressive error process or the subset of autoregressive error lags to be fitted. Note that NLAG=3 is the same as NLAG=(1 2 3). If the NLAG= option is not specified, PROC AUTOREG does not fit an autoregressive model.

GARCH Estimation Options

GARCH= ( option-list )
Specifies a GARCH-type conditional heteroscedasticity model. The GARCH= option in the MODEL statement specifies the family of ARCH models to be estimated. The GARCH(1,1) regression model is specified in the following statement:

model y = x1 x2 / garch=(q=1,p=1);
When you want to estimate the subset of ARCH terms, for example, ARCH(1 3), you can write the SAS statement as follows:

```sas
model y = x1 x2 / garch=(q=(1 3));
```

With the TYPE= option, you can specify various GARCH models. The IGARCH(2,1) model without trend in variance is estimated as follows:

```sas
model y = / garch=(q=2,p=1,type=integ,noint);
```

The following options can be used in the GARCH=() option. The options are listed within parentheses and separated by commas.

**Q=** *number*

*Q= (number-list)*

specifies the order of the process or the subset of ARCH terms to be fitted.

**P=** *number*

*P= (number-list)*

specifies the order of the process or the subset of GARCH terms to be fitted. If only the P= option is specified, Q=1 is assumed.

**TYPE=** *value*

specifies the type of GARCH model. The values of the TYPE= option are

- EXP specifies the exponential GARCH or EGARCH model.
- INTEGRATED specifies the integrated GARCH or IGARCH model.
- NELSON | NELSONCAO specifies the Nelson-Cao inequality constraints.
- NONNEG specifies the GARCH model with nonnegativity constraints.
- STATIONARY constrains the sum of GARCH coefficients to be less than 1.

The default is TYPE=NELSON.

**MEAN=** *value*

specifies the functional form of the GARCH-M model. The values of the MEAN= option are

- LINEAR specifies the linear function.
  
  \[
  y_t = x_t'\beta + \delta h_t + \epsilon_t 
  \]

- LOG specifies the log function.
  
  \[
  y_t = (\ln h_t) + \epsilon_t 
  \]

- SQRT specifies the square root function.
  
  \[
  y_t = \sqrt{h_t} \epsilon_t + \epsilon_t
  \]
Part 2. General Information

**DIST= value**
specifies the distribution assumed for the error term. The values of the DIST= option are

T specifies Student’s t distribution.
NORMAL specifies the standard normal distribution. The default is DIST=NORMAL.

**NOINT**
suppresses the intercept parameter in the conditional variance model. This option is valid only with the TYPE=INTEG option.

**STARTUP= MSE | ESTIMATE**
STARTUP=ESTIMATE requests that the positive constant $c$ for the start-up values of the GARCH conditional error variance process be estimated. By default or if STARTUP=MSE is specified, the value of the mean squared error is used as the default constant.

**TR**
uses the trust region method for GARCH estimation. This algorithm is numerically stable, though computation is expensive. The double quasi-Newton method is the default.

**Printing Options**

**ALL**
requests all printing options.

**ARCHTEST**
requests the Q and LM statistics testing for the absence of ARCH effects.

**CHOW= (obs1,...,obsn)**
The CHOW= option computes Chow tests to evaluate the stability of the regression coefficient. The Chow test is also called the analysis-of-variance test.

**COEF**
prints the transformation coefficients for the first $p$ observations. These coefficients are formed from a scalar multiplied by the inverse of the Cholesky root of the Toeplitz matrix of autocovariances.

**CORRB**
prints the estimated correlations of the parameter estimates.

**COVB**
prints the estimated covariances of the parameter estimates.

**COVEST= OP | HESSIAN | QML**
The COVEST= option specifies the type of covariance matrix for the GARCH or heteroscedasticity model. When COVEST=OP is specified, the outer product matrix is used to compute the covariance matrix of the parameter estimates. The COVEST=HESSIAN option produces the covariance matrix using the Hessian matrix. The quasi-maximum likelihood estimates are computed with COVEST=QML. The default is COVEST=OP.
**DW= n**
prints Durbin-Watson statistics up to the order \( n \). The default is DW=1. When the LAGDEP option is specified, the Durbin-Watson statistic is not printed unless the DW= option is explicitly specified.

**DWPROB**
The DWPROB option now produces \( p \)-values for the generalized Durbin-Watson test statistics for large sample sizes. Previously, the Durbin-Watson probabilities were calculated only for small sample sizes. The new method of calculating Durbin-Watson probabilities is based on the algorithm of Ansley, Kohn, and Shively (1992).

**GINV**
prints the inverse of the Toeplitz matrix of autocovariances for the Yule-Walker solution. See "Computational Methods" later in this chapter for details.

**GODFREY**
**GODFREY= r**
The GODFREY option produces Godfrey’s general Lagrange multiplier test against ARMA errors.

**ITPRINT**
prints the objective function and parameter estimates at each iteration. The objective function is the full log likelihood function for the maximum likelihood method, while the error sum of squares is produced as the objective function of unconditional least squares. For the ML method, the ITPRINT option prints the value of the full log likelihood function, not the concentrated likelihood.

**LAGDEP**
**LAGDV**
prints the Durbin \( t \) statistic, which is used to detect residual autocorrelation in the presence of lagged dependent variables. See "Generalized Durbin-Watson Tests" later in this chapter for details.

**LAGDEP= name**
**LAGDV= name**
prints the Durbin \( h \) statistic for testing the presence of first-order autocorrelation when regressors contain the lagged dependent variable whose name is specified as LAGDEP= name. If the Durbin \( h \) statistic cannot be computed, the asymptotically equivalent \( t \) statistic is printed instead. See "Generalized Durbin-Watson Tests" for details.

When the regression model contains several lags of the dependent variable, specify the lagged dependent variable for the smallest lag in the LAGDEP= option, for example,

\[
\text{model } y = x_1 \ x_2 \ ylag2 \ ylag3 / \text{lagdep}=ylag2;
\]

**LOGLIK**
The LOGLIK option prints the log likelihood value of the regression model, assuming normally distributed errors.
Part 2. General Information

NOPRINT
suppresses all printed output.

NORMAL
The NORMAL option specifies the Jarque-Bera’s normality test statistic for regression residuals.

PARTIAL
prints partial autocorrelations.

PCHOW= (obs1 ... obsn)
The PCHOW= option computes the predictive Chow test. The form of the PCHOW= option is the same as the CHOW= option; see the discussion of the CHOW= option earlier in this chapter.

RESET
The RESET option produces Ramsey’s RESET test statistics. The RESET option tests the null model

\[ y_t = \mathbf{x}_t \beta + u_t \]

against the alternative

\[ y_t = \mathbf{x}_t \beta + \sum_{j=2}^{p} \phi_j \hat{y}_t^j + u_t \]

where \( \hat{y}_t \) is the predicted value from the OLS estimation of the null model. The RESET option produces three RESET test statistics for \( p = 2, 3, \) and 4.

STATIONARITY= (PHILLIPS)
STATIONARITY= (PHILLIPS=(value ... value))
The STATIONARITY= option specifies tests of stationarity or unit roots. The STATIONARITY= option provides Phillips-Perron tests.

The PHILLIPS or PHILLIPS= suboption of the STATIONARITY= option produces the Phillips-Perron unit root test when there are no regressors in the MODEL statement. When the model includes regressors, the PHILLIPS option produces the Phillips-Ouliaris cointegration test. The PHILLIPS option can be abbreviated as PP.

The PHILLIPS option performs the Phillips-Perron test for three null hypothesis cases: zero mean, single mean, and deterministic trend. For each case, the PHILLIPS option computes two test statistics, \( Z(\hat{\alpha}) \) and \( Z(t_\alpha) \), and reports their \( p \)-values. These test statistics have the same limiting distributions as the corresponding Dickey-Fuller tests.

The three types of the Phillips-Perron unit root test reported by the PHILLIPS option are as follows.
Zero Mean computes the Phillips-Perron test statistic based on the zero mean autoregressive model

\[ y_t = \alpha y_{t-1} + u_t \]

Single Mean computes the Phillips-Perron test statistic based on the autoregressive model with a constant term

\[ y_t = \mu + \alpha y_{t-1} + u_t \]

Trend computes the Phillips-Perron test statistic based on the autoregressive model with constant and time trend terms

\[ y_t = \mu + \alpha y_{t-1} + \delta t + u_t \]

You can specify several truncation points \( l \) for weighted variance estimators using the PHILLIPS=(l1:ln) specification.

**URSQ**

The URSQ option prints the uncentered regression \( R^2 \). The uncentered regression \( R^2 \) is useful to compute Lagrange multiplier test statistics, since most LM test statistics are computed as \( T^*\text{URSQ} \), where \( T \) is the number of observations used in estimation.

**Stepwise Selection Options**

**BACKSTEP**

removes insignificant autoregressive parameters. The parameters are removed in order of least significance. This backward elimination is done only once on the Yule-Walker estimates computed after the initial ordinary least-squares estimation. The BACKSTEP option can be used with all estimation methods since the initial parameter values for other estimation methods are estimated using the Yule-Walker method.

**SLSTAY=** value

specifies the significance level criterion to be used by the BACKSTEP option. The default is SLSTAY=.05.

**Estimation Control Options**

**CONVERGE=** value

specifies the convergence criterion. If the maximum absolute value of the change in the autoregressive parameter estimates between iterations is less than this amount, then convergence is assumed. The default is CONVERGE=.001.

**INITIAL=** ( initial-values )

**START=** ( initial-values )

The INITIAL= option specifies initial values for some or all of the parameter estimates. The values specified are assigned to model parameters in the same order as the parameter estimates are printed in the AUTOREG procedure output. The order of values in the INITIAL= or START= option is: the intercept, the regressor coefficients, the autoregressive parameters, the ARCH parameters, the GARCH parameters, the inverted degrees of freedom for Student’s \( t \) distribution, the start-up value
for conditional variance, and the heteroscedasticity model parameters \( \eta \) specified by the HETERO statement.

The following is an example of specifying initial values for an AR(1)-GARCH(1,1) model with regressors X1 and X2:

```sas
model y = w x / nlag=1 garch=(p=1,q=1)
   initial=(1 1 .5 .8 1 .6);
```

The model specified by this MODEL statement is

\[
y_t = \beta_0 + \beta_1 w_t + \beta_2 x_t + \nu_t
\]

\[
\nu_t = \epsilon_t - \phi_1 \nu_{t-1}
\]

\[
\epsilon_t = \sqrt{h_t} \epsilon_t
\]

\[
h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 h_{t-1}
\]

\[
\epsilon_t \sim N(0, \sigma^2_t)
\]

The initial values for the regression parameters, INTERCEP \( \beta_0 \), X1 \( \beta_1 \), and X2 \( \beta_2 \), are specified as 1. The initial value of the AR(1) coefficient \( \phi_1 \) is specified as 0.5. The initial value of ARCH0 \( \omega \) is 0.8, the initial value of ARCH1 \( \alpha_1 \) is 0.1, and the initial value of GARCH1 \( \gamma_1 \) is 0.6.

When you use the RESTRICT statement, the initial values specified by the INITIAL= option should satisfy the restrictions specified for the parameter estimates. If they do not, the initial values you specify are adjusted to satisfy the restrictions.

**LDW**

The LDW option specifies that p-values for the Durbin-Watson test be computed using a linearized approximation of the design matrix when the model is nonlinear due to the presence of an autoregressive error process. (The Durbin-Watson tests of the OLS linear regression model residuals are not affected by the LDW option.) Refer to White (1992) for Durbin-Watson testing of nonlinear models.

**MAXITER= number**

sets the maximum number of iterations allowed. The default is MAXITER=50.

**METHOD= value**

requests the type of estimates to be computed. The values of the METHOD= option are

- METHOD=ML specifies maximum likelihood estimates
- METHOD=ULS specifies unconditional least-squares estimates
- METHOD=YW specifies Yule-Walker estimates
- METHOD=ITYW specifies iterative Yule-Walker estimates
If the GARCH= or LAGDEP option is specified, the default is METHOD=ML. Otherwise, the default is METHOD=YW.

**NOMISS**
requests the estimation to the first contiguous sequence of data with no missing values. Otherwise, all complete observations are used.

**OPTMETHOD= QN | TR**
The OPTMETHOD= option specifies the optimization technique when the GARCH or heteroscedasticity model is estimated. The OPTMETHOD=QN option specifies the quasi-Newton method. The OPTMETHOD=TR option specifies the trust region method. The default is OPTMETHOD=QN.

**HETERO Statement**
The HETERO statement specifies variables that are related to the heteroscedasticity of the residuals and the way these variables are used to model the error variance of the regression.

The syntax of the HETERO statement is

```
HETERO variables / options ;
```

The heteroscedastic regression model supported by the HETERO statement is

\[
y_t = x_t \hat{\beta} + \epsilon_t
\]

\[
\epsilon_t \sim N(0, \sigma_t^2)
\]

\[
\sigma_t^2 = \sigma^2 h_t
\]

\[
h_t = l(z_t' \eta)
\]

The HETERO statement specifies a model for the conditional variance \( h_t \). The vector \( z_t \) is composed of the variables listed on the HETERO statement, \( \eta \) is a parameter vector, and \( l(\cdot) \) is a link function that depends on the value of the LINK= option.

The keyword XBETA can be used in the variables list to refer to the model predicted value \( x_t' \hat{\beta} \).

The errors \( \epsilon_t \) are assumed to be uncorrelated – the heteroscedasticity models specified by the HETERO statement cannot be combined with an autoregressive model for the errors. Thus, the HETERO statement cannot be used if the NLAG= option is specified in the MODEL statement.

You can specify the following options in the HETERO statement:
**LINK= value**

The LINK= option specifies the functional form of the heteroscedasticity model. If you want to estimate the GARCH model whose conditional error variance contains exogenous variables, you do not need to specify the LINK= option. The default is LINK=EXP. Values of the LINK= option are

- **EXP** specifies the exponential link function. The following model is estimated when you specify LINK=EXP:

  \[ h_t = \exp(z_t' \eta) \]

- **SQUARE** specifies the square link function. The following model is estimated when you specify LINK=SQUARE:

  \[ h_t = (1 + z_t' \eta)^2 \]

- **LINEAR** specifies the linear function; that is, the HETERO statement variables predict the error variance linearly. The following model is estimated when you specify LINK=LINEAR:

  \[ h_t = (1 + z_t' \eta) \]

**COEF= value**

The COEF= option imposes constraints on the estimated parameters \( \eta \) of the heteroscedasticity model. The values of the COEF= option are

- **NONNEG** specifies that the estimated heteroscedasticity parameters \( \eta \) must be nonnegative. When the HETERO statement is used in conjunction with the GARCH= option, the default is COEF=NONNEG.
- **UNIT** constrains all heteroscedasticity parameters \( \eta \) to equal 1.
- **ZERO** constrains all heteroscedasticity parameters \( \eta \) to equal 0.
- **UNREST** specifies unrestricted estimation of \( \eta \). When the GARCH= option is not specified, the default is COEF=UNREST.

**STD= value**

The STD= option imposes constraints on the estimated standard deviation \( \sigma \) of the heteroscedasticity model. The values of the STD= option are

- **NONNEG** specifies that the estimated standard deviation parameter \( \sigma \) must be nonnegative.
- **UNIT** constrains the standard deviation parameter \( \sigma \) to equal 1.
- **UNREST** specifies unrestricted estimation of \( \sigma \). This is the default.
TEST= **LM**
The TEST=LM option produces a Lagrange multiplier test for heteroscedasticity. The null hypothesis is homoscedasticity; the alternative hypothesis is heteroscedasticity of the form specified by the HETERO statement. The power of the test depends on the variables specified in the HETERO statement.

The test may give different results depending on the functional form specified by the LINK= option. However, in many cases the test does not depend on the LINK= option. The test is invariant to the form of \( h_t \) when \( h_t(0) = 1 \) and \( h_t(0) \neq 0 \). (The condition \( h_t(0) = 1 \) is satisfied except when the NOCONST option is specified with LINK=SQUARE or LINK=LINEAR.)

**NOCONST**
The NOCONST option specifies that the heteroscedasticity model does not include the unit term for the LINK=SQUARE and LINK=LINEAR options. For example, the following model is estimated when you specify the options LINK=SQUARE NOCONST:

\[
h_t = (z_t^t \eta)^2
\]

**RESTRICT Statement**
The RESTRICT statement provides constrained estimation.

```
RESTRICT equation , ... , equation ;
```

The RESTRICT statement places restrictions on the parameter estimates for covariates in the preceding MODEL statement. Any number of RESTRICT statements can follow a MODEL statement. Several restrictions can be specified in a single RESTRICT statement by separating the individual restrictions with commas.

Each restriction is written as a linear equation composed of constants and parameter names. Refer to model parameters by the name of the corresponding regressor variable. Each name used in the equation must be a regressor in the preceding MODEL statement. Use the keyword INTERCEPT to refer to the intercept parameter in the model.

The following is an example of a RESTRICT statement:

```
model y = a b c d;
restrict a+b=0, 2*d-c=0;
```

When restricting a linear combination of parameters to be 0, you can omit the equal sign. For example, the following RESTRICT statement is equivalent to the preceding example:

```
restrict a+b, 2*d-c;
```
The following RESTRICT statement constrains the parameters estimates for three regressors (X1, X2, and X3) to be equal:

```
restrict x1 = x2, x2 = x3;
```

The preceding restriction can be abbreviated as follows.

```
restrict x1 = x2 = x3;
```

Only simple linear combinations of parameters can be specified in RESTRICT statement expressions; complex expressions involving parentheses, division, functions, or complex products are not allowed.

**TEST Statement**

The AUTOREG procedure now supports a TEST statement for linear hypothesis tests.

```
TEST equation , ... , equation / option ;
```

The TEST statement tests hypotheses about the covariates in the model estimated by the preceding MODEL statement. Each equation specifies a linear hypothesis to be tested. If more than one equation is specified, the equations are separated by commas. Each test is written as a linear equation composed of constants and parameter names. Refer to parameters by the name of the corresponding regressor variable. Each name used in the equation must be a regressor in the preceding MODEL statement. Use the keyword INTERCEPT to refer to the intercept parameter in the model.

You can specify the following option in the TEST statement:

**TYPE=** value

The TYPE= option specifies the test statistics to use, F or Wald. TYPE=F produces an F-test. TYPE=WALD produces a Wald test. The default is TYPE=F.

The following example of a TEST statement tests the hypothesis that the coefficients of two regressors A and B are equal:

```
model y = a b c d;
    test a = b;
```

To test separate null hypotheses, use separate TEST statements. To test a joint hypothesis, specify the component hypotheses on the same TEST statement, separated by commas.

For example, consider the following linear model:

```
y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t
```
The following statements test the two hypotheses $H_0 : \beta_0 = 1$ and $H_0 : \beta_1 + \beta_2 = 0$:

```sas
model y = x1 x2;
test intercept = 1;
test x1 + x2 = 0;
```

The following statements test the joint hypothesis $H_0 : \beta_0 = 1$ and $\beta_1 + \beta_2 = 0$:

```sas
model y = x1 x2;
test intercept = 1, x1 + x2 = 0;
```

### OUTPUT Statement

```
OUTPUT OUT= SAS-data-set keyword = options . . . ;
```

The OUTPUT statement creates an output SAS data set as specified by the following options:

- **OUT= SAS-data-set**
  names the output SAS data set containing the predicted and transformed values. If the OUT= option is not specified, the new data set is named according to the DATA convention.

- **ALPHACL= number**
  sets the confidence limit size for the estimates of future values of the response time series. The ALPHACL= value must be between 0 and 1. The resulting confidence interval has $1 - \text{number}$ confidence. The default is ALPHACL=.05, corresponding to a 95% confidence interval.

- **ALPHACLM= number**
  sets the confidence limit size for the estimates of the structural or regression part of the model. The ALPHACLM= value must be between 0 and 1. The resulting confidence interval has $1 - \text{number}$ confidence. The default is ALPHACLM=.05, corresponding to a 95% confidence interval.

- **ALPHACSM= .01 | .05 | .10**
  The ALPHACSM= option specifies the significance level for the upper and lower bounds of the CUSUM and CUSUMSQ statistics output by the CUSUMLB=, CUSUMUB=, CUSUMSQLB=, and CUSUMSQUB= options. The significance level specified by the ALPHACSM= option can be .01, .05, or .10. Other values are not supported.

The following options are of the form **KEYWORD=name**, where **KEYWORD** specifies the statistic to include in the output data set and **name** gives the name of the variable in the OUT= data set containing the statistic.

- **BLUS= variable**
  The BLUS= option specifies the name of a variable to contain the values of the Theil’s BLUS residuals. Refer to Theil (1971) for more information on BLUS residuals.
Part 2. General Information

**CEV= variable**

The CEV= option writes to the output data set the value of the error variance \( \sigma_t^2 \) from the heteroscedasticity model specified by the HETERO statement or the value of the conditional error variance \( h_t \) by the GARCH= option in the MODEL statement.

**HT= variable**

The HT= variable writes the conditional prediction error variance to the output data set. The value of conditional prediction error variance is equal to that of the conditional error variance when there are no autoregressive parameters. For the exponential GARCH model, conditional prediction error variance cannot be calculated. See "Predicted Values" later in this chapter for details.

**CPEV= variable**

writes the conditional prediction error variance to the output data set. The value of conditional prediction error variance is equal to that of the conditional error variance when there are no autoregressive parameters. For the exponential GARCH model, conditional prediction error variance cannot be calculated. See "Predicted Values" later in this chapter for details.

**CONSTANT= variable**

writes the transformed intercept to the output data set. The details of the transformation are described in "Computational Methods" later in this chapter.

**CUSUM= variable**

The CUSUM= option specifies the name of a variable to contain the CUSUM statistics.

**CUSUMSQ= variable**

The CUSUMSQ= option specifies the name of a variable to contain the CUSUMSQ statistics.

**CUSUMUB= variable**

The CUSUMUB= option specifies the name of a variable to contain the upper confidence bound for the CUSUM statistic.

**CUSUMLB= variable**

The CUSUMLB= option specifies the name of a variable to contain the lower confidence bound for the CUSUM statistic.

**CUSUMSQUB= variable**

The CUSUMSQUB= option specifies the name of a variable to contain the upper confidence bound for the CUSUMSQ statistic.

**CUSUMSQLB= variable**

The CUSUMSQLB= option specifies the name of a variable to contain the lower confidence bound for the CUSUMSQ statistic.

**LCL= name**

writes the lower confidence limit for the predicted value (specified in the PREDICTED= option) to the output data set. The size of the confidence interval is set by the ALPHACL= option. When a GARCH model is estimated, the lower confidence limit is calculated assuming that the disturbances have homoscedastic conditional variance. See "Predicted Values" later in this chapter for details.

**LCLM= name**

writes the lower confidence limit for the structural predicted value (specified in the PREDICTEDM= option) to the output data set under the name given. The size of the confidence interval is set by the ALPHACL= option.
Chapter 8. Syntax

**PREDICTED=** name

**P=** name

writes the predicted values to the output data set. These values are formed from both the structural and autoregressive parts of the model. See "Predicted Values" later in this chapter for details.

**PREDICTEDM=** name

**PM=** name

writes the structural predicted values to the output data set. These values are formed from only the structural part of the model. See "Predicted Values" later in this chapter for details.

**RECPEV=** variable

The RECPEV= option specifies the name of a variable to contain the part of the predictive error variance ($v_t$) that is used to compute the recursive residuals.

**RECRES=** variable

The RECRES= option specifies the name of a variable to contain recursive residuals. The recursive residuals are used to compute the CUSUM and CUSUMSQ statistics.

**RESIDUAL=** name

**R=** name

writes the residuals from the predicted values based on both the structural and time series parts of the model to the output data set.

**RESIDUALM=** name

**RM=** name

writes the residuals from the structural prediction to the output data set.

**TRANSFORM=** variables

transforms the specified variables from the input data set by the autoregressive model and writes the transformed variables to the output data set. The details of the transformation are described in "Computational Methods" later in this chapter. If you need to reproduce the data suitable for reestimation, you must also transform an intercept variable. To do this, transform a variable that is all 1s or use the CONSTANT= option.

**UCL=** name

writes the upper confidence limit for the predicted value (specified in the PREDICTED= option) to the output data set. The size of the confidence interval is set by the ALPHACLI= option. When the GARCH model is estimated, the upper confidence limit is calculated assuming that the disturbances have homoscedastic conditional variance. See "Predicted Values" later in this chapter for details.

**UCLM=** name

writes the upper confidence limit for the structural predicted value (specified in the PREDICTEDM= option) to the output data set. The size of the confidence interval is set by the ALPHACLM= option.
Details

Missing Values

PROC AUTOREG skips any missing values at the beginning of the data set. If the NOMISS option is specified, the first contiguous set of data with no missing values is used; otherwise, all data with nonmissing values for the independent and dependent variables are used. Note, however, that the observations containing missing values are still needed to maintain the correct spacing in the time series. PROC AUTOREG can generate predicted values when the dependent variable is missing.

Autoregressive Error Model

The regression model with autocorrelated disturbances is as follows:

\[ y_t = \mathbf{x}_t\beta + \nu_t \]

\[ \nu_t = \epsilon_t - \varphi_1 \nu_{t-1} - \cdots - \varphi_m \nu_{t-m} \]

\[ \epsilon_t \sim N(0, \sigma^2) \]

In these equations, \( y_t \) are the dependent values, \( \mathbf{x}_t \) is a column vector of regressor variables, \( \beta \) is a column vector of structural parameters, and \( \epsilon_t \) is normally and independently distributed with a mean of 0 and a variance of \( \sigma^2 \). Note that in this parameterization, the signs of the autoregressive parameters are reversed from the parameterization documented in most of the literature.

PROC AUTOREG offers four estimation methods for the autoregressive error model. The default method, Yule-Walker (YW) estimation, is the fastest computationally. The Yule-Walker method used by PROC AUTOREG is described in Gallant and Goebel (1976). Harvey (1981) calls this method the two-step full transform method. The other methods are iterated YW, unconditional least squares (ULS), and maximum likelihood (ML). The ULS method is also referred to as nonlinear least squares (NLS) or exact least squares (ELS).

You can use all of the methods with data containing missing values, but you should use ML estimation if the missing values are plentiful. See the section "Alternative Autocorrelation Correction Methods" later in this chapter for further discussion of the advantages of different methods.

The Yule-Walker Method

Let \( \varphi \) represent the vector of autoregressive parameters

\[ \varphi = (\varphi_1, \varphi_2, \ldots, \varphi_m)' \]

and let the variance matrix of the error vector \( \nu = (\nu_1, \ldots, \nu_N)' \) be \( \Sigma \)

\[ E(\nu\nu') = \Sigma = \sigma^2 \mathbf{V} \]
If the vector of autoregressive parameters $\varphi$ is known, the matrix $V$ can be computed from the autoregressive parameters. $\Sigma$ is then $\sigma^2V$. Given $\Sigma$, the efficient estimates of regression parameters $\beta$ can be computed using generalized least squares (GLS). The GLS estimates then yield the unbiased estimate of the variance $\sigma^2$.

The Yule-Walker method alternates estimation of $\beta$ using generalized least squares with estimation of $\varphi$ using the Yule-Walker equations applied to the sample autocorrelation function. The YW method starts by forming the OLS estimate of $\beta$. Next, $\varphi$ is estimated from the sample autocorrelation function of the OLS residuals using the Yule-Walker equations. Then $V$ is estimated from the estimate of $\varphi$, and $\Sigma$ is estimated from $V$ and the OLS estimate of $\sigma^2$. The autocorrelation corrected estimates of the regression parameters $\beta$ are then computed by GLS using the estimated $\Sigma$ matrix. These are the Yule-Walker estimates.

If the ITER option is specified, the Yule-Walker residuals are used to form a new sample autocorrelation function, the new autocorrelation function is used to form a new estimate of $\varphi$ and $V$, and the GLS estimates are recomputed using the new variance matrix. This alternation of estimates continues until either the maximum change in the $\hat{\varphi}$ estimate between iterations is less than the value specified by the CONVERGE= option or the maximum number of allowed iterations is reached. This produces the Iterated Yule-Walker estimates. Iteration of the estimates may not yield much improvement.

The Yule-Walker equations, solved to obtain $\hat{\varphi}$ and a preliminary estimate of $\sigma^2$, are

$$R\hat{\varphi} = -r$$

Here $r = (r_1, \ldots, r_m)'$, where $r_i$ is the lag $i$ sample autocorrelation. The matrix $R$ is the Toeplitz matrix whose $i,j$th element is $r_{i-j}$. If you specify a subset model, then only the rows and columns of $R$ and $r$ corresponding to the subset of lags specified are used.

If the BACKSTEP option is specified, for purposes of significance testing, the matrix $[R \ r]$ is treated as a sum-of-squares-and-crossproducts matrix arising from a simple regression with $N - k$ observations, where $k$ is the number of estimated parameters.

**The Unconditional Least Squares and Maximum Likelihood Methods**

Define the transformed error, $e$, as

$$e = L^{-1} n$$

where $n = y - X\beta$.

The unconditional sum of squares for the model, $S$, is

$$S = n'V^{-1}n = e'e$$

The ULS estimates are computed by minimizing $S$ with respect to the parameters $\beta$ and $\varphi_i$. 

343
The full log likelihood function for the autoregressive error model is

\[
l = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|V|) - \frac{S}{2\sigma^2}
\]

where $|V|$ denotes determinant of $V$. For the ML method, the likelihood function is maximized by minimizing an equivalent sum-of-squares function.

Maximizing $l$ with respect to $\sigma^2$ (and concentrating $\sigma^2$ out of the likelihood) and dropping the constant term $-\frac{N}{2}\ln(2\pi) + 1 - \ln(N)$ produces the concentrated log likelihood function

\[
l_c = -\frac{N}{2}\ln(S|V|^{1/N})
\]

Rewriting the variable term within the logarithm gives

\[S_{ml} = |L|^{1/N}e'e|L|^{1/N}\]

PROC AUTOREG computes the ML estimates by minimizing the objective function $S_{ml} = |L|^{1/N}e'e|L|^{1/N}$.

The maximum likelihood estimates may not exist for some data sets (Anderson and Mentz 1980). This is the case for very regular data sets, such as an exact linear trend.

**Computational Methods**

**Sample Autocorrelation Function**

The sample autocorrelation function is computed from the structural residuals or noise $n_t = y_t - x'_t\beta$, where $\beta$ is the current estimate of $\beta$. The sample autocorrelation function is the sum of all available lagged products of $n_t$ of order $j$ divided by $\ell + j$, where $\ell$ is the number of such products.

If there are no missing values, then $\ell + j = N$, the number of observations. In this case, the Toeplitz matrix of autocorrelations, $R$, is at least positive semidefinite. If there are missing values, these autocorrelation estimates of $r$ can yield an $R$ matrix that is not positive semidefinite. If such estimates occur, a warning message is printed, and the estimates are tapered by exponentially declining weights until $R$ is positive definite.

**Data Transformation and the Kalman Filter**

The calculation of $V$ from $\varphi$ for the general AR($m$) model is complicated, and the size of $V$ depends on the number of observations. Instead of actually calculating $V$ and performing GLS in the usual way, in practice a Kalman filter algorithm is used to transform the data and compute the GLS results through a recursive process.

In all of the estimation methods, the original data are transformed by the inverse of the Cholesky root of $V$. Let $L$ denote the Cholesky root of $V$, that is $V = LL'$ with $L$ lower triangular. For an AR($m$) model, $L^{-1}$ is a band diagonal matrix with $m$ anomalous rows at the beginning and the autoregressive parameters along the remaining
rows. Thus, if there are no missing values, after the first \( m-1 \) observations the data are transformed as

\[ z_t = x_t + \phi_1 x_{t-1} + \ldots + \phi_m x_{t-m} \]

The transformation is carried out using a Kalman filter, and the lower triangular matrix \( L \) is never directly computed. The Kalman filter algorithm, as it applies here, is described in Harvey and Phillips (1979) and Jones (1980). Although \( L \) is not computed explicitly, for ease of presentation the remaining discussion is in terms of \( L \). If there are missing values, then the submatrix of \( L \) consisting of the rows and columns with nonmissing values is used to generate the transformations.

**Gauss-Newton Algorithms**

The ULS and ML estimates employ a Gauss-Newton algorithm to minimize the sum of squares and maximize the log likelihood, respectively. The relevant optimization is performed simultaneously for both the regression and AR parameters. The OLS estimates of \( \beta \) and the Yule-Walker estimates of \( \varphi \) are used as starting values for these methods.

The Gauss-Newton algorithm requires the derivatives of \( e \) or \( |L|^{1/N} e \) with respect to the parameters. The derivatives with respect to the parameter vector \( \beta \) are

\[ \frac{\partial e}{\partial \beta} = -L^{-1}X \]

\[ \frac{\partial |L|^{1/N} e}{\partial \beta} = -L^{1/N}L^{-1}X \]

These derivatives are computed by the transformation described previously. The derivatives with respect to \( \varphi \) are computed by differentiating the Kalman filter recurrences and the equations for the initial conditions.

**Variance Estimates and Standard Errors**

For the Yule-Walker method, the estimate of the error variance, \( s^2 \), is the error sum of squares from the last application of GLS, divided by the error degrees of freedom (number of observations \( N \) minus the number of free parameters).

The variance-covariance matrix for the components of \( b \) is taken as \( s^2(X'V^{-1}X)^{-1} \) for the Yule-Walker method. For the ULS and ML methods, the variance-covariance matrix of the parameter estimates is computed as \( s^2(J'J)^{-1} \). For the ULS method, \( J \) is the matrix of derivatives of \( e \) with respect to the parameters. For the ML method, \( J \) is the matrix of derivatives of \( |L|^{1/N} e \) divided by \( |L|^{1/N} \). The estimate of the variance-covariance matrix of \( b \) assuming that \( \varphi \) is known is \( s^2(X'V^{-1}X)^{-1} \). Park and Mitchell (1980) investigated the small sample performance of the standard error estimates obtained from some of these methods. In particular, simulating an AR(1) model for the noise term, they found that the standard errors calculated using GLS with an estimated autoregressive parameter underestimated the true standard errors. These estimates of standard errors are the ones calculated by PROC AUTOREG with the Yule-Walker method.
The estimates of the standard errors calculated with the ULS or ML methods take into account the joint estimation of the AR and the regression parameters and may give more accurate standard-error values than the YW method. At the same values of the autoregressive parameters, the ULS and ML standard errors will always be larger than those computed from Yule-Walker. However, simulations of the models used by Park and Mitchell suggest that the ULS and ML standard error estimates can also be underestimates. Caution is advised, especially when the estimated autocorrelation is high and the sample size is small.

High autocorrelation in the residuals is a symptom of lack of fit. An autoregressive error model should not be used as a nostrum for models that simply do not fit. It is often the case that time series variables tend to move as a random walk. This means that an AR(1) process with a parameter near one absorbs a great deal of the variation. See Example 8.3 later in this chapter, which fits a linear trend to a sine wave.

For ULS or ML estimation, the joint variance-covariance matrix of all the regression and autoregression parameters is computed. For the Yule-Walker method, the variance-covariance matrix is computed only for the regression parameters.

### Lagged Dependent Variables

The Yule-Walker estimation method is not directly appropriate for estimating models that include lagged dependent variables among the regressors. Therefore, the maximum likelihood method is the default when the LAGDEP or LAGDEP= option is specified in the MODEL statement. However, when lagged dependent variables are used, the maximum likelihood estimator is not exact maximum likelihood but is conditional on the first few values of the dependent variable.

### Alternative Autocorrelation Correction Methods

Autocorrelation correction in regression analysis has a long history, and various approaches have been suggested. Moreover, the same method may be referred to by different names.

Pioneering work in the field was done by Cochrane and Orcutt (1949). The Cochrane-Orcutt method refers to a more primitive version of the Yule-Walker method that drops the first observation. The Cochrane-Orcutt method is like the Yule-Walker method for first-order autoregression, except that the Yule-Walker method retains information from the first observation. The iterative Cochrane-Orcutt method is also in use.

The Yule-Walker method used by PROC AUTOREG is also known by other names. Harvey (1981) refers to the Yule-Walker method as the two-step full transform method. The Yule-Walker method can be considered as generalized least squares using the OLS residuals to estimate the covariances across observations, and Judge et al. (1985) use the term estimated generalized least squares (EGLS) for this method. For a first-order AR process, the Yule-Walker estimates are often termed Prais-Winsten estimates (Prais and Winsten 1954). There are variations to these methods that use different estimators of the autocorrelations or the autoregressive parameters.
The unconditional least squares (ULS) method, which minimizes the error sum of squares for all observations, is referred to as the nonlinear least squares (NLS) method by Spitzer (1979).

The Hildreth-Lu method (Hildreth and Lu 1960) uses nonlinear least squares to jointly estimate the parameters with an AR(1) model, but it omits the first transformed residual from the sum of squares. Thus, the Hildreth-Lu method is a more primitive version of the ULS method supported by PROC AUTOREG in the same way Cochrane-Orcutt is a more primitive version of Yule-Walker.

The maximum likelihood method is also widely cited in the literature. Although the maximum likelihood method is well defined, some early literature refers to estimators that are called maximum likelihood but are not full unconditional maximum likelihood estimates. The AUTOREG procedure produces full unconditional maximum likelihood estimates.

Harvey (1981) and Judge et al. (1985) summarize the literature on various estimators for the autoregressive error model. Although asymptotically efficient, the various methods have different small sample properties. Several Monte Carlo experiments have been conducted, although usually for the AR(1) model.

Harvey and McAvinchey (1978) found that for a one-variable model, when the independent variable is trending, methods similar to Cochrane-Orcutt are inefficient in estimating the structural parameter. This is not surprising since a pure trend model is well modeled by an autoregressive process with a parameter close to 1.

Harvey and McAvinchey (1978) also made the following conclusions:

- The Yule-Walker method appears to be about as efficient as the maximum likelihood method. Although Spitzer (1979) recommended ML and NLS, the Yule-Walker method (labeled Prais-Winsten) did as well or better in estimating the structural parameter in Spitzer’s Monte Carlo study (table A2 in their article) when the autoregressive parameter was not too large. Maximum likelihood tends to do better when the autoregressive parameter is large.
- For small samples, it is important to use a full transformation (Yule-Walker) rather than the Cochrane-Orcutt method, which loses the first observation. This was also demonstrated by Maeshiro (1976), Chipman (1979), and Park and Mitchell (1980).
- For large samples (Harvey used 100), losing the first few observations does not make much difference.

**GARCH, IGARCH, EGARCH, and GARCH-M Models**

Consider the series $y_t$, which follows the GARCH process. The conditional distribution of the series $Y$ for time $t$ is written

$$y_t \mid \Psi_{t-1} \sim N(0, h_t)$$
where $\Psi_{t-1}$ denotes all available information at time $t - 1$. The conditional variance $h_t$ is

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j}$$

where

$$p \geq 0, q > 0$$

$$\omega > 0, \alpha_i \geq 0, \gamma_j \geq 0$$

The GARCH(p,q) model reduces to the ARCH(q) process when $p = 0$. At least one of the ARCH parameters must be nonzero ($q > 0$). The GARCH regression model can be written

$$y_t = x_t \beta + \epsilon_t$$

$$\epsilon_t = \sqrt{h_t} \epsilon_t$$

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j}$$

where $\epsilon_t \sim \text{IN}(0, 1)$. In addition, you can consider the model with disturbances following an autoregressive process and with the GARCH errors. The AR(m)-GARCH(p,q) regression model is denoted

$$y_t = x_t \beta + \nu_t$$

$$\nu_t = \epsilon_t - \varphi_1 \nu_{t-1} - \cdots - \varphi_m \nu_{t-m}$$

$$\epsilon_t = \sqrt{h_t} \epsilon_t$$

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j}$$
**GARCH Estimation with Nelson-Cao Inequality Constraints**

The GARCH(p,q) model is written in ARCH(\(\infty\)) form as

\[
ht = \left( 1 - \sum_{j=1}^{p} \gamma_j B^j \right)^{-1} \left[ \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \right] = \omega^* + \sum_{i=1}^{\infty} \phi_i \epsilon_{t-i}^2
\]

where \(B\) is a backshift operator. Therefore, \(ht \geq 0\) if \(\omega^* \geq 0\) and \(\phi_i \geq 0, \forall i\). Assume that the roots of the following polynomial equation are inside the unit circle:

\[
\sum_{j=0}^{p} -\gamma_j Z^{p-j}
\]

where \(\gamma_0 = -1\) and \(Z\) is a complex scalar. \(\sum_{j=0}^{p} -\gamma_j Z^{p-j}\) and \(\sum_{i=1}^{q} \alpha_i Z^{q-i}\), do not share common factors. Under these conditions, \(|\omega^*| < \infty, |\phi_i| < \infty\), and these coefficients of the ARCH(\(\infty\)) process are well defined.

Define \(n = max(p,q)\). The coefficient \(\phi_i\) is written

\[
\begin{align*}
\phi_0 &= \alpha_1 \\
\phi_1 &= \gamma_1 \phi_0 + \alpha_2 \\
\ldots \\
\phi_{n-1} &= \gamma_1 \phi_{n-2} + \gamma_2 \phi_{n-3} + \cdots + \gamma_{n-1} \phi_0 + \alpha_n \\
\phi_k &= \gamma_1 \phi_{k-1} + \gamma_2 \phi_{k-2} + \cdots + \gamma_k \phi_{k-n} \text{ for } k \geq n
\end{align*}
\]

where \(\alpha_i = 0\) for \(i > q\) and \(\gamma_j = 0\) for \(j > p\).

Nelson and Cao (1992) proposed the finite inequality constraints for GARCH(1,q) and GARCH(2,q) cases. However, it is not straightforward to derive the finite inequality constraints for the general GARCH(p,q) model.

For the GARCH(1,q) model, the nonlinear inequality constraints are

\[
\begin{align*}
\omega &\geq 0 \\
\gamma_1 &\geq 0 \\
\phi_k &\geq 0 \text{ for } k = 0, 1, \ldots, q - 1
\end{align*}
\]

For the GARCH(2,q) model, the nonlinear inequality constraints are

\[
\begin{align*}
\Delta_i &\in R \text{ for } i = 1, 2 \\
\omega^* &\geq 0
\end{align*}
\]
Part 2. General Information

\[
\Delta_1 > 0 \\
\sum_{j=0}^{q-1} \Delta_1^{-j} \alpha_{j+1} > 0 \\
\phi_k \geq 0 \text{ for } k = 0, 1, \ldots, q
\]

where \( \Delta_1 \) and \( \Delta_2 \) are the roots of \(( Z^2 - \gamma_1 Z - \gamma_2 )\).

For the GARCH(p,q) model with \( p > 2 \), only \( \max(q-1,p)+1 \) nonlinear inequality constraints \( \phi_k \geq 0 \) for \( k = 0 \) to \( \max(q-1,p) \) are imposed, together with the in-sample positivity constraints of the conditional variance \( h_t \).

**IGARCH and Stationary GARCH Model**

The condition \( \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \gamma_j < 1 \) implies that the GARCH process is weakly stationary since the mean, variance, and autocovariance are finite and constant over time. However, this condition is not sufficient for weak stationarity in the presence of autocorrelation. For example, the stationarity condition for an AR(1)-GARCH(p,q) process is

\[
\frac{1}{1 - \varphi_1} \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \gamma_j < 1
\]

When the GARCH process is stationary, the unconditional variance of \( \epsilon_t \) is computed as

\[
V(\epsilon_t) = \frac{\omega}{(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \gamma_j)}
\]

where \( \epsilon_t = \sqrt{h_t} \epsilon_t \) and \( h_t \) is the GARCH(p,q) conditional variance.

Sometimes, the multistep forecasts of the variance do not approach the unconditional variance when the model is integrated in variance; that is, \( \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \gamma_j = 1 \).

The unconditional variance for the IGARCH model does not exist. However, it is interesting that the IGARCH model can be strongly stationary even though it is not weakly stationary. Refer to Nelson (1990) for details.

**E-GARCH Model**

The E-GARCH model was proposed by Nelson (1991). Nelson and Cao (1992) argue that the nonnegativity constraints in the linear GARCH model are too restrictive. The GARCH model imposes the nonnegative constraints on the parameters, \( \alpha_i \) and \( \gamma_j \), while there are no restrictions on these parameters in the E-GARCH model. In the E-GARCH model, the conditional variance, \( h_t \), is an asymmetric function of lagged disturbances \( \epsilon_{t-i} \):

\[
\ln(h_t) = \omega + \sum_{i=1}^{q} \alpha_i g(z_{t-i}) + \sum_{j=1}^{p} \gamma_j \ln(h_{t-j})
\]
where

\[ g(z_t) = \theta z_t + \gamma |z_t| - E|z_t| \]

\[ z_t = \epsilon_t / \sqrt{h_t} \]

The coefficient of the second term in \( g(z_t) \) is set to be 1 (\( \gamma = 1 \)) in our formulation. Note that \( E|z_t| = (2/\pi)^{1/2} \) if \( z_t \sim N(0,1) \). The properties of the EGARCH model are summarized as follows:

- The function \( g(z_t) \) is linear in \( z_t \) with slope coefficient \( \theta + 1 \) if \( z_t \) is positive while \( g(z_t) \) is linear in \( z_t \) with slope coefficient \( \theta - 1 \) if \( z_t \) is negative.
- Suppose that \( \theta = 0 \). Large innovations increase the conditional variance if \( |z_t| - E|z_t| > 0 \) and decrease the conditional variance if \( |z_t| - E|z_t| < 0 \).
- Suppose that \( \theta < 1 \). The innovation in variance, \( g(z_t) \), is positive if the innovations \( z_t \) are less than \( (2/\pi)^{1/2}/(\theta - 1) \). Therefore, the negative innovations in returns, \( \epsilon_t \), cause the innovation to the conditional variance to be positive if \( \theta \) is much less than 1.

**GARCH-in-Mean**

The GARCH-M model has the added regressor that is the conditional standard deviation:

\[ y_t = \beta^T \beta + \delta \sqrt{h_t} + \epsilon_t \]

\[ \epsilon_t = \sqrt{h_t} \epsilon_t \]

where \( h_t \) follows the ARCH or GARCH process.

**Maximum Likelihood Estimation**

The family of GARCH models are estimated using the maximum likelihood method. The log-likelihood function is computed from the product of all conditional densities of the prediction errors.

When \( \epsilon_t \) is assumed to have a standard normal distribution (\( \epsilon_t \sim N(0,1) \)), the likelihood function is given by

\[ l = \sum_{t=1}^{N} \frac{1}{2} \left[ -\ln(2\pi) - \ln(h_t) - \frac{\epsilon_t^2}{h_t} \right] \]

where \( \epsilon_t = y_t - \beta^T \beta \) and \( h_t \) is the conditional variance. When the GARCH(p,q)-M model is estimated, \( \epsilon_t = y_t - \beta^T \beta - \delta \sqrt{h_t} \). When there are no regressors, the residuals \( \epsilon_t \) are denoted as \( y_t \) or \( y_t - \delta \sqrt{h_t} \).
If \( e_t \) has the standardized Student’s \( t \) distribution the log likelihood function for the conditional \( t \) distribution is

\[
\ell = \sum_{t=1}^{N} \left[ \log \left( \Gamma \left( \frac{\nu + 1}{2} \right) \right) - \log \left( \Gamma \left( \frac{\nu}{2} \right) \right) - \frac{1}{2} \log((\nu - 2) h_t) \right] \\
- \frac{1}{2}(\nu + 1) \log \left( 1 + \frac{e_t^2}{h_t(\nu - 2)} \right)
\]

where \( \Gamma(\cdot) \) is the gamma function and \( \nu \) is the degree of freedom (\( \nu > 2 \)). Under the conditional \( t \) distribution, the additional parameter \( 1/\nu \) is estimated. The log likelihood function for the conditional \( t \) distribution converges to the log likelihood function of the conditional normal GARCH model as \( 1/\nu \to 0 \).

The likelihood function is maximized via either the dual quasi-Newton or trust region algorithm. The default is the dual quasi-Newton algorithm. The starting values for the regression parameters \( \beta \) are obtained from the OLS estimates. When there are autoregressive parameters in the model, the initial values are obtained from the Yule-Walker estimates. The starting value \( 1.0^{-6} \) is used for the GARCH process parameters.

The variance-covariance matrix is computed using the Hessian matrix. The dual quasi-Newton method approximates the Hessian matrix while the quasi-Newton method gets an approximation of the inverse of Hessian. The trust region method uses the Hessian matrix obtained using numerical differentiation. When there are active constraints, that is, \( q(\theta) = 0 \), the variance-covariance matrix is given by

\[
V(\hat{\theta}) = H^{-1}[I - Q'(QH^{-1}Q')^{-1}QH^{-1}]
\]

where \( H = -\frac{\partial^2 l}{\partial \theta \partial \theta'} \) and \( Q = \frac{\partial q(\theta)}{\partial \theta} \). Therefore, the variance-covariance matrix without active constraints reduces to \( V(\hat{\theta}) = H^{-1} \).

### \( R^2 \) Statistics and Other Measures of Fit

This section discusses various goodness-of-fit statistics produced by the AUTOREG procedure.

#### Total \( R^2 \)

The total \( R^2 \) statistic (Total Rsq) is computed as

\[
R^2_{\text{tot}} = 1 - \frac{SSE}{SST}
\]

where \( SST \) is the sum of squares for the original response variable corrected for the mean and \( SSE \) is the final error sum of squares. The Total Rsq is a measure of how well the next value can be predicted using the structural part of the model and the
past values of the residuals. If the NOINT option is specified, SST is the uncorrected sum of squares.

**Regression \( R^2 \)**

The regression \( R^2 \) (Reg RSQ) is computed as

\[
R_{\text{reg}}^2 = 1 - \frac{TSSE}{TSST}
\]

where \( TSST \) is the total sum of squares of the transformed response variable corrected for the transformed intercept, and \( TSSE \) is the error sum of squares for this transformed regression problem. If the NOINT option is requested, no correction for the transformed intercept is made. The Reg RSQ is a measure of the fit of the structural part of the model after transforming for the autocorrelation and is the \( R^2 \) for the transformed regression.

The regression \( R^2 \) and the total \( R^2 \) should be the same when there is no autocorrelation correction (OLS regression).

**Calculation of Recursive Residuals and CUSUM Statistics**

The recursive residuals \( w_t \) are computed as

\[
w_t = \frac{e_t}{\sqrt{v_t}}
\]

\[
v_t = 1 + x_t' \left[ \sum_{i=1}^{t-1} x_i x_i' \right]^{-1} x_t
\]

Note that the forecast error variance of \( e_t \) is the scalar multiple of \( v_t \) such that \( V(e_t) = \sigma^2 v_t \).

The CUSUM and CUSUMSQ statistics are computed using the preceding recursive residuals.

\[
\text{CUSUM}_t = \sum_{i=k+1}^{t} \frac{w_i}{\sigma_w}
\]

\[
\text{CUSUMSQ}_t = \frac{\sum_{i=k+1}^{t} w_i^2}{\sum_{i=k+1}^{T} w_i^2}
\]

where \( w_i \) are the recursive residuals,

\[
\sigma_w = \sqrt{\frac{\sum_{i=k+1}^{T} (w_i - \hat{w})^2}{(T - k - 1)}}
\]

\[
\hat{w} = \frac{1}{T - k} \sum_{i=k+1}^{T} w_i
\]
and $k$ is the number of regressors.

The CUSUM statistics can be used to test for misspecification of the model. The upper and lower critical values for CUSUM$_t$ are

$$\pm a \left[ \sqrt{T - k} + 2 \frac{(t - k)}{(T - k)^{1/2}} \right]$$

where $a = 1.143$ for a significance level .01, 0.948 for .05, and 0.850 for .10. These critical values are output by the CUSUMLB= and CUSUMUB= options for the significance level specified by the ALPHACSM= option.

The upper and lower critical values of CUSUMSQ$_t$ are given by

$$\pm a + \frac{(t - k)}{T - k}$$

where the value of $a$ is obtained from the table by Durbin (1969) if the $\frac{1}{2}(T - k) - 1 \leq 60$. Edgerton and Wells (1994) provided the method of obtaining the value of $a$ for large samples.

These critical values are output by the CUSUMSQLB= and CUSUMSQUB= options for the significance level specified by the ALPHACSM= option.

**Information Criteria AIC and SBC**

The Akaike’s information criterion (AIC) and the Schwarz’s Bayesian information criterion (SBC) are computed as follows:

$$\text{AIC} = -2\ln(L) + 2k$$

$$\text{SBC} = -2\ln(L) + \ln(N)k$$

In these formulas, $L$ is the value of the likelihood function evaluated at the parameter estimates, $N$ is the number of observations, and $k$ is the number of estimated parameters. Refer to Judge et al. (1985) and Schwarz (1978) for additional details.

**Generalized Durbin-Watson Tests**

Consider the following linear regression model:

$$Y = X\beta + \nu$$

where $X$ is an $N \times k$ data matrix, $\beta$ is a $k \times 1$ coefficient vector, and $\nu$ is a $N \times 1$ disturbance vector. The error term $\nu$ is assumed to be generated by the $j$th order autoregressive process $\nu_t = \epsilon_t - \varphi_j \nu_{t-j}$ where $|\varphi_j| < 1$, $\epsilon_t$ is a sequence of independent normal error terms with mean 0 and variance $\sigma^2$. Usually, the Durbin-Watson statistic
is used to test the null hypothesis \( H_0 : \varphi_1 = 0 \) against \( H_1 : -\varphi_1 > 0 \). Vinod (1973) generalized the Durbin-Watson statistic:

\[
d_j = \frac{\sum_{t=j+1}^{N} (\hat{\nu}_t - \hat{\nu}_{t-j})^2}{\sum_{t=1}^{N} \hat{\nu}_t^2}
\]

where \( \hat{\nu} \) are OLS residuals. Using the matrix notation,

\[
d_j = \frac{\nu' \mathbf{M} \mathbf{A}_j \mathbf{A}_j' \mathbf{M} \nu}{\nu' \mathbf{M} \nu}
\]

where \( \mathbf{M} = \mathbf{I}_N - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \) and \( \mathbf{A}_j \) is a \((N - j) \times N\) matrix:

\[
\mathbf{A}_j = \begin{bmatrix}
-1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
0 & -1 & 0 & \cdots & 0 & 1 & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & -1 & 0 & \cdots & 0 & 1
\end{bmatrix}
\]

and there are \( j - 1 \) zeros between -1 and 1 in each row of matrix \( \mathbf{A}_j \).

The QR factorization of the design matrix \( \mathbf{X} \) yields an \( N \times N \) orthogonal matrix \( \mathbf{Q} \)

\[
\mathbf{X} = \mathbf{Q} \mathbf{R}
\]

where \( \mathbf{R} \) is a \( N \times k \) upper triangular matrix. There exists a \( N \times (N - k) \) submatrix of \( \mathbf{Q} \) such that \( \mathbf{Q}_1 \mathbf{Q}_1' = \mathbf{M} \) and \( \mathbf{Q}_1' \mathbf{Q}_1 = \mathbf{I}_{N-k} \). Consequently, the generalized Durbin-Watson statistic is stated as a ratio of two quadratic forms:

\[
d_j = \frac{\sum_{t=1}^{n} \lambda_{jt} \xi_t^2}{\sum_{t=1}^{n} \xi_t^2}
\]

where \( \lambda_{j1} \ldots \lambda_{jn} \) are upper \( n \) eigenvalues of \( \mathbf{M} \mathbf{A}_j' \mathbf{A}_j \mathbf{M} \) and \( \xi_t \) is a standard normal variate, and \( n = \min(N - k, N - j) \). These eigenvalues are obtained by a singular value decomposition of \( \mathbf{Q}_1' \mathbf{A}_j \) (Golub and Loan 1989; Savin and White 1978).

The marginal probability (or \( p \)-value) for \( d_j \) given \( c_0 \) is

\[
\text{Prob}\left(\frac{\sum_{t=1}^{n} \lambda_{jt} \xi_t^2}{\sum_{t=1}^{n} \xi_t^2} < c_0\right) = \text{Prob}(q_j < 0)
\]

where

\[
q_j = \sum_{t=1}^{n} (\lambda_{jt} - c_0) \xi_t^2
\]
When the null hypothesis $H_0 : \varphi_j = 0$ holds, the quadratic form $q_j$ has the characteristic function

$$\phi_j(t) = \prod_{l=1}^{n} (1 - 2(\lambda_{jl} - \alpha_0)it)^{-1/2}$$

The distribution function is uniquely determined by this characteristic function:

$$F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_{0}^{\infty} e^{itx} \phi_j(-t) - e^{-itx} \phi_j(t) \frac{dt}{it}$$

For example, to test $H_0 : \varphi_4 = 0$ given $\varphi_1 = \varphi_2 = \varphi_3 = 0$ against $H_1 : -\varphi_4 > 0$, the marginal probability ($p$-value) can be used:

$$F(0) = \frac{1}{2} + \frac{1}{2\pi} \int_{0}^{\infty} (\phi_4(-t) - \phi_4(t)) \frac{dt}{it}$$

where

$$\phi_4(t) = \prod_{l=1}^{n} (1 - 2(\lambda_{4l} - \hat{d}_4)it)^{-1/2}$$

and $\hat{d}_4$ is the calculated value of the fourth-order Durbin-Watson statistic.

In the Durbin-Watson test, the marginal probability indicates positive autocorrelation ($-\varphi_j > 0$) if it is less than the level of significance ($\alpha$), while you can conclude that a negative autocorrelation ($-\varphi_j < 0$) exists if the marginal probability based on the computed Durbin-Watson statistic is greater than $1-\alpha$. Wallis (1972) presented tables for bounds tests of fourth-order autocorrelation and Vinod (1973) has given tables for a five percent significance level for orders two to four. Using the AUTOREG procedure, you can calculate the exact $p$-values for the general order of Durbin-Watson test statistics. Tests for the absence of autocorrelation of order $p$ can be performed sequentially; at the $j$th step, test $H_0 : \varphi_j = 0$ given $\varphi_1 = \ldots = \varphi_{j-1} = 0$ against $\varphi_j \neq 0$. However, the size of the sequential test is not known.

The Durbin-Watson statistic is computed from the OLS residuals, while that of the autoregressive error model uses residuals that are the difference between the predicted values and the actual values. When you use the Durbin-Watson test from the residuals of the autoregressive error model, you must be aware that this test is only an approximation. See “Regression with Autoregressive Errors” earlier in this chapter. If there are missing values, the Durbin-Watson statistic is computed using all the nonmissing values and ignoring the gaps caused by missing residuals. This does not affect the significance level of the resulting test, although the power of the test against certain
alternatives may be adversely affected. Savin and White (1978) have examined the use of the Durbin-Watson statistic with missing values.

**Enhanced Durbin-Watson Probability Computation**

The Durbin-Watson probability calculations have been enhanced to compute the p-value of the generalized Durbin-Watson statistic for large sample sizes. Previously, the Durbin-Watson probabilities were only calculated for small sample sizes.

Consider the following linear regression model:

\[
Y = X\beta + u
\]

\[
u_t + \varphi_j u_{t-j} = \epsilon_t, \quad t = 1, \ldots, N
\]

where \(X\) is an \(N \times k\) data matrix, \(\beta\) is a \(k \times 1\) coefficient vector, \(u\) is a \(N \times 1\) disturbance vector, \(\epsilon_t\) is a sequence of independent normal error terms with mean 0 and variance \(\sigma^2\).

The generalized Durbin-Watson statistic is written as

\[
DW_j = \frac{\hat{u}'A_j^tA_j\hat{u}}{\hat{u}'\hat{u}}
\]

where \(\hat{u}\) is a vector of OLS residuals and \(A_j\) is a \((T - j) \times T\) matrix. The generalized Durbin-Watson statistic \(DW_j\) can be rewritten as

\[
DW_j = \frac{u'MA_j^tA_jMu}{u'Mu} = \frac{\eta'(Q_1'A_j^tA_jQ_1)\eta}{\eta'\eta}
\]

where \(Q_1'I_{T-k}Q_1'X = 0\), and \(\eta = Q_1'\hat{u}\).

The marginal probability for the Durbin-Watson statistic is

\[
Pr(DW_j < c) = Pr(h < 0)
\]

where \(h = \eta'(Q_1'A_j^tA_jQ_1 - cI)\eta\).

The \(p\)-value or the marginal probability for the generalized Durbin-Watson statistic is computed by numerical inversion of the characteristic function \(\phi(u)\) of the quadratic form \(h = \eta'(Q_1'A_j^tA_jQ_1 - cI)\eta\). The trapezoidal rule approximation to the marginal probability \(Pr(h < 0)\) is

\[
Pr(h < 0) = \frac{1}{2} - \sum_{k=0}^{K} \frac{\text{Im} [\phi((k + \frac{1}{2})\Delta)]}{\pi(k + \frac{1}{2})} + E_I(\Delta) + E_T(K)
\]

where \(\text{Im} [\phi(\cdot)]\) is the imaginary part of the characteristic function, \(E_I(\Delta)\) and \(E_T(K)\) are integration and truncation errors, respectively. Refer to Davies (1973) for numerical inversion of the characteristic function.
Ansley, Kohn, and Shively (1992) proposed a numerically efficient algorithm which requires $O(N)$ operations for evaluation of the characteristic function $\phi(u)$. The characteristic function is denoted as

\[
\phi(u) = |I - 2iu(Q_j^tA_jQ_1 - cI_{N-k})|^{-1/2}
\]

where $\mathbf{V} = (1 + 2ic\mathbf{U})I - 2iu\mathbf{A}_j^t\mathbf{A}_j$ and $i = \sqrt{-1}$. By applying the Cholesky decomposition to the complex matrix $\mathbf{V}$, you can obtain the lower triangular matrix $\mathbf{G}$ which satisfies $\mathbf{V} = \mathbf{GG'}$. Therefore, the characteristic function can be evaluated in $O(N)$ operations using the following formula:

\[
\phi(u) = |G|^{-1} |X^tX^*|^{-1/2} |X'X|^{1/2}
\]

where $X^* = G^{-1}X$. Refer to Ansley, Kohn, and Shively (1992) for more information on evaluation of the characteristic function.

**Tests for Serial Correlation with Lagged Dependent Variables**

When regressors contain lagged dependent variables, the Durbin-Watson statistic ($d_1$) for the first-order autocorrelation is biased toward 2 and has reduced power. Wallis (1972) shows that the bias in the Durbin-Watson statistic ($d_4$) for the fourth-order autocorrelation is smaller than the bias in $d_1$ in the presence of a first-order lagged dependent variable. Durbin (1970) proposed two alternative statistics ($h$ and $t$) that are asymptotically equivalent. The $h$ statistic is written as

\[
h = \hat{\rho} \sqrt{N/(1 - NV)}
\]

where $\hat{\rho} = \sum_{t=2}^{N} \hat{\nu}_t\hat{\nu}_{t-1}/\sum_{t=1}^{N} \hat{\nu}_t^2$ and $\hat{\nu}$ is the least-squares variance estimate for the coefficient of the lagged dependent variable. Durbin’s $t$-test consists of regressing the OLS residuals $\hat{\nu}_t$ on explanatory variables and $\hat{\nu}_{t-1}$ and testing the significance of the estimate for coefficient of $\hat{\nu}_{t-1}$.

Inder (1984) shows that the Durbin-Watson test for the absence of first-order autocorrelation is generally more powerful than the $h$-test in finite samples. Refer to Inder (1986) and King and Wu (1991) for the Durbin-Watson test in the presence of lagged dependent variables.

**Testing**

**Heteroscedasticity and Normality Tests**

**Portmanteau $Q$-Test**

For nonlinear time series models, the portmanteau test statistic based on squared residuals is used to test for independence of the series (McLeod and Li 1983):

\[
Q(q) = N(N + 2) \sum_{i=1}^{q} \frac{r(i; \hat{\nu}_i^2)}{(N-i)}
\]
where

\[ r(i, \hat{\nu}_t^2) = \frac{\sum_{t=i+1}^{N} (\hat{\nu}_t^2 - \sigma^2)(\hat{\nu}_{t-i}^2 - \sigma^2)}{\sum_{t=1}^{N} (\hat{\nu}_t^2 - \sigma^2)^2} \]

\[ \sigma^2 = \frac{1}{N} \sum_{t=1}^{N} \hat{\nu}_t^2 \]

This \( Q \) statistic is used to test the nonlinear effects (for example, GARCH effects) present in the residuals. The GARCH(p,q) process can be considered as an ARMA(max(p,q),p) process. See the section "Predicting the Conditional Variance" later in this chapter. Therefore, the \( Q \) statistic calculated from the squared residuals can be used to identify the order of the GARCH process.

**Lagrange Multiplier Test for ARCH Disturbances**

Engle (1982) proposed a Lagrange multiplier test for ARCH disturbances. The test statistic is asymptotically equivalent to the test used by Breusch and Pagan (1979). Engle’s Lagrange multiplier test for the \( q \)th order ARCH process is written

\[ LM(q) = \frac{NW'Z(Z'Z)^{-1}Z'W}{W'W} \]

where

\[ W = \left( \frac{\hat{\nu}_1^2}{\sigma^2}, \ldots, \frac{\hat{\nu}_N^2}{\sigma^2} \right)' \]

and

\[ Z = \begin{bmatrix}
1 & \hat{\nu}_0^2 & \cdots & \hat{\nu}_{q+1}^2 \\
: & : & \cdots & :
\end{bmatrix} \]

The presample values \( (\hat{\nu}_0^2, \ldots, \hat{\nu}_{-q+1}^2) \) have been set to 0. Note that the LM(\( q \)) tests may have different finite sample properties depending on the presample values, though they are asymptotically equivalent regardless of the presample values. The LM and \( Q \) statistics are computed from the OLS residuals assuming that disturbances are white noise. The \( Q \) and LM statistics have an approximate \( \chi^2_{(q)} \) distribution under the white-noise null hypothesis.

**Normality Test**

Based on skewness and kurtosis, Bera and Jarque (1982) calculated the test statistic

\[ T_N = \left[ \frac{N}{6} \hat{\nu}_1^2 + \frac{N}{24} (b_2 - 3)^2 \right] \]
Part 2. General Information

where

\[
b_1 = \frac{\sqrt{N} \sum_{t=1}^{N} \hat{u}_t^3}{\left( \sum_{t=1}^{N} \hat{u}_t^2 \right)^{3/2}}
\]

\[
b_2 = \frac{N \sum_{t=1}^{N} \hat{u}_t^4}{\left( \sum_{t=1}^{N} \hat{u}_t^2 \right)^{3/2}}
\]

The $\chi^2(2)$-distribution gives an approximation to the normality test $T_N$.

When the GARCH model is estimated, the normality test is obtained using the standardized residuals $\hat{u}_t = \epsilon_t / \sqrt{\hat{h}_t}$. The normality test can be used to detect misspecification of the family of ARCH models.

**Computation of the Chow Test**

Consider the linear regression model

\[ y = X \beta + u \]

where the parameter vector $\beta$ contains $k$ elements.

Split the observations for this model into two subsets at the break point specified by the CHOW= option, so that $y = (y_1', y_2')'$.

\[ X = (X_1', X_2')', \]

\[ u = (u_1', u_2')'. \]

Now consider the two linear regressions for the two subsets of the data modeled separately,

\[ y_1 = X_1 \beta_1 + u_1 \]

\[ y_2 = X_2 \beta_2 + u_2 \]

where the number of observations from the first set is $n_1$ and the number of observations from the second set is $n_2$.

The Chow test statistic is used to test the null hypothesis $H_0 : \beta_1 = \beta_2$ conditional on the same error variance $V(u_1) = V(u_2)$. The Chow test is computed using three sums of square errors.

\[ F_{chow} = \frac{(\hat{u}' \hat{u} - \hat{u}_1' \hat{u}_1 - \hat{u}_2' \hat{u}_2)/k}{(\hat{u}_1' \hat{u}_1 + \hat{u}_2' \hat{u}_2)/(n_1 + n_2 - 2k)} \]

where $\hat{u}$ is the regression residual vector from the full set model, $\hat{u}_1$ is the regression residual vector from the first set model, and $\hat{u}_2$ is the regression residual vector from the second set model.
the second set model. Under the null hypothesis, the Chow test statistic has an $F$-distribution with $k$ and $(n_1 + n_2 - 2k)$ degrees of freedom, where $k$ is the number of elements in $\beta$.

Chow (1960) suggested another test statistic that tests the hypothesis that the mean of prediction errors is 0. The predictive Chow test can also be used when $n_2 < k$.

The PCHOW= option computes the predictive Chow test statistic

$$F_{pchow} = \frac{(\hat{u}'\hat{u} - \hat{u}'_1 \hat{u}_1)/n_2}{\hat{u}'_1 \hat{u}_1/(n_1 + k)}$$

The predictive Chow test has an $F$-distribution with $n_2$ and $(n_1 - k)$ degrees of freedom.

**Unit Root and Cointegration Testing**

Consider the random walk process

$$y_t = y_{t-1} + u_t$$

where the disturbances might be serially correlated with possible heteroscedasticity. Phillips and Perron (1988) proposed the unit root test of the OLS regression model.

Let $s^2 = \frac{1}{T-K} \sum_{t=1}^{T} \hat{u}_t^2$ and let $\hat{\sigma}^2$ be the variance estimate of the OLS estimator $\hat{\alpha}$, where $\hat{u}_t$ is the OLS residual. You can estimate the asymptotic variance of $\frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2$ using the truncation lag $l$.

$$\hat{\lambda} = \sum_{j=0}^{l} \kappa_j [1 - j/(l+1)] \hat{\gamma}_j$$

where $\kappa_0 = 1$, $\kappa_j = 2$ for $j > 0$, and $\hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}$.

Then the Phillips-Perron $Z(\hat{\alpha})$ test (zero mean case) is written

$$Z(\hat{\alpha}) = T(\hat{\alpha} - 1) - \frac{1}{2} T^2 \hat{\sigma}^2(\hat{\lambda} - \hat{\gamma}_0)/s^2$$

and has the following limiting distribution:

$$\frac{1}{\sqrt{T}} \left\{ \frac{B(1)^2 - 1}{\int_0^1 [B(x)]^2 dx} \right\}$$

where $B(\cdot)$ is a standard Brownian motion. Note that the realization $Z(x)$ from the the stochastic process $B(\cdot)$ is distributed as $N(0,x)$ and thus $B(1)^2 \sim \chi_1^2$. 
Therefore, you can observe that $P(\alpha < 1) \approx 0.68$ as $T \to \infty$, which shows that the limiting distribution is skewed to the left.

Let $t_\alpha$ be the $t$-test statistic for $\alpha$. The Phillips-Perron $Z(t_\alpha)$ test is written

$$Z(t_\alpha) = (\hat{\gamma}_0 / \hat{\lambda})^{1/2} t_\alpha - \frac{1}{2} T \hat{\sigma}(\hat{\lambda} - \gamma_0)/(s \hat{\lambda}^{1/2})$$

and its limiting distribution is derived as

$$\frac{1}{2} \{ [B(1)]^2 - 1 \} \left[ \int_0^1 B(x)dx \right]^{1/2}$$

When you test the regression model $y_t = \mu + \alpha y_{t-1} + u_t$ for the true random walk process (single mean case), the limiting distribution of the statistic $Z(\hat{\alpha})$ is written

$$\frac{1}{2} \{ [B(1)]^2 - 1 \} - B(1) \int_0^1 B(x)dx \left[ \int_0^1 B(x)dx \right]^{1/2}$$

while the limiting distribution of the statistic $Z(t_\alpha)$ is given by

$$\frac{1}{2} \{ [B(1)]^2 - 1 \} - B(1) \int_0^1 B(x)dx \left[ \int_0^1 B(x)dx \right]^{1/2}$$

Finally, the limiting distribution of the Phillips-Perron test for the random walk with drift process $y_t = \mu + y_{t-1} + u_t$ (trend case) can be derived as

$$\begin{bmatrix} 0 & c & 0 \end{bmatrix} V^{-1} \begin{bmatrix} B(1) \\ \frac{[B(1)]^2 - 1}{2} \\ B(1) - \int_0^1 B(x)dx \end{bmatrix}$$

where $c = 1$ for $Z(\hat{\alpha})$ and $c = \frac{1}{\sqrt{Q}}$ for $Z(t_\alpha)$,

$$V = \begin{bmatrix} 1 & \int_0^1 B(x)dx & \frac{1}{2} \\ \int_0^1 B(x)dx & \int_0^1 B(x)^2dx & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

$$Q = \begin{bmatrix} 0 & c & 0 \end{bmatrix} V^{-1} \begin{bmatrix} 0 \\ c \\ 0 \end{bmatrix}$$

When several variables $z_t = (z_{1t}, \ldots, z_{kt})'$ are cointegrated, there exists
a \((k \times 1)\) cointegrating vector \(c\) such that \(c' z_t\) is stationary and \(c\) is a nonzero vector. The residual based cointegration test is based on the following regression model:

\[ y_t = \beta_1 + x_t' \beta + u_t \]

where \(y_t = z_{1t}, \ x_t = (z_{2t}, \ldots, z_{kt})'\), and \(\beta = (\beta_2, \ldots, \beta_k)'\). You can estimate the consistent cointegrating vector using OLS if all variables are difference stationary, that is, I(1). The Phillips-Ouliaris test is computed using the OLS residuals from the preceding regression model, and it performs the test for the null hypothesis of no cointegration. The estimated cointegrating vector is \(\hat{c} = (1, -\hat{\beta_2}, \ldots, -\hat{\beta_k})'\).

Since the AUTOREG procedure does not produce the \(p\)-value of the cointegration test, you need to refer to the tables by Phillips and Ouliaris (1990). Before you apply the cointegration test, you might perform the unit root test for each variable.

**Predicted Values**

The AUTOREG procedure can produce two kinds of predicted values for the response series and corresponding residuals and confidence limits. The residuals in both cases are computed as the actual value minus the predicted value. In addition, when GARCH models are estimated, the AUTOREG procedure can output predictions of the conditional error variance.

**Predicting the Unconditional Mean**

The first type of predicted value is obtained from only the structural part of the model, \(x_t'\beta\). These are useful in predicting values of new response time series, which are assumed to be described by the same model as the current response time series. The predicted values, residuals, and upper and lower confidence limits for the structural predictions are requested by specifying the PREDICTEDM=, RESIDUALM=, UCLM=, or LCLM= options in the OUTPUT statement. The ALPHACLTM= option controls the confidence level for UCLM= and LCLM=. These confidence limits are for estimation of the mean of the dependent variable, \(x_t'\beta\), where \(x_t\) is the column vector of independent variables at observation \(t\).

The predicted values are computed as

\[ \hat{y}_t = x_t'\beta \]

and the upper and lower confidence limits as

\[ \hat{u}_t = \hat{y}_t + t_{\alpha/2} v \]

\[ \hat{l}_t = \hat{y}_t - t_{\alpha/2} v \]

where \(v^2\) is an estimate of the variance of \(\hat{y}_t\) and \(t_{\alpha/2}\) is the upper \(\alpha/2\) percentage point of the \(t\) distribution.

\[ \text{Prob}(T > t_{\alpha/2}) = \alpha/2 \]
where $T$ is an observation from a $t$ distribution with $q$ degrees of freedom. The value of $\alpha$ can be set with the ALPHACLIM= option. The degrees of freedom parameter, $q$, is taken to be the number of observations minus the number of free parameters in the regression and autoregression parts of the model. For the YW estimation method, the value of $v$ is calculated as

$$v = \sqrt{s^2 x_t' (X'V^{-1}X)^{-1} x_t}$$

where $s^2$ is the error sum of squares divided by $q$. For the ULS and ML methods, it is calculated as

$$v = \sqrt{s^2 x_t' W x_t}$$

where $W$ is the $k \times k$ submatrix of $(J'J)^{-1}$ that corresponds to the regression parameters. For details, see "Computational Methods" earlier in this chapter.

**Predicting Future Series Realizations**

The other predicted values use both the structural part of the model and the predicted values of the error process. These conditional mean values are useful in predicting future values of the current response time series. The predicted values, residuals, and upper and lower confidence limits for future observations conditional on past values are requested by the PREDICTED=, RESIDUAL=, UCL=, or LCL= options in the OUTPUT statement. The ALPHACLIM= option controls the confidence level for UCL= and LCL=. These confidence limits are for the predicted value,

$$\hat{y}_t = x_t' b + \nu_{t-1}$$

where $x_t$ is the vector of independent variables and $\nu_{t-1}$ is the minimum variance linear predictor of the error term given the available past values of $\nu_{t-j}, j = 1, 2, \ldots, t - 1$, and the autoregressive model for $\nu_t$. If the $m$ previous values of the structural residuals are available, then

$$\nu_{t-1} = -\hat{\phi}_1 \nu_{t-1} - \cdots - \hat{\phi}_m \nu_{t-m}$$

where $\hat{\phi}_1, \ldots, \hat{\phi}_m$ are the estimated AR parameters. The upper and lower confidence limits are computed as

$$\bar{y}_t = \hat{y}_t + t_{\alpha/2} v$$

$$\tilde{y}_t = \hat{y}_t - t_{\alpha/2} v$$

where $v$, in this case, is computed as

$$v = \sqrt{s^2 (x_t' (X'V^{-1}X)^{-1} x_t + r)}$$
where the value $rs^2$ is the estimate of the variance of $\nu_{|t-1}$. At the start of the series, and after missing values, $r$ is generally greater than 1. See "Predicting the Conditional Variance" for computational details of $r$. The plot of residuals and confidence limits in Example 8.4 later in this chapter illustrates this behavior.

Except to adjust the degrees of freedom for the error sum of squares, the preceding formulas do not account for the fact that the autoregressive parameters are estimated. In particular, the confidence limits are likely to be somewhat too narrow. In large samples, this is probably not an important effect, but it may be appreciable in small samples. Refer to Harvey (1981) for some discussion of this problem for AR(1) models.

Note that at the beginning of the series (the first $m$ observations, where $m$ is the value of the NLAG= option) and after missing values, these residuals do not match the residuals obtained by using OLS on the transformed variables. This is because, in these cases, the predicted noise values must be based on less than a complete set of past noise values and, thus, have larger variance. The GLS transformation for these observations includes a scale factor as well as a linear combination of past values. Put another way, the $L^{-1}$ matrix defined in the section "Computational Methods" has the value 1 along the diagonal, except for the first $m$ observations and after missing values.

**Predicting the Conditional Variance**

The GARCH process can be written

$$
\epsilon_t^2 = \omega + \sum_{i=1}^{n} (\alpha_i + \gamma_i) \epsilon_{t-i}^2 - \sum_{j=1}^{p} \gamma_j \eta_{t-j} + \eta_t
$$

where $\eta_t = \epsilon_t^2 - h_t$ and $n = \max(p, q)$. This representation shows that the squared residual $\epsilon_t^2$ follows an ARMA($n$, $p$) process. Then for any $d > 0$, the conditional expectations are as follows:

$$
E(\epsilon_{t+d}^2|\Psi_t) = \omega + \sum_{i=1}^{n} (\alpha_i + \gamma_i) E(\epsilon_{t+d-i}^2|\Psi_t) - \sum_{j=1}^{p} \gamma_j E(\eta_{t+d-j}|\Psi_t)
$$

The $d$-step-ahead prediction error, $\xi_{t+d} = y_{t+d} - y_{t+d|t}$, has the conditional variance

$$
V(\xi_{t+d}|\Psi_t) = \sum_{j=0}^{d-1} g_j^2 \sigma_{t+d-j|t}^2
$$

where

$$
\sigma_{t+d-j|t}^2 = E(\epsilon_{t+d-j}^2|\Psi_t)
$$

Coefficients in the conditional $d$-step prediction error variance are calculated recursively using the following formula:

$$
g_j = -\varphi_1 g_{j-1} - \cdots - \varphi_m g_{j-m}
$$
Part 2. General Information

where \( g_0 = 1 \) and \( g_j = 0 \) if \( j < 0 \); \( \varphi_1, \ldots, \varphi_m \) are autoregressive parameters. Since the parameters are not known, the conditional variance is computed using the estimated autoregressive parameters. The \( d \)-step-ahead prediction error variance is simplified when there are no autoregressive terms:

\[
V(\xi_{t+d}|\Psi_t) = \sigma_{t+d}^2
\]

Therefore, the one-step-ahead prediction error variance is equivalent to the conditional error variance defined in the GARCH process:

\[
h_t = E(\varepsilon_t^2|\Psi_{t-1}) = \sigma_{it-1}^2
\]

Note that the conditional prediction error variance of the EGARCH and GARCH-M models cannot be calculated using the preceding formula. Therefore, the confidence intervals for the predicted values are computed assuming the homoscedastic conditional error variance. That is, the conditional prediction error variance is identical to the unconditional prediction error variance:

\[
V(\xi_{t+d}|\Psi_t) = V(\xi_{t+d}) = \sigma^2 \sum_{j=0}^{d-1} g_j^2
\]

since \( \sigma_{t+d-j}^2 = \sigma^2 \). You can compute \( \sigma^2 r \), which is the second term of the variance for the predicted value \( \hat{y}_t \) explained previously in "Predicting Future Series Realizations," using the formula \( \sigma^2 \sum_{j=0}^{d-1} g_j^2 \); \( r \) is estimated from \( \sum_{j=0}^{d-1} g_j^2 \) using the estimated autoregressive parameters.

Consider the following conditional prediction error variance:

\[
V(\xi_{t+d}|\Psi_t) = \sigma^2 \sum_{j=0}^{d-1} g_j^2 + \sum_{j=0}^{d-1} g_j^2 (\sigma_{t+d-j}^2 - \sigma^2)
\]

The second term in the preceding equation can be interpreted as the noise from using the homoscedastic conditional variance when the errors follow the GARCH process. However, it is expected that if the GARCH process is covariance stationary, the difference between the conditional prediction error variance and the unconditional prediction error variance disappears as the forecast horizon \( d \) increases.

**OUT= Data Set**

The output SAS data set produced by the OUTPUT statement contains all the variables in the input data set and the new variables specified by the OUTPUT statement options. See the section "OUTPUT Statement" earlier in this chapter for information on the output variables that can be created. The output data set contains one observation for each observation in the input data set.
OUTEST= Data Set

The OUTEST= data set contains all the variables used in any MODEL statement. Each regressor variable contains the estimate for the corresponding regression parameter in the corresponding model. In addition, the OUTEST= data set contains the following variables:

- **_A_**<sub>i</sub> the \( i \)th order autoregressive parameter estimate. There are \( m \) such variables _A_1 through _A_m, where \( m \) is the value of the NLAG= option.
- **_AH_**<sub>i</sub> the \( i \)th order ARCH parameter estimate, if the GARCH= option is specified. There are \( q \) such variables _AH_1 through _AH_q, where \( q \) is the value of the Q= option. The variable _AH_0 contains the estimate of \( \omega \).
- **_DELTA_** the estimated mean parameter for the GARCH-M model, if a GARCH-in-mean model is specified
- **_DEPVAR_** the name of the dependent variable
- **_GH_**<sub>i</sub> the \( i \)th order GARCH parameter estimate, if the GARCH= option is specified. There are \( p \) such variables _GH_1 through _GH_p, where \( p \) is the value of the P= option.
- **INTERCEP** the intercept estimate. INTERCEP contains a missing value for models for which the NOINT option is specified.
- **_METHOD_** the estimation method that is specified in the METHOD= option
- **_MODEL_** the label of the MODEL statement if one is given, or blank otherwise
- **_MSE_** the value of the mean square error for the model
- **_NAME_** the name of the row of covariance matrix for the parameter estimate, if the COVOUT option is specified
- **_LIKL_** the log likelihood value of the GARCH model
- **_SSE_** the value of the error sum of squares
- **_STDERR_** standard error of the parameter estimate, if the COVOUT option is specified
- **_THETA_** the estimate of the \( \theta \) parameter in the EGARCH model, if an EGARCH model is specified
- **_TYPE_** OLS for observations containing parameter estimates, or COV for observations containing covariance matrix elements.

The OUTEST= data set contains one observation for each MODEL statement giving the parameter estimates for that model. If the COVOUT option is specified, the OUTEST= data set includes additional observations for each MODEL statement giving the rows of the covariance of parameter estimates matrix. For covariance observations, the value of the _TYPE_ variable is COV, and the _NAME_ variable identifies the parameter associated with that row of the covariance matrix.
Printed Output

The AUTOREG procedure prints the following items:

1. the name of the dependent variable
2. the ordinary least-squares estimates
3. estimates of autocorrelations, which include the estimates of the autocovariances, the autocorrelations, and (if there is sufficient space) a graph of the autocorrelation at each LAG
4. if the PARTIAL option is specified, the partial autocorrelations
5. the preliminary MSE, which results from solving the Yule-Walker equations. This is an estimate of the final MSE.
6. the estimates of the autoregressive parameters (Coefficient), their standard errors (Std Error), and the ratio of estimate to standard error (t Ratio).
7. the statistics of fit are printed for the final model. These include the error sum of squares (SSE), the degrees of freedom for error (DFE), the mean square error (MSE), the root mean square error (Root MSE), the Schwarz information criterion (SBC), the Akaike information criterion (AIC), the regression $R^2$ (Reg Rsq), and the total $R^2$ (Total Rsq). For GARCH models, the following additional items are printed:
   - the value of the log likelihood function
   - the number of observations that are used in estimation (OBS)
   - the unconditional variance (UVAR)
   - the normality test statistic and its p-value
8. the parameter estimates for the structural model (B Value), a standard error estimate (Std Error), the ratio of estimate to standard error (t Ratio), and an approximation to the significance probability for the parameter being 0 (Approx Prob)
9. the regression parameter estimates, printed again assuming that the autoregressive parameter estimates are known to be correct. The Std Error and related statistics for the regression estimates will, in general, be different when the autoregressive parameters are assumed to be given.

ODS Table Names

PROC AUTOREG assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table. For more information on ODS, see Chapter 6, “Using the Output Delivery System.”
### Table 8.1. ODS Tables Produced in PROC AUTOREG

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FitSummary</strong></td>
<td>Summary of regression</td>
<td>default</td>
</tr>
<tr>
<td><strong>SummaryDepVarCen</strong></td>
<td>Summary of regression (centered dependent var)</td>
<td>CENTER</td>
</tr>
<tr>
<td><strong>SummaryNoIntercept</strong></td>
<td>Summary of regression (no intercept)</td>
<td>NOINT</td>
</tr>
<tr>
<td><strong>YWIterSSE</strong></td>
<td>Yule-Walker iteration sum of squared error</td>
<td>METHOD=ITYW</td>
</tr>
<tr>
<td><strong>PreMSE</strong></td>
<td>Preliminary MSE</td>
<td>NLAG=</td>
</tr>
<tr>
<td><strong>Dependent</strong></td>
<td>Dependent variable</td>
<td>default</td>
</tr>
<tr>
<td><strong>DependenceEquations</strong></td>
<td>Linear dependence equation</td>
<td></td>
</tr>
<tr>
<td><strong>ARCHTest</strong></td>
<td>Q and LM Tests for ARCH</td>
<td>ARCHTEST</td>
</tr>
<tr>
<td><strong>ChowTest</strong></td>
<td>Chow Test and Predictive Chow Test</td>
<td>CHOW=</td>
</tr>
<tr>
<td><strong>Godfrey</strong></td>
<td>Godfrey’s Serial Correlation Test</td>
<td>GODFREY=</td>
</tr>
<tr>
<td><strong>PhilPerron</strong></td>
<td>Phillips-Perron Unit Root Test</td>
<td>STATIONARITY=</td>
</tr>
<tr>
<td></td>
<td>(no regressor)</td>
<td>(PHILIPS&lt;=())</td>
</tr>
<tr>
<td><strong>PhilOul</strong></td>
<td>Phillips-Ouliaris Cointegration Test</td>
<td>STATIONARITY=</td>
</tr>
<tr>
<td></td>
<td>(has regressor)</td>
<td>(PHILIPS&lt;=())</td>
</tr>
<tr>
<td><strong>ResetTest</strong></td>
<td>Ramsey’s RESET Test</td>
<td>RESET</td>
</tr>
<tr>
<td><strong>ARParameterEstimates</strong></td>
<td>Estimates of Autoregressive Parameters</td>
<td>NLAG=</td>
</tr>
<tr>
<td><strong>CorrGraph</strong></td>
<td>Estimates of Autocorrelations</td>
<td>NLAG=</td>
</tr>
<tr>
<td><strong>BackStep</strong></td>
<td>Backward Elimination of Autoregressive Terms</td>
<td>BACKSTEP</td>
</tr>
<tr>
<td><strong>ExpAutocorr</strong></td>
<td>Expected Autocorrelations</td>
<td>NLAG=</td>
</tr>
<tr>
<td><strong>IterHistory</strong></td>
<td>Iteration History</td>
<td>ITPRINT</td>
</tr>
<tr>
<td><strong>ParameterEstimates</strong></td>
<td>Parameter Estimates</td>
<td>default</td>
</tr>
<tr>
<td><strong>ParameterEstimatesGivenAR</strong></td>
<td>Parameter estimates assuming AR parameters are given</td>
<td>NLAG=</td>
</tr>
<tr>
<td><strong>PartialAutoCorr</strong></td>
<td>Partial autocorrelation</td>
<td>PARTIAL</td>
</tr>
<tr>
<td><strong>CovB</strong></td>
<td>Covariance of Parameter Estimates</td>
<td>COVB</td>
</tr>
<tr>
<td><strong>CorrB</strong></td>
<td>Correlation of Parameter Estimates</td>
<td>CORRB</td>
</tr>
<tr>
<td><strong>CholeskyFactor</strong></td>
<td>Cholesky Root of Gamma</td>
<td>ALL</td>
</tr>
<tr>
<td><strong>Coefficients</strong></td>
<td>Coefficients for First NLAG Observations</td>
<td>COEF</td>
</tr>
<tr>
<td><strong>GammaInverse</strong></td>
<td>Gamma Inverse</td>
<td>GINV</td>
</tr>
<tr>
<td><strong>ConvergenceStatus</strong></td>
<td>Convergence Status</td>
<td>default</td>
</tr>
<tr>
<td><strong>DWTTestProb</strong></td>
<td>Durbin-Watson Statistics</td>
<td>DW=</td>
</tr>
</tbody>
</table>
## Table 8.1. (continued)

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ODS Tables Created by the Restrict Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restrict</td>
<td>Restriction table</td>
<td>default</td>
</tr>
<tr>
<td><strong>ODS Tables Created by the Test Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTest</td>
<td>F test</td>
<td>default</td>
</tr>
<tr>
<td>WaldTest</td>
<td>Wald test</td>
<td>TYPE=WALD</td>
</tr>
</tbody>
</table>

## Examples

### Example 8.1. Analysis of Real Output Series

In this example, the annual real output series is analyzed over the period 1901 to 1983 (Gordon 1986, pp 781-783). With the DATA step, the original data is transformed using the natural logarithm, and the differenced series DY is created for further analysis. The log of real output is plotted in Output 8.1.1.

```sas
title 'Analysis of Real GNP';
data gnp;
  date = intnx( 'year', '01jan1901'd, _n_-1 );
  format date year4.;
  input x @@;
  y = log(x);
  dy = dif(y);
  t = _n_;
  label y = 'Real GNP'
           dy = 'First Difference of Y'
           t = 'Time Trend';
datelines;
... datalines omitted ...
;
proc gplot data=gnp;
  plot y * date / haxis='01jan1901'd '01jan1911'd '01jan1921'd '01jan1931'd '01jan1941'd '01jan1951'd '01jan1961'd '01jan1971'd '01jan1981'd '01jan1991'd;
  symbol i=join;
run;
```
Output 8.1.1.  Real Output Series: 1901 - 1983

The (linear) trend-stationary process is estimated using the following form:

\[ y_t = \beta_0 + \beta_1 t + \nu_t \]

where

\[ \nu_t = \epsilon_t - \varphi_1 \nu_{t-1} - \varphi_2 \nu_{t-2} \]

\[ \epsilon_t \sim \text{IN}(0, \sigma^2) \]

The preceding trend-stationary model assumes that uncertainty over future horizons is bounded since the error term, \( \nu_t \), has a finite variance. The maximum likelihood AR estimates are shown in Output 8.1.2.

```sas
proc autoreg data=gnp;
   model y = t / nlag=2 method=ml;
run;
```
Part 2. General Information

Output 8.1.2. Estimating the Linear Trend Model

The AUTOREG Procedure

Maximum Likelihood Estimates

| Variable   | DF | Estimate  | Error    | t Value  | Pr > |t| Label      |
|------------|----|-----------|----------|----------|-------|------------|
| Intercept  | 1  | 4.8206    | 0.0661   | 72.88    | <.0001|            |
| t          | 1  | 0.0302    | 0.001346 | 22.45    | <.0001| Time Trend |
| AR1        | 1  | -1.2041   | 0.1040   | -11.58   | <.0001|            |
| AR2        | 1  | 0.3748    | 0.1039   | 3.61     | 0.0005|            |

Autoregressive parameters assumed given.

| Variable   | DF | Estimate  | Error    | t Value  | Pr > |t| Label      |
|------------|----|-----------|----------|----------|-------|------------|
| Intercept  | 1  | 4.8206    | 0.0661   | 72.88    | <.0001|            |
| t          | 1  | 0.0302    | 0.001346 | 22.45    | <.0001| Time Trend |

Nelson and Plosser (1982) failed to reject the hypothesis that macroeconomic time series are nonstationary and have no tendency to return to a trend line. In this context, the simple random walk process can be used as an alternative process:

\[ y_t = \beta_0 + y_{t-1} + \nu_t \]

where \( \nu_t = \epsilon_t \) and \( y_0 = 0 \). In general, the difference-stationary process is written as

\[ \phi(L)(1-L)y_t = \beta_0 \phi(1) + \theta(L) \epsilon_t \]

where \( L \) is the lag operator. You can observe that the class of a difference-stationary process should have at least one unit root in the AR polynomial \( \phi(L)(1-L) \).

The Dickey-Fuller procedure is used to test the null hypothesis that the series has a unit root in the AR polynomial. Consider the following equation for the augmented Dickey-Fuller test:

\[ \Delta y_t = \beta_0 + \delta t + \beta_1 y_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta y_{t-i} + \epsilon_t \]

where \( \Delta = 1 - L \). The test statistic \( \tau_r \) is the usual \( t \) ratio for the parameter estimate \( \beta_1 \), but the \( \tau_r \) does not follow a \( t \) distribution.

The %DFTEST macro computes the test statistic \( \tau_r \) and its \( p \) value to perform the Dickey-Fuller test. The default value of \( m \) is 3, but you can specify \( m \) with the AR= option. The option TREND=2 implies that the Dickey-Fuller test equation contains linear time trend. See Chapter 4, “SAS Macros and Functions,” for details.
The augmented Dickey-Fuller test indicates that the output series may have a difference-stationary process. The statistic _TAU_ has a value of -2.61903 and its p-value is 0.29104. See Output 8.1.3.

**Output 8.1.3.** Augmented Dickey-Fuller Test Results

The AR(1) model for the differenced series DY is estimated using the maximum likelihood method for the period 1902 to 1983. The difference-stationary process is written

\[ \Delta y_t = \beta_0 + \nu_t \]

\[ \nu_t = \epsilon_t - \varphi_1 \nu_{t-1} \]

The estimated value of \( \varphi_1 \) is -0.297 and that of \( \beta_0 \) is 0.0293. All estimated values are statistically significant.
Output 8.1.4.  Estimating the Differenced Series with AR(1) Error

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|------|---|
| Intercept| 1  | 0.0293   | 0.009093       | 3.22    | 0.0018 |
| AR1      | 1  | -0.2967  | 0.1067         | -2.78   | 0.0067 |

Autoregressive parameters assumed given.

Example 8.2. Comparing Estimates and Models

In this example, the Grunfeld series are estimated using different estimation methods. Refer to Maddala (1977) for details of the Grunfeld investment data set. For comparison, the Yule-Walker method, the ULS method, and maximum likelihood method estimates are shown. With the DWPROB option, the $p$-value of the Durbin-Watson statistic is printed. The Durbin-Watson test indicates the positive autocorrelation of the regression residuals.

```sas
data grunfeld;
  input year gei gef gec;
  label gei = 'Gross investment GE'
    gec = 'Lagged Capital Stock GE'
    gef = 'Lagged Value of GE shares';
datalines;
... data lines omitted ...
;
proc autoreg data=grunfeld;
  model gei = gef gec / nlag=1 dwprob;
  model gei = gef gec / nlag=1 method=uls;
  model gei = gef gec / nlag=1 method=ml;
run;
```

The printed output produced by each of the MODEL statements is shown in Output 8.2.1 through Output 8.2.4.
### Output 8.2.1. OLS Analysis of Residuals

**Grunfeld’s Investment Models Fit with Autoregressive Errors**

The AUTOREG Procedure

**Dependent Variable**: 

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-9.9563</td>
<td>31.3742</td>
<td>-0.32</td>
<td>0.7548</td>
<td></td>
</tr>
<tr>
<td>gef</td>
<td>1</td>
<td>0.0266</td>
<td>0.0156</td>
<td>1.71</td>
<td>0.1063</td>
<td>Lagged Value of GE shares</td>
</tr>
<tr>
<td>gec</td>
<td>1</td>
<td>0.1517</td>
<td>0.0257</td>
<td>5.90</td>
<td>&lt;.0001</td>
<td>Lagged Capital Stock GE</td>
</tr>
</tbody>
</table>

**Estimates of Autocorrelations**

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>660.8</td>
<td>1.000000</td>
</tr>
<tr>
<td>1</td>
<td>304.6</td>
<td>0.460867</td>
</tr>
</tbody>
</table>

**Preliminary MSE**: 520.5

### Output 8.2.2. Regression Results Using Default Yule-Walker Method

**Grunfeld’s Investment Models Fit with Autoregressive Errors**

The AUTOREG Procedure

**Estimates of Autoregressive Parameters**

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.460867</td>
<td>0.221867</td>
<td>-2.08</td>
</tr>
</tbody>
</table>

**Yule-Walker Estimates**

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-18.2318</td>
<td>33.2511</td>
<td>-0.55</td>
<td>0.5911</td>
<td></td>
</tr>
<tr>
<td>gef</td>
<td>1</td>
<td>0.0322</td>
<td>0.0158</td>
<td>2.10</td>
<td>0.0523</td>
<td>Lagged Value of GE shares</td>
</tr>
<tr>
<td>gec</td>
<td>1</td>
<td>0.1392</td>
<td>0.0383</td>
<td>3.63</td>
<td>0.0022</td>
<td>Lagged Capital Stock GE</td>
</tr>
</tbody>
</table>
Output 8.2.3. Regression Results Using Unconditional Least Squares Method

Grunfeld's Investment Models Fit with Autoregressive Errors

The AUTOREG Procedure

Estimates of Autoregressive Parameters

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.460867</td>
<td>0.221867</td>
<td>-2.08</td>
</tr>
</tbody>
</table>

Algorithm converged.

Unconditional Least Squares Estimates

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>10220.8455</td>
</tr>
<tr>
<td>MSE</td>
<td>638.80284</td>
</tr>
<tr>
<td>SBC</td>
<td>193.756692</td>
</tr>
<tr>
<td>DFE</td>
<td>16</td>
</tr>
<tr>
<td>Root MSE</td>
<td>25.27455</td>
</tr>
<tr>
<td>AIC</td>
<td>189.773763</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.5511</td>
</tr>
<tr>
<td>Total R-Square</td>
<td>0.7721</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>1.3523</td>
</tr>
</tbody>
</table>

Standard Approx

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| Variable Label |
|----------|----|----------|----------------|---------|-------|----------------|
| Intercept| 1  | -18.6582 | 34.8101        | -0.54   | 0.5993|
| gef      | 1  | 0.0339   | 0.0179         | 1.89    | 0.0769|
| gec      | 1  | 0.1369   | 0.0449         | 3.05    | 0.0076|
| AR1      | 1  | -0.4996  | 0.2592         | -1.93   | 0.0718|

Autoregressive parameters assumed given.

Standard Approx

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| Variable Label |
|----------|----|----------|----------------|---------|-------|----------------|
| Intercept| 1  | -18.6582 | 33.7567        | -0.55   | 0.5881|
| gef      | 1  | 0.0339   | 0.0159         | 2.13    | 0.0486|
| gec      | 1  | 0.1369   | 0.0404         | 3.39    | 0.0037|
Output 8.2.4.  Regression Results Using Maximum Likelihood Method

Grunfeld's Investment Models Fit with Autoregressive Errors

The AUTOREG Procedure

Estimates of Autoregressive Parameters

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.460867</td>
<td>0.221867</td>
<td>-2.08</td>
</tr>
</tbody>
</table>

Algorithm converged.

Maximum Likelihood Estimates

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>10229.2303</td>
<td>DFE</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>639.32689</td>
<td>Root MSE</td>
<td>25.28491</td>
<td></td>
</tr>
<tr>
<td>SBC</td>
<td>193.738877</td>
<td>AIC</td>
<td>189.755947</td>
<td></td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.5656</td>
<td>Total R-Square</td>
<td>0.7719</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>1.3385</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Regress R-Square 0.5656 Total R-Square 0.7719

Durbin-Watson 1.3385

| Variable | DF | Estimate | Error | t Value | Pr > |t| |
|----------|----|----------|-------|---------|-------|---|
| Intercept | 1  | -18.3751 | 34.5941 | -0.53   | 0.6026 |
| gef      | 1  | 0.0334   | 0.0179  | 1.87    | 0.0799 |
| gec      | 1  | 0.1385   | 0.0428  | 3.23    | 0.0052 |
| AR1      | 1  | -0.4728  | 0.2582  | -1.83   | 0.0858 |

Autoregressive parameters assumed given.

| Variable | DF | Estimate | Error | t Value | Pr > |t| |
|----------|----|----------|-------|---------|-------|---|
| Intercept | 1  | -18.3751 | 33.3931| -0.55   | 0.5897 |
| gef      | 1  | 0.0334   | 0.0158 | 2.11    | 0.0512 |
| gec      | 1  | 0.1385   | 0.0389 | 3.56    | 0.0026 |

Example 8.3. Lack of Fit Study

Many time series exhibit high positive autocorrelation, having the smooth appearance of a random walk. This behavior can be explained by the partial adjustment and adaptive expectation hypotheses.

Short-term forecasting applications often use autoregressive models because these models absorb the behavior of this kind of data. In the case of a first-order AR process where the autoregressive parameter is exactly 1 (a random walk), the best prediction of the future is the immediate past.

PROC AUTOREG can often greatly improve the fit of models, not only by adding additional parameters but also by capturing the random walk tendencies. Thus, PROC AUTOREG can be expected to provide good short-term forecast predictions.

However, good forecasts do not necessarily mean that your structural model contributes anything worthwhile to the fit. In the following example, random noise is fit to part of a sine wave. Notice that the structural model does not fit at all, but the autoregressive process does quite well and is very nearly a first difference (A(1) = -.976).


```sas
title1 'Lack of Fit Study';
title2 'Fitting White Noise Plus Autoregressive Errors to a Sine Wave';

data a;
  pi=3.14159;
  do time = 1 to 75;
    if time > 75 then y = .;
    else y = sin( pi * ( time / 50 ) );
    x = ranuni( 1234567 );
    output;
  end;
run;

proc autoreg data=a;
  model y = x / nlag=1;
  output out=b p=pred pm=xbeta;
run;

proc gplot data=b;
  plot y*time=1 pred*time=2 xbeta*time=3 / overlay;
  symbol1 v='none' i=spline;
  symbol2 v=triangle;
  symbol3 v=circle;
run;

The printed output produced by PROC AUTOREG is shown in Output 8.3.1 and Output 8.3.2. Plots of observed and predicted values are shown in Output 8.3.3.
```
Output 8.3.1.  Results of OLS Analysis: No Autoregressive Model Fit

Lake of Fit Study
Fitting White Noise Plus Autoregressive Errors to a Sine Wave

The AUTOREG Procedure

Dependent Variable  y

Ordinary Least Squares Estimates

<table>
<thead>
<tr>
<th>SSE</th>
<th>34.8061005</th>
<th>DFE</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.47680</td>
<td>Root MSE</td>
<td>0.69050</td>
</tr>
<tr>
<td>SBC</td>
<td>163.898598</td>
<td>AIC</td>
<td>159.263622</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.0008</td>
<td>Total R-Square</td>
<td>0.0008</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>0.0057</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|-------|---|
| Intercept| 1  | 0.2383   | 0.1584         | 1.50    | 0.1367 |
| x        | 1  | -0.0665  | 0.2771         | -0.24   | 0.8109 |

Estimates of Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4641</td>
<td>1.000000</td>
</tr>
<tr>
<td>1</td>
<td>0.4531</td>
<td>0.976386</td>
</tr>
</tbody>
</table>

Preliminary MSE  0.0217

Output 8.3.2.  Regression Results with AR(1) Error Correction

Lake of Fit Study
Fitting White Noise Plus Autoregressive Errors to a Sine Wave

The AUTOREG Procedure

Estimates of Autoregressive Parameters

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.976386</td>
<td>0.025460</td>
<td>-38.35</td>
</tr>
</tbody>
</table>

Yule-Walker Estimates

<table>
<thead>
<tr>
<th>SSE</th>
<th>0.18304264</th>
<th>DFE</th>
<th>72</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.00254</td>
<td>Root MSE</td>
<td>0.05042</td>
</tr>
<tr>
<td>SBC</td>
<td>-222.30643</td>
<td>AIC</td>
<td>-229.2589</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.0001</td>
<td>Total R-Square</td>
<td>0.9947</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>0.0942</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|-------|---|
| Intercept| 1  | -0.1473  | 0.1702         | -0.87   | 0.3898 |
| x        | 1  | -0.001219| 0.0141         | -0.09   | 0.9315 |
Output 8.3.3. Plot of Autoregressive Prediction

Example 8.4. Missing Values

In this example, a pure autoregressive error model with no regressors is used to generate 50 values of a time series. Approximately fifteen percent of the values are randomly chosen and set to missing. The following statements generate the data.

```sas
title 'Simulated Time Series with Roots:';
title2 ' (X-1.25)(X**4-1.25)';
title3 'With 15% Missing Values';
data ar;
do i=1 to 550;
e = rannor(12345);
n = sum( e, .8*n1, .8*n4, -.64*n5 ); /* ar process */
y = n;
if ranuni(12345) > .85 then y = .; /* 15% missing */
n5=n4; n4=n3; n3=n2; n2=n1; n1=n; /* set lags */
if i>500 then output;
end;
run;
```

The model is estimated using maximum likelihood, and the residuals are plotted with 99% confidence limits. The PARTIAL option prints the partial autocorrelations. The following statements fit the model:

```sas
proc autoreg data=ar partial;
    model y = / nlag=(1 4 5) method=ml;
    output out=a predicted=p residual=r ucl=u lcl=l alphacli=.01;
run;
```
Chapter 8. Examples

The printed output produced by the AUTOREG procedure is shown in Output 8.4.1.

**Output 8.4.1. Autocorrelation-Corrected Regression Results**

```
Simulated Time Series with Roots:
(X-1.25)(X**4-1.25)
With 15% Missing Values

The AUTOREG Procedure

Dependent Variable y

Ordinary Least Squares Estimates

<table>
<thead>
<tr>
<th>SSE</th>
<th>182.972379</th>
<th>DFE</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>4.57431</td>
<td>Root MSE</td>
<td>2.13876</td>
</tr>
<tr>
<td>SBC</td>
<td>181.39282</td>
<td>AIC</td>
<td>179.679248</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.0000</td>
<td>Total R-Square</td>
<td>0.0000</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>1.3962</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Variable  | DF | Estimate | Error | t Value | Pr > |t| |
|-----------|----|----------|-------|---------|------|----|
| Intercept | 1  | -2.2387  | 0.3340| -6.70   | <.0001 |

Estimates of Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4.4627</td>
<td>1.000000</td>
</tr>
<tr>
<td>1</td>
<td>1.4241</td>
<td>0.319109</td>
</tr>
<tr>
<td>2</td>
<td>1.6505</td>
<td>0.369829</td>
</tr>
<tr>
<td>3</td>
<td>0.6808</td>
<td>0.152551</td>
</tr>
<tr>
<td>4</td>
<td>2.9167</td>
<td>0.653556</td>
</tr>
<tr>
<td>5</td>
<td>-0.3816</td>
<td>-0.085519</td>
</tr>
</tbody>
</table>

Partial Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.319109</td>
</tr>
<tr>
<td>4</td>
<td>0.619288</td>
</tr>
<tr>
<td>5</td>
<td>-0.821179</td>
</tr>
</tbody>
</table>
```
### Estimates of Autoregressive Parameters

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.733182</td>
<td>0.089966</td>
<td>-8.15</td>
</tr>
<tr>
<td>4</td>
<td>-0.803754</td>
<td>0.071849</td>
<td>-11.19</td>
</tr>
<tr>
<td>5</td>
<td>0.821179</td>
<td>0.093818</td>
<td>8.75</td>
</tr>
</tbody>
</table>

### Expected Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Autocorr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.0000</td>
</tr>
<tr>
<td>1</td>
<td>0.4204</td>
</tr>
<tr>
<td>2</td>
<td>0.2480</td>
</tr>
<tr>
<td>3</td>
<td>0.3160</td>
</tr>
<tr>
<td>4</td>
<td>0.6903</td>
</tr>
<tr>
<td>5</td>
<td>0.0228</td>
</tr>
</tbody>
</table>

Algorithm converged.

### Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>48.4396756</td>
</tr>
<tr>
<td>MSE</td>
<td>1.30918</td>
</tr>
<tr>
<td>SBC</td>
<td>146.879013</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.0000</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>2.9457</td>
</tr>
</tbody>
</table>

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|------|---|
| Intercept| 1  | -2.2370  | 0.5239         | -4.27   | 0.0001|
| AR1      | 1  | -0.6201  | 0.1129         | -5.49   | <.0001|
| AR4      | 1  | -0.7237  | 0.0914         | -7.92   | <.0001|
| AR5      | 1  | 0.6550   | 0.1202         | 5.45    | <.0001|
The AUTOREG Procedure

Expected Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Autocorr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.0000</td>
</tr>
<tr>
<td>1</td>
<td>0.4204</td>
</tr>
<tr>
<td>2</td>
<td>0.2423</td>
</tr>
<tr>
<td>3</td>
<td>0.2958</td>
</tr>
<tr>
<td>4</td>
<td>0.6318</td>
</tr>
<tr>
<td>5</td>
<td>0.0411</td>
</tr>
</tbody>
</table>

Autoregressive parameters assumed given.

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|------|---|
| Intercept| 1  | -2.2370  | 0.5225         | -4.28   | 0.0001 |

The following statements plot the residuals and confidence limits:

```sas
data reshape1;
set a;
miss = .;
if r=. then do;
   miss = p;
p = .;
end;
run;

title 'Predicted Values and Confidence Limits';
proc gplot data=reshape1;
plot l*i=1 miss*i=2 p*i=3 u*i=4 / overlay;
symbol1 i=join v=none l=2;
symbol2 i=needle v=X;
symbol3 i=needle v=circle;
symbol4 i=join v=none l=2;
run;
```

The plot of the predicted values and the upper and lower confidence limits is shown in Output 8.4.2. Note that the confidence interval is wider at the beginning of the series (when there are no past noise values to use in the forecast equation) and after missing values where, again, there is an incomplete set of past residuals.
The following example estimates the log-log money demand equation using the maximum likelihood method. The money demand model contains four explanatory variables. The lagged nominal money stock M1 is divided by the current price level GDF to calculate a new variable M1CP since the money stock is assumed to follow the partial adjustment process. The variable M1CP is then used to estimate the coefficient of adjustment. All variables are transformed using the natural logarithm with a DATA step. Refer to Balke and Gordon (1986) for data description.

The first eight observations are printed using the PRINT procedure and are shown in Output 8.5.1. Note that the first observation of the variables M1CP and INFR are missing. Therefore, the money demand equation is estimated for the period 1968:2 to 1983:4 since PROC AUTOREG ignores the first missing observation.
\[ \text{infr} = \text{'Rate of Prices Changes'}; \]
\[
\text{datalines;}
\]
\[
\]

**Output 8.5.1.**  Money Demand Data Series – First 8 Observations

<table>
<thead>
<tr>
<th>Obs</th>
<th>date</th>
<th>ml</th>
<th>gnp</th>
<th>gdf</th>
<th>ycb</th>
<th>m</th>
<th>mlcp</th>
<th>y</th>
<th>intr</th>
<th>infr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1968:1</td>
<td>187.15</td>
<td>1036.22</td>
<td>81.18</td>
<td>6.84</td>
<td>5.44041</td>
<td>6.49337</td>
<td>1.92279</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1968:2</td>
<td>190.63</td>
<td>1056.02</td>
<td>82.12</td>
<td>6.97</td>
<td>5.44732</td>
<td>5.42890</td>
<td>6.96226</td>
<td>1.94162</td>
<td>1.15127</td>
</tr>
<tr>
<td>3</td>
<td>1968:3</td>
<td>194.30</td>
<td>1068.72</td>
<td>82.80</td>
<td>6.98</td>
<td>5.45815</td>
<td>5.43908</td>
<td>6.97422</td>
<td>1.94305</td>
<td>0.82465</td>
</tr>
<tr>
<td>4</td>
<td>1968:4</td>
<td>198.55</td>
<td>1071.28</td>
<td>84.04</td>
<td>6.84</td>
<td>5.46492</td>
<td>5.44328</td>
<td>6.97661</td>
<td>1.92279</td>
<td>1.48648</td>
</tr>
<tr>
<td>5</td>
<td>1969:1</td>
<td>201.73</td>
<td>1084.15</td>
<td>84.97</td>
<td>7.32</td>
<td>5.46980</td>
<td>5.45391</td>
<td>6.98855</td>
<td>1.99061</td>
<td>1.10054</td>
</tr>
<tr>
<td>6</td>
<td>1969:2</td>
<td>203.18</td>
<td>1088.73</td>
<td>86.10</td>
<td>7.54</td>
<td>5.46375</td>
<td>5.45659</td>
<td>6.99277</td>
<td>2.02021</td>
<td>1.32112</td>
</tr>
<tr>
<td>7</td>
<td>1969:3</td>
<td>204.18</td>
<td>1091.90</td>
<td>87.49</td>
<td>7.70</td>
<td>5.45265</td>
<td>5.44774</td>
<td>6.99567</td>
<td>2.04122</td>
<td>1.60151</td>
</tr>
<tr>
<td>8</td>
<td>1969:4</td>
<td>206.10</td>
<td>1085.53</td>
<td>88.62</td>
<td>8.22</td>
<td>5.44917</td>
<td>5.43981</td>
<td>6.98982</td>
<td>2.10657</td>
<td>1.28331</td>
</tr>
</tbody>
</table>

The money demand equation is first estimated using OLS. The DW=4 option produces generalized Durbin-Watson statistics up to the fourth order. Their exact marginal probabilities ($p$-values) are also calculated with the DWPROB option. The Durbin-Watson test indicates positive first-order autocorrelation at, say, the 10% confidence level. You can use the Durbin-Watson table, which is available only for 1% and 5% significance points. The relevant upper ($d_U$) and lower ($d_L$) bounds are $d_U = 1.731$ and $d_L = 1.471$, respectively, at 5% significance level. However, the bounds test is inconvenient since sometimes you may get the statistic in the inconclusive region while the interval between the upper and lower bounds becomes smaller with the increasing sample size.

```sas
* Title for money demand equation.
proc autoreg data=money outest=est covout;
    model m = mlcp y intr infr / dw=4 dwprob;
run;
```
Part 2. General Information

Output 8.5.2. OLS Estimation of the Partial Adjustment Money Demand Equation

```
Partial Adjustment Money Demand Equation
Quarterly Data - 1968:2 to 1983:4

The AUTOREG Procedure

Dependent Variable m
    Real Money Stock (M1)

Ordinary Least Squares Estimates

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.00271902</td>
<td>DFE</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0000469</td>
<td>Root MSE</td>
</tr>
<tr>
<td>SBC</td>
<td>-433.68709</td>
<td>AIC</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.9546</td>
<td>Total R-Square</td>
</tr>
</tbody>
</table>

Durbin-Watson Statistics

<table>
<thead>
<tr>
<th>Order</th>
<th>DW</th>
<th>Pr &lt; DW</th>
<th>Pr &gt; DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7355</td>
<td>0.0607</td>
<td>0.9393</td>
</tr>
<tr>
<td>2</td>
<td>2.1058</td>
<td>0.5519</td>
<td>0.4481</td>
</tr>
<tr>
<td>3</td>
<td>2.0286</td>
<td>0.5002</td>
<td>0.4998</td>
</tr>
<tr>
<td>4</td>
<td>2.2835</td>
<td>0.8880</td>
<td>0.1120</td>
</tr>
</tbody>
</table>

Standard Approx

| Variable | DF | Estimate | Error | t Value | Pr > |t| Variable Label |
|----------|----|----------|-------|---------|------|----------------|
| Intercept| 1  | 0.3084   | 0.2359| 1.31    | 0.1963|
| m1cp     | 1  | 0.8952   | 0.0439| 20.38   | <.0001| Lagged M1/Current GDP |
| y        | 1  | 0.0476   | 0.0122| 3.89    | 0.0003| Real GNP |
| intr     | 1  | -0.0238  | 0.00793| -3.00   | 0.0040| Yield on Corporate Bonds |
| infr     | 1  | -0.005646| 0.001584| -3.56   | 0.0007| Rate of Prices Changes |
```

The autoregressive model is estimated using the maximum likelihood method. Though the Durbin-Watson test statistic is calculated after correcting the autocorrelation, it should be used with care since the test based on this statistic is not justified theoretically.

```
proc autoreg data=money;
    model m = m1cp y intr infr / nlag=1 method=ml maxit=50;
    output out=a p=p pm=pm r=r rm=rm ucl=ucl lcl=lcl uclm=uclm lclm=lclm;
run;

proc print data=a(obs=8);
    var p pm r rm ucl lcl uclm lclm;
run;
```

A difference is shown between the OLS estimates in Output 8.5.2 and the AR(1)-ML estimates in Output 8.5.3. The estimated autocorrelation coefficient is significantly negative (-0.88345). Note that the negative coefficient of A(1) should be interpreted as a positive autocorrelation.

Two predicted values are produced dash predicted values computed for the structural model and predicted values computed for the full model. The full model includes both the structural and error-process parts. The predicted values and residuals are stored in the output data set A, as are the upper and lower 95% confidence limits for the predicted values. Part of the data set A is shown in Output 8.5.4. The first
observation is missing since the explanatory variables, M1CP and INFR, are missing for the corresponding observation.

**Output 8.5.3. Estimated Partial Adjustment Money Demand Equation**

```
Partial Adjustment Money Demand Equation
Quarterly Data - 1968:2 to 1983:4

The AUTOREG Procedure

Estimates of Autoregressive Parameters

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.126273</td>
<td>0.131393</td>
<td>-0.96</td>
</tr>
</tbody>
</table>

Algorithm converged.

Maximum Likelihood Estimates

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSR</td>
<td>0.00226719</td>
<td>DFE</td>
<td>57</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0005398</td>
<td>Root MSE</td>
<td>0.00631</td>
</tr>
<tr>
<td>SBC</td>
<td>-439.47665</td>
<td>AIC</td>
<td>-452.33545</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.6954</td>
<td>Total R-Square</td>
<td>0.9621</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>2.1778</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```

```
Variable  DF  Estimate  Error  t Value  Pr > |t|  Variable Label
Intercept 1  2.4121  0.4880  4.94  <.0001  
m1cp  1  0.4086  0.0908  4.50  <.0001  Lagged M1/Current GDF
y  1  0.1509  0.0411  3.67  0.0005  Real GNP
intr  1  0.1101  0.0159  6.92  <.0001  Yield on Corporate Bonds
infr  1  0.006348  0.001834  3.46  0.0010  Rate of Prices Changes
AR1  1  0.8835  0.0686  12.89  <.0001  
```

Autoregressive parameters assumed given.

```
Variable  DF  Estimate  Error  t Value  Pr > |t|  Variable Label
Intercept 1  2.4121  0.4880  4.94  <.0001  
m1cp  1  0.4086  0.0908  4.50  <.0001  Lagged M1/Current GDF
y  1  0.1509  0.0411  3.67  0.0005  Real GNP
intr  1  0.1101  0.0159  6.92  <.0001  Yield on Corporate Bonds
infr  1  0.006348  0.001834  3.46  0.0010  Rate of Prices Changes
```

**Output 8.5.4. Partial List of the Predicted Values**

```
Obs p  pm  r  rm  ucl  lcl  uclm  lclm
1  
2  5.45962  5.45962  -0.005763043  -0.012301  5.49319  5.42606  5.47962  5.43962
3  5.45663  5.46750  0.00511258  -0.009395  5.47987  5.43340  5.48700  5.44800
4  5.45934  5.46761  0.005774104  -0.002691  5.48267  5.43601  5.48723  5.44799
5  5.46636  5.46874  0.003442075  0.001064  5.48903  5.44369  5.48757  5.44991
6  5.46675  5.46581  -0.002994443  -0.002054  5.48925  5.44424  5.48444  5.44718
7  5.45672  5.45854  -0.004074196  -0.005889  5.47882  5.43462  5.47667  5.44040
8  5.44404  5.44924  0.005136019  0.000066  5.46604  5.42203  5.46726  5.43122
```

**Example 8.6. Estimation of ARCH(2) Process**

Stock returns show a tendency for small changes to be followed by small changes while large changes are followed by large changes. The plot of daily price changes of the IBM common stock (Box and Jenkins 1976, p 527) are shown in Output 8.6.1.
The time series look serially uncorrelated, but the plot makes us skeptical of their independence.

With a DATA step, the stock (capital) returns are computed from the closing prices. To forecast the conditional variance, an additional 46 observations with missing values are generated.

```sas
title 'IBM Stock Returns (daily)';
title2 '29jun1959 - 30jun1960';

data ibm;
  infile datalines eof=last;
  input x @@;
  r = dif( log( x ) );
  time = _n_-1;
  output;
  return;
last:
  do i = 1 to 46;
    r = .;
    time + 1;
    output;
  end;
  return;
datalines;
;
proc gplot data=ibm;
  plot r*time / vref=0;
  symbol1 i=join v=none;
run;
```
Output 8.6.1. IBM Stock Returns: Daily

The simple ARCH(2) model is estimated using the AUTOREG procedure. The MODEL statement option GARCH=(Q=2) specifies the ARCH(2) model. The OUTPUT statement with the CEV= option produces the conditional variances \( V \). The conditional variance and its forecast is calculated using parameter estimates:

\[
 h_t = \hat{\omega} + \hat{\alpha}_1 \epsilon_{t-1}^2 + \hat{\alpha}_2 \epsilon_{t-2}^2
\]

\[
 E(\epsilon_{t+d}^2 | \Psi_t) = \hat{\omega} + \sum_{i=1}^{2} \hat{\alpha}_i E(\epsilon_{t+d-i}^2 | \Psi_t)
\]

where \( d > 1 \).

```sas
proc autoreg data=ibm maxit=50;  
    model r = / noint garch=(q=2);  
    output out=a cev=v;  
run;
```

The parameter estimates for \( \omega, \alpha_1, \) and \( \alpha_2 \) are 0.00011, 0.04136, and 0.06976, respectively. The normality test indicates that the conditional normal distribution may not fully explain the leptokurtosis in the stock returns (Bollerslev 1987).

The ARCH model estimates are shown in Output 8.6.2, and conditional variances are also shown in Output 8.6.3.
### Output 8.6.2. ARCH(2) Estimation Results

The AUTOREG Procedure

Dependent Variable r

**Ordinary Least Squares Estimates**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.03214307</td>
<td>DFE</td>
<td>254</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0001265</td>
<td>Root MSE</td>
<td>0.01125</td>
</tr>
<tr>
<td>SBC</td>
<td>-1558.802</td>
<td>AIC</td>
<td>-1558.802</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.0000</td>
<td>Total R-Square</td>
<td>0.0000</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>2.1377</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** No intercept term is used. R-squares are redefined.

Algorithm converged.

**GARCH Estimates**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Observations</th>
<th>254</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.03214307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>0.0001265</td>
<td>Uncond Var</td>
<td>0.00012632</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>781.017441</td>
<td>Total R-Square</td>
<td>0.0000</td>
</tr>
<tr>
<td>SBC</td>
<td>-1545.4229</td>
<td>AIC</td>
<td>-1556.0349</td>
</tr>
<tr>
<td>Normality Test</td>
<td>105.8557</td>
<td>Pr &gt; ChiSq</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**NOTE:** No intercept term is used. R-squares are redefined.

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|-------|---|
| ARCH0    | 1  | 0.000112 | 7.5608E-6      | 14.85   | <.0001|
| ARCH1    | 1  | 0.0413   | 0.0511         | 0.81    | 0.4181|
| ARCH2    | 1  | 0.0697   | 0.0432         | 1.62    | 0.1062|
Output 8.6.3. Conditional Variance for IBM Stock Prices
Part 2. General Information

References


Hildreth, C. and Lu, J.Y. (1960), "Demand Relations with Autocorrelated Disturbances," Michigan State University Agricultural Experiment Station Technical Bulletin 276, East Lansing, MI.

Part 2. General Information


