SAS/ETS® 14.2 User's Guide
The AUTOREG Procedure

This document is an individual chapter from *SAS/ETS® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/ETS® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/ETS® 14.2 User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2016

Chapter 8
The AUTOREG Procedure

Contents

Overview: AUTOREG Procedure

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* (*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 Q test statistics (McLeod and Li 1983), Engle's Lagrange multiplier tests (Engle 1982), tests from Lee and King (1993), and tests from Wong and Li (1995). Test statistics and significance p-values are reported for conditional heteroscedasticity at lags 1 through 12. The Jarque-Bera normality test statistic and its significance level are also reported to test for conditional nonnormality of residuals. The following tests for independence are also supported by the AUTOREG procedure for residual analysis and diagnostic checking: Brock-Dechert-Scheinkman (BDS) test, runs test, turning point test, and the rank version of the von Neumann ratio test.

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 it 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)
- quadratic GARCH (QGARCH)
- threshold GARCH (TGARCH)
- power GARCH (PGARCH)
- GARCH-in-mean (GARCH-M)

For GARCH-type models, the AUTOREG procedure produces the conditional prediction error variances in addition to 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 in addition to 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.

Getting Started: AUTOREG Procedure

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:

$$
\begin{aligned}
y_t &= \mathbf{x}_t'\beta + v_t \\
v_t &= -\varphi_1 v_{t-1} - \varphi_2 v_{t-2} - \cdots - \varphi_m v_{t-m} + \epsilon_t \\
\epsilon_t &\sim \mathrm{IN}(0, \sigma^2)
\end{aligned}
$$

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

By simultaneously estimating the regression coefficients β and the autoregressive error model parameters φ_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:

```
/* Regression with Autocorrelated Errors */
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

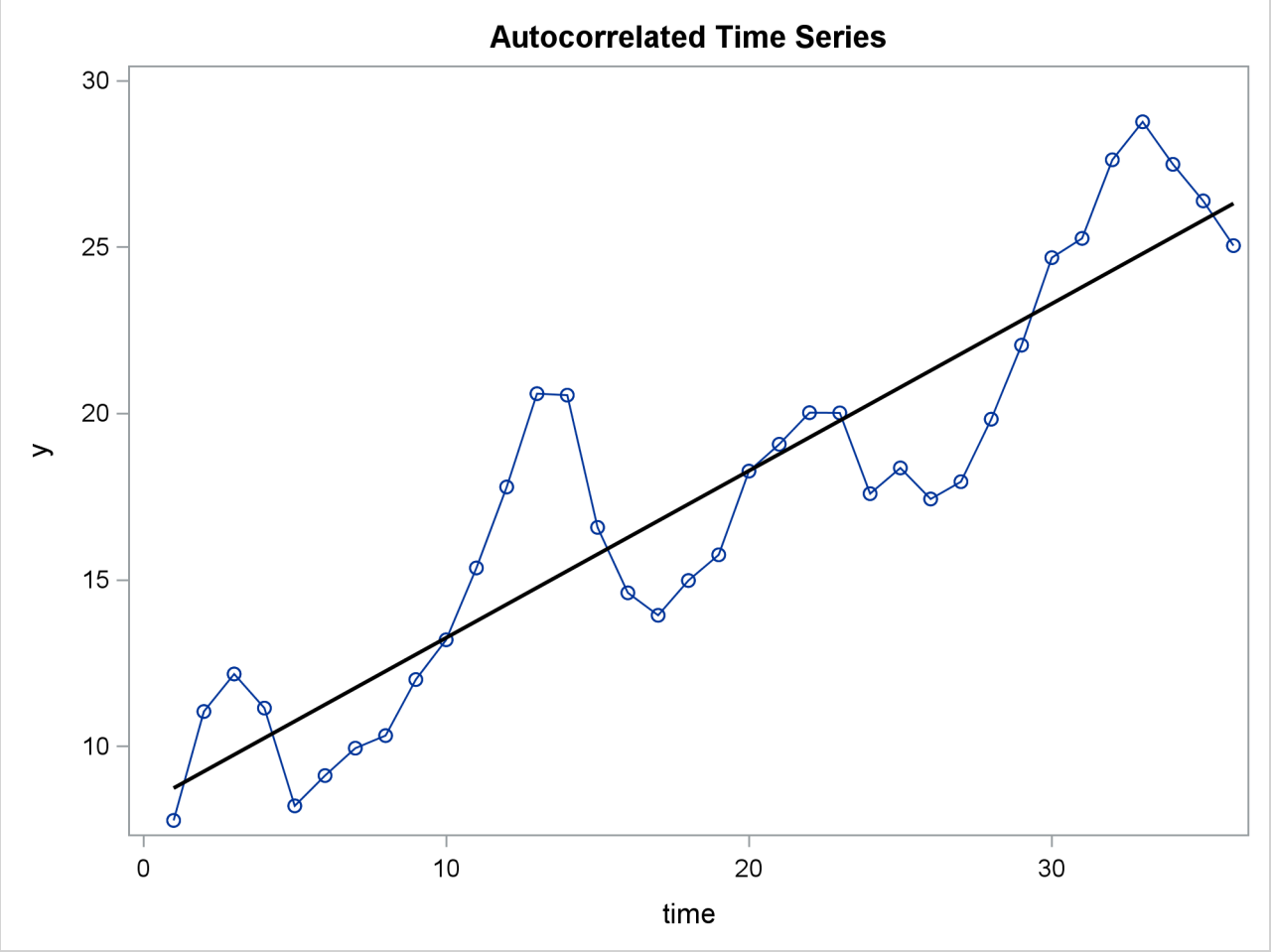
$$y_t = 10 + 0.5t + v_t$$
$$v_t = 1.3v_{t-1} - 0.5v_{t-2} + \epsilon_t$$
$$\epsilon_t \sim \text{IN}(0, 4)$$

The following statements plot the simulated time series Y. A linear regression trend line is shown for reference.

```
title 'Autocorrelated Time Series';
proc sgplot data=a noautolegend;
   series x=time y=y / markers;
   reg x=time y=y/  lineattrs=(color=black);
run;
```

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

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 by using ordinary least squares:

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

The AUTOREG procedure output is shown in Figure 8.2.

Figure 8.2 PROC AUTOREG Results for OLS Estimation

Autocorrelated Time Series

The AUTOREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	214.953429	DFE	34
MSE	6.32216	Root MSE	2.51439
SBC	173.659101	AIC	170.492063
MAE	2.01903356	AICC	170.855699
MAPE	12.5270666	HQC	171.597444
Durbin-Watson	0.4752	Total R-Square	0.8200

Parameter Estimates

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

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 (Total R-Square) is 0.82.

Other statistics shown are the sum of square errors (SSE), mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), error degrees of freedom (DFE, the number of observations minus the number of parameters), the information criteria SBC, HQC, AIC, and AICC, and the Durbin-Watson statistic. (Durbin-Watson statistics, MAE, MAPE, SBC, HQC, AIC, and AICC are discussed in the section "Goodness-of-Fit Measures and Information Criteria" on page 380.)

The output then shows a table of regression coefficients, with standard errors and t tests. The estimated model is

$$y_t = 8.23 + 0.502t + \epsilon_t$$
$$\text{Est. Var}(\epsilon_t) = 6.32$$

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.

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

The first part of the results is 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.

Figure 8.3 Preliminary Estimate for AR(2) Error Model

Autocorrelated Time Series

The AUTOREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	214.953429	DFE	34
MSE	6.32216	Root MSE	2.51439
SBC	173.659101	AIC	170.492063
MAE	2.01903356	AICC	170.855699
MAPE	12.5270666	HQC	171.597444
Durbin-Watson	0.4752	Total R-Square	0.8200

Parameter Estimates

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

Estimates of Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	5.9709	1.000000	||********************|
1	4.5169	0.756485	||*************** |
2	2.0241	0.338995	||******* |

Preliminary MSE	1.7943

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.

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

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-1.169057	0.148172	-7.89
2	0.545379	0.148172	3.68

Algorithm converged.

Maximum Likelihood Estimates

SSE	54.7493022	**DFE**	32
MSE	1.71092	**Root MSE**	1.30802
SBC	133.476508	**AIC**	127.142432
MAE	0.98307236	**AICC**	128.432755
MAPE	6.45517689	**HQC**	129.353194
Log Likelihood	-59.571216	**Transformed Regression R-Square**	0.7280
Durbin-Watson	2.2761	**Total R-Square**	0.9542
		Observations	36

Parameter Estimates

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

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 σ^2, while the OLS MSE overestimates σ^2.

Notice that the total R^2 statistic computed from the autoregressive model residuals is 0.954, reflecting the improved fit from the use of past residuals to help predict the next Y value. The transformed regression R^2 0.728 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 more information, see the section "Goodness-of-Fit Measures and Information Criteria" on page 380, 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 AR(1) and AR(2).

The estimated model is

$$y_t = 7.88 + 0.5096t + v_t$$
$$v_t = 1.25v_{t-1} - 0.628v_{t-2} + \epsilon_t$$
$$\text{Est. Var}(\epsilon_t) = 1.71$$

Note that the signs of the autoregressive parameters shown in this equation for v_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, $\mathbf{x}_t'\mathbf{b}$. 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 value includes 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.)

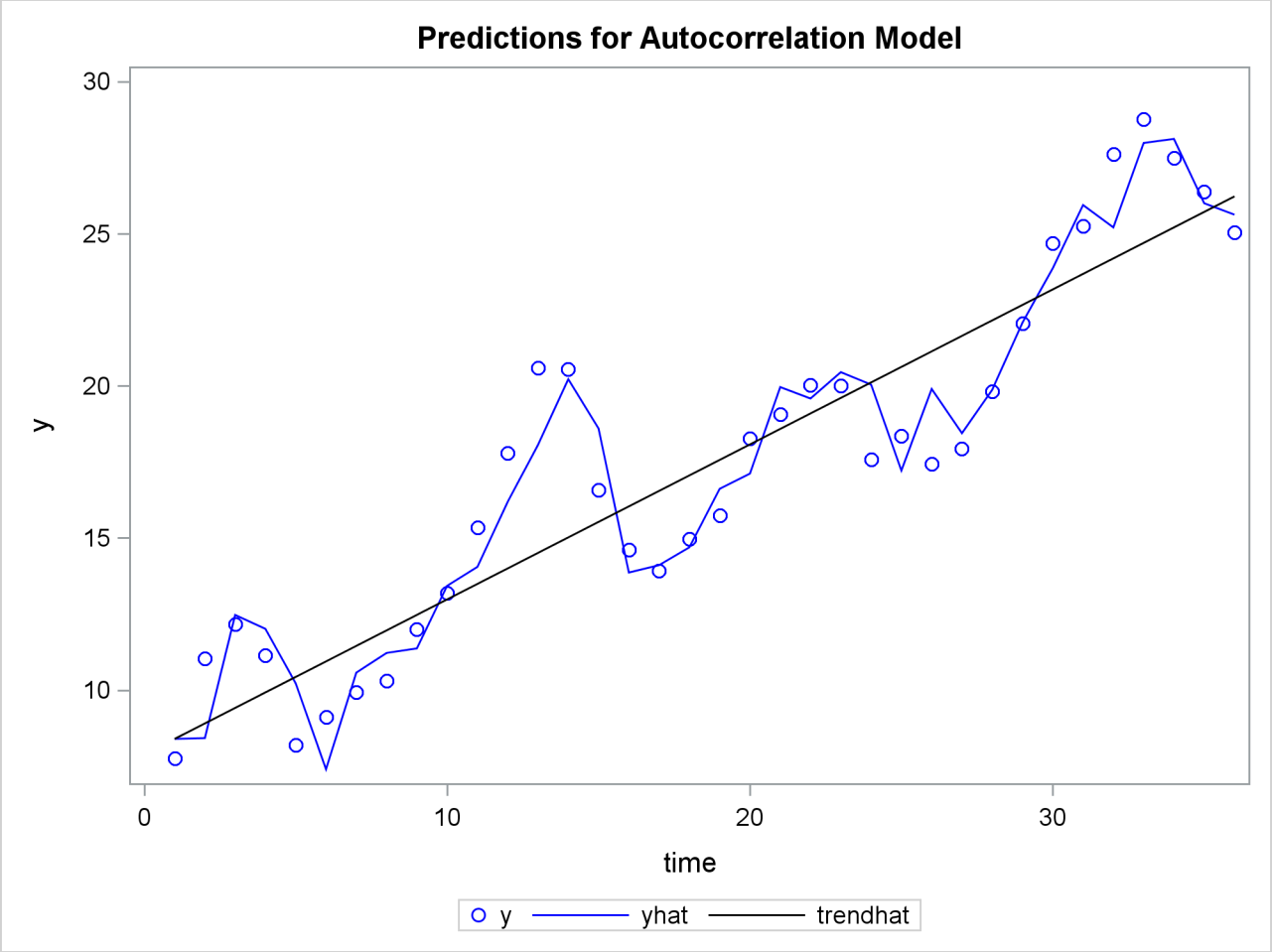
```
proc autoreg data=a;
   model y = time / nlag=2 method=ml;
   output out=p p=yhat pm=trendhat;
run;
```

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

```
title 'Predictions for Autocorrelation Model';
proc sgplot data=p;
   scatter x=time y=y / markerattrs=(color=blue);
   series x=time y=yhat / lineattrs=(color=blue);
   series x=time y=trendhat / lineattrs=(color=black);
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:

```
data b;
   y = .;
   do time = 37 to 46; output; end;
run;

data b;
   merge a b;
   by time;
run;
```

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.

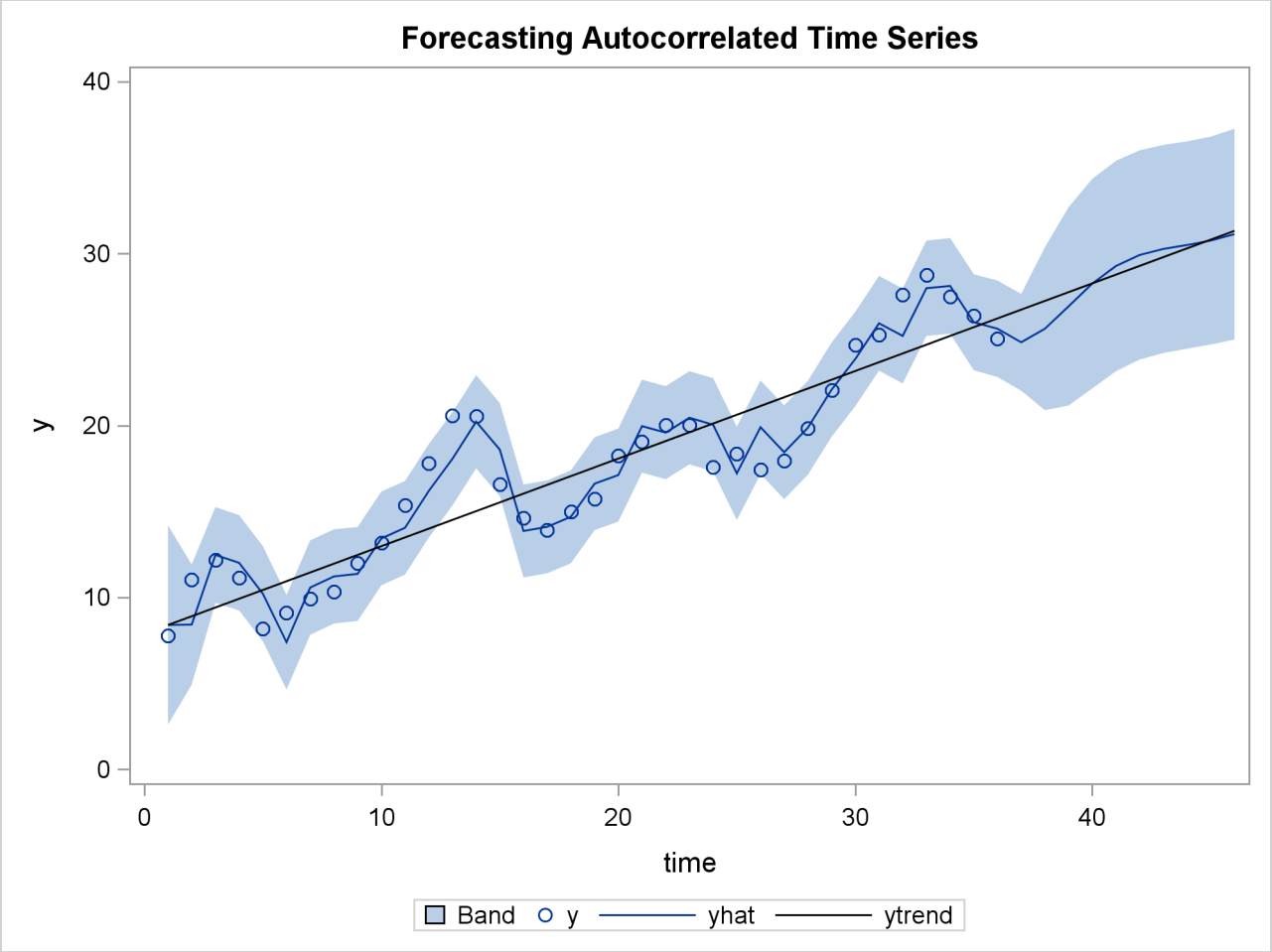
```
proc autoreg data=b;
   model y = time / nlag=2 method=ml;
   output out=p p=yhat pm=ytrend
                lcl=lcl ucl=ucl;
run;
```

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.

```
title 'Forecasting Autocorrelated Time Series';
proc sgplot data=p;
   band x=time upper=ucl lower=lcl;
   scatter x=time y=y;
   series x=time y=yhat;
   series x=time y=ytrend / lineattrs=(color=black);
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.

Figure 8.6 PROC AUTOREG Forecasts



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 tests only 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.

```
/*-- Durbin-Watson test for autocorrelation --*/
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.

Figure 8.7 Durbin-Watson Test Results for OLS Residuals

Forecasting Autocorrelated Time Series

The AUTOREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	214.953429	DFE	34
MSE	6.32216	Root MSE	2.51439
SBC	173.659101	AIC	170.492063
MAE	2.01903356	AICC	170.855699
MAPE	12.5270666	HQC	171.597444
		Total R-Square	0.8200

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	0.4752	<.0001	1.0000
2	1.2935	0.0137	0.9863
3	2.0694	0.6545	0.3455
4	2.5544	0.9818	0.0182

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

Parameter Estimates					
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

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 0.0027, indicating autocorrelation.

Figure 8.8 Durbin *h* Test with a Lagged Dependent Variable

Forecasting Autocorrelated Time Series

The AUTOREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	97.711226	DFE	33
MSE	2.96095	Root MSE	1.72074
SBC	142.369787	AIC	139.259091
MAE	1.29949385	AICC	139.634091
MAPE	8.1922836	HQC	140.332903
		Total R-Square	0.9109

Miscellaneous Statistics			
Statistic	Value	Prob	Label
Durbin h	2.7814	0.0027	Pr > h

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

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:

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

The results are shown in Figure 8.9.

Figure 8.9 Stepwise Autoregression

Forecasting Autocorrelated Time Series

The AUTOREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	214.953429	**DFE**	34
MSE	6.32216	**Root MSE**	2.51439
SBC	173.659101	**AIC**	170.492063
MAE	2.01903356	**AICC**	170.855699
MAPE	12.5270666	**HQC**	171.597444
Durbin-Watson	0.4752	**Total R-Square**	0.8200

Parameter Estimates

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

Estimates of Autocorrelations

Lag	Covariance	Correlation	−1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	5.9709	1.000000	|********************|
1	4.5169	0.756485	|*************** |
2	2.0241	0.338995	|******* |
3	-0.4402	-0.073725	|* |
4	-2.1175	-0.354632	|******* |
5	-2.8534	-0.477887	|********** |

Figure 8.9 *continued*

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

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 0.05 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 a second-order 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 lower 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 12th-order subset model for the seasonality, with a single parameter at lag 12. This results in a 13th-order subset model with nonzero parameters at lags 1, 12, and 13. For further discussion of subset and factored autoregressive models, see Chapter 7, "The ARIMA Procedure."

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:

```
/*-- specifying the lags --*/
proc autoreg data=a;
   model y = time / nlag=(1 4 5);
run;
```

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

$$y_t = a + bt + v_t$$

$$v_t = -\varphi_1 v_{t-1} - \varphi_4 v_{t-4} - \varphi_5 v_{t-5} + \epsilon_t$$

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*.

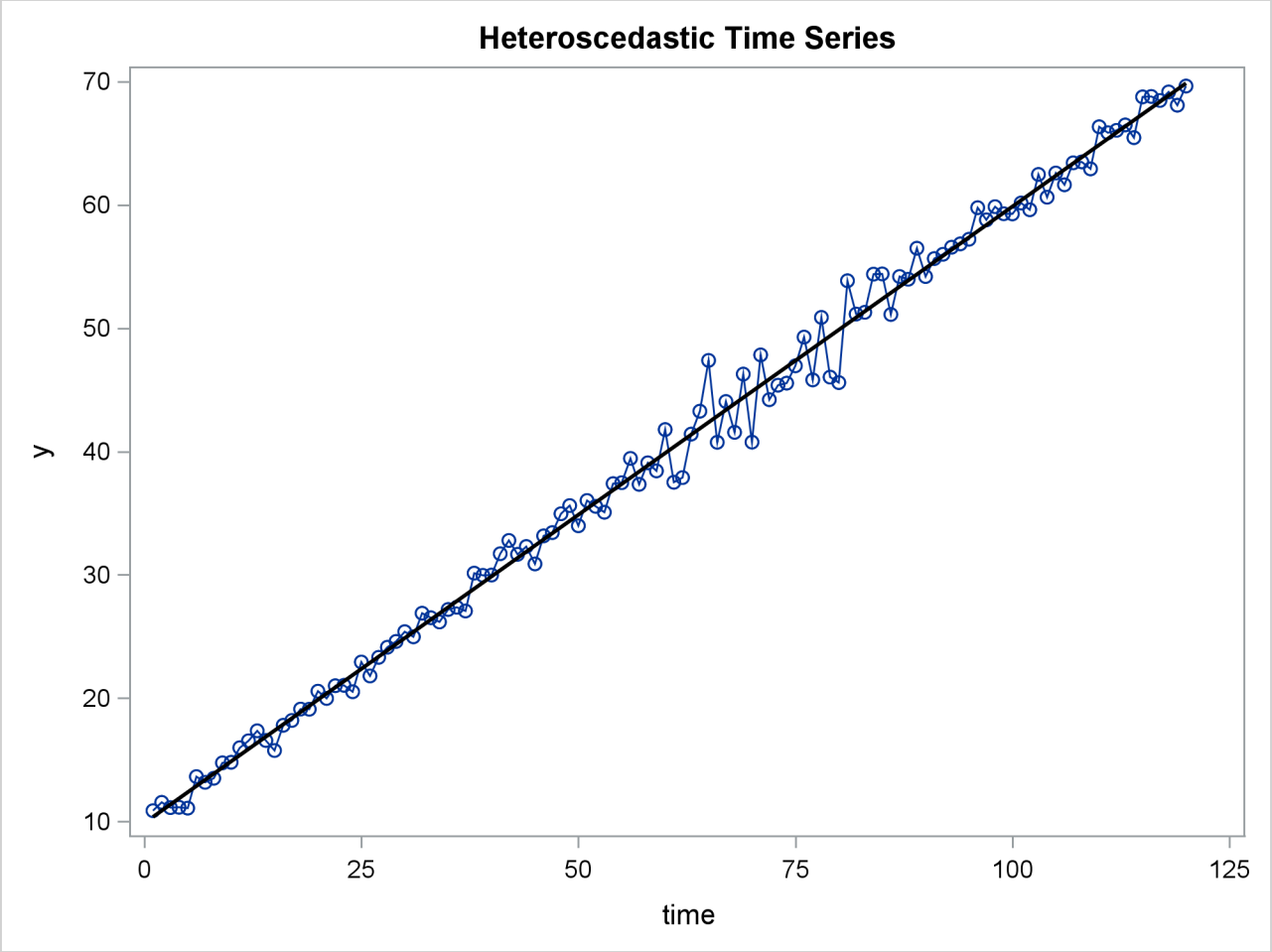
Since ordinary least squares regression assumes constant error variance, heteroscedasticity causes the OLS estimates to be inefficient. Models that take into account the changing variance can make more efficient use of the data. Also, heteroscedasticity can make the OLS forecast error variance inaccurate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

To illustrate heteroscedastic time series, the following statements 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.

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

title 'Heteroscedastic Time Series';
proc sgplot data=a noautolegend;
   series x=time y=y / markers;
   reg x=time y=y / lineattrs=(color=black);
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:

```
/*-- test for heteroscedastic OLS residuals --*/
proc autoreg data=a;
   model y = time / archtest;
   output out=r r=yresid;
run;
```

The PROC AUTOREG output is shown in Figure 8.11. The Q statistics test for changes in variance across time by using lag windows that range from 1 through 12. (For more information, see the section "Testing for Nonlinear Dependence: Heteroscedasticity Tests" on page 401.) The p-values for the test statistics 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 that is appropriate for modeling the heteroscedasticity, assuming that the changing variance follows an autoregressive conditional heteroscedasticity model.

Figure 8.11 Heteroscedasticity Tests

Heteroscedastic Time Series

The AUTOREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	223.645647	**DFE**	118
MSE	1.89530	**Root MSE**	1.37670
SBC	424.828766	**AIC**	419.253783
MAE	0.97683599	**AICC**	419.356347
MAPE	2.73888672	**HQC**	421.517809
Durbin-Watson	2.4444	**Total R-Square**	0.9938

Tests for ARCH Disturbances Based on OLS Residuals

Order	Q	Pr > Q	LM	Pr > LM
1	19.4549	<.0001	19.1493	<.0001
2	21.3563	<.0001	19.3057	<.0001
3	28.7738	<.0001	25.7313	<.0001
4	38.1132	<.0001	26.9664	<.0001
5	52.3745	<.0001	32.5714	<.0001
6	54.4968	<.0001	34.2375	<.0001
7	55.3127	<.0001	34.4726	<.0001
8	58.3809	<.0001	34.4850	<.0001
9	68.3075	<.0001	38.7244	<.0001
10	73.2949	<.0001	38.9814	<.0001
11	74.9273	<.0001	39.9395	<.0001
12	76.0254	<.0001	40.8144	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	9.8684	0.2529	39.02	<.0001
time	1	0.5000	0.003628	137.82	<.0001

The tests of Lee and King (1993) and Wong and Li (1995) can also be applied to check the absence of ARCH effects. The following example shows that Wong and Li's test is robust to detect the presence of ARCH effects with the existence of outliers:

```
/*-- data with outliers at observation 10 --*/
data b;
   do time = -10 to 120;
      s = 1 + (time >= 60 & time < 90);
      u = s*rannor(12346);
      y = 10 + .5 * time + u;
      if time = 10 then
         do; y = 200; end;
      if time > 0 then output;
   end;
run;
/*-- test for heteroscedastic OLS residuals --*/
proc autoreg data=b;
   model y = time / archtest=(qlm) ;
   model y = time / archtest=(lk,wl) ;
run;
```

As shown in Figure 8.12, the *p*-values of Q or LM statistics for all lag windows are above 90%, which fails to reject the null hypothesis of the absence of ARCH effects. Lee and King's test, which rejects the null hypothesis for lags more than 8 at 10% significance level, works better. Wong and Li's test works best, rejecting the null hypothesis and detecting the presence of ARCH effects for all lag windows.

Figure 8.12 Heteroscedasticity Tests

Heteroscedastic Time Series

The AUTOREG Procedure

Tests for ARCH Disturbances Based on OLS Residuals

Order	Q	Pr > Q	LM	Pr > LM
1	0.0076	0.9304	0.0073	0.9319
2	0.0150	0.9925	0.0143	0.9929
3	0.0229	0.9991	0.0217	0.9992
4	0.0308	0.9999	0.0290	0.9999
5	0.0367	1.0000	0.0345	1.0000
6	0.0442	1.0000	0.0413	1.0000
7	0.0522	1.0000	0.0485	1.0000
8	0.0612	1.0000	0.0565	1.0000
9	0.0701	1.0000	0.0643	1.0000
10	0.0701	1.0000	0.0742	1.0000
11	0.0701	1.0000	0.0838	1.0000
12	0.0702	1.0000	0.0939	1.0000

Figure 8.12 *continued*

Tests for ARCH Disturbances Based on OLS Residuals				
Order	LK	Pr > |LK|	WL	Pr > WL
1	-0.6377	0.5236	34.9984	<.0001
2	-0.8926	0.3721	72.9542	<.0001
3	-1.0979	0.2723	104.0322	<.0001
4	-1.2705	0.2039	139.9328	<.0001
5	-1.3824	0.1668	176.9830	<.0001
6	-1.5125	0.1304	200.3388	<.0001
7	-1.6385	0.1013	238.4844	<.0001
8	-1.7695	0.0768	267.8882	<.0001
9	-1.8881	0.0590	304.5706	<.0001
10	-2.2349	0.0254	326.3658	<.0001
11	-2.2380	0.0252	348.8036	<.0001
12	-2.2442	0.0248	371.9596	<.0001

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

$$y_t = \mathbf{x}_t'\beta + v_t$$

$$v_t = \epsilon_t - \varphi_1 v_{t-1} - \cdots - \varphi_m v_{t-m}$$

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

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

$$e_t \sim \mathrm{IN}(0,1)$$

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

The tests for the presence of ARCH effects (namely, Q and LM tests, tests from Lee and King (1993) and tests from Wong and Li (1995)) can help determine the order of the ARCH model appropriate for the data. For example, the Lagrange multiplier (LM) tests shown in Figure 8.11 are significant ($p < 0.0001$) 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 that is 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 data set is the same as in the section "Testing for Heteroscedasticity" on page 325.

```
data c;
   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 'AR(2)-GARCH(1,1) model for the Y series regressed on TIME';
proc autoreg data=c;
   model y = time / nlag=2 garch=(q=1,p=1) maxit=50;
   output out=out cev=vhat;
run;
```

The results for the GARCH model are shown in Figure 8.13. (The preliminary estimates are not shown.)

Figure 8.13 AR(2)-GARCH(1, 1) Model

AR(2)-GARCH(1,1) model for the Y series regressed on TIME

The AUTOREG Procedure

GARCH Estimates			
SSE	218.861036	**Observations**	120
MSE	1.82384	**Uncond Var**	1.6299733
Log Likelihood	-187.44013	**Total R-Square**	0.9941
SBC	408.392693	**AIC**	388.88025
MAE	0.97051406	**AICC**	389.88025
MAPE	2.75945337	**HQC**	396.804343
		Normality Test	0.0838
		Pr > ChiSq	0.9590

Figure 8.13 *continued*

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	8.9301	0.7456	11.98	<.0001
time	1	0.5075	0.0111	45.90	<.0001
AR1	1	-1.2301	0.1111	-11.07	<.0001
AR2	1	0.5023	0.1090	4.61	<.0001
ARCH0	1	0.0850	0.0780	1.09	0.2758
ARCH1	1	0.2103	0.0873	2.41	0.0159
GARCH1	1	0.7375	0.0989	7.46	<.0001

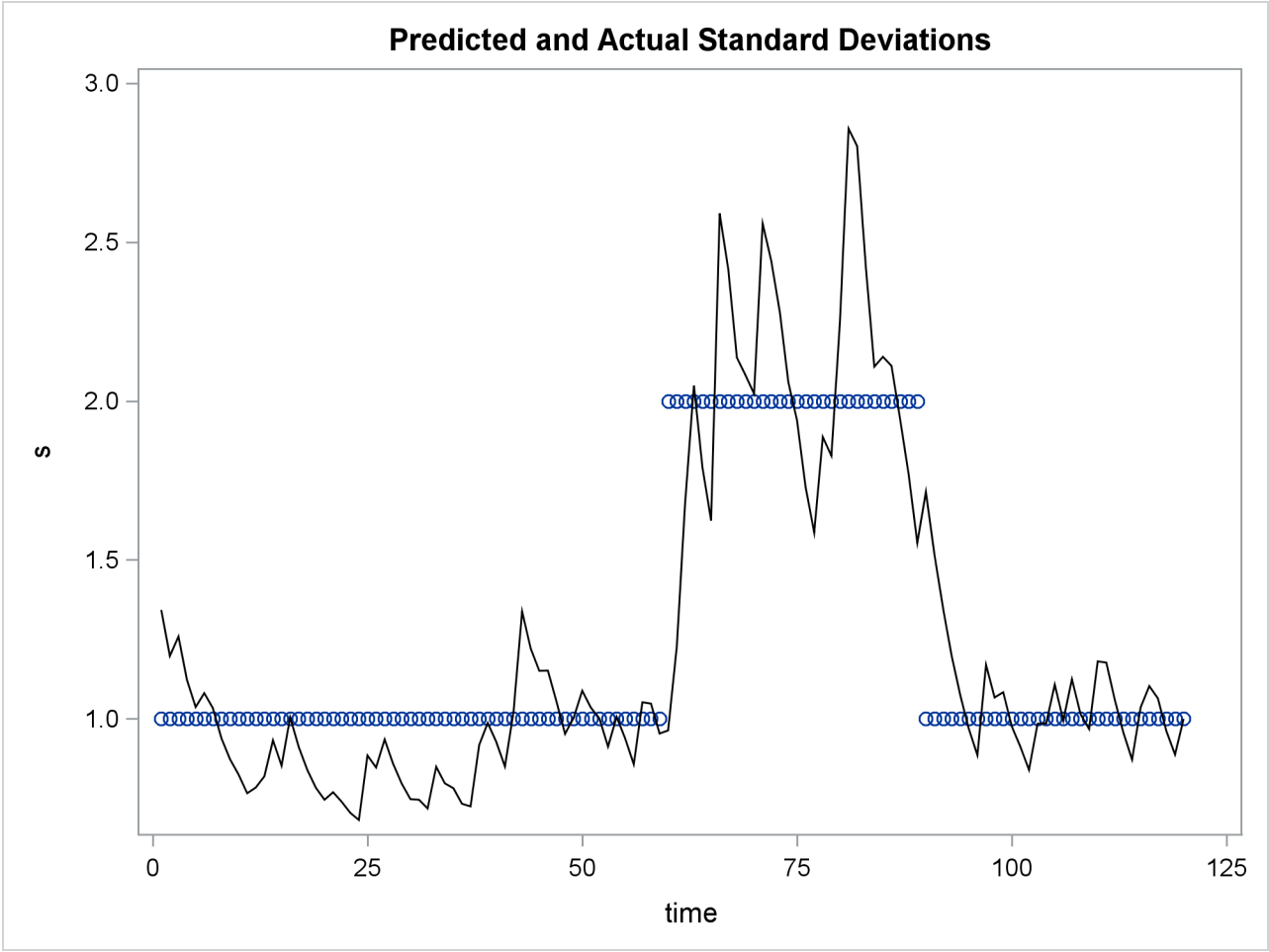
The normality test is not significant ($p = 0.959$), which is consistent with the hypothesis that the residuals from the GARCH model, $\epsilon_t / \sqrt{h_t}$, are normally distributed. The parameter estimates table includes rows for the GARCH parameters. ARCH0 represents the estimate for the parameter ω, ARCH1 represents α_1, and GARCH1 represents γ_1.

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

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

title 'Predicted and Actual Standard Deviations';
proc sgplot data=out noautolegend;
   scatter x=time y=s;
   series x=time y=shat/ lineattrs=(color=black);
run;
```

The plot is shown in Figure 8.14.

Figure 8.14 Estimated and Actual Error Standard Deviation Series



In this example note that the form of heteroscedasticity used in generating the simulated series Y does not fit the GARCH model. The GARCH model assumes *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 8.14, 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 might 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 market data.

GARCH Models

The AUTOREG procedure supports several variations of GARCH models.

Using the TYPE= option along with the GARCH= option enables you to control the constraints placed on the estimated GARCH parameters. You can specify unconstrained, nonnegativity-constrained (default), stationarity-constrained, or integration-constrained models. The integration constraint produces the integrated GARCH (IGARCH) model.

You can also use the TYPE= option to specify the exponential form of the GARCH model, called the EGARCH model, or other types of GARCH models, namely the quadratic GARCH (QGARCH), threshold GARCH (TGARCH), and power GARCH (PGARCH) models. The MEAN= option along with the GARCH= option specifies the GARCH-in-mean (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.)

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

For more information, see the section "GARCH Models" on page 371.

Syntax: AUTOREG Procedure

The AUTOREG procedure is controlled by the following statements:

PROC AUTOREG *options* ;
BY *variables* ;
CLASS *variables* ;
MODEL *dependent = regressors / options* ;
HETERO *variables / options* ;
NLOPTIONS *options* ;
OUTPUT < **OUT=***SAS-data-set* > < *options* > < *keyword=name* > ;
RESTRICT *equation , . . . , equation* ;
TEST *equation , . . . , equation / option* ;

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 Table 8.1.

Table 8.1 AUTOREG Functional Summary

Description	Statement	Option
Data Set Options		
Specify the input data set	AUTOREG	DATA=
Write parameter estimates to an output data set	AUTOREG	OUTEST=

Table 8.1 *continued*

Description	Statement	Option
Include covariances in the OUTEST= data set	AUTOREG	COVOUT
Requests that the procedure produce graphics via the Output Delivery System	AUTOREG	PLOTS=
Write predictions, residuals, and confidence limits to an output data set	OUTPUT	OUT=
Declaring the Role of Variables		
Specify BY-group processing	BY	
Specify classification variables	CLASS	
Printing Control Options		
Request all printing options	MODEL	ALL
Print transformed coefficients	MODEL	COEF
Print correlation matrix of the estimates	MODEL	CORRB
Print covariance matrix of the estimates	MODEL	COVB
Print DW statistics up to order *j*	MODEL	DW=*j*
Print marginal probability of the generalized Durbin-Watson test statistics for large sample sizes	MODEL	DWPROB
Print the *p*-values for the Durbin-Watson test be computed using a linearized approximation of the design matrix	MODEL	LDW
Print inverse of Toeplitz matrix	MODEL	GINV
Print the Godfrey LM serial correlation test	MODEL	GODFREY=
Print details at each iteration step	MODEL	ITPRINT
Print the Durbin *t* statistic	MODEL	LAGDEP
Print the Durbin *h* statistic	MODEL	LAGDEP=
Print the log-likelihood value of the regression model	MODEL	LOGLIKL
Print the Jarque-Bera normality test	MODEL	NORMAL
Print the tests for the absence of ARCH effects	MODEL	ARCHTEST=
Print BDS tests for independence	MODEL	BDS=
Print rank version of von Neumann ratio test for independence	MODEL	VNRRANK=
Print runs test for independence	MODEL	RUNS=
Print the turning point test for independence	MODEL	TP=
Print the Lagrange multiplier test	HETERO	TEST=LM
Print Bai-Perron tests for multiple structural changes	MODEL	BP=
Print the Chow test for structural change	MODEL	CHOW=
Print the predictive Chow test for structural change	MODEL	PCHOW=
Suppress printed output	MODEL	NOPRINT
Print partial autocorrelations	MODEL	PARTIAL

Table 8.1 *continued*

Description	Statement	Option
Print Ramsey's RESET test	MODEL	RESET
Print augmented Dickey-Fuller tests for stationarity or unit roots	MODEL	STATIONARITY=(ADF=)
Print ERS tests for stationarity or unit roots	MODEL	STATIONARITY=(ERS=)
Print KPSS tests or Shin tests for stationarity or cointegration	MODEL	STATIONARITY=(KPSS=)
Print Ng-Perron tests for stationarity or unit roots	MODEL	STATIONARITY=(NP=)
Print Phillips-Perron tests for stationarity or unit roots	MODEL	STATIONARITY=(PHILLIPS=)
Print tests of linear hypotheses	TEST	
Specify the test statistics to use	TEST	TYPE=
Print the uncentered regression R^2	MODEL	URSQ
Options to Control the Optimization Process		
Specify the optimization options	NLOPTIONS	See Chapter 6, "Nonlinear Optimization Methods."
Model Estimation Options		
Specify the order of autoregressive process	MODEL	NLAG=
Center the dependent variable	MODEL	CENTER
Suppress the intercept parameter	MODEL	NOINT
Remove nonsignificant AR parameters	MODEL	BACKSTEP
Specify significance level for BACKSTEP	MODEL	SLSTAY=
Specify the convergence criterion	MODEL	CONVERGE=
Specify the type of covariance matrix	MODEL	COVEST=
Set the initial values of parameters used by the iterative optimization algorithm	MODEL	INITIAL=
Specify iterative Yule-Walker method	MODEL	ITER
Specify maximum number of iterations	MODEL	MAXITER=
Specify the estimation method	MODEL	METHOD=
Use only first sequence of nonmissing data	MODEL	NOMISS
Specify the optimization technique	MODEL	OPTMETHOD=
Imposes restrictions on the regression estimates	RESTRICT	
Estimate and test heteroscedasticity models	HETERO	
GARCH Related Options		
Specify order of GARCH process	MODEL	GARCH=(Q=,P=)
Specify type of GARCH model	MODEL	GARCH=(...,TYPE=)
Specify various forms of the GARCH-M model	MODEL	GARCH=(...,MEAN=)
Suppress GARCH intercept parameter	MODEL	GARCH=(...,NOINT)
Specify the trust region method	MODEL	GARCH=(...,TR)
Estimate the GARCH model for the conditional *t* distribution	MODEL	GARCH=(...) DIST=
Estimate the start-up values for the conditional variance equation	MODEL	GARCH=(...,STARTUP=)

Table 8.1 *continued*

Description	Statement	Option
Specify the functional form of the heteroscedasticity model	HETERO	LINK=
Specify that the heteroscedasticity model does not include the unit term	HETERO	NOCONST
Impose constraints on the estimated parameters in the heteroscedasticity model	HETERO	COEF=
Impose constraints on the estimated standard deviation of the heteroscedasticity model	HETERO	STD=
Output conditional error variance	OUTPUT	CEV=
Output conditional prediction error variance	OUTPUT	CPEV=
Specify the flexible conditional variance form of the GARCH model	HETERO	
Output Control Options		
Specify confidence limit size	OUTPUT	ALPHACLI=
Specify confidence limit size for structural predicted values	OUTPUT	ALPHACLM=
Specify the significance level for the upper and lower bounds of the CUSUM and CUSUMSQ statistics	OUTPUT	ALPHACSM=
Specify the name of a variable to contain the values of the Theil's BLUS residuals	OUTPUT	BLUS=
Output the value of the error variance σ_t^2	OUTPUT	CEV=
Output transformed intercept variable	OUTPUT	CONSTANT=
Specify the name of a variable to contain the CUSUM statistics	OUTPUT	CUSUM=
Specify the name of a variable to contain the CUSUMSQ statistics	OUTPUT	CUSUMSQ=
Specify the name of a variable to contain the upper confidence bound for the CUSUM statistic	OUTPUT	CUSUMUB=
Specify the name of a variable to contain the lower confidence bound for the CUSUM statistic	OUTPUT	CUSUMLB=
Specify the name of a variable to contain the upper confidence bound for the CUSUMSQ statistic	OUTPUT	CUSUMSQUB=
Specify the name of a variable to contain the lower confidence bound for the CUSUMSQ statistic	OUTPUT	CUSUMSQLB=
Output lower confidence limit	OUTPUT	LCL=
Output lower confidence limit for structural predicted values	OUTPUT	LCLM=
Output predicted values	OUTPUT	P=
Output predicted values of structural part	OUTPUT	PM=
Output residuals	OUTPUT	R=
Output residuals from structural predictions	OUTPUT	RM=

Table 8.1 *continued*

Description	Statement	Option
Specify the name of a variable to contain the part of the predictive error variance (v_t)	OUTPUT	RECPEV=
Specify the name of a variable to contain recursive residuals	OUTPUT	RECRES=
Output transformed variables	OUTPUT	TRANSFORM=
Output upper confidence limit	OUTPUT	UCL=
Output upper confidence limit for structural predicted values	OUTPUT	UCLM=

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. For information about the contents of this data set, see the section "OUTEST= Data Set" on page 410.

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

PLOTS<(*global-plot-options***)> < = *(specific plot options)*>**
requests that the AUTOREG procedure produce statistical graphics via the Output Delivery System, provided that the ODS GRAPHICS statement has been specified. For general information about ODS Graphics, see Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*). The *global-plot-options* apply to all relevant plots generated by the AUTOREG procedure. The *global-plot-options* supported by the AUTOREG procedure follow.

Global Plot Options

ONLY suppresses the default plots. Only the plots specifically requested are produced.

UNPACKPANEL | UNPACK displays each graph separately. (By default, some graphs can appear together in a single panel.)

Specific Plot Options

ALL requests that all plots appropriate for the particular analysis be produced.

ACF produces the autocorrelation function plot.

IACF produces the inverse autocorrelation function plot of residuals.

PACF produces the partial autocorrelation function plot of residuals.

FITPLOT plots the predicted and actual values.

COOKSD produces the Cook's *D* plot.

QQ Q-Q plot of residuals.

RESIDUAL | RES plots the residuals.

STUDENTRESIDUAL plots the studentized residuals. For the models with the NLAG= or GARCH= options in the MODEL statement or with the HETERO statement, this option is replaced by the STANDARDRESIDUAL option.

STANDARDRESIDUAL plots the standardized residuals.

WHITENOISE plots the white noise probabilities.

RESIDUALHISTOGRAM | RESIDHISTOGRAM plots the histogram of residuals.

NONE suppresses all plots.

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.

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.

CLASS Statement (Experimental)

CLASS *variables* **;**

The CLASS statement names the classification variables to be used in the analysis. Classification variables can be either character or numeric.

In PROC AUTOREG, the CLASS statement enables you to output classification variables to a data set that contains a copy of the original data.

Class levels are determined from the formatted values of the CLASS variables. Thus, you can use formats to group values into levels. For more information, see the discussion of the FORMAT procedure in *SAS Language Reference: Dictionary*.

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 valid only 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

DIST=*value*
specifies the distribution assumed for the error term in GARCH-type estimation. If no GARCH= option is specified, the option is ignored. If EGARCH is specified, the distribution is always the normal distribution. The *values* of the DIST= option are as follows:

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

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, such as ARCH(1, 3), you can write the SAS statement as follows:

```
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:

```
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, P= option is ignored and Q=1 is assumed.

TYPE=*value*

specifies the type of GARCH model. The *values* of the TYPE= option are as follows:

EXP | EGARCH specifies the exponential GARCH, or EGARCH, model.

INTEGRATED | IGARCH specifies the integrated GARCH, or IGARCH, model.

NELSON | NELSONCAO specifies the Nelson-Cao inequality constraints.

NOCONSTRAINT specifies the GARCH model with no constraints.

NONNEG specifies the GARCH model with nonnegativity constraints.

POWER | PGARCH specifies the power GARCH, or PGARCH, model.

QUADR | QUADRATIC | QGARCH specifies the quadratic GARCH, or QGARCH, model.

STATIONARY constrains the sum of GARCH coefficients to be less than 1.

THRES | THRESHOLD | TGARCH | GJR | GJRGARCH specifies the threshold GARCH, or TGARCH, model.

The default is TYPE=NELSON.

MEAN=*value*

specifies the functional form of the GARCH-M model. You can specify the following *values*:

LINEAR specifies the linear function:

$$y_t = \mathbf{x}_t'\beta + \delta h_t + \epsilon_t$$

LOG specifies the log function:

$$y_t = \mathbf{x}_t'\beta + \delta \ln(h_t) + \epsilon_t$$

SQRT specifies the square root function:

$$y_t = \mathbf{x}_t'\beta + \delta \sqrt{h_t} + \epsilon_t$$

NOINT
suppresses the intercept parameter in the conditional variance model. This option is valid only when you also specify the TYPE=INTEG option.

STARTUP=MSE | 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 you specify STARTUP=MSE, 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, although computation is expensive. The double quasi-Newton method is the default.

Printing Options

ALL
requests all printing options.

ARCHTEST

ARCHTEST=*(option-list)*
specifies tests for the absence of ARCH effects. The following options can be used in the ARCHTEST=( ) option. The options are listed within parentheses and separated by commas.

QLM | QLMARCH	requests the Q and Engle's LM tests.
LK | LKARCH	requests Lee and King's ARCH tests.
WL | WLARCH	requests Wong and Li's ARCH tests.
ALL	requests all ARCH tests, namely Q and Engle's LM tests, Lee and King's tests, and Wong and Li's tests.

If ARCHTEST is defined without additional suboptions, it requests the Q and Engle's LM tests. That is, the statement

```
model return = x1 x2 / archtest;
```

is equivalent to the statement

```
model return = x1 x2 / archtest=(qlm);
```

The following statement requests Lee and King's tests and Wong and Li's tests:

```
model return = / archtest=(lk,wl);
```

BDS

BDS=*(option-list)*
specifies Brock-Dechert-Scheinkman (BDS) tests for independence. The following options can be used in the BDS=( ) option. The options are listed within parentheses and separated by commas.

M=*number*
specifies the maximum number of the embedding dimension. The BDS tests with embedding dimension from 2 to M are calculated. M must be an integer between 2 and 20. The default value of the M= suboption is 20.

D=*number*
specifies the parameter to determine the radius for BDS test. The BDS test sets up the radius as $r = D * \sigma$, where σ is the standard deviation of the time series to be tested. By default, D=1.5.

PVALUE=DIST | SIM
specifies the way to calculate the *p*-values. By default or if PVALUE=DIST is specified, the *p*-values are calculated according to the asymptotic distribution of BDS statistics (that is, the standard normal distribution). Otherwise, for samples of size less than 500, the *p*-values are obtained though Monte Carlo simulation.

Z=*value*
specifies the type of the time series (residuals) to be tested. You can specify the following *values*:

Y specifies the regressand.

RO specifies the OLS residuals.

R specifies the residuals of the final model.

RM specifies the structural residuals of the final model.

SR specifies the standardized residuals of the final model, defined by residuals over the square root of the conditional variance.

The default is Z=Y.

If BDS is defined without additional suboptions, all suboptions are set as default values. That is, the following two statements are equivalent:

```
model return = x1 x2 / nlag=1 BDS;
```

```
model return = x1 x2 / nlag=1 BDS=(M=20, D=1.5, PVALUE=DIST, Z=Y);
```

To do the specification check of a GARCH(1,1) model, you can write the SAS statement as follows:

```
model return = / garch=(p=1,q=1) BDS=(Z=SR);
```

BP

BP=*(option-list)*
specifies Bai-Perron (BP) tests for multiple structural changes, introduced in Bai and Perron (1998). You can specify the following *options* in the BP=( ) option, in parentheses and separated by commas.

EPS=*number*

specifies the minimum length of regime; that is, if EPS=ε, for any $i, i = 1, \ldots, M$, $T_i - T_{i-1} \geq T\varepsilon$, where T is the sample size; $(T_1 \ldots T_M)$ are the break dates; and $T_0 = 0$ and $T_{M+1} = T$. The default is EPS=0.05.

ETA=*number*

specifies that the second method is to be used in the calculation of the $supF(l+1|l)$ test, and the minimum length of regime for the new additional break date is $(T_i - T_{i-1})\eta$ if ETA=η and the new break date is in regime i for the given break dates $(T_1 \ldots T_l)$. The default value of the ETA= suboption is the missing value; that is, the first method is to be used in the calculation of the $supF(l+1|l)$ test and, no matter which regime the new break date is in, the minimum length of regime for the new additional break date is $T\varepsilon$ when EPS=ε.

HAC*<(option-list)>*

specifies that the heteroscedasticity- and autocorrelation-consistent estimator be applied in the estimation of the variance covariance matrix and the confidence intervals of break dates. When the HAC option is specified, you can specify the following options within parentheses and separated by commas:

KERNEL=*value*

specifies the type of kernel function. You can specify the following *values*:

BARTLETT specifies the Bartlett kernel function.

PARZEN specifies the Parzen kernel function.

QUADRATICSPECTRAL | QS specifies the quadratic spectral kernel function.

TRUNCATED specifies the truncated kernel function.

TUKEYHANNING | TUKEY | TH specifies the Tukey-Hanning kernel function.

The default is KERNEL=QUADRATICSPECTRAL.

KERNELLB=*number*

specifies the lower bound of the kernel weight value. Any kernel weight less than this lower bound is regarded as zero, which accelerates the calculation for big samples, especially for the quadratic spectral kernel. The default is KERNELLB=0.

BANDWIDTH=*value*

specifies the fixed bandwidth value or bandwidth selection method to use in the kernel function. You can specify the following *values*:

ANDREWS91 | ANDREWS

specifies the Andrews (1991) bandwidth selection method.

NEWEYWEST94 | NW94 <(C=*number***)>**

specifies the Newey and West (1994) bandwidth selection method. You can specify the C= option in parentheses to calculate the lag selection parameter; the default is C=12.

SAMPLESIZE | SS <(*option-list*)>
specifies that the bandwidth be calculated according to the following equation, based on the sample size:

$$b = \gamma T^r + c$$

where b is the bandwidth parameter and T is the sample size, and γ, r, and c are values specified by the following options within parentheses and separated by commas.

GAMMA=*number*
specifies the coefficient γ in the equation. The default is $\gamma = 0.75$.

RATE=*number*
specifies the growth rate r in the equation. The default is $r = 0.3333$.

CONSTANT=*number*
specifies the constant c in the equation. The default is $c = 0.5$.

INT
specifies that the bandwidth parameter must be integer; that is, $b = \lfloor \gamma T^r + c \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

number
specifies the fixed value of the bandwidth parameter.

The default is BANDWIDTH=ANDREWS91.

PREWHITENING
specifies that prewhitening is required in the calculation.

In the calculation of the HAC estimator, the adjustment for degrees of freedom is always applied. For more information about the HAC estimator, see the section "Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrix Estimator" on page 376.

HE
specifies that the errors are assumed to have heterogeneous distribution across regimes in the estimation of covariance matrix.

HO
specifies that Ω_is in the calculation of confidence intervals of break dates are different across regimes.

HQ
specifies that Q_is in the calculation of confidence intervals of break dates are different across regimes.

HR
specifies that the regressors are assumed to have heterogeneous distribution across regimes in the estimation of covariance matrix.

M=*number*

specifies the number of breaks. For a given M, the following tests are to be performed: (1) the $supF$ tests of no break versus the alternative hypothesis that there are i breaks, $i = 1, \ldots, M$; (2) the $UDmaxF$ and $WDmaxF$ double maximum tests of no break versus the alternative hypothesis that there are unknown number of breaks up to M; and (3) the $supF(l+1|l)$ tests of l versus $l+1$ breaks, $l = 0, \ldots, M$. The default is M=5.

NTHREADS=*number*

specifies the number of threads to be used for parallel computing. The default is the number of CPUs available.

P=*number*

specifies the number of covariates that have coefficients unchanged over time in the partial structural change model. The first P=p independent variables that are specified in the MODEL statement have unchanged coefficients; the rest of the independent variables have coefficients that change across regimes. The default is P=0; that is, the pure structural change model is estimated.

PRINTEST=ALL | **BIC** | **LWZ** | **NONE** | **SEQ***<(number)>* | *number*

specifies in which structural change models the parameter estimates are to be printed. You can specify the following option values:

ALL specifies that the parameter estimates in all structural change models with m breaks, $m = 0, \ldots, M$, be printed.

BIC specifies that the parameter estimates in the structural change model that minimizes the BIC information criterion be printed.

LWZ specifies that the parameter estimates in the structural change model that minimizes the LWZ information criterion be printed.

NONE specifies that none of the parameter estimates be printed.

SEQ specifies that the parameter estimates in the structural change model that is chosen by sequentially applying $supF(l+1|l)$ tests, l from 0 to M, be printed. If you specify the SEQ option, you can also specify the significance level in the parentheses, for example, SEQ(0.10). The first l such that the p-value of $supF(l+1|l)$ test is greater than the significance level is selected as the number of breaks in the structural change model. By default, the significance level 5% is used for the SEQ option; that is, specifying SEQ is equivalent to specifying SEQ(0.05).

number specifies that the parameter estimates in the structural change model with the specified number of breaks be printed. If the specified number is greater than the number specified in the M= option, none of the parameter estimates are printed; that is, it is equivalent to specifying the NONE option.

The default is PRINTEST=ALL.

If you define the BP option without additional suboptions, all suboptions are set as default values. That is, the following two statements are equivalent:

```
model y = z1 z2 / BP;
```

```
model y = z1 z2 / BP=(M=5, P=0, EPS=0.05, PRINTEST=ALL);
```

To apply the HAC estimator with the Bartlett kernel function and print only the parameter estimates in the structural change model selected by the LWZ information criterion, you can write the SAS statement as follows:

```
model y = z1 z2 / BP=(HAC(KERNEL=BARTLETT), PRINTEST=LWZ);
```

To specify a partial structural change model, you can write the SAS statement as follows:

```
model y = x1 x2 x3 z1 z2 / NOINT BP=(P=3);
```

CHOW=($obs_1 \ldots obs_n$)
computes Chow tests to evaluate the stability of the regression coefficient. The Chow test is also called the analysis-of-variance test.

Each value obs_i listed on the CHOW= option specifies a break point of the sample. The sample is divided into parts at the specified break point, with observations before obs_i in the first part and obs_i and later observations in the second part, and the fits of the model in the two parts are compared to whether both parts of the sample are consistent with the same model.

The break points obs_i refer to observations within the time range of the dependent variable, ignoring missing values before the start of the dependent series. Thus, CHOW=20 specifies the 20th observation after the first nonmissing observation for the dependent variable. For example, if the dependent variable Y contains 10 missing values before the first observation with a nonmissing Y value, then CHOW=20 actually refers to the 30th observation in the data set.

When you specify the break point, you should note the number of presample missing values.

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 | HC0 | HC1 | HC2 | HC3 | HC4 | HAC <(...)> | NEWEYWEST <(...)>
specifies the type of covariance matrix. You can specify the following values (by default, COVEST=OP):

OP

uses the outer product matrix to compute the covariance matrix of the parameter estimates. When the final model is an OLS or AR error model, this option is ignored; the method to calculate the estimate of covariance matrix is illustrated in the section "Variance Estimates and Standard Errors" on page 369.

HESSIAN

produces the covariance matrix by using the Hessian matrix. When the final model is an OLS or AR error model, this option is ignored; the method to calculate the estimate of covariance matrix is illustrated in the section "Variance Estimates and Standard Errors" on page 369.

QML

computes the quasi–maximum likelihood estimates. This option is equivalent to COVEST=HC0. When the final model is an OLS or AR error model, this option is ignored; the method to calculate the estimate of covariance matrix is illustrated in the section "Variance Estimates and Standard Errors" on page 369.

HC*n*

calculates the heteroscedasticity-consistent covariance matrix estimator (HCCME) that corresponds to *n*, where $n = 0, 1, 2, 3, 4$.

HAC<(*options***)>**

specifies the heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator. When you specify this option, you can specify the following *options* in parentheses and separate them with commas:

KERNEL=*value*

specifies the type of kernel function. You can specify the following *values*:

BARTLETT specifies the Bartlett kernel function.

PARZEN specifies the Parzen kernel function.

QUADRATICSPECTRAL | QS specifies the quadratic spectral kernel function.

TRUNCATED specifies the truncated kernel function.

TUKEYHANNING | TUKEY | TH specifies the Tukey-Hanning kernel function.

By default, KERNEL=QUADRATICSPECTRAL.

KERNELLB=*number*

specifies the lower bound of the kernel weight value. Any kernel weight less than *number* is regarded as zero, which accelerates the calculation for big samples, especially for the quadratic spectral kernel. By default, KERNELLB=0.

BANDWIDTH=*value*

specifies the fixed bandwidth value or bandwidth selection method to use in the kernel function. You can specify the following *values*:

ANDREWS91 | ANDREWS specifies the Andrews (1991) bandwidth selection method.

NEWEYWEST94 | NW94 <(C=*number*)> specifies the Newey and West (1994) bandwidth selection method. You can specify the C= option in the parentheses to calculate the lag selection parameter; the default is C=12.

SAMPLESIZE | SS <(*option-list*)> calculates the bandwidth according to the following equation, based on the sample size:

$$b = \gamma T^r + c$$

where b is the bandwidth parameter; T is the sample size; and γ, r, and c are values specified by the following options within parentheses and separated by commas.

GAMMA=*number*
specifies the coefficient γ in the equation. The default is $\gamma = 0.75$.

RATE=*number*
specifies the growth rate r in the equation. The default is $r = 0.3333$.

CONSTANT=*number*
specifies the constant c in the equation. The default is $c = 0.5$.

INT
specifies that the bandwidth parameter must be integer; that is, $b = \lfloor \gamma T^r + c \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

number specifies the fixed value of the bandwidth parameter.

By default, BANDWIDTH=ANDREWS91.

PREWHITENING
specifies that prewhitening is required in the calculation.

ADJUSTDF
specifies that the adjustment for degrees of freedom be required in the calculation.

NEWEYWEST<(*options*)>
specifies the well-known Newey-West estimator, which is a special HAC estimator with (1) the Bartlett kernel; (2) the bandwidth parameter determined by the equation based on the sample size, $b = \lfloor \gamma T^r + c \rfloor$; and (3) no adjustment for degrees of freedom and no prewhitening. By default, the bandwidth parameter for the Newey-West estimator is $\lfloor 0.75T^{0.3333} + 0.5 \rfloor$, as shown in equation (15.17) in Stock and Watson (2002). You can specify the following *options* in parentheses and separate them with commas:

GAMMA=*number*
specifies the coefficient γ in the equation. The default is $\gamma = 0.75$.

RATE=*number*
specifies the growth rate r in the equation. The default is $r = 0.3333$.

CONSTANT=*number*
specifies the constant c in the equation. The default is $c = 0.5$.

The following two statements are equivalent:

```
model y = x / COVEST=NEWEYWEST;

model y = x / COVEST=HAC(KERNEL=BARTLETT,
                         BANDWIDTH=SAMPLESIZE(GAMMA=0.75,
                                              RATE=0.3333,
                                              CONSTANT=0.5,
                                              INT));
```

Another popular sample-size-dependent bandwidth, $\left\lfloor T^{1/4} + 1.5 \right\rfloor$, as mentioned in Newey and West (1987), can be specified by the following statement:

```
model y = x / COVEST=NEWEYWEST(GAMMA=1,RATE=0.25,CONSTANT=1.5);
```

For more information about the HC0 to HC4, HAC, and Newey-West estimators, see the section "Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrix Estimator" on page 376. By default, COVEST=OP.

DW=*n*
prints Durbin-Watson statistics up to the order *n*. When the LAGDEP option is specified, the Durbin-Watson statistic is not printed unless the DW= option is explicitly specified. By default, DW=1.

DWPROB
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. For more information, see the section "Computational Methods" on page 368.

GODFREY
GODFREY=*r*
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. For more information, see the section "Generalized Durbin-Watson Tests" on page 397.

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. For more information, see the section "Generalized Durbin-Watson Tests" on page 397.

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:

```
model y = x1 x2 ylag2 ylag3 / lagdep=ylag2;
```

LOGLIKL

prints the log-likelihood value of the regression model, assuming normally distributed errors.

NOPRINT

suppresses all printed output.

NORMAL

specifies the Jarque-Bera's normality test statistic for regression residuals.

PARTIAL

prints partial autocorrelations.

PCHOW=*(obs*$_1$ *... obs*$_n$ *)*

computes the predictive Chow test. The form of the PCHOW= option is the same as the form of the CHOW= option; see the discussion of the CHOW= option.

RESET

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.

RUNS

RUNS=*(Z=value)*

specifies the runs test for independence. The Z= suboption specifies the type of the time series or residuals to be tested. The values of the Z= suboption are as follows:

Y specifies the regressand. The default is Z=Y.

RO specifies the OLS residuals.

R specifies the residuals of the final model.

RM specifies the structural residuals of the final model.

SR specifies the standardized residuals of the final model, defined by residuals over the square root of the conditional variance.

STATIONARITY=(*test***<=(***test-options***)><,** *test***<=(***test-options***)>>...<,** *test***<=(***test-options***)>>)**

specifies tests of stationarity or unit roots. You can specify one of more of the following *tests* along with their *test-options*. For example, the following statement tests the stationarity of a variable by using the augmented Dickey-Fuller unit root test and the KPSS test in which the quadratic spectral kernel is applied:

```
model y=  / stationarity = (adf, kpss=(kernel=qs));
```

STATIONARITY=(ADF)

STATIONARITY=(ADF=(*value . . . value***)**

produces the augmented Dickey-Fuller unit root test (Dickey and Fuller 1979). As in the Phillips-Perron test, three regression models can be specified for the null hypothesis for the augmented Dickey-Fuller test (zero mean, single mean, and trend). These models assume that the disturbances are distributed as white noise. The augmented Dickey-Fuller test can account for the serial correlation between the disturbances in some way. The model, with the time trend specification for example, is

$$y_t = \mu + \rho y_{t-1} + \delta t + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + u_t$$

This formulation has the advantage that it can accommodate higher-order autoregressive processes in u_t. The test statistic follows the same distribution as the Dickey-Fuller test statistic. For more information, see the section "PROBDF Function for Dickey-Fuller Tests" on page 163.

In the presence of regressors, the ADF option tests the cointegration relation between the dependent variable and the regressors. Following Engle and Granger (1987), a two-step estimation and testing procedure is carried out, in a fashion similar to the Phillips-Ouliaris test. The OLS residuals of the regression in the MODEL statement are used to compute the t statistic of the augmented Dickey-Fuller regression in a second step. Three cases arise based on which type of deterministic terms are included in the first step of regression. Only the constant term and linear trend cases are practically useful (Davidson and MacKinnon 1993, page 721), and therefore are computed and reported. The test statistic, as shown in Phillips and Ouliaris (1990), follows the same distribution as the $\hat{Z}_t$ statistic in the Phillips-Ouliaris cointegration test. The asymptotic distribution is tabulated in tables IIa–IIc of Phillips and Ouliaris (1990), and the finite sample distribution is obtained in Table 2 and Table 3 in Engle and Yoo (1987) by Monte Carlo simulation.

STATIONARITY=(ERS)

STATIONARITY=(ERS=(*value*)

STATIONARITY=(NP)

STATIONARITY=(NP=(*value*)

provides a class of *efficient unit root tests*, because they reduce the size distortion and improve the power compared with traditional unit root tests such as the augmented Dickey-Fuller and Phillips-Perron tests. Two test statistics are reported with the ERS= suboption: the point optimal test and the DF-GLS test, which are originally proposed in Elliott, Rothenberg, and Stock (1996). Elliott, Rothenberg, and Stock suggest using the Schwarz Bayesian information criterion to select the optimal lag length in the augmented Dickey-Fuller regression. The maximum lag length can be specified by ERS=*value*. The minimum lag length is 3 and the default maximum lag length is 8.

Six tests, namely MZ_α, MSB, MZ_t, the modified point optimal test, the point optimal test, and the DF-GLS test, which are discussed in Ng and Perron (2001), are reported with the NP= suboption. Ng and Perron suggest using the modified AIC to select the optimal lag length in the augmented Dickey-Fuller regression by using GLS detrended data. The maximum lag length can be specified by NP=*value*. The default maximum lag length is 8. The maximum lag length in the ERS tests and Ng-Perron tests cannot exceed $T/2 - 2$, where T is the sample size.

STATIONARITY=(KPSS)

STATIONARITY=(KPSS=(KERNEL=(*type*))

STATIONARITY=(KPSS=(KERNEL=(*type* TRUNCPOINTMETHOD))

produce the Kwiatkowski, Phillips, Schmidt, and Shin (1992) (KPSS) unit root test or Shin (1994) cointegration test.

Unlike the null hypothesis of the Dickey-Fuller and Phillips-Perron tests, the null hypothesis of the KPSS states that the time series is stationary. As a result, it tends to reject a random walk more often. If the model does not have an intercept, the KPSS option performs the KPSS test for three null hypothesis cases: zero mean, single mean, and deterministic trend. Otherwise, it reports the single mean and deterministic trend only. It computes a test statistic and provides p-value (Hobijn, Franses, and Ooms 2004) for the hypothesis that the random walk component of the time series is equal to zero in the following cases (for more information, see the section "Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Unit Root Test and Shin Cointegration Test" on page 392):

Zero mean — computes the KPSS test statistic based on the zero mean autoregressive model.

$$y_t = u_t$$

Single mean — computes the KPSS test statistic based on the autoregressive model with a constant term.

$$y_t = \mu + u_t$$

Trend — computes the KPSS test statistic based on the autoregressive model with constant and time trend terms.

$$y_t = \mu + \delta t + u_t$$

This test depends on the long-run variance of the series being defined as

$$\sigma_{Tl}^2 = \frac{1}{T}\sum_{i=1}^{T}\hat{u}_i^2 + \frac{2}{T}\sum_{s=1}^{l} w_{sl} \sum_{t=s+1}^{T} \hat{u}_t \hat{u}_{t-s}$$

where w_{sl} is a kernel, s is a maximum lag (truncation point), and $\hat{u}_t$ are OLS residuals or original data series. You can specify two types of the kernel:

KERNEL=NW | BART Newey-West (or Bartlett) kernel

$$w(s,l) = 1 - \frac{s}{l+1}$$

KERNEL=QS Quadratic spectral kernel

$$w(s/l) = w(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$$

You can set the truncation point l by using three different methods:

SCHW=c Schwert maximum lag formula

$$l = \max\left\{1, \text{floor}\left[c\left(\frac{T}{100}\right)^{1/4}\right]\right\}$$

LAG=*l* LAG=*l* manually defined number of lags.

AUTO Automatic bandwidth selection (Hobijn, Franses, and Ooms 2004) (for more information, see the section "Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Unit Root Test and Shin Cointegration Test" on page 392).

If STATIONARITY=KPSS is defined without additional parameters, the Newey-West kernel is used. For the Newey-West kernel the default is the Schwert truncation point method with $c = 12$. For the quadratic spectral kernel the default is AUTO.

The KPSS test can be used in general time series models because its limiting distribution is derived in the context of a class of weakly dependent and heterogeneously distributed data. The limiting probability for the KPSS test is computed assuming that error disturbances are normally distributed. The p-values that are reported are based on the simulation of the limiting probability for the KPSS test.

To test for stationarity of a variable, y, by using default KERNEL=NW and SCHW=12, you can use the following statements:

```
/*-- test for stationarity of regression residuals --*/
proc autoreg data=a;
   model y=  / stationarity = (KPSS);
run;
```

To test for stationarity of a variable, y, by using quadratic spectral kernel and automatic bandwidth selection, you can use the following statements:

```
/*-- test for stationarity using quadratic
     spectral kernel and automatic bandwidth selection --*/
proc autoreg data=a;
   model y= /
   stationarity = (KPSS=(KERNEL=QS AUTO));
run;
```

If there are regressors in the MODEL statement except for the intercept, the Shin (1994) cointegration test, an extension of the KPSS test, is carried out. The limiting distribution of the tests, and then the reported p-values, are different from those in the KPSS tests. For more information, see the section "Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Unit Root Test and Shin Cointegration Test" on page 392.

STATIONARITY=(PHILLIPS)

STATIONARITY=(PHILLIPS=(*value . . . value***)**

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, $\hat{Z}_\rho$ and $\hat{Z}_t$—in the original paper, Phillips and Perron (1988), they are referred to as $\hat{Z}_\alpha$ and $\hat{Z}_t$)—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:

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

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

$$y_t = \mu + \rho y_{t-1} + u_t$$

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

$$y_t = \mu + \rho y_{t-1} + \delta t + u_t$$

You can specify several truncation points l for weighted variance estimators by using the PHILLIPS=($l_1 \ldots l_n$) specification. The statistic for each truncation point l is computed as

$$\sigma^2_{Tl} = \frac{1}{T}\sum_{i=1}^{T}\hat{u}_i^2 + \frac{2}{T}\sum_{s=1}^{l} w_{sl} \sum_{t=s+1}^{T} \hat{u}_t \hat{u}_{t-s}$$

where $w_{sl} = 1 - s/(l+1)$ and $\hat{u}_t$ are OLS residuals. If you specify the PHILLIPS option without specifying truncation points, the default truncation point is $\max(1, \sqrt{T}/5)$, where T is the number of observations.

The Phillips-Perron test can be used in general time series models because its limiting distribution is derived in the context of a class of weakly dependent and heterogeneously distributed data. The marginal probability for the Phillips-Perron test is computed assuming that error disturbances are normally distributed.

When there are regressors in the MODEL statement, the PHILLIPS option computes the Phillips-Ouliaris cointegration test statistic by using the least squares residuals. The normalized cointegrating vector is estimated using OLS regression. Therefore, the cointegrating vector estimates might vary with the regressand (normalized element) unless the regression R-square is 1. You can define the truncation points in the calculation of weighted variance estimators, $\sigma^2_{Tl}, l = l_1 \ldots l_n$, in the same way as you define the truncation points for the Phillips-Perron test—by using the PHILLIPS=($l_1 \ldots l_n$) option.

The marginal probabilities for cointegration testing are not produced. You can refer to Phillips and Ouliaris (1990) tables Ia–Ic for the $\hat{Z}_\alpha$ test and tables IIa–IIc for the $\hat{Z}_t$ test. The standard residual-based cointegration test can be obtained using the NOINT option in the MODEL statement, and the demeaned test is computed by including the intercept term. To obtain the demeaned and detrended cointegration tests, you should include the time trend variable in the regressors. For information about the Phillips-Ouliaris cointegration test, see Phillips and Ouliaris (1990) or Hamilton (1994, Tbl. 19.1). Note that Hamilton (1994, Tbl. 19.1) uses Z_ρ and Z_t instead of the original Phillips and Ouliaris (1990) notation. This chapter adopts the notation introduced in Hamilton. To distinguish from Student's t distribution, these two statistics are named accordingly as ρ (rho) and τ (tau).

TP

TP=*(***Z=***value)*

specifies the turning point test for independence. The Z= suboption specifies the type of the time series or residuals to be tested. You can specify the following *values*:

Y specifies the regressand. The default is Z=Y.

RO specifies the OLS residuals.

R specifies the residuals of the final model.

RM specifies the structural residuals of the final model.

SR specifies the standardized residuals of the final model, defined by residuals over the square root of the conditional variance.

URSQ

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 *URSQ, where T is the number of observations used in estimation.

VNRRANK

VNRRANK=*(option-list)*

specifies the rank version of the von Neumann ratio test for independence. You can specify the following options in the VNRRANK=( ) option. The options are listed within parentheses and separated by commas.

PVALUE=DIST | SIM

specifies how to calculate the *p*-value. You can specify the following values:

DIST calculates the *p*-value according to the asymptotic distribution of the statistic (that is, the standard normal distribution).

SIM calculates the *p*-value as follows:

- If the sample size is less than or equal to 10, the *p*-value is calculated according to the exact CDF of the statistic.
- If the sample size is between 11 and 100, the *p*-value is calculated according to Monte Carlo simulation of the distribution of the statistic.
- If the sample size is more than 100, the *p*-value is calculated according to the standard normal distribution because the simulated distribution of the statistic in this case is almost the same as the standard normal distribution.

By default, PVALUE=DIST.

Z=*value*

specifies the type of the time series or residuals to be tested. You can specify the following *values*:

Y specifies the regressand.

RO specifies the OLS residuals.

R specifies the residuals of the final model.

RM specifies the structural residuals of the final model.

SR specifies the standardized residuals of the final model, defined by residuals over the square root of the conditional variance.

By default, Z=Y.

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. You can use the BACKSTEP option with all estimation methods because the initial parameter values for other estimation methods are estimated by using the Yule-Walker method.

SLSTAY=*value*

specifies the significance level criterion to be used by the BACKSTEP option. By default, 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 criterion, then convergence is assumed. By default, CONVERGE=.001.

If you specify the GARCH= option or the HETERO statement, convergence is assumed when the absolute maximum gradient is smaller than the value specified by the CONVERGE= option or when the relative gradient is smaller than 1E–8. By default, CONVERGE=1E–5.

INITIAL=*(initial-values)*

START=*(initial-values)*

specifies initial values for some or all of the parameter estimates. This option is not applicable when the Yule-Walker method or iterative Yule-Walker method is used. The specified values are assigned to model parameters in the same order in which the parameter estimates are printed in the AUTOREG procedure output. The order of values in the INITIAL= or START= option is as follows: 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 η 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:

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

The model that is specified by this MODEL statement is

$$y_t = \beta_0 + \beta_1 w_t + \beta_2 x_t + v_t$$

$$v_t = \epsilon_t - \phi_1 v_{t-1}$$

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

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

$$\epsilon_t \ \mathrm{N}(0, \sigma_t^2)$$

The initial values for the regression parameters, INTERCEPT (β_0), X1 (β_1), and X2 (β_2), are specified as 1. The initial value of the AR(1) coefficient (ϕ_1) is specified as 0.5. The initial value of ARCH0 (ω) is 0.8, the initial value of ARCH1 (α_1) is 0.1, and the initial value of GARCH1 (γ_1) is 0.6.

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

LDW
specifies that *p*-values for the Durbin-Watson test be computed by using a linearized approximation of the design matrix when the model is nonlinear because an autoregressive error process is present. (The Durbin-Watson tests of the OLS linear regression model residuals are not affected by the LDW option.) For information about Durbin-Watson testing of nonlinear models, see White (1992).

MAXITER=*number*
sets the maximum number of iterations allowed. The default is MAXITER=50. When you specify both the GARCH= option in the MODEL statement and the MAXITER= option in the NLOPTIONS statement, the MAXITER= option in the MODEL statement is ignored. This option is not applicable when the Yule-Walker method is used.

METHOD=*value*
requests the type of estimates to be computed. You can specify the following *values*:

ML specifies maximum likelihood estimates.

ULS specifies unconditional least squares estimates.

YW specifies Yule-Walker estimates.

ITYW specifies iterative Yule-Walker estimates.

The default is defined as follows:

- When the GARCH= option or the HETERO statement is specified, METHOD=ML by default.
- When the GARCH= option and the HETERO statement are not specified but the NLAG= option and the LAGDEP option are specified, METHOD=ML by default.
- When the GARCH= option, the LAGDEP option, and the HETERO statement are not specified, but the NLAG= option is specified, METHOD=YW by default.
- When none of the NLAG= option, the GARCH= option, and the HETERO statement is specified (that is, only the OLS model is to be estimated), then the estimates are calculated through the OLS method and the METHOD= option is ignored.

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

OPTMETHOD=QN | TR
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

HETERO *variables / options* **;**

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 heteroscedastic regression model supported by the HETERO statement is

$$y_t = \mathbf{x}_t\beta + \epsilon_t$$

$$\epsilon_t \sim \mathrm{N}(0, \sigma_t^2)$$

$$\sigma_t^2 = \sigma^2 h_t$$

$$h_t = l(\mathbf{z}'_t\eta)$$

The HETERO statement specifies a model for the conditional variance h_t. The vector $\mathbf{z}_t$ is composed of the variables listed in the HETERO statement, η is a parameter vector, and $l(\cdot)$ is a link function that depends on the value of the LINK= option. In the printed output, $HET0$ represents the estimate of sigma, while $HET1$ - $HETn$ are the estimates of parameters in the η vector.

The keyword XBETA can be used in the *variables* list to refer to the model predicted value $\mathbf{x}'_t\beta$. If XBETA is specified in the *variables* list, other variables in the HETERO statement will be ignored. In addition, XBETA cannot be specified in the GARCH process.

For heteroscedastic regression models without GARCH effects, the errors ϵ_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, when a HETERO statement is used, the NLAG= option cannot be specified unless the GARCH= option is also specified.

You can specify the following options in the HETERO statement.

LINK=*value*

specifies the functional form of the heteroscedasticity model. By default, LINK=EXP. If you specify a GARCH model with the HETERO statement, the model is estimated using LINK=LINEAR only. For more information, see the section "Using the HETERO Statement with GARCH Models" on page 373. Values of the LINK= option are as follows:

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

$$h_t = \exp(\mathbf{z}'_t\eta)$$

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

$$h_t = (1 + \mathbf{z}'_t\eta)^2$$

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:

$$h_t = (1 + \mathbf{z}'_t\eta)$$

COEF=*value*

imposes constraints on the estimated parameters η of the heteroscedasticity model. You can specify the following *values*:

NONNEG	specifies that the estimated heteroscedasticity parameters η must be nonnegative.
UNIT	constrains all heteroscedasticity parameters η to equal 1.
ZERO	constrains all heteroscedasticity parameters η to equal 0.
UNREST	specifies unrestricted estimation of η.

If you specify the GARCH= option in the MODEL statement, the default is COEF=NONNEG. If you do not specify the GARCH= option in the MODEL statement, the default is COEF=UNREST.

STD=*value*

imposes constraints on the estimated standard deviation σ of the heteroscedasticity model. You can specify the following *values*:

NONNEG	specifies that the estimated standard deviation parameter σ must be nonnegative.
UNIT	constrains the standard deviation parameter σ to equal 1.
UNREST	specifies unrestricted estimation of σ.

The default is STD=UNREST.

TEST=LM

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

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:

$$h_t = (\mathbf{z}'_t \eta)^2$$

NLOPTIONS Statement

NLOPTIONS < *options* > **;**

PROC AUTOREG uses the nonlinear optimization (NLO) subsystem to perform nonlinear optimization tasks when the GARCH= option is specified. If the GARCH= option is not specified, the NLOPTIONS statement is ignored. For a list of all the options of the NLOPTIONS statement, see Chapter 6, "Nonlinear Optimization Methods."

OUTPUT Statement

OUTPUT < **OUT=***SAS-data-set* > < *options* > < *keyword=name* > **;**

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 to contain the predicted and transformed values. If the OUT= option is not specified, the new data set is named according to the DATA*n* convention.

You can specify any of the following *options*:

ALPHACLI=*number*
sets the confidence limit size for the estimates of future values of the response time series. The ALPHACLI= value must be between 0 and 1. The resulting confidence interval has 1–*number* confidence. The default is ALPHACLI=0.05, which corresponds 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–*number* confidence. The default is ALPHACLM=0.05, which corresponds to a 95% confidence interval.

ALPHACSM=0.01 | **0.05** | **0.10**
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 0.01, 0.05, or 0.10. Other values are not supported.

You can specify the following values for *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 to contain the statistic.

BLUS=*variable*
specifies the name of a variable to contain the values of the Theil's BLUS residuals. For more information about BLUS residuals, see Theil (1971).

CEV=*variable*

HT=*variable*
writes to the output data set the value of the error variance σ_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.

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 more information, see the section "Predicted Values" on page 406.

CONSTANT=*variable*
writes the transformed intercept to the output data set. For information about the transformation, see the section "Computational Methods" on page 368.

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

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

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

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

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

CUSUMSQLB=*variable*
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 ALPHACLI= option. For more information, see the section "Predicted Values" on page 406.

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 ALPHACLM= option.

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. For more information, see the section "Predicted Values" on page 406.

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. For more information, see the section "Predicted Values" on page 406.

RECPEV=*variable*
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*
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. For information about the transformation, see the section "Computational Methods" on page 368. If you need to reproduce the data suitable for re-estimation, 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. For more information, see the section "Predicted Values" on page 406.

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.

RESTRICT Statement

RESTRICT *equation , . . . , equation* **;**

The RESTRICT statement provides constrained estimation and places restrictions on the parameter estimates for covariates in the preceding MODEL statement. The AR, GARCH, and HETERO parameters are also supported in the RESTRICT statement when you specify the GARCH= option. Any number of RESTRICT statements can follow a MODEL statement. To specify more than one restriction in a single RESTRICT statement, separate them 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 that is 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. For the names of these parameters, see the section "OUTEST= Data Set" on page 410. Inequality constraints are supported only when you specify the GARCH= option. For non-GARCH models, if inequality signs are specified, they are treated as equality signs.

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;
```

The following example shows how to specify AR, GARCH, and HETERO parameters in the RESTRICT statement:

```
model y = a b / nlag=2 garch=(p=2,q=3,mean=sqrt);
hetero c d;
restrict _A_1=0,_AH_2=0.2,_HET_2=1,_DELTA_=0.1;
```

You can specify only simple linear combinations of parameters in RESTRICT statement expressions. You cannot specify complex expressions that involve parentheses, division, functions, or complex products.

TEST Statement

The AUTOREG procedure supports a TEST statement for linear hypothesis tests. The syntax of the TEST statement is

TEST *equation , . . . , equation / option* **;**

The TEST statement tests hypotheses about the covariates in the model that are estimated by the preceding MODEL statement. The AR, GARCH, and HETERO parameters are also supported in the TEST statement when you specify the GARCH= option. Each equation specifies a linear hypothesis to be tested. If you specify more than one equation, separate them with 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 that is 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. For the names of these parameters, see the section "OUTEST= Data Set" on page 410.

You can specify the following options in the TEST statement:

TYPE=*value*
specifies the test statistics to use. The default is TYPE=F. The following values for the TYPE= option are available:

F produces an *F* test. This option is supported for all models specified in the MODEL statement.

WALD produces a Wald test. This option is supported for all models specified in the MODEL statement.

LM produces a Lagrange multiplier test. This option is supported only when the GARCH= option is specified (for example, when there is a statement like MODEL Y = C D I / GARCH=(Q=2)).

LR produces a likelihood ratio test. This option is supported only when the GARCH= option is specified (for example, when there is a statement like MODEL Y = C D I / GARCH=(Q=2)).

ALL produces all tests applicable for a particular model. For non-GARCH-type models, only *F* and Wald tests are output. For all other models, all four tests (LR, LM, *F*, and Wald) are computed.

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:

$$y_t = \beta_0 + \beta_1 x1_t + \beta_2 x2_t + \epsilon_t$$

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

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

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

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

To illustrate the TYPE= option, consider the following examples:

```
model Y = C D I / garch=(q=2);
test C + D = 1;
```

The preceding statements produce only one default test, the *F* test.

```
model Y = C D I / garch=(q=2);
test C + D = 1 / type = LR;
```

The preceding statements produce one of four tests applicable for GARCH-type models, the likelihood ratio test.

```
model Y = C D I / nlag = 2;
test C + D = 1 / type = LM;
```

The preceding statements produce the warning and do not output any test because the Lagrange multiplier test is not applicable for non-GARCH models.

```
model Y = C D I / nlag=2;
test C + D = 1 / type = ALL;
```

The preceding statements produce all tests that are applicable for non-GARCH models (namely, the *F* and Wald tests). The TYPE= prefix is optional. Thus the test statement in the previous example could also have been written as

```
test C + D = 1 / ALL;
```

The following example shows how to test AR, GARCH, and HETERO parameters:

```
model y = a b / nlag=2 garch=(p=2,q=3,mean=sqrt);
hetero c d;
test _A_1=0,_AH_2=0.2,_HET_2=1,_DELTA_=0.1;
```

Details: AUTOREG Procedure

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:

$$y_t = \mathbf{x}_t'\beta + \nu_t$$

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

$$\epsilon_t \ \mathrm{N}(0, \sigma^2)$$

In these equations, y_t are the dependent values, $\mathbf{x}_t$ is a column vector of regressor variables, β is a column vector of structural parameters, and ϵ_t is normally and independently distributed with a mean of 0 and a variance of σ^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. For further discussion of the advantages of different methods, see the section "Alternative Autocorrelation Correction Methods" on page 370, later in this chapter.

The Yule-Walker Method

Let $\boldsymbol{\varphi}$ represent the vector of autoregressive parameters,

$$\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_m)'$$

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

$$E(\nu\nu') = \boldsymbol{\Sigma} = \sigma^2\mathbf{V}$$

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

The Yule-Walker method alternates estimation of β using generalized least squares with estimation of $\boldsymbol{\varphi}$ using the Yule-Walker equations applied to the sample autocorrelation function. The YW method starts by forming the OLS estimate of β. Next, $\boldsymbol{\varphi}$ is estimated from the sample autocorrelation function of the OLS residuals by using the Yule-Walker equations. Then $\mathbf{V}$ is estimated from the estimate of $\boldsymbol{\varphi}$, and $\boldsymbol{\Sigma}$ is estimated from $\mathbf{V}$ and the OLS estimate of σ^2. The autocorrelation corrected estimates of the regression parameters β are then computed by GLS, using the estimated $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\varphi}$ and $\mathbf{V}$, and the GLS estimates are recomputed using the new variance matrix. This alternation of estimates continues until either the maximum change in the $\widehat{\varphi}$ 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 $\widehat{\varphi}$ and a preliminary estimate of σ^2, are

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

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

If the BACKSTEP option is specified, for purposes of significance testing, the matrix $[\mathbf{R}\ \mathbf{r}]$ 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, $\mathbf{e}$, as

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

where $\mathbf{n} = \mathbf{y} - \mathbf{X}\beta$.

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

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

The ULS estimates are computed by minimizing S with respect to the parameters β and φ_i.

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

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

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

Maximizing l with respect to σ^2 (and concentrating σ^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

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

Rewriting the variable term within the logarithm gives

$$S_{ml} = |\mathbf{L}|^{1/N}\mathbf{e}'\mathbf{e}|\mathbf{L}|^{1/N}$$

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

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 $\mathbf{n_t} = y_t - \mathbf{x}_t'\mathbf{b}$, where $\mathbf{b}$ is the current estimate of β. The sample autocorrelation function is the sum of all available lagged products of $\mathbf{n}_t$ of order j divided by $\ell + j$, where ℓ 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, $\mathbf{R}$, is at least positive semidefinite. If there are missing values, these autocorrelation estimates of r can yield an $\mathbf{R}$ 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 $\mathbf{R}$ is positive definite.

Data Transformation and the Kalman Filter

The calculation of $\mathbf{V}$ from $\boldsymbol{\varphi}$ for the general AR(m) model is complicated, and the size of $\mathbf{V}$ depends on the number of observations. Instead of actually calculating $\mathbf{V}$ 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 $\mathbf{V}$. Let $\mathbf{L}$ denote the Cholesky root of $\mathbf{V}$—that is, $\mathbf{V} = \mathbf{L}\mathbf{L}'$ with $\mathbf{L}$ lower triangular. For an AR(m) model, $\mathbf{L}^{-1}$ 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

$$z_t = x_t + \hat{\varphi}_1 x_{t-1} + \cdots + \hat{\varphi}_m x_{t-m}$$

The transformation is carried out using a Kalman filter, and the lower triangular matrix $\mathbf{L}$ is never directly computed. The Kalman filter algorithm, as it applies here, is described in Harvey and Phillips (1979) and Jones (1980). Although $\mathbf{L}$ is not computed explicitly, for ease of presentation the remaining discussion is in terms of $\mathbf{L}$. If there are missing values, then the submatrix of $\mathbf{L}$ 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 β and the Yule-Walker estimates of $\boldsymbol{\varphi}$ are used as starting values for these methods.

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

$$\frac{\partial \mathbf{e}}{\partial \beta'} = -\mathbf{L}^{-1}\mathbf{X}$$

$$\frac{\partial |\mathbf{L}|^{1/N}\mathbf{e}}{\partial \boldsymbol{\beta}'} = -|\mathbf{L}|^{1/N}\mathbf{L}^{-1}\mathbf{X}$$

These derivatives are computed by the transformation described previously. The derivatives with respect to $\boldsymbol{\varphi}$ 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 $\mathbf{b}$ is taken as $s^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ for the Yule-Walker method. For the ULS and ML methods, the variance-covariance matrix of the parameter estimates is computed as $s^2(\mathbf{J}'\mathbf{J})^{-1}$. For the ULS method, $\mathbf{J}$ is the matrix of derivatives of $\mathbf{e}$ with respect to the parameters. For the ML method, $\mathbf{J}$ is the matrix of derivatives of $|\mathbf{L}|^{1/N}\mathbf{e}$ divided by $|\mathbf{L}|^{1/N}$. The estimate of the variance-covariance matrix of $\mathbf{b}$ assuming that $\boldsymbol{\varphi}$ is known is $s^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. For OLS model, the estimate of the variance-covariance matrix is $s^2(\mathbf{X}'\mathbf{X})^{-1}$.

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 method 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 (1980) 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, 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 and McAvinchey used 100), losing the first few observations does not make much difference.

GARCH Models

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

$$y_t|\Psi_{t-1} \sim \mathrm{N}(0, h_t)$$

where Ψ_{t-1} denotes all available information at time $t-1$. The conditional variance h_t is

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

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

$$y_t = \mathbf{x}_t'\beta + \epsilon_t$$

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

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

where $e_t \sim \mathrm{IN}(0,1)$.

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

$$y_t = \mathbf{x}_t'\beta + \nu_t$$

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

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

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

GARCH Estimation with Nelson-Cao Inequality Constraints

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

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

where B is a backshift operator. Therefore, $h_t \geq 0$ 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(∞) process are well defined.

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

$$\begin{aligned} \phi_0 &= \alpha_1 \\ \phi_1 &= \gamma_1 \phi_0 + \alpha_2 \\ &\cdots \\ \phi_{n-1} &= \gamma_1 \phi_{n-2} + \gamma_2 \phi_{n-3} + \cdots + \gamma_{n-1} \phi_0 + \alpha_n \\ \phi_k &= \gamma_1 \phi_{k-1} + \gamma_2 \phi_{k-2} + \cdots + \gamma_n \phi_{k-n} \text{ for } k \geq n \end{aligned}$$

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

$$\begin{aligned} \omega &\geq 0 \\ \gamma_1 &\geq 0 \\ \phi_k &\geq 0 \text{ for } k = 0, 1, \ldots, q-1 \end{aligned}$$

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

$$\begin{aligned} \Delta_i &\in R \text{ for } i = 1, 2 \\ \omega^* &\geq 0 \\ \Delta_1 &> 0 \\ \sum_{j=0}^{q-1} \Delta_1^{-j} \alpha_{j+1} &> 0 \\ \phi_k &\geq 0 \text{ for } k = 0, 1, \ldots, q \end{aligned}$$

where Δ_1 and Δ_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.

Using the HETERO Statement with GARCH Models

The HETERO statement can be combined with the GARCH= option in 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

$$\begin{aligned} \epsilon_t &= \sqrt{h_t} e_t \\ h_t &= \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 h_{t-1} + \eta_1 D1_t + \eta_2 D2_t \end{aligned}$$

The following statements estimate this GARCH model:

```
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:

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

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. When the GARCH process is stationary, the unconditional variance of ϵ_t is computed as

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

where $\epsilon_t = \sqrt{h_t} e_t$ 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. For more information, see Nelson (1990).

EGARCH Model

The EGARCH 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, α_i and γ_j, while there are no restrictions on these parameters in the EGARCH model. In the EGARCH model, the conditional variance, h_t, is an asymmetric function of lagged disturbances ϵ_{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 (γ=1) in our formulation. Note that $E|z_t| = (2/\pi)^{1/2}$ if $z_t \sim \mathrm{N}(0, 1)$. 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, ϵ_t, cause the innovation to the conditional variance to be positive if θ is much less than 1.

QGARCH, TGARCH, and PGARCH Models

As shown in many empirical studies, positive and negative innovations have different impacts on future volatility. There is a long list of variations of GARCH models that consider the asymmetricity. Three typical variations are the quadratic GARCH (QGARCH) model (Engle and Ng 1993), the threshold GARCH (TGARCH) model (Glosten, Jaganathan, and Runkle 1993; Zakoian 1994), and the power GARCH (PGARCH) model (Ding, Granger, and Engle 1993). For more information about the asymmetric GARCH models, see Engle and Ng (1993).

In the QGARCH model, the lagged errors' centers are shifted from zero to some constant values:

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

In the TGARCH model, there is an extra slope coefficient for each lagged squared error,

$$h_t = \omega + \sum_{i=1}^{q} (\alpha_i + 1_{\epsilon_{t-i}<0}\psi_i)\epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j}$$

where the indicator function $1_{\epsilon_t<0}$ is one if $\epsilon_t < 0$; otherwise, zero.

The PGARCH model not only considers the asymmetric effect, but also provides another way to model the long memory property in the volatility,

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

where $\lambda > 0$ and $|\psi_i| \leq 1, i = 1, \ldots, q$.

Note that the implemented TGARCH model is also well known as GJR-GARCH (Glosten, Jaganathan, and Runkle 1993), which is similar to the threshold GARCH model proposed by Zakoian (1994) but not exactly the same. In Zakoian's model, the conditional standard deviation is a linear function of the past values of the white noise. Zakoian's version can be regarded as a special case of the PGARCH model when $\lambda = 1/2$.

GARCH-in-Mean

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

$$y_t = \mathbf{x}_t'\beta + \delta\sqrt{h_t} + \epsilon_t$$

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

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 e_t is assumed to have a standard normal distribution ($e_t \sim \mathrm{N}(0,1)$), the log-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 $\epsilon_t = y_t - \mathbf{x}_t'\beta$ and h_t is the conditional variance. When the GARCH(p,q)-M model is estimated, $\epsilon_t = y_t - \mathbf{x}_t'\beta - \delta\sqrt{h_t}$. When there are no regressors, the residuals ϵ_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

$$\ell = \sum_{t=1}^{N}\left[\ln\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}\ln((\nu-2)\pi h_t)\right.$$

$$\left. -\frac{1}{2}(\nu+1)\ln\left(1 + \frac{\epsilon_t^2}{h_t(\nu-2)}\right)\right]$$

where $\Gamma(\cdot)$ is the gamma function and ν 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 $1/\nu \rightarrow 0$.

The likelihood function is maximized via either the dual quasi-Newton or the trust region algorithm. The default is the dual quasi-Newton algorithm. The starting values for the regression parameters β 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, $\mathbf{q}(\theta) = \mathbf{0}$, the variance-covariance matrix is given by

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

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

Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrix Estimator

The heteroscedasticity-consistent covariance matrix estimator (HCCME), also known as the sandwich (or robust or empirical) covariance matrix estimator, has been popular in recent years because it gives the consistent estimation of the covariance matrix of the parameter estimates even when the heteroscedasticity structure might be unknown or misspecified. White (1980) proposes the concept of HCCME, known as HC0. However, the small-sample performance of HC0 is not good in some cases. Davidson and MacKinnon (1993) introduce more improvements to HC0, namely HC1, HC2 and HC3, with the degrees-of-freedom or leverage adjustment. Cribari-Neto (2004) proposes HC4 for cases that have points of high leverage.

HCCME can be expressed in the following general "sandwich" form,

$$\Sigma = B^{-1}MB^{-1}$$

where B, which stands for "bread," is the Hessian matrix and M, which stands for "meat," is the outer product of gradient (OPG) with or without adjustment. For HC0, M is the OPG without adjustment; that is,

$$M_{\mathrm{HC0}} = \sum_{t=1}^{T} g_t g_t'$$

where T is the sample size and g_t is the gradient vector of tth observation. For HC1, M is the OPG with the degrees-of-freedom correction; that is,

$$M_{\mathrm{HC1}} = \frac{T}{T-k}\sum_{t=1}^{T} g_t g_t'$$

where k is the number of parameters. For HC2, HC3, and HC4, the adjustment is related to leverage, namely,

$$M_{\mathrm{HC2}} = \sum_{t=1}^{T} \frac{g_t g_t'}{1-h_{tt}} \qquad M_{\mathrm{HC3}} = \sum_{t=1}^{T} \frac{g_t g_t'}{(1-h_{tt})^2} \qquad M_{\mathrm{HC4}} = \sum_{t=1}^{T} \frac{g_t g_t'}{(1-h_{tt})^{\min(4, Th_{tt}/k)}}$$

The leverage h_{tt} is defined as $h_{tt} \equiv j_t'(\sum_{t=1}^{T} j_t j_t')^{-1} j_t$, where j_t is defined as follows:

- For an OLS model, j_t is the tth observed regressors in column vector form.
- For an AR error model, j_t is the derivative vector of the tth residual with respect to the parameters.
- For a GARCH or heteroscedasticity model, j_t is the gradient of the tth observation (that is, g_t).

The heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator can also be expressed in "sandwich" form,

$$\Sigma = B^{-1}MB^{-1}$$

where B is still the Hessian matrix, but M is the kernel estimator in the following form:

$$M_{\mathrm{HAC}} = a\left(\sum_{t=1}^{T} g_t g_t' + \sum_{j=1}^{T-1} k\left(\frac{j}{b}\right)\sum_{t=1}^{T-j}\left(g_t g_{t+j}' + g_{t+j} g_t'\right)\right)$$

where T is the sample size, g_t is the gradient vector of tth observation, $k(.)$ is the real-valued kernel function, b is the bandwidth parameter, and a is the adjustment factor of small-sample degrees of freedom (that is, $a = 1$ if ADJUSTDF option is not specified and otherwise $a = T/(T-k)$, where k is the number of parameters). The types of kernel functions are listed in Table 8.2.

Table 8.2 Kernel Functions

Kernel Name	Equation
Bartlett	$k(x) = \begin{cases} 1-\lvert x\rvert & \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$
Parzen	$k(x) = \begin{cases} 1-6x^2+6\lvert x\rvert^3 & 0 \le \lvert x\rvert \le 1/2 \\ 2(1-\lvert x\rvert)^3 & 1/2 \le \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$
Quadratic spectral	$k(x) = \frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$
Truncated	$k(x) = \begin{cases} 1 & \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$
Tukey-Hanning	$k(x) = \begin{cases} (1+\cos(\pi x))/2 & \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$

When you specify BANDWIDTH=ANDREWS91, according to Andrews (1991) the bandwidth parameter is estimated as shown in Table 8.3.

Table 8.3 Bandwidth Parameter Estimation

Kernel Name	Bandwidth Parameter
Bartlett	$b = 1.1447(\alpha(1)T)^{1/3}$
Parzen	$b = 2.6614(\alpha(2)T)^{1/5}$
Quadratic spectral	$b = 1.3221(\alpha(2)T)^{1/5}$
Truncated	$b = 0.6611(\alpha(2)T)^{1/5}$
Tukey-Hanning	$b = 1.7462(\alpha(2)T)^{1/5}$

Let $\{g_{at}\}$ denote each series in $\{g_t\}$, and let (ρ_a, σ_a^2) denote the corresponding estimates of the autoregressive and innovation variance parameters of the AR(1) model on $\{g_{at}\}$, $a = 1, \ldots, k$, where the AR(1) model is parameterized as $g_{at} = \rho g_{at-1} + \epsilon_{at}$ with $Var(\epsilon_{at}) = \sigma_a^2$. The factors $\alpha(1)$ and $\alpha(2)$ are estimated with the formulas

$$\alpha(1) = \frac{\sum_{a=1}^{k} \frac{4\rho_a^2\sigma_a^4}{(1-\rho_a)^6(1+\rho_a)^2}}{\sum_{a=1}^{k} \frac{\sigma_a^4}{(1-\rho_a)^4}} \qquad \alpha(2) = \frac{\sum_{a=1}^{k} \frac{4\rho_a^2\sigma_a^4}{(1-\rho_a)^8}}{\sum_{a=1}^{k} \frac{\sigma_a^4}{(1-\rho_a)^4}}$$

When you specify BANDWIDTH=NEWEYWEST94, according to Newey and West (1994) the bandwidth parameter is estimated as shown in Table 8.4.

Table 8.4 Bandwidth Parameter Estimation

Kernel Name	Bandwidth Parameter
Bartlett	$b = 1.1447(\{s_1/s_0\}^2T)^{1/3}$
Parzen	$b = 2.6614(\{s_1/s_0\}^2T)^{1/5}$
Quadratic spectral	$b = 1.3221(\{s_1/s_0\}^2T)^{1/5}$
Truncated	$b = 0.6611(\{s_1/s_0\}^2T)^{1/5}$
Tukey-Hanning	$b = 1.7462(\{s_1/s_0\}^2T)^{1/5}$

The factors s_1 and s_0 are estimated with the following formulas:

$$s_1 = 2\sum_{j=1}^{n} j\sigma_j \qquad\qquad s_0 = \sigma_0 + 2\sum_{j=1}^{n} \sigma_j$$

where n is the lag selection parameter and is determined by kernels, as listed in Table 8.5.

Table 8.5 Lag Selection Parameter Estimation

Kernel Name	Lag Selection Parameter
Bartlett	$n = c(T/100)^{2/9}$
Parzen	$n = c(T/100)^{4/25}$
Quadratic spectral	$n = c(T/100)^{2/25}$
Truncated	$n = c(T/100)^{1/5}$
Tukey-Hanning	$n = c(T/100)^{1/5}$

The factor c in Table 8.5 is specified by the C= option; by default it is 12.

The factor σ_j is estimated with the equation

$$\sigma_j = T^{-1}\sum_{t=j+1}^{T}\left(\sum_{a=i}^{k} g_{at}\sum_{a=i}^{k} g_{at-j}\right), j = 0,\ldots,n$$

where i is 1 if the NOINT option in the MODEL statement is specified (otherwise, it is 2), and g_{at} is the same as in the Andrews method.

If you specify BANDWIDTH=SAMPLESIZE, the bandwidth parameter is estimated with the equation

$$b = \begin{cases} \lfloor \gamma T^r + c \rfloor & \text{if BANDWIDTH=SAMPLESIZE(INT) option is specified} \\ \gamma T^r + c & \text{otherwise} \end{cases}$$

where T is the sample size; $\lfloor x \rfloor$ is the largest integer less than or equal to x; and γ, r, and c are values specified by the BANDWIDTH=SAMPLESIZE(GAMMA=, RATE=, CONSTANT=) options, respectively.

If you specify the PREWHITENING option, g_t is prewhitened by the VAR(1) model,

$$g_t = Ag_{t-1} + w_t$$

Then M is calculated by

$$M_{\mathrm{HAC}} = a(I-A)^{-1}\left(\sum_{t=1}^{T} w_t w_t' + \sum_{j=1}^{T-1} k\left(\frac{j}{b}\right)\sum_{t=1}^{T-j}\left(w_t w_{t+j}' + w_{t+j} w_t'\right)\right)\left((I-A)^{-1}\right)'$$

The bandwidth calculation is also based on the prewhitened series w_t.

Goodness-of-Fit Measures and Information Criteria

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

Total R-Square Statistic

The total R-square statistic (Total Rsq) is computed as

$$\mathrm{R}^2_{\mathrm{tot}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$$

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.

Transformed Regression R-Square Statistic

The transformed regression R-square statistic is computed as

$$\mathrm{R}^2_{\mathrm{tr}} = 1 - \frac{\mathrm{TSSE}}{\mathrm{TSST}}$$

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 transformed regression R-square statistic is a measure of the fit of the structural part of the model after transforming for the autocorrelation and is the R-square for the transformed regression.

Mean Absolute Error and Mean Absolute Percentage Error

The mean absolute error (MAE) is computed as

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T} |e_t|$$

where e_t are the estimated model residuals and T is the number of observations.

The mean absolute percentage error (MAPE) is computed as

$$\mathrm{MAPE} = \frac{1}{T'}\sum_{t=1}^{T} \delta_{y_t \neq 0} \frac{|e_t|}{|y_t|}$$

where e_t are the estimated model residuals, y_t are the original response variable observations, $\delta_{y_t \neq 0} = 1$ if $y_t \neq 0$, $\delta_{y_t \neq 0}\,|e_t / y_t| = 0$ if $y_t = 0$, and T' is the number of nonzero original response variable observations.

Calculation of Recursive Residuals and CUSUM Statistics

The recursive residuals w_t are computed as

$$w_t = \frac{e_t}{\sqrt{v_t}}$$

$$e_t = y_t - x_t' \beta^{(t)}$$

$$\beta^{(t)} = \left[\sum_{i=1}^{t-1} \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \left(\sum_{i=1}^{t-1} \mathbf{x}_i y_i\right)$$

$$v_t = 1 + \mathbf{x}'_t \left[\sum_{i=1}^{t-1} \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \mathbf{x}_t$$

Note that the first $\beta^{(t)}$ can be computed for $t = p + 1$, where p is the number of regression coefficients. As a result, first p recursive residuals are not defined. Note also 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

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

where $a = 1.143$ for a significance level 0.01, 0.948 for 0.05, and 0.850 for 0.10. 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, AICC, SBC, and HQC

Akaike's information criterion (AIC), the corrected Akaike's information criterion (AICC), Schwarz's Bayesian information criterion (SBC), and the Hannan-Quinn information criterion (HQC) are computed as follows:

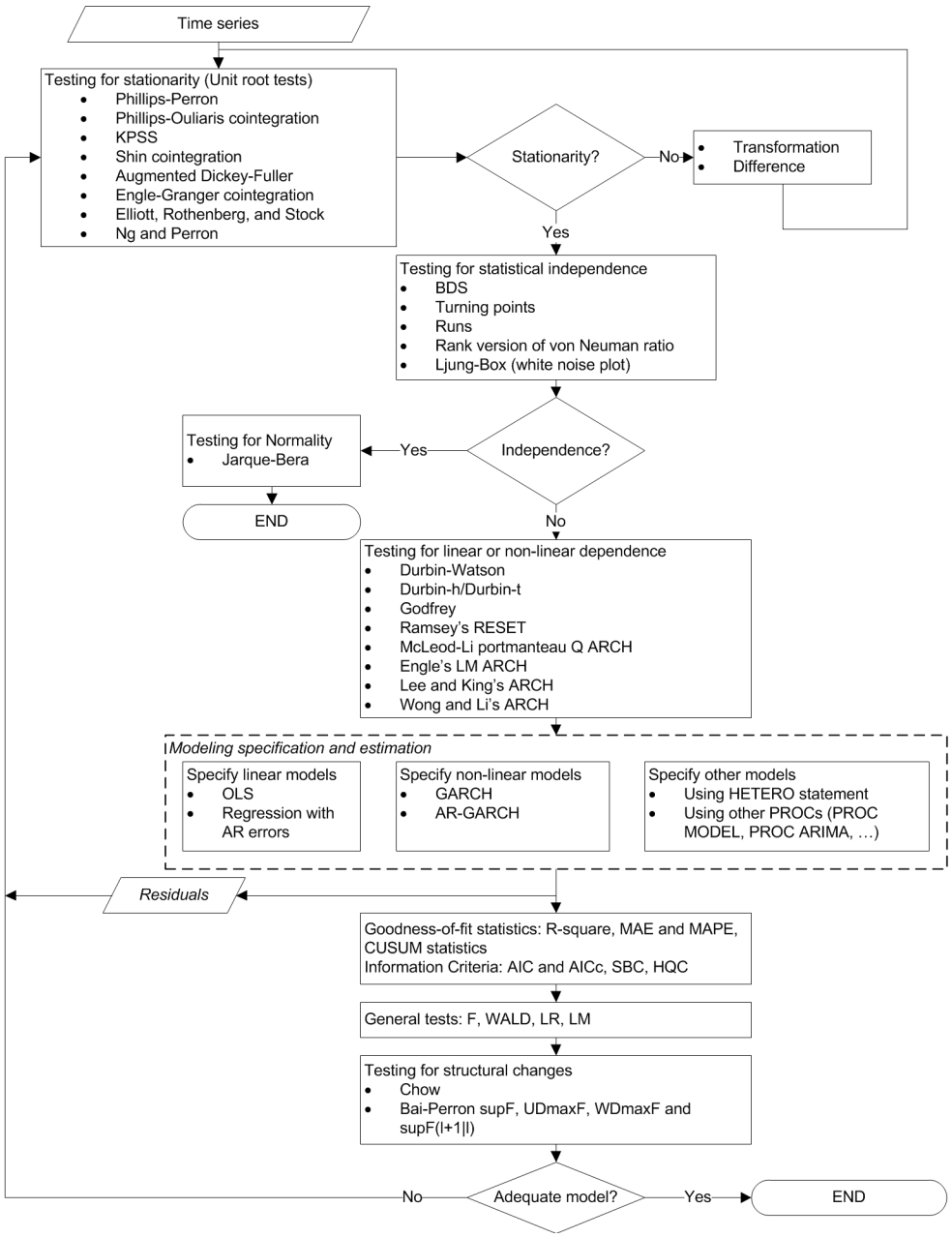
$$\begin{aligned}\text{AIC} &= -2\ln(L)+2k\\ \text{AICC} &= \text{AIC}+2\frac{k(k+1)}{N-k-1}\\ \text{SBC} &= -2\ln(L)+\ln(N)k\\ \text{HQC} &= -2\ln(L)+2\ln(\ln(N))k\end{aligned}$$

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. For more information, see Judge et al. (1985), Hurvich and Tsai (1989), Schwarz (1978) and Hannan and Quinn (1979).

Testing

The modeling process consists of four stages: identification, specification, estimation, and diagnostic checking (Cromwell, Labys, and Terraza 1994). The AUTOREG procedure supports tens of statistical tests for identification and diagnostic checking. Figure 8.15 illustrates how to incorporate these statistical tests into the modeling process.

Figure 8.15 Statistical Tests in the AUTOREG Procedure



Testing for Stationarity

Most of the theories of time series require stationarity; therefore, it is critical to determine whether a time series is stationary. Two nonstationary time series are fractionally integrated time series and autoregressive series with random coefficients. However, more often some time series are nonstationary due to an upward trend over time. The trend can be captured by either of the following two models.

- The *difference stationary* process

$$(1 - L)y_t = \delta + \psi(L)\epsilon_t$$

where L is the lag operator, $\psi(1) \neq 0$, and ϵ_t is a white noise sequence with mean zero and variance σ^2. Hamilton (1994) also refers to this model the *unit root* process.

- The *trend stationary* process

$$y_t = \alpha + \delta t + \psi(L)\epsilon_t$$

When a process has a unit root, it is said to be integrated of order one or I(1). An I(1) process is stationary after differencing once. The trend stationary process and difference stationary process require different treatment to transform the process into stationary one for analysis. Therefore, it is important to distinguish the two processes. Bhargava (1986) nested the two processes into the following general model:

$$y_t = \gamma_0 + \gamma_1 t + \alpha(y_{t-1} - \gamma_0 - \gamma_1(t-1)) + \psi(L)\epsilon_t$$

However, a difficulty is that the right-hand side is nonlinear in the parameters. Therefore, it is convenient to use a different parameterization:

$$y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + \psi(L)\epsilon_t$$

The test of null hypothesis that $\alpha = 1$ against the one-sided alternative of $\alpha < 1$ is called a *unit root test*.

Dickey-Fuller unit root tests are based on regression models similar to the previous model,

$$y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + \epsilon_t$$

where ϵ_t is assumed to be white noise. The t statistic of the coefficient α does not follow the normal distribution asymptotically. Instead, its distribution can be derived using the functional central limit theorem. Three types of regression models including the preceding one are considered by the Dickey-Fuller test. The deterministic terms that are included in the other two types of regressions are either null or constant only.

An assumption in the Dickey-Fuller unit root test is that it requires the errors in the autoregressive model to be white noise, which is often not true. There are two popular ways to account for general serial correlation between the errors. One is the augmented Dickey-Fuller (ADF) test, which uses the lagged difference in the regression model. This was originally proposed by Dickey and Fuller (1979) and later studied by Said and Dickey (1984) and Phillips and Perron (1988). Another method is proposed by Phillips and Perron (1988); it is called Phillips-Perron (PP) test. The tests adopt the original Dickey-Fuller regression with intercept, but modify the test statistics to take account of the serial correlation and heteroscedasticity. It is called nonparametric because no specific form of the serial correlation of the errors is assumed.

A problem of the augmented Dickey-Fuller and Phillips-Perron unit root tests is that they are subject to size distortion and low power. It is reported in Schwert (1989) that the size distortion is significant when the series contains a large moving average (MA) parameter. DeJong et al. (1992) find that the ADF has power around one third and PP test has power less than 0.1 against the trend stationary alternative, in some common settings. Among some more recent unit root tests that improve upon the size distortion and the low power are the tests described by Elliott, Rothenberg, and Stock (1996) and Ng and Perron (2001). These tests involve a step of detrending before constructing the test statistics and are demonstrated to perform better than the traditional ADF and PP tests.

Most testing procedures specify the unit root processes as the null hypothesis. Tests of the null hypothesis of stationarity have also been studied, among which Kwiatkowski et al. (1992) is very popular.

Economic theories often dictate that a group of economic time series are linked together by some long-run equilibrium relationship. Statistically, this phenomenon can be modeled by *cointegration*. When several

nonstationary processes $\mathbf{z}_t = (z_{1t}, \ldots, z_{kt})'$ are cointegrated, there exists a $(k \times 1)$ cointegrating vector $\mathbf{c}$ such that $\mathbf{c}'\mathbf{z}_t$ is stationary and $\mathbf{c}$ is a nonzero vector. One way to test the relationship of cointegration is the *residual based cointegration test*, which assumes the regression model

$$y_t = \beta_1 + \mathbf{x}_t'\beta + u_t$$

where $y_t = z_{1t}$, $\mathbf{x}_t = (z_{2t}, \ldots, z_{kt})'$, and $\beta = (\beta_2, \ldots, \beta_k)'$. The OLS residuals from the regression model are used to test for the null hypothesis of no cointegration. Engle and Granger (1987) suggest using ADF on the residuals while Phillips and Ouliaris (1990) study the tests using PP and other related test statistics.

Augmented Dickey-Fuller Unit Root and Engle-Granger Cointegration Testing

Common unit root tests have the null hypothesis that there is an autoregressive unit root $H_0 : \alpha = 1$, and the alternative is $H_a : |\alpha| < 1$, where α is the autoregressive coefficient of the time series

$$y_t = \alpha y_{t-1} + \epsilon_t$$

This is referred to as the zero mean model. The standard Dickey-Fuller (DF) test assumes that errors ϵ_t are white noise. There are two other types of regression models that include a constant or a time trend as follows:

$$y_t = \mu + \alpha y_{t-1} + \epsilon_t$$
$$y_t = \mu + \beta t + \alpha y_{t-1} + \epsilon_t$$

These two models are referred to as the constant mean model and the trend model, respectively. The constant mean model includes a constant mean μ of the time series. However, the interpretation of μ depends on the stationarity in the following sense: the mean in the stationary case when $\alpha < 1$ is the trend in the integrated case when $\alpha = 1$. Therefore, the null hypothesis should be the joint hypothesis that $\alpha = 1$ and $\mu = 0$. However, for the unit root tests, the test statistics are concerned with the null hypothesis of $\alpha = 1$. The joint null hypothesis is not commonly used. This issue is addressed in Bhargava (1986) with a different nesting model.

There are two types of test statistics. The conventional t ratio is

$$DF_\tau = \frac{\hat{\alpha} - 1}{sd(\hat{\alpha})}$$

and the second test statistic, called ρ-test, is

$$T(\hat{\alpha} - 1)$$

For the zero mean model, the asymptotic distributions of the Dickey-Fuller test statistics are

$$T(\hat{\alpha} - 1) \Rightarrow \left(\int_0^1 W(r)dW(r)\right)\left(\int_0^1 W(r)^2 dr\right)^{-1}$$
$$DF_\tau \Rightarrow \left(\int_0^1 W(r)dW(r)\right)\left(\int_0^1 W(r)^2 dr\right)^{-1/2}$$

For the constant mean model, the asymptotic distributions are

$$T(\hat{\alpha} - 1) \Rightarrow \left([W(1)^2 - 1]/2 - W(1)\int_0^1 W(r)dr\right)\left(\int_0^1 W(r)^2 dr - \left(\int_0^1 W(r)dr\right)^2\right)^{-1}$$
$$DF_\tau \Rightarrow \left([W(1)^2 - 1]/2 - W(1)\int_0^1 W(r)dr\right)\left(\int_0^1 W(r)^2 dr - \left(\int_0^1 W(r)dr\right)^2\right)^{-1/2}$$

For the trend model, the asymptotic distributions are

$$T(\hat{\alpha}-1) \Rightarrow \left[W(r)dW + 12\left(\int_0^1 rW(r)dr - \frac{1}{2}\int_0^1 W(r)dr\right)\left(\int_0^1 W(r)dr - \frac{1}{2}W(1)\right) - W(1)\int_0^1 W(r)dr \right] D^{-1}$$

$$DF_\tau \Rightarrow \left[W(r)dW + 12\left(\int_0^1 rW(r)dr - \frac{1}{2}\int_0^1 W(r)dr\right)\left(\int_0^1 W(r)dr - \frac{1}{2}W(1)\right) - W(1)\int_0^1 W(r)dr \right] D^{1/2}$$

where

$$D = \int_0^1 W(r)^2 dr - 12\left(\int_0^1 r(W(r)dr\right)^2 + 12\int_0^1 W(r)dr\int_0^1 rW(r)dr - 4\left(\int_0^1 W(r)dr\right)^2$$

One problem of the Dickey-Fuller and similar tests that employ three types of regressions is the difficulty in the specification of the deterministic trends. Campbell and Perron (1991) claimed that "the proper handling of deterministic trends is a vital prerequisite for dealing with unit roots." However, the "proper handling" is not obvious since the distribution theory of the relevant statistics about the deterministic trends is not available. Hayashi (2000) suggests using the constant mean model when you think there is no trend, and using the trend model when you think otherwise. However, no formal procedure is provided.

The null hypothesis of the Dickey-Fuller test is a random walk, possibly with drift. The differenced process is not serially correlated under the null of I(1). There is a great need for the generalization of this specification. The augmented Dickey-Fuller (ADF) test, originally proposed in Dickey and Fuller (1979), adjusts for the serial correlation in the time series by adding lagged first differences to the autoregressive model,

$$\Delta y_t = \mu + \delta t + \alpha y_{t-1} + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + \epsilon_t$$

where the deterministic terms δt and μ can be absent for the models without drift or linear trend. As previously, there are two types of test statistics. One is the OLS t value

$$\frac{\hat{\alpha}}{sd(\hat{\alpha})}$$

and the other is given by

$$\frac{T\hat{\alpha}}{1-\hat{\alpha}_1-\cdots-\hat{\alpha}_p}$$

The asymptotic distributions of the test statistics are the same as those of the standard Dickey-Fuller test statistics.

Nonstationary multivariate time series can be tested for cointegration, which means that a linear combination of these time series is stationary. Formally, denote the series by $\mathbf{z}_t = (z_{1t}, \ldots, z_{kt})'$. The null hypothesis of cointegration is that there exists a vector $\mathbf{c}$ such that $\mathbf{c}'\mathbf{z}_t$ is stationary. Residual-based cointegration tests were studied in Engle and Granger (1987) and Phillips and Ouliaris (1990). The latter are described in the next subsection. The first step regression is

$$y_t = \mathbf{x}_t'\beta + u_t$$

where $y_t = z_{1t}$, $\mathbf{x}_t = (z_{2t}, \ldots, z_{kt})'$, and $\beta = (\beta_2, \ldots, \beta_k)'$. This regression can also include an intercept or an intercept with a linear trend. The residuals are used to test for the existence of an autoregressive unit root. Engle and Granger (1987) proposed augmented Dickey-Fuller type regression without an intercept on the residuals to test the unit root. When the first step OLS does not include an intercept, the asymptotic distribution of the ADF test statistic DF_τ is given by

$$
\begin{aligned}
DF_\tau &\Longrightarrow \int_0^1 \frac{Q(r)}{(\int_0^1 Q^2)^{1/2}} dS \\
Q(r) &= W_1(r) - \int_0^1 W_1 W_2' \left(\int_0^1 W_2 W_2' \right)^{-1} W_2(r) \\
S(r) &= \frac{Q(r)}{(\kappa'\kappa)^{1/2}} \\
\kappa' &= \left(1, -\int_0^1 W_1 W_2' \left(\int_0^1 W_2 W_2' \right)^{-1} \right)
\end{aligned}
$$

where $W(r)$ is a k vector standard Brownian motion and

$$W(r) = \Big(W_1(r), W_2(r) \Big)$$

is a partition such that $W_1(r)$ is a scalar and $W_2(r)$ is $k-1$ dimensional. The asymptotic distributions of the test statistics in the other two cases have the same form as the preceding formula. If the first step regression includes an intercept, then $W(r)$ is replaced by the demeaned Brownian motion $\overline{W}(r) = W(r) - \int_0^1 W(r)dr$. If the first step regression includes a time trend, then $W(r)$ is replaced by the detrended Brownian motion. The critical values of the asymptotic distributions are tabulated in Phillips and Ouliaris (1990) and MacKinnon (1991).

The residual based cointegration tests have a major shortcoming. Different choices of the dependent variable in the first step OLS might produce contradictory results. This can be explained theoretically. If the dependent variable is in the cointegration relationship, then the test is consistent against the alternative that there is cointegration. On the other hand, if the dependent variable is not in the cointegration system, the OLS residual $y_t - \mathbf{x}_t'\beta$ do not converge to a stationary process. Changing the dependent variable is more likely to produce conflicting results in finite samples.

Phillips-Perron Unit Root and Cointegration Testing

Besides the ADF test, there is another popular unit root test that is valid under general serial correlation and heteroscedasticity, developed by Phillips (1987) and Phillips and Perron (1988). The tests are constructed using the AR(1) type regressions, unlike ADF tests, with corrected estimation of the long run variance of Δy_t. In the case without intercept, consider the driftless 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,

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

Denote the OLS residual by $\hat{u}_t$. The asymptotic variance of $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2$ can be estimated by 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 $\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}$. This is a consistent estimator suggested by Newey and West (1987).

The variance of u_t can be estimated by $s^2 = \frac{1}{T-k}\sum_{t=1}^{T}\hat{u}_t^2$. Let $\hat{\sigma}^2$ be the variance estimate of the OLS estimator $\hat{\rho}$. Then the Phillips-Perron $\hat{Z}_\rho$ test (zero mean case) is written

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

The $\hat{Z}_\rho$ statistic is just the ordinary Dickey-Fuller $\hat{Z}_\alpha$ statistic with a correction term that accounts for the serial correlation. The correction term goes to zero asymptotically if there is no serial correlation.

Note that $P(\hat{\rho} < 1) \approx 0.68$ as $T \to \infty$, which shows that the limiting distribution is skewed to the left.

Let τ_ρ be the τ statistic for $\hat{\rho}$. The Phillips-Perron $\hat{Z}_t$ (defined here as $\hat{Z}_\tau$) test is written

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

To incorporate a constant intercept, the regression model $y_t = \mu + \rho y_{t-1} + u_t$ is used (single mean case) and null hypothesis the series is a driftless random walk with nonzero unconditional mean. To incorporate a time trend, we used the regression model $y_t = \mu + \delta t + \rho y_{t-1} + u_t$ and under the null the series is a random walk with drift.

The limiting distributions of the test statistics for the zero mean case are

$$\hat{Z}_\rho \Rightarrow \frac{\frac{1}{2}\{B(1)^2 - 1\}}{\int_0^1 [B(s)]^2 ds}$$

$$\hat{Z}_\tau \Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\}}{\{\int_0^1 [B(x)]^2 dx\}^{1/2}}$$

where $B(\cdot)$ is a standard Brownian motion.

The limiting distributions of the test statistics for the intercept case are

$$\hat{Z}_\rho \Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\} - B(1)\int_0^1 B(x)dx}{\int_0^1 [B(x)]^2 dx - \left[\int_0^1 B(x)dx\right]^2}$$

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

Finally, the limiting distributions of the test statistics for the trend case are can be derived as

$$\begin{bmatrix} 0 & c & 0 \end{bmatrix} V^{-1} \begin{bmatrix} B(1) \\ \left(B(1)^2 - 1\right)/2 \\ B(1) - \int_0^1 B(x)dx \end{bmatrix}$$

where $c = 1$ for $\hat{Z}_\rho$ and $c = \frac{1}{\sqrt{Q}}$ for $\hat{Z}_\tau$,

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

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

The finite sample performance of the PP test is not satisfactory (see Hayashi 2000).

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

$$y_t = \beta_1 + \mathbf{x}_t'\beta + u_t$$

where $y_t = z_{1t}$, $\mathbf{x}_t = (z_{2t}, \ldots, z_{kt})'$, and $\beta = (\beta_2, \ldots, \beta_k)'$. You can estimate the consistent cointegrating vector by using OLS if all variables are difference stationary—that is, I(1). The estimated cointegrating vector is $\hat{\mathbf{c}} = (1, -\hat{\beta}_2, \ldots, -\hat{\beta}_k)'$. The Phillips-Ouliaris test is computed using the OLS residuals from the preceding regression model, and it uses the PP unit root tests $\hat{Z}_\rho$ and $\hat{Z}_\tau$ developed in Phillips (1987), although in Phillips and Ouliaris (1990) the asymptotic distributions of some other leading unit root tests are also derived. The null hypothesis is no cointegration.

You need to refer to the tables by Phillips and Ouliaris (1990) to obtain the p-value of the cointegration test. Before you apply the cointegration test, you might want to perform the unit root test for each variable (see the option STATIONARITY=).

As in the Engle-Granger cointegration tests, the Phillips-Ouliaris test can give conflicting results for different choices of the regressand. There are other cointegration tests that are invariant to the order of the variables, including Johansen (1988), Johansen (1991), Stock and Watson (1988).

ERS and Ng-Perron Unit Root Tests

As mentioned earlier, ADF and PP both suffer severe size distortion and low power. There is a class of newer tests that improve both size and power. These are sometimes called efficient unit root tests, and among them tests by Elliott, Rothenberg, and Stock (1996) and Ng and Perron (2001) are prominent.

Elliott, Rothenberg, and Stock (1996) consider the data generating process

$$y_t = \beta' z_t + u_t$$
$$u_t = \alpha u_{t-1} + v_t, t = 1, \ldots, T$$

where $\{z_t\}$ is either $\{1\}$ or $\{(1, t)\}$ and $\{v_t\}$ is an unobserved stationary zero-mean process with positive spectral density at zero frequency. The null hypothesis is $H_0 : \alpha = 1$, and the alternative is $H_a : |\alpha| < 1$. The key idea of Elliott, Rothenberg, and Stock (1996) is to study the asymptotic power and asymptotic power envelope of some new tests. Asymptotic power is defined with a sequence of local alternatives. For a fixed alternative hypothesis, the power of a test usually goes to one when sample size goes to infinity; however, this says nothing about the finite sample performance. On the other hand, when the data generating process under the alternative moves closer to the null hypothesis as the sample size increases, the power does not necessarily converge to one. The local-to-unity alternatives in ERS are

$$\alpha = 1 + \frac{c}{T}$$

and the power against the local alternatives has a limit as T goes to infinity, which is called asymptotic power. This value is strictly between 0 and 1. Asymptotic power indicates the adequacy of a test to distinguish small deviations from the null hypothesis.

Define

$$y_\alpha = (y_1, (1-\alpha L)y_2, \ldots, (1-\alpha L)y_T)$$
$$z_\alpha = (z_1, (1-\alpha L)z_2, \ldots, (1-\alpha L)z_T)$$

Let $S(\alpha)$ be the sum of squared residuals from a least squares regression of y_α on z_α. Then the *point optimal test* against the local alternative $\bar{\alpha} = 1 + \bar{c}/T$ has the form

$$P_T^{GLS} = \frac{S(\bar{\alpha}) - \bar{\alpha}S(1)}{\hat{\omega}^2}$$

where $\hat{\omega}^2$ is an estimator for $\omega^2 = \sum_{k=-\infty}^{\infty} E v_t v_{t-k}$. The autoregressive (AR) estimator is used for $\hat{\omega}^2$ (Elliott, Rothenberg, and Stock 1996, equations 13 and 14),

$$\hat{\omega}^2 = \frac{\hat{\sigma}_\eta^2}{(1 - \sum_{i=1}^{p} \hat{a}_i)^2}$$

where $\hat{\sigma}_\eta^2$ and $\hat{a}_i$ are OLS estimates from the regression

$$\Delta y_t = a_0 y_{t-1} + \sum_{i=1}^{p} a_i \Delta y_{t-i} + a_{p+1} + \eta_t$$

where p is selected according to the Schwarz Bayesian information criterion. The test rejects the null when P_T is small. The asymptotic power function for the point optimal test that is constructed with $\bar{c}$ under local alternatives with c is denoted by $\pi(c, \bar{c})$. Then the power envelope is $\pi(c, c)$ because the test formed with $\bar{c}$ is the most powerful against the alternative $c = \bar{c}$. In other words, the asymptotic function $\pi(c, \bar{c})$ is always below the power envelope $\pi(c)$ except that at one point, $c = \bar{c}$, they are tangent. Elliott, Rothenberg, and Stock (1996) show that choosing some specific values for $\bar{c}$ can cause the asymptotic power function $\pi(c, \bar{c})$ of the point optimal test to be very close to the power envelope. The optimal $\bar{c}$ is -7 when $z_t = 1$, and -13.5 when $z_t = (1, t)'$. This choice of $\bar{c}$ corresponds to the tangent point where $\pi = 0.5$. This is also true of the DF-GLS test.

Elliott, Rothenberg, and Stock (1996) also propose the *DF-GLS test*, given by the t statistic for testing $\psi_0 = 0$ in the regression

$$\Delta y_t^d = \psi_0 y_{t-1}^d + \sum_{j=1}^{p} \psi_j \Delta y_{t-j}^d + \epsilon_{tp}$$

where y_t^d is obtained in a first step detrending

$$y_t^d = y_t - \hat{\beta}_{\bar{\alpha}}' z_t$$

and $\hat{\beta}_{\bar{\alpha}}$ is least squares regression coefficient of y_α on z_α. Regarding the lag length selection, Elliott, Rothenberg, and Stock (1996) favor the Schwarz Bayesian information criterion. The optimal selection of the lag length p and the estimation of ω^2 is further discussed in Ng and Perron (2001). The lag length is selected from the interval $[0, p_{max}]$ for some fixed p_{max} by using the modified Akaike's information criterion,

$$\mathrm{MAIC}(p) = \log(\hat{\sigma}_p^2) + \frac{2(\tau_T(p) + p)}{T - p_{max}}$$

where $\tau_T(p) = (\hat{\sigma}_p^2)^{-1}\hat{\psi}_0^2 \sum_{t=p_{max}+1}^{T-1}(y_t^d)^2$ and $\hat{\sigma}_p^2 = (T - p_{max} - 1)^{-1}\sum_{t=p_{max}+1}^{T-1}\hat{\epsilon}_{tp}^2$. For fixed lag length p, an estimate of ω^2 is given by

$$\hat{\omega}^2 = \frac{(T-1-p)^{-1}\sum_{t=p+2}^{T}\hat{\epsilon}_{tp}^2}{\left(1-\sum_{j=1}^{p}\hat{\psi}_j\right)^2}$$

DF-GLS is indeed a superior unit root test, according to Stock (1994), Schwert (1989), and Elliott, Rothenberg, and Stock (1996). In terms of the size of the test, DF-GLS is almost as good as the ADF t test DF_τ and better than the PP $\hat{\mathrm{Z}}_\rho$ and $\hat{\mathrm{Z}}_\tau$ test. In addition, the power of the DF-GLS test is greater than that of both the ADF t test and the ρ-test.

Ng and Perron (2001) also apply GLS detrending to obtain the following M-tests:

$$MZ_\alpha = ((T-1)^{-1}(y_T^d)^2 - \hat{\omega}^2)\left(2(T-1)^{-2}\sum_{t=1}^{T-1}(y_t^d)^2\right)^{-1}$$

$$MSB = \left(\frac{\sum_{t=1}^{T-1}(y_t^d)^2}{(T-1)^2\hat{\omega}^2}\right)^{1/2}$$

$$MZ_t = MZ_\alpha \times MSB$$

The first one is a modified version of the Phillips-Perron Z_ρ test,

$$MZ_\rho = \mathrm{Z}_\rho + \frac{T}{2}(\hat{\alpha}-1)^2$$

where the detrended data $\{y_t^d\}$ is used. The second is a modified Bhargava (1986) R_1 test statistic. The third can be perceived as a modified Phillips-Perron Z_τ statistic because of the relationship $\mathrm{Z}_\tau = MSB \times \mathrm{Z}_\rho$.

The modified point optimal tests that use the GLS detrended data are

$$MP_T^{GLS} = \frac{\bar{c}^2(T-1)^{-2}\sum_{t=1}^{T-1}(y_t^d)^2 - \bar{c}(T-1)^{-1}(y_T^d)^2}{\hat{\omega}^2} \quad \text{for } z_t = 1$$

$$MP_T^{GLS} = \frac{\bar{c}^2(T-1)^{-2}\sum_{t=1}^{T-1}(y_t^d)^2 + (1-\bar{c})(T-1)^{-1}(y_T^d)^2}{\hat{\omega}^2} \quad \text{for } z_t = (1,t)$$

The DF-GLS test and the MZ_t test have the same limiting distribution:

$$\text{DF-GLS} \approx MZ_t \Rightarrow 0.5\frac{(J_c(1)^2-1)}{\left(\int_0^1 J_c(r)^2dr\right)^{1/2}} \quad \text{for } z_t = 1$$

$$\text{DF-GLS} \approx MZ_t \Rightarrow 0.5\frac{(V_{c,\bar{c}}(1)^2-1)}{\left(\int_0^1 V_{c,\bar{c}}(r)^2dr\right)^{1/2}} \quad \text{for } z_t = (1,t)$$

The point optimal test and the modified point optimal test have the same limiting distribution,

$$P_T^{GLS} \approx MP_T^{GLS} \Rightarrow \bar{c}^2\int_0^1 J_c(r)^2dr - \bar{c}J_c(1)^2 \quad \text{for } z_t = 1$$

$$P_T^{GLS} \approx MP_T^{GLS} \Rightarrow \bar{c}^2\int_0^1 V_{c,\bar{c}}(r)^2dr + (1-\bar{c})V_{c,\bar{c}}(1)^2 \quad \text{for } z_t = (1,t)$$

where $W(r)$ is a standard Brownian motion and $J_c(r)$ is an Ornstein-Uhlenbeck process defined by $dJ_c(r) = cJ_c(r)dr + dW(r)$ with $J_c(0) = 0$, $V_{c,\bar{c}}(r) = J_c(r) - r\left[\lambda J_c(1) + 3(1-\lambda)\int_0^1 sJ_c(s)ds\right]$, and $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$.

Overall, the M-tests have the smallest size distortion, with the ADF t test having the next smallest. The ADF ρ-test, $\hat{\mathrm{Z}}_\rho$, and $\hat{\mathrm{Z}}_\tau$ have the largest size distortion. In addition, the power of the DF-GLS and M-tests is greater than that of the ADF t test and ρ-test. The ADF $\hat{\mathrm{Z}}_\rho$ has more severe size distortion than the ADF $\hat{\mathrm{Z}}_\tau$, but it has more power for a fixed lag length.

Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Unit Root Test and Shin Cointegration Test

There are fewer tests available for the null hypothesis of trend stationarity I(0). The main reason is the difficulty of theoretical development. The KPSS test was introduced in Kwiatkowski et al. (1992) to test the null hypothesis that an observable series is stationary around a deterministic trend. For consistency, the notation used here differs from the notation in the original paper. The setup of the problem is as follows: it is assumed that the series is expressed as the sum of the deterministic trend, random walk r_t, and stationary error u_t; that is,

$$y_t = \mu + \delta t + r_t + u_t$$
$$r_t = r_{t-1} + e_t$$

where $e_t \sim$iid $(0, \sigma_e^2)$, and an intercept μ (in the original paper, the authors use r_0 instead of μ; here we assume $r_0 = 0$.) The null hypothesis of trend stationarity is specified by $H_0 : \sigma_e^2 = 0$, while the null of level stationarity is the same as above with the model restriction $\delta = 0$. Under the alternative that $\sigma_e^2 \neq 0$, there is a random walk component in the observed series y_t.

Under stronger assumptions of normality and iid of u_t and e_t, a one-sided LM test of the null that there is no random walk ($e_t = 0, \forall t$) can be constructed as follows:

$$\widehat{LM} = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{s^2(l)}$$
$$s^2(l) = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 + \frac{2}{T} \sum_{s=1}^{l} w(s,l) \sum_{t=s+1}^{T} \hat{u}_t \hat{u}_{t-s}$$
$$S_t = \sum_{\tau=1}^{t} \hat{u}_\tau$$

Under the null hypothesis, $\hat{u}_t$ can be estimated by ordinary least squares regression of y_t on an intercept and the time trend. Following the original work of Kwiatkowski et al. (1992), under the null ($\sigma_e^2 = 0$), the $\widehat{LM}$ statistic converges asymptotically to three different distributions depending on whether the model is trend-stationary, level-stationary ($\delta = 0$), or zero-mean stationary ($\delta = 0$, $\mu = 0$). The trend-stationary model is denoted by subscript τ and the level-stationary model is denoted by subscript μ. The case when there is no trend and zero intercept is denoted as 0. The last case, although rarely used in practice, is considered in Hobijn, Franses, and Ooms (2004),

$$y_t = u_t : \quad \widehat{LM}_0 \xrightarrow{D} \int_0^1 B^2(r)dr$$
$$y_t = \mu + u_t : \quad \widehat{LM}_\mu \xrightarrow{D} \int_0^1 V^2(r)dr$$
$$y_t = \mu + \delta t + u_t : \quad \widehat{LM}_\tau \xrightarrow{D} \int_0^1 V_2^2(r)dr$$

with

$$V(r) = B(r) - rB(1)$$
$$V_2(r) = B(r) + (2r - 3r^2)B(1) + (-6r + 6r^2) \int_0^1 B(s)ds$$

where $B(r)$ is a Brownian motion (Wiener process) and $\xrightarrow{D}$ is convergence in distribution. $V(r)$ is a standard Brownian bridge, and $V_2(r)$ is a second-level Brownian bridge.

Using the notation of Kwiatkowski et al. (1992), the $\widehat{LM}$ statistic is named as $\hat{\eta}$. This test depends on the computational method used to compute the long-run variance $s(l)$; that is, the window width l and the kernel type $w(\cdot,\cdot)$. You can specify the kernel used in the test by using the KERNEL option:

- Newey-West/Bartlett (KERNEL=NW | BART) (this is the default)

$$w(s,l) = 1 - \frac{s}{l+1}$$

- quadratic spectral (KERNEL=QS)

$$w(s,l) = \tilde{w}\left(\frac{s}{l}\right) = \tilde{w}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos\left(\frac{6}{5}\pi x\right)\right)$$

You can specify the number of lags, l, in three different ways:

- Schwert (SCHW = c) (default for NW, c=12)

$$l = \max\left\{1, \text{floor}\left[c\left(\frac{T}{100}\right)^{1/4}\right]\right\}$$

- manual (LAG = l)
- automatic selection (AUTO) (default for QS), from Hobijn, Franses, and Ooms (2004). The number of lags, l, is calculated as in the following table:

KERNEL=NW	**KERNEL=QS**
$l = \min(T, \text{floor}(\hat{\gamma} T^{1/3}))$	$l = \min(T, \text{floor}(\hat{\gamma} T^{1/5}))$
$\hat{\gamma} = 1.1447\left\{\left(\frac{\hat{s}^{(1)}}{\hat{s}^{(0)}}\right)^2\right\}^{1/3}$	$\hat{\gamma} = 1.3221\left\{\left(\frac{\hat{s}^{(2)}}{\hat{s}^{(0)}}\right)^2\right\}^{1/5}$
$\hat{s}^{(j)} = \delta_{0,j}\hat{\gamma}_0 + 2\sum_{i=1}^{n} i^j \hat{\gamma}_i$	$\hat{s}^{(j)} = \delta_{0,j}\hat{\gamma}_0 + 2\sum_{i=1}^{n} i^j \hat{\gamma}_i$
$n = \text{floor}(T^{2/9})$	$n = \text{floor}(T^{2/25})$

where T is the number of observations, $\delta_{0,j} = 1$ if $j = 0$ and 0 otherwise, and $\hat{\gamma}_i = \frac{1}{T}\sum_{t=1}^{T-i} u_t u_{t+i}$.

Simulation evidence shows that the KPSS has size distortion in finite samples. For an example, see Caner and Kilian (2001). The power is reduced when the sample size is large; this can be derived theoretically (see Breitung 1995). Another problem of the KPSS test is that the power depends on the truncation lag used in the Newey-West estimator of the long-run variance $s^2(l)$.

Shin (1994) extends the KPSS test to incorporate the regressors to be a cointegration test. The cointegrating regression becomes

$$y_t = \mu + \delta t + X_t'\beta + r_t + u_t$$
$$r_t = r_{t-1} + e_t$$

where y_t and X_t are scalar and m-vector $I(1)$ variables. There are still three cases of cointegrating regressions: without intercept and trend, with intercept only, and with intercept and trend. The null hypothesis of the cointegration test is the same as that for the KPSS test, $H_0 : \sigma_e^2 = 0$. The test statistics for cointegration in the three cases of cointegrating regressions are exactly the same as those in the KPSS test; these test statistics are then ignored here. Under the null hypothesis, the statistics converge asymptotically to three different distributions,

$$y_t = X_t'\beta + u_t : \qquad \widehat{LM}_0 \xrightarrow{D} \int_0^1 Q_1^2(r)dr$$

$$y_t = \mu + X_t'\beta + u_t : \qquad \widehat{LM}_\mu \xrightarrow{D} \int_0^1 Q_2^2(r)dr$$

$$y_t = \mu + \delta t + X_t'\beta + u_t : \qquad \widehat{LM}_\tau \xrightarrow{D} \int_0^1 Q_3^2(r)dr$$

with

$$Q_1(r) = B(r) - \left(\int_0^r \mathbf{B_m}(x)dx\right)\left(\int_0^1 \mathbf{B_m}(x)\mathbf{B_m}'(x)dx\right)^{-1}\left(\int_0^1 \mathbf{B_m}(x)dB(x)\right)$$

$$Q_2(r) = V(r) - \left(\int_0^r \bar{\mathbf{B}}_\mathbf{m}(x)dx\right)\left(\int_0^1 \bar{\mathbf{B}}_\mathbf{m}(x)\bar{\mathbf{B}}_\mathbf{m}'(x)dx\right)^{-1}\left(\int_0^1 \bar{\mathbf{B}}_\mathbf{m}(x)dB(x)\right)$$

$$Q_3(r) = V_2(r) - \left(\int_0^r \mathbf{B}^*_\mathbf{m}(x)dx\right)\left(\int_0^1 \mathbf{B}^*_\mathbf{m}(x)\mathbf{B}^{*\prime}_\mathbf{m}(x)dx\right)^{-1}\left(\int_0^1 \mathbf{B}^*_\mathbf{m}(x)dB(x)\right)$$

where $B(.)$ and $\mathbf{B_m}(.)$ are independent scalar and m-vector standard Brownian motion, and $\xrightarrow{D}$ is convergence in distribution. $V(r)$ is a standard Brownian bridge, $V_2(r)$ is a Brownian bridge of a second-level, $\bar{\mathbf{B}}_\mathbf{m}(r) = \mathbf{B_m}(r) - \int_0^1 \mathbf{B_m}(x)dx$ is an m-vector standard demeaned Brownian motion, and $\mathbf{B}^*_\mathbf{m}(r) = \mathbf{B_m}(r) + (6r - 4)\int_0^1 \mathbf{B_m}(x)dx + (-12r + 6)\int_0^1 x\mathbf{B_m}(x)dx$ is an m-vector standard demeaned and detrended Brownian motion.

The p-values that are reported for the KPSS test and Shin test are calculated via a Monte Carlo simulation of the limiting distributions, using a sample size of 2,000 and 1,000,000 replications.

Testing for Statistical Independence

Independence tests are widely used in model selection, residual analysis, and model diagnostics because models are usually based on the assumption of independently distributed errors. If a given time series (for example, a series of residuals) is independent, then no deterministic model is necessary for this completely random process; otherwise, there must exist some relationship in the series to be addressed. In the following section, four independence tests are introduced: the BDS test, the runs test, the turning point test, and the rank version of von Neumann ratio test.

BDS Test

Brock, Dechert, and Scheinkman (1987) propose a test (BDS test) of independence based on the correlation dimension. Brock et al. (1996) show that the first-order asymptotic distribution of the test statistic is independent of the estimation error provided that the parameters of the model under test can be estimated $\sqrt{n}$-consistently. Hence, the BDS test can be used as a model selection tool and as a specification test.

Given the sample size T, the embedding dimension m, and the value of the radius r, the BDS statistic is

$$S_{\text{BDS}}(T,m,r) = \sqrt{T-m+1}\frac{c_{m,m,T}(r) - c_{1,m,T}^{m}(r)}{\sigma_{m,T}(r)}$$

where

$$c_{m,n,N}(r) = \frac{2}{(N-n+1)(N-n)}\sum_{s=n}^{N}\sum_{t=s+1}^{N}\prod_{j=0}^{m-1} I_r(z_{s-j}, z_{t-j})$$

$$I_r(z_s, z_t) = \begin{cases} 1 & \text{if } |z_s - z_t| < r \\ 0 & \text{otherwise} \end{cases}$$

$$\sigma_{m,T}^2(r) = 4\left(k^m + 2\sum_{j=1}^{m-1} k^{m-j}c^{2j} + (m-1)^2 c^{2m} - m^2 k c^{2m-2}\right)$$

$$c = c_{1,1,T}(r)$$

$$k = k_T(r) = \frac{6}{T(T-1)(T-2)}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\sum_{l=s+1}^{T} h_r(z_t, z_s, z_l)$$

$$h_r(z_t, z_s, z_l) = \frac{1}{3}\left(I_r(z_t, z_s)I_r(z_s, z_l) + I_r(z_t, z_l)I_r(z_l, z_s) + I_r(z_s, z_t)I_r(z_t, z_l)\right)$$

The statistic has a standard normal distribution if the sample size is large enough. For small sample size, the distribution can be approximately obtained through simulation. Kanzler (1999) has a comprehensive discussion on the implementation and empirical performance of BDS test.

Runs Test and Turning Point Test

The runs test and turning point test are two widely used tests for independence (Cromwell, Labys, and Terraza 1994).

The runs test needs several steps. First, convert the original time series into the sequence of signs, $\{+ + - - \cdots + - - -\}$, that is, map $\{z_t\}$ into $\{sign(z_t - z_M)\}$ where z_M is the sample mean of z_t and $sign(x)$ is "+" if x is nonnegative and "−" if x is negative. Second, count the number of runs, R, in the sequence. A run of a sequence is a maximal non-empty segment of the sequence that consists of adjacent equal elements. For example, the following sequence contains $R = 8$ runs:

$$\underbrace{+++}_{1}\underbrace{---}_{1}\underbrace{++}_{1}\underbrace{--}_{1}\underbrace{+}_{1}\underbrace{-}_{1}\underbrace{+++++}_{1}\underbrace{--}_{1}$$

Third, count the number of pluses and minuses in the sequence and denote them as N_+ and N_-, respectively. In the preceding example sequence, $N_+ = 11$ and $N_- = 8$. Note that the sample size $T = N_+ + N_-$. Finally, compute the statistic of runs test,

$$S_{\text{runs}} = \frac{R - \mu}{\sigma}$$

where

$$\mu = \frac{2N_+N_-}{T} + 1$$

$$\sigma^2 = \frac{(\mu - 1)(\mu - 2)}{T - 1}$$

The statistic of the turning point test is defined as

$$S_{\mathrm{TP}} = \frac{\sum_{t=2}^{T-1} TP_t - 2(T-2)/3}{\sqrt{(16T - 29)/90}}$$

where the indicator function of the turning point TP_t is 1 if $z_t > z_{t\pm1}$ or $z_t < z_{t\pm1}$ (that is, both the previous and next values are greater or less than the current value); otherwise, 0.

The statistics of both the runs test and the turning point test have the standard normal distribution under the null hypothesis of independence.

Rank Version of the von Neumann Ratio Test

Because the runs test completely ignores the magnitudes of the observations, Bartels (1982) proposes a rank version of the von Neumann ratio test for independence,

$$S_{\mathrm{RVN}} = \frac{\sqrt{T}}{2}\left(\frac{\sum_{t=1}^{T-1}(R_{t+1} - R_t)^2}{(T(T^2-1)/12)} - 2\right)$$

where R_t is the rank of tth observation in the sequence of T observations. For large samples, the statistic follows the standard normal distribution under the null hypothesis of independence. For small samples of size between 11 and 100, the critical values that have been simulated would be more precise. For samples of size less than or equal to 10, the exact CDF of the statistic is available. Hence, the VNRRANK=(PVALUE=SIM) option is recommended for small samples whose size is no more than 100, although it might take longer to obtain the p-value than if you use the VNRRANK=(PVALUE=DIST) option.

Testing for Normality

Based on skewness and kurtosis, Jarque and Bera (1980) calculated the test statistic

$$T_N = \left[\frac{N}{6}b_1^2 + \frac{N}{24}(b_2 - 3)^2\right]$$

where

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

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

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 $\hat{u}_t = \hat{\epsilon}_t/\sqrt{h_t}$. The normality test can be used to detect misspecification of the family of ARCH models.

Testing for Linear Dependence

Generalized Durbin-Watson Tests

Consider the linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\nu}$$

where $\mathbf{X}$ is an $N \times k$ data matrix, β is a $k \times 1$ coefficient vector, and $\boldsymbol{\nu}$ is an $N \times 1$ disturbance vector. The error term $\boldsymbol{\nu}$ is assumed to be generated by the jth-order autoregressive process $\nu_t = \epsilon_t - \varphi_j \nu_{t-j}$ where $|\varphi_j| < 1$, ϵ_t is a sequence of independent normal error terms with mean 0 and variance σ^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 $\hat{\boldsymbol{\nu}}$ are OLS residuals. Using the matrix notation,

$$d_j = \frac{\mathbf{Y}'\mathbf{M}\mathbf{A}_j'\mathbf{A}_j\mathbf{M}\mathbf{Y}}{\mathbf{Y}'\mathbf{M}\mathbf{Y}}$$

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,

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

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 R is an $N \times k$ upper triangular matrix. There exists an $N \times (N - k)$ 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,

$$d_j = \frac{\sum_{l=1}^{n} \lambda_{jl} {\xi_l}^2}{\sum_{l=1}^{n} \xi_l^2}$$

where $\lambda_{j1} \ldots \lambda_{jn}$ are upper n eigenvalues of $\mathbf{M}\mathbf{A}'_j\mathbf{A}_j\mathbf{M}$ and ξ_l 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 Van Loan 1989; Savin and White 1978).

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

$$\text{Prob}\left(\frac{\sum_{l=1}^{n} \lambda_{jl}\xi_l^2}{\sum_{l=1}^{n} \xi_l^2} < c_0\right) = \text{Prob}(q_j < 0)$$

where

$$q_j = \sum_{l=1}^{n} (\lambda_{jl} - c_0)\xi_l^2$$

When the null hypothesis $H_0 : \varphi_j = 0$ holds, the quadratic form q_j has the characteristic function

$$\phi_j(t) = \prod_{l=1}^{n} (1 - 2(\lambda_{jl} - c_0)it)^{-1/2}$$

The distribution function is uniquely determined by this characteristic function:

$$F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{e^{itx}\phi_j(-t) - e^{-itx}\phi_j(t)}{it} dt$$

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,

$$F(0) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{(\phi_4(-t) - \phi_4(t))}{it} dt$$

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 (α), 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 5% 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 jth step, test $H_0 : \varphi_j = 0$ given $\varphi_1 = \cdots = \varphi_{j-1} = 0$ 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 the section "Autoregressive Error Model" on page 366. 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.

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 linear regression model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$$

$$u_t + \varphi_j u_{t-j} = \epsilon_t, \qquad t = 1, \ldots, N$$

where $\mathbf{X}$ is an $N \times k$ data matrix, β is a $k \times 1$ coefficient vector, $\mathbf{u}$ is an $N \times 1$ disturbance vector, and ϵ_t is a sequence of independent normal error terms with mean 0 and variance σ^2.

The generalized Durbin-Watson statistic is written as

$$\mathrm{DW}_j = \frac{\hat{\mathbf{u}}'\mathbf{A}_j'\mathbf{A}_j\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}}$$

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

$$\mathrm{DW}_j = \frac{\mathbf{Y}'\mathbf{M}\mathbf{A}_j'\mathbf{A}_j\mathbf{M}\mathbf{Y}}{\mathbf{Y}'\mathbf{M}\mathbf{Y}} = \frac{\eta'(\mathbf{Q}_1'\mathbf{A}_j'\mathbf{A}_j\mathbf{Q}_1)\eta}{\eta'\eta}$$

where $\mathbf{Q}_1'\mathbf{Q}_1 = \mathbf{I}_{T-k}$, $\mathbf{Q}_1'\mathbf{X} = 0$, and $\eta = \mathbf{Q}_1'\mathbf{u}$.

The marginal probability for the Durbin-Watson statistic is

$$\Pr(\mathrm{DW}_j < c) = \Pr(h < 0)$$

where $h = \eta'(\mathbf{Q}_1'\mathbf{A}_j'\mathbf{A}_j\mathbf{Q}_1 - c\mathbf{I})\eta$.

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 $h = \eta'(\mathbf{Q}_1'\mathbf{A}_j'\mathbf{A}_j\mathbf{Q}_1 - c\mathbf{I})\eta$. The trapezoidal rule approximation to the marginal probability $\Pr(h < 0)$ is

$$\Pr(h < 0) = \frac{1}{2} - \sum_{k=0}^{K} \frac{\mathrm{Im}\left[\phi((k+\frac{1}{2})\Delta)\right]}{\pi(k+\frac{1}{2})} + \mathrm{E}_I(\Delta) + \mathrm{E}_T(K)$$

where $\mathrm{Im}\left[\phi(\cdot)\right]$ is the imaginary part of the characteristic function, $\mathrm{E}_I(\Delta)$ and $\mathrm{E}_T(K)$ are integration and truncation errors, respectively. For numerical inversion of the characteristic function, see Davies (1973).

Ansley, Kohn, and Shively (1992) proposed a numerically efficient algorithm that requires O(N) operations for evaluation of the characteristic function $\phi(u)$. The characteristic function is denoted as

$$\begin{aligned}\phi(u) &= \left|\mathbf{I} - 2iu(\mathbf{Q}_1'\mathbf{A}_j'\mathbf{A}_j\mathbf{Q}_1 - c\mathbf{I}_{N-k})\right|^{-1/2} \\ &= |\mathbf{V}|^{-1/2}\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right|^{-1/2}\left|\mathbf{X}'\mathbf{X}\right|^{1/2}\end{aligned}$$

where $\mathbf{V} = (1+2iuc)\mathbf{I} - 2iu\mathbf{A}_j'\mathbf{A}_j$ and $i = \sqrt{-1}$. By applying the Cholesky decomposition to the complex matrix $\mathbf{V}$, you can obtain the lower triangular matrix $\mathbf{G}$ that satisfies $\mathbf{V} = \mathbf{G}\mathbf{G}'$. Therefore, the characteristic function can be evaluated in O(N) operations by using the formula

$$\phi(u) = |\mathbf{G}|^{-1}\left|\mathbf{X}^{*\prime}\mathbf{X}^*\right|^{-1/2}\left|\mathbf{X}'\mathbf{X}\right|^{1/2}$$

where $\mathbf{X}^* = \mathbf{G}^{-1}\mathbf{X}$. For more information about evaluation of the characteristic function, see Ansley, Kohn, and Shively (1992).

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) proposes two alternative statistics (Durbin h and t) that are asymptotically equivalent. The h statistic is written as

$$h = \hat{\rho}\sqrt{N/(1 - N\hat{V})}$$

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

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. For information about the Durbin-Watson test in the presence of lagged dependent variables, see Inder (1986) and King and Wu (1991).

Godfrey LM test

The GODFREY= option in the MODEL statement produces the Godfrey Lagrange multiplier test for serially correlated residuals for each equation (Godfrey 1978b, a). r is the maximum autoregressive order, and specifies that Godfrey's tests be computed for lags 1 through r. The default number of lags is four.

Testing for Nonlinear Dependence: Ramsey's Reset Test

Ramsey's reset test is a misspecification test associated with the functional form of models to check whether power transforms need to be added to a model. The original linear model, henceforth called the restricted model, is

$$y_t = \mathbf{x_t}\beta + u_t$$

To test for misspecification in the functional form, the unrestricted model is

$$y_t = \mathbf{x_t}\beta + \sum_{j=2}^{p} \phi_j \hat{y}_t^j + u_t$$

where $\hat{y}_t$ is the predicted value from the linear model and p is the power of $\hat{y}_t$ in the unrestricted model equation starting from 2. The number of higher-ordered terms to be chosen depends on the discretion of the analyst. The RESET option produces test results for $p = 2$, 3, and 4.

The reset test is an F statistic for testing $H_0 : \phi_j = 0$, for all $j = 2, \ldots, p$, against $H_1 : \phi_j \neq 0$ for at least one $j = 2, \ldots, p$ in the unrestricted model and is computed as

$$F_{(p-1,n-k-p+1)} = \frac{(\mathrm{SSE}_R - \mathrm{SSE}_U)/(p-1)}{\mathrm{SSE}_U/(n-k-p+1)}$$

where SSE_R is the sum of squared errors due to the restricted model, SSE_U is the sum of squared errors due to the unrestricted model, n is the total number of observations, and k is the number of parameters in the original linear model.

Ramsey's test can be viewed as a linearity test that checks whether any nonlinear transformation of the specified independent variables has been omitted, but it need not help in identifying a new relevant variable other than those already specified in the current model.

Testing for Nonlinear Dependence: Heteroscedasticity 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),

$$Q(q) = N(N+2)\sum_{i=1}^{q}\frac{r(i;\hat{\nu}_t^2)}{(N-i)}$$

where

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

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

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" on page 408. Therefore, the Q statistic calculated from the squared residuals can be used to identify the order of the GARCH process.

Engle's 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 qth order ARCH process is written

$$LM(q) = \frac{N\mathbf{W}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}}{\mathbf{W}'\mathbf{W}}$$

where

$$\mathbf{W} = \left(\frac{\hat{\nu}_1^2}{\hat{\sigma}^2}-1,\ldots,\frac{\hat{\nu}_N^2}{\hat{\sigma}^2}-1\right)'$$

and

$$\mathbf{Z} = \begin{bmatrix} 1 & \hat{\nu}_0^2 & \cdots & \hat{\nu}_{-q+1}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \hat{\nu}_{N-1}^2 & \cdots & \hat{\nu}_{N-q}^2 \end{bmatrix}$$

The presample values ($\nu_0^2,\ldots,\nu_{-q+1}^2$) have been set to 0. Note that the LM(q) tests might have different finite-sample properties depending on the presample values, though they are asymptotically equivalent regardless of the presample values.

Lee and King's Test for ARCH Disturbances

Engle's Lagrange multiplier test for ARCH disturbances is a two-sided test; that is, it ignores the inequality constraints for the coefficients in ARCH models. Lee and King (1993) propose a one-sided test and prove that the test is locally most mean powerful. Let $\varepsilon_t, t = 1, \ldots, T$, denote the residuals to be tested. Lee and King's test checks

$$H_0 : \alpha_i = 0, i = 1, \ldots, q$$

$$H_1 : \alpha_i > 0, i = 1, \ldots, q$$

where $\alpha_i, i = 1, \ldots, q$, are in the following ARCH(q) model:

$$\varepsilon_t = \sqrt{h_t} e_t, e_t \; iid(0, 1)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$$

The statistic is written as

$$S = \frac{\sum_{t=q+1}^{T} (\frac{\varepsilon_t^2}{h_0} - 1) \sum_{i=1}^{q} \varepsilon_{t-i}^2}{\left[2 \sum_{t=q+1}^{T} (\sum_{i=1}^{q} \varepsilon_{t-i}^2)^2 - \frac{2(\sum_{t=q+1}^{T} \sum_{i=1}^{q} \varepsilon_{t-i}^2)^2}{T-q} \right]^{1/2}}$$

Wong and Li's Test for ARCH Disturbances

Wong and Li (1995) propose a rank portmanteau statistic to minimize the effect of the existence of outliers in the test for ARCH disturbances. They first rank the squared residuals; that is, $R_t = rank(\varepsilon_t^2)$. Then they calculate the rank portmanteau statistic

$$Q_R = \sum_{i=1}^{q} \frac{(r_i - \mu_i)^2}{\sigma_i^2}$$

where r_i, μ_i, and σ_i^2 are defined as follows:

$$r_i = \frac{\sum_{t=i+1}^{T} (R_t - (T+1)/2)(R_{t-i} - (T+1)/2)}{T(T^2-1)/12}$$

$$\mu_i = -\frac{T-i}{T(T-1)}$$

$$\sigma_i^2 = \frac{5T^4 - (5i+9)T^3 + 9(i-2)T^2 + 2i(5i+8)T + 16i^2}{5(T-1)^2 T^2 (T+1)}$$

The Q, Engle's LM, Lee and King's, and Wong and Li's statistics are computed from the OLS residuals, or residuals if the NLAG= option is specified, assuming that disturbances are white noise. The Q, Engle's LM, and Wong and Li's statistics have an approximate $\chi_{(q)}^2$ distribution under the white-noise null hypothesis, while the Lee and King's statistic has a standard normal distribution under the white-noise null hypothesis.

Testing for Structural Change

Chow Test

Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

where the parameter vector β contains k elements.

Split the observations for this model into two subsets at the break point specified by the CHOW= option, so that

$$\begin{aligned} \mathbf{y} &= (\mathbf{y'}_1, \mathbf{y'}_2)' \\ \mathbf{X} &= (\mathbf{X'}_1, \mathbf{X'}_2)' \\ \mathbf{u} &= (\mathbf{u'}_1, \mathbf{u'}_2)' \end{aligned}$$

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

$$\mathbf{y}_1 = \mathbf{X}_1\beta_1 + \mathbf{u}_1$$

$$\mathbf{y}_2 = \mathbf{X}_2\beta_2 + \mathbf{u}_2$$

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 $V(\mathbf{u}_1) = V(\mathbf{u}_2)$. The Chow test is computed using three sums of square errors,

$$\mathrm{F}_{chow} = \frac{(\hat{\mathbf{u}}'\hat{\mathbf{u}} - \hat{\mathbf{u}}'_1\hat{\mathbf{u}}_1 - \hat{\mathbf{u}}'_2\hat{\mathbf{u}}_2)/k}{(\hat{\mathbf{u}}'_1\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}'_2\hat{\mathbf{u}}_2)/(n_1 + n_2 - 2k)}$$

where $\hat{\mathbf{u}}$ is the regression residual vector from the full set model, $\hat{\mathbf{u}}_1$ is the regression residual vector from the first set model, and $\hat{\mathbf{u}}_2$ is the regression residual vector from 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 β.

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

$$\mathrm{F}_{pchow} = \frac{(\hat{\mathbf{u}}'\hat{\mathbf{u}} - \hat{\mathbf{u}}'_1\hat{\mathbf{u}}_1)/n_2}{\hat{\mathbf{u}}'_1\hat{\mathbf{u}}_1/(n_1 - k)}$$

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

Bai and Perron's Multiple Structural Change Tests

Bai and Perron (1998) propose several kinds of multiple structural change tests: (1) the test of no break versus a fixed number of breaks ($supF$ test), (2) the equal and unequal weighted versions of double maximum tests of no break versus an unknown number of breaks given some upper bound ($UDmaxF$ test and $WDmaxF$ test), and (3) the test of l versus $l + 1$ breaks ($supF_{l+1|l}$ test). Bai and Perron (2003a, b, 2006) also show how to implement these tests, the commonly used critical values, and the simulation analysis on these tests.

Consider the following partial structural change model with m breaks ($m+1$ regimes):

$$y_t = x_t'\beta + z_t'\delta_j + u_t, \qquad t = T_{j-1}+1, \ldots, T_j,\ j = 1, \ldots, m$$

Here, y_t is the dependent variable observed at time t, $x_t(p \times 1)$ is a vector of covariates with coefficients β unchanged over time, and $z_t(q \times 1)$ is a vector of covariates with coefficients δ_j at regime j, $j = 1, \ldots, m$. If $p = 0$ (that is, there are no x regressors), the regression model becomes the pure structural change model. The indices $(T_1, \ldots, T_m)$ (that is, the break dates or break points) are unknown, and the convenient notation $T_0 = 0$ and $T_{m+1} = T$ applies. For any given m-partition $(T_1, \ldots, T_m)$, the associated least squares estimates of β and δ_j, $j = 1, \ldots, m$, are obtained by minimizing the sum of squared residuals (SSR),

$$S_T(T_1, \ldots, T_m) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - x_t'\beta - z_t'\delta_i)^2$$

Let $\hat{S}_T(T_1, \ldots, T_m)$ denote the minimized SSR for a given $(T_1, \ldots, T_m)$. The estimated break dates $(\hat{T}_1, \ldots, \hat{T}_m)$ are such that

$$(\hat{T}_1, \ldots, \hat{T}_m) = \arg \min_{T_1, \ldots, T_m} \hat{S}_T(T_1, \ldots, T_m)$$

where the minimization is taken over all partitions $(T_1, \ldots, T_m)$ such that $T_i - T_{i-1} \geq T\epsilon$. Bai and Perron (2003a) propose an efficient algorithm, based on the principle of dynamic programming, to estimate the preceding model.

In the case that the data are nontrending, as stated in Bai and Perron (1998), the limiting distribution of the break dates is

$$\frac{(\Delta_i' Q_i \Delta_i)^2}{(\Delta_i' \Omega_i \Delta_i)}(\hat{T}_i - T_i^0) \Rightarrow \arg\max_s V^{(i)}(s), \qquad i = 1, \ldots, m$$

where

$$V^{(i)}(s) = \begin{cases} W_1^{(i)}(-s) - |s|/2 & \text{if } s \leq 0 \\ \sqrt{\eta_i}(\phi_{i,2}/\phi_{i,1}) W_2^{(i)}(s) - \eta_i |s|/2 & \text{if } s > 0 \end{cases}$$

and

$$\begin{aligned} \Delta T_i^0 &= T_i^0 - T_{i-1}^0 \\ \Delta_i &= \delta_{i+1}^0 - \delta_i^0 \\ Q_i &= \lim (\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_i^0} E(z_t z_t') \\ \Omega_i &= \lim (\Delta T_i^0)^{-1} \sum_{r=T_{i-1}^0+1}^{T_i^0} \sum_{t=T_{i-1}^0+1}^{T_i^0} E(z_r z_t' u_r u_t) \\ \eta_i &= \Delta_i' Q_{i+1} \Delta_i / \Delta_i' Q_i \Delta_i \\ \phi_{i,1}^2 &= \Delta_i' \Omega_i \Delta_i / \Delta_i' Q_i \Delta_i \\ \phi_{i,2}^2 &= \Delta_i' \Omega_{i+1} \Delta_i / \Delta_i' Q_{i+1} \Delta_i \end{aligned}$$

Also, $W_1^{(i)}(s)$ and $W_2^{(i)}(s)$ are independent standard Weiner processes that are defined on $[0, \infty)$, starting at the origin when $s = 0$; these processes are also independent across i. The cumulative distribution function of $\arg\max_s V^{(i)}(s)$ is shown in Bai (1997). Hence, with the estimates of Δ_i, Q_i, and Ω_i, the relevant critical values for confidence interval of break dates T_i can be calculated. The estimate of Δ_i is $\hat{\delta}_{i+1} - \hat{\delta}_i$. The estimate of Q_i is either

$$\hat{Q}_i = (\Delta \hat{T}_i)^{-1} \sum_{t=\hat{T}_{i-1}^0+1}^{\hat{T}_i^0} z_t z_t'$$

if the regressors are assumed to have heterogeneous distributions across regimes (that is, the HQ option is specified), or

$$\hat{Q}_i = \hat{Q} = (T)^{-1} \sum_{t=1}^{T} z_t z_t'$$

if the regressors are assumed to have identical distributions across regimes (that is, the HQ option is not specified). The estimate of Ω_i can also be constructed with data over regime i only or the whole sample, depending on whether the vectors $z_t \hat{u}_t$ are heterogeneously distributed across regimes (that is, the HO option is specified). If the HAC option is specified, $\hat{\Omega}_i$ is estimated through the heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator applied to vectors $z_t \hat{u}_t$.

The $supF$ test of no structural break ($m = 0$) versus the alternative hypothesis that there are a fixed number, $m = k$, of breaks is defined as

$$supF(k) = \frac{1}{T}\left(\frac{T - (k+1)q - p}{kq}\right)(R\hat{\theta})'(R\hat{V}(\hat{\theta})R')^{-1}(R\hat{\theta})$$

where

$$R_{(kq)\times(p+(k+1)q)} = \begin{pmatrix} 0_{q\times p} & I_q & -I_q & 0 & 0 & \cdots & 0 \\ 0_{q\times p} & 0 & I_q & -I_q & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \ddots & \ddots & \ddots & \cdots \\ 0_{q\times p} & 0 & \cdots & \cdots & 0 & I_q & -I_q \end{pmatrix}$$

and I_q is the $q \times q$ identity matrix; $\hat{\theta}$ is the coefficient vector $(\hat{\beta}'\hat{\delta}_1' \ldots \hat{\delta}_{k+1})'$, which together with the break dates $(\hat{T}_1 \ldots \hat{T}_k)$ minimizes the global sum of squared residuals; and $\hat{V}(\hat{\theta})$ is an estimate of the variance-covariance matrix of $\hat{\theta}$, which could be estimated by using the HAC estimator or another way, depending on how the HAC, HR, and HE options are specified. The output $supF$ test statistics are scaled by q, the number of regressors, to be consistent with the limiting distribution; Bai and Perron (2003b, 2006) take the same action.

There are two versions of double maximum tests of no break against an unknown number of breaks given some upper bound M: the $UDmaxF$ test,

$$UDmaxF(M) = \max_{1\le m\le M} supF(m)$$

and the $WDmaxF$ test,

$$WDmaxF(M, \alpha) = \max_{1\le m\le M} \frac{c_\alpha(1)}{c_\alpha(m)} supF(m)$$

where α is the significance level and $c_\alpha(m)$ is the critical value of $supF(m)$ test given the significance level α. Four kinds of $WDmaxF$ tests that correspond to $\alpha = 0.100, 0.050, 0.025$, and 0.010 are implemented.

The $supF(l+1|l)$ test of l versus $l+1$ breaks is calculated in two ways that are asymptotically the same. In the first calculation, the method amounts to the application of $(l+1)$ tests of the null hypothesis of no structural change versus the alternative hypothesis of a single change. The test is applied to each segment that contains the observations $\hat{T}_{i-1}$ to $\hat{T}_i$ $(i = 1, \ldots, l+1)$. The $supF(l+1|l)$ test statistics are the maximum of these $(l+1)$ $supF$ test statistics. In the second calculation, for the given l breaks $(\hat{T}_1, \ldots, \hat{T}_l)$, the new break $\hat{T}^{(N)}$ is to minimize the global SSR:

$$\hat{T}^{(N)} = \arg\min_{T^{(N)}} SSR(\hat{T}_1, \ldots, \hat{T}_l; T^{(N)})$$

Then,

$$supF(l+1|l) = (T-(l+1)q-p)\frac{SSR(\hat{T}_1, \ldots, \hat{T}_l) - SSR(\hat{T}_1, \ldots, \hat{T}_l; \hat{T}^{(N)})}{SSR(\hat{T}_1, \ldots, \hat{T}_l)}$$

The p-value of each test is based on the simulation of the limiting distribution of that test.

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, $\mathbf{x}_t'\mathbf{b}$. 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= option in the OUTPUT statement. The ALPHACLM= option controls the confidence level for UCLM= and LCLM=. These confidence limits are for estimation of the mean of the dependent variable, $\mathbf{x}_t'\mathbf{b}$, where $\mathbf{x}_t$ is the column vector of independent variables at observation t.

The predicted values are computed as

$$\hat{y}_t = \mathbf{x}_t'\mathbf{b}$$

and the upper and lower confidence limits as

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

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

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.

$$\mathrm{Prob}(T > t_{\alpha/2}) = \alpha/2$$

where T is an observation from a t distribution with q degrees of freedom. The value of α can be set with the ALPHACLM= option. The degrees of freedom parameter, q, is taken to be the number of observations minus the number of free parameters in the final model. For the YW estimation method, the value of v is calculated as

$$\mathrm{v} = \sqrt{s^2 \mathbf{x}_t' (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{x}_t}$$

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

$$\mathrm{v} = \sqrt{s^2 \mathbf{x}_t' \mathbf{W} \mathbf{x}_t}$$

where $\mathbf{W}$ is the $k \times k$ submatrix of $(\mathbf{J}'\mathbf{J})^{-1}$ that corresponds to the regression parameters. For more information, see the section "Computational Methods" on page 368.

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= option in the OUTPUT statement. The ALPHACLI= option controls the confidence level for UCL= and LCL=. These confidence limits are for the predicted value,

$$\tilde{y}_t = \mathbf{x}_t' \mathbf{b} + \nu_{t|t-1}$$

where $\mathbf{x}_t$ is the vector of independent variables if all independent variables at time t are nonmissing, and $\nu_{t|t-1}$ is the minimum variance linear predictor of the error term, which is defined in the following recursive way given the autoregressive model, AR(m) model, for ν_t,

$$\nu_{s|t} = \begin{cases} -\sum_{i=1}^{m} \hat{\varphi}_i \nu_{s-i|t} & s > t \text{ or observation } s \text{ is missing} \\ y_s - \mathbf{x}_s' \mathbf{b} & 0 < s \leq t \text{ and observation } s \text{ is nonmissing} \\ 0 & s \leq 0 \end{cases}$$

where $\hat{\varphi}_i, i = 1, \ldots, m$, are the estimated AR parameters. Observation s is considered to be missing if the dependent variable or at least one independent variable is missing. If some of the independent variables at time t are missing, the predicted $\tilde{y}_t$ is also missing. With the same definition of $\nu_{s|t}$, the prediction method can be easily extended to the multistep forecast of $\tilde{y}_{t+d}, d > 0$:

$$\tilde{y}_{t+d} = \mathbf{x}_{t+d}' \mathbf{b} + \nu_{t+d|t-1}$$

The prediction method is implemented through the Kalman filter.

If $\tilde{y}_t$ is not missing, the upper and lower confidence limits are computed as

$$\tilde{u}_t = \tilde{y}_t + t_{\alpha/2} \mathrm{v}$$

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

where v, in this case, is computed as

$$\mathrm{v} = \sqrt{\mathbf{z}_t' \mathbf{V}_\beta \mathbf{z}_t + s^2 r}$$

where $\mathbf{V}_\beta$ is the variance-covariance matrix of the estimation of regression parameter β; $\mathbf{z}_t$ is defined as

$$\mathbf{z}_t = \mathbf{x}_t + \sum_{i=1}^{m} \hat{\varphi}_i \mathbf{x}_{t-i|t-1}$$

and $\mathbf{x}_{s|t}$ is defined in a similar way as $\nu_{s|t}$:

$$\mathbf{x}_{s|t} = \begin{cases} -\sum_{i=1}^{m} \hat{\varphi}_i \mathbf{x}_{s-i|t} & s > t \text{ or observation } s \text{ is missing} \\ \mathbf{x}_s & 0 < s \leq t \text{ and observation } s \text{ is nonmissing} \\ 0 & s \leq 0 \end{cases}$$

The formula for computing the prediction variance v is deducted based on Baillie (1979).

The value $s^2 r$ is the estimate of the conditional prediction error variance. At the start of the series, and after missing values, r is usually greater than 1. For the computational details of r, see the section "Predicting the Conditional Variance" on page 408. The plot of residuals and confidence limits in Example 8.4 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 might be appreciable in small samples. For some discussion of this problem for AR(1) models, see Harvey (1981).

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 in addition to a linear combination of past values. Put another way, the $\mathbf{L}^{-1}$ matrix defined in the section "Computational Methods" on page 368 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 as

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

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

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

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

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

where

$$\sigma^2_{t+d-j|t} = \mathbf{E}(\epsilon^2_{t+d-j}|\Psi_t)$$

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

$$g_j = -\varphi_1 g_{j-1} - \cdots - \varphi_m g_{j-m}$$

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:

$$\mathbf{V}(\xi_{t+d}|\Psi_t) = \sigma^2_{t+d|t}$$

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

$$h_t = \mathbf{E}(\epsilon^2_t|\Psi_{t-1}) = \sigma^2_{t|t-1}$$

The multistep forecast of conditional error variance of the EGARCH, QGARCH, TGARCH, PGARCH, and GARCH-M models cannot be calculated using the preceding formula for the GARCH model. The following formulas are recursively implemented to obtain the multistep forecast of conditional error variance of these models:

- for the EGARCH(p, q) model:

$$\ln(\sigma^2_{t+d|t}) = \omega + \sum_{i=d}^{q} \alpha_i g(z_{t+d-i}) + \sum_{j=1}^{d-1} \gamma_j \ln(\sigma^2_{t+d-j|t}) + \sum_{j=d}^{p} \gamma_j \ln(h_{t+d-j})$$

  where

$$g(z_t) = \theta z_t + |z_t| - E|z_t|$$

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

- for the QGARCH(p, q) model:

$$\begin{aligned}\sigma^2_{t+d|t} = \omega &+ \sum_{i=1}^{d-1} \alpha_i (\sigma^2_{t+d-i|t} + \psi_i^2) + \sum_{i=d}^{q} \alpha_i (\epsilon_{t+d-i} - \psi_i)^2 \\ &+ \sum_{j=1}^{d-1} \gamma_j \sigma^2_{t+d-j|t} + \sum_{j=d}^{p} \gamma_j h_{t+d-j}\end{aligned}$$

- for the TGARCH(p, q) model:

$$\begin{aligned}\sigma^2_{t+d|t} = \omega &+ \sum_{i=1}^{d-1} (\alpha_i + \psi_i/2) \sigma^2_{t+d-i|t} + \sum_{i=d}^{q} (\alpha_i + 1_{\epsilon_{t+d-i}<0} \psi_i) \epsilon^2_{t+d-i} \\ &+ \sum_{j=1}^{d-1} \gamma_j \sigma^2_{t+d-j|t} + \sum_{j=d}^{p} \gamma_j h_{t+d-j}\end{aligned}$$

- for the PGARCH(p, q) model:

$$(\sigma^2_{t+d|t})^\lambda = \omega + \sum_{i=1}^{d-1} \alpha_i((1+\psi_i)^{2\lambda} + (1-\psi_i)^{2\lambda})(\sigma^2_{t+d-i|t})^\lambda/2 + \sum_{i=d}^{q} \alpha_i(|\epsilon_{t+d-i}| - \psi_i\epsilon_{t+d-i})^{2\lambda} + \sum_{j=1}^{d-1} \gamma_j(\sigma^2_{t+d-j|t})^\lambda + \sum_{j=d}^{p} \gamma_j h^\lambda_{t+d-j}$$

- for the GARCH-M model: ignoring the mean effect and directly using the formula of the corresponding GARCH model.

If the conditional error variance is homoscedastic, the conditional prediction error variance is identical to the unconditional prediction error variance

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

since $\sigma^2_{t+d-j|t} = \sigma^2$. You can compute $s^2 r$ (which is the second term of the variance for the predicted value $\tilde{y}_t$ explained in the section "Predicting Future Series Realizations" on page 407) by using the formula $\sigma^2 \sum_{j=0}^{d-1} g_j^2$, and r is estimated from $\sum_{j=0}^{d-1} g_j^2$ by using the estimated autoregressive parameters.

Consider the following conditional prediction error variance:

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

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. For information about the output variables that can be created, see the section "OUTPUT Statement" on page 361. 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_*i* — the *i*th order autoregressive parameter estimate. There are *m* such variables _A_1 through _A_*m*, where *i*s the value of the NLAG= option.

_AH_*i* — 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 ω.

_AHP_*i* — the estimate of the ψ_i parameter in the PGARCH model, if a PGARCH model is specified. There are q such variables _AHP_1 through _AHP_q, where q is the value of the Q= option.

_AHQ_*i* — the estimate of the ψ_i parameter in the QGARCH model, if a QGARCH model is specified. There are q such variables _AHQ_1 through _AHQ_q, where q is the value of the Q= option.

_AHT_*i* — the estimate of the ψ_i parameter in the TGARCH model, if a TGARCH model is specified. There are q such variables _AHT_1 through _AHT_q, where q is the value of the Q= option.

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_*i* — 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.

_HET_*i* — the *i*th heteroscedasticity model parameter specified by the HETERO statement

INTERCEPT — the intercept estimate. INTERCEPT 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

LAMBDA — the estimate of the power parameter λ in the PGARCH model, if a PGARCH model is specified.

LIKLHD — the log-likelihood value of the GARCH model

SSE — the value of the error sum of squares

START — the estimated start-up value for the conditional variance when GARCH= (STARTUP=ESTIMATE) option is specified

STATUS — This variable indicates the optimization status. _STATUS_ $= 0$ indicates that there were no errors during the optimization and the algorithm converged. _STATUS_ $= 1$ indicates that the optimization could not improve the function value and means that the results should be interpreted with caution. _STATUS_ $= 2$ indicates that the optimization failed due to the number of iterations exceeding either the maximum default or the specified number of iterations or the number of function calls allowed. _STATUS_ $= 3$ indicates that an error occurred during the optimization process. For example, this error message is obtained when a function or its derivatives cannot be calculated at the initial values or

	during the iteration process, when an optimization step is outside of the feasible region or when active constraints are linearly dependent.
STDERR	standard error of the parameter estimate, if the COVOUT option is specified.
TDFI	the estimate of the inverted degrees of freedom for Student's t distribution, if DIST=T is specified.
THETA	the estimate of the θ parameter in the EGARCH model, if an EGARCH model is specified.
TYPE	PARM 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 (Standard Error), and the ratio of estimate to standard error (t Value)
7. the statistics of fit 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 mean absolute error (MAE), the mean absolute percentage error (MAPE), the root mean square error (Root MSE), the Schwarz information criterion (SBC), the Hannan-Quinn information criterion (HQC), Akaike's information criterion (AIC), the corrected Akaike's information criterion (AICC), the Durbin-Watson statistic (Durbin-Watson), the transformed regression R^2 (Transformed Regress R-Square), and the total R^2 (Total R-Square). For GARCH models, the following additional items are printed:
   - the value of the log-likelihood function (Log Likelihood)
   - the number of observations that are used in estimation (Observations)
   - the unconditional variance (Uncond Var)
   - the normality test statistic and its p-value (Normality Test and Pr > ChiSq)

8. the parameter estimates for the structural model (Estimate), a standard error estimate (Standard Error), the ratio of estimate to standard error (t Value), and an approximation to the significance probability for the parameter being 0 (Approx Pr > |t|)

9. If the NLAG= option is specified with METHOD=ULS or METHOD=ML, the regression parameter estimates are printed again, assuming that the autoregressive parameter estimates are known. In this case, the Standard Error and related statistics for the regression estimates will, in general, be different from the case when they are estimated. Note that from a standpoint of estimation, Yule-Walker and iterated Yule-Walker methods (NLAG= with METHOD=YW, ITYW) generate only one table, assuming AR parameters are given.

10. If you specify the NORMAL option, the Jarque-Bera normality test statistics are printed. If you specify the LAGDEP option, Durbin's *h* or Durbin's *t* is printed.

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 Table 8.6.

Table 8.6 ODS Tables Produced in PROC AUTOREG

ODS Table Name	Description	Option
ODS Tables Created by the MODEL Statement		
ClassLevels	Class levels	Default
FitSummary	Summary of regression	Default
SummaryDepVarCen	Summary of regression (centered dependent var)	CENTER
SummaryNoIntercept	Summary of regression (no intercept)	NOINT
YWIterSSE	Yule-Walker iteration sum of squared error	METHOD=ITYW
PreMSE	Preliminary MSE	NLAG=
Dependent	Dependent variable	Default
DependenceEquations	Linear dependence equation	
ARCHTest	Tests for ARCH disturbances based on OLS residuals	ARCHTEST=
ARCHTestAR	Tests for ARCH disturbances based on residuals	ARCHTEST= (with NLAG=)
BDSTest	BDS test for independence	BDS<=()>
RunsTest	Runs test for independence	RUNS<=()>
TurningPointTest	Turning point test for independence	TP<=()>

Table 8.6 *continued*

ODS Table Name	Description	Option
VNRRankTest	Rank version of von Neumann ratio test for independence	VNRRANK<=()>
FitSummarySCBP	Fit summary of Bai and Perron's multiple structural change models	BP=
BreakDatesSCBP	Break dates of Bai and Perron's multiple structural change models	BP=
SupFSCBP	supF tests of Bai and Perron's multiple structural change models	BP=
UDmaxFSCBP	UDmaxF test of Bai and Perron's multiple structural change models	BP=
WDmaxFSCBP	WDmaxF tests of Bai and Perron's multiple structural change models	BP=
SeqFSCBP	supF(l+1|l) tests of Bai and Perron's multiple structural change models	BP=
ParameterEstimatesSCBP	Parameter estimates of Bai and Perron's multiple structural change models	BP=
ChowTest	Chow test and predictive Chow test	CHOW= PCHOW=
Godfrey	Godfrey's serial correlation test	GODFREY<=>
PhilPerron	Phillips-Perron unit root test	STATIONARITY= (PHILIPS<=()>) (no regressor)
PhilOul	Phillips-Ouliaris cointegration test	STATIONARITY= (PHILIPS<=()>) (has regressor)
ADF	Augmented Dickey-Fuller unit root test	STATIONARITY= (ADF<=()>) (no regressor)
EngleGranger	Engle-Granger cointegration test	STATIONARITY= (ADF<=()>) (has regressor)
ERS	ERS unit root test	STATIONARITY= (ERS<=()>)
NgPerron	Ng-Perron Unit root tests	STATIONARITY= (NP=<()>)
KPSS	Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test or Shin cointegration test	STATIONARITY= (KPSS<=()>)
ResetTest	Ramsey's RESET test	RESET

Table 8.6 *continued*

ODS Table Name	Description	Option
ARParameterEstimates	Estimates of autoregressive parameters	NLAG=
CorrGraph	Estimates of autocorrelations	NLAG=
BackStep	Backward elimination of autoregressive terms	BACKSTEP
ExpAutocorr	Expected autocorrelations	NLAG=
IterHistory	Iteration history	ITPRINT
ParameterEstimates	Parameter estimates	Default
ParameterEstimatesGivenAR	Parameter estimates assuming AR parameters are given	NLAG=, METHOD= ULS | ML
PartialAutoCorr	Partial autocorrelation	PARTIAL
CovB	Covariance of parameter estimates	COVB
CorrB	Correlation of parameter estimates	CORRB
CholeskyFactor	Cholesky root of gamma	ALL
Coefficients	Coefficients for first NLAG observations	COEF
GammaInverse	Gamma inverse	GINV
ConvergenceStatus	Convergence status table	Default
MiscStat	Durbin *t* or Durbin *h*, Jarque-Bera normality test	LAGDEP=; NORMAL
DWTest	Durbin-Watson statistics	DW=
ODS Tables Created by the RESTRICT Statement		
Restrict	Restriction table	Default
ODS Tables Created by the TEST Statement		
FTest	*F* test	Default, TYPE=ALL
WaldTest	Wald test	TYPE=WALD | ALL
LMTest	LM test	TYPE=LM | ALL (only supported with GARCH= option)
LRTest	LR test	TYPE=LR | ALL (only supported with GARCH= option)

ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*).

Before you create graphs, ODS Graphics must be enabled (for example, with the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" in that chapter.

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section "A Primer on ODS Statistical Graphics" in that chapter.

This section describes the use of ODS for creating graphics with the AUTOREG procedure.

To request these graphs, you must specify the ODS GRAPHICS statement. By default, only the residual, predicted versus actual, and autocorrelation of residuals plots are produced. If, in addition to the ODS GRAPHICS statement, you also specify the ALL option in either the PROC AUTOREG statement or MODEL statement, all plots are created. For HETERO, GARCH, and AR models studentized residuals are replaced by standardized residuals. For the autoregressive models, the conditional variance of the residuals is computed as described in the section "Predicting Future Series Realizations" on page 407. For the GARCH and HETERO models, residuals are assumed to have h_t conditional variance invoked by the HT= option of the OUTPUT statement. For all these cases, the Cook's D plot is not produced.

ODS Graph Names

PROC AUTOREG assigns a name to each graph it creates using ODS. You can use these names to reference the graphs when using ODS. The names are listed in Table 8.7.

Table 8.7 ODS Graphics Produced in PROC AUTOREG

ODS Table Name	Description	PLOTS= Option
DiagnosticsPanel	All applicable plots	
ACFPlot	Autocorrelation of residuals	ACF
FitPlot	Predicted versus actual plot	FITPLOT, default
CooksD	Cook's D plot	COOKSD (no NLAG=)
IACFPlot	Inverse autocorrelation of residuals	IACF
QQPlot	Q-Q plot of residuals	QQ
PACFPlot	Partial autocorrelation of residuals	PACF
ResidualHistogram	Histogram of the residuals	RESIDUALHISTOGRAM or RESIDHISTOGRAM
ResidualPlot	Residual plot	RESIDUAL or RES, default
StudentResidualPlot	Studentized residual plot	STUDENTRESIDUAL (no NLAG=, GARCH=, or HETERO)
StandardResidualPlot	Standardized residual plot	STANDARDRESIDUAL
WhiteNoiseLogProbPlot	Tests for white noise residuals	WHITENOISE

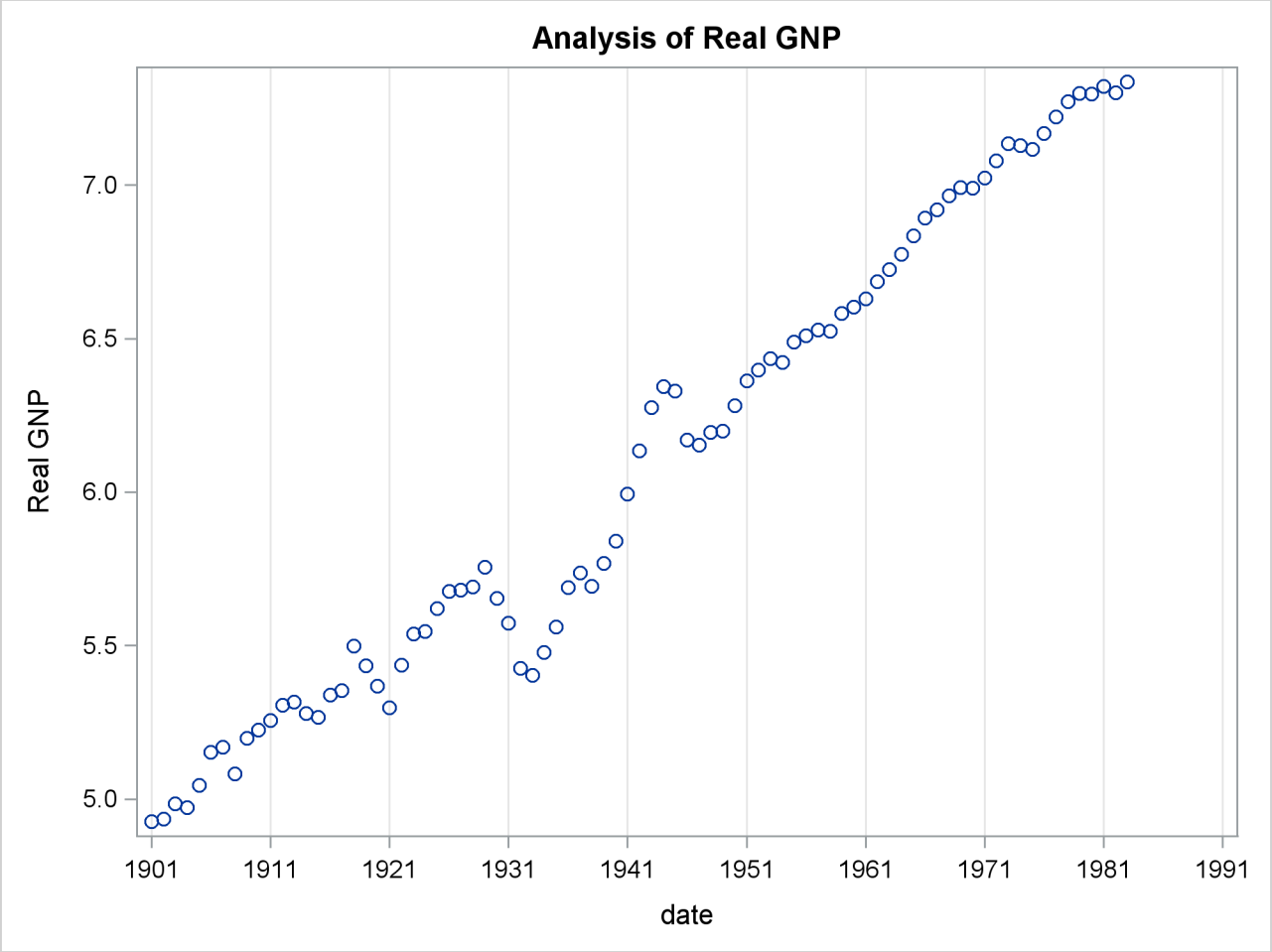
Examples: AUTOREG Procedure

Example 8.1: Analysis of Real Output Series

In this example, the annual real output series is analyzed over the period 1901 to 1983 (Balke and Gordon 1986, pp. 581–583). With the following DATA step, the original data are 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.

```
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';
datalines;
  137.87  139.13  146.10  144.21  155.04  172.97  175.61  161.22

   ... more lines ...


proc sgplot data=gnp noautolegend;
   scatter x=date y=y;
   xaxis grid values=('01jan1901'd '01jan1911'd '01jan1921'd '01jan1931'd
             '01jan1941'd '01jan1951'd '01jan1961'd '01jan1971'd
             '01jan1981'd '01jan1991'd);
run;
```

Output 8.1.1 Real Output Series: 1901–1983



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

$$y_t = \beta_0 + \beta_1 t + \nu_t$$

where

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

$$\epsilon_t \sim \mathrm{IN}(0, \sigma_\epsilon)$$

The preceding trend-stationary model assumes that uncertainty over future horizons is bounded since the error term, ν_t, has a finite variance. The maximum likelihood AR estimates from the statements that follow are shown in Output 8.1.2:

```
proc autoreg data=gnp;
   model y = t / nlag=2 method=ml;
run;
```

Output 8.1.2 Estimating the Linear Trend Model

Analysis of Real GNP

The AUTOREG Procedure

Maximum Likelihood Estimates			
SSE	0.23954331	**DFE**	79
MSE	0.00303	**Root MSE**	0.05507
SBC	-230.39355	**AIC**	-240.06891
MAE	0.04016596	**AICC**	-239.55609
MAPE	0.69458594	**HQC**	-236.18189
Log Likelihood	124.034454	**Transformed Regression R-Square**	0.8645
Durbin-Watson	1.9935	**Total R-Square**	0.9947
		Observations	83

Parameter Estimates						
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Variable 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**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Variable 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,

$$y_t = \beta_0 + y_{t-1} + v_t$$

where $v_t = \epsilon_t$ 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 τ_τ is the usual t ratio for the parameter estimate $\hat{\beta}_1$, but the τ_τ does not follow a t distribution.

The following code performs the augmented Dickey-Fuller test with $m = 3$ and we are interesting in the test results in the linear time trend case since the previous plot reveals there is a linear trend.

```
proc autoreg data = gnp;
   model y = / stationarity =(adf =3);
run;
```

The augmented Dickey-Fuller test indicates that the output series may have a difference-stationary process. The statistic Tau with linear time trend has a value of -2.6190 and its p-value is 0.2732. The statistic Rho has a p-value of 0.0817, which also indicates the null of unit root is accepted at the 5% level. (See Output 8.1.3.)

Output 8.1.3 Augmented Dickey-Fuller Test Results

Analysis of Real GNP

The AUTOREG Procedure

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	3	0.3827	0.7732	3.3342	0.9997		_
Single Mean	3	-0.1674	0.9465	-0.2046	0.9326	5.7521	0.0211
Trend	3	-18.0246	0.0817	-2.6190	0.2732	3.4472	0.4957

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 φ_1 is -0.297 and that of β_0 is 0.0293. All estimated values are statistically significant. The PROC step follows:

```
proc autoreg data=gnp;
   model dy = / nlag=1 method=ml;
run;
```

The printed output produced by the PROC step is shown in Output 8.1.4.

Output 8.1.4 Estimating the Differenced Series with AR(1) Error

Analysis of Real GNP

The AUTOREG Procedure

Maximum Likelihood Estimates			
SSE	0.27107673	DFE	80
MSE	0.00339	Root MSE	0.05821
SBC	-226.77848	AIC	-231.59192
MAE	0.04333026	AICC	-231.44002
MAPE	153.637587	HQC	-229.65939
Log Likelihood	117.795958	Transformed Regression R-Square	0.0000
Durbin-Watson	1.9268	Total R-Square	0.0900
		Observations	82

Output 8.1.4 *continued*

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx 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

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.0293	0.009093	3.22	0.0018

Example 8.2: Comparing Estimates and Models

In this example, the Grunfeld series are estimated using different estimation methods. For information about the Grunfeld investment data set, see Maddala (1977). For comparison, the Yule-Walker method, 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. The DATA and PROC steps follow:

```
title 'Grunfeld''s Investment Models Fit with Autoregressive Errors';
data grunfeld;
   input year gei gef gec;
   label gei = 'Gross investment GE'
         gec = 'Lagged Capital Stock GE'
         gef = 'Lagged Value of GE shares';
datalines;
1935     33.1      1170.6     97.8

   ... more lines ...


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	gei
	Gross investment GE

Output 8.2.1 *continued*

Ordinary Least Squares Estimates			
SSE	13216.5878	**DFE**	17
MSE	777.44634	**Root MSE**	27.88272
SBC	195.614652	**AIC**	192.627455
MAE	19.9433255	**AICC**	194.127455
MAPE	23.2047973	**HQC**	193.210587
Durbin-Watson	1.0721	**Total R-Square**	0.7053

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-9.9563	31.3742	-0.32	0.7548	
gef	1	0.0266	0.0156	1.71	0.1063	Lagged Value of GE shares
gec	1	0.1517	0.0257	5.90	<.0001	Lagged Capital Stock GE

Estimates of Autocorrelations

Lag	Covariance	Correlation	−1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	660.8	1.000000	|********************|
1	304.6	0.460867	|********* |

Preliminary MSE	520.5

Output 8.2.2 Regression Results Using Default Yule-Walker Method

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.460867	0.221867	-2.08

Yule-Walker Estimates			
SSE	10238.2951	**DFE**	16
MSE	639.89344	**Root MSE**	25.29612
SBC	193.742396	**AIC**	189.759467
MAE	18.0715195	**AICC**	192.426133
MAPE	21.0772644	**HQC**	190.536976
Durbin-Watson	1.3321	**Transformed Regression R-Square**	0.5717
		Total R-Square	0.7717

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-18.2318	33.2511	-0.55	0.5911	
gef	1	0.0332	0.0158	2.10	0.0523	Lagged Value of GE shares
gec	1	0.1392	0.0383	3.63	0.0022	Lagged Capital Stock GE

Output 8.2.3 Regression Results Using Unconditional Least Squares Method

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.460867	0.221867	-2.08

Algorithm converged.

Unconditional Least Squares Estimates

SSE	10220.8455	**DFE**	16
MSE	638.80284	**Root MSE**	25.27455
SBC	193.756692	**AIC**	189.773763
MAE	18.1317764	**AICC**	192.44043
MAPE	21.149176	**HQC**	190.551273
Durbin-Watson	1.3523	**Transformed Regression R-Square**	0.5511
		Total R-Square	0.7721

Parameter Estimates

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

Autoregressive parameters assumed given

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-18.6582	33.7567	-0.55	0.5881	
gef	1	0.0339	0.0159	2.13	0.0486	Lagged Value of GE shares
gec	1	0.1369	0.0404	3.39	0.0037	Lagged Capital Stock GE

Output 8.2.4 Regression Results Using Maximum Likelihood Method

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.460867	0.221867	-2.08

Algorithm converged.

Maximum Likelihood Estimates

SSE	10229.2303	**DFE**	16
MSE	639.32689	**Root MSE**	25.28491
SBC	193.738877	**AIC**	189.755947
MAE	18.0892426	**AICC**	192.422614
MAPE	21.0978407	**HQC**	190.533457
Log Likelihood	-90.877974	**Transformed Regression R-Square**	0.5656
Durbin-Watson	1.3385	**Total R-Square**	0.7719
		Observations	20

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-18.3751	34.5941	-0.53	0.6026	
gef	1	0.0334	0.0179	1.87	0.0799	Lagged Value of GE shares
gec	1	0.1385	0.0428	3.23	0.0052	Lagged Capital Stock GE
AR1	1	-0.4728	0.2582	-1.83	0.0858	

Autoregressive parameters assumed given

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-18.3751	33.3931	-0.55	0.5897	
gef	1	0.0334	0.0158	2.11	0.0512	Lagged Value of GE shares
gec	1	0.1385	0.0389	3.56	0.0026	Lagged Capital Stock GE

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 (AR(1) = −.976). The DATA step, PROC AUTOREG step, and PROC SGPLOT step follow:

```
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 plots;
   model y = x / nlag=1;
   output out=b p=pred pm=xbeta;
run;

proc sgplot data=b;
   scatter y=y x=time / markerattrs=(color=black);
   series y=pred x=time / lineattrs=(color=blue);
   series y=xbeta x=time / lineattrs=(color=red);
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 and Output 8.3.4. Note: the plot Output 8.3.3 can be viewed in the Autoreg.Model.FitDiagnosticPlots category by selecting **View►Results**.

Output 8.3.1 Results of OLS Analysis: No Autoregressive Model Fit

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

The AUTOREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	34.8061005	DFE	73
MSE	0.47680	Root MSE	0.69050
SBC	163.898598	AIC	159.263622
MAE	0.59112447	AICC	159.430289
MAPE	117894.045	HQC	161.114317
Durbin-Watson	0.0057	Total R-Square	0.0008

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx 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

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	0.4641	1.000000	|********************|
1	0.4531	0.976386	|******************** |

Preliminary MSE	0.0217

Output 8.3.2 Regression Results with AR(1) Error Correction

Estimates of Autoregressive Parameters

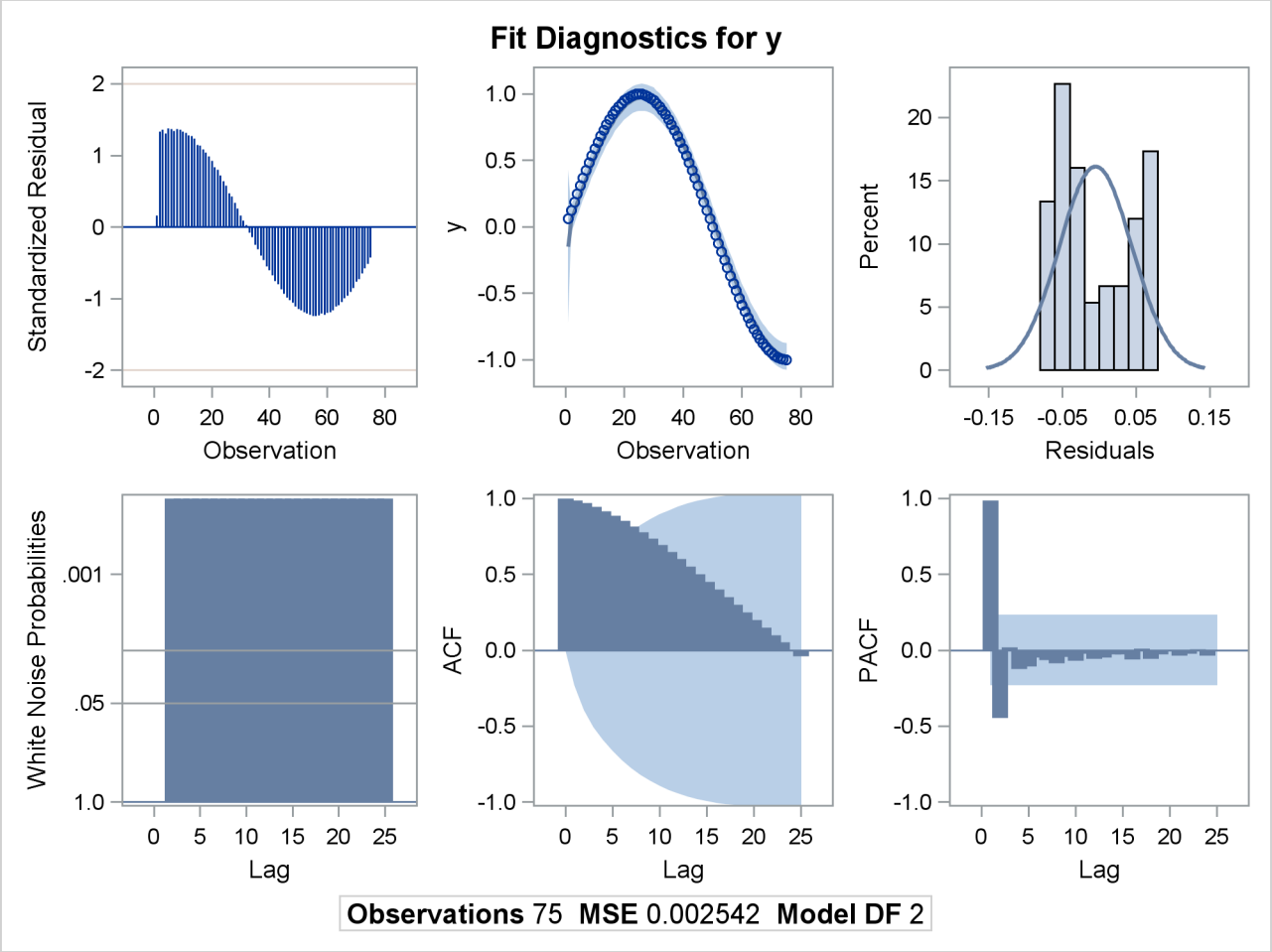
Lag	Coefficient	Standard Error	t Value
1	-0.976386	0.025460	-38.35

Yule-Walker Estimates			
SSE	0.18304264	DFE	72
MSE	0.00254	Root MSE	0.05042
SBC	-222.30643	AIC	-229.2589
MAE	0.04551667	AICC	-228.92087
MAPE	29145.3526	HQC	-226.48285
Durbin-Watson	0.0942	Transformed Regression R-Square	0.0001
		Total R-Square	0.9947

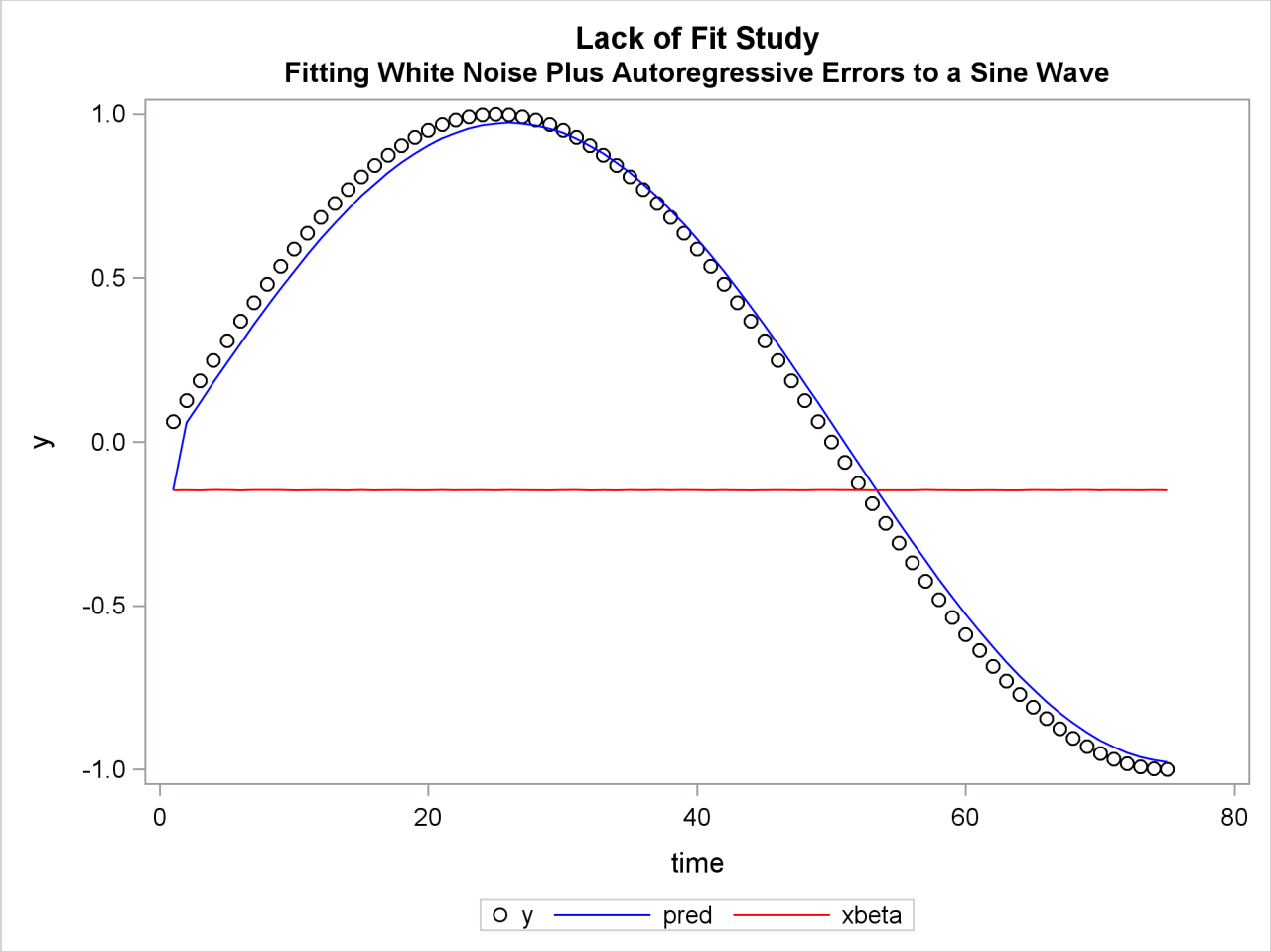
Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx 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 Diagnostics Plots



Output 8.3.4 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 15% of the values are randomly chosen and set to missing. The following statements generate the data:

```
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:

```
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;
```

The printed output produced by the AUTOREG procedure is shown in Output 8.4.1 and Output 8.4.2. Note: the plot Output 8.4.2 can be viewed in the Autoreg.Model.FitDiagnosticPlots category by selecting **View►Results**.

Output 8.4.1 Autocorrelation-Corrected Regression Results

Simulated Time Series with Roots:
(X-1.25)(X4-1.25)**
With 15% Missing Values

The AUTOREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	182.972379	**DFE**	40
MSE	4.57431	**Root MSE**	2.13876
SBC	181.39282	**AIC**	179.679248
MAE	1.80469152	**AICC**	179.781813
MAPE	270.104379	**HQC**	180.303237
Durbin-Watson	1.3962	**Total R-Square**	0.0000

Output 8.4.1 *continued*

Parameter Estimates

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

Estimates of Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	4.4627	1.000000	|********************|
1	1.4241	0.319109	|****** |
2	1.6505	0.369829	|******* |
3	0.6808	0.152551	|*** |
4	2.9167	0.653556	|************* |
5	-0.3816	-0.085519	| **|

Partial Autocorrelations	
1	0.319109
4	0.619288
5	-0.821179

Preliminary MSE 0.7609

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.733182	0.089966	-8.15
4	-0.803754	0.071849	-11.19
5	0.821179	0.093818	8.75

Expected Autocorrelations

Lag	Autocorr
0	1.0000
1	0.4204
2	0.2480
3	0.3160
4	0.6903
5	0.0228

Algorithm converged.

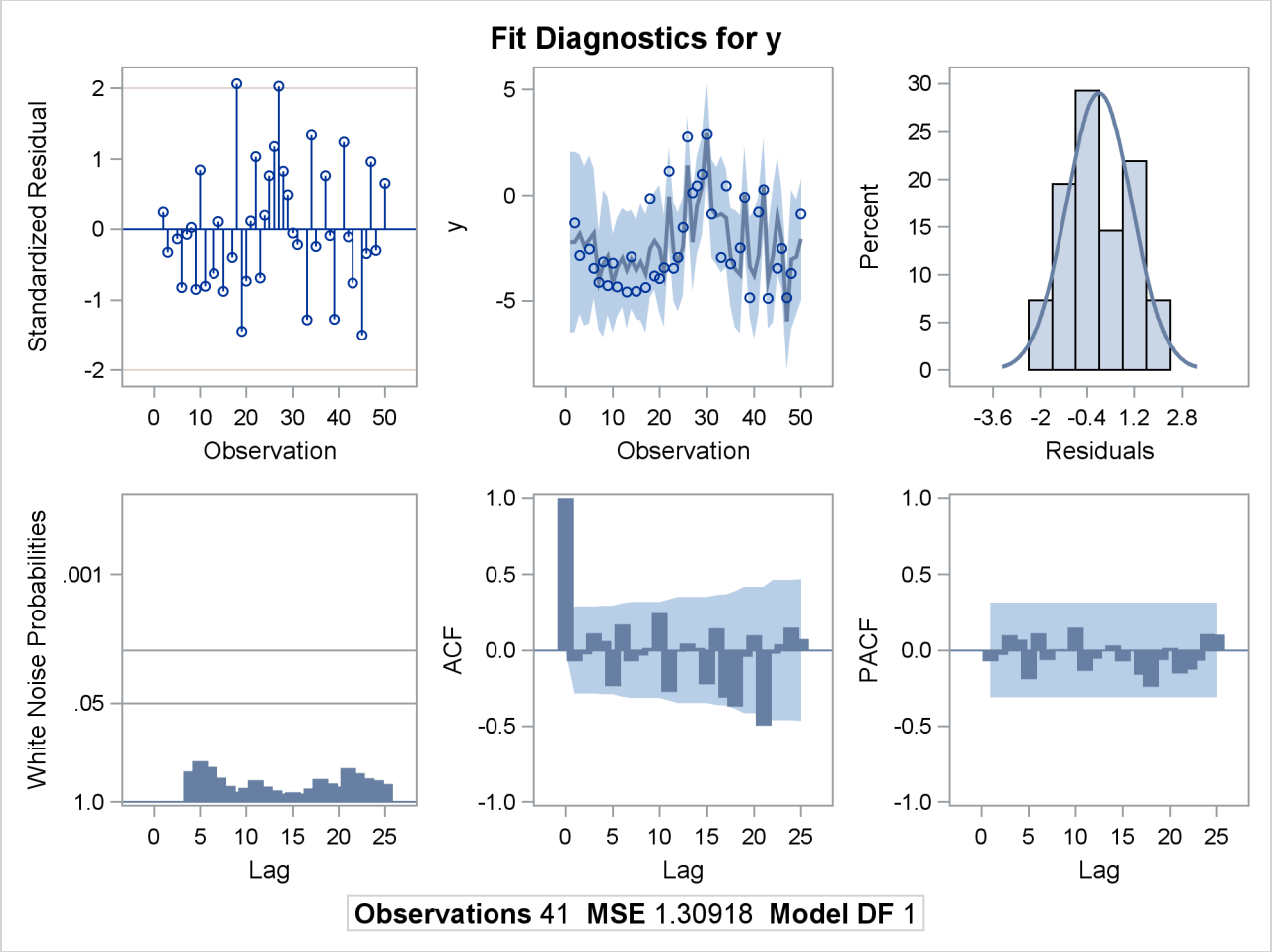
Maximum Likelihood Estimates			
SSE	48.4396756	**DFE**	37
MSE	1.30918	**Root MSE**	1.14419
SBC	146.879013	**AIC**	140.024725
MAE	0.88786192	**AICC**	141.135836
MAPE	141.377721	**HQC**	142.520679
Log Likelihood	-66.012362	**Transformed Regression R-Square**	0.0000
Durbin-Watson	2.9457	**Total R-Square**	0.7353
		Observations	41

Parameter Estimates					
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx 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

Expected Autocorrelations	
Lag	**Autocorr**
0	1.0000
1	0.4204
2	0.2423
3	0.2958
4	0.6318
5	0.0411

Autoregressive parameters assumed given					
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**
Intercept	1	-2.2370	0.5225	-4.28	0.0001

Output 8.4.2 Diagnostic Plots

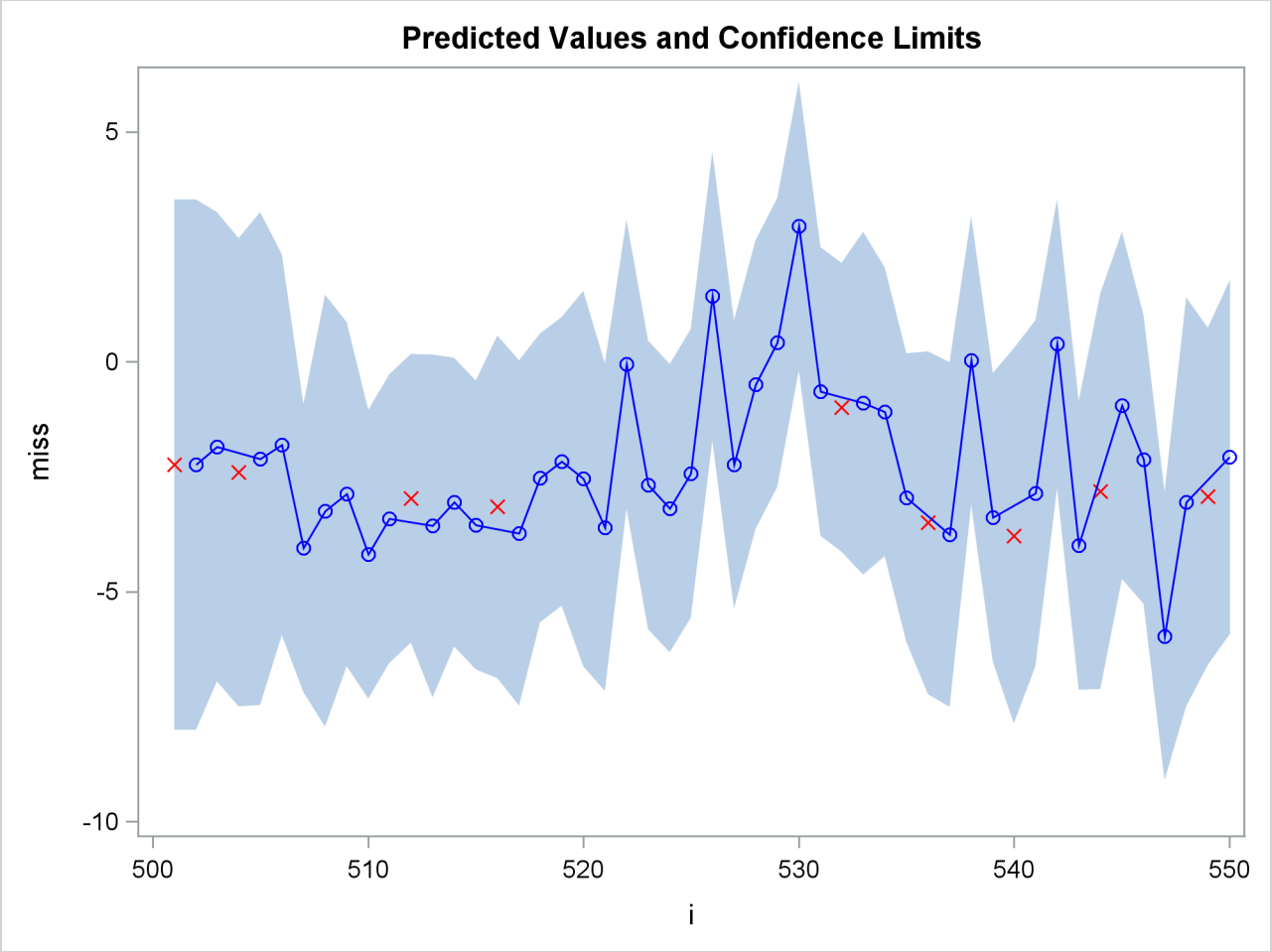


The following statements plot the residuals and confidence limits:

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

title 'Predicted Values and Confidence Limits';

proc sgplot data=reshape1 NOAUTOLEGEND;
   band x=i upper=u lower=l;
   scatter y=miss x=i/ MARKERATTRS =(symbol=x color=red);
   series y=p x=i/markers MARKERATTRS =(color=blue) lineattrs=(color=blue);
run;
```

The plot of the predicted values and the upper and lower confidence limits is shown in Output 8.4.3. 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.

Output 8.4.3 Plot of Predicted Values and Confidence Interval



Example 8.5: Money Demand Model

This example estimates the log-log money demand equation by 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. For a data description, see Balke and Gordon (1986).

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. The DATA step that follows generates the transformed variables:

```
data money;
   date = intnx( 'qtr', '01jan1968'd, _n_-1 );
   format date yyqc6.;
   input m1 gnp gdf ycb @@;
   m = log( 100 * m1 / gdf );
   m1cp = log( 100 * lag(m1) / gdf );
   y = log( gnp );
   intr = log( ycb );
   infr = 100 * log( gdf / lag(gdf) );
   label m    = 'Real Money Stock (M1)'
         m1cp = 'Lagged M1/Current GDF'
         y    = 'Real GNP'
         intr = 'Yield on Corporate Bonds'
         infr = 'Rate of Prices Changes';
datalines;
  187.15 1036.22   81.18  6.84

   ... more lines ...
```

Output 8.5.1 Money Demand Data Series – First 8 Observations

Predicted Values and Confidence Limits

Obs	date	m1	gnp	gdf	ycb	m	m1cp	y	intr	infr
1	1968:1	187.15	1036.22	81.18	6.84	5.44041	.	6.94333	1.92279	.
2	1968:2	190.63	1056.02	82.12	6.97	5.44732	5.42890	6.96226	1.94162	1.15127
3	1968:3	194.30	1068.72	82.80	6.98	5.45815	5.43908	6.97422	1.94305	0.82465
4	1968:4	198.55	1071.28	84.04	6.84	5.46492	5.44328	6.97661	1.92279	1.48648
5	1969:1	201.73	1084.15	84.97	7.32	5.46980	5.45391	6.98855	1.99061	1.10054
6	1969:2	203.18	1088.73	86.10	7.54	5.46375	5.45659	6.99277	2.02022	1.32112
7	1969:3	204.18	1091.90	87.49	7.70	5.45265	5.44774	6.99567	2.04122	1.60151
8	1969:4	206.10	1085.53	88.62	8.22	5.44917	5.43981	6.98982	2.10657	1.28331

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. The PROC step follows:

```
title 'Partial Adjustment Money Demand Equation';
title2 'Quarterly Data - 1968:2 to 1983:4';

proc autoreg data=money outest=est covout;
   model m = m1cp y intr infr / dw=4 dwprob;
run;
```

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			
SSE	0.00271902	**DFE**	58
MSE	0.0000469	**Root MSE**	0.00685
SBC	-433.68709	**AIC**	-444.40276
MAE	0.00483389	**AICC**	-443.35013
MAPE	0.08888324	**HQC**	-440.18824
		Total R-Square	0.9546

Durbin-Watson Statistics			
Order	**DW**	**Pr < DW**	**Pr > DW**
1	1.7355	0.0607	0.9393
2	2.1058	0.5519	0.4481
3	2.0286	0.5002	0.4998
4	2.2835	0.8880	0.1120

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

Parameter Estimates						
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx 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 GDF
y	1	0.0476	0.0122	3.89	0.0003	Real GNP
intr	1	-0.0238	0.007933	-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. The PROC step follows:

```
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 AR(1) should be interpreted as a positive autocorrelation.

Two predicted values are produced: 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

Lag	Coefficient	Standard Error	t Value
1	-0.126273	0.131393	-0.96

Algorithm converged.

Maximum Likelihood Estimates

SSE	0.00226719	**DFE**	57
MSE	0.0000398	**Root MSE**	0.00631
SBC	-439.47665	**AIC**	-452.33545
MAE	0.00506044	**AICC**	-450.83545
MAPE	0.09302277	**HQC**	-447.27802
Log Likelihood	232.167727	**Transformed Regression R-Square**	0.6954
Durbin-Watson	2.1778	**Total R-Square**	0.9621
		Observations	63

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx 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	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	2.4121	0.4685	5.15	<.0001	
m1cp	1	0.4086	0.0840	4.87	<.0001	Lagged M1/Current GDF
y	1	0.1509	0.0402	3.75	0.0004	Real GNP
intr	1	-0.1101	0.0155	-7.08	<.0001	Yield on Corporate Bonds
infr	1	-0.006348	0.001828	-3.47	0.0010	Rate of Prices Changes

Output 8.5.4 Partial List of the Predicted Values

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

Obs	p	pm	r	rm	ucl	lcl	uclm	lclm
1	.	.	.	.	.	.	.	.
2	5.45962	5.45962	-.005763043	-0.012301	5.49319	5.42606	5.47962	5.43962
3	5.45663	5.46750	0.001511258	-0.009356	5.46954	5.44373	5.48700	5.44800
4	5.45934	5.46761	0.005574104	-0.002691	5.47243	5.44626	5.48723	5.44799
5	5.46636	5.46874	0.003442075	0.001064	5.47944	5.45328	5.48757	5.44991
6	5.46675	5.46581	-.002994443	-0.002054	5.47959	5.45390	5.48444	5.44718
7	5.45672	5.45854	-.004074196	-0.005889	5.46956	5.44388	5.47667	5.44040
8	5.44404	5.44924	0.005136019	-0.000066	5.45704	5.43103	5.46726	5.43122

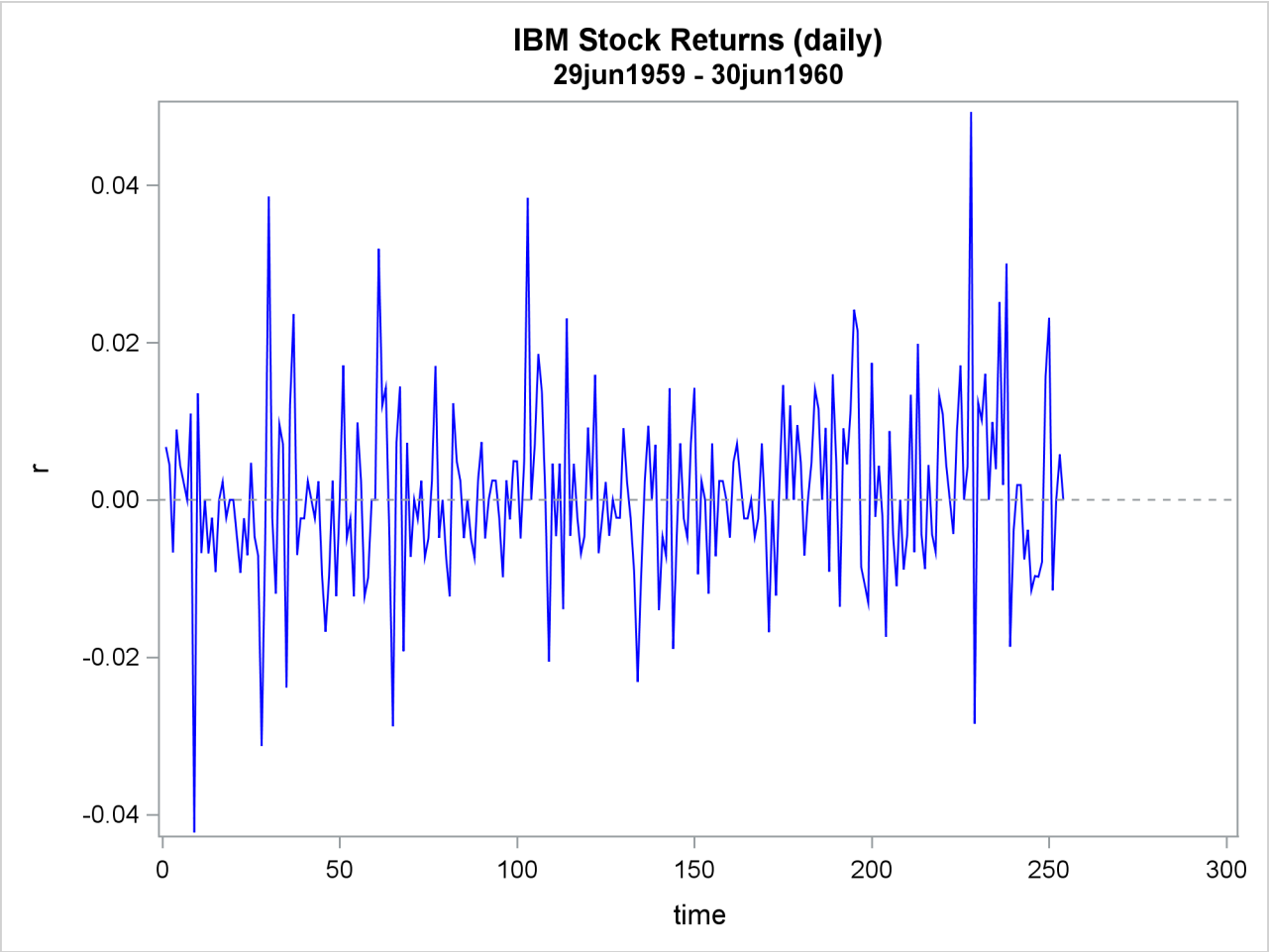
Example 8.6: Estimation of ARCH(2) Process

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 IBM common stock (Box and Jenkins 1976, p. 527) is shown in Output 8.6.1. The time series look serially uncorrelated, but the plot makes us skeptical of their independence.

With the following 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.

```
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;
445 448 450 447 451 453 454 454 459 440 446 443 443 440

   ... more lines ...


proc sgplot data=ibm;
   series y=r x=time/lineattrs=(color=blue);
         refline 0/ axis = y LINEATTRS = (pattern=ShortDash);
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 are calculated using parameter estimates,

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

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

where $d > 1$. This model can be estimated as follows:

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

The parameter estimates for ω, α_1, and α_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. The following code generates Output 8.6.3:

```
data b; set a;
   length type $ 8.;
   if r ^= . then do;
      type = 'ESTIMATE'; output; end;
   else do;
      type = 'FORECAST'; output; end;
run;
proc sgplot data=b;
   series x=time y=v/group=type;
   refline 254/ axis = x LINEATTRS = (pattern=ShortDash);
run;
```

Output 8.6.2 ARCH(2) Estimation Results

IBM Stock Returns (daily)
29jun1959 - 30jun1960

The AUTOREG Procedure

Dependent Variable r

Ordinary Least Squares Estimates			
SSE	0.03214307	**DFE**	254
MSE	0.0001265	**Root MSE**	0.01125
SBC	-1558.802	**AIC**	-1558.802
MAE	0.00814086	**AICC**	-1558.802
MAPE	100.378566	**HQC**	-1558.802
Durbin-Watson	2.1377	**Total R-Square**	0.0000

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

Algorithm converged.

GARCH Estimates			
SSE	0.03214307	**Observations**	254
MSE	0.0001265	**Uncond Var**	0.00012632
Log Likelihood	781.017441	**Total R-Square**	0.0000
SBC	-1545.4229	**AIC**	-1556.0349
MAE	0.00805675	**AICC**	-1555.9389
MAPE	100	**HQC**	-1551.7658
		Normality Test	105.8587
		Pr > ChiSq	<.0001

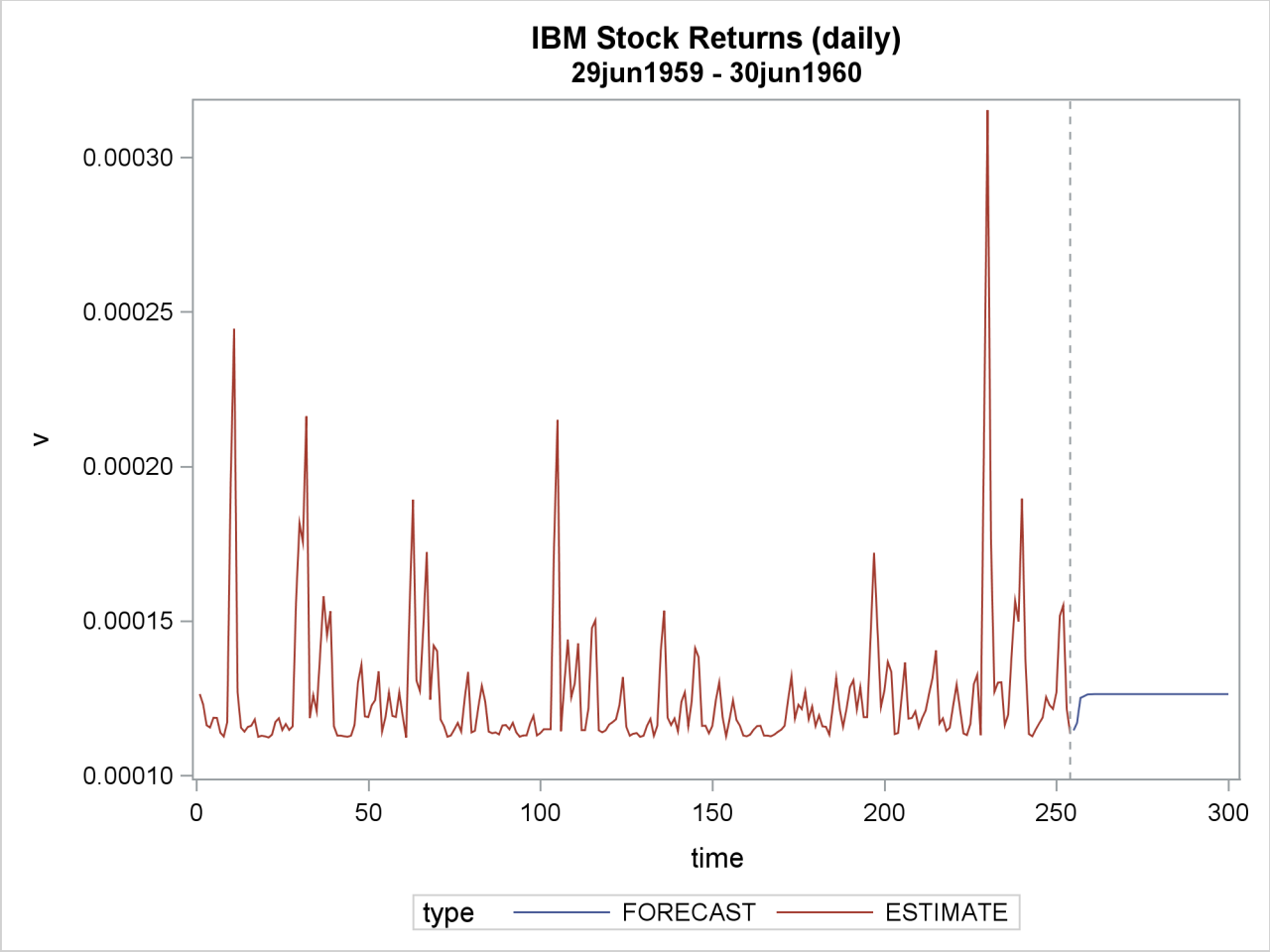
NOTE: No intercept term is used.
R-squares are redefined.

Output 8.6.2 *continued*

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
ARCH0	1	0.000112	7.6059E-6	14.76	<.0001
ARCH1	1	0.0414	0.0514	0.81	0.4208
ARCH2	1	0.0698	0.0434	1.61	0.1082

Output 8.6.3 Conditional Variance for IBM Stock Prices



Example 8.7: Estimation of GARCH-Type Models

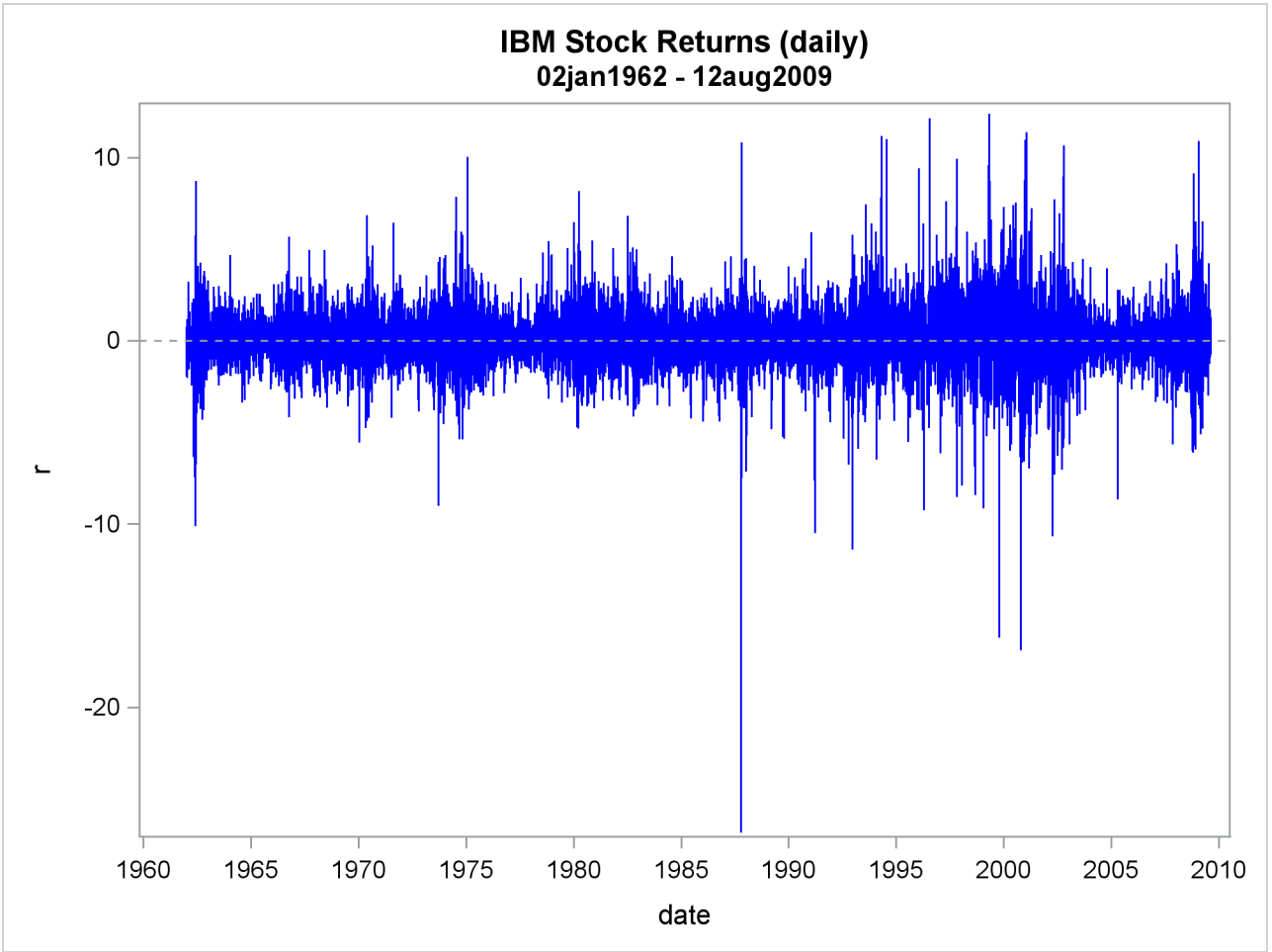
This example extends Example 8.6 to include more volatility models and to perform model selection and diagnostics.

Following are the data of daily IBM stock prices for the long period from 1962 to 2009:

```
data ibm_long;
   infile datalines;
   format date MMDDYY10.;
   input date:MMDDYY10. price_ibm;
   r = 100*dif( log( price_ibm ) );
datalines;
01/02/1962 2.68
01/03/1962 2.7
01/04/1962 2.67
01/05/1962 2.62
01/08/1962 2.57

   ... more lines ...

08/12/2009 119.29
;
```

The time series of IBM returns is depicted graphically in Output 8.7.1.

Output 8.7.1 IBM Stock Returns: Daily



The following statements perform estimation of different kinds of GARCH-type models. First, ODS listing output that contains fit summary tables for each single model is captured by using an ODS OUTPUT statement with the appropriate ODS table name assigned to a new SAS data set. Along with these new data sets, another one that contains parameter estimates is created by using the OUTEST= option in the AUTOREG statement.

```
/* Capturing ODS tables into SAS data sets */
ods output Autoreg.ar_1.FinalModel.FitSummary
           =fitsum_ar_1;
ods output Autoreg.arch_2.FinalModel.Results.FitSummary
           =fitsum_arch_2;
ods output Autoreg.garch_1_1.FinalModel.Results.FitSummary
           =fitsum_garch_1_1;
ods output Autoreg.st_garch_1_1.FinalModel.Results.FitSummary
           =fitsum_st_garch_1_1;
ods output Autoreg.ar_1_garch_1_1.FinalModel.Results.FitSummary
           =fitsum_ar_1_garch_1_1;
ods output Autoreg.igarch_1_1.FinalModel.Results.FitSummary
           =fitsum_igarch_1_1;
ods output Autoreg.garchm_1_1.FinalModel.Results.FitSummary
           =fitsum_garchm_1_1;
ods output Autoreg.egarch_1_1.FinalModel.Results.FitSummary
           =fitsum_egarch_1_1;
ods output Autoreg.qgarch_1_1.FinalModel.Results.FitSummary
           =fitsum_qgarch_1_1;
ods output Autoreg.tgarch_1_1.FinalModel.Results.FitSummary
           =fitsum_tgarch_1_1;
ods output Autoreg.pgarch_1_1.FinalModel.Results.FitSummary
           =fitsum_pgarch_1_1;

/* Estimating multiple GARCH-type models */
title "GARCH family";
proc autoreg data=ibm_long outest=garch_family;
   ar_1 :           model r = / noint nlag=1 method=ml;
   arch_2 :         model r = / noint garch=(q=2);
   garch_1_1 :      model r = / noint garch=(p=1,q=1);
   st_garch_1_1 :   model r = / noint garch=(p=1,q=1,type=stationary);
   ar_1_garch_1_1 : model r = / noint nlag=1 garch=(p=1,q=1);
   igarch_1_1 :     model r = / noint garch=(p=1,q=1,type=integ,noint);
   egarch_1_1 :     model r = / noint garch=(p=1,q=1,type=egarch);
   garchm_1_1 :     model r = / noint garch=(p=1,q=1,mean=log);
   qgarch_1_1 :     model r = / noint garch=(p=1,q=1,type=qgarch);
   tgarch_1_1 :     model r = / noint garch=(p=1,q=1,type=tgarch);
   pgarch_1_1 :     model r = / noint garch=(p=1,q=1,type=pgarch);
run;
```

The following statements print partial contents of the data set GARCH_FAMILY. The columns of interest are explicitly specified in the VAR statement.

```
/* Printing summary table of parameter estimates */
title "Parameter Estimates for Different Models";
proc print data=garch_family;
   var _MODEL_ _A_1 _AH_0 _AH_1 _AH_2
       _GH_1  _AHQ_1 _AHT_1 _AHP_1 _THETA_ _LAMBDA_ _DELTA_;
run;
```

These statements produce the results shown in Output 8.7.2.

Output 8.7.2 GARCH-Family Estimation Results

Parameter Estimates for Different Models

Obs	_MODEL_	_A_1	_AH_0	_AH_1	_AH_2	_GH_1	_AHQ_1	_AHT_1	_AHP_1
1	ar_1	0.017112	.	.	.	.	.	.	.
2	arch_2	.	1.60288	0.23235	0.21407	.	.	.	.
3	garch_1_1	.	0.02730	0.06984	.	0.92294	.	.	.
4	st_garch_1_1	.	0.02831	0.06913	.	0.92260	.	.	.
5	ar_1_garch_1_1	-0.005995	0.02734	0.06994	.	0.92282	.	.	.
6	igarch_1_1	.	.	0.00000	.	1.00000	.	.	.
7	egarch_1_1	.	0.01541	0.12882	.	0.98914	.	.	.
8	garchm_1_1	.	0.02897	0.07139	.	0.92079	.	.	.
9	qgarch_1_1	.	0.00120	0.05792	.	0.93458	0.66461	.	.
10	tgarch_1_1	.	0.02706	0.02966	.	0.92765	.	0.074815	.
11	pgarch_1_1	.	0.01623	0.06724	.	0.93952	.	.	0.43445

Obs	_THETA_	_LAMBDA_	_DELTA_
1	.	.	.
2	.	.	.
3	.	.	.
4	.	.	.
5	.	.	.
6	.	.	.
7	-0.41706	.	.
8	.	.	0.094773
9	.	.	.
10	.	.	.
11	.	0.53625	.

The table shown in Output 8.7.2 is convenient for reporting the estimation result of multiple models and their comparison.

The following statements merge multiple tables that contain fit statistics for each estimated model, leaving only columns of interest, and rename them:

```
/* Merging ODS output tables and extracting AIC and SBC measures */
data sbc_aic;
   set fitsum_arch_2 fitsum_garch_1_1 fitsum_st_garch_1_1
       fitsum_ar_1 fitsum_ar_1_garch_1_1 fitsum_igarch_1_1
       fitsum_egarch_1_1 fitsum_garchm_1_1
       fitsum_tgarch_1_1 fitsum_pgarch_1_1 fitsum_qgarch_1_1;
   keep Model SBC AIC;
   if Label1="SBC" then do; SBC=input(cValue1,BEST12.4); end;
   if Label2="SBC" then do; SBC=input(cValue2,BEST12.4); end;
   if Label1="AIC" then do; AIC=input(cValue1,BEST12.4); end;
   if Label2="AIC" then do; AIC=input(cValue2,BEST12.4); end;
   if not (SBC=.) then output;
run;
```

Next, sort the models by one of the criteria—for example, by AIC:

```
/* Sorting data by AIC criterion */
proc sort data=sbc_aic;
   by AIC;
run;
```

Finally, print the sorted data set:

```
title "Selection Criteria for Different Models";
proc print data=sbc_aic;
   format _NUMERIC_ BEST12.4;
run;
```

The result is given in Output 8.7.3.

Output 8.7.3 GARCH-Family Model Selection on the Basis of AIC and SBC

Selection Criteria for Different Models

Obs	Model	SBC	AIC
1	pgarch_1_1	42907.7292	42870.7722
2	egarch_1_1	42905.9616	42876.3959
3	tgarch_1_1	42995.4893	42965.9236
4	qgarch_1_1	43023.106	42993.5404
5	garchm_1_1	43158.4139	43128.8483
6	garch_1_1	43176.5074	43154.3332
7	ar_1_garch_1_1	43185.5226	43155.957
8	st_garch_1_1	43178.2497	43156.0755
9	arch_2	44605.4332	44583.259
10	ar_1	45922.0721	45914.6807
11	igarch_1_1	45925.5828	45918.1914

According to the smaller-is-better rule for the information criteria, the PGARCH(1,1) model is the leader by AIC while the EGARCH(1,1) is the model of choice according to SBC.

Next, check whether the power GARCH model is misspecified, especially, if dependence exists in the standardized residuals that correspond to the assumed independently and identically distributed (iid) disturbance. The following statements reestimate the power GARCH model and use the BDS test to check the independence of the standardized residuals:

```
proc autoreg data=ibm_long;
   model r = / noint garch=(p=1,q=1,type=pgarch) BDS=(Z=SR,D=2.0);
run;
```

The partial results listing of the preceding statements is given in Output 8.7.4.

Output 8.7.4 Diagnostic Checking of the PGARCH(1,1) Model

Selection Criteria for Different Models

The AUTOREG Procedure

BDS Test for Independence			
Distance	Embedding Dimension	BDS	Pr > |BDS|
2.0000	2	2.9691	0.0030
	3	3.3810	0.0007
	4	3.1299	0.0017
	5	3.3805	0.0007
	6	3.3368	0.0008
	7	3.1888	0.0014
	8	2.9576	0.0031
	9	2.7386	0.0062
	10	2.5553	0.0106
	11	2.3510	0.0187
	12	2.1520	0.0314
	13	1.9373	0.0527
	14	1.7210	0.0852
	15	1.4919	0.1357
	16	1.2569	0.2088
	17	1.0647	0.2870
	18	0.9635	0.3353
	19	0.8678	0.3855
	20	0.7660	0.4437

The results in Output 8.7.4 indicate that when embedded size is greater than 9, you fail to reject the null hypothesis of independence at 1% significance level, which is a good indicator that the PGARCH model is not misspecified.

Example 8.8: Illustration of ODS Graphics

This example illustrates the use of ODS GRAPHICS. This is a continuation of the section "Forecasting Autoregressive Error Models" on page 319.

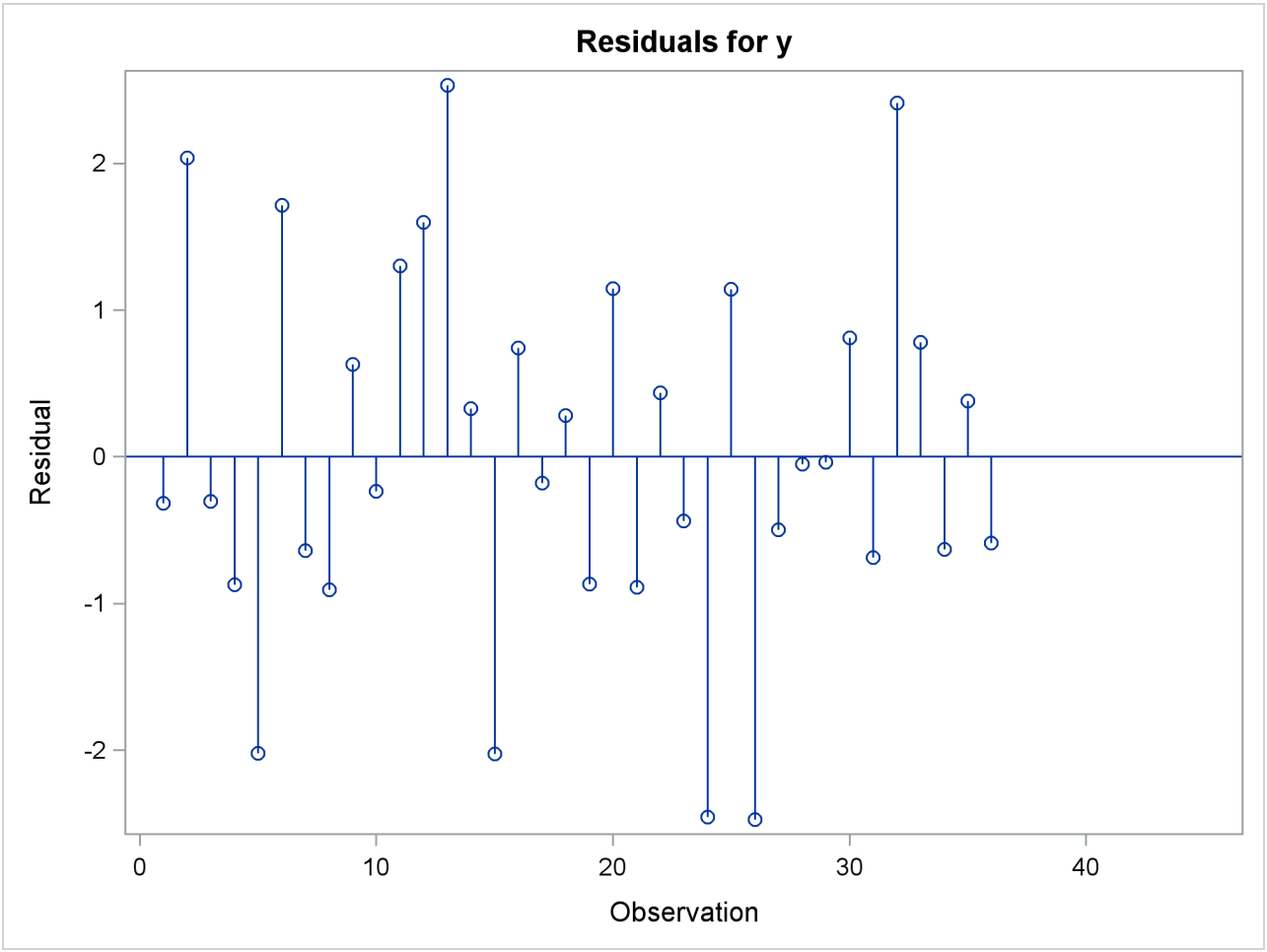
These graphical displays are requested by specifying the ODS GRAPHICS statement. For information about the graphs available in the AUTOREG procedure, see the section "ODS Graphics" on page 416.

The following statements show how to generate ODS GRAPHICS plots with the AUTOREG procedure. In this case, all plots are requested using the ALL option in the PROC AUTOREG statement, in addition to the ODS GRAPHICS statement. The plots are displayed in Output 8.8.1 through Output 8.8.8. Note: these plots can be viewed in the Autoreg.Model.FitDiagnosticPlots category by selecting **View►Results**.

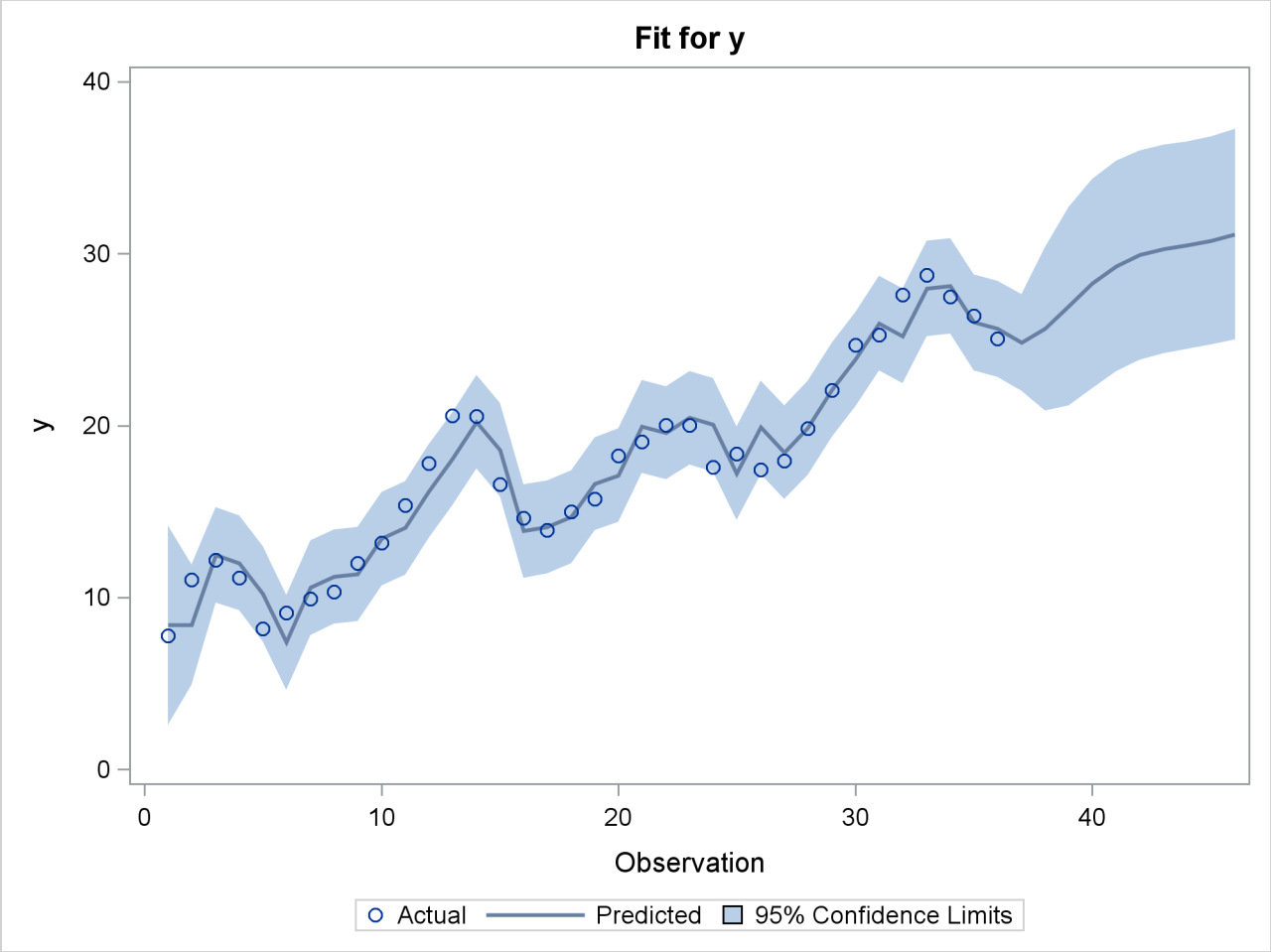
```
data a;
   ul = 0; ull = 0;
   do time = -10 to 36;
      u = + 1.3 * ul - .5 * ull + 2*rannor(12346);
```

```
      y = 10 + .5 * time + u;
      if time > 0 then output;
      ull = ul; ul = u;
   end;
run;

data b;
   y = .;
   do time = 37 to 46; output; end;
run;

data b;
   merge a b;
   by time;
run;

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

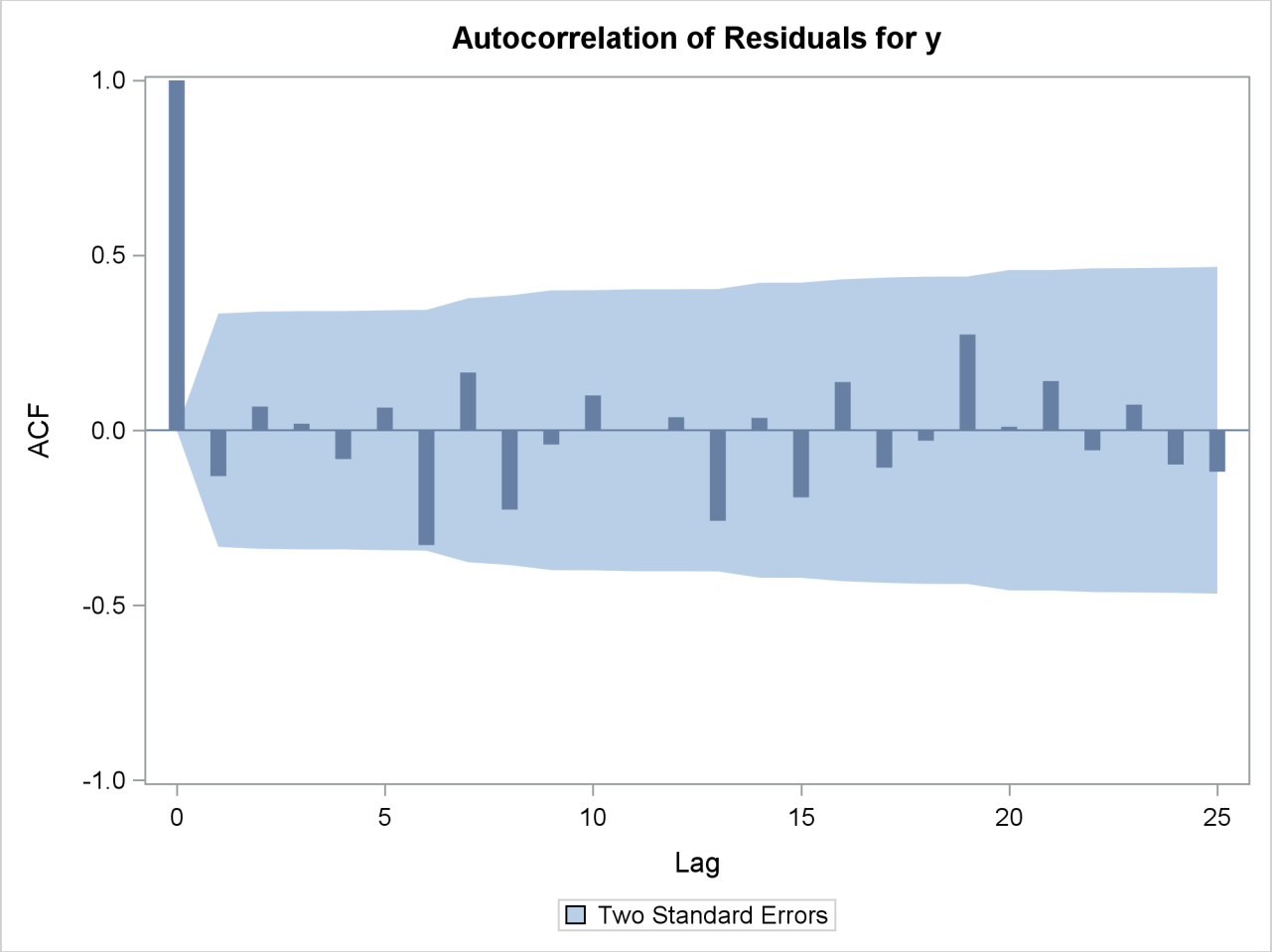
Output 8.8.1 Residuals Plot



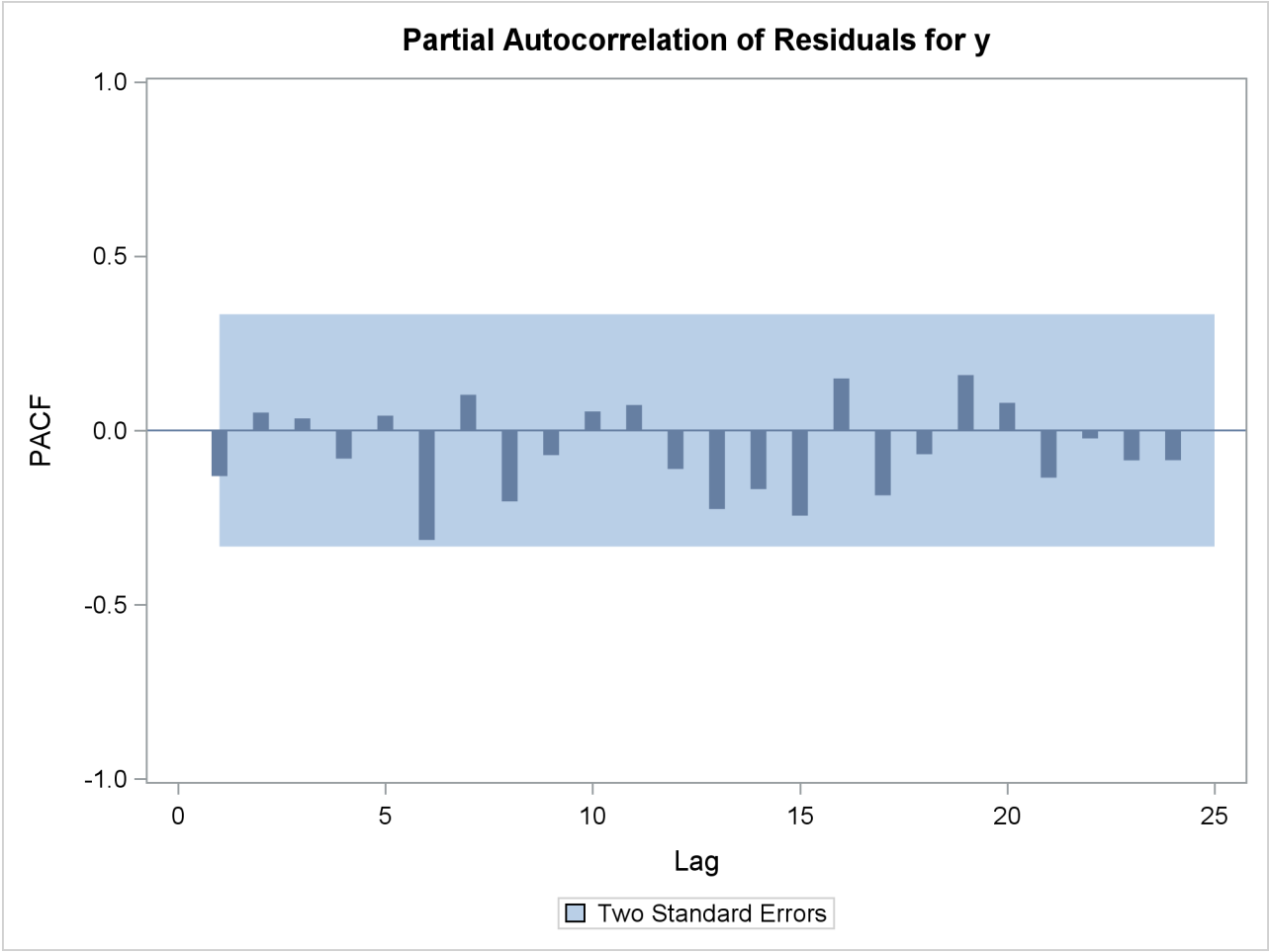
Output 8.8.2 Predicted versus Actual Plot



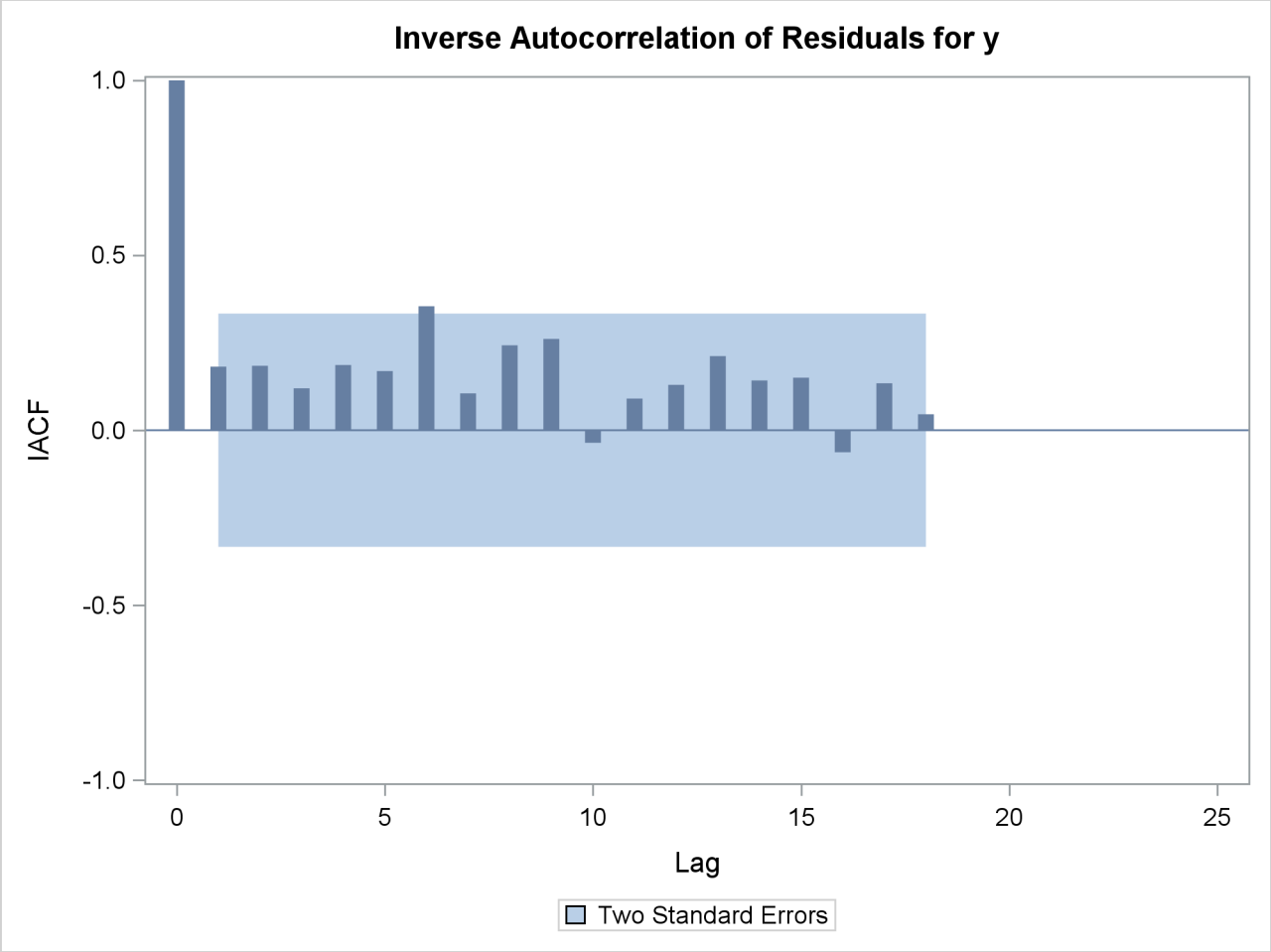
Output 8.8.3 Autocorrelation of Residuals Plot



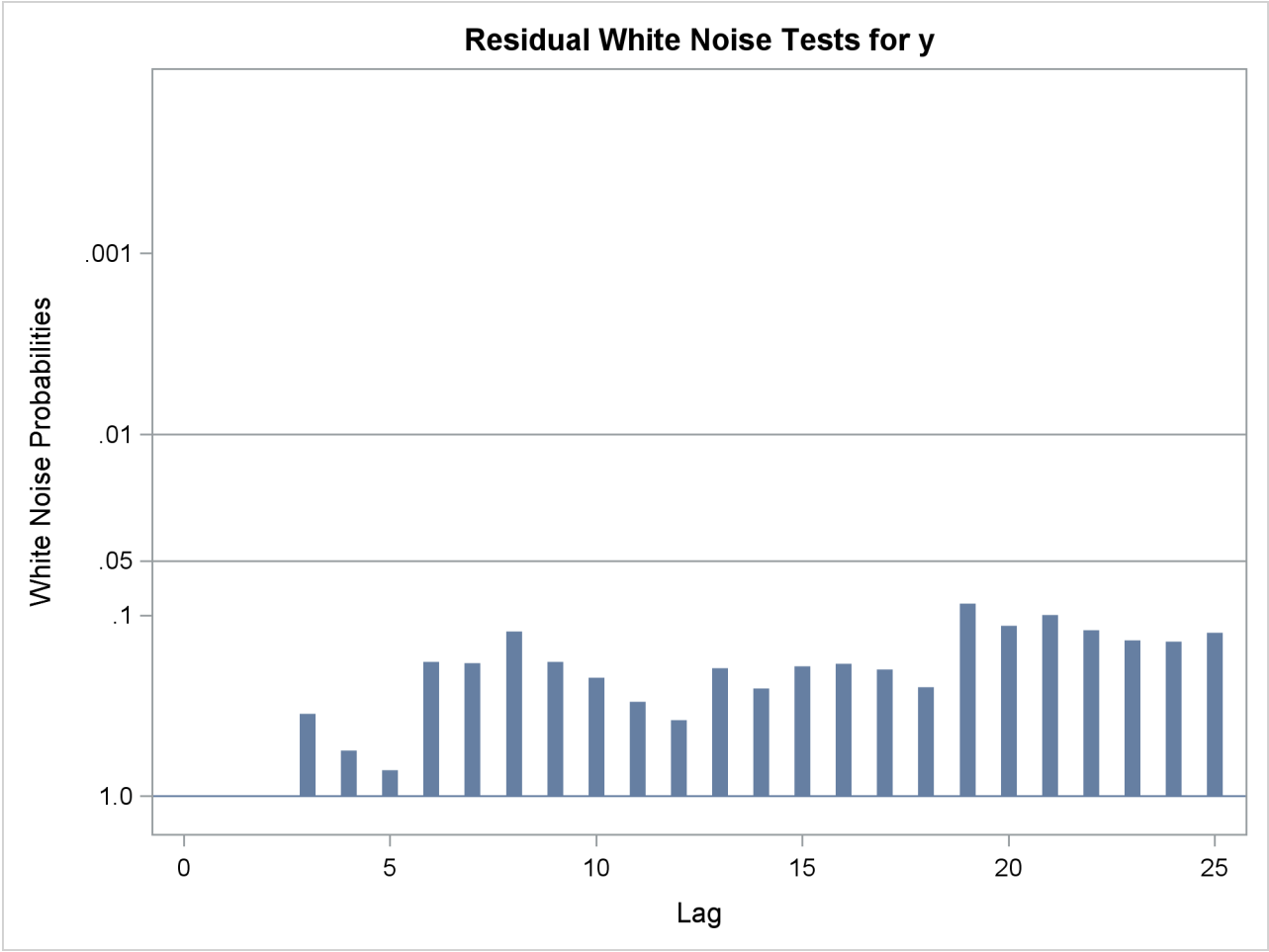
Output 8.8.4 Partial Autocorrelation of Residuals Plot



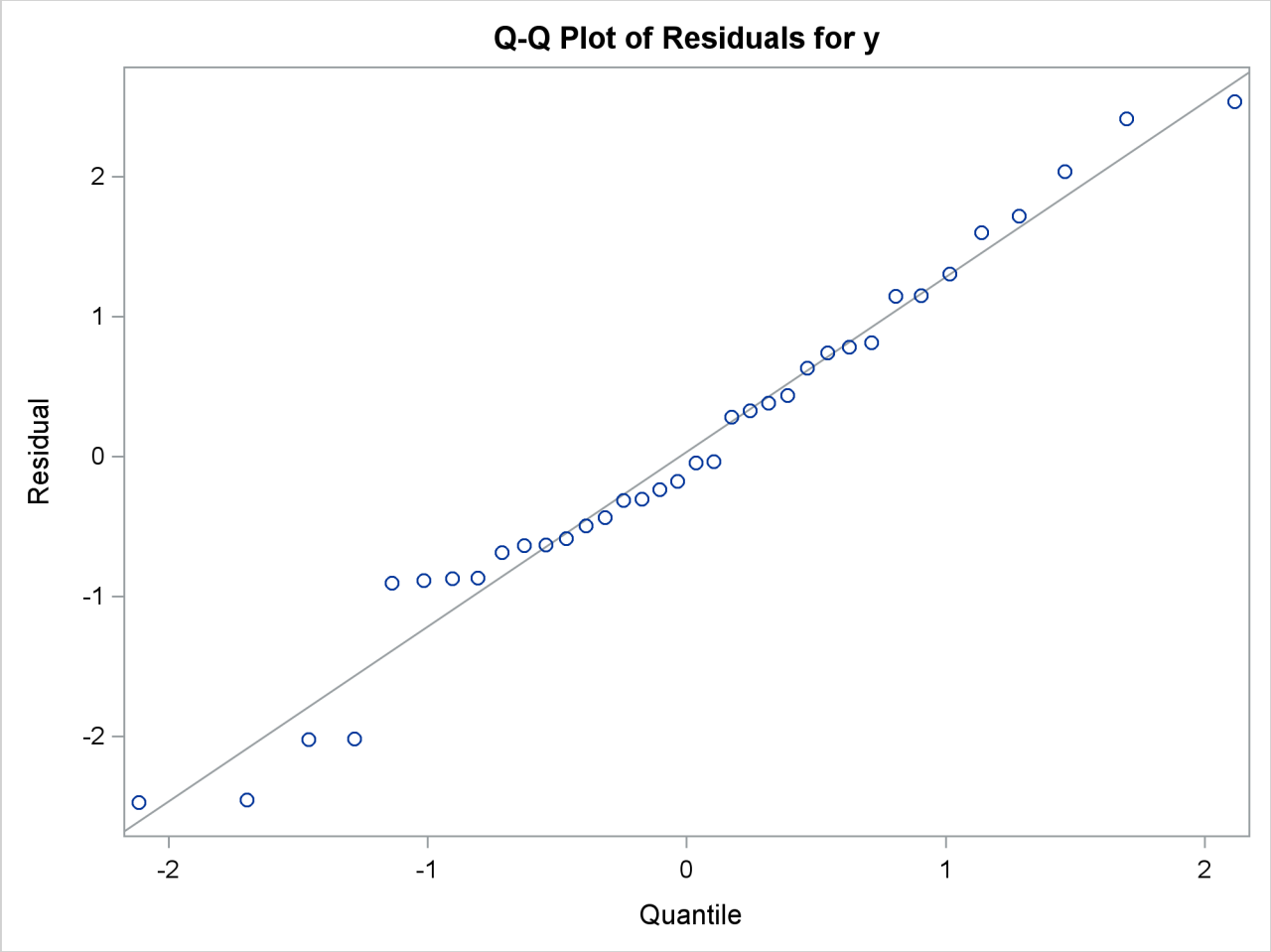
Output 8.8.5 Inverse Autocorrelation of Residuals Plot



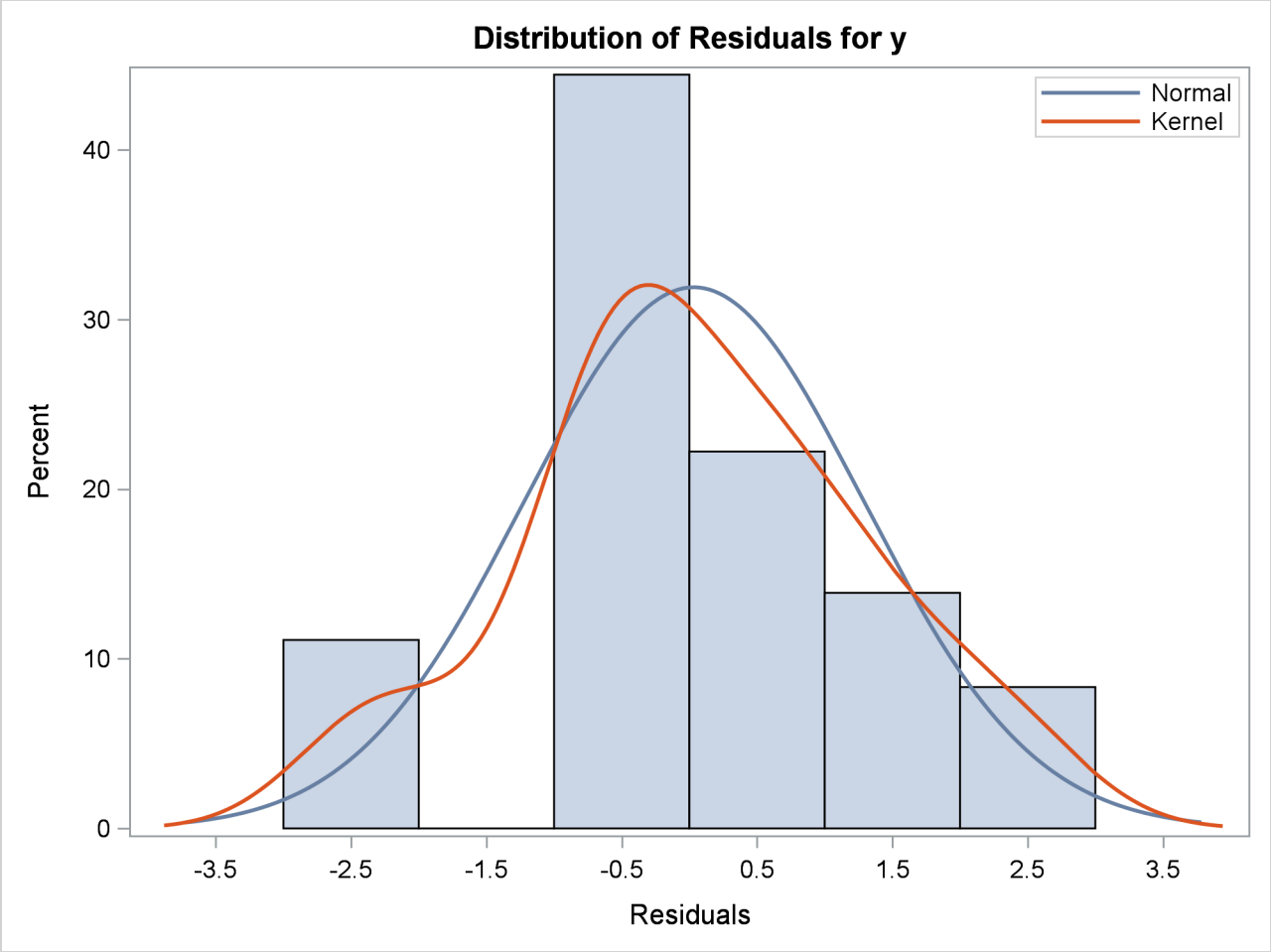
Output 8.8.6 Tests for White Noise Residuals Plot



Output 8.8.7 Q-Q Plot of Residuals



Output 8.8.8 Histogram of Residuals



References

Anderson, T. W., and Mentz, R. P. (1980). "On the Structure of the Likelihood Function of Autoregressive and Moving Average Models." *Journal of Time Series* 1:83–94.

Andrews, D. W. K. (1991). "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59:817–858.

Ansley, C. F., Kohn, R., and Shively, T. S. (1992). "Computing p-Values for the Generalized Durbin-Watson and Other Invariant Test Statistics." *Journal of Econometrics* 54:277–300.

Bai, J. (1997). "Estimation of a Change Point in Multiple Regression Models." *Review of Economics and Statistics* 79:551–563.

Bai, J., and Perron, P. (1998). "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66:47–78. http://www.jstor.org/stable/2998540.

Bai, J., and Perron, P. (2003a). "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics* 18:1–22. http://dx.doi.org/10.1002/jae.659.

Bai, J., and Perron, P. (2003b). "Critical Values for Multiple Structural Change Tests." *Econometrics Journal* 6:72–78. http://dx.doi.org/10.1111/1368-423X.00102.

Bai, J., and Perron, P. (2006). "Multiple Structural Change Models: A Simulation Analysis." In *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, edited by D. Corbae, S. N. Durlauf, and B. E. Hansen, 212–237. Cambridge: Cambridge University Press.

Baillie, R. T. (1979). "The Asymptotic Mean Squared Error of Multistep Prediction from the Regression Model with Autoregressive Errors." *Journal of the American Statistical Association* 74:175–184.

Baillie, R. T., and Bollerslev, T. (1992). "Prediction in Dynamic Models with Time-Dependent Conditional Variances." *Journal of Econometrics* 52:91–113.

Balke, N. S., and Gordon, R. J. (1986). "Historical Data." In *The American Business Cycle*, edited by R. J. Gordon, 781–850. Chicago: University of Chicago Press.

Bartels, R. (1982). "The Rank Version of von Neumann's Ratio Test for Randomness." *Journal of the American Statistical Association* 77:40–46.

Beach, C. M., and MacKinnon, J. G. (1978). "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors." *Econometrica* 46:51–58.

Bhargava, A. (1986). "On the Theory of Testing for Unit Roots in Observed Time Series." *Review of Economic Studies* 53:369–384.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31:307–327.

Bollerslev, T. (1987). "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *Review of Economics and Statistics* 69:542–547.

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Rev. ed. San Francisco: Holden-Day.

Breitung, J. (1995). "Modified Stationarity Tests with Improved Power in Small Samples." *Statistical Papers* 36:77–95.

Breusch, T. S., and Pagan, A. R. (1979). "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* 47:1287–1294.

Brock, W. A., Dechert, W. D., and Scheinkman, J. A. (1987). "A Test for Independence Based on the Correlation Dimension." Departments of Economics, University of Wisconsin–Madison, University of Houston, and University of Chicago.

Brock, W. A., Scheinkman, J. A., Dechert, W. D., and LeBaron, B. (1996). "A Test for Independence Based on the Correlation Dimension." *Econometric Reviews* 15:197–235.

Campbell, J. Y., and Perron, P. (1991). "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots." In *NBER Macroeconomics Annual*, edited by O. Blanchard and S. Fisher, 141–201. Cambridge, MA: MIT Press.

Caner, M., and Kilian, L. (2001). "Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate." *Journal of International Money and Finance* 20:639–657.

Chipman, J. S. (1979). "Efficiency of Least Squares Estimation of Linear Trend When Residuals Are Autocorrelated." *Econometrica* 47:115–128.

Chow, G. (1960). "Tests of Equality between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28:531–534.

Cochrane, D., and Orcutt, G. H. (1949). "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms." *Journal of the American Statistical Association* 44:32–61.

Cribari-Neto, F. (2004). "Asymptotic Inference under Heteroskedasticity of Unknown Form." *Computational Statistics and Data Analysis* 45:215–233.

Cromwell, J. B., Labys, W. C., and Terraza, M. (1994). *Univariate Tests for Time Series Models*. Thousand Oaks, CA: Sage Publications.

Davidson, R., and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Davies, R. B. (1973). "Numerical Inversion of a Characteristic Function." *Biometrika* 60:415–417.

DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992). "The Power Problems of Unit Root Rest in Time Series with Autoregressive Errors." *Journal of Econometrics* 53:323–343.

Dickey, D. A., and Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74:427–431.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). "A Long Memory Property of Stock Market Returns and a New Model." *Journal of Empirical Finance* 1:83–106.

Duffie, D. (1989). *Futures Markets*. Englewood Cliffs, NJ: Prentice-Hall.

Durbin, J. (1969). "Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least-Squares Residuals." *Biometrika* 56:1–15.

Durbin, J. (1970). "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables." *Econometrica* 38:410–421.

Edgerton, D., and Wells, C. (1994). "Critical Values for the CUSUMSQ Statistic in Medium and Large Sized Samples." *Oxford Bulletin of Economics and Statistics* 56:355–365.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). "Efficient Tests for an Autoregressive Unit Root." *Econometrica* 64:813–836.

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50:987–1007.

Engle, R. F., and Bollerslev, T. (1986). "Modelling the Persistence of Conditional Variances." *Econometric Reviews* 5:1–50.

Engle, R. F., and Granger, C. W. J. (1987). "Co-integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55:251–276.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). "Estimating Time Varying Risk in the Term Structure: The ARCH-M Model." *Econometrica* 55:391–407.

Engle, R. F., and Ng, V. K. (1993). "Measuring and Testing the Impact of News on Volatility." *Journal of Finance* 48:1749–1778.

Engle, R. F., and Yoo, B. S. (1987). "Forecasting and Testing in Co-integrated Systems." *Journal of Econometrics* 35:143–159.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: John Wiley & Sons.

Gallant, A. R., and Goebel, J. J. (1976). "Nonlinear Regression with Autoregressive Errors." *Journal of the American Statistical Association* 71:961–967.

Glosten, L., Jaganathan, R., and Runkle, D. (1993). "Relationship between the Expected Value and Volatility of the Nominal Excess Returns on Stocks." *Journal of Finance* 48:1779–1802.

Godfrey, L. G. (1978a). "Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables." *Econometrica* 46:1293–1301.

Godfrey, L. G. (1978b). "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables." *Econometrica* 46:1303–1310.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics*. Cambridge: Cambridge University Press.

Golub, G. H., and Van Loan, C. F. (1989). *Matrix Computations*. 2nd ed. Baltimore: Johns Hopkins University Press.

Greene, W. H. (1993). *Econometric Analysis*. 2nd ed. New York: Macmillan.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hannan, E. J., and Quinn, B. G. (1979). "The Determination of the Order of an Autoregression." *Journal of the Royal Statistical Society, Series B* 41:190–195.

Harvey, A. C. (1981). *The Econometric Analysis of Time Series*. New York: John Wiley & Sons.

Harvey, A. C. (1990). *The Econometric Analysis of Time Series*. 2nd ed. Cambridge, MA: MIT Press.

Harvey, A. C., and McAvinchey, I. D. (1978). "The Small Sample Efficiency of Two-Step Estimators in Regression Models with Autoregressive Disturbances." Discussion Paper No. 78-10, University of British Columbia.

Harvey, A. C., and Phillips, G. D. A. (1979). "Maximum Likelihood Estimation of Regression Models with Autoregressive–Moving Average Disturbances." *Biometrika* 66:49–58.

Hayashi, F. (2000). *Econometrics*. Princeton, NJ: Princeton University Press.

Hildreth, C., and Lu, J. Y. (1960). *Demand Relations with Autocorrelated Disturbances*. Technical Report 276, Michigan State University Agricultural Experiment Station.

Hobijn, B., Franses, P. H., and Ooms, M. (2004). "Generalization of the KPSS-Test for Stationarity." *Statistica Neerlandica* 58:483–502.

Hurvich, C. M., and Tsai, C.-L. (1989). "Regression and Time Series Model Selection in Small Samples." *Biometrika* 76:297–307.

Inder, B. A. (1984). "Finite-Sample Power of Tests for Autocorrelation in Models Containing Lagged Dependent Variables." *Economics Letters* 14:179–185.

Inder, B. A. (1986). "An Approximation to the Null Distribution of the Durbin-Watson Statistic in Models Containing Lagged Dependent Variables." *Econometric Theory* 2:413–428.

Jarque, C. M., and Bera, A. K. (1980). "Efficient Tests for Normality, Homoskedasticity, and Serial Independence of Regression Residuals." *Economics Letters* 6:255–259.

Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12:231–254.

Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59:1551–1580.

Johnston, J. (1972). *Econometric Methods*. 2nd ed. New York: McGraw-Hill.

Jones, R. H. (1980). "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations." *Technometrics* 22:389–396.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*. 2nd ed. New York: John Wiley & Sons.

Kanzler, L. (1999). "Very Fast and Correctly Sized Estimation of the BDS Statistic." Working paper, Department of Economics, Christ Church, University of Oxford.

King, M. L., and Wu, P. X. (1991). "Small-Disturbance Asymptotics and the Durbin-Watson and Related Tests in the Dynamic Regression Model." *Journal of Econometrics* 47:145–152.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root." *Journal of Econometrics* 54:159–178.

Lee, J. H., and King, M. L. (1993). "A Locally Most Mean Powerful Based Score Test for ARCH and GARCH Regression Disturbances." *Journal of Business and Economic Statistics* 11:17–27.

L'Esperance, W. L., Chall, D., and Taylor, D. (1976). "An Algorithm for Determining the Distribution Function of the Durbin-Watson Test Statistic." *Econometrica* 44:1325–1326.

MacKinnon, J. G. (1991). "Critical Values for Cointegration Tests." In *Long-Run Economic Relationships: Readings in Cointegration*, edited by R. F. Engle and C. W. J. Granger, 267–276. Oxford: Oxford University Press.

Maddala, G. S. (1977). *Econometrics*. New York: McGraw-Hill.

Maeshiro, A. (1976). "Autoregressive Transformation, Trended Independent Variables and Autocorrelated Disturbance Terms." *Review of Economics and Statistics* 58:497–500.

McLeod, A. I., and Li, W. K. (1983). "Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations." *Journal of Time Series Analysis* 4:269–273.

Nelson, C. R., and Plosser, C. I. (1982). "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics* 10:139–162.

Nelson, D. B. (1990). "Stationarity and Persistence in the GARCH(1,1) Model." *Econometric Theory* 6:318–334.

Nelson, D. B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica* 59:347–370.

Nelson, D. B., and Cao, C. Q. (1992). "Inequality Constraints in the Univariate GARCH Model." *Journal of Business and Economic Statistics* 10:229–235.

Newey, W. K., and West, D. W. (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55:703–708.

Newey, W. K., and West, D. W. (1994). "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies* 61:631–653.

Ng, S., and Perron, P. (2001). "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power." *Econometrica* 69:1519–1554.

Park, R. E., and Mitchell, B. M. (1980). "Estimating the Autocorrelated Error Model with Trended Data." *Journal of Econometrics* 13:185–201.

Phillips, P. C. B. (1987). "Time Series Regression with a Unit Root." *Econometrica* 55:277–301.

Phillips, P. C. B., and Ouliaris, S. (1990). "Asymptotic Properties of Residual Based Tests for Cointegration." *Econometrica* 58:165–193.

Phillips, P. C. B., and Perron, P. (1988). "Testing for a Unit Root in Time Series Regression." *Biometrika* 75:335–346.

Prais, S. J., and Winsten, C. B. (1954). "Trend Estimators and Serial Correlation." Cowles Commission Discussion Paper No. 383.

Ramsey, J. B. (1969). "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis." *Journal of the Royal Statistical Society, Series B* 31:350–371.

Said, S. E., and Dickey, D. A. (1984). "Testing for Unit Roots in ARMA Models of Unknown Order." *Biometrika* 71:599–607.

Savin, N. E., and White, K. J. (1978). "Testing for Autocorrelation with Missing Observations." *Econometrica* 46:59–67.

Schwarz, G. (1978). "Estimating the Dimension of a Model." *Annals of Statistics* 6:461–464.

Schwert, G. (1989). "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business and Economic Statistics* 7:147–159.

Shin, Y. (1994). "A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration." *Econometric Theory* 10:91–115. http://dx.doi.org/10.1017/S0266466600008240.

Shively, T. S. (1990). "Fast Evaluation of the Distribution of the Durbin-Watson and Other Invariant Test Statistics in Time Series Regression." *Journal of the American Statistical Association* 85:676–685.

Spitzer, J. J. (1979). "Small-Sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimators in the Context of Autocorrelated Errors." *Journal of the American Statistical Association* 74:41–47.

Stock, J. H. (1994). "Unit Roots, Structural Breaks, and Trends." In *Handbook of Econometrics*, edited by R. F. Engle and D. L. McFadden, 2739–2841. Amsterdam: North-Holland.

Stock, J. H., and Watson, M. W. (1988). "Testing for Common Trends." *Journal of the American Statistical Association* 83:1097–1107.

Stock, J. H., and Watson, M. W. (2002). *Introduction to Econometrics*. 3rd ed. Reading, MA: Addison-Wesley.

Theil, H. (1971). *Principles of Econometrics*. New York: John Wiley & Sons.

Vinod, H. D. (1973). "Generalization of the Durbin-Watson Statistic for Higher Order Autoregressive Process." *Communications in Statistics* 2:115–144.

Wallis, K. F. (1972). "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations." *Econometrica* 40:617–636.

White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48:817–838.

White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50:1–25.

White, K. J. (1992). "The Durbin-Watson Test for Autocorrelation in Nonlinear Models." *Review of Economics and Statistics* 74:370–373.

Wong, H., and Li, W. K. (1995). "Portmanteau Test for Conditional Heteroscedasticity, Using Ranks of Squared Residuals." *Journal of Applied Statistics* 22:121–134.

Zakoian, J. M. (1994). "Threshold Heteroscedastic Models." *Journal of Economic Dynamics and Control* 18:931–955.

Subject Index

Syntax Index