Paper 338-2012

Weighted Portmanteau Tests Revisited: Detecting Heteroscedasticity, Fitting Nonlinear and Multivariate Time Series

Thomas J. Fisher, Department of Mathematics & Statistics, University of Missouri-Kansas City, Kansas City, MO, 64110, USA

ABSTRACT

In the 2011 SAS® Global Forum, two weighted portmanteau tests were introduced for goodness-of-fit of an Autoregressive-Moving Average (ARMA) time series process. This result is summarized and extended for use as a diagnostic tool in detecting nonlinear and variance-changing processes such as the Generalized Autoregressive Conditional Heteroscedasticity process. The efficacy of the weighting scheme is shown in simulation experiments and analysis of stock market data. The statistics are easy to implement in SAS® and source code is provided. Lastly, the versatility of this methodology is discussed for a fitted GARCH, Vector ARMA, and other time series processes.

INTRODUCTION

After fitting a regression model with respect to time, or modeling the trend and seasonality of a time series, a plethora of situations are known to occur in which the residual error terms of the model are not independent. This violates the underlying assumptions of most statistical testing and modeling and can lead to a multitude of problems. Modeling the serial-correlation of the errors does not address all of the issues but allows us to have better forecast (predictive models). However, much like checking the adequacy of the regression through an F-test, checking the adequacy of the fitted time-series model is of the utmost importance.

Let $\{X_t\}$ be an observed time series for $t = 1, \dots, n$ where n is the number of observations. Suppose $\{X_t\}$ is generated by a stationary and invertible ARMA(p, q) process of the form

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \qquad (1)$$

where $\varepsilon \sim N(0, \sigma^2)$ are white-noise residuals. A model of this form is typically fitted with the autoregressive and moving average parameters, φ_i and θ_j, respectively, estimated by their maximum likelihood or conditional least squares counterparts, $\hat{\varphi}_i$ and $\hat{\theta}_j$. After we have fit the model for given orders p and q, testing for the adequacy of the fitted model follows:

$$H_0: \text{Model is Adequate} \quad vs \quad H_A: \text{Model is Inadequate}$$

Most diagnostic goodness-of-fit procedures are based on the residual autocorrelation coefficients (i.e., the observed serial correlation) provided by

$$r_k = \frac{\sum_{t=k+1}^{n} \varepsilon_t \varepsilon_{t-k}}{\sum_{t=1}^{n} \varepsilon_t^2} \qquad (2)$$

where ε_t is the residual at time t. Each r_k essentially measures the amount of correlation in the series between some time t and time $t - k$ where k is called the lag. The hypothesis test can be written:

$$H_0: r_k = 0 \text{ for all } k = 1, \dots, m \quad vs \quad H_A: r_k \neq 0 \text{ for some } k = 1, \dots, m$$

where m is the largest lag considered for autocorrelation. If we correctly identified, or possibly overestimated, the correct orders of the ARMA process, the value of each autocorrelation should be approximately zero. However, if we underestimate the ARMA model some of the values of the autocorrelations will deviate from zero towards ±1. Of course in practice the autocorrelation coefficients are unknown and must be estimated based on the sample residuals,

$$\hat{\varepsilon}_t = X_t - \left(\sum_{i=1}^{p} \hat{\varphi}_i X_{t-i} - \sum_{j=1}^{q} \hat{\theta}_j \hat{\varepsilon}_{t-j} \right)$$

and $\hat{r}_k$ can be found by using the above equation (2) with ε_t replaced by $\hat{\varepsilon}_t$.

338-2012 Weighted Portmanteau Test Revisited, continued

PORTMANTEAU TEST

The first widely used testing method based on the autocorrelation coefficients is the Box-Pierce (1970) statistic, provided by

$$Q = n \sum_{k=1}^{m} \hat{r}_k^2.$$

In most modern applications, it has been replaced by the Ljung-Box (1978) statistic

$$\tilde{Q} = n(n+2) \sum_{k=1}^{m} \frac{\hat{r}_k^2}{n-k}$$

that includes the standardizing term $\frac{n+2}{n-k}$ on each squared autocorrelation coefficient. These statistics are used to test for significant correlation up to lag m. It is well known that for independent and identically distributed residuals, as $n \to \infty$ the autocorrelations behave as independent jointly normal distributed random variables, and therefore under the null hypothesis (correctly fitted model) both Q and $\tilde{Q}$ are shown to be asymptotically distributed chi-squared random variables with $m - (p+q)$ degrees of freedom, where p and q are the order of autoregressive and moving average terms estimated in the fitted model, respectively. If Q or $\tilde{Q}$ are large, compared to the chi-squared critical value at significance level α, we have evidence to suggest the fitted ARMA process does not adequately model the correlation in the data.

The Ljung-Box statistic is provided in the SAS procedure **PROC ARIMA** for an assortment of lags m. For large n, the Box-Pierce and Ljung-Box statistics are essentially equivalent. The Ljung-Box (1978) statistic is typically used since it better approximates a chi-squared random variable for smaller n.

A similar statistic to the Ljung-Box statistic was introduced by Monti (1994) that uses the standardized partial autocorrelation function up to lag m:

$$\tilde{M} = n(n+2) \sum_{k=1}^{m} \frac{\hat{\pi}_k^2}{n-k}$$

where $\hat{\pi}_k$ is the residual partial autocorrelation at lag k. Recently, Peña and Rodríguez (2002) proposed a statistic based on the determinant of the residual autocorrelation matrix:

$$\hat{R}_m = \begin{pmatrix} 1 & \hat{r}_1 & \cdots & \hat{r}_m \\ \hat{r}_1 & 1 & \cdots & \hat{r}_{m-1} \\ \vdots & \cdots & \ddots & \vdots \\ \hat{r}_m & \cdots & \hat{r}_1 & 1 \end{pmatrix}.$$

Under the null hypothesis that we have fitted an adequate model for the ARMA process, each $\hat{r}_k \approx 0$. Hence the matrix $\hat{R}_m$ should be approximately the identity matrix. Testing for model adequacy is equivalent to testing if $\hat{R}_m$ is approximately the identity matrix. They show

$$D = n \left(1 - \left|\hat{R}_m\right|^{1/m}\right)$$

is asymptotically distributed as a linear combination of chi-squared random variables and is approximately a Gamma distributed random variable for large values of m. In practice, they recommend the matrix $\hat{R}_m$ be constructed using the standardized autocorrelations as this improves the Gamma distribution approximation. This essentially includes the $\frac{n+2}{n-k}$ terms from the Ljung-Box and Monti statistics and creates a matrix of the form $\hat{R}_m$. In Peña and Rodríguez (2006) they show that the log of the determinant follows the same asymptotic distribution as D and can be better in small sample time series. The statistic D determines whether the matrix $\hat{R}_m$ is an identity matrix, or equivalent, if the fitted model is adequate, using methodology similar to the Likelihood Ratio Test in multivariate analysis.

It has been demonstrated that both $\tilde{M}$ and D improve over the Ljung-Box and Box-Pierce statistics in some situations; see Monti (1994) or Peña and Rodríguez (2002, 2006). However, neither appears to be frequently implemented in applications of time series. Particularly, the Peña and Rodríguez statistic may be difficult to implement since it involves calculating the determinant of a matrix. As pointed out in Lin and McLeod (2006), the statistic D constructed using the standardized residual autocorrelations may be degenerate in practice since the matrix could be ill-conditioned or singular. They, along with Mahdi and McLeod (2011), recommend a Monte Carlo method to determine the distributional properties of the statistic. This recommended Monte Carlo method improves performance but creates additional computational issues, and for the typical SAS user, would require the need for IML and some familiarity with Monte Carlo simulations.

WEIGHTED PORTMANTEAU TEST

At the 2011 SAS Global Forum, the contributed paper by Fisher (2011) outlined two new statistics that are easy to implement and improve over the frequently used Ljung-Box and Box-Pierce statistics. The test statistics are defined in Fisher and Gallagher (2012) as

$$\tilde{Q}_W = n(n+2) \sum_{k=1}^{m} \frac{(m-k+1)}{m} \frac{\hat{r}_k^2}{n-k} \tag{3}$$

and

$$\widetilde{M}_W = n(n+2) \sum_{k=1}^{m} \frac{(m-k+1)}{m} \frac{\hat{\pi}_k^2}{n-k}. \tag{4}$$

The two statistics look similar to the Ljung-Box and Monti statistics with the exception a weight, $\frac{m-k+1}{m}$, on each sample squared residual autocorrelation or squared partial autocorrelation. The weights are derived using multivariate analysis techniques on the matrix of autocorrelations or matrix of partial autocorrelations (similar to that in Peña and Rodríguez). Note that the sample autocorrelation at lag 1, $\hat{r}_1$, is given weight $\frac{m}{m} = 1$. The sample autocorrelation at lag 2, $\hat{r}_2$, is given weight $\frac{m-1}{m} < 1$. We can interpret the weights as putting more emphasis on the first autocorrelation, and the least emphasis on the autocorrelation at lag m (corresponding weight $\frac{1}{m}$). This matches typical intuition about statistical estimators. The first autocorrelation $\hat{r}_1$ is calculated using information from all n observations. The second autocorrelation is based on $n-1$ observations, and the mth autocorrelation is based on $n-m$ observations. Intuitively, it makes sense to put more emphasis on the first autocorrelation as it should be the most accurate. This idea also holds true for the partial autocorrelations.

The two statistics are asymptotically distributed as a linear combination of chi-squared random variables. This is the same asymptotic distribution as the statistics in Peña and Rodríguez (2002, 2006) and similar to that in Mahdi and McLeod (2012). The Weighted Ljung-Box $\tilde{Q}_W$ and Weighted Monti $\widetilde{M}_W$ statistics are asymptotically equivalent to D but have the added benefit of easy calculation and computational stability. When a small number of parameters have been fit under the null hypothesis of an adequate model, the statistics $\tilde{Q}_W$ and $\widetilde{M}_W$ are approximately distributed as Gamma random variables with shape parameter

$$\gamma = \frac{3}{4} \frac{m(m+1)^2}{2m^2 + 3m + 1 - 6m(p+q)}$$

and scale parameter

$$\lambda = \frac{2}{3} \frac{2m^2 + 3m + 1 - 6m(p+q)}{m(m+1)}.$$

The Gamma approximation is constructed to have a similar theoretical mean and variance as the true asymptotic distribution but are chosen to give the statistics conservative Type I error performance in practice.

The 2011 SAS Global Forum contributed article by Fisher (2011) includes a simulation study showing the effectiveness of the weighted portmanteau statistics in detecting an underfit ARMA model. Fisher and Gallagher (2012) provide a larger study that shows, in general, the statistics $\tilde{Q}_W$ and $\widetilde{M}_W$ have conservative Type I error performance and are comparable to the methods of in Peña and Rodríguez (2002, 2006) and Mahdi and McLeod (2012). All of which tend to be more powerful than the traditionally used Ljung and Box (1978) statistics. Furthermore, studies in the literature generally suggest the statistics based on the partial autocorrelation coefficients (Monti and Weighted Monti) are better at detecting under fit Moving-Average while those based on the autocorrelation values are better at detecting under fit Autoregressive processes.

Implementation in SAS

The two statistics, $\tilde{Q}_W$ and $\widetilde{M}_W$, require no difficult matrix calculations and are easy to implement in SAS. Although their asymptotic distribution is difficult to express, an easy approximation is possible with the Gamma distribution. Source code is provided below that calculates both, $\tilde{Q}_W$ and $\widetilde{M}_W$, in addition to the Ljung-Box and Monti statistics, for a given m after $p+q$ (the variable `pq`) parameters were fit. In the example we analyze the monthly log returns of IBM stock from 1926 through 1999; that is, $X_t = \ln Y_t - \ln Y_{t-1}$ where Y_t is the monthly closing price of the stock. The series is displayed in Figure 1.

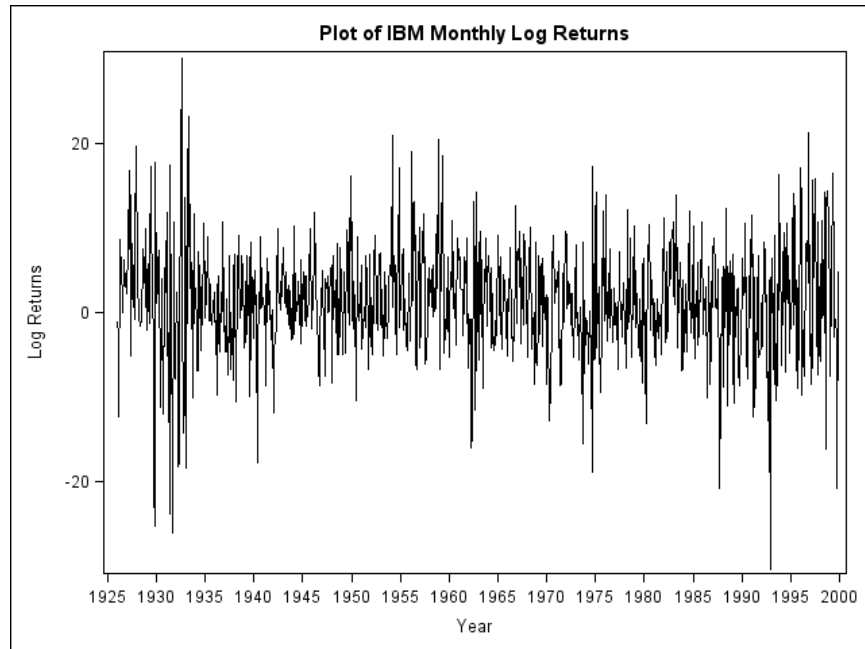338-2012 Weighted Portmanteau Test Revisited, continued

Figure 1. IBM Monthly Returns

This series was chosen as a representative example of stock market returns. We are simply interested in modeling the serial correlation of the time series. Tsay (2010) analyzes the data in his book and it can be obtained from his website (http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts/) or from almost any financial site, such as *Yahoo! Finance* or *Google Finance*. By quickly examining the correlogram in Figure 2, we see some minor-indication of the presence of serial correlation, hence an ARMA model may be appropriate. Furthermore the Ljung-Box (p-value 0.018) and Weighted Ljung-Box (p-value 0.045) test statistics in Output 1 suggest the series is not white noise; that is, it contains some serial correlation.
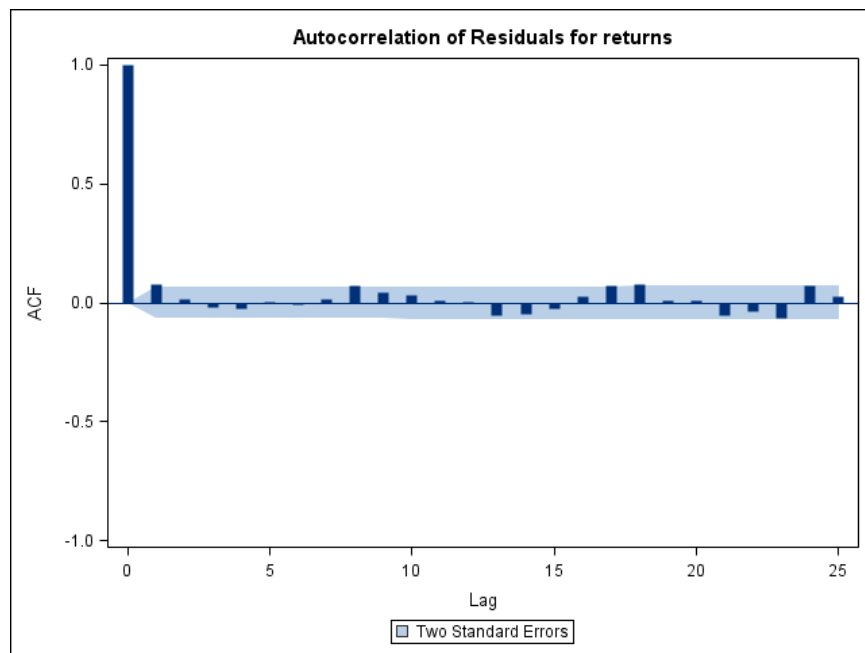

Figure 2. Correlogram for IBM Monthly Returns

Output 1 shows the results of the final **PROC PRINT** statement in the source code below when no ARMA model is fit; i.e. $p + q = 0$ and the autocorrelations (and partial autocorrelations) are found on the original series. This particular source code is not provided but is analogous to the results below.

338-2012 Weighted Portmanteau Test Revisited, continued

Obs	m	deg Free	QLB	pValQLB	Mon	pValMon	WQLB	pVal WQLB	WM	pValWM
31	30	30	48.3872	0.018131	39.2462	0.12031	24.0347	0.045030	20.1296	0.15374

Output 1. Output from the final PROC PRINT Statement when no ARMA model is fit

Since the two statistics based on the autocorrelations reject the null hypothesis, we attempt to fit an autoregressive model. This follows the work of Tsay (2010) where an ARMA(1,0), or AR(1), model is fit on the monthly log returns series we call `IBMRET`; we then test the models adequacy at lag $m = 30$, where $p + q = 1$ parameters were estimated in the model. The following code will calculate the Ljung-Box, Monti, Weighted Ljung-Box and Weighted Monti statistics and their associated p-values:

```
PROC ARIMA DATA=IBMRET;
  IDENTIFY VAR=returns NLAG=30 NOPRINT;
  ESTIMATE p=1 q=0 METHOD=ML;
  FORECAST OUT=getResids;

PROC ARIMA DATA=getResids;
  IDENTIFY VAR=RESIDUAL NLAG=30 OUTCOV=acfs;

DATA acfs;
  SET acfs;
  IF lag = 0 THEN r=0;
    ELSE r = corr*corr;
  IF lag = 0 THEN pr=0;
    ELSE pr = partcorr*partcorr;
  rLB = r/N;
  prM = pr/N;
  sampSize = N+lag;

  QtmpLB + rLb;
  QtmpLB = QtmpLB;
  QLB = QtmpLB*sampSize*(sampSize+2);
  Mtmp + prM;
  Mon = Mtmp*sampSize*(sampSize+2);
  degFree = lag - 1;
  m = 30;
  pq = 1;
  shape = (3/4)*m*(m+1)*(m+1)/(2*m*m + 3*m + 1 -6*m*pq);
  scale = (2/3)*(2*m*m + 3*m + 1 - 6*m*pq)/(m*(m+1));
  weights = (m - lag + 1)/(m);
  WQLBtmp + (weights*rLb);
  WQLBtmp = WQLBtmp;
  WQLB = WQLBtmp*sampSize*(sampSize+2);
  WMtmp + (weights*prM);
  WMtmp = WMtmp;
  WM = WMtmp*sampSize*(sampSize+2);
  pValQLB = SDF('chisquared',QLB,degFree);
  pValMon = SDF('chisquared',Mon,degFree);
  pValWQLB = SDF('gamma',WQLB, shape, scale);
  pValWM = SDF('gamma',WM, shape, scale);

PROC PRINT DATA=acfs(where=(lag=30));
  VAR m degFree QLB pValQLB Mon pValMon WQLB pValWQLB WM pValWM;
RUN;
```

Output 2 shows the results of the final PRINT statement above where an ARMA(1,0) is fit to the series.

338-2012 Weighted Portmanteau Test Revisited, continued

Obs	m	deg Free	QLB	pValQLB	Mon	pValMon	WQLB	pVal WQLB	WM	pValWM
31	30	29	39.4493	0.093351	34.0854	0.23615	16.5881	0.36793	14.9842	0.51015

Output 2. Output from the final PROC PRINT Statement

Notice the similarities in the two output statements, but the degrees of freedom have changed and now we fail to reject using all test statistics. Based on this analysis it appears an AR(1) is an adequate model in capturing the serial correlation in the IBM Monthly Log return series. This is typical of many stock market return series, an AR(1) model is generally adequate, although sometimes an ARMA(1,1) may be needed or no ARMA model at all. In our case, SAS has estimated the ARMA(1,0) model as $X_t = 1.24 + 0.076 X_{t-1}$, which is very similar to the estimated model in Tsay (2010). The discrepancy is strictly due to the optimization routine in R versus SAS. We note that the estimated autoregressive parameter is quite small (0.076), which corresponds to correlogram in Figure 2 that shows hardly any significant serial correlation. This also supports the results of the test statistics in Output 1; significant evidence of serial correlation is found, but not strongly significant.

NONLINEAR PROCESSES

In many applications of time series, the effectiveness of the ARMA model suggested in (1) may be limited and a nonlinear model may be preferred. For example, the residual errors may be uncorrelated but not necessarily independent. A nonlinear function of the residuals may explain the dependence structure. In particular we consider a model for the residual error terms of the form

$$\varepsilon_t = g(h_t)\eta_t \tag{5}$$

where the η_t terms are independent and identically distributed with mean 0 and variance 1, the series h_t follows an ARMA type process, see (1), and the function $g(\cdot)$ is nonlinear. Typical examples include $g(x) = \sqrt{x}$, which is the case of the Generalized Autoregressive Conditional Heteroscedasticity (ARCH and GARCH) process of Engle (1982) and Bollerslev (1986), and $g(x) = exp\{x\}$, which is the Stochastic Volatility (SV) model of Taylor (1986). The concepts of these models were discussed at the 2011 SAS Global Forum by Dickey (2011). This leads to two problems of interest: 1. How to detect a nonlinear process and 2. After modeling a nonlinear process can we determine the goodness-of-fit of the model?

DETECTING NONLINEAR PROCESSES

Many authors have studied the problem of detecting a nonlinear residual process by transforming the sample residuals of a fitted ARMA model. That is, consider transforming $\hat{\varepsilon}_t$ by either taking squares, absolute values or the logarithm of squares. McLeod and Li (1983) provide the first diagnostic tool for detecting nonlinear processes utilizing this method. Consider the autocorrelation function constructed from the squared sample residuals

$$r_k(\varepsilon^2) = \frac{\sum_{t=k+1}^{n}(\varepsilon_t^2 - \bar{\varepsilon}^2)(\varepsilon_{t-k}^2 - \bar{\varepsilon}^2)}{\sum_{t=1}^{n}(\varepsilon_t^2 - \bar{\varepsilon}^2)^2} \tag{6}$$

where $\bar{\varepsilon}^2 = \sum \varepsilon_t^2 / n$ is the average of the squared residuals. McLeod and Li (1983) suggest a Ljung-Box type statistic where the autocorrelations are replaced by (6). By a similar result involving the partial autocorrelation coefficients, the statistics of Monti (1994), Peña and Rodríguez (2002, 2006) and Mahdi and McLeod (2012) can be adapted to utilize the squared residuals as well. McLeod and Li (1983) showed that the asymptotic distribution of $r_k(\varepsilon^2)$ does not depend on the number of ARMA parameters fit. To a practitioner of statistics, this simplifies the distribution under the null hypothesis and makes implementation easier. Likewise the statistics can be built utilizing absolute and logarithms of squared residuals. Simulation studies in the literature suggest that, generally, those based on squared and absolute residuals are the most effective.

WEIGHTED PORTMANTEAU STATISTICS FOR DETECTING NONLINEAR PROCESSES

The weighted statistics $\tilde{Q}_W$ and $\widetilde{M}_W$ can easily be adapted to detect nonlinear processes. Simply by replacing $\hat{r}_k^2$ in equation (3) with $\hat{r}_k^2(\hat{\varepsilon}^2)$ or $\hat{r}_k^2(|\hat{\varepsilon}|)$ and $\hat{\pi}_k^2$ in equation (4) with $\hat{\pi}_k^2(\hat{\varepsilon}^2)$ or $\hat{\pi}_k^2(|\hat{\varepsilon}|)$ we have two statistics that are effective in detecting nonlinear processes. Fisher and Gallagher (2012) show the statistics are asymptotically distributed as a linear combination of chi-square random variables and that the distribution can be well approximated by a Gamma distribution with shape and scale parameters

$$\gamma = \frac{3}{4}\frac{m(m+1)}{2m+1}$$

and

338-2012 Weighted Portmanteau Test Revisited, continued

$$\lambda = \frac{2}{3}\frac{2m+1}{m}$$

respectively. Note this distribution does not depend on the number of ARMA parameters fit to the model. This simplifies the implementation of the statistic.

Simulation Study

Fisher and Gallagher (2012) provide a thorough simulation study showing the weighted test statistics are very effective in detecting nonlinear models. In particular, they appear to be the most powerful in detecting long-memory processes. Here, we provide a brief simulation study showing the effectiveness of the weighted test statistics.

We consider a time series generated from a GARCH process with long persistence. The GARCH(q, p) process is defined using equation (5) with

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j}^{p} \beta_j h_{t-j} \tag{7}$$

and $g(x) = \sqrt{x}$. Note the similarities between equations (1) and (7). Two examples are considered, GF-1: $a_0 = 1, \alpha_1 = 0.05$, $\beta_1 = 0.90$ and GF-2: $a_0 = 1, \alpha_1 = 0.15$, $\beta_1 = 0.80$ with a sample of size 250. 1000 replicates are generated and the test statistics for detecting nonlinear models are calculated, the power is reported as the number of times (out of the 1000) the test statistic detects the nonlinear process. We see in Table 1 that the Weighted McLeod Li (Ljung-Box) type test tends to be the most powerful. It also appears the test based on $\hat{\varepsilon}^2$ are more effective than those based on $|\hat{\varepsilon}|$ but there are many cases where the opposite is true.

Empirical Power at $m = 12$					
Model	$r_k(\cdot)$	$\tilde{Q}_W$	$\widetilde{M}_W$	$\tilde{Q}$	$\widetilde{M}$
GF-1	ε^2	**0.317**	0.280	0.285	0.251
---	$|\varepsilon|$	**0.246**	0.202	0.230	0.161
GF-2	ε^2	**0.877**	0.844	0.847	0.794
---	$|\varepsilon|$	**0.846**	0.802	0.813	0.744

Table 1. Power Level of McLeod Li, Monti and Weighted Portmanteau Tests

Implementation in SAS

The implementation in SAS is straight forward and follows the previous example. After fitting an ARMA process we can look for nonlinear terms in the residuals by utilizing the square or absolute residuals. We simply must add a data step in between the two PROC ARIMA statements and adjust the shape and scale parameters in the data step that calculates the test statistics. In this code we calculate the test statistics based on the squared residuals. The code can easily be modified to utilize the absolute residuals. We explore the correlogram of the autocorrelations based on the squared residuals. First we fit an ARMA(1,0) model to the IBM return series, then we find and plot the autocorrelations based on the square of the residuals of the fitted ARMA series.

```
PROC ARIMA DATA=IBMRET;
   IDENTIFY VAR=returns NLAG=30 NOPRINT;
   ESTIMATE p=1 q=0 METHOD=ML;
   FORECAST OUT=getResids;

DATA getResids;
   SET getResids;
   SET IBMRET;
   date=date;
   SqrResiduals = RESIDUAL*RESIDUAL;

PROC AUTOREG DATA=getResids PLOTS(unpackpanel only) = ACFPLOT;
   MODEL SqrResiduals = date;
```

338-2012 Weighted Portmanteau Test Revisited, continued

As you can see in Figure 3, there is clear evidence that there is serial correlation in the squared residuals of the fitted ARMA(1,0) process. A similar phenomenon can be seen utilizing the absolute residuals.
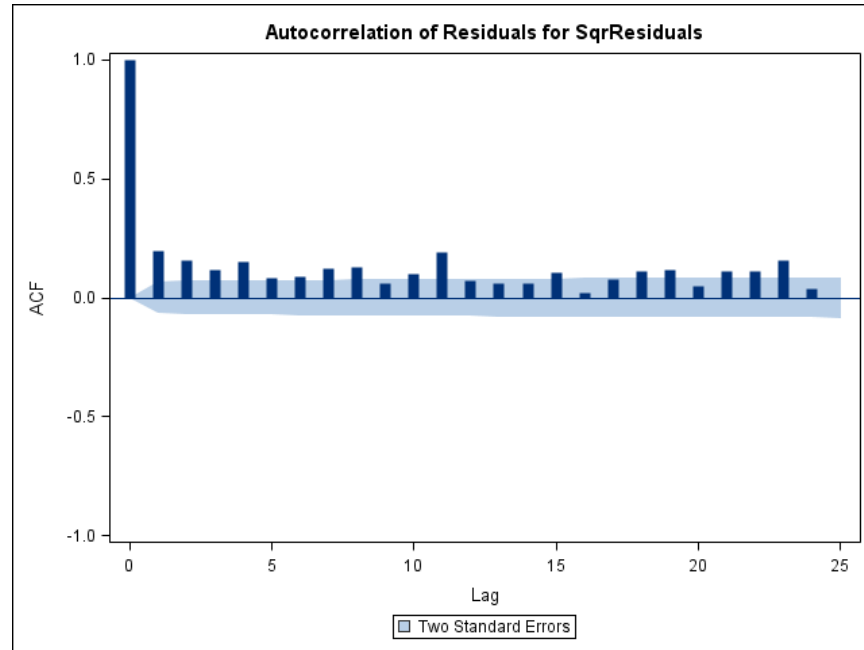

Figure 3. Correlogram for Squared Residuals

The correlogram in Figure 3 strongly suggests autocorrelation in the squared residuals. However, as pracitioners of statistics a procedural method based on hypothsis testing is preferred. The source code below will calculate the McLeod Li (Ljung-Box type), Monti, and Weighted McLeod Li and Weighted Monti type statistics at lag $m = 30$ to detect nonlinear processes.

```
PROC ARIMA DATA=getResids;
  IDENTIFY VAR=SqrResiduals NLAG=30 OUTCOV=acfsSqr;

DATA acfsSqr;
  SET acfsSqr;
  IF lag = 0 THEN r=0;
    ELSE r = corr*corr;
  IF lag = 0 THEN pr=0;
    ELSE pr = partcorr*partcorr;
  rLB = r/N;
  prM = pr/N;
  sampSize = N+lag;

  QtmpLB + rLb;
  QtmpLB = QtmpLB;
  QLB = QtmpLB*sampSize*(sampSize+2);
  Mtmp + prM;
  Mon = Mtmp*sampSize*(sampSize+2);
  degFree = lag;
  m = 30;
  shape = (3/4)*m*(m+1)/(2*m + 1);
  scale = (2/3)*(2*m + 1)/m;
  weights = (m - lag + 1)/(m);
  WQLBtmp + (weights*rLb);
  WQLBtmp = WQLBtmp;
  WQLB = WQLBtmp*sampSize*(sampSize+2);
```

338-2012 Weighted Portmanteau Test Revisited, continued

```
    WMtmp + (weights*prM);
    WMtmp = WMtmp;
    WM = WMtmp*sampSize*(sampSize+2);
    pValQLB = SDF('chisquared',QLB,degFree);
    pValMon = SDF('chisquared',Mon,degFree);
    pValWQLB = SDF('gamma',WQLB, shape, scale);
    pValWM = SDF('gamma',WM, shape, scale);

  PROC PRINT DATA=acfsSqr(where=(lag=30));
    VAR m degFree QLB pValQLB Mon pValMon WQLB pValWQLB WM pValWM;
  RUN;
```

Output 3 shows the results of the final PRINT statement above suggesting that the AR(1) model is not fully adequate. All of the statistics strongly suggest that a nonlinear process is necessary in fully modeling the correlation structure in the IBM series.

Obs	m	deg Free	QLB	pValQLB	Mon	pValMon	WQLB	pValWQLB	WM	pValWM
31	30	30	282.904	6.042E-43	132.768	6.9017E-15	188.205	1.2156E-45	94.0713	1.3704E-18

Output 3. Output from the final PROC PRINT Statement

ARCH GOODNESS-OF-FIT

Much like checking the fit of an ARMA process, after fitting a GARCH or ARCH process in equation (7), a practitioner should check the adequacy of that model. Li and Mak (1994) derive the distribution of the autocorrelation function developed from the standardized squared residuals, which are $\hat{r}_k(\hat{\varepsilon}_t^2/\hat{h}_t)$ using a slight modification to equation (6). Using some fundamental results about quadratic forms from multivariate analysis, Li and Mak suggest a statistic to check for the adequacy of a fitted GARCH process. Furthermore, they show under the null hypothesis of an adequately fitted ARCH model, the test statistic simplifies into a Ljung-Box type statistc

$$L(q,m) = n \sum_{k=q+1}^{m} \hat{r}_k(\hat{\varepsilon}_t^2/\hat{h}_t). \tag{8}$$

The weighting scheme utilized in the previous sections can be adapted for testing the adequacy of a fitted GARCH or ARCH process. To keep the results of this paper straight forward we will only concentrate on the fitted ARCH process. The test statistics for a fitted GARCH process can be implemented using IML in SAS and the technical of its derivation details are available in Li and Mak (1994) and Fisher and Gallagher (2012). For a fitted ARCH*(q)* process, Fisher and Gallagher (2012) suggest the Weighted Li Mak type test

$$L_W(q,m) = n \sum_{k=q+1}^{m} \frac{m-k+q+1}{m} \hat{r}_k(\hat{\varepsilon}_t^2/\hat{h}_t). \tag{9}$$

They show it is distributed as a linear combination of chi squared random variables. They recommend its distribution be approximated by a Gamma distribution with respective shape and scale parameters,

$$\gamma = \frac{3}{4} \frac{(m+q+1)^2(m-q)}{2m^2+3m+2mq+2q^2+3q+1}$$

and

$$\lambda = \frac{2}{3} \frac{2m^2+3m+2mq+2q^2+3q+1}{m(m+q+1)}.$$

Simulation Studies

The studies in Fisher and Gallagher (2012) show that both statistics, see equations (8) and (9), for a fitted ARCH model have conservative Type I performance. Fisher and Gallagher (2012) demonstrate that generally the weighting scheme improves the power of the test statistics in terms of detecting an underfit ARCH process. Additional studies suggest the same results for the fitted and underfitted GARCH process. Here we provide a brief simulation study showing the effectiveness of the statistics. A time series is simulated from an AR(1) with ARCH(2) errors, the specific

parameters are $\varphi_1 = 0.2$, $\alpha_0 = 0.2$, $\alpha_1 = 0.2$, $\alpha_2 = 0.2$ and an AR(1) with ARCH(1) errors is improperly fit to the model. Whence, we are properly fitting the linear process (autoregressive of order 1) but inadequately fitting the heteroscedastic process (the ARCH). Table 2 shows the results for increasing sample sizes at two lags. We see in the results that the Weighting scheme seems to improve power. Furthermore, both statistics tends to be consistent (i.e. power increases to 1) and that there is a reduction in power as the lag increases, although the amount of reduction seems to diminish with the weighting scheme.

	$m = 6$		$m = 12$	
n	$L_W(q,m)$	$L(q,m)$	$L_W(q,m)$	$L(q,m)$
100	**0.185**	0.134	**0.139**	0.081
200	**0.393**	0.330	**0.338**	0.233
300	**0.579**	0.491	**0.511**	0.399
400	**0.753**	0.662	**0.676**	0.553

Table 2. Power Level of Li Mak and Weighted Portmanteau Test

Implementation in SAS

Unlike the test statistics for a fitted GARCH process, which require linear algebra and the IML package in SAS, the calculation of the stastistics for a fitted ARCH is relatively straight forward. In the source code below we calculate the test from Li and Mak (1994) and Fisher and Gallagher (2012). We utilize the **AUTOREG** procedure to fit an ARCH(1) process to the residuals of the fitted ARMA(1,0) IBM return series. The residuals errors and fitted conditional variance series, $\hat{h}_t$, are output. A data step is used to calculate the standardized squared residuals. The remainder of the code follows the previous examples. Here we check at lag $m = 10$.

```
PROC AUTOREG DATA=getResids;
  MODEL RESIDUAL = /garch = (q=1);
  OUTPUT OUT=archResids CPEV=ht RESIDUAL=errors;

DATA archResids;
  SET archResids;
  sqrStdResids = errors*errors/ht;

PROC ARIMA DATA=archResids;
  IDENTIFY VAR=sqrStdREsids NLAG=20 OUTCOV=acfsARCH;

DATA acfsARCH;
  SET acfsARCH;
  IF lag < 2 THEN r=0;
    ELSE r = corr*corr;

  sampSize = N+lag;
  QtmpLB + r;
  QtmpLB = QtmpLB;
  LiMak = QtmpLB*sampSize;
  pq = 1;
  degFree = lag-pq;
  m = 10;
  shape = (3/4)*(m+pq+1)*(m+pq+1)*(pq)/(2*m*m+3*m+2*m*pq+2*pq*pq+3*pq+1);
  scale = (2/3)*(2*m*m + 3*m + 2*m*pq + 2*pq*pq + 3*pq + 1)/(m*(m+pq+1));
  weights = (m - lag + pq + 1)/(m);
  WLMtmp + (weights*r);
  WLMtmp = WLMtmp;
  WLiMak = WLMtmp*sampSize;
  pValLM = SDF('chisquared',LiMak,degFree);
```

338-2012 Weighted Portmanteau Test Revisited, continued

```
    pValWLM = SDF('gamma',WLiMak, shape, scale);

  PROC PRINT DATA=acfsARCH(where=(lag=10));
    VAR m degFree LiMak pValLM WLiMak pValWLM;
  RUN;
```

Output 3 shows the results of the final PRINT statement providing the test statistics and p-values for the Li Mak test and the weighted variant. As can be seen in the below output, we have very strong evidence that the ARCH(1) model is not adequate in capturing the nonlinear, or heteroscedastic, process in the IBM series.

Obs	m	deg Free	LiMak	pValLM	WLiMak	pValWLM
11	10	9	51.9836	4.5528E-8	35.6539	2.5413E-8

Output 3. Output from the final PROC PRINT Statement

A not included GARCH(1,1) fit is performed and both the Li Mak and Weighted Li Mak test suggest it is adequate in modeling the heteroscedastic process in the IBM series. Overall we conclude that an ARMA(1,0) is adequate for the serial correlation and a GARCH(1,1) is adequate in modeling the change of variance. This is typical of stock returns, an AR(1) or ARMA(1,1) is an adequate linear fit, and a GARCH(1,1) is adequate for the nonlinear term. As mentioned before, this follows an example in Tsay (2010) and our findings agree with the results in the text.

FUTURE CONSIDERATIONS

The weighting scheme presented in this article and developed in Fisher and Gallagher (2012) is quite versatile. Currently the methodology is being extended for use in other time series models. The author, along with some collaborators, is developing diagnostic procedures for noncausal time series with inPartfinite variance. That is, an observation may depend on the future, not just the past (strictly causal), and may exhibit large spikes in the series (extreme value/infinite variance). Fitting models of this type is notoriously difficult and very little diagnostic procedures are currently available.

The author is also extending the results presented in this paper for applications to Vector ARMA (VARMA) processes. This is a common modeling technique utilized in multivariate time series; that is, consider the simultaneous analysis of more than one time series. Not only will each series exhibit some of the properties outlines in this paper, but they may also interact.

CONCLUSION

This article outlined the techniques of fitting time series in SAS. It introduced some Weighted Portmanteau Statistics based on the recent theoretical work of Fisher and Gallagher (2012) and demonstrated how they can be used in diagnostic testing time series models. The results were extended to the problem of detecting nonlinear processes and for diagnostic procedures on a fitted ARCH Model. Some brief simulation studies are provided showing the efficacy of the weighting scheme. SAS source code was provided throughout using the common procedures **PROC ARIMA** and **PROC AUTOREG** along with the data step. Lastly, the article closes by discussing some future considerations currently being addressed by the author.

REFERENCES

Bollerslev, T. (1986) "Generalized autoregressive conditional heterskedasticity," *Journal of Econometrics*, Vol. 31, No. 3, 307—327.

Box, G.E.P. and Pierce, D.A. (1970). "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, Vol. 65, No. 332, pp. 1509—1526.

Dickey, D. A. (2011). "SAS/ETS® and the Nobel Prize." *Invited Paper 328-2011 to the 2011 SAS Global Forum*, Las Vegas, NV, April 2011.

Engle, R. F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica,* Vol. 50, No. 4, 987—1007.

Fisher, T. J. (2011). "Testing the Adequacy of ARMA Models using a Weighted Portmanteau Test on Residual Autocorrelations," *Contributed Paper 327-2011 to 2011 SAS Global Forum*, Las Vegas, NV, April 2011.

338-2012 Weighted Portmanteau Test Revisited, continued

Fisher, T. J. and Gallagher, C. M. (2012). "New Weighted Portmanteau Statistics for Time Series Goodness-of-Fit Testing," in revision with the *Journal of the American Statistical Association*.

Li, W. K. and Mak, T. K. (1994). "On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity." *Journal of Time Series Analysis*, Vol. 15, No. 6, 627—636.

Lin, J.W. and McLeod, A.I. (2006). "Improved Peña-Rodríguez portmanteau test," *Computational Statistics and Data Analysis*, Vol. 51, No. 3, pp. 1731—1738.

Ljung, G.M. and Box, G.E.P. (1978). "On a measure of lack of fit in time series models," *Biometrika*, Vol. 65, No. 2, pp. 297—303.

Mahdi, E. and McLeod, A. I. (2012). "Improved Multivariate Portmanteau Diagnostic Testing," *Journal of Time Series Analysis*, Vol 33, No 2, 211-222.

McLeod, A. I. and Li, W. K. (1983). "Diagnostic checking ARMA time series models using squared-residual autocorrelations," *Journal of Time Series Analysis*, Vol. 4, No. 4, 269—273.

Monti, A.C. (1994). "A Proposal for a Residual Autocorrelation Test in Linear Models," *Biometrika*, Vol. 81, No. 4, pp. 776—790.

Peña, D. and Rodríguez, J. (2002). "A Powerful Portmanteau Test of Lack of Fit for Time Series," *Journal of the American Statistical Association*, Vol. 97, No. 458, pp. 601—610.

Peña, D. and Rodríguez, J. (2006). "The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series," *Journal of Statistical Planning and Inference*, Vol. 136, No. 8, pp. 2706—2718.

Taylor, S. J. (1986). "Modelling Financial Time Series," Wiley, New York.

Tsay, R. S. (2010). "Analysis of Financial Time Series". Third Edition. Wiley, New Jersey.

Tsay, R. S. "Webpage for Analysis of Financial Time Series". 14 March 2012.
http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts/

ACKNOWLEDGMENTS

RECOMMENDED READING

The theoretical results of this paper are currently under review in the Journal of the American Statistical Association. Those interested in obtaining a draft of the work are encouraged to contact the author or Dr. Colin Gallagher of Clemson University.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Dr. Thomas Fisher
Department of Mathematics and Statistics
University of Missouri-Kansas City
5100 Rockhill Road
Kansas City, MO 64110
816-235-2853 (Office)
816-235-5517 (Fax)
fishertho@umkc.edu
http://f.web.umkc.edu/fishertho