Alternative Tests for Time Series Dependence Based on Autocorrelation Coefficients

Richard M. Levich and Rosario C. Rizzo [*]

Current Draft: December 1998

Abstract: When autocorrelation is small, existing statistical techniques may not be powerful enough to reject the hypothesis that a series is free of autocorrelation. We propose two new and simple statistical tests (*RHO* and *PHI*) based on the unweighted sum of autocorrelation and partial autocorrelation coefficients. We analyze a set of simulated data to show the higher power of *RHO* and *PHI* in comparison to conventional tests for autocorrelation, especially in the presence of small but persistent autocorrelation. We show an application of our tests to data on currency futures to demonstrate their practical use. Finally, we indicate how our methodology could be used for a new class of time series models (the Generalized Autoregressive, or *GAR* models) that take into account the presence of small but persistent autocorrelation.

An earlier version of this paper was presented at the Symposium on Global Integration and Competition, sponsored by the Center for Japan-U.S. Business and Economic Studies, Stern School of Business, New York University, March 27-28, 1997. We thank Paul Samuelson and the other participants at that conference for useful comments.

---

[*] Stern School of Business, New York University and Research Department, Bank of Italy, respectively.

1. Introduction

Economic time series are often characterized by positive autocorrelation. For macroeconomic data (such as gross domestic product, unemployment, housing starts, and so forth), such persistence is commonly associated with business cycle phenomenon and periods of expansion and recession. For financial market data, most studies by economists find little evidence of significant autocorrelation in price changes.[1] In contrast, many market analysts rely heavily on technical trading signals, especially for short-term projections.[2] These trading signals are often based upon the assumption of positive autocorrelation ("trending behavior") in financial asset's returns.

The presence of autocorrelation in economic data is important for at least two reasons. First, for modeling or forecasting macroeconomic data, the classical regression model assumes that error terms are free of autocorrelation. When autocorrelation is present, the classical model is mispecified, so an alternative model would produce a superior description of the data and possibly improved forecasts. Second, when analyzing financial price data, the presence of autocorrelation in price changes could lead to a profitable trading rule based on publicly available information. Autocorrelation of this extent would constitute a violation of the informationally efficient market hypothesis.

We show that, when the autocorrelation of price changes is *small but persistently positive* (or negative), standard tests for the significance of autocorrelation at an individual lag (like the Bartlett test) or joint tests for significance at several lags (like the Box-Pierce *Q-test*) are unlikely to reject the null hypothesis of absence of serial correlation. We propose two new test statistics, indicated as "P" (the Greek capital letter "*RHO*") and "F" (the Greek capital letter "*PHI*"), which analyze the progression of, respectively, the autocorrelation and partial autocorrelation functions over successive time lags. Our new tests combine some ideas contained in the traditional Q test, in the recently developed Variance Ratio test for autocorrelation, which can be seen as a weighted sum of autocorrelations, and in the CUSUM test of residuals, which measures the tendency of a structural model to break down over time.

Using Monte Carlo simulations, we show that the P and F tests have greater power than conventional tests to identify small, but persistent, autocorrelation. While the Variance Ratio test picks up the presence of autocorrelation in our simulations, the P and F tests seem more precise in identifying the correct order of the autocorrelation process. We present an application of the P

---

[1] See Fama (1991) for a survey of the efficient capital markets literature. See Meese (1990) for the view that foreign exchange rate changes are essentially random and unpredictable. However, Mussa (1979, 1990) concludes that the exchange rate is only approximately a random walk. Recently, several authors have demonstrated that the random walk representation of certain time series can be rejected, if the alternative model comes from a class of non-linear models. See Bilson (1990, 1993) and Engel and Hamilton (1990).

[2] A survey of London foreign exchange traders by Taylor and Allen (1992) finds that nearly 90% report the use of technical models for short-run foreign exchange forecasts. Frankel and Froot (1990) also discuss the widespread use of chartist techniques in short-run exchange rate forecasting.

and F statistics to data on financial futures, where our tests identify significant autocorrelation that would not have been identified by conventional statistical tests.

As an extension, we indicate how our methodology could be used for a new class of time series models ("Generalized Autoregressive", or GAR models) that take into account the presence of small but persistent autocorrelation. The GAR model uses the descriptive properties of P and F to estimate time series models that exploit the piece-wise nature of the cumulative autocorrelation function. The model appear to have explanatory power on an in-sample basis for some series of currency futures returns, as it finds significant positive autocorrelation. These results are consistent with prior studies that show the profitability of some trading rules commonly used by "technical" analysts in the foreign exchange market.

Overall, our results suggest that the P and F statistics may be useful additions to the econometrician's toolkit for identifying the presence and nature of autocorrelation. The P and F approach is general and could be applied to any time series of economic data, regression residuals, or financial asset returns.

The plan of our paper is as follows. Section 2 reviews various tests of autocorrelation from the traditional statistics literature. In Section 3, the P and F tests are developed. Section 4 reports the empirical distribution and the power function of the P and F tests based on Monte Carlo simulations and compares them with other tests for the presence of autocorrelation. In Section 5, we show an empirical application of the P tests using daily series of currency futures prices. Section 6 shows how to extend the P and F approach by estimating a piece-wise time series model (the *Generalized Autoregressive, or GAR model*) for two currencies. Concluding remarks are in the final section.

2. Tests of autocorrelation in econometric analysis

Autocorrelation (or serial correlation) is important in econometric analysis for at least two reasons. First, when autocorrelation is present, a fundamental assumption of the OLS method is violated, and the estimates of coefficient variance will be biased downward and the t-statistics upward. Second, when a time series shows significant autocorrelation, it is possible to represent it as a time series model (e.g. an ARIMA model) which will yield non-trivial forecasts for the future observations of the series. In this section, we briefly review the most widely used tests of autocorrelation and outline the advantages and limitations of each of them.

2.a The Durbin-Watson, the Durbin *h* and *m* tests

The *Durbin-Watson test*, published in 1950, is probably the best known test for serial correlation.[3] Consider the residuals e_t from an OLS regression with T observations. To test the null

[3] J. Durbin and G.S. Watson, "Testing for Serial Correlation in Least Squares Regression," *Biometrika*, 1950, pp. 409-29, and 1951, pp. 159-78.

hypothesis of no first-order autocorrelation, the DW test statistic is defined as :

$$d = \sum_{t=2}^{T} (e_t - e_{t-1})^2 \; / \; \sum_{t=1}^{T} e_t^{\,2} \qquad (1.)$$

Compared to most estimators in classical statistics, the cutoff between the acceptance and the rejection region is not clear, as an inconclusive region is present. The DW test relies on two limits, d_l and d_u . In a two-tailed test, where both positive and negative autocorrelation are tested, we reject the null hypothesis (of absence of serial correlation) for values of *d* below d_l or above 4-d_l, and we fail to reject the null hypothesis for values between d_u and 4-d_u . However, the test is inconclusive for values between d_l and d_u or between 4-d_u and 4-d_l .

A second drawback of the DW test is that it tests the null hypothesis of absence of autocorrelation against the alternative hypothesis of first-order autocorrelation only, while real economic time series often present autocorrelation of higher order.

In addition to the limitations mentioned in the previous section, the DW statistic cannot be used to test for autocorrelation of residuals when an explanatory variable in the regression is a lagged dependent variable. In this situation, the DW statistic will be biased toward 2 (acceptance of the null hypothesis) even when the errors are serially correlated. In order to solve this problem, Durbin (1970) proposed a modification, the *h* test, which under the null hypothesis is approximately normally distributed with unit variance. The test statistic is defined as:

$$h = \left(1 - \frac{d}{2}\right) \times \left(\frac{T}{1 - T\, s_b^{2}}\right)^{1/2} \qquad (2.)$$

where s_b^2 is the estimated variance of the coefficient β of the lagged dependent variable, and T is the sample size in the regression.

Durbin's *h* test cannot be used for $T s_b^2 \geq 1$. In this case, Durbin (1970) proposed another statistic, the *m* test. The latter consists of performing the original OLS regression, collecting the residuals and running the following regression:

$$e_t = b_1 + b_2 X_t + b_3 Y_{t-1} + b_4 e_{t-1} + u_t \qquad (3.)$$

where Y_{t-1} is the lagged dependent variable, and X_t is the (vector of) independent variable(s). Serial correlation is tested by using the t-value of the β_4 coefficient. The main advantages of the *m* test are that (1) it does not include an indeterminate region, (2) it does not suffer the restriction of the *h* test when $T s_b^2 \geq 1$, and (3) it can be extended to test against the hypothesis of serial correlation of higher order by including further lagged residuals and testing the joint hypothesis that all the β coefficients are equal to zero. However, the approach used in the *m* test is similar to the autocorrelation and partial autocorrelation functions, and these are more frequently used in econometric analysis.

2.b Autocorrelation and partial autocorrelation functions: the Bartlett test

The autocorrelation function (*ACF*) is a widely known statistic for detecting the presence of serial correlation. The *ACF* is more useful than the tests mentioned above, as it provides a more detailed description of the underlying process.

In the following sections, we will consider the stationary process $\{y_t\}$, a more general case than the residuals $\{e_t\}$ from an *OLS* regression.[4] We recall that for a process to be *stationary*, the mean (m) and the variance (s^2) of the series, as well as the covariance between y_t and y_{t-k}, must be invariant with respect to time.

The *autocorrelation coefficient with lag k* (r_k) indicates the degree of correlation between values of the series $\{y_t\}$ separated by k lags, and it is defined as the covariance between y_t and y_{t-k} divided by the product of their standard deviations:

$$r_k = \frac{cov\,(y_t\,, y_{t-k})}{s\,(y_t)\;\; s\,(y_{t-k})} = \frac{E[(y_t - m)\,(y_{t-k} - m)]}{E[(y_t - m)^2\,]} \qquad (4.)$$

For any stochastic process, $r_0 = 1$. Moreover, the *ACF* is symmetric about zero, as $r_k = r_{-k}$ for all k. Therefore, the graph of the *ACF* against k, called the *correlogram*, considers only positive values of k.

The autocorrelation coefficient r_k measures both the direct correlation between y_t and y_{t-k} and the indirect correlation due to the intervening coefficients r_1 , r_2 , ... , r_{k-1} . For example, when only first-order serial correlation is present, the autocorrelation coefficient r_2 is equal to the correlation between y_t and y_{t-1} times the correlation between y_{t-1} and y_{t-2}, therefore $r_2 = r_1^2$. The partial autocorrelation function (*PACF*) is thus useful to net out the effects of the intervening coefficients and to measure the direct correlation between y_t and y_{t-k}. The *partial autocorrelation (coefficient) with lag k* (ϕ_k) is defined as the k-th coefficient a_k in a linear projection of y_t on its k most recent values:[5]

$$y_t = a_0 + a_1\, y_{t-1} + a_2\, y_{t-2} + \ldots\ldots + a_k\, y_{t-k} \qquad (5.)$$
$$f_k = a_k$$

The *ACF* and *PACF* are extremely useful as they help us to identify the correct specification for an ARIMA model that describes the stochastic process $\{y_t\}$. In particular, if the process is white noise, all autocorrelation and partial autocorrelation coefficients equal zero. If the process is an AR(p), the *PACF* will equal zero for all lags $k>p$, while if the process is a MA(q) the *ACF* will

[4] The analysis holds for either a weakly or a strictly stationary process.

[5] See J. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.

equal zero for all lags $k>q$.[6]

The *ACF* and *PACF* are theoretical functions and apply to stochastic processes. When we analyze real time series with finite observations, we have to estimate them. The sample *ACF* is defined as:

$$\hat{\rho}_k = \sum_{t=k+1}^{T} (y_t - \bar{\mu})(y_{t-k} - \bar{\mu}) \Big/ \sum_{t=1}^{N} (y_t - \bar{\mu})^2 \qquad (6.)$$

where $\bar{\mu}$ is the sample mean of the series.

The *PACF* $\hat{\phi}_k$ can be estimated by the OLS regression[7]:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2} + \ldots\ldots + \hat{\beta}_k y_{t-k} + \varepsilon_t \qquad (7.)$$
$$\hat{\phi}_k = \hat{\beta}_k$$

To test whether a sample autocorrelation (and partial autocorrelation) coefficient equals zero, we need to specify the distribution of the standard estimators of the autocorrelation (and partial autocorrelation) coefficients. This test was developed by Bartlett (1946), who showed that if a series is generated by a white noise process, the estimators are approximately normally distributed random variables with mean zero and variance $1/T$, where T is the total number of observations. The same characteristics hold for the distribution of the estimators of the partial autocorrelation coefficients.[8] These conclusions are valid for large values of T. For small samples, Fuller (1976) shows that the mean has a negative bias and the variance is smaller than 1/T. However, in the following sections we will maintain the commonly used approximations. Therefore, to make a 95% confidence test of the null hypothesis of no autocorrelation or partial autocorrelation at lag k, we simply need to compare the value of the sample coefficient with the critical values $\pm 1.96/T^{1/2}$. If the value falls outside the bands, the null hypothesis is rejected at the 95% level.

2.c The Box-Pierce *Q*-statistic

The Bartlett confidence interval tests the null hypothesis that a *single* autocorrelation coefficient is equal to zero. However, if a stochastic process is white noise, then *all* the autocorrelation coefficients (ρ) will be zero. Similarly, if a stochastic process is an MA(q) process, then *all* ρ_k for lags k>q will be equal to zero. Thus, to test for the overall and simultaneous absence of autocorrelation at multiple lags, we need a test of the joint null hypothesis

[6] See G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, 1976.

[7] See J. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.

[8] The value of the variance was shown by M.H. Quenouille in "Approximate Tests of Correlation in Time Series", *Journal of the Royal Statistical Society*, n.68, 1949.

that a group of coefficients are equal to zero.

Box and Pierce (1970) developed such a test based on the fact that, for a white noise process, autocorrelation coefficients are approximately independently distributed normal random variables.[9] The Box-Pierce test (commonly known as the *Q-statistic*, and also referred to as a "portmanteau test") tests the null hypothesis of zero autocorrelation at lags *1* to *N*. *Q* is defined as follows (in the ensuing expressions, when it will be unambiguous, the hatted symbol " ^ " for estimated values will be omitted):

$$Q = \sum_{i=1}^{N} (r_i / (1/T^{1/2}))^2 = T \sum_{i=1}^{N} r_i^{\,2} \qquad (8.)$$

For a white noise process, expression (8.) is a sum of squares of approximately independent normal random variables with mean zero and unit variance (the estimated coefficients ρ_i are normalized by dividing them by their standard deviation $1/T^{1/2}$). As a consequence, *Q* is approximately distributed as a chi-square with N degrees of freedom.[10] The null hypothesis of white noise is often tested by comparing *Q* with critical values for N=5, 10, 20 and 30 lags. If the statistic is significant, the null is rejected and a better model for the series should be specified.

The Box-Pierce *Q*-statistic turns out to be significant when more than the expected number of coefficients (r_i) are significant according to the Bartlett test, or even when none is significant but several are almost significant. In the latter case, *Q* is more powerful, as it allows for a rejection of the null hypothesis whereas the Bartlett test on individual coefficients fails to reject.

Various studies have been critical of the generally low power of the *Q*-statistic in identifying departures from the white noise process.[11] However, a more serious criticism is the claim that the *Q*-statistic (like the original DW test) is inappropriate for models with lagged dependent variables.

2.d Lagrange Multiplier (LM) test

A general test for the presence of higher-order serial correlation was developed by Breusch (1978) and Godfrey (1978a, 1978b) based on the Lagrange multiplier principle. The test is similar to Durbin's m-test discussed earlier. Again, let e_t represent the residuals from a linear regression of the form:

$$y_t = \sum^{k} x_{it} b_t + e_t \qquad (9.)$$

---

[9] See G. Box and D. Pierce, "Distribution of Residual Autocorrelations in ARIMA Time Series Models", *Journal of the American Statistical Association*, n.65, December 1970, pp. 1509-1526.

[10] For the residuals of an ARIMA(p,q) process, the Box-Pierce statistic is distributed as a chi-square with N-*P*-*Q* degrees of freedom. See R. Pindyck and D. Rubinfeld, *Econometric Models and Economic Forecasts*, McGraw-Hill, 1991. Ljung and Box (1978) and Kmenta (1986, *p*. 332) developed a modified *Q*-statistic suitable for moderate size samples.

[11] For example, see the discussion in Maddala (1992, pp. 539-42) and the references therein.

$i=1$

and assume that the errors are specified as:

$$e_t \;=\; r_1\, e_{t-1} \;+\; r_2\, e_{t-2} \;+\; \times\times\times + \; r_P\, e_{t-P} \;+\; u_t \qquad (10.)$$

where u_t is normally distributed with mean zero and constant variance. To test the null hypothesis that $r_i=0$ for all i, we estimate the regression:

$$e_t \;=\; \underset{i=1}{\overset{k}{S}}\, x_{it}\, g \;+\; \underset{i=1}{\overset{p}{S}}\, u_{t-i}\, r_i \;+\; h_t \qquad (11.)$$

and test whether the r_i are jointly zero.

The LM test uses the F-statistic from expression (11) and examines p ´F as chi-square with p degrees of freedom. By specifying expression (11) appropriately, the error series could be tested against autoregressive or moving average processes of any order. Despite the advantages of the LM test, many studies of economic time series continue to report *Q*-statistics and interpret insignificant values as evidence that the residuals in a regression or an ARIMA model are free of autocorrelation.[12]

2e. The Variance Ratio (VR) Test

While variance ratio tests are typically associated with tests for equal variance across samples, Lo and MacKinlay (1989) describe how variance ratios can be used to test for autocorrelation.[13] Under the joint null hypothesis of no serial correlation in returns and constant variance of returns, the variance of two-period compounded returns, $r_t(2) \equiv r_t + r_{t+1}$, should be twice the variance of one-period returns r_t .More formally, the two-period variance ratio, VR(2),

[12] Maddala (1992, *p*. 542) suggests that studies continue to rely on *Q-statistic*s because they are automatically reported in most time-series computer software. While the standard RATS output includes the DW and Ljung-Box *Q-test*s for serial correlation, users are encouraged to use the optional general LM tests when the model contains a lagged dependent variable. The standard MICROFIT output includes the DW, Durbin's *h* test, and LM tests, but there is no option to compute a *Q-statistic*.

[13] See Campbell, Lo, and MacKinlay (1997, pp. 48-55) for a thorough discussion of the VR test.

can be written as:

$$VR(2) = \frac{Var[r_t\,(2)]}{2\,Var[r_t]} = \frac{Var[r_t + r_{t+1}]}{2\,Var[r_t]}$$

$$= \frac{2\,Var[r_t] + 2\;Cov[r_t\,, r_{t-1}]}{2\,Var[r_t]}$$

$$VR(2) \quad = 1 + 2\;\rho(1) \qquad (12.)$$

where ρ(1) is the first order serial correlation of returns. When returns are positively correlated, VR(2) > 1, and when returns are negatively correlated, VR(2) < 1. Significance tests can be based on the measure (VR(2)-1)/√2 which is distributed as a standard normal variate.

The variance ratio test can be generalized to test for higher order autocorrelation. The general *q*-period variance ratio statistic VR(q) can be written as

$$VR(q) = \frac{Var[r_t\,(q)]}{q\,Var[r_t]} = 1 + 2\sum_{k=1}^{q-1}(1-\tfrac{k}{q})\;\rho(k) \qquad (13.)$$

where $r_t(k) \equiv r_t + r_{t-1} + \cdots + r_{t-k+1}$ and ρ(k) is the *k*th order autocorrelation coefficient of $\{r_t\}$. Note that VR(q) tests for autocorrelation at lag *q-1*. Note also that VR(q) is a summation of ρ(k) with linearly declining weights. Thus VR(q) is similar to our RHO (defined in section 3) which is an unweighted summation of ρ(k). The differences we find empirically between VR and RHO are likely the result of these different weights.

Under the null hypothesis of zero serial correlation at all lags, VR(q) = 1. As in the two-period example, VR(q) will exceed one when there is positive autocorrelation of returns. VR(q) will level off at lags where ρ(k) becomes zero. Significance tests about VR(q) can be based on the measure √nq (VR(q)-1) which is distributed approximately normal with mean zero and variance 2(q-1).[14] Other alternative hypotheses about ρ(k) can be specified precisely and tested against the data.

3. New tests of autocorrelation in time series

While there are numerous techniques to test for autocorrelation in residuals or time series

---

[14] This test assumes constant variance. See Campbell, Lo, and MacKinlay (1997, pp. 52-5) for the case of heteroskedasticity.

data, earlier research has pointed out the limitations of many of the standard tests. We conjecture that there is another limitation, namely, low power in identifying autocorrelation that is small but persistent. This type of autocorrelation is particularly important when it occurs in financial prices, because of its implications for investment trading strategies and tests of market efficiency. In this section, we develop a new test for autocorrelation that is powerful in identifying this type of autocorrelation.

3.a The *RHO (P)* statistic

Previously, we argued that the Box-Pierce *Q-statistic* is more powerful than the Bartlett test because it permits us to reject the null hypothesis even when the Bartlett test is not significant. However, Q has an important limitation: it does not take into account the sign of the r_i coefficients, as they are squared.[15]

The sign of the r_i is an important piece of information that is lost in the *Q-statistic*. In fact, if the coefficients of a white noise process are approximately independent, normally distributed random variables with zero mean, the probability that the first N are all positive (or all negative) is much lower than the probability that they are about half positive and half negative. Indeed, most real macroeconomic and financial time series show a systematically higher percentage of positive autocorrelation coefficients in the initial lags.

In order to retain the information contained in the signs of the r_i, it is necessary to develop a statistic where the ρ_i are not raised to an even power, as in the Q-test. The simplest such statistic is the unweighted sum of the first N coefficients:

$$P_N = \sum_{i=1}^{N} r_i \qquad (14.)$$

Under the null hypothesis, P (the Greek capital letter "*RHO*") is the sum of N approximately independent, normally distributed random variables each with mean zero and, for large T, constant variance $1/T$. Therefore, P_N will be approximately normally distributed with mean zero and variance equal to the sum of the variances, in this case N/T. The normality of the P statistic allows us to define standard confidence intervals, which (for the 95% level) will be equal to ±1.96 standard deviations (± 1.96 ´ $(N/T)^{1/2}$).

Our P test has several advantages over the Box-Pierce *Q-test*. First, the *Q-statistic*, even when it is significant, does not tell us if there is *positive or negative autocorrelation*, while P does. Second, the *Q-statistic* tests the null hypothesis against a generic alternative of "autocorrelation" (of either sign), while with the P statistic we can test the null hypothesis against the generic alternative or against the alternative hypothesis of *autocorrelation of a specific sign* (usually positive). In practice, both two-tailed and one-tailed tests are possible. Third, as we will demonstrate, in several cases P is more powerful than Q. With the P test, in order to show

---

[15] Liu and He (1991, p. 778, footnote 6) make essentially the same point in their comparison of the VR and Q tests.

significant autocorrelation, it is not necessary that several r_i coefficients are large; it is sufficient that the r_i are *small but systematically of the same sign*.

Our P test may also have advantages relative to the VR test. First, as an unweighted sum, the P test is more intuitive and computationally more simple than the VR test. Second, our *P* test does not constrain the variance of returns to be equal across periods. Implicitly, the *P* test only assumes that the r_i are unbiased and efficient estimates of the true r_i.

Third, as a computational matter, if all $r_i = 0$, then any weighted or unweighted summation of r_i will equal zero. However, if several r_i are non-zero, then an evenly weighted summation (like *P*) may be more powerful and more able to precisely estimate the order of autocorrelation in the underlying process than the VR approach which places declining weights on successive r_i.

Fourth, the graphic representation of *P* against N (the *P* function), gives a clearer picture of the underlying stochastic process and may be better able to highlight intervals of positive autocorrelation, followed by zero or negative correlation, and so on. And fifth, as we will see later on, the *P* function allows us to introduce a new class of time series models, the *GAR (Generalized Autoregressive)* and *GARI (Generalized Autoregressive Integrated)* models, that reflect the time varying nature of autocorrelation.

3.b The *PHI* (F) statistic

As shown in the previous paragraph, the *P* statistic is the sum of the autocorrelation coefficients r_i . In the same way, we can extend the analysis to the partial autocorrelation coefficients f_i . In fact, we know that for a white noise process the f_i are also approximately normally distributed with same mean and variance as the r_i . Therefore, the statistic F (the Greek capital letter "*PHI*"), defined as:

$$F_N = \sum_{i=1}^{N} f_i \qquad (15.)$$

is also approximately normally distributed with mean zero and, for large *T*, variance *N/T*. Similarly, the confidence intervals for the 95% level will be equal to $\pm 1.96 \times (N/T)^{1/2}$.

For white noise or for any process with relatively low autocorrelation, the r_i and f_i coefficients will tend to be small and very similar, and the same will be true for the *P* and F statistics. Therefore, in the following sections with examples of financial historical time series, we will analyze only the *P* statistic. However, it is important to point out that when we analyze a series with substantial autocorrelation or when we want to specify a GAR or GARI model (see Section 6), the F statistic can give us some additional information.

4. Asymptotic distributions and power of tests for autocorrelation

In the previous section, we concluded that, under the null hypothesis of white noise, the *P* statistic is normally distributed with mean zero and standard deviation *N/T*. In this section, we perform a Monte Carlo simulation to verify if the actual shape of the distribution is indeed in line with the theoretical one. We generated time series of 1,000 observations as realizations of the following stochastic process:

$$\begin{aligned} y_t \;&=\; r_1\, y_{t-1} \;+\; r_2\, y_{t-1} \;+\; \ldots \;+\; r_p\, y_{t-p} \;+\; e_t \\ e_t \;&\sim\; N(0;1) \end{aligned} \qquad (16.)$$

using the following four specifications of the autocorrelation coefficients[16]

(S1)	White noise with no drift	($p=10$ and $r_1 = r_2 = \cdots = r_{10} = 0$)
(S2)	AR(10), with small autocorrelation	($p=10$ and $r_1 = \cdots = r_{10} = 0.5/1000^{½} = 0.0158$)
(S3)	AR(10), with large autocorrelation	($p=10$ and $r_1 = \cdots = r_{10} = 1/1000^{½} = 0.0316$)
(S4)	AR(20), with small autocorrelation	($p=20$ and $r_1 = \cdots = r_{20} = 0.5/1000^{½} = 0.0158$)

For each series, we estimated the autocorrelation (r_i), partial autocorrelation (f_i), *Q-statistic*s (based on both r_i and f_i), as well as our new statistics, P_i and F_i, at lags i=1, ... 100. We generated 1,000 replications for each process and used them to compute representative values for each test statistic at their 2.5, 50, and 97.5 percentiles of the 1,000 replications.

To begin, consider the white noise process (series S1). In Figure 1, we show the theoretical values of the 95% confidence bounds for P_i as widening solid lines. The expected value of P_i under the null hypothesis is zero. The simulation estimates of the P_i at their 2.5, 50, and 97.5 percentiles are plotted with the symbol •.

The graph shows that the *P*-statistics for the simulated series have the same general shape as predicted by their theoretical functions. The values of P_i at the 97.5 percentile are extremely close to their theoretical values. The values for P_i at the 2.5 percentile tend to drift apart, but are still reasonably close. The values for P_i at the 50 percentile are extremely close, and confirm the well-known negative bias in white noise.[17]

The next question is whether *P* and *F* can be used to identify series of autocorrelated observations. In Figure 2, we examine our Monte Carlo estimates of P_i (at the 50th percentile) for series S2, S3, and S4 that are constructed to have autocorrelation. S2 is a series with a small

[16] We use a "small" and "large" measure of autocorrelation defined as ½ and one standard deviation (respectively) of the sample autocorrelation statistic for a white noise series of 1,000 observations.

[17] The results for the Φ_i statistic (not shown) are virtually identical to the results for P_i .

amount of autocorrelation maintained for a relatively short period (10 lags). Notice that the median value of P_i for this series falls short of the 97.5% boundary. S3 is a series with a larger amount of autocorrelation maintained for a similar time (10 lags). In this case, the 50th percentile of P_i is well beyond the 97.5% boundary. Finally, S4 is a series with a small amount of autocorrelation, but maintained over a longer period (20) lags. In this final case, the 50th percentile of P_i again is well beyond the 97.5% boundary. Thus we see that our *P*-statistic is able to identify a small measure of autocorrelation that persists for a long enough time, or a somewhat larger measure of autocorrelation that persists for a shorter time.[18]

Notice also that the *slope* of the P_i function changes abruptly at the lag given by the AR process. While P_i continues to climb somewhat after the lag 10 (for series S2 and S3) and after lag 20 (for series S4), this movement is basically sideways (mimicking the pattern of the 50th percentile of white noise in Figure 1) and indicating an absence of autocorrelation after these lags.

From Figure 2, it is clear that a large percentage of the S2, S3, and S4 series would be identified as containing significant autocorrelation and at a lag length near 10 for S2 and S3, and near 20 for S4. An important question is whether the *P*-statistic would be more powerful in identifying the presence of autocorrelation than traditional tests. We now turn to this question.

For each of our four series, we use the 1,000 simulations to calculate a bootstrap estimate of the percentage of significant test statistics (at the 95% level) for testing the null hypothesis of no autocorrelation at lags 1-100. In the white noise case (series S1) shown in Figure 3, all the test statistics reject the null hypothesis about 5% of the time. There is some minor variation across the various test statistics. For example, our *P*-statistic tends to reject only about 3% of the time, while Φ tends to reject about 6-7% of the time at long lags. But we view this as normal sampling variation.

In Figure 4, we show the results for S2, an AR(10) process with small autocorrelation. Notice that the traditional ρ and ϕ measures occasionally identify the AR process, but only about 10% of the time in our sample of 1,000 replications. This is hardly higher than their performance with respect to a white noise process. The *Q-statistic* does a bit better; it rejects the null hypothesis about 15-17% of the time. However, the *P* and *F* statistics demonstrate much greater power, and reject the null hypothesis between 35-40% of the time.

Series S3 is also an AR(10) process, but with a larger ρ linking the observations. In Figure 5 we see that all of the test statistics gain power in identifying the presence of autocorrelation in S3. But still, the traditional ρ and ϕ measures reject the null only about 15-25% of the time. The *Q*-statistic does substantially better -- it rejects the null hypothesis about 65-70% of the time. However, *P* and *F* demonstrate still greater power, and reject the null hypothesis 90% of the time at lag 10.

Now examine Series S4 where a small amount of autocorrelation persists over 20 periods. In Figure 6, we see that the traditional ρ and ϕ measures reject the null only about 6-10% of the time. This is hardly better than their performance with respect to a white noise process. The *Q*-

[18] Again, the results for the Φ_i statistic (not shown) are virtually identical to the results for P_i .

statistic does better -- it rejects the null hypothesis about 20-30% of the time. But this is less than half the power of the P and F statistics that reject the null hypothesis 60-70% of the time at lag 20.

These simulation results show clearly that P and F are considerably more powerful than traditional tests. The power gain appears most sizable when autocorrelation is small (Figure 4) or small and persistent (Figure 6). These are extremely important types of autocorrelation in financial markets where profit opportunities may be earned by exploiting autocorrelation patterns.

Finally, in addition to comparing the P and F statistics to the traditional ρ and ϕ measures, we also examine the ability of the variance ratio (VR) test to identify small but persistent autocorrelation. In these tests we generated 500 observations for series S1, S2 and S3 described earlier. We again use a bootstrap procedure to estimate significance levels for testing the null hypothesis of no autocorrelation at all lags.

The results in Figure 7 show that when the underlying series is white noise (S1) the VR test rejects the null hypothesis about 5% of the time, very much as we found for the P and F statistics and the traditional ρ and ϕ tests. For series S2 and S3 that have serial correlation through 10 lags, the VR test rejects the null hypothesis (looking at lag 11) about 37% and 88% of the time respectively -- similar to the power of the P and F statistics (See Figures 4 and 5).[19] However, Figure 7 reveals an important difference -- the peak of the power function for S2 and S3 occurs at lags 17 and 16 respectively. In comparison, the P and F statistics spiked clearly at lag 10. Thus, while VR seems powerful in identifying persistent autocorrelation in these examples, it appears less precise than P and *?* in identifying the correct order of the autocorrelation process. Moreover P and F are more intuitive to interpret and easier to calculate than VR.

5. An application of the *P* statistic to currency futures

5.a Data source and description

In this section, we present an application of the P statistic to test for autocorrelation in time series of currency futures. If the prices in each series are determined in an efficient market with constant equilibrium returns, then the series of the returns should be white noise.[20] Thus, if the P statistic falls outside the 95% confidence intervals, then the efficient market hypothesis is rejected and the evidence favors an alternative hypothesis of autocorrelation.

Our data set is daily closing prices on the five most liquid currency futures traded on the Chicago Mercantile Exchange, Japanese yen (JY), British pound (BP), Deutsche mark (DM),

[19] The VR(q) tests for autocorrelation at lag *q-1*.

[20] Because tests of market efficiency are tests of a joint hypothesis, a finding of significant serial correlation in returns does not necessarily imply market inefficiency.

Swiss franc (SF) and Canadian dollar (CD).[21] We choose to examine futures markets because they are regarded as examples of efficient financial markets, due to high liquidity and low transaction costs. Moreover, since futures prices embody the time value of money, the prices do not follow any drift due to the interest rate differential.

Our sample period spans 15 years, from January 2, 1981 to December 29, 1995 and yields 3791 trading days. In order to create a continuous price series from the quarterly contacts (March, June, September, and December), we chain them together using a procedure called "back-adjustment." For any trading day, we use the return on the most liquid contract (usually the closest to expiration); the returns are then integrated to form a continuous price series. This procedure produces a true series of actual daily returns, as it eliminates the typical "jumps" on the four roll-over days per year (March to June, June to September, and so forth). All returns are logarithmic returns.

5.b Empirical results

For each currency, we calculate the autocorrelation coefficients, $\hat{r}_i$ for i = 1, ... 200 daily lags. The results are shown in Figures 8-12 where for each currency we graph the autocorrelation function (plotted as a sequence of squares), the Bartlett 95% confidence interval (±1.96 standard deviations, parallel bands at approximately ±0.03), the *P*-statistic (plotted as a sequence of triangles), and its 95% confidence intervals for a two-tailed test (±1.96 standard deviations, widening bands). A summary of the classic test results is in Table 1.

Table 1
Summary Statistics for Standard Tests of Autocorrelation in Currency Futures
Sample Period: January 2, 1981 - December 29, 1995, N=3791

	Bartlett Test: Number of Significant ρ(i)			Box-Pierce *Q*-test		
Currency	Lags 1-10	Lags 1-20	Lags 1-30	*Q*(10)	*Q*(20)	*Q*(30)
BP	0	1	3	4.46	15.83	37.38
CD	3	4	4	36.79**	53.34**	62.30**
DM	0	2	3	4.70	20.15	43.70*
JY	2	3	4	17.19	28.55	49.44**
SF	0	2	2	2.90	15.52	34.30

Note: ** Significant at 95% level
* Significant at 90% level

[21] The data are from a CD-ROM provided by Prophet Information Services, Inc. For other details on the vendor and the data set, see their WWW site at http://www.prophetdata.com .

In Table 1, autocorrelation appears greatest for the Canadian dollar which has the largest number of significant $\hat{r}_i$ at lags 1-20, and highly significant *Q*-statistics. For the other four currencies, some individual $\hat{r}_i$ are significant but not enough to produce significant *Q*-statistics, except for the DM and the Japanese yen with N=30 lags.

Next we calculate variance ratio (VR) statistics for the same sample of currency futures data and autocorrelations. We use equation (13) with standard errors estimated from a bootstrap simulation based on 1,000 replications of series with 3,791 observations. The results are summarized in Table 2.

Table 2

Summary Statistics for Variance Ratio Tests of Autocorrelation in Currency Futures

Sample Period: January 2, 1981 - December 29, 1995, N=3791

	Number of Significant VR(i) *			
Currency	Lags 1-10	Lags 1-20	Lags 1-30	Lags 1-40
BP	0	0	0	0
CD	2	2	2	2
DM	0	0	0	0
JY	0	0	0	7
SF	0	0	0	0

* Significant at the 95% level

The results in Table 2 show no evidence of autocorrelation at any lag in three currencies (BP, DM and SF) with some autocorrelation in the CD at very short lags. The VR test reveals some autocorrelation in the JY at lags 31-40, and then a stretch of significant autocorrelation at lags 65-100 days as shown in Figure 7A. Our results for the Yen are consistent with Liu and He (1991) who find significant VR(i) at lags of 4, 8, and 16 weeks using spot exchange rate data over roughly the same sample period. However, Liu and He also found some significant VR results for the DM and Pound that our analysis of currency futures does not confirm.

Overall, Table 1 gives the impression of substantial autocorrelation in the CD, with little evidence of autocorrelation for the other currencies up to lags 10 and 20. Table 2 supports the impression of little autocorrelation with the exception of the Yen. These impressions change when we take a closer look at the $\hat{r}_i$ and examine the *P*-functions for the individual currencies.

Japanese Yen. Figure 8 shows the results for the Japanese yen. The ρ_i are *slightly* significant at lags 8, 9 and 14. Not surprisingly, the *Q*-statistics for lags 10 and 20 are not significant at the 95% level. However, the *Q*-statistic for lag 30 is significant, mainly due to the large autocorrelation coefficient at lag 24. These results would not permit us to specify an ARIMA model for returns. Considering the first 200 lags, only 14 coefficients (or 7% of all 200) are significant, not an unreasonable number for a 95% confidence interval. Overall, analysis of the classic autocorrelation function does not allow us to convincingly reject the white noise process for returns. Moreover, the slight evidence from the *Q*-statistic does not say anything about the sign of the autocorrelation.

In the Japanese yen case, the sign of the autocorrelation coefficients is extremely important. Fifteen of the first 20 coefficients are positive, a low probability outcome if they were independent random variables with mean zero. In addition, the positive ρ_i are larger (in absolute value) than the negative ones, and the three significant ρ_i are all positive. We conjecture that a small but persistent positive autocorrelation is present.

The *P*-function, by calculating the cumulative sum of the ρ_i, clearly shows this non-random characteristic. The *P*-function begins near zero at lags 1 and 2. It starts to increase from lag 3 and crosses above the upper 95% confidence band at lag 10. Continuing on, the *P*-function remains outside the confidence boundary, and continues to increase up to a local peak of 0.21 at lag 15, and another local maximum of 0.50 at lag 101. The *P*-function does not return within the confidence interval until lag 133. Over all 200 lags, 152 values of the *P*-function (or 76%) are outside the 95% confidence boundary.

The results allow us to draw several important conclusions. (1) The *P*-function leads us to reject the hypothesis of no autocorrelation in returns for the Japanese yen at the 95% level (and also at the 99% level, as many values exceed even 3 standard deviations). (2) The *P*-function shows that the autocorrelation is positive. (3) The *P*-function shows that autocorrelation is not significant for the first few lags, but that it is small and persistently increasing up to lag 100. The local maximum at lag 101 and the generally sideways movement of the *P*-function thereafter is notable, because it indicates that the autocorrelation after this point develops in a fairly random pattern with mean zero -- i.e. the evidence points to an *absence of autocorrelation* after lag 101. This information cannot be provided by either the classical autocorrelation function or by the Box-Pierce statistic.

British Pound. The results for the British pound are shown in Figure 9. In this case also, analysis of individual ρ_i does not allow us to reject the null hypothesis of zero autocorrelation. Of the first 20 coefficients, only one (at lag 15) is significant. While 16 of the first 200 ρ_i, (or 8%) are significant at the 95% level, the Box-Pierce Q-statistics for lags 10, 20 and 30 are not significant. With these results, no ARIMA model can be specified.

However, once again in the BP case, the sign of the autocorrelation coefficients is very important. Of the 16 significant r_i, 13 are positive and only 3 are negative. Moreover, the positive coefficients are larger (in absolute value) than the negative ones.

In Figure 9, the *P*-function remains very close to zero for lags 1-19, but then it begins to increase and crosses above the upper 95% confidence band at lag 37. The *P*-function then moves sideways, inside and outside of the confidence interval, and then returns to an increasing pattern, up to the maximum of 0.37 at lag 75. Beyond lag 75, the *P*-function decreases, toward a local minimum of 0.13 at lag 158. Over all 200 lags, 24 values of the *P*-function (or 12%) are outside the 95% confidence boundary.

Once again, evidence from the *P*-function leads us to reject the null hypothesis of no autocorrelation in returns at the 95% level. The data indicate that the autocorrelation is positive, that it is not present for the first lags, and that it accumulates slowly thereafter between lags 20 and 75. The decreasing pattern of the *P*-function between lags 75 and 158 could suggest that the British pound follows a trend (positive autocorrelation) in the short term (up to 3 months) and a mean-reverting pattern (negative autocorrelation) in the medium term (7 months). This pattern is consistent with many theories of exchange rate behavior.

Deutsche Mark and Swiss Franc. The results for the Deutsche mark and Swiss franc (shown in Figures 10 and 11) are quite similar. The standard analysis of individual ρ_i does not allow us to reject the null hypothesis, and the *Q*-statistics for 10 and 20 lags are not significant, although *Q*(30) is slightly significant for the DM at the 90% level. No ARIMA model can be specified for either currency. In this case, the *P*-function does not reject the null hypothesis either, as no value exceeds the 95% confidence intervals. However, the results for the DM and SF are significant at the 90% or even at the 95% level if we apply a one-tailed rather than a two-tailed test. The former test can be supported on the grounds that many economic series exhibit positive autocorrelation, and evidence from trend-following strategies leads us to suspect positive correlation in currencies.

Canadian Dollar. The evidence for the Canadian dollar (in Figure 12) is somewhat different. The autocorrelation function indicates that the first three ρ_i coefficients are significant, with *t*-values +4.15, -2.01 and -2.99 respectively, and, as a consequence, the *Q*-statistics are significant also. This degree of autocorrelation would allow us to specify an ARIMA model with some explanatory power. The other coefficients between lags 4 and 200 do not seem to be particularly important. This is the only case where the *P*-function does not indicate a persistent pattern of autocorrelation. In fact, *P* is significant only at lag 1 (where it is coincident with the autocorrelation function), but afterwards remains well inside the 95% confidence interval.

5.c Comparison to results for technical trading rules

The evidence in Figures 8-12 suggests that positive autocorrelation was present in the return series for the BP, DM, JY and SF. Using the *P*-function, we found that positive autocorrelation reached significant levels at lags 10 (for the JY), 14 (for the DM and SF) and 37 (for the BP), and that in each case it persisted and grew larger at longer lags. Earlier research found that technical trading rules, which rely on trend-following and persistence, were particularly profitable for these same four currencies.[22] Some authors have suggested that non-linear methods are needed to understand how trend-following strategies could be profitable in markets whose prices apparently do not exhibit autocorrelation.[23] Our results suggest that small but persistent autocorrelation in daily price changes (that is not easily detected using traditional techniques) may provide an alternative explanation for the profitability of currency trend-following strategies.

Moreover, earlier studies found that these strategies were not particularly profitable in the

[22] For a summary see Levich (1998, pp. 222-231) and for specific results see Levich and Thomas (1993).

[23] See Engle and Hamilton (1990) and Bilson (1990, 1993) for examples of non-linear approaches to predicting currency movements.

case of Canadian dollars, and attributed this result to the relatively low volatility of this currency. Our results show that in addition, the Canadian dollar has lacked the kind of autocorrelation (positive and persistent) of the other four currencies.

6. The Generalized Autoregressive (GAR) model

As we suggested earlier, the *P*-function allows us to specify a new class of time series models, named "*Generalized Autoregressive models" (GAR)*. We saw that *P* is more powerful than both the *ACF* and the *Q*-statistic when the time series presents a small autocorrelation that persists for several lags. In this case, the generic term y_t does not depend directly on any specific past value y_{t-k} (as in the ARIMA model), but it depends on the *general tendency* of a set of past values. The GAR model takes into account this "general tendency," without focusing on any specific value, but considering the *sum* of the set of past values. For example, if we believe that an observation y_t has a positive autocorrelation with the previous, say, 10 values, we can express the observation as a linear function of a single explanatory variable, $S_{t1=1,10}\ y_{t-t1}$. In this way, *each past value has the same weight as the others*. It can happen that some specific y_{t-t1} will be negatively correlated with y_t, but if the autocorrelation is generally positive, the sum will have a significant coefficient. The GAR model has the important advantage of being generally more *parsimonious* than the classical ARIMA model.

In order to explain more clearly how the GAR model works, let's use the example of the British pound. As we saw in Figure 9, the *P*-function increases sharply between lags 20 and 75, moving from +0.02 to +0.37, above the 95% confidence interval. The *F* function (not shown) has a very similar pattern. Therefore, there is a general tendency of small positive correlation between the observation y_t and *any* observation in the interval (y_{t-20} , y_{t-75}). Based on this information, we can specify the following GAR(20-75) model for the returns on British pound futures:

$$y_t = a_1 \mathop{S}\limits_{t1=20}^{75} y_{t-t1} + u_t \qquad (17.)$$

In the model, the return is a linear function of a single variable, the sum of the returns between t-20 and t-75, and $\{u_t\}$ is a white noise process.

The *P*-function gives us still further information. As we can see in Figure 9, the function decreases from lag 97 to lag 158, moving from +0.36 to +0.12. This means that there is a general tendency of small negative correlation (or a "mean-reverting" pattern) between observation y_t and any observation in the interval (y_{t-97}, y_{t-158}). Therefore, we can improve the explanatory power of the GAR model, by using the following specification:

$$y_t = a_1 \mathop{S}\limits^{75} y_{t-t1} + a_2 \mathop{S}\limits^{158} y_{t-t2} + u_t \qquad (18.)$$

$t1=20 \qquad\qquad t2=97$

In this case, α_1 is expected to be positive and α_2 negative. An OLS regression on the British pound returns yields the following results (t-statistics in parentheses):

$$y_t = \underset{(2.03)}{+0.00417} \sum_{t1=20}^{75} y_{t-t1} \underset{(-1.83)}{-0.00340} \sum_{t2=97}^{158} y_{t-t2} + u_t \qquad (19.)$$

The F-statistic for this regression is 3.56, significant at the 95% level. The GAR(3-101, 123-140) model of the returns on the Japanese yen leads to even more significant results:

$$y_t = \underset{(2.30)}{+0.00303} \sum_{t1=3}^{101} y_{t-t1} \underset{(-2.69)}{-0.00990} \sum_{t2=123}^{140} y_{t-t2} + u_t \qquad (20.)$$

In this case, the *F*-statistic is 5.71, significant at the 99% level.

The GAR model describes stationary time series (as the ARMA model); non-stationary time series (as price levels) are described by GARI models, where the “I” stands for “Integrated”. More generally, a *GARI*(a_1 -b_1 , a_2 -b_2 , ... , a_p -b_p , ; *d*) *model* is a stochastic process whose *d-th* difference is equal to:

$$y_t = a_1 \sum_{t1=a1}^{b1} y_{t-t1} + a_2 \sum_{t2=a2}^{b2} y_{t-t2} + \ . \ . \ . \ + a_P \sum_{tp=ap}^{bp} y_{t-tp} + u_t \qquad (21.)$$

where: $a_1 \geq 1$, $b_i \geq a_i$, $b_{i+1} > a_i$, and $\{u_t\}$ is a white noise process.

The reason why the model has been called “Generalized” Autoregressive depends on the fact that the classical AR(p) model is a special case of the GAR model, where:

$a_1=b_1=1$, $a_2 = b_2 = 1$; .... ; $a_p = b_p = p$

For example, an AR(5) model can be expressed as a GAR(1-1, 2-2, 3-3, 4-4, 5-5) model.

Another special case is the “*momentum*” oscillator, an indicator used by the technical analysts of financial markets. This indicator generates a trading rule to buy a “trending” security when the price has gone up during the last n periods, and vice versa for a sell signal. In other words, it forecasts that the return for the following period will have the same sign of the cumulated returns over the past n periods. Thus, the momentum of order *n* is simply a GAR(1-n) model where $\alpha_1>0$. This consideration suggests interesting similarities between the forecasting methods of econometric theory and technical analysis.

7. Conclusions

This paper was motivated by the conjecture that some financial series may contain an amount of autocorrelation that is small, but persistent for several lags. Technical trading systems based on trend-following (or positive-feedback) strategies have been able to exploit this persistence in currencies, but traditional linear statistical methods typically cannot identify meaningful patterns of autocorrelation in the underlying series. Our review of these traditional methods pointed out several short-comings, such as tests which do not consider the joint distribution of the autocorrelation coefficients (ρ_i) or tests that do not consider the sign of the coefficients (Q).

We proposed two simple tests, *RHO* (P) and *PHI* (F), that depend on *both* the sign and magnitude of ρ_i and ϕ_i and their cumulative behavior (i.e. persistence). In large samples, P and F are approximately normally distributed and therefore easy to implement. Simulations based on white noise and autocorrelated stochastic processes showed that the empirical distributions of P and F matched closely to their theoretical values. We evaluated P and F against the Bartlett test and the *Q-test* and found that the former were more powerful in rejecting the null hypothesis. The power of the P and F tests is greatest when ρ is persistent over about 10-20 lags. The power of the P and F tests was similar to the VR test, but P and F appeared *more precise* in identifying the correct order of autocorrelation.

We showed an empirical application of P and F to currency futures data, and in four out of five cases found evidence to reject the null hypothesis of no autocorrelation in returns and support the alternative of positive autocorrelation.

Another important advantage of P and F is the higher information content of their visual representation. The cumulative sum of the autocorrelation coefficients showed patterns that were not easily identified by analyzing the graphs of *ACF*, *PACF* and *Q*. Thanks to these patterns, we can specify a new class of simple and parsimonious linear models, named *Generalized Autoregressive models*, of which the classical *AR* model represents a special case.

This paper reflects several shortcomings. We have not taken into account the effect of non-normality and heteroskedasticity on P and F. Second, our Monte Carlo simulations have analyzed only a small number of autocorrelation patterns. Further work is needed to determine how P and F compare to *VR* which is the nearest competitor to our new tests. It may be that both *P/F* and VR tests should be calculated to identify the presence of certain patterns of autocorrelation.[24]

Despite these limitations, we find the initial results in this paper encouraging. They suggest that the P and F statistics, which are general and could be applied to any time series, may be useful additions to the econometrician's toolkit for identifying the presence and nature of autocorrelation.

[24] See Miller and Newbold (1995) for conditions under which it may be appropriate to use both VR and Q- tests.

References

Bartlett M.S., On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series, *Journal of the Royal Statistical Society*, ser. B8, vol. 27, 1946.

Bilson, John F.O. "'Technical' Currency Trading," in Lee R. Thomas (ed.), *The Currency-Hedging Debate*, (London: IFR Publishing), 1990.

Bilson, John F.O. "Value, Yield, and Trend: A Composite Forecasting Approach to Foreign Exchange Trading", *Strategic Currency Investing*, (Chicago: Probus Publishing), 1993.

Box G.E., Jenkins G.M., *Time Series Analysis: Forecasting and Control*, (San Francisco: Holden-Day), 1970.

Box, G.E., Pierce D. "Distribution of Residual Autocorrelations in ARIMA Time Series Models", *Journal of the American Statistical Association*, n.65, December 1970, pp. 1509-1526.

Breusch, T.S. "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17 (1978): 334-55.

Campbell J.Y., Lo A.W., Craig MacKinlay A., *The Econometrics of Financial Markets*, (Princeton: Princeton University Press), 1997.

Durbin, J. "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors Are Lagged Dependent Variables", *Econometrica*, n.38, May 1970, pp. 410-21.

Durbin, J. and G.S. Watson. "Testing for Serial Correlation in Least Squares Regression," *Biometrika*, 1950, pp. 409-29, and 1951, pp. 159-78.

Engel, Charles and James D. Hamilton. "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?" *American Economic Review*, Vol. 80, No. 4, (1990): 689-713.

Fama, Eugene F. "Efficient Capital Markets: II," *Journal of Finance*, 46, No. 5 (December 1991): 1575-1617.

Frankel, Jeffrey A. and Kenneth A. Froot. "Chartists, Fundamentalists, and Trading in the Foreign Exchange Market," *American Economic Review* 80, no. 2 (May 1990): 181-5.

Godfrey, L.G. "Testing Against General AR and MA Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46 (1978a): 1293-302.

Godfrey, L.G. "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46 (1978b): 1303-10.

Kmenta, J. *Elements of Econometrics*, (New York: Macmillan), 1986.

LeBaron, B. "Technical trading rule profitability and foreign exchange intervention", Working Paper, Department of Economics, University of Wisconsin (1996)

Levich, Richard M. *International Financial Markets: Prices and Policies*, (New York: McGraw-Hill), 1998.

Lee, C.I. and Mathur, I. "Trading rule profits in European currency spot cross-rates", *Journal of Banking and Finance*, 20 (1996), 949-962

Levich, Richard M. and Lee R. Thomas. "The Significance of Technical Trading-Rule Profits in the Foreign Exchange Market: A Bootstrap Approach," *Journal of International Money and Finance* 12, No. 5 (1993): 451-74.

Liu, Christina Y. and Jia He. "A Variance-Ratio Test of Random Walks in the Foreign Exchange Market," *Journal of Finance* 46, no. 2 (June 1991): 773-85.

Ljung, G.M. and G. Box. "On a Measure of a Lack of Fit in Time Series Models," *Biometrika,* 65, (1978): 297-303.

Lo A.W., Craig MacKinlay A. "The Size and Power of the Variance Ratio Test in Finite Samples," *Journal of Econometrics*, 40, no. 2 (February 1989): 203-38.

Maddala, G.S. *Introduction to Econometrics*, Second Edition, (New York: Macmillan), 1992.

Mathur, I. and Szakmary, A., "Central bank intervention and trading rule profits in foreign exchange markets", Journal of International Money and Finance, 4 (1997): 513-535

Meese, Richard. "Currency Fluctuations in the Post-Bretton Woods Era," *Journal of Economic Perspectives,* 4, no. 1 (Winter 1990):117-34.

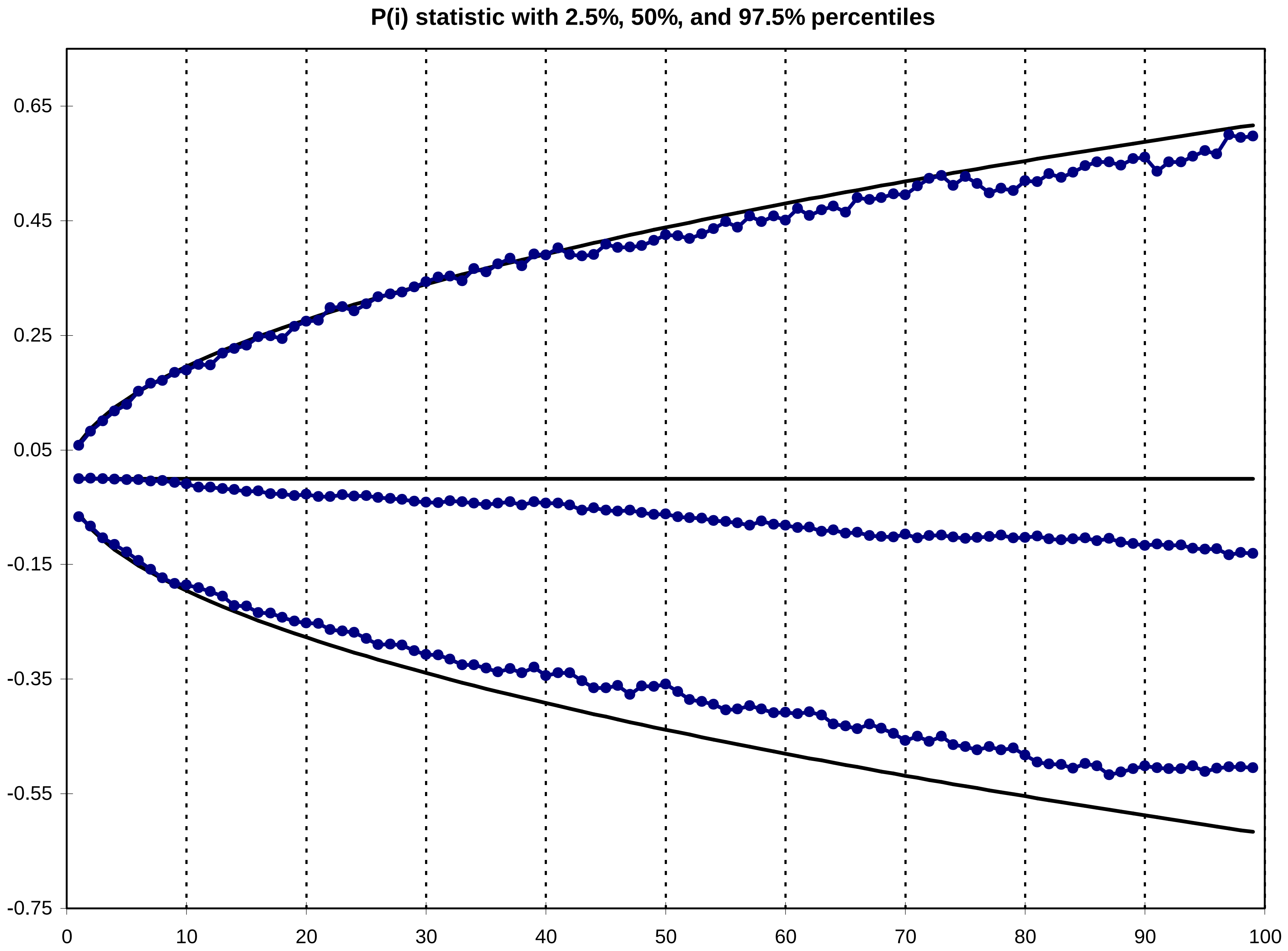
Miller, C. and P. Newbold. "A Generalized Variance Ratio Test of ARIMA(p,1,q) Model Specification," *Journal of Times Series Analysis* 16, no. 4 (July 1995): 333-49.

Mussa, Michael. "Empirical Regularities in the Behavior of Exchange Rates and Theories of the Foreign Exchange Market" in *Policies for Employment, Prices and Exchange Rates*, Carnegie-Rochester Conference 11 (Amsterdam: North-Holland), 1979, pp. 9-57.

Mussa, Michael. "Exchange Rates in Theory and Reality," *Princeton Essays in International Finance*, No. 179, December 1990.

Pindyck, R. and D. Rubinfeld. *Econometric Models and Economic Forecasts*, (New York: McGraw-Hill), 1991.

Taylor, Mark and Helen Allen. "The Use of Technical Analysis in the Foreign Exchange Market," *Journal of International Money and Finance*, (June 1992): 304-14.


P(i) statistic with 2.5%, 50%, and 97.5% percentiles
0.65
0.45
0.25
0.05
-0.15
-0.35
-0.55
-0.75
0
10
20
30
40
50
60
70
80
90
100

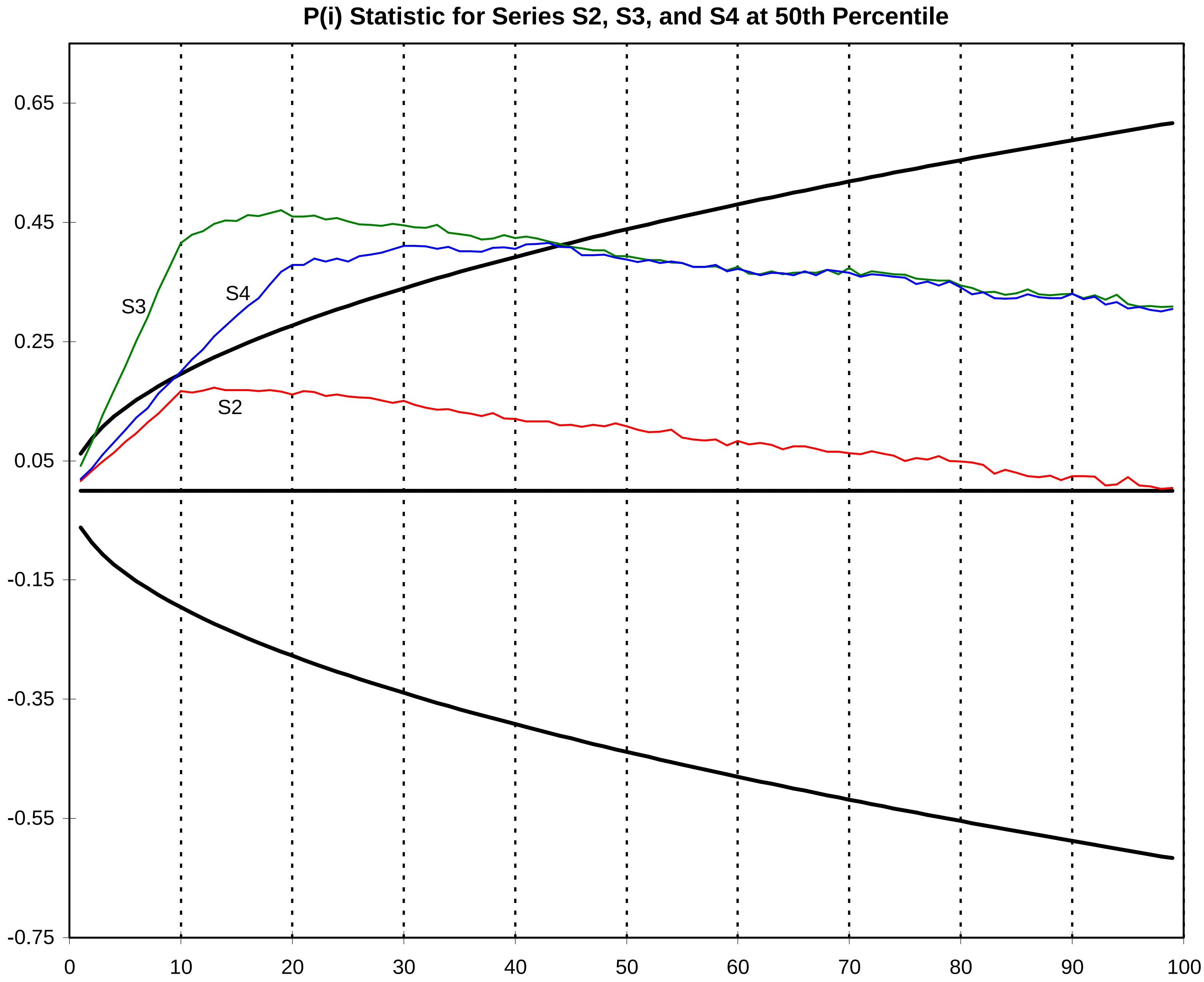
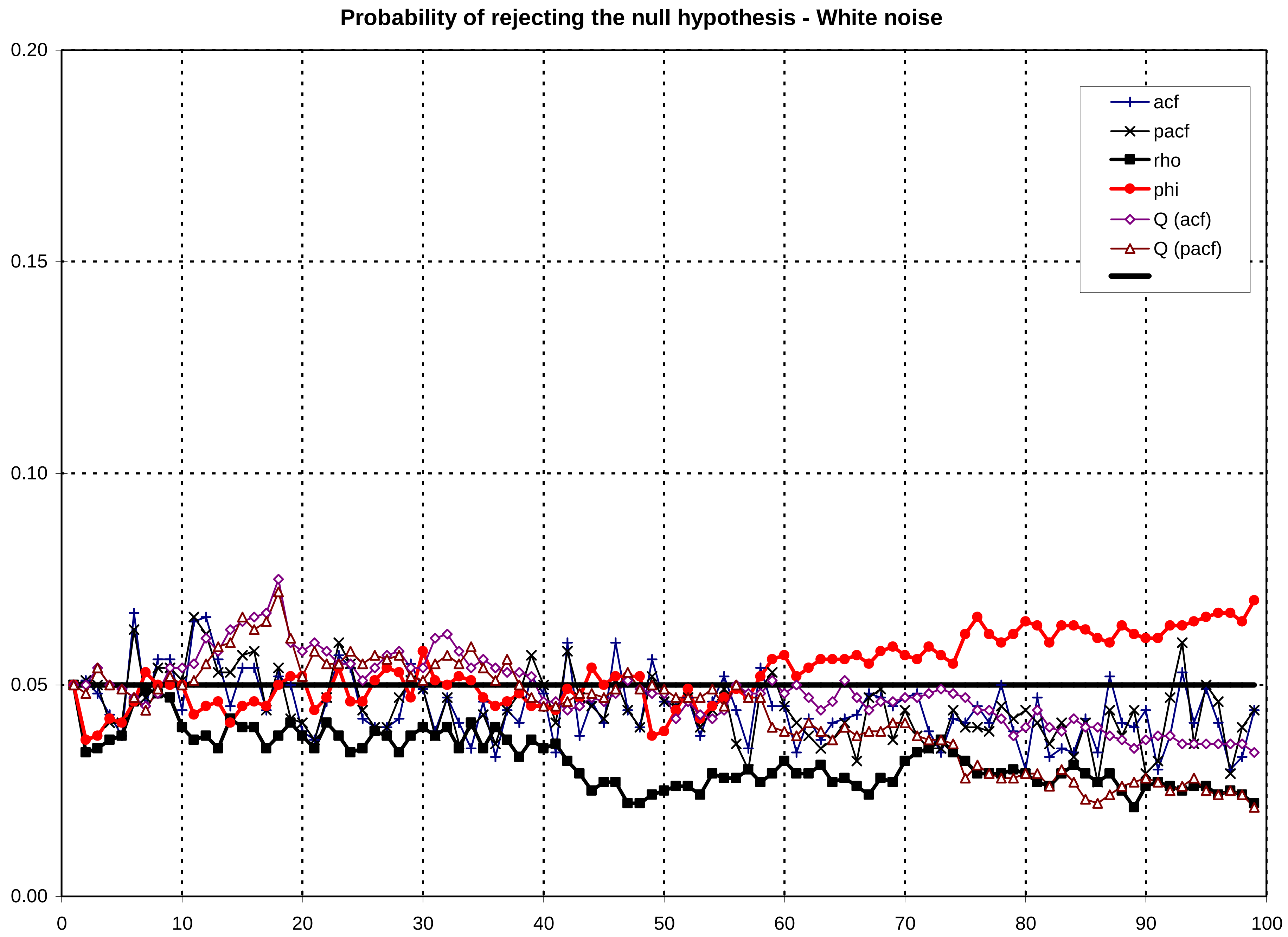
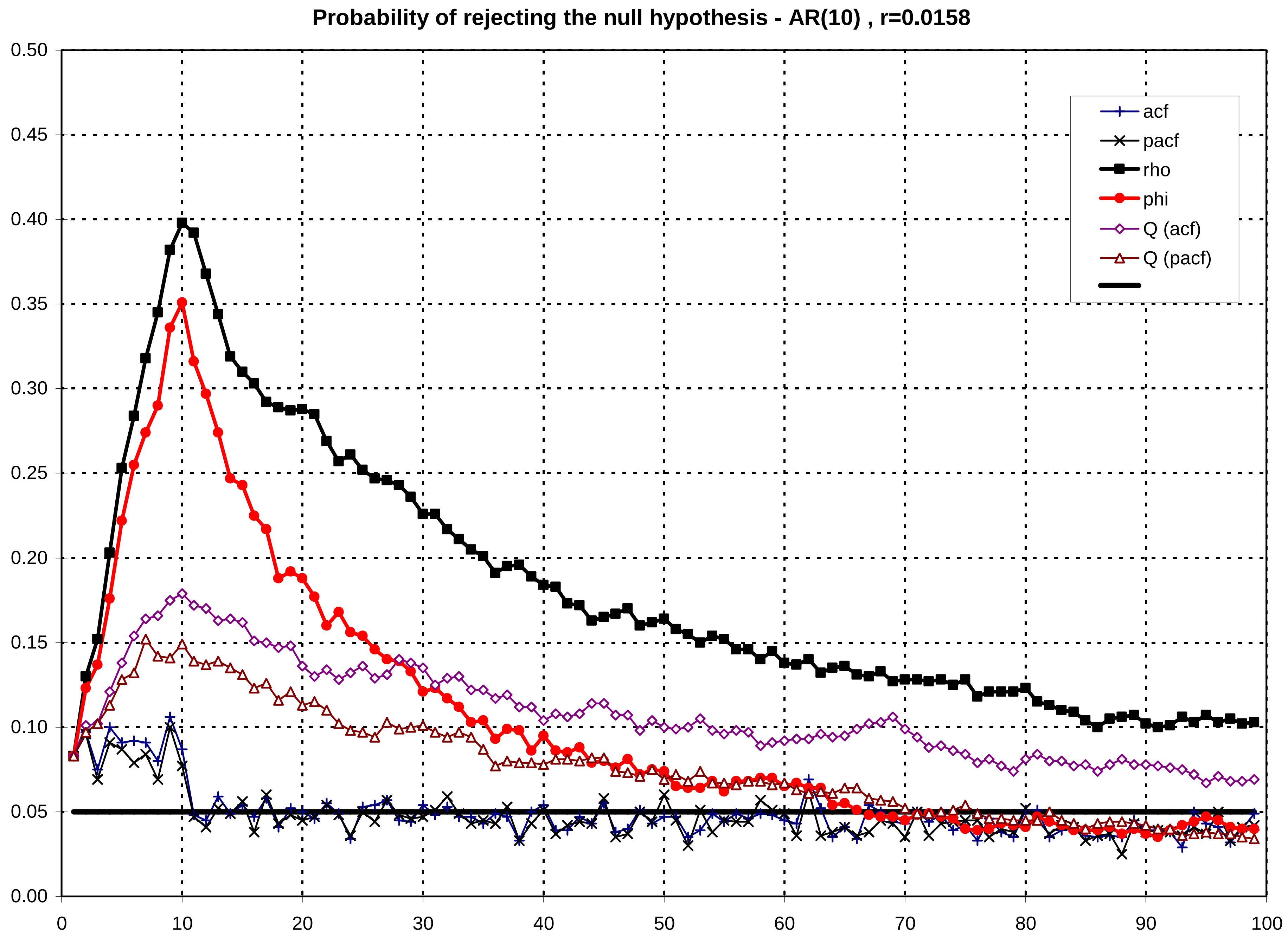
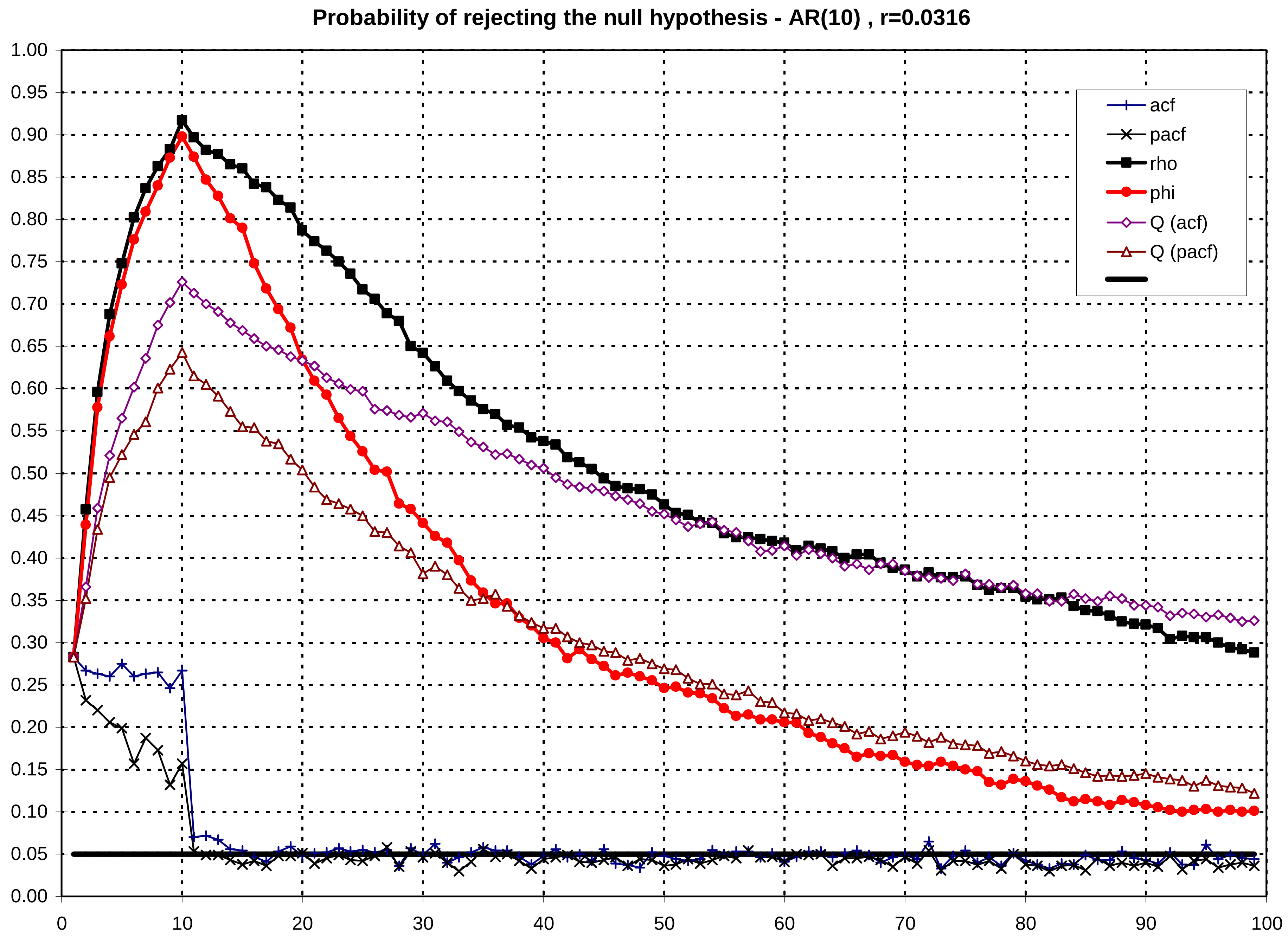

P(i) Statistic for Series S2, S3, and S4 at 50th Percentile
0.65
0.45
0.25
0.05
-0.15
-0.35
-0.55
-0.75
0
10
20
30
40
50
60
70
80
90
100
S3
S4
S2

Figure 3


Probability of rejecting the null hypothesis - White noise
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
70
80
90
100
acf
pacf
rho
phi
Q (acf)
Q (pacf)


Probability of rejecting the null hypothesis - AR(10) , r=0.0158
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
70
80
90
100
acf
pacf
rho
phi
Q (acf)
Q (pacf)


Probability of rejecting the null hypothesis - AR(10) , r=0.0316
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
70
80
90
100
acf
pacf
rho
phi
Q (acf)
Q (pacf)

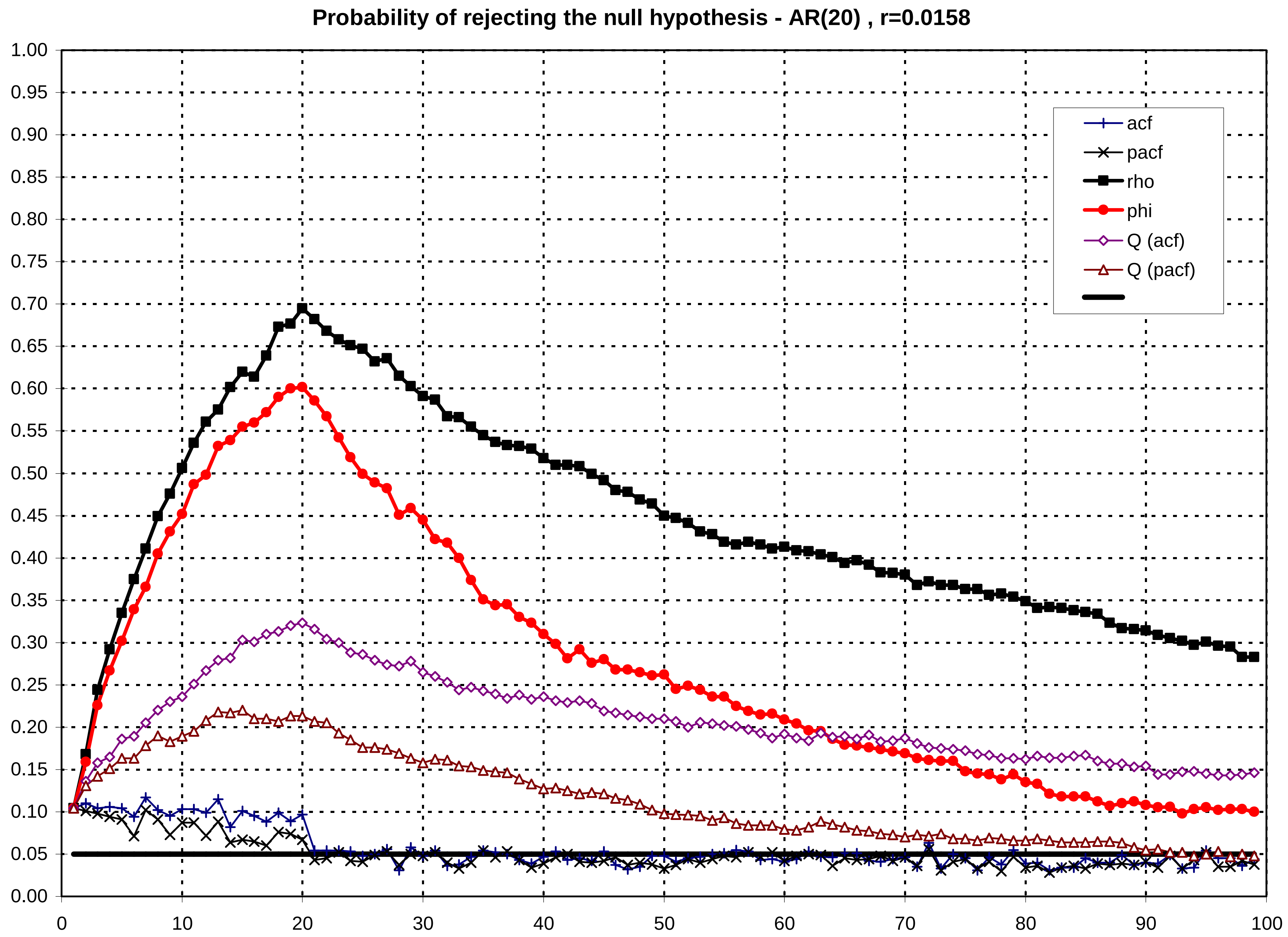

Probability of rejecting the null hypothesis - AR(20) , r=0.0158
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
70
80
90
100
acf
pacf
rho
phi
Q (acf)
Q (pacf)

Figure 7

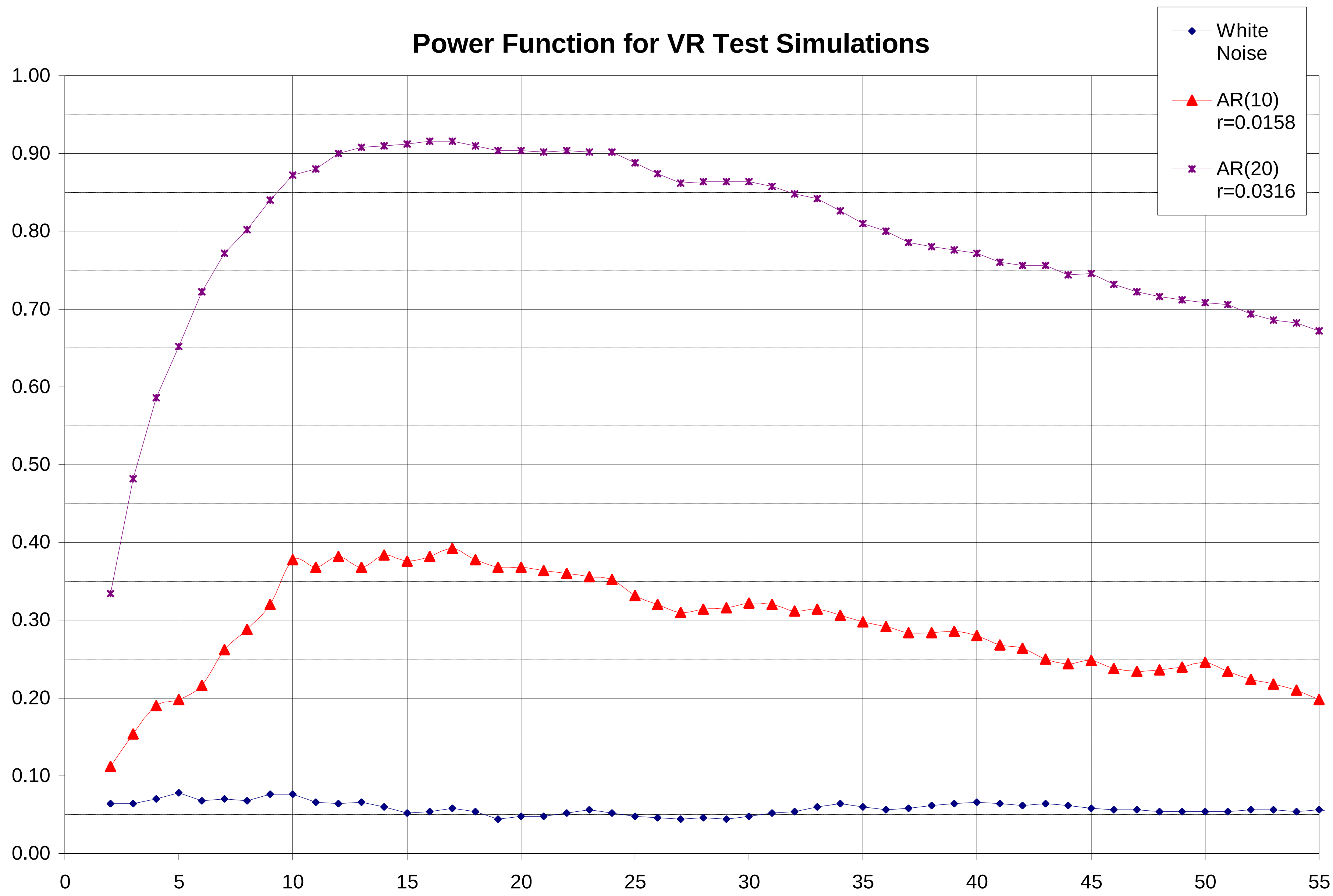

Power Function for VR Test Simulations
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
5
10
15
20
25
30
35
40
45
50
55
White Noise
AR(10) r=0.0158
AR(20) r=0.0316

Figure 7A

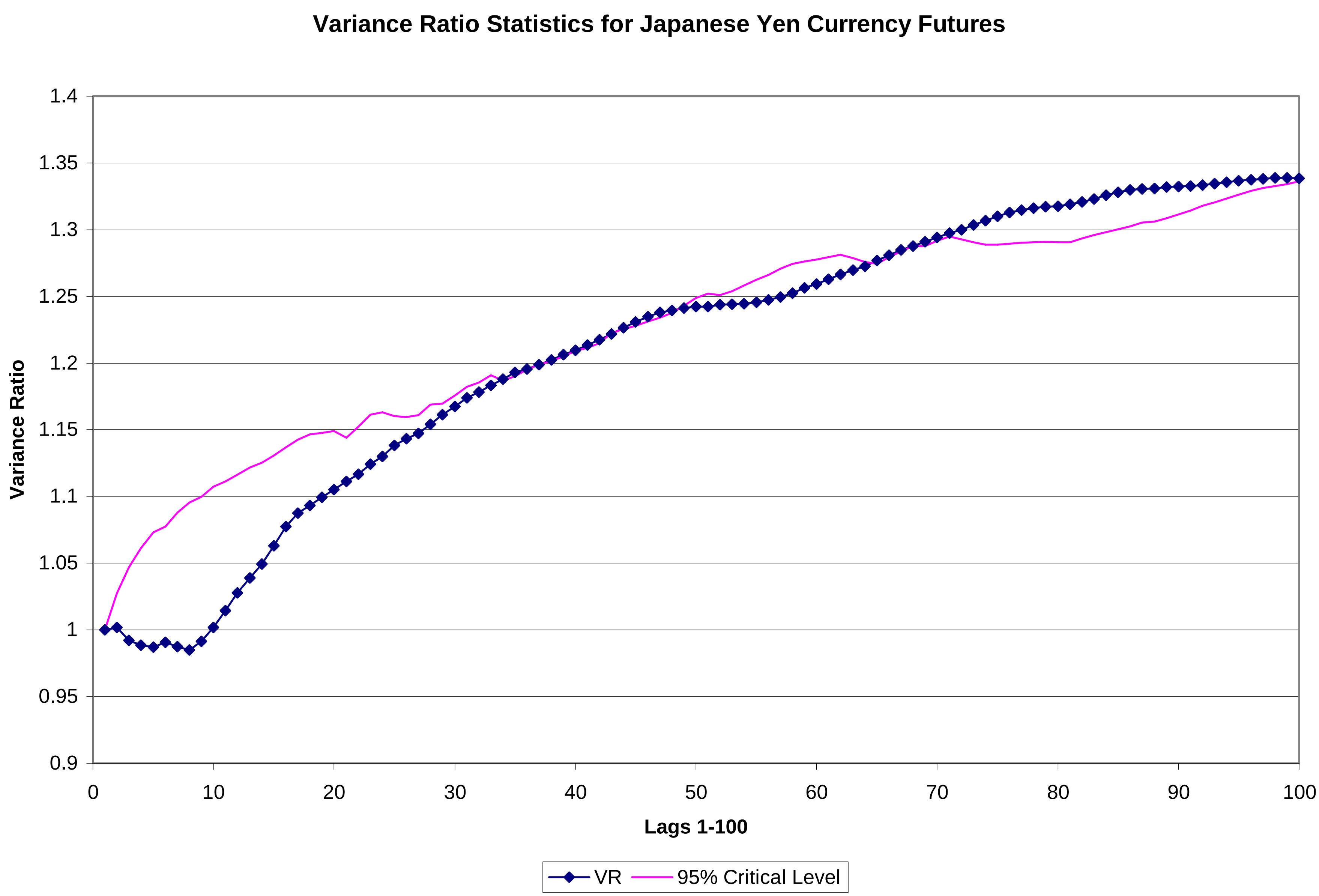
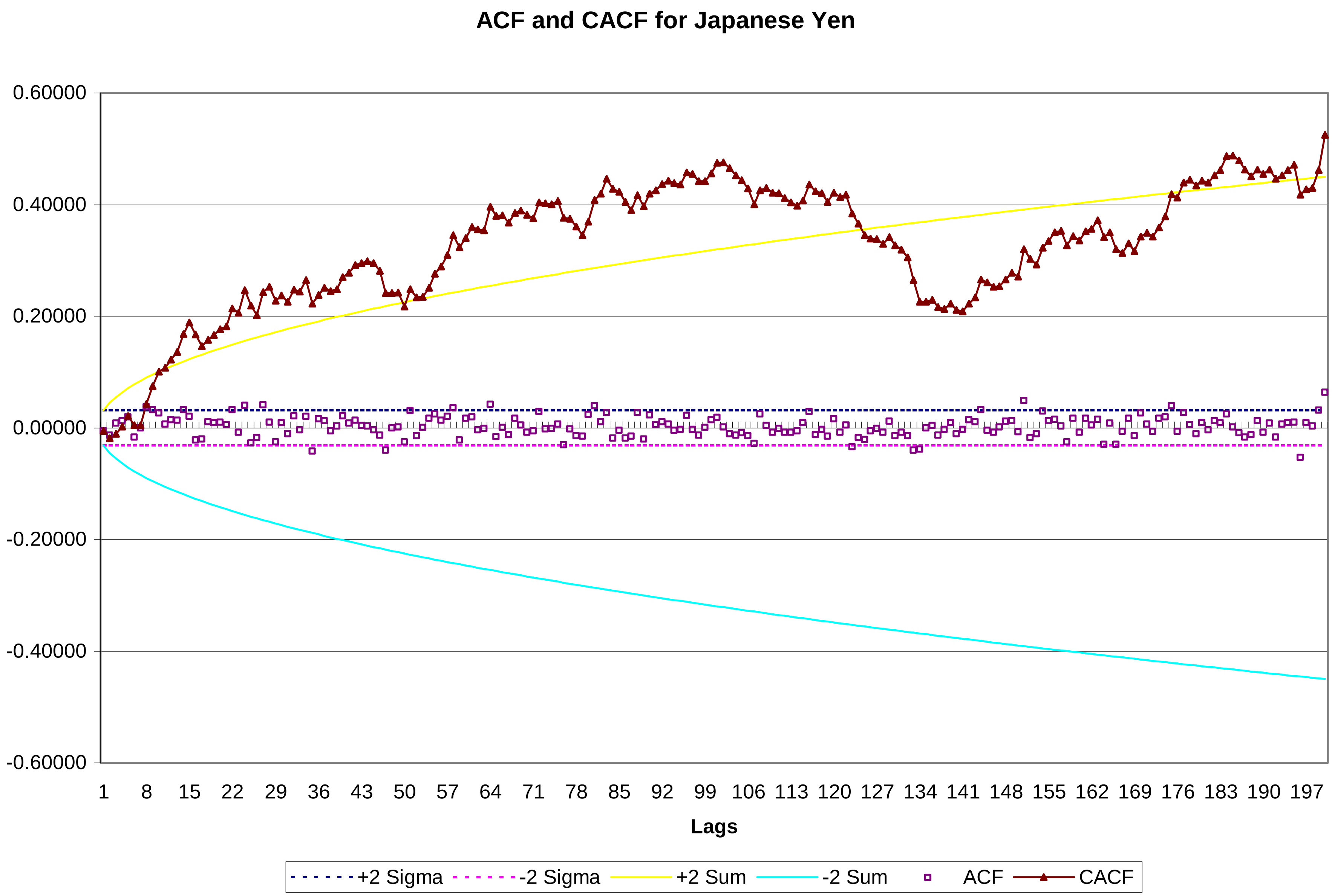
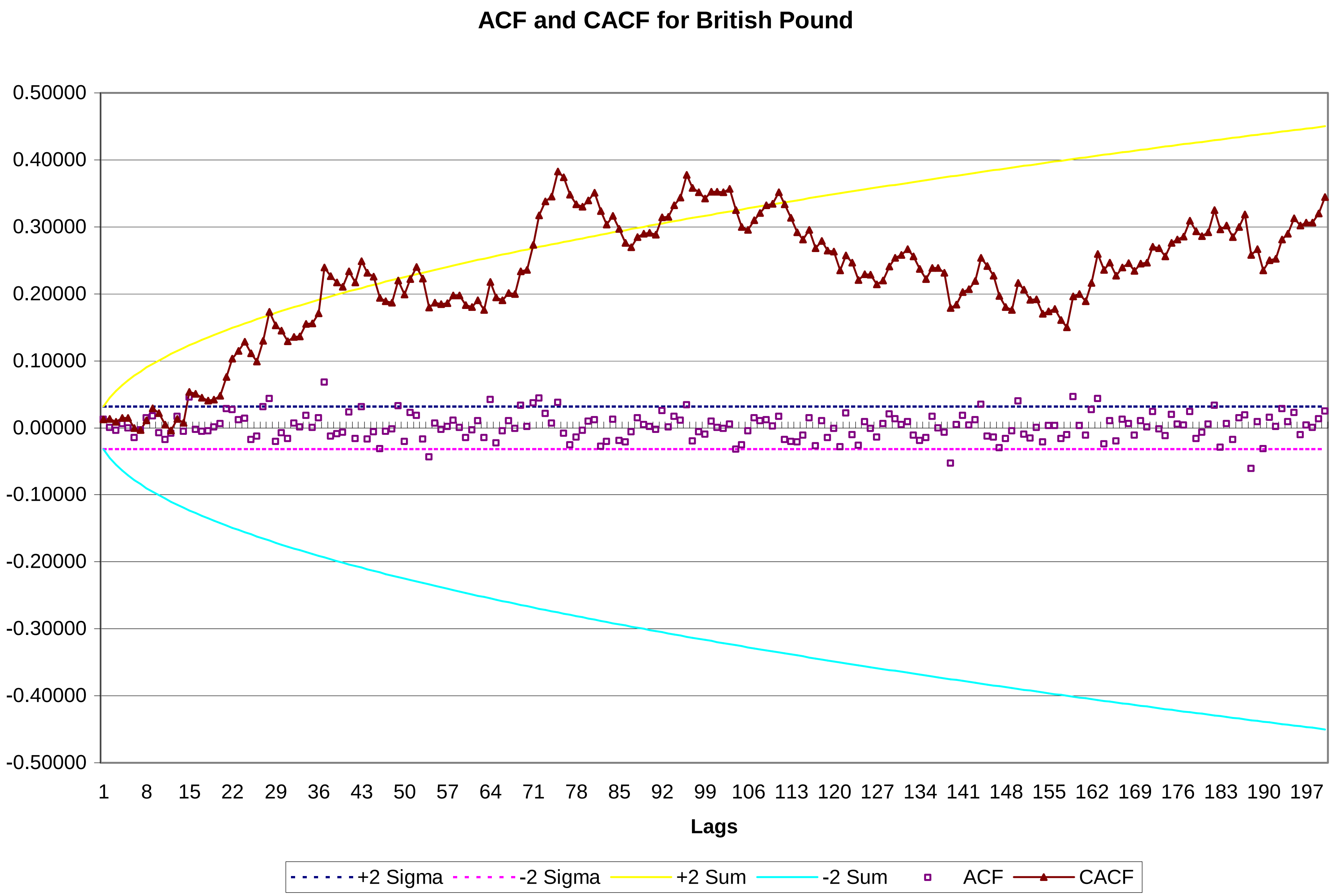

Variance Ratio Statistics for Japanese Yen Currency Futures
1.4
1.35
1.3
1.25
1.2
1.15
1.1
1.05
1
0.95
0.9
Variance Ratio
0
10
20
30
40
50
60
70
80
90
100
Lags 1-100
VR
95% Critical Level

Figure 8



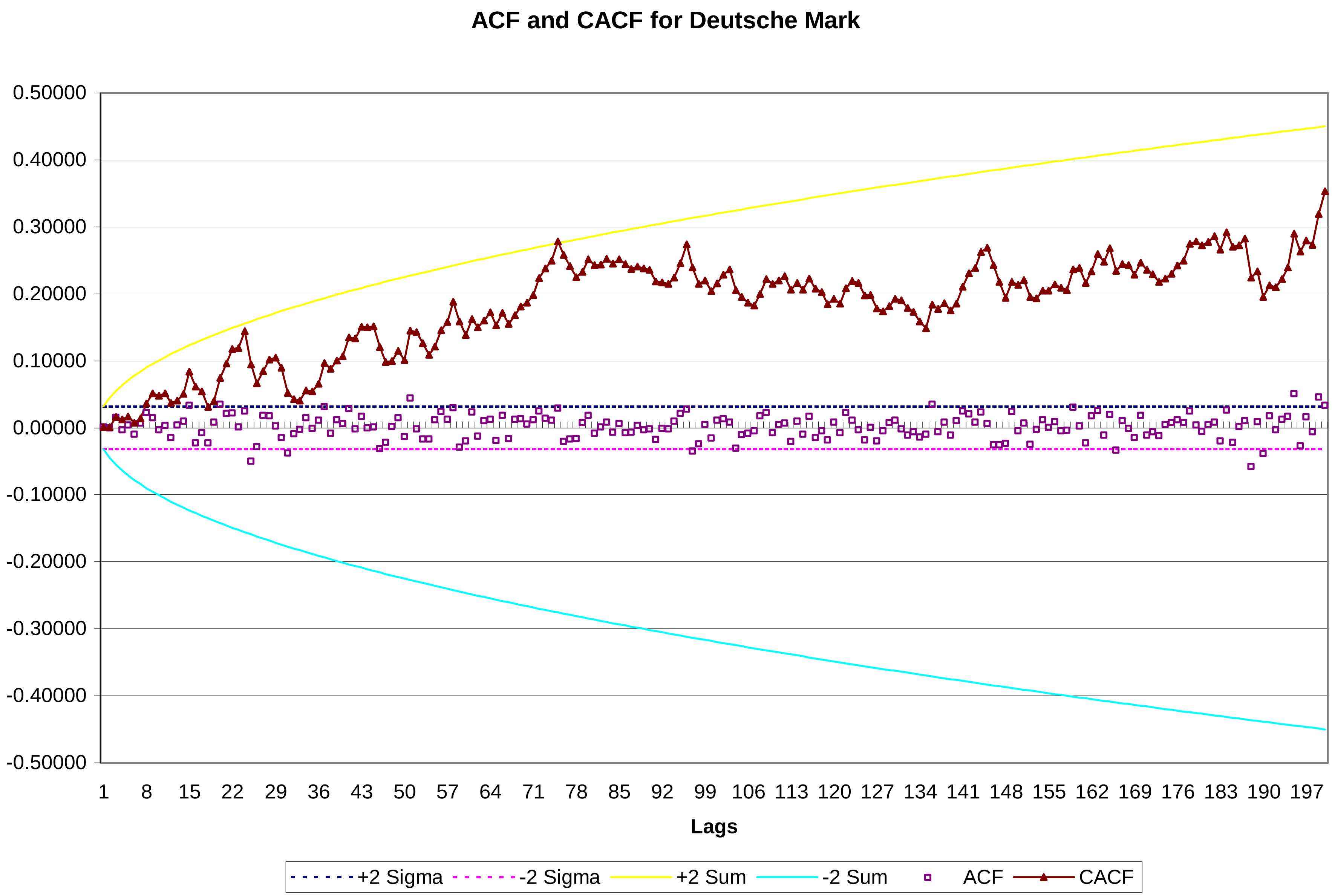
ACF and CACF for Japanese Yen
0.60000
0.40000
0.20000
0.00000
-0.20000
-0.40000
-0.60000
1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155 162 169 176 183 190 197
Lags
+2 Sigma
-2 Sigma
+2 Sum
-2 Sum
ACF
CACF

Figure 9



ACF and CACF for British Pound
0.50000
0.40000
0.30000
0.20000
0.10000
0.00000
-0.10000
-0.20000
-0.30000
-0.40000
-0.50000
1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155 162 169 176 183 190 197
Lags
+2 Sigma
-2 Sigma
+2 Sum
-2 Sum
ACF
CACF

Figure 10



ACF and CACF for Deutsche Mark
0.50000
0.40000
0.30000
0.20000
0.10000
0.00000
-0.10000
-0.20000
-0.30000
-0.40000
-0.50000
1
8
15
22
29
36
43
50
57
64
71
78
85
92
99
106
113
120
127
134
141
148
155
162
169
176
183
190
197
Lags
+2 Sigma
-2 Sigma
+2 Sum
-2 Sum
ACF
CACF

Figure 11

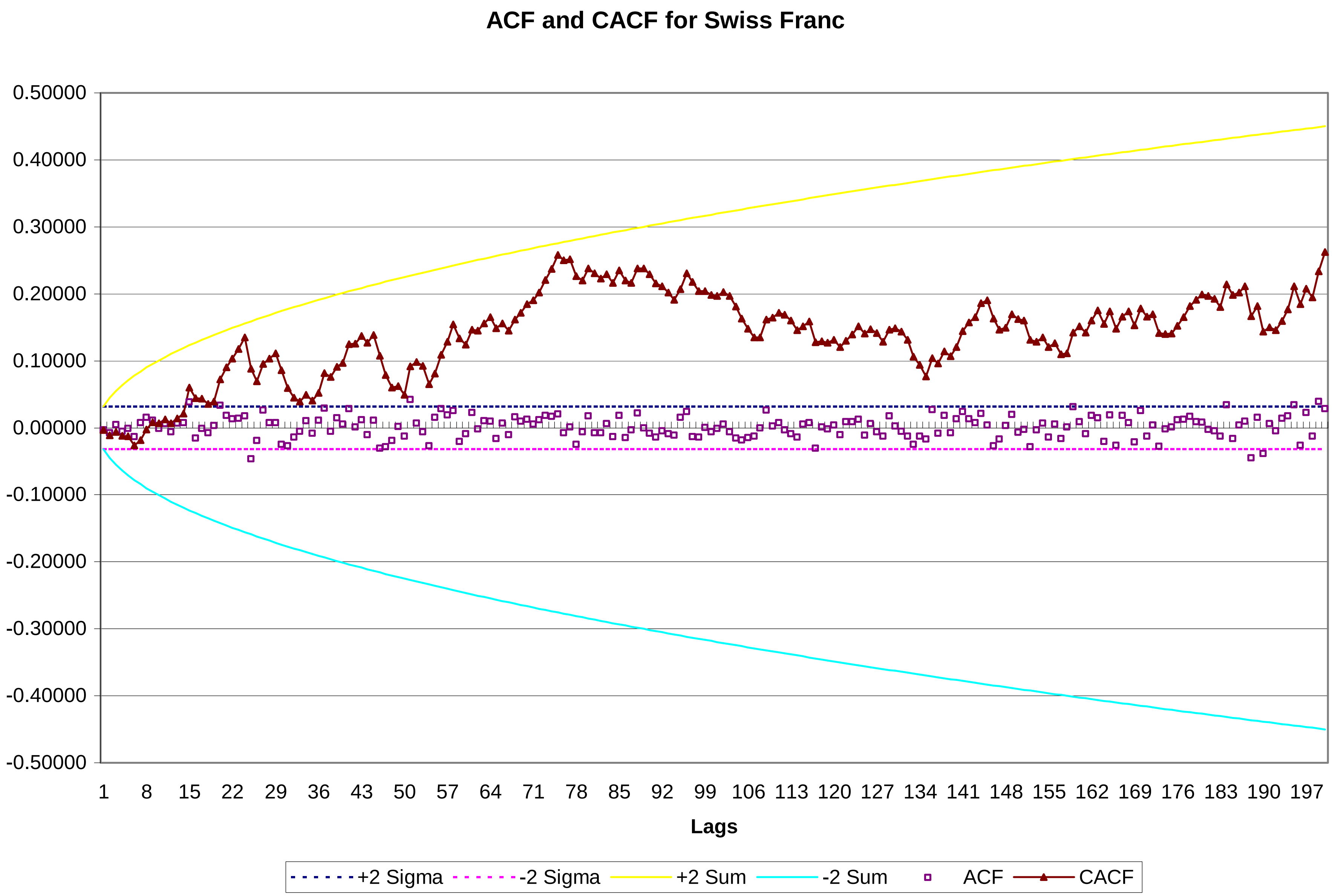
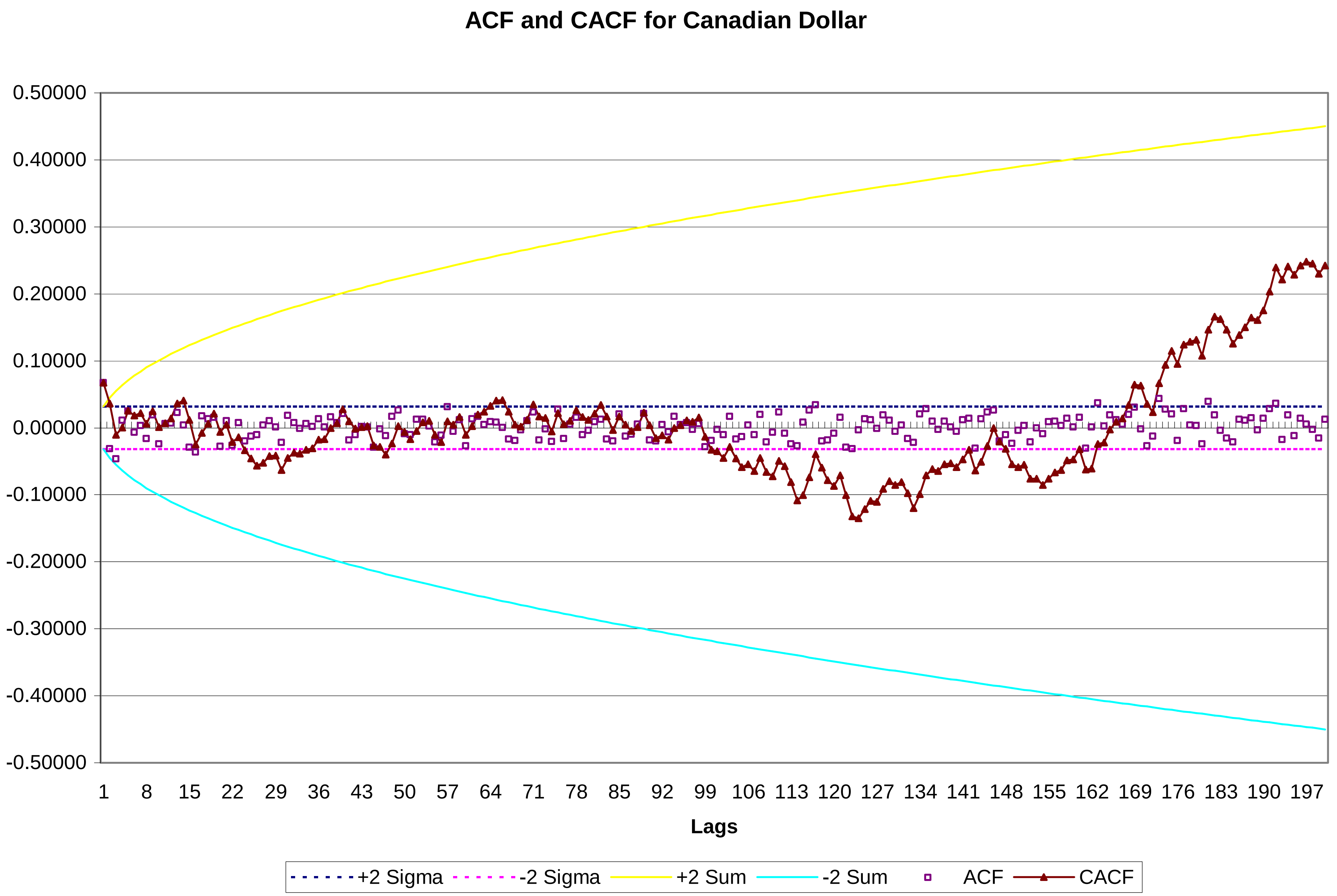

ACF and CACF for Swiss Franc
0.50000
0.40000
0.30000
0.20000
0.10000
0.00000
-0.10000
-0.20000
-0.30000
-0.40000
-0.50000
1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155 162 169 176 183 190 197
Lags
+2 Sigma
-2 Sigma
+2 Sum
-2 Sum
ACF
CACF

Figure 12


ACF and CACF for Canadian Dollar
0.50000
0.40000
0.30000
0.20000
0.10000
0.00000
-0.10000
-0.20000
-0.30000
-0.40000
-0.50000
1
8
15
22
29
36
43
50
57
64
71
78
85
92
99
106
113
120
127
134
141
148
155
162
169
176
183
190
197
Lags
+2 Sigma
-2 Sigma
+2 Sum
-2 Sum
ACF
CACF