ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mizrach, Bruce

Working Paper

Forecast Comparison in L2

Working Paper, No. 1995-24

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Mizrach, Bruce (1996) : Forecast Comparison in L2, Working Paper, No. 1995-24, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/94277

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Comments Solicited

Forecast Comparison in L_2

Bruce Mizrach*
Department of Economics
Rutgers University
First Draft: March 1991
This Draft: January 1995

Abstract:

This paper provides a comprehensive framework for comparing predictors of univariate time series in the mean square norm. Initially, the forecast errors are assumed to be unbiased, independent, and normally distributed. Each of these is progressively relaxed. A new heteroscedasticity and autocorrelation consistent statistic for forecast comparison is derived. Finite sample distributions are tabulated in a sequence of Monte Carlo exercises. Power is examined by comparing forecast errors from a moving average model with misspecified autoregressive alternatives.

KEY WORDS: Mean squared prediction error; robust forecast comparison;

---

*Address for editorial correspondence: Department of Economics, Rutgers University, 424 New Jersey Hall, New Brunswick, NJ 08903. e-mail: mizrach@gandalf.rutgers.edu, (908) 932-8261 (voice) and (908) 932-7416 (fax). I would like to thank Frank Diebold and Craig MacKinlay for helpful comments on an earlier draft. An earlier version of this paper was written while I was a visitor at the Wharton School and was presented at the 1992 North American Winter Meetings of the Econometric Society. A FORTRAN program implementing the robust forecast comparison is available upon request from the author.

1. Introduction

Econometric models are frequently compared on the basis of forecast accuracy.[1] Unfortunately, nearly all these evaluations rely upon heuristic approaches, devoid of any formal statistical theory. This paper seeks to provide statistical foundations for forecast comparisons when mean squared prediction error (MSPE) is the criterion for predictive accuracy.

Applied forecasting exercises generally make comparisons with is a ratio of MSPEs, often called the Theil (1966) U-statistic. I show below that if the forecast errors are correlated, the U-statistic has very different properties than the standard F-distribution for the ratio of variances. This ratio can yield very misleading inference.

Granger and Newbold (1986) and Meese and Rogoff (1988) were among the first papers to go beyond heuristic statistics like the Theil-U. Granger and Newbold looked at a very restricted case, that of normally distributed, serially uncorrelated and unbiased forecast errors. Meese and Rogoff (1988) then extended Granger and Newbold's work to the time series case. This paper generalizes mean square norm inference by relaxing all three assumptions.

I propose a new robust forecast comparison statistic that only assumes bounds on the first four moments and mixing. The robust statistic performs admirably in finite samples. The statistic is properly sized in normal and non-normal populations, even in the case of leptokurtosis that badly distorts statistics that assume normality.

I examine the power of the new statistic by comparing forecasts from a moving average model with misspecified autoregressive alternatives. The robust statistic rejects nearly $q\%$ of the time alternatives that increase the forecast error variance by $q\%$ in a sample of 25. In a large sample of 250, these results are two to four time better.

I state the decision problem formally in Section 2. For loss functions in L_2 the risk associated with the forecast is the MSPE. The null hypothesis is then to differentiate sample measures of the risk. Section 3 takes up the case looked at by Granger and Newbold. I begin by briefly examining the workhorse of the empirical literature, the Theil-U. Section 4 is the extension to the time series case. Section 5 discusses bias. Section 6 develops the robust statistic. Sections 7 and 8 present the Monte Carlo and power comparisons. A conclusion and guide to further research is in Section 9. Some derivations involving bivariate moments and cumulants are relegated to an appendix.

2. Mean Squared Prediction Error

In many exercises in applied econometrics, the data analyst is interested in evaluating a number of candidate models. Particularly with time series models, approximations

[1] Some representative examples include Armstrong's (1983) analysis of stock company earnings, McNees' (1990) scorecard for macroeconometric models, and Mizrach's (1992b) effort to improve upon the random walk in forecasting foreign exchange rates.

to the data generating mechanism are compared based on their ability to predict either in or out of sample. I will use the notation y to denote the vector of variables to be forecasted, and let z_i be the ith predictor of y. The corresponding forecast errors will be denoted e_i. In this paper, I follow the applied literature in looking at loss functions in the mean square norm. Consider the function,

$$\ell(y - z_i) = \langle e_i, e_i \rangle . \tag{1}$$

where $\langle ., . \rangle$ denotes the inner product on R^n,

$$\langle e_i, e_i \rangle = \sum_{j=1}^{n} e_{ij}^2. \tag{2}$$

The *risk function* for the data analyst's problem is just the expectation of (2),

$$R(z_i) = E\left[\sum_{j=1}^{n} e_{ij}^2\right]. \tag{3}$$

The objective of the analyst's statistical decision problem is to choose a predictor that will minimize the risk (3). As is widely known, the conditional expectation of y will minimize quadratic loss. The sample analog of (3) is commonly known as the mean squared prediction error (MSPE) of the forecast z_i,

$$MPSE_i = 1/n \sum_{j=1}^{n} e_{ij}^2. \tag{4}$$

Given two estimators of the conditional mean, z_1 and z_2, the data analyst faces the question of whether the predictor with the lower risk, z_1 by our convention, is significantly better than the other. This problem can be stated as the null hypothesis,

$$H_0 : R(z_1) = R(z_2). \tag{5}$$

.

Typically, the data exercise will involve forecasts from rival predictors. H_0 will be used to evaluate whether the forecast improvement is statistically significant.[2] The next four sections are concerned with testing the hypothesis (5) under progressively weaker distributional assumptions. Section 3 begins with a baseline case of a bivariate normal, serially uncorrelated, zero mean population.

3. The Baseline Case

I initially assume that the two forecast errors are draws from a bivariate normal population (E_1, E_2) with common means $\mu_1 = \mu_2 = 0$, own variances σ_1^2 and σ_2^2 and covariance, $\sigma_{12} = \rho_E \sigma_1 \sigma_2$, where ρ_E is the correlation coefficient. In part 1, I show

[2]Mean squared error as a forecast criterion has been criticized by Clements and Hendry (1992) for ignoring the complete predictive density. As is typical in many statistical problems, we must sacrifice efficiency to develop a statistic robust to a large range of population assumptions.

that the Theil-U can result in very misleading inference, and in parts 2, 3 and 4, I introduce a number of tests for this baseline case. In part 5, I undertake a comparison of finite sample properties.

3.1 The Theil-U Statistic

A common way of representing the outcome of applied forecasting exercises is to report the ratio of the MSPEs,

$$TU = \frac{1/n \sum_{j=1}^{n} e_{1j}^2}{1/n \sum_{j=1}^{n} e_{2j}^2}. \tag{6}$$

If z_2 is the no change forecast, then $e_{2j} = y_j$ for all j, and (6) becomes the Theil (1966) U-statistic.[3]

With unbiased errors, taking expectations in the numerator and denominator of (6) yields the ratio of the population variances,

$$\frac{E\left[1/n \sum_{j=1}^{n} e_{1j}^2\right]}{E\left[1/n \sum_{j=1}^{n} e_{2j}^2\right]} = \frac{\sigma_1^2}{\sigma_2^2}. \tag{7}$$

The null hypothesis (5) can be restated as

$$H_0 : \sigma_1^2 = \sigma_2^2. \tag{8}$$

While the numerator and denominator of (7) are each distributed as chi-square random variables with n degrees of freedom, the ratio (7) will only have the standard F-distribution if the forecast errors are independent. When $\rho_E \neq 0$, the statistic TU is distorted from the F-distribution by a factor proportional to the square of the correlation between e_1 and e_2.

The joint frequency distribution of the sample forecast error variances, s_1 and s_2, and correlation coefficient, r, is proportional to (see e.g. Kendall and Stuart (1963) p.385)

$$dF \propto \exp\left[-\frac{n^2}{2(1-\rho_E^2)}\left\{\frac{s_1^2}{\sigma_1^2} - \frac{2\rho_E^2 r s_1 s_2}{\sigma_1 \sigma_2} + \frac{s_2^2}{\sigma_2^2}\right\}\right] s_1^{n-2} s_2^{n-2} (1-r^2)^{(n-4)/2} ds_1 ds_2 dr \tag{9}$$

.

Evaluating the Jacobian of the transformation gives us the constant term,

$$dF = \frac{n^{n-1}}{\pi \sigma_1^{n-1} \sigma_2^{n-1} (1-\rho_E^2)^{(n-1)/2} \Gamma(n-2)} \exp[.] ds_1 ds_2 dr \tag{10}$$

[3] I will refer throughout the paper to any ratio of MSPEs as the Theil-U. This matches Theil's usage only in the case where z_2 is the no change forecast.

Define the change of variables $\nu = (s_1/s_2)$. Under the null, ν is just the ratio of the root mean squared errors, the square root of the Theil-U. (10) simplifies to

$$dF \propto \frac{\nu^{n-2}(1-r^2)^{(n-4)/2}}{(1-2\rho_E r\nu+\nu^2)^{n-1}} dr d\nu \tag{11}$$

Integrating out $\int_{-1}^{1} r dr$, Bose (1935) and Finney (1938) have obtained the result

$$dF = \frac{2(1-\rho_E^2)^{(n-1)/2}}{Beta\left\{(n-1)/2,(n-1)/2\right\}} \frac{\nu^{n-2}}{(1+\nu^2)^{n-1}} \left\{1 - \frac{4\rho_E^2\nu^2}{(1+\nu^2)^2}\right\} d\nu. \tag{12}$$

Mizrach (1992a) shows that as $\rho_E \to 0$, we approach the standard F-distribution, but otherwise, dF in (12) is strictly less than the standard-F in the right tail of the distribution, for $\rho_E \neq 0$. The empirical size is well below the nominal size of the standard-F at traditional significance levels. Monte Carlo results in Mizrach (1992a) indicate a high rate of Type II error for the typical 95% confidence test. With $\rho_E = 0.9$, the F-statistic never rejects the null hypothesis in samples of 10, 25 and 50 observations. Although one could construct a conservative test using the F-distribution, the Theil-U will not generally be very useful in data analysis, even merely as a summary measure. We can now turn to more statistically rigorous testing procedures.

3.2 Tests Relying on the Sample Covariance

Granger and Newbold (1986) and Meese and Rogoff (1988) noticed an easy way to test the null through the following transformation. Let $U = E_1 - E_2$, and $V = E_1 + E_2$. Then, (U, V) has a bivariate normal distribution with parameters $E[U] = \mu_1 - \mu_2 = 0$, , $E[V] = \mu_1 + \mu_2 = 0$, $var(U) \equiv \sigma_U^2 = \sigma_1^2 + \sigma_2^2 - 2\rho_E\sigma_1\sigma_2$, $var(V) \equiv \sigma_V^2 = \sigma_1^2 + \sigma_2^2 + 2\rho_E\sigma_1\sigma_2$, and $cov(U,V) \equiv \sigma_{UV} = \rho\sigma_U\sigma_V$. In terms of the original population, $\sigma_{UV} = \sigma_1^2 - \sigma_2^2$. If the mean squared prediction errors in the original population are equal, then the covariance in the transformed population must be zero. Now restate the null hypothesis in terms of the transformed population,

$$H_0: \; cov(U,V) \; = \; 0. \tag{13}$$

A direct way to test the null is to use the sample covariance around the population means,

$$1/n \sum_{j=1}^{n} (u_j - [\mu_1 - \mu_2])(v_j - [\mu_1 + \mu_2]) = 1/n \sum_{j=1}^{n} u_j v_j \equiv s_{UV}, \tag{14}$$

where $u_j = (e_{1j} - e_{2j})$, $v_j = (e_{1j} + e_{2j})$. The variance of this covariance is

$$var(s_{UV}) = 1/n^2 E\left[\left(\sum_{j=1}^{n} u_j v_j\right)^2\right] - \sigma_{UV}^2. \tag{15}$$

With independent draws from the transformed population, $E[u_j u_{j\neq k}] = 0$, $E[v_j v_{j\neq k}] = 0$, and $E[u_j v_k] = 0$, for all $j, k = 1, ..., n$. This simplifies (15) to

$$var(s_{UV}) = 1/n^2 E\left[\sum_{j=1}^{n} u_j^2 v_j^2\right]. \tag{16}$$

In terms of the original population,

$$var(s_{UV}) = 1/n^2 E\left[\sum_{j=1}^{n} e_{1j}^4 + 2e_{1j}^2 e_{2j}^2 + e_{2j}^4\right] = \left[3(\sigma_1^2)^2 + 2(1+2\rho_E^2)\sigma_1^2\sigma_2^2 + 3(\sigma_2^2)^2\right]/n \tag{17}$$

It follows that the statistic

$$\frac{s_{UV}}{\left[\sum_{j=1}^{n} u_j^2 v_j^2/n^2\right]^{1/2}} \overset{asy.}{\sim} N(0,1). \tag{18}$$

Dealing with the sample means complicates things a bit. However, the asymptotic variance of

$$\bar{s}_{UV} \equiv 1/n \sum_{j=1}^{n} (u_j - \bar{u})(v_j - \bar{v}) \tag{19}$$

where $\bar{u} = 1/n \sum_{j=1}^{n} u_j$, $\bar{v} = 1/n \sum_{j=1}^{n} v_j$, is the same as that of s_{UV}. The finite sample approximation is accurate to $o(n^{-1/2})$.

The standard precautions are called for in using a standard error to test a statistical hypothesis. It is only justified in the case where the test statistic tends to normality. In financial applications, this may not be satisfied. Secondly, a test on a moment depends on population moments of twice that order. Fourth moment restrictions will often be needed to test the equality of the sample MSPEs.

3.3 Tests Using the Sample Correlation Coefficient

Given our assumption of normality, a necessary and sufficient condition for $cov(U, V) = 0$ is that $\rho = 0$. An asymptotically equivalent testing procedure can therefore be devised around the sample correlation coefficient in place of the covariance. The advantage here is that we have some existing finite sample theory here to rely upon. Define the (noncentral) sample correlation coefficient as

$$r \equiv \frac{\sum_{j=1}^{n} u_j v_j}{\left[\sum_{j=1}^{n} (u_j)^2 \sum_{j=1}^{n} (v_j)^2\right]^{1/2}}. \tag{20}$$

The statistic

$$r/\sqrt{var(r)}, \tag{21}$$

using the standard result, $var(r) = 1/n(1-\rho^2)^2$, could be utilized to test the null, but the use of a standard error is not recommended here as the sampling distribution

of r tends to normality very slowly, particularly for large values of ρ.[4] For large values of ρ and moderate sample sizes, the distribution is highly skewed. For the case $\rho = 0$, Fisher (1915) derived the exact finite sample distribution of r. Fisher showed that

$$T = r\sqrt{n-2}/\sqrt{1-r^2}, \tag{22}$$

has the Student t-distribution, with $n-1$ degrees of freedom. [5] Meese and Rogoff make use of this statistic. Granger and Newbold note that this test is uniformly most powerful unbiased under our assumptions in this section. In the case where is non-zero, Fisher (1921) has made use of a transformation[6] of ρ which tends to normality considerably faster than r. Let

$$\eta = 1/2 \log[(1+r)/(1-r)] \tag{23}$$

Fisher shows that in large samples that η is distributed $N(\mu_\eta, \sigma_\eta^2)$, where $\mu_\eta = 1/2 \log[(1+\rho)/(1-\rho)]$ and $\sigma_\eta = 1/\sqrt{n-3}$. It follows that

$$(\eta - \mu_\eta)/\sigma_\eta \overset{asy.}{\sim} N(0,1). \tag{24}$$

To the best of my knowledge, the statistic (24) first appeared in Mizrach (1990).

3.4 A Likelihood Ratio Test

For completeness, I chose to include a likelihood ratio test. For the bivariate normal case, the log likelihood is given by

$$\log L(u,v;\theta = (\rho, \sigma_U, \sigma_V)) = -\log\left[2\pi\sigma_U\sigma_V(1-\rho^2)^{-1/2}\right]^n + \left[-\frac{1}{2(1-p)^2}\left\{\sum_{j=1}^n u_j^2/\sigma_U^2 - 2\rho\sum_{j=1}^n u_j v_j/\sigma_U\sigma_V + \sum_{j=1}^n v_j^2/\sigma_V^2\right\}\right], \tag{25}$$

where I have used the population assumptions $\mu_1 = \mu_2 = 0$. The maximum likelihood estimates, denoted with hats, are the sample moments. The null hypothesis is that $H_0 : \sigma_U = \hat{\sigma}_U; \sigma_V = \hat{\sigma}_V; \rho = \hat{\rho}$; which I test against the alternative $H_1 : \sigma_U = \sigma_V = \sqrt{2}; \rho = 0$; Denote the sample likelihood evaluated at the maximum likelihood estimates as $L_n(u, v, \hat{\theta}(\hat{\rho}, \hat{\sigma}_U, \hat{\sigma}_V))$. Using (25), define

[4] See the example in Hoel (1954) p.123.

[5] By using the common population means in place of the sample means, we pick up a degree of freedom.

[6] In Fisher (1921), the expansion actually made use of the hyperbolic tangents of η. Let $\xi = 1/2 \log[(1+\rho)/(1-\rho)]$. Fisher showed that $\mu_1'(\eta - \xi) = \tanh(\xi)/2(n-1)$ and that $var(\eta - \xi) = 1/(n-1) + [4 - \tanh^2(\xi)/2(n-1)^2]$ which for small ρ and large n, $1/(n-1) + 2/(n-1)^2 \approx 1/(n-3)$, yielding the equation in the text.

$$\Lambda_n \equiv \frac{L_n(u, v; \theta = (0, \sqrt{2}, \sqrt{2}))}{L_n(u, v; \hat{\theta} = (\hat{\rho}, \hat{\sigma}_U, \hat{\sigma}_V))}. \tag{26}$$

The statistic

$$-2\log(\Lambda_n) \overset{asy.}{\sim} \chi^2(3). \tag{27}$$

is distributed asymptotically, χ^2 with 3 degrees of freedom.

3.5 Finite Sample Comparisons

Apart from the statistic (22) for the sample correlation coefficient in an independent bivariate normal population, the distribution theory obtained for our test statistics is all asymptotic. I will need to examine the finite sample properties of five statistics, one for the covariance, (18), three for the correlation coefficient, (21), (22), and (24), and one for the likelihood ratio, (25). I included the exact finite sample statistic (22) as a check on our results. I generated 20,000 samples of size $n = 10$, 25, 50, 100 and 250 using the IMSL subroutines CHFAC and RMNVN. CHFAC obtains the Cholesky factorization of the covariance matrix of the transformed errors,

$$cov(U, V) = \begin{bmatrix} \sigma_U^2 & \sigma_{VU} \\ \sigma_{UV} & \sigma_V^2 \end{bmatrix}. \tag{28}$$

By choosing $\sigma_U^2 = \sigma_V^2 = 1$, the correlation coefficient equals the covariance, $\rho = \sigma_{UV}$. I look only at the case in which $\rho = 0$. The subroutine RMNVN then generates samples of random numbers using user supplied seed values. I found that choosing the optimization level at 5 through the subroutine RNOPT was critical for even large samples to closely approximate population values. The results of the Monte Carlo simulation are in Tables 1.1 and 1.2. All of the statistics are within sampling error of their nominal sizes, even in samples as small as 10. The skewness in ρ that sometimes distorts (21) is apparently not a problem for small values of ρ. The LR statistic (25) performs as well as any of the statistics with any asymptotic normal distribution. There is little basis to discriminate between the five statistics, but as we turn to more complicated population assumptions, the choice between statistics will become more sharp.

4. The Time Series Case

The statistics in Section 3 are in general only valid for one step ahead forecasts. In multi-step prediction, even optimal forecasts will be serially correlated. An important extension then will be to consider statistics for the time series case. In the case of a $(k+1)$-step ahead forecast, the data analyst can be relatively confident of the

autocorrelation structure of the forecast errors.[7] I will proceed under the assumption that the forecast errors are a dependent process of order k. This information will be used to construct a consistent estimate of the variance of $cov(U,V)$. A large sample result is then obtained. Without the orthogonality assumptions made in arriving at (16), the variance of the covariance in the transformed population will include terms involving cross products of lagged $U's$ and $V's$.

$$var(s_{UV}) = 1/n^2 E\left[\sum_{j=1}^{n}\sum_{m=1}^{n} u_j v_j u_m v_m\right] - E[s_{UV}]^2 \tag{29}$$

The bracketed quadruple product in (29) will be crucial for the next three sections. I expand this term in the Appendix. Normality enables me to drop out the first and second lines of (A.8). Unbiasedness eliminates the fourth and fifth lines. After making the change in variables, I find

$$var(s_{UV}) = 1/nE\left[\sum_{j=1}^{n} \gamma_{UU}(t)\gamma_{VV}(t) + \gamma_{UV}(t)\gamma_{VU}(t)\right], \tag{30}$$

where for

$$\begin{array}{lll} & t>0 & t\leq 0 \\ \gamma_{UU}(t) \equiv & /n\sum_{j=t+1}^{n} E[u_j u_{j-t}], & \gamma_{UU}(t) \equiv 1/n\sum_{j=-t+1}^{n} E[u_{j+t}u_j], \\ \gamma_{VV}(t) \equiv & 1/n\sum_{j=t+1}^{n} E[v_j v_{j-t}], & \gamma_{VV}(t) \equiv 1/n\sum_{j=-t+1}^{n} E[v_{j+t}v_j], \\ \gamma_{VV}(t) \equiv & 1/n\sum_{j=t+1}^{n} E[u_j v_{j-t}], & \gamma_{UV}(t) \equiv 1/n\sum_{j=-t+1}^{n} E[u_{j+t}v_j], \\ \gamma_{VU}(t) \equiv & 1/n\sum_{j=t+1}^{n} E[v_j u_{j-t}], & \gamma_{VU}(t) \equiv 1/n\sum_{j=-t+1}^{n} E[v_{j+t}u_j]. \end{array}$$

The next equation is:

$$\frac{s_{UV}}{\left[1/n\sum_{t=-k}^{k}\left[1-(|t|/n)\hat{\gamma}_{UU}(t)\hat{\gamma}_{VV}(t)+\hat{\gamma}_{UV}(t)\hat{\gamma}_{VU}(t)\right]\right.} \overset{asy.}{\sim} N(0,1). \tag{31}$$

Note that $\gamma_{UV}(0) = \sigma_{UV}$, $\gamma_{UU}(0) = \sigma_U^2$, and $\gamma_{VV}(0) = \sigma_V^2$. If the process $\{u_j v_j\}$ is stationary, then $\gamma_{UU}(t) = \gamma_{UU}(-t)$, etc. Given our dependence assumptions, terms for which $t > k$ will be zero. Denoting with hats the sample analogs[8] of equations (31), it follows that

$$\frac{s_{UV}}{\left[1/n\sum_{t=-k}^{k}\left[1-(|t|/n)\hat{\gamma}_{UU}(t)\hat{\gamma}_{VV}(t)+\hat{\gamma}_{UV}(t)\hat{\gamma}_{VU}(t)\right]\right.} \overset{asy.}{\sim} N(0,1) \tag{32}$$

.

If the order of dependence is unknown, the analyst must use the robust procedures described in Section 6. A closely related statistic for correlated normal populations

[7] It is a standard result that for a $(k+1)$-step ahead forecast, the forecast errors will have a moving average representation of order k. See e.g. Box and Jenkins (1976) p.158-60.

[8] I use the biased estimates of the autocovariances, e.g. $\hat{\gamma}_{UU}(t) = 1/n\sum_{j=t+1}^{n} u_j u_{j-t}$, and the weighting structure $(1-(|t|/n))$ because then the expression in the numerator is a positive semi-definite function. See Priestley (1981) p.323-4 and the discussion at the end of Section 6. Although not acknowledged notationally, clearly the γ are functions of the sample size, n

is derived in both Meese and Rogoff (1988) and Diebold and Rudebush (1991). In Sections 5 and 6, I move beyond the existing literature, progressively weakening the two remaining baseline assumptions.

5. Biased Forecasts

Bias will arise in most cases because of misspecification. Although bias may be ubiquitous, the adjustments needed are in general quite minor. Let $E[e_1] \equiv \mu_1$ and $E[e_2] \equiv \mu_2$ and assume that μ_1 and μ_2 are not equal to zero. In the transformed population, $E[U] \equiv \mu_U = \mu_1 - \mu_2$, and $E[V] \equiv \mu_V = \mu_1 + \mu_2$. The covariance of U and V becomes

$$cov(U,V) = E\left[E_1^2 - E_2^2\right] - \mu_U \mu_V. \tag{33}$$

We see there is no longer the one-to-one correspondence between the null hypothesis (5) on the MSPEs in the original population, and the covariance, (13), of the transformed population. Expanding the second term in (33),

$$\mu_U \mu_V = \mu_1^2 - \mu_2^2. \tag{34}$$

The covariance will equal zero in all cases for which $\mu_1 = \mu_2$, not simply when both are equal to zero. The analysis of Section 3 would go through unmodified in this instance.[9] Even if $\mu_1 \neq \mu_2$, we can still make use of much of the analysis of Section 3 by working with noncentral moments.[10] We can restate the null hypothesis as

$$H_0 : E[UV] = 0. \tag{35}$$

To test (35), begin by defining the noncentral sample covariance,

$$s'_{UV} = 1/n\left[\sum_{j=1}^{n} u_j v_j\right]. \tag{36}$$

Under the null, $E[s'_{UV}] = \left(E\left[\sum_{j=1}^{n} e_{1j}^2\right] - E\left[\sum_{j=1}^{n} e_{2j}^2\right]\right)/n = 0$. The variance of this noncentral moment is given by,

$$var(s'_{UV}) = 1/n^2 E\left[\sum_{j=1}^{n}\sum_{m=1}^{n} u_j v_j u_m v_m\right] - E[s'_{UV}]^2. \tag{37}$$

which is identical to the unbiased case (29) since we chose the origin as the mean of our baseline population. As with (29), normality eliminates the fourth cumulants

[9] Ashley, Granger and Schmalensee (1980) run the regression $e_{1j} - e_{2j} = \beta_1 + \beta_2\left[(\Sigma e_{1j})^2 - (\Sigma e_{2j})^2\right]$ as a test for Granger causality. The equality of MSPEs in the biased case implies that $\beta_1 = \beta_2 = 0$. Interpreting AGS's results when one or both β's are found to be negative can be tricky. Asymptotically, their approach is equivalent to the one I take in this section.

[10] I could also test the equality of the noncentral second moments in the untransformed population, $H'_0 : \sigma'_1 = \sigma'_2$, where $E\left[e_i^2\right] = \sigma'_i$.

and triple products in (A.8). The bias leaves in lines four and five, but only for the on diagonal terms as $E[u_j u_k] = E[v_j v_k] = E[u_j v_k] = 0$, for all $j \neq k$. Noting that under the null hypothesis, $\gamma_{UV}^2(0) = 0$, (37) simplifies to

$$var(s'_{UV}) = 1/n \left[6\mu_U^2 \mu_V^2 - 2\gamma_{UU}(0)\mu_V^2 - 2\gamma_{VV}(0)\mu_U^2 + \gamma_{UU}(0)\gamma_{VV}(0) \right]. \tag{38}$$

If we were still in the unbiased case, $\mu_U = \mu_V = 0$ and (38) would further simplify to (30) with $t = 0$. Normality and the absence of serial correlation are helpful in obtaining a simple expression for the variance. Having now developed the necessary preliminaries, I am now ready to tackle the problem in full generality.

6. A Robust Test for Heteroscedastic Populations

In this section, I examine the general case where $(e_1,\ e_2)$ are dependent draws from a bivariate population, with joint density $f(e_1, e_2; \theta)$. As in the biased case, the first moments are assumed stationary, $E[e_{1j}] = \mu_1$ and $E[e_{2j}] = \mu_2$ for all j. I relax the assumption that the higher moments are stationary, allowing for heteroscedasticity. For any given j then, $E\left[e_{1j}^2 - e_{2j}^2\right]$ does not necessarily equal zero. On average though, the mean squared prediction errors are assumed equal. We can restate the null as

$$H_0 : E[s'_{UV}] \equiv \gamma_{UV}(0) = 0. \tag{39}$$

I will derive two heteroscedasticity and autocorrelation consistent (HAC) test statistics. [11] The first statistic will be for the case in which the order of dependence is known to be k, and the second case, the order of dependence will be unknown. To test the null hypothesis (39), I will again need the variance of this noncentral moment. In expression (A.8) of the appendix, I expand the quadratic from that arises in this variance. Begin by defining the non-central sample analog of γ_{UVUV},

$$s'_{UVUV}(t) = 1/n \sum_{j=t+1}^{n} u_j v_j u_{j-t} v_{j-t}, 1/n \sum_{j=-t+1}^{n} u_{j+t} v_{j+t} u_j v_j. \tag{40}$$

To obtain a statistic with a limiting normal distribution, I will only have to bound the on diagonal terms of the variance and the degree of auto- and cross-correlation.

PROPOSITION 1: If $\sup_{t\geq 1} E \left\| e_{it} \right\|^{4\nu} < \infty$, $i = 1, 2$, for some $\nu > 1$, and $\gamma_{UVUV}(t) = 0$ for some $\infty > t > k$, then

[11] Diebold and Mariano (1991) use an estimate of the spectral density at frequency zero to estimate the variance of s'_{UV}. Though not stated, the authors implicitly assume covariance stationarity, making Proposition 2 below more general. Diebold and Mariano also do not prove asymptotic normality.

$$\sqrt{n}\frac{1/n\sum_{j=1}^{n}u_jv_j}{\left[\sum_{t=-k}^{k}s'_{UVUV}(t)\right]^{1/2}}\overset{asy.}{\sim}N(0,1) \tag{41}$$

This proposition is really a corollary to the more general Proposition 2 as the k-dependence is a stronger condition than the -mixing assumption below.[12] Hence, I will forego any proof. A difficulty with implementing the statistic (41) is that there is no guarantee that for a given k and finite n that the variance estimate will be positive. This is a well-known result for the sample autocovariance function that will apply to the bivariate process $\{u_jv_j\}$ as well. The Bartlett window used by Newey-West (1987) can easily handle this contingency, that is, $\sum_{t=-k}^{k}(1-[|t|/(k+1)])s'_{UVUV}(t)>0$, for all n,k. It follows that

$$\sqrt{n}\frac{1/n\sum_{j=1}^{n}u_jv_j}{\left[\sum_{t=-k}^{k}(1-[|t|/(k+1)])\,s'_{UVUV}(t)\right]^{1/2}}\overset{asy.}{\sim}N(0,1). \tag{42}$$

I can also relax the assumption that the order of dependence is known, replacing it with a mixing condition. Accompanied with a rule for expanding k with the sample size, we can extend Proposition 1 to the case of heteroscedasticity of unknown form.

PROPOSITION 2: Assume that:

(i) The processes $\{e_i\}$, $i=1,2$, are an α-mixing sequence of random variables, satisfying: $\sup_{t\geq 1}E\left\|e_{it}\right\|^{4\nu}<\infty$; and $\sum_{j=1}^{n}j^2\alpha(j)^{(\nu-1)/\nu}$ for some $\nu>1$;

(ii) The weights $\omega(t,k),(k=1,2,...,t=1,..,k)$ are finite and for each t, $\lim_{k\to\infty}\omega(t,k)=1$;

Then if $k(n)$ is chosen to be a function of sample size such that $\lim_{n\to\infty}k(n)=+\infty$ and $\lim_{n\to\infty}k(n)/n^{1/2}=0$, it follows that

$$\sqrt{n}\frac{1/n\sum_{j=1}^{n}u_jv_j}{\left[\sum_{t=-k(n)}^{k(n)}(1-[|t|/(k+1)])\,s'_{UVUV}(t)\right]^{1/2}}\overset{asy.}{\sim}N(0,1). \tag{43}$$

Proof: H_0, $\sqrt{n}\left(1/n\sum_{j=1}^{n}u_jv_j-0\right)=O_p(1)$. Consider now the term in the denominator. Under the α-mixing assumption, trivially satisfied if the process were k-dependent), Andrews (1991) shows that

$$\sum_{j=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\kappa_{UVUV}(0,0,j,m)<\infty,$$

$$\sum_{j=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\gamma_{abc}(0,j,m)<\infty,\ a,b,c\in(U,V)$$

[12]The mixing coefficients $\alpha(j)$ determine how much dependence exists in two realizations of the random sequence j-periods apart. By allowing for these coefficients to approach zero as j grows large, α-mixing implies an asymptotic independence. For a k-dependent process, the coefficients $\alpha(j)$ are zero for $j>k$. For more discussion, see White (1984, p.45)

$$\sum_{j=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \gamma_{ab}(j) < \infty, \ a, b \in (U, V).$$

With all of these terms bounded and finite, then

$$\lim_{n\to\infty} 1/n \sum_{j=t+1}^{n} u_j v_j u_{j-t} v_{j-t} = \gamma_{UVUV}(t) < \infty, t \geq 0$$

$$\lim_{n\to\infty} 1/n \sum_{j=-t+1}^{n} u_{j+t} v_{j+t} u_j v_j = \gamma_{UVUV}(t) < \infty, t < 0$$

With the Bartlett weights, clearly $\omega(t,k) \to 1$ for each t, and $[t/(k(n)+1)]^{1/2} \to 0$ if t grows no faster than $n^{1/2}$. A multivariate central limit theorem for dependent processes (see e.g. Gallant (1987) Theorem 2, p. 519) completes the proof.

7. Robust Statistic: Finite Sample Properties

This section presents two Monte Carlo exercises. The first exercise compares the Meese-Rogoff statistic and the new robust statistic in a time series example. I then compare the robust statistic with a fixed number of autocorrelations k to the case in which k grows with the sample.

7.1 A Time Series Example

In this section, I contrast the Meese and Rogoff (1988) statistic (32) to the robust statistic (42). The data generating mechanism was the model,

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} 0.50 & -0.45 & -0.15 \\ 0.05 & 0.00 & 0.35 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \varepsilon_{1t-2} \\ \varepsilon_{1t-3} \end{bmatrix} + \begin{bmatrix} 0.25 & 0.00 \\ -0.35 & 0.50 \end{bmatrix} \begin{bmatrix} \varepsilon_{2t-1} \\ \varepsilon_{2t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{44}$$

In the first design, the $\varepsilon's$ were drawn from a standard normal distribution. Results are in Table 2.1 for samples sizes of $n = 25, 50$ and 250, with $k = 3$ in both instances. Both statistics are properly sized in all three samples. There does not seems to be any substantial loss in finite sample accuracy in using the robust statistic rather than Meese-Rogoff, even when the data satisfy our baseline assumption of normality. In many applications in finance though, this assumption is violated. Dependence in the first and second conditional moments and unconditional leptokurtosis are stylized facts for many asset returns. Heteroscedasticity is also frequently encountered. I chose to capture these features in my second design by using a mixture of

normals distribution.[13] Every fifth observation, I had the random number generator sample from a $N(0,4)$ distribution in place of an $N(0,1)$. This produced an excess kurtosis of around 3 in small samples and around 10 in large samples. Results for 20,000 replications are in the first two panels of Table 2.2. The HAC statistic (42) is undersized in the small samples, but properly sized for n greater than 100. The Meese-Rogoff statistic rejects anywhere from 3 to 5 times too often, with performance deteriorating as sample size increases. While the HAC statistic works well in normal and non-normal populations, the Meese-Rogoff statistic does not show a similar versatility. Since non-normality is likely to matter in financial time series, the robust statistic is clearly to be preferred in these applications.

7.2 Heteroscedasticity of Unknown Form

There is little basis on which to choose a rate of growth for k in practice. Newey and West (1987) suggest that this is an important topic for future research. While an exhaustive examination is beyond our scope here, I wanted to compare the robust statistic with k known to the case in which k is unknown. To examine the implications of heteroscedasticity of unknown form, I allowed $k(n)$ to grow according to the rule $k = int[n^{1/3}] + 1$, where $int[.]$, denotes the integer part of the argument.[14] This obviously satisfies the conditions of Proposition 2. For $n = 25$, $k = 3, n = 50, k = 4, n = 250, k = 7$. I again used the data generating mechanism (44). Results are in the bottom panel of Table 2.2. Finite sample performance is essentially equivalent to the case in which k is known. As long as the ratio of autocorrelations to data points is fairly high, allowing k to grow with the sample size seems to have little impact on the inference. Obviously though when k is known, one should incorporate that population information.

8. Power of the Robust Statistics

Having now dealt with size, I take up the issue of power in this section. In 8.1, I motivate a problem similar to that faced by data analysts in that I consider models known to be misspecified. The power calculations using the robust statistic (42) are in 8.2.

8.1 An Illustrative Example

In time series analysis, one often begins only with representation theory as a priori knowledge. Given stationarity, a Wold representation exists; with invertibility, an

[13] An ARCH model or some member of the Paretian densities are the obvious alternatives, but the former has rather restrictive moment conditions, and the latter family has no moments at all for many parameterizations.

[14] Andrews (1991) presents evidence that this is the optimal rate of growth for the Bartlett window.

autoregressive representation may also be obtained. Since the pioneering work of Box and Jenkins (1976), time series analysts have refined the art and science of identifying time series models solely with data analysis. With either time series or structural models, the primary goal of the data analysis is often prediction. Even misspecified models may be considered if they are particularly useful in forecasting. I consider as an illustrative example the MA(1) data generating mechanism

$$y_t = \varepsilon_t - 0.8\varepsilon_{t-1} \tag{45}$$

where ε_t is a white noise disturbance. (45) is invertible and I can then re-write the DGM as an infinite order autoregressive process,

$$y_t(1 - 0.8L)^{-1} = \varepsilon_t \tag{46}$$

Expanding the polynomial in the lag operator (1-0.8L),

$$y_t = -0.8y_{t-1} - (0.8)^2 y_{t-2} - \cdots - (0.8)^n y_{t-n} - \cdots + \varepsilon_t \tag{47}$$

I consider forecasts of the process (46) from misspecified $AR(p)$ alternatives. I wish to compare mean squared prediction errors for k step ahead forecasts.

Davies and Newbold (1980) have tabulated percentage increases in forecast error variances for $k = 1, 2, 3, 4$ and $p = 1$ to 8. For instance, the forecast error variance using an AR(1) model for the one step ahead forecast ($k = 1$) is 25% larger than the correctly specified MA(1) model. For four steps ahead though, the variance is only 0.3% larger. For an AR(4) model, the one step ahead variance increase is only 4.3%, and for an AR(8), only 0.3%. With an 8th order model forecasting 8 steps ahead, the models are virtually indistinguishable (0.1% increase in forecast error variance.) With the misspecified model little is lost in large samples when forecasting either 4 steps ahead or with an 8th order AR model. Theoretically, power falls off a little bit more rapidly as we expand k than p.

8.2 The Robust Statistic: Power Against AR(p) Alternatives

I utilized the robust statistic (42) for all the power comparisons. I look at both normal and mixture of normals distributions for the errors in the MA(1) model. I look at models of $p = 1, 4$, and 8 and forecast horizons of $k = 1, 4$, and 8 periods, and sample sizes of $n = 25$ and 250. All exercises have 20,000 replications. In Table 3.1, I use normally distributed errors, and in Table 3.2, a mixture of $N(0, 1)$ and $N(0, 4)$ errors. The ratio of sample MSPEs is reported as a rough measure of the percentage increase in the population forecast error variance. For the case of $p = 1$, $k = 1$, the MSPE ratio is 0.743, indicative of the 25% larger population forecast error variance of the misspecified model. In both tables, for a given value of p, power falls off monotonically with the forecast horizon. Although the errors are correlated, the Theil-U does describe the power fall off for fixed p. For $k = 1$ and $k = 4$,

power also falls off as the order of the AR approximation increases. Interestingly though, for $k = 8$ and $n = 250$, power increases slightly from $p = 4$ to $p = 8$ in both Tables 3.1 and 3.2. This same phenomenon was noted by Davies and Newbold (1980). Taking into account the imprecision of the parameter estimates, the forecast error variances actually start to increase at all forecast horizons when $p > 4$. The leptokurtosis of the mixture of normal errors, if anything, enhances the power of the robust statistic. In the large sample, there are sizable power improvements at $k = 1$ and 4 for $p = 1$ and 4. The only substantial reduction occurs in the small sample with $k = 1$and $p = 1$. Overall, the robust statistic proves its utility in the power comparisons. In the small sample, it rejects nearly $q\%$ of the time alternatives that increase the forecast error variance by $q\%$. For example, with $p = 4$, and $k = 4$, the theoretical increase is 0.9%, and the robust statistic rejects 1.1% of the time. In large samples, (42) rejects 99% of the time against the alternative with the highest variance increase ($p = 1, k = 1, n = 250$) with either normal or mixture errors. Even at $p = 4, k = 4$, it still rejects almost 60% of the time. While power is considerably lower against the AR(8) model, recall that this model's theoretical variance is only 0.6% larger than the correct model even at a one period horizon. The robust statistic does the job of warning us about badly misspecified models in small samples. In large samples, it does the job even against close alternatives.

9. Conclusion

I hope that this paper will help to close the distance between the applied and theoretical time series literature. Data analysts now can construct simple tests of the null hypothesis of the equality of mean squared errors under extremely weak population assumptions. In the finance literature, this should help resolve questions like which stock price newsletter is truly doing the best job or whether nonlinear models can better the random walk.[15] With macroeconometric models, it may help both model builders and model critics to identify areas in need of improvement. The Monte Carlo and power comparisons are useful diagnostics. Size and power depend critically in our examples upon dependence assumptions and higher population moments. Our finite sample analysis indicates that ignoring departures from the original Granger-Newbold framework can result in misleading inference. The obvious extension of this work is to the L_1 norm (mean absolute deviations). Given the asymptotic normality of most population fractiles, at least a few basic results should be obtainable.

[15]Mizrach (1992b) improved upon a random walk forecast for the Italian Lira by nearly 5% using a nonparametric technique. Assuming normality, the statistic (24) indicated an improvement at the 95% critical level. Upon finding excess kurtosis in the forecast errors though, I used the robust statistic, and instead found a p-value of only 0.34.

Appendix: Cumulants for Multivariate Distributions

There is a very simple relationship between univariate and multivariate moments and cumulants. Consider the fourth cumulant of some random variable, $\kappa_4 = \mu_4 - 3\mu_2^2$. We can, following Kendall and Stuart (1963 p.83), treat this as a function of r,

$$\kappa(r^4) = \mu(r^4) - 3\mu(r^2)^2 \tag{A.1}$$

Applying the operator $s\partial\kappa/\partial r$, we obtain the corresponding bivariate cumulant,

$$4\kappa(r^3 s) = 4\mu(r^3 s) - 12\mu(r^2)\mu(rs). \tag{A.2}$$

After dividing through by 4, we obtain

$$\kappa_{31} = \mu_{31} - 3\mu_{20}\mu_{11}. \tag{A.3}$$

A similar operator may be applied to the moments as well. The cumulant that arises in the case of the non-normally distributed, noncentral covariance is κ_{1111}. Using the operator from (A.2), $\kappa_{31} = s\partial\kappa_4/\partial r$, $\kappa_{211} = t\partial\kappa_{31}/\partial s$, and $\kappa_{1111} = u\partial\kappa_{2111}/\partial t$. It is a well known result that the fourth cumulant of the univariate normal distribution is zero. It follows that the first order quadravariate cumulant is zero as well. From any order cumulant (or moment) in n variables, it is straightforward to compute the corresponding expression in terms of the moments of the n-variates. Denote the four variate cumulant as κ_{abcd}. The expression for the cumulant in terms of the moments is just a permutation of the underlying moments,

$$\kappa_{abcd} = \mu'_{abcd} + \alpha_2 \sum^4 \mu'_a \mu'_{bcd} + \alpha_2 \sum^3 \mu'_{ab}\mu'_{cd} + \alpha_3 \sum^6 \mu'_a \mu'_b \mu'_{cd} + \alpha_4 \mu'_a \mu'_b \mu'_c \mu'_d. \tag{A.4}$$

The superscript above the summation sign denotes the number of permutations of the underlying indices, $\alpha_p = (p-1)(-1)^{p-1}$, where p is the number of moments in the expression, and μ'_i denotes the first noncentral moments of the variables $i = a, b, c, d$ and $\mu'_{ab} = E[ab]$, etc.

Having now developed these preliminaries, I can now expand the quadratic form that arises in calculating the variance of the noncentral covariance in (29) and (37),

$$var(s'_{UV}) = 1/nE\left[1/n \sum_{j=1}^{n} \sum_{m=1}^{n} u_j v_j u_m v_m\right]. \tag{A.5}$$

By a change of variables, I can write (A.5) as

$$var(s'_{UV}) = 1/n \sum_{t=-n+1}^{n-1} \gamma_{UVUV}(t), \tag{A.6}$$

where

$$\gamma_{UVUV}(t) \equiv 1/n \sum_{j=t+1}^{n} E\left[u_j v_j u_{j-t} v_{j-t}\right].\ 1/n \sum_{j=-t+1}^{n} E\left[u_{j+t} v_{j+t} u_j v_j\right]. \tag{A.7}$$

To describe the sufficient conditions for consistency of the estimated variance, we must further analyze (A.7).

$$\begin{aligned}
\gamma_{UVUV}(t) &= \kappa_{UVUV}(0,0,t,t) + \gamma_{UVU}(0,0,t)\mu_V + \gamma_{UVV}(0,0,t)\mu_U \\
&\quad +\gamma_{UUV}(0,t,t)\mu_V + \gamma_{VUV}(0,t,t)\mu_U \\
&\quad +\gamma_{UV}^2(0) + \gamma_{UU}(t)\gamma_{VV}(t) + \gamma_{VU}(t)\gamma_{UV}(t) \\
&\quad -2\mu_U\mu_V\left[\gamma_{UV}(0) + \gamma_{UV}(t) + \gamma_{VU}(t) + \gamma_{UV}(0)\right] \\
&\quad -2\left[\gamma_{UU}(t)\mu_V^2 + \gamma_{VV}(t)\mu_U^2\right] + 6\left[\mu_U^2\mu_V^2\right],
\end{aligned} \tag{A.8}$$

where $\kappa_{UVUV}(0,0,t,t)$ denotes for $t > 0$,

$$1/n\sum_{j=t+1}^{n} E\left[(u_j - E[u_j])(v_j - E[v_j])(u_{j-t} - E[u_{j-t}])(v_{j-t} - E[v_{j-t}])\right],$$

for $t \leq 0$,

$$1/n\sum_{j=-t+1}^{n} E\left[(u_{j+t} - E[u_{j+t}])(v_{j+t} - E[v_{j+t}])(u_j - E[u_j])(v_j - E[v_j])\right], \tag{A.9}$$

and for

$$\begin{array}{lll}
 & t > 0 & t \leq 0 \\
\gamma_{UVU}(0,0,t) \equiv & 1/n\sum_{j=t+1}^{n} E\left[u_j v_j u_{j-t}\right], & 1/n\sum_{j=-t+1}^{n} E\left[u_{j+t} v_{j+t} u_j\right], \\
\gamma_{UVU}(0,0,t) \equiv & 1/n\sum_{j=t+1}^{n} E\left[u_j v_j v_{j-t}\right], & 1/n\sum_{j=-t+1}^{n} E\left[u_{j+t} v_{j+t} v_j\right], \\
\gamma_{UVU}(0,t,t) \equiv & 1/n\sum_{j=t+1}^{n} E\left[u_j u_{j-t} v_{j-t}\right], & 1/n\sum_{j=-t+1}^{n} E\left[u_{j+t} u_j v_j\right], \\
\gamma_{UVU}(0,t,t) \equiv & 1/n\sum_{j=t+1}^{n} E\left[v_j u_{j-t} v_{j-t}\right], & 1/n\sum_{j=-t+1}^{n} E\left[v_{j+t} u_j v_j\right].
\end{array} \tag{A.10}$$

The remaining cross products, γ_{UU} etc., are defined in the text as equations (31). For the normal distribution, the fourth cumulant κ_{UVUV} will be zero, and for symmetric distributions, the triple products (A.10) will be zero. In the unbiased case, the fourth and fifth lines of (A.8) are zero. Under a dependence assumption of order k, all terms for which $t > k$ are zero.

References

ANDREWS, D. (1991), "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 307-45.

ARMSTRONG, J.S. (1983), "Relative Accuracy of Judgmental and Extrapolative Methods in Forecasts of Annual Earnings," *Journal of Forecasting* 2, 437-47.

ASHLEY, R., C.W.J. GRANGER and R. SCHMALANSEE, (1980) "Advertising and Aggregate Consumption: An Analysis of Causality," *Econometrica* 48, 1149-67.

BOSE, S.S. (1935) "On the Distribution of the Ratio of Variances of Two Samples Drawn From a Given Normal Bivariate Correlated Population," *Sankhya* 1, 65.

BOX, G. and G.W. JENKINS (1976), *Time Series Analysis*, San Francisco: Holden Day.

CLEMENTS, M.P. and D.F. HENDRY (1992), "On the Limitations of Comparing Mean Square Forecast Errors," *Journal of Forecasting*, forthcoming.

DAVIES, N. and P. NEWBOLD (1980), "Forecasting with Misspecified Models," *Applied Statistics* 29, 87-92.

DIEBOLD, F. and R. MARIANO (1991), "Comparing Predictive Accuracy I: Asymptotic Tests," Working Paper, Department of Economics, University of Pennsylvania.

DIEBOLD, F. and G. RUDEBUSH (1991), "Forecasting Output with the Composite leading Index: A Real Time Analysis," *Journal of the American Statistical Association*, forthcoming.

FINNEY, D.J. (1938), "The Distribution of the Ratio of Estimates of the Two Variances in a Sample From a Normal Bivariate Population," *Biometrika* 30, 190-92.

FISHER, R. (1915), "Frequency Distribution of the Values of Correlation Coefficients in Samples from an Indefinitely Large Population," *Biometrika* 10, 507-21.

FISHER, R. (1921), "On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample," *Metron* 1, 3-32.

GALLANT, A.R. (1987) Nonlinear Statistical Models, New York: John Wiley.

GRANGER, C.W.J. and P. NEWBOLD (1986), *Forecasting in Business and Economic Time Series*, San Diego: Academic Press.

HOEL, G. (1954), *Introduction to Mathematical Statistics*, New York: Wiley.

KENDALL, M.G. and A. STUART (1963), *The Advanced Theory of Statistics Volume I: Distribution Theory*, London: Charles Griffin.

McNEES, S.K. (1990), "The Role of Judgmental Macroeconomic Forecasting Accuracy," *International Journal of Forecasting* 6, 287-99.

MEESE, R. and K. ROGOFF (1983), "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?", *Journal of International Economics* 14, 3-24.

MEESE, R. and K. ROGOFF(1988), "Was It Real? The Exchange Rate-Interest Differential Relation over the Modern Floating Rate Period," *Journal of Finance* 43, 933-48.

MIZRACH, B. (1992a) "The Distribution of the Theil-U Statistic in Bivariate Normal Populations," *Economics Letters* 38, 1637-67.

MIZRACH, B. (1992b), "Multivariate Nearest Neighbors Forecasts of EMS Exchange Rates," *Journal of Applied Econometrics* 7 Supplement, S151-63.

MIZRACH, B., (1990), "The Information in the Term Structure: A Nonparametric Investigation," Working Paper, Department of Finance, The Wharton School, forthcoming in the *Journal of Forecasting*.

NEWEY, W. and K. WEST (1987), "A Simple Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-8.

PRIESTLEY, M.B. (1981), *Spectral Analysis and Time Series*, San Diego: Academic Press.

THEIL, H. (1966), *Applied Economic Forecasting*, Amsterdam: North Holland.

WHITE, H. (1984), *Asymptotic Theory for Econometricians*, Orlando: Academic Press.

Table 1.1
Statistics for Bivariate Normal Populations*

Sample Size n = 10

Fractiles	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
Variance of Covariance	-2.056	-1.608	-1.340	1.336	1.595	2.056	0.03
Variance of Corr.	-2.202	-1.693	-1.367	1.372	1.707	2.215	0.05
Corr. Exact Finite	-2.745	-1.794	-1.356	1.362	1.813	2.777	0.05
Fisher's Transformation	-2.276	-1.582	-1.225	1.229	1.597	2.298	0.04
Likelihood Ratio	0.124	0.367	0.610	6.262	7.881	11.729	0.05
Student T 9DF	-2.764	-1.812	-1.372	1.372	1.812	2.764	

Sample Size n = 25

Fractiles	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
Variance of Covariance	-2.222	-1.629	-1.311	1.292	1.624	2.213	0.04
Variance of Corr.	-2.448	-1.759	-1.395	1.382	1.759	2.437	0.07
Corr. Exact Finite	-2.694	-1.803	-1.394	1.379	1.802	2.676	0.06
Fisher's Transformation	-2.512	-1.724	-1.345	1.331	1.723	2.498	0.06
Likelihood Ratio	0.120	0.369	0.599	6.615	8.341	12.158	0.06
Student T 9DF	-2.485	-1.708	-1.308	1.316	1.708	2.485	

Sample Size n = 50

Fractiles	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
Variance of Covariance	-2.243	-1.655	-1.293	1.300	1.651	2.278	0.05
Variance of Corr.	-2.426	-1.747	-1.363	1.353	1.745	2.446	0.07
Corr. Exact Finite	-2.531	-1.767	-1.361	1.351	1.765	2.554	0.06
Fisher's Transformation	-2.452	-1.730	-1.338	1.329	1.728	2.473	0.06
Likelihood Ratio	0.123	0.363	0.594	6.534	8.283	12.188	0.06
Normal (0.1)	-2.326	-1.645	-1.282	1.282	1.645	2.326	
$\chi^2(3)$	0.114	0.352	0.584	6.251	7.816	11.345	

*The variance of the covariance is given by (18). The variance of the correlation coefficient is equation (21). Fisher's transformation of the sample correlation coefficient is given by (24). These all have Normal (0,1) distributions in large samples. The exact finite sample statistic, (22) has a Student-t distribution with n-1 degrees of freedom. The likelihood ratio statistic, (27) is distributed χ^2 with 3 degrees of freedom. All exercises are based on 20,000 replications.

Table 1.2
Statistics for Bivariate Normal Populations*

Sample Size n = 100

Fractiles	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
Variance of Covariance	-2.227	-1.627	-1.286	1.308	1.661	2.278	0.05
Variance of Corr.	-2.306	-1.656	-1.302	1.330	1.700	2.349	0.05
Corr. Exact Finite	-2.346	-1.663	-1.300	1.329	1.708	2.392	0.05
Fisher's Transformation	-2.313	-1.647	-1.290	1.318	1.691	2.357	0.05
Likelihood Ratio	0.110	0.331	0.549	6.058	7.602	11.299	0.05

Sample Size n = 250

Fractiles	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
Variance of Covariance	-2.307	-1.626	-1.288	1.292	1.652	2.286	0.05
Variance of Corr.	-2.312	-1.606	-1.277	1.275	1.632	2.319	0.05
Corr. Exact Finite	-2.328	-1.608	-1.276	1.274	1.635	2.335	0.05
Fisher's Transformation	-2.315	-1.602	-1.272	1.270	1.628	2.322	0.05
Likelihood Ratio	0.113	0.364	0.594	6.248	7.813	11.161	0.05
Normal (0,1)	-2.326	-1.645	-1.282	1.282	1.645	2.326	
$\chi^2(3)$	0.114	0.352	0.584	6.251	7.815	11.345	

*The variance of the covariance is given by (18). The variance of the correlation coefficient is equation (21). Fisher's transformation of the sample correlation coefficient is given by (24). These all have Normal (0,1) distributions in large samples. The exact finite sample statistic, (22) has a Student-t distribution with n-1 degrees of freedom. The likelihood ratio statistic, (27) is distributed χ^2 with 3 degrees of freedom. All exercises are based on 20,000 replications.

Table 2.1
Statistics for the Sample Covariance: The Time Series Case
Bivariate Normal Populations*

Fractiles of Meese and Rogoff (1988) Statistic

Sample Size	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
n = 25	-1.819	-1.404	-1.151	1.302	1.586	2.028	0.02
n = 50	-1.894	-1.390	-1.105	1.363	1.677	2.180	0.03
n = 250	-2.189	-1.515	-1.160	1.431	1.783	2.472	0.05

Fractiles of Robust Statistic

Sample Size	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
n = 25	-1.916	-1.563	-1.328	1.385	1.623	1.964	0.02
n = 50	-2.058	-1.559	-1.252	1.459	1.763	2.228	0.04
n = 250	-2.293	-1.581	-1.225	1.486	1.836	2.519	0.06

*The Meese and Rogoff (1988) statistics is given by (32) with k = 3. the robust statistic is given by (42), again for k = 3. The data generating mechanism is given by expression (44) in the text. All exercises are based on 20,000 replications.

Table 2.2
Statistics for the Sample Covariance in Heteroscedastic Populations
Mixture of Normals N(0,1) and N(0,4)*

Fractiles of Meese and Rogoff (1988) Statistic

Sample Size	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
n = 25	-2.201	-1.793	-1.500	1.714	2.000	2.385	0.09
n = 50	-2.678	-2.083	-1.686	1.953	2.336	2.912	0.16
n = 250	-3.550	-12.584	-2.060	2.267	2.849	3.861	0.25

Fractiles of Robust Statistic, k= 3

Sample Size	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
n = 25	-1.809	-1.507	-1.311	1.332	1.509	1.785	0.01
n = 50	-1.956	-1.595	-1.338	1.418	1.645	2.001	0.02
n = 250	-2.348	-1.707	-1.352	1.400	1.755	2.335	0.06

Fractiles of Robust Statistic, k= int[$n^{1/3}$]+1

Sample Size	0.01	0.05	0.10	0.90	0.95	0.99	Size of .05 Test
n = 25	-1.809	-1.507	-1.311	1.332	1.509	1.785	0.01
n = 50	-1.918	-1.566	-1.322	1.406	1.629	1.969	0.02
n = 250	-2.242	-1.644	-1.302	1.361	1.708	2.280	0.05

*The Meese and Rogoff (1988) statistics is given by (32) with k = 3. the robust statistic is given by (42), again for k = 3. The data generating mechanism is given by expression (44) in the text. All exercises are based on 20,000 replications.

Table 3.1
Power Against AR(p) Alternatives
Robust Forecast Statistic
Bivariate Normal Errors*

Sample Size n = 25

	Forecast	Horizon	
Model	1	4	8
$p = 1$	0.230	0.027	0.000
(MSPE Ratio)	(0.743)	(0.883)	(0.890)
$p = 4$	0.078	0.011	0.000
(MSPE Ratio)	(0.921)	(0.924)	(0.992)
$p = 8$	0.043	>0.000	0.000
(MSPE Ratio)	(0.986)	(0.987)	(0.988)

Sample Size n = 250

	Forecast	Horizon	
Model	1	4	8
$p = 1$	0.999	0.758	0.176
(MSPE Ratio)	(0.712)	(0.859)	(0.973)
$p = 4$	0.715	0.567	0.100
(MSPE Ratio)	(0.905)	(0.905)	(0.989)
$p = 8$	0.176	0.148	0.130
(MSPE Ratio)	(0.982)	(0.983)	(0.983)

*In column 1, p stands for the order of the autoregressive approximation. The % rejection of the null for forecasts at 1, 4 and 8 steps ahead are on each row. The MSPE ratio is reported in parentheses below. The data generating mechanism is the MA(1) model $y_t = -0.8\varepsilon_{t-1} + \varepsilon_t$. The theoretical increases in forecast error variance are (from Granger and Newbold (1980)) for k = 1 and k = 4, with p = 1, (25.0%,0.3%), p = 4, (4.3%,0.9%), and p = 8, (0.6%,0.1%).

Table 3.2
Power Against AR(p) Alternatives
Robust Forecast Statistic
Mixture of Normals N(0,1) and N(0,4)*

Sample Size n = 25

	Forecast	Horizon	
Model	1	4	8
$p = 1$	0.177	0.031	0.000
(MSPE Ratio)	(0.678)	(0.846)	(0.976)
$p = 4$	0.066	0.005	0.000
(MSPE Ratio)	(0.905)	(0.941)	(0.996)
$p = 8$	0.026	>0.000	0.000
(MSPE Ratio)	(0.978)	(0.985)	(0.985)

Sample Size n = 250

	Forecast	Horizon	
Model	1	4	8
$p = 1$	0.999	0.870	0.244
(MSPE Ratio)	(0.706)	(0.855)	(0.972)
$p = 4$	0.835	0.697	0.064
(MSPE Ratio)	(0.902)	(0.904)	(0.990)
$p = 8$	0.247	0.116	0.106
(MSPE Ratio)	(0.982)	(0.983)	(0.983)

*In column 1, p stands for the order of the autoregressive approximation. The % rejection of the null for forecasts at 1, 4 and 8 steps ahead are on each row. The MSPE ratio is reported in parentheses below. The data generating mechanism is the MA(1) model $y_t = -0.8\varepsilon_{t-1} + \varepsilon_t$. The theoretical increases in forecast error variance are (from Granger and Newbold (1980)) for k = 1 and k = 4, with p = 1, (25.0%,0.3%), p = 4, (4.3%,0.9%), and p = 8, (0.6%,0.1%).