Economics Letters 38 (1992) 163–167
North-Holland

The distribution of the Theil *U*-statistic in bivariate normal populations

Bruce Mizrach

The Wharton School, Philadelphia PA, USA

Received 13 September 1991
Accepted 18 November 1991

This note provides an exact finite sample distribution for the Theil (1966) *U*-statistic in correlated bivariate normal populations. The type II error rate using the *F*-distribution is shown to increase with the square of the correlation coefficient.

1. Introduction

Applied econometricians often engage in data analytic horse races, assessing the relative merits of a number of forecasting models. [1] Methods of inference for in-sample comparisons are relatively well developed, but surprisingly, statistics for evaluating forecasts are rather sparse. Recent work by Mizrach (1991) has attempted to close this gap.

In the forecast competitions, data analysts have been forced to rely upon heuristic criteria. Typically, bias and mean squared error are tabulated. A ubiquitous statistic is the ratio of mean squared errors from rival forecasts. When a comparison is made against a forecast of no change, the ratio is commonly called the Theil (1966) *U*-statistic.

This note provides an exact finite sample distribution for the Theil-*U* (actually any mean squared error ratio) under the assumption of bivariate normality. I show below that the ratio of squared normal variates differs from the usual *F*-distribution when the errors are correlated. The probability mass in the tails shrinks in proportion to the square of the correlation between the errors. Using the *F*-distribution, as is often done naively in the literature, will lead the data analyst to accept far too often the null hypothesis that the mean squared forecast errors are *not* significantly different.

In section 2, I derive the distribution theory. I evaluate the integrals numerically in Monte Carlo exercises in section 3. A short conclusion closes the paper in section 4.

2. The Theil-*U* statistic

Denote by $y = (y_1, y_2, \dots, y_n)$ the vector of variables being forecasted, and let z_i be the ith predictor of y. The corresponding forecast errors will be denoted e_i. In the applied literature, a

Correspondence to: Bruce Mizrach, Department of Finance, The Wharton School, Philadelphia, PA 19104, USA. (215) 898-2761.

[1] Some representative examples are Meese and Rogoff (1983) for exchange rates, Armstrong (1983) for company earnings, and McNees (1990) for macroeconometric models.

common criterion by which to compare models is the *mean squared prediction error* (*MSPE*) of the forecast z_i,

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

Typically, the data exercise will not present a formal hypothesis test between rival predictors. Rather, the ratio of the MSPEs,

$$TU \equiv \left(1/n \sum_{j=1}^{n} e_{1j}^2\right) \Bigg/ \left(1/n \sum_{j=1}^{n} e_{2j}^2\right) \equiv s_1^2/s_2^2, \tag{2}$$

is presented as a heuristic. If z_2 is the no change forecast, then $e_{2j} = y_j$ for all j, and (2) becomes the Theil (1966) *U*-statistic. [2] It is the purpose of this note to show that the *U*-statistic can be very misleading, even as a diagnostic.

To derive the distribution of the Theil-*U*, I make the following population assumptions: 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 $\sigma_1^2 = \sigma_2^2$, and covariance, $\sigma_{12} = \rho\sigma_1\sigma_2$, where ρ is the correlation coefficient.

With unbiased errors, taking expectations in the numerator and denominator of (2) yields the ratio of the population variances,

$$\mathrm{E}\left[1/n \sum_{j=1}^{n} e_{1j}^2\right] / \mathrm{E}\left[1/n \sum_{j=1}^{n} e_{2j}^2\right] = \sigma_1^2/\sigma_2^2. \tag{3}$$

While the numerator and denominator of (3) are each distributed as χ^2 random variables with n degrees of freedom [3], one might guess that (3) has the standard variance ratio or *F*-distribution,

$$g(F) = [\Gamma(n)/\Gamma(n/2)\Gamma(n/2)] F^{(n-2)/2}(1+F)^{-n}, \tag{4}$$

where $F = (s_1^2/s_2^2)$. I show below that the ratio (3) will only have the standard *F*-distribution if the forecast errors are independent. When $\rho \neq 0$, the statistic *TU*'s distribution differs 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 variances, s_1^2 and s_2^2, and correlation coefficient, r, is proportional to [see e.g. Kendall and Stuart (1963, p. 385)]

$$\mathrm{d}F \propto \exp\left[-\frac{n^2}{2(1-\rho^2)}\left\{\frac{s_1^2}{\sigma_1^2} - \frac{2\rho^2 r s_1 s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\right\}\right] s_1^{n-1} s_2^{n-1} (1-r^2)^{(n-2)/2}\, \mathrm{d}s_1\, \mathrm{d}s_2\, \mathrm{d}r. \tag{5}$$

[2] 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.

[3] By using the population means in place of the sample means, we gain a degree of freedom.

Define the change of variables $H \equiv (s_1/s_2)$. H is just the ratio of the root mean squared errors, the square root of the Theil-U. (5) simplifies to

$$\mathrm{d}F \propto \frac{H^{n-1}(1-r^2)^{(n-2)/2}}{(1-2\rho rH+H^2)^n}\,\mathrm{d}r\,\mathrm{d}H. \tag{6}$$

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

$$\mathrm{d}F = \frac{2(1-\rho^2)^{n/2}}{\mathrm{Beta}\{n/2,\,n/2\}}\,\frac{H^{n-1}}{(1+H^2)^n}\left\{1-\frac{4\rho^2H^2}{(1+H^2)^2}\right\}^{-n/2}\mathrm{d}H. \tag{7}$$

(7) looks quite a bit different than the standard-F, but I will show that as $\rho \to 0$, we approach the ratio of independent normal variates.

It will take several steps to arrive there. Note first the following property of the Beta function, $\mathrm{Beta}(a+b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, giving us

$$\mathrm{d}F = \frac{\Gamma(n)}{\Gamma(n/2)\Gamma(n/2)}\,\frac{2(1-\rho^2)^{n/2}H^{n-1}}{(1+H^2)^n}\left\{1-\frac{4\rho^2H^2}{(1+H^2)^2}\right\}^{-n/2}\mathrm{d}H. \tag{8}$$

If the errors are independent, then obviously $\rho = 0$. Upon simplifying,

$$h(H) = \frac{\Gamma(n)}{\Gamma(n/2)\Gamma(n/2)}\,\frac{2H^{n-1}}{(1+H^2)^n}. \tag{9}$$

Consider the ratio of (4)/(9),

$$\frac{g(F)}{h(H)} = \frac{F^{(n-2)/2}(1+F)^{-n}}{2H^{n-1}(1+H^2)^{-1}}. \tag{10}$$

Now make the change of variables, $H = F^{1/2}$, noting that $f(F) = h(H)|\mathrm{d}H/\mathrm{d}F|$,

$$f(F) = 2F^{(n-1)/2}(1+F)^{-n}\,|F^{-1/2}/2|. \tag{11}$$

Collecting terms,

$$f(F) = F^{(n-2)/2}(1+F)^{-n} = g(F), \tag{12}$$

this verifies that the distribution of the Theil-U approaches the standard-F as $\rho \to 0$. For $\rho \neq 0$, the distribution collapses around its center. To evaluate this effect on the size of the traditional F-test, I turn to Monte Carlo simulations of the integral (7).

3. Monte Carlo evaluation of the exact finite sample distribution

For $\rho \neq 0$, (7) is strictly less than (4), the standard F, in the right tail of the distribution. The empirical size is well below the nominal size at traditional significance levels. This will show up in

Table 1
Distribution of the Theil U-statistic correlated bivariate normal populations. [a]

	Fractiles						Size of 0.5 test
	0.01	0.05	0.10	0.90	0.95	0.99	
Sample size $N = 10$							
F Distribution 10DF	0.206	0.336	0.431	2.323	2.978	4.849	0.05
Theil-U $\rho = 0.0$	0.213	0.331	0.426	2.310	2.969	4.912	0.05
Theil-U $\rho = 0.25$	0.224	0.342	0.438	2.268	2.919	4.674	0.05
Theil-U $\rho = 0.50$	0.254	0.380	0.476	2.106	2.614	4.100	0.03
Theil-U $\rho = 0.90$	0.473	0.605	0.684	1.458	1.654	2.083	0.00
Sample size $N = 25$							
F Distribution 25DF	0.384	0.511	0.594	1.683	1.955	2.604	0.05
Theil-U $\rho = 0.0$	0.386	0.507	0.592	1.690	1.976	2.659	0.05
Theil-U $\rho = 0.25$	0.395	0.522	0.603	1.664	1.921	2.565	0.05
Theil-U $\rho = 0.50$	0.434	0.559	0.638	1.569	1.793	2.355	0.03
Theil-U $\rho = 0.90$	0.652	0.745	0.797	1.257	1.343	1.535	0.00
Sample size $N = 50$							
F Distribution 50DF	0.513	0.625	0.694	1.441	1.599	1.949	0.05
Theil-U $\rho = 0.0$	0.512	0.628	0.697	1.447	1.609	1.958	0.05
Theil-U $\rho = 0.25$	0.523	0.640	0.705	1.426	1.579	1.917	0.05
Theil-U $\rho = 0.50$	0.563	0.670	0.731	1.368	1.503	1.793	0.03
Theil-U $\rho = 0.90$	0.747	0.816	0.854	1.173	1.232	1.352	0.00

[a] All exercises are based on 20,000 replications. The population is a bivariate $N(0, 1)$ with correlation coefficient ρ. The size is the empirical size of a 0.05 test using the F-distribution with (N, N) degrees of freedom.

the Monte Carlo exercises as a high rate of Type II error. I find that a 5% test never rejects the null hypothesis in samples with a correlation of $\rho = 0.9$.

I generated 20,000 samples of size $N = 10$, 25 and 50 using the IMSL subroutines CHFAC and RMNVN. CHFAC obtains the Cholesky factorization of the covariance matrix of the transformed errors,

$$\Sigma \equiv \begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix}. \tag{13}$$

By choosing $\sigma_U^2 = \sigma_V^2 = 1$, the correlation coefficient equals the covariance, $\rho = \sigma_{UV}$. The subroutine RMNVN then generates samples of random numbers using user supplied seed values. Since (7) varies with the square of the correlation coefficient, only the positive values $\rho = 0.0$, 0.25, 0.5 and 0.9 are reported in table 1.

As implied theoretically, the Theil-U with $\rho = 0$ matches the standard F-distribution. For all sample sizes, as ρ increases, the range of the [0.01, 0.99] interval of the distribution shrinks steadily. Using the F-distribution for $\rho = 0.25$ does not distort the size too drastically. The size falls to 0.03 in the case of $\rho = 0.5$; with $\rho = 0.9$ though, the F-statistic never rejects the null hypothesis.

4. Conclusion

Particularly with time series data, forecast errors are likely to be highly correlated. [4] Improper use of the Theil-U can result in very misleading inference. This paper has provided data analysts with the properly sized statistics for evaluating MSPE ratios. Given that the type II error rate is monotonic in the correlation coefficient, I have also shown that the F-distribution can safely be used as a conservative rule of thumb for comparisons.

References

Armstrong, J.S., 1983, Relative accuracy of judgmental, extrapolative methods in forecasts of annual earnings, Journal of Forecasting 2, 437–447.

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.

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, 190–192.

Kendall, M.G. and A. Stuart, 1963, The advanced theory of statistics, vol. 1, Distribution theory (Charles Griffin, London).

McNees, S.K., 1990, The role of judgmental macroeconomic forecasting accuracy, International Journal of Forecasting 6, 287–299.

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.

Mizrach, B., 1991, Forecast comparison in L_2, Working paper (University of Pennsylvania, PA).

Theil, H., 1966, Applied economic forecasting (North-Holland, Amsterdam).

[4] In the power exercises of Mizrach (1991), I compare autoregressive forecasts of an MA(1) data generating mechanism to the optimal MA(1) predictor. Correlations of 0.8 and 0.9 are typical.