A Simple Nonparametric Test of Predictive Performance

Author(s): M. Hashem Pesaran and Allan Timmermann

Source: *Journal of Business & Economic Statistics*, Oct., 1992, Vol. 10, No. 4 (Oct., 1992), pp. 461-465

Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association

Stable URL: https://www.jstor.org/stable/1391822

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/1391822?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Taylor & Francis, Ltd. and *American Statistical Association* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Business & Economic Statistics*

This content downloaded from
195.128.0.89 on Tue, 17 Feb 2026 15:24:39 UTC
All use subject to https://about.jstor.org/terms

© 1992 American Statistical Association

A Simple Nonparametric Test of Predictive Performance

M. Hashem Pesaran
Faculty of Economics and Politics, University of Cambridge, Cambridge CB3 9DD, United Kingdom, and Department of Economics, University of California, Los Angeles, CA 90024-1477

Allan Timmermann
Economics Department, Birkbeck College, London WCIE 7HX, United Kingdom

This article develops a distribution-free procedure for testing the accuracy of forecasts when the focus of the analysis is on the correct prediction of the direction of change in the variable under consideration. The proposed test is of particular interest in situations in which the underlying probability distribution of forecasts is difficult to derive analytically as in large nonlinear dynamic macroeconometric models or when the forecasts are made available only in the form of qualitative data as in most surveys on expectations. [Notable examples of these types of expectations surveys include the trend surveys carried out by the IFO-Institute of Munich, the Institut National de la Statistique et des Etudes Economiques (INSEE) in France, and the Confederation of British Industries (CBI). See, for example, Pesaran (1987, chap. 8).] The plan of the article is as follows: Section 1 introduces the test statistic in the 2 × 2 case and derives its distribution under fairly general conditions. Section 2 generalizes the test to the $m \times m$ case. Section 3 provides two applications of the test, a dichotomous version of the test to the CBI's Industrial Trends surveys of price changes and businessmen's expectations of these changes in the manufacturing sector and a trichotomous version of the test to the demand data from business surveys of French manufacturing industry conducted by INSEE.

1. THE TEST IN THE 2 × 2 CASE

Let $x_t = \hat{E}(y_t \mid \Omega_{t-1})$ be the predictor of y_t formed with respect to the information available at time $t - 1$, Ω_{t-1}, and suppose that there are n observations on (y_t, x_t). The predictive-failure test that we propose here is based on the proportion of times that the direction of change in y_t is correctly predicted in the sample. The test does not require quantitative information on y and uses only information on the signs of y_t and x_t. Furthermore, it is valued for a wide class of underlying probability distributions and only requires that the probability of changes in the direction of y_t (and x_t) is time-invariant and does not take the extreme values of 0 and 1. [This requirement is clearly satisfied if the joint distribution of (y_t, x_t) is continuous and stationary. The stationarity assumption, however, is not a necessary condition for the test to be valid. For example, the test is applicable to nonstationary distributions as long as y_t and x_t are symmetrically distributed around 0.]

Consider first the relatively simple case in which the focus of interest is on the correct prediction of the signs of y_t and x_t. Introduce the indicator variables

$$\begin{aligned} Y_t &= 1 \quad \text{if } y_t > 0, \\ &= 0 \quad \text{otherwise}, \\ X_t &= 1 \quad \text{if } x_t > 0 \\ &= 0 \quad \text{otherwise}, \end{aligned}$$

and

$$\begin{aligned} Z_t &= 1 \quad \text{if } y_t x_t > 0 \\ &= 0 \quad \text{otherwise}. \end{aligned}$$

Let $P_y = \Pr(y_t > 0)$ and $P_x = \Pr(x_t > 0)$, and denote by $\hat{P}$ the proportion of times that the sign of y_t is predicted correctly; then $\hat{P} = n^{-1} \Sigma_{t=1}^{n} Z_t = \overline{Z}$. Throughout we assume that $1 > P_y$ and $P_x > 0$. (This assumption is likely to be satisfied when y_t and x_t refer to changes in economic variables.) On the assumption that y_t and x_t are independently distributed (i.e., x_t has no power in predicting y_t), $n\hat{P}$ has a binomial distribution with mean nP_*, where

$$\begin{aligned} P_* = \Pr(Z_t = 1) &= \Pr(y_t x_t > 0) \\ &= \Pr(y_t > 0, x_t > 0) \\ &\quad + \Pr(y_t < 0, x_t < 0) \\ &= P_y P_x + (1 - P_y)(1 - P_x). \end{aligned} \quad (1)$$

In the case in which P_y and P_x are known, the predictive-failure test can be based on the standardized binomial variate

$$S_n = \left\{ \frac{P_*(1 - P_*)}{n} \right\}^{-1/2} (\hat{P} - P_*),$$

which is asymptotically distributed as $N(0, 1)$. (An example of known P_y and P_x arises when it is known that

This content downloaded from
195.128.0.89 on Tue, 17 Feb 2026 15:24:39 UTC
All use subject to https://about.jstor.org/terms

y_t and x_t are distributed symmetrically around 0, in which case $P_y = P_x = \frac{1}{2}$.)

When the probabilities P_y and P_x are not known, the test can be based on $S_n = \hat{P} - \hat{P}_*$, which is a Hausman (1978) type statistic, and compares $\hat{P}$ with the estimator of its expectations obtained under the null—namely, $\hat{P}_*$. Using (1), we have $\hat{P}_* = \hat{P}_y\hat{P}_x + (1 - \hat{P}_y)(1 - \hat{P}_x)$, and on the null hypothesis that y_t and x_t are independently distributed, P_y and P_x are efficiently estimated by $\hat{P}_y = \Sigma_{t=1}^n Y_t/n = \overline{Y}$ and $\hat{P}_x = \Sigma_{t=1}^n X_t/n = \overline{X}$. In this case, S_n does not have a binomial distribution. We have

$$S_n = (\overline{Z} - P_*) + (1 - 2P_y)(\overline{X} - P_x) + (1 - 2P_x)(\overline{Y} - P_y) - 2(\overline{Y} - P_y)(\overline{X} - P_x), \quad (2)$$

where it is easily seen that $n^{1/2}(\overline{Z} - P_*) \underset{a}{\sim} N(0, P_*(1 - P_*))$, $n^{1/2}(\overline{X} - P_x) \underset{a}{\sim} N(0, P_x(1 - P_x))$, $n^{1/2}(\overline{Y} - P_y) \underset{a}{\sim} N(0, P_y(1 - P_y))$, and plim $(X - P_x) =$ plim $(\overline{Y} - P_y) = 0$. Hence using these results in (2) yields

$$\sqrt{n}S_n \underset{a}{\sim} N(0, V_S), \quad (3)$$

where V_S denotes the asymptotic variance of $\sqrt{n}S_n$. To derive V_S, we first note that since $\hat{P}_*$ is based on efficient estimators of P_y and P_x, it is itself an efficient estimator of P_*. [Asymptotically, the minimum value of var$(\sqrt{n}\hat{P}_*)$, say V_*, is given by the Cramer–Rao lower bound,

$$V_* = (\partial P_*/\partial\theta)' I_\theta^{-1}(\partial P_*/\partial\theta),$$

where $\theta' = (P_y, P_x)$, $I_\theta = E[(-n^{-1})\ \partial^2\ell(\theta)/\partial\theta\partial\theta']$ is the information matrix, and $\ell(\theta)$ is the log-likelihood function: $n^{-1}\ell(\theta) = \overline{Y}\log P_y + (1 - \overline{Y})\log(1 - P_y) + \overline{X}\log P_x + (1 - \overline{X})\log(1 - P_x)$. It is now easily seen that under the null hypothesis $V_* = (2P_y - 1)^2P_x(1 - P_x) + (2P_x - 1)^2P_y(1 - P_y)$.] Therefore, using a result due to Hausman (1978), we have $n^{-1}V_S = \text{var}(S_n) = \text{var}(\hat{P}) - \text{var}(\hat{P}_*)$, where $\text{var}(\hat{P}) = n^{-1}P_*(1 - P_*)$. To obtain the expression for var$(\hat{P}_*)$, we first write $\hat{P}_*$ as $\hat{P}_* = 2\overline{Y}\overline{X} - \overline{Y} - \overline{X} + 1$, or

$$\text{var}(\hat{P}_*) = 4\ \text{var}(\overline{Y}\overline{X}) + \text{var}(\overline{Y}) + \text{var}(\overline{X}) - 4\ \text{cov}(\overline{Y}\overline{X}, \overline{Y}) - 4\ \text{cov}(\overline{Y}\overline{X}, \overline{X}) + 2\ \text{cov}(\overline{Y}, \overline{X}). \quad (4)$$

But under the null hypothesis, $\text{cov}(\overline{Y}, \overline{X}) = 0$, $\text{cov}(\overline{X}\overline{Y}, \overline{X}) = E(\overline{Y})\ \text{var}(\overline{X})$, $\text{cov}(\overline{X}\overline{Y}, \overline{Y}) = E(\overline{X})\ \text{var}(\overline{Y})$, $\text{var}(\overline{Y}\overline{X}) = E(\overline{Y}^2)E(\overline{X}^2) - [E(\overline{Y})E(\overline{X})]^2$, $E(\overline{Y}) = P_y$, $E(\overline{X}) = P_x$, $\text{var}(\overline{Y}) = n^{-1}P_y(1 - P_y)$, and $\text{var}(\overline{X}) = n^{-1}P_x(1 - P_x)$. Using these results in (4) yields

$$n^{-1}\ \text{var}(\hat{P}_*) = (2P_y - 1)^2P_x(1 - P_x) + (2P_x - 1)^2P_y(1 - P_y) + 4n^{-1}P_yP_x(1 - P_y)(1 - P_x), \quad (5)$$

which asymptotically achieves the Cramer–Rao lower bound given previously.

In the general case, a nonparametric test of predictive performance of x_t can be based on the standardized statistic

$$S_n = \frac{\hat{P} - \hat{P}_*}{\{\hat{\text{var}}(\hat{P}) - \hat{\text{var}}(\hat{P}_*)\}^{1/2}} \underset{a}{\sim} N(0, 1), \quad (6)$$

where $\hat{\text{var}}(\hat{P}) = n^{-1}\hat{P}_*(1 - \hat{P}_*)$ and $\hat{\text{var}}(\hat{P}_*) = n^{-1}(2\hat{P}_y - 1)^2\hat{P}_x(1 - \hat{P}_x) + n^{-1}(2\hat{P}_x - 1)^2\hat{P}_y(1 - \hat{P}_y) + 4n^{-2}\hat{P}_y\hat{P}_x(1 - \hat{P}_y)(1 - \hat{P}_x)$. The last term in the expression for $\hat{\text{var}}(\hat{P}_*)$ is asymptotically negligible. This test that compares the signs of y_t and x_t should not be confused with the familiar *sign test*, which is often used to test the hypothesis that y_t and x_t have the same distribution functions. The sign test proper is based on the proportion of times that the sign of $y_t - x_t$ changes in the sample. [For a more detailed discussion of the sign test, see, for example, Bickel and Doksum (1977, sec. 9.2).]

2. A GENERALIZATION

The s_n test can be readily extended to situations in which the information on y_t and x_t are categorized into more than two classes. Suppose that in the general case there are m categories into which the realizations of y and x may fall, and let n_{ij} denote the number of observations in the $Y_i \times X_j$ category cross-tabulated in the form of an $m \times m$ contingency table. Moreover, let n_{io} and n_{oj} represent the ith row and the jth column totals. The S_n statistic for this general formulation may now be written as

$$S_n = \hat{P} - \hat{P}_* = \sum_{i=1}^{m} (\hat{P}_{ii} - \hat{P}_{io}\hat{P}_{oi}), \quad (7)$$

where $\hat{P}_{ij} = n_{ij}/n$, $\hat{P}_{io} = n_{io}/n$, $\hat{P}_{oj} = n_{oj}/n$, and n is the total number of observations. As before, $\hat{P} = \Sigma_{i=1}^m \hat{P}_{ii}$ measures the proportion of times that the direction of change in y_t (say, $+$, $=$, $-$) is correctly predicted, and $\hat{P}_* = \Sigma_{i=1}^m \hat{P}_{io}\hat{P}_{oi}$. The implicit null hypothesis underlying (7) is given by

$$H_o^*: \sum_{i=1}^{m} (P_{ii} - P_{io}P_{oi}) = 0,$$

and in general is far less restrictive than the null hypothesis of independence, H_o: $P_{ij} = P_{io}P_{oj}$ for all (i, j). It is easily seen that in the 2×2 case the predictive failure hypothesis, H_o^*, and the independence hypothesis, H_o, coincide. [The equivalence of H_o^* and H_o for $m = 2$ follows simply from the fact that in this case both hypotheses imply the same *single* restriction, $P_{11}P_{22} = P_{21}P_{12}$. For the purpose of testing predictive performance there is no merit in focusing on a particular cell in the 2×2 contingency table. This is because in this case the test of predictive performance has only 1 df when the row and column frequencies are unknown. It is important to note that this test differs from that derived by Henriksson and Merton (1981) and Merton (1981) which assumes known row and column frequen-

This content downloaded from
195.128.0.89 on Tue, 17 Feb 2026 15:24:39 UTC
All use subject to https://about.jstor.org/terms

cies. For a discussion of the relative merits of the two testing procedures, see Pesaran and Timmermann (1992).] It also follows that the predictive failure test proposed in this article and the familiar χ^2 goodness-of-fit test based on contingency tables,

$$\chi^2 = \sum_{i,j=1}^{m} \frac{\left(n_{ij} - \frac{n_{io}n_{oj}}{n}\right)}{\frac{n_{io}n_{oj}}{n}} \underset{a}{\sim} \chi^2_{(m-1)^2}, \tag{8}$$

are asymptotically equivalent for the 2×2 case. [A direct proof of this proposition can be obtained from the authors on request, one in which it is shown that under the null hypothesis the difference between s_n^2 and the goodness-of-fit statistics in finite samples lies in the presence of the asymptotically negligible term $4n^{-1}P_yP_x(1 - P_y)(1 - P_x)$ in (5), which does not enter the χ^2 statistic.]

In general, however, the predictive failure and the χ^2 tests are not equivalent and can lead to different conclusions even in large samples. For $m > 2$, the independence hypothesis, H_o, implies H_o^*, but not vice versa, and the χ^2 test in general will be rather conservative as a test of predictive performance. A formal comparison of the local power functions of the S_n and the χ^2 tests is given in the Appendix. (Apart from H_o and H_o^*, there are also other hypotheses of interest that focus on the diagonal elements of the contingency table either individually or jointly.)

To derive the asymptotic distribution of S_n, let $\mathbf{P}' = (P_{11}, P_{12}, \ldots, P_{1m}; P_{21}, P_{22}, \ldots; P_{m1}, P_{m2}, \ldots, P_{mm})$ and, using familiar results on the maximum likelihood ML estimator of P_{ij}'s, note that

$$\sqrt{n}(\hat{\mathbf{P}} - \hat{\mathbf{P}}_o) \underset{a}{\sim} N(0, \Psi_o - \mathbf{P}_o\mathbf{P}_o'), \tag{9}$$

where $\hat{\mathbf{P}}$ is the ML estimator of $\mathbf{P}$, $\mathbf{P}_o$ is its "true" value, and Ψ_o is an $m^2 \times m^2$ diagonal matrix with $\mathbf{P}_o$ as its diagonal elements. For example, in the 2×2 case,

$$\Omega = \Psi - \mathbf{PP}'$$

$$= \begin{bmatrix} P_{11}(1-P_{11}) & -P_{11}P_{12} & -P_{11}P_{21} & -P_{11}P_{22} \\ -P_{11}P_{12} & P_{12}(1-P_{12}) & -P_{12}P_{21} & -P_{12}P_{22} \\ -P_{11}P_{21} & -P_{12}P_{21} & P_{21}(1-P_{12}) & -P_{21}P_{22} \\ -P_{11}P_{22} & -P_{12}P_{22} & -P_{21}P_{22} & P_{22}(1-P_{22}) \end{bmatrix}.$$

Since S_n is a well-defined function of $\hat{\mathbf{P}}$, say $f(\hat{\mathbf{P}}) = \Sigma_{i=1}^m (\hat{P}_{ii} - \hat{P}_{io}\hat{P}_{oi})$, it is easily seen that under $H_o^* \sqrt{n}S_n \underset{a}{\sim} N(0, V_s)$, where

$$V_s = \left(\frac{\partial f(\mathbf{P}_o)}{\partial \mathbf{P}}\right) (\Psi - \mathbf{P}_o\mathbf{P}_o') \left(\frac{\partial f(\mathbf{P}_o)}{\partial \mathbf{P}}\right)$$

and

$$\frac{\partial f(\mathbf{P})}{\partial P_{ij}} = 1 - P_{oi} - P_{io} \quad \text{for } i = j$$

$$= \; - P_{jo} - P_{oi} \quad \text{for } i \neq j.$$

For example, in the 2×2 case we have

$$\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}} = \begin{bmatrix} 1 - P_{o1} - P_{1o} \\ -P_{o1} - P_{2o} \\ -P_{o2} - P_{1o} \\ 1 - P_{o2} - P_{2o} \end{bmatrix}.$$

In the general case, the test of predictive performance can be based on the standardized statistic $s_n = \sqrt{n}V_n^{-1/2}S_n \underset{a}{\sim} N(0, 1)$, where

$$\hat{V}_n = \left(\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}\right)_{\mathbf{P}=\hat{\mathbf{P}}} (\hat{\Psi} - \hat{\mathbf{P}}\hat{\mathbf{P}}') \left(\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}\right)_{\mathbf{P}=\hat{\mathbf{P}}}.$$

This test statistic is relatively easy to compute and is particularly appropriate when the focus of the analysis is on predicting the overall changes in a variable rather than on the general hypothesis of statistical independence.

3. TWO APPLICATIONS

Here we initially provide an application of the test to the CBI Industrial Trends Surveys. These surveys cover a significant proportion of the firms in the British manufacturing sector and ask each firm about past and expected trends in a number of variables, including their average selling prices. The survey results are in the form of qualitative responses reporting whether firms expect their prices to "go up," "stay the same," or to "go down." Each firm is also asked about past trends in prices again in the form of "up," "same," and "down" categories. The test developed in this article can be applied either directly to the responses at the firm level, or alternatively it could be applied to the "average" responses of firms over time. (Unfortunately, the survey results published by the CBI are in the form of average responses and do not give data on individual firms. But see what follows for an application of the test to individual responses from the INSEE surveys.) Different methods of averaging across individual responses were discussed in the literature by Anderson (1952), Theil (1952), Carlson and Parkin (1975), Pesaran (1984), and others. (A review of these procedures can be found in Pesaran [1987, chap. 8].) A popular measure is the *balance* statistic, defined as the difference between the percentage of respondents who report (or expect) a price increase and the percentage of respondents who report (or expect) a price decrease. Let ${}_{t-1}R_t^e$ be the proportion of firms (appropriately weighted to account for their size differences) that at time $t - 1$ expect their prices to rise over the period $t - 1$ to t, and ${}_{t-1}F_t^e$ be the proportion of firms that expect prices to fall. Moreover, denote the past actual trends corresponding to ${}_{t-1}R_t^e$ by R_t and ${}_{t-1}F_t^e$ by F_t. The balance statistics ${}_{t-1}B_t^e = {}_{t-1}R_t^e - {}_{t-1}F_t^e$ and $B_t = R_t - F_t$ may now be used to measure the expected and the actual rate of price changes. The accuracy of the balance statistics as quantitative measures of "average" price changes, however, depends on the underlying distribution of price changes across firms. (See Pesaran 1987, ex. 3.1; Theil 1952). The application of the nonpara-

This content downloaded from
195.128.0.89 on Tue, 17 Feb 2026 15:24:39 UTC
All use subject to https://about.jstor.org/terms

Table 1. Contingency Table for Actual Versus Expected Changes in Demand (INSEE, January 1985)

D/D^e	+	=	−
+	122	149	39
=	87	503	153
−	37	203	335

metric test developed in Section 1 allows one to test the performance of $x_t = {}_{t-1}B_t^e - B_{t-1}$ as a predictor of $y_t = B_t - B_{t-1}$ without having to make any distributional assumptions about the processes generating price changes across firms.

Using the CBI Industrial Trends Surveys of Pesaran (1987, table A.1), we obtain the following estimates for the period 1959(1)–1985(4):

$$\hat{P}_y = .5579, \quad \hat{P}_x = .7474, \quad \hat{P} = .7684,$$

$$\hat{P}_* = .5286, \hat{V}(\hat{P}) = .00263, \quad \hat{V}(\hat{P}_*) = .00068.$$

Substituting these estimates in (6) now yields $s_n = 5.44$, which is well above the 95% critical value of a standard normal variate and strongly rejects the hypothesis that x_t and y_t are independently distributed. There is a clear evidence of predictive power in survey expectations for the actual movement of price changes.

The square of the s_n statistic in this case is directly comparable with the χ^2 goodness-of-fit statistic given in (8). The value of this latter statistic turned out to be 29.3, which is only marginally different from the value of $s_n^2 = 29.6$. (Recall that in the 2×2 case, the s_n and the χ^2 tests are asymptotically equivalent.)

Our second application relates to survey responses at the level of individual firms and applies the s_n test to the realizations (D) and expectations (D^e) of changes in demand classified into the three categories—"increase" (+), "stay the same" (=), and "decrease" (−). The source of the data is INSEE, and they were kindly provided to us by Marc Ivaldi. The number and identity of participating firms varies over time, and the survey is based on periodic random replacement. (See Ivaldi [1991] for a more detailed description of the data.)

Table 1 shows the observed frequencies of actual changes in demand versus the expected changes in demand for January 1985. We also considered the data for March, June, and October 1985 to see whether the test outcome is sensitive to seasonal effects. The results are summarized in Table 2, and give both the s_n and the χ^2 statistics. For ease of comparison, we give the square of the s_n statistic with the relevant 5% critical values in parentheses. Both tests clearly reject the hypothesis that there are no predictable relationships between expected and actual changes in demand. In a way, as far as the illustration of the differences of the two tests is concerned the present application is rather unfortunate. We would have preferred an application with a less clear-cut outcome, but we believe that the present application still adequately highlights the main differences that exist between the proposed test and the goodness-of-fit test. The two tests focus on different aspects of the relationships between realized and expected changes in demand and can in principle result in different conclusions.

Table 2. Summary Statistics for the 3 × 3 Case (demand expectations, INSEE Surveys)

Date	Predictive-failure statistic (s_n^2) (χ_1^2)	Goodness-of-fit statistic (χ_4^2)
January 1985	285.6 (3.84)	410.7 (9.49)
March 1985	194.6 (3.84)	254.1 (9.49)
June 1985	182.8 (3.84)	276.1 (9.49)
October 1985	175.6 (3.84)	294.5 (9.49)

Finally, it is perhaps worth emphasizing that the test statistics in the preceding two applications measure the degree of association between realized and expected changes in prices and demand and do not provide a test for the rationality of expectations. [For direct tests of rational expectations hypothesis applied to the French Survey data, see, for example, Gourieroux and Pradel (1986) and Ivaldi (1991).] It is possible to imagine cases in which the rational-expectations hypothesis is not rejected, and yet expectations have little predictive power (e.g. stock-market returns under the efficient-market hypothesis). Similarly, examples may occur in which there is evidence of substantial predictive power in the expectations, although the rational-expectations hypothesis is rejected. This case arises, for example, when the information used to produce the forecasts is not exploited optimally.

ACKNOWLEDGMENTS

Helpful comments from Nick Kiefer, Richard Smith, and an anonymous referee of the journal are gratefully acknowledged.

APPENDIX: ASYMPTOTIC LOCAL POWER OF TESTS FOR HIERARCHICAL HYPOTHESES

Consider the two hypotheses H_o: $f(\boldsymbol{\theta}) = 0$ and H_o^*: $h(f(\boldsymbol{\theta})) = 0$, where $\boldsymbol{\theta}$ is a $k \times 1$ vector of unknown parameters, $f(\cdot)$ and $h(\cdot)$ are vector functions of dimensions s and r such that $k > s \geq r > 0$, and $h(0) = 0$. Assume also that the matrix derivatives $\partial f(\theta)/\partial\theta$ and $\partial h/\partial f$ are well defined and have full ranks.

The hypotheses H_o and H_o^* have a hierarchical structure in the sense that H_o implies H_o^* but not the reverse. (Clearly, it is possible to extend the structure to include further hypotheses and branches.) The independence hypothesis and the predictive-failure hypothesis discussed in Section 2 are examples of such a hierarchical structure. In this appendix, we compare the local power functions of the Wald tests of H_o and H_o^* and in par-

This content downloaded from
195.128.0.89 on Tue, 17 Feb 2026 15:24:39 UTC
All use subject to https://about.jstor.org/terms

ticular examine the problem of choosing between the two tests when the hypothesis of interest is H_o^*. (In the case in which H_o is the hypothesis of interest, the choice problem does not arise, and a test of H_o^* will not yield the correct size as far as testing H_o is concerned.)

Let $\hat{\theta}_n$ be the ML estimator of θ based on n observations, and suppose that $\sqrt{n}(\hat{\theta}_n - \theta_o)$ has an asymptotic normal distribution with mean 0 and the nonsingular variance–covariance matrix $V(\boldsymbol{\theta}_o)$. The Wald statistic for tests of H_o and H_o^* are then given by

$$W_n = nf'(\hat{\boldsymbol{\theta}}_n)(F_n V(\hat{\boldsymbol{\theta}}_n)F_n')^{-1}f(\hat{\boldsymbol{\theta}}_n) \qquad \text{(A.1)}$$

and

$$W_n^* = nh'(f(\hat{\boldsymbol{\theta}}_n))(H_n F_n V(\hat{\boldsymbol{\theta}}_n)F_n' H_n')^{-1}h(f(\hat{\boldsymbol{\theta}}_n)), \qquad \text{(A.2)}$$

where $\boldsymbol{\theta}_o$ is the true value of $\boldsymbol{\theta}$, and

$$F_n = \left.\frac{\partial f}{\partial \theta}\right|_{\theta - \theta_*}, \quad H_n = \left.\frac{\partial h}{\partial f}\right|_{\theta = \theta_*}$$

are $s \times k$ and $r \times s$ matrices of derivatives evaluated at $\hat{\boldsymbol{\theta}}_n$.

Consider now the local alternatives

$$H_{1n}\colon \boldsymbol{\theta} = \boldsymbol{\theta}_o + n^{-1/2}\boldsymbol{\delta}, \qquad \text{(A.3)}$$

where $\boldsymbol{\delta}$ is a fixed vector of constants such that $n^{-1}(\boldsymbol{\delta}'\boldsymbol{\delta})$ has a finite limit. It is well known that under local alternatives, H_{1n}, the asymptotic distributions of the Wald statistic W_n and W_n^* tend to noncentral chi-squared variates with s and r degrees of freedom, respectively. The noncentrality parameters corresponding to these distributions are given by

$$w = \operatorname*{plim}_{n\to\infty} (W_n \mid H_{1n}) \quad \text{and} \quad w^* = \operatorname*{plim}_{n\to\infty} (W_n^* \mid H_{1n}), \qquad \text{(A.4)}$$

respectively. To derive these noncentrality parameters, we first consider the first-order expansions $f(\hat{\theta}_n) = f(\theta_o) + F(\theta_o)(\hat{\theta}_n - \theta_o) + o_P(n^{-1})$ and $h(f(\hat{\theta}_n)) = h(f(\theta_o)) + H(\theta_o)F(\theta_o)(\hat{\theta}_n - \theta_o) + o_P(n^{-1})$, where $F(\theta_o) = \partial f/\partial\theta$ and $H(\theta_o) = \partial h/\partial f$ evaluated at $\theta = \theta_o$. Note also that $f(\theta_o) = 0$ and hence $h(f(\theta_o)) = 0$. Therefore, using (A.4) under local alternatives, we have

$$w = \delta' F'(FVF')^{-1}F\delta$$

and

$$w^* = \delta' F' H'(HFVF'H')^{-1}HF\delta,$$

where for ease of exposition the explicit dependence of F, H, and V on θ_o is suppressed.

To compare the noncentrality parameters of the two tests, let $\eta = F\delta$ and $A = FVF'$, and write

$$\begin{aligned} w - w^* &= \eta' A^{-1/2}[I - A^{1/2}H'(HAH')^{-1}HA^{1/2}]A^{-1/2}\eta \\ &= \eta_*'[I - X(X'X)^{-1}X']\eta_* \ge 0, \end{aligned} \qquad \text{(A.5)}$$

where $\eta_* = A^{-1/2}\eta$ and $X = A^{1/2}H'$. Hence, in general, the χ^2 test of H_o based on W_n has a larger noncentrality parameter than that of the test based on W_n^*, but it does not follow from this that the W_n test is preferable to the W_n^* test. This is because the degrees of freedom of the W_n are generally larger than that of the W_n^* test. As shown, for example, by Das Gupta and Perlman (1974), the asymptotic power of the χ^2 test is strictly decreasing in its degrees of freedom and increasing in its noncentrality parameter. It is, therefore, not possible to rank the W_n and W_n^* tests on the basis of their local power functions when the aim is to test H_n^*.

REFERENCES

Anderson, O. (1952), "The Business Test of the IFO-Institute for Economic Research, Munich, and its Theoretical Model," *Revue de l'Institut International de Statistique*, 20, 1–17.

Bickel, P. J., and Doksum, K. A. (1977), *Mathematical Statistics: Basic Ideas and Selected Topics*, Oakland, CA: Holden-Day.

Carlson, J. A., and Parkin, M. (1975), "Inflation Expectations," *Economica*, 42, 123–138.

Das Gupta, S., and Perlman, M. D. (1974), "Power of the Noncentral *F*-Test: Effect of Additional Variates on Hotelling's T^2-Test," *Journal of the American Statistical Association*, 69, 174–180.

Gourieroux, C., and Pradel, J. (1986), "Direct Tests of the Rational Expectation Hypothesis," *European Economic Review*, 30, 265–284.

Hausman, J. A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1272.

Henriksson, R. D., and Merton, R. C. (1981), "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business*, 54, 513–533.

Ivaldi, M. (1991), "Survey Evidence on the Rationality of Expectations," mimeo, GREMAQ, Toulouse.

Merton, R. C. (1981), "On Market Timing and Investment Performance. I. An Equilibrium Theory of Value for Market Forecasts," *Journal of Business*, 54, 363–406.

Pesaran, M. H. (1984), "Expectations Formations and Macroeconomic Modelling," in *Contemporary Macroeconomic Modelling*, eds. P. Malgrange and P. A. Muetz, Oxford, U.K.: Basil Blackwell.

——— (1987), *The Limits to Rational Expectations* (reprinted with corrections, 1989), Oxford, U.K.: Basil Blackwell.

Pesaran, M. H., and Timmermann, A. G. (1992), "A Note on the Non-parametric Henriksson–Merton Test of Market Timing," unpublished manuscript.

Theil, H. (1952), "On the Time Shape of Economic Microvariables and the Munich Business Test," *Revue de l'Institut International de Statistique*, 20, 105–120.

This content downloaded from
195.128.0.89 on Tue, 17 Feb 2026 15:24:39 UTC
All use subject to https://about.jstor.org/terms