A Consistent Nonparametric Test of Symmetry in Linear Regression Models

Yanqin FAN and Ramazan GENCAY*

This article proposes a *consistent* nonparametric test of the hypothesis that the disturbance in a linear regression model is distributed symmetrically around zero. Simulation results show that the test has good size and power properties for sample sizes as small as 50. We illustrate the use of the test in a cross-country model of inflation and monetary growth.

KEY WORDS: Discrepancy measure; Monte Carlo simulation; Nonparametric kernel estimation.

1. INTRODUCTION AND MOTIVATION

Consider the following linear regression model:

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where $\mathbf{x}_i$ is a $d \times 1$ vector of regressors that may include a constant, β is a $d \times 1$ vector of parameters, $(\mathbf{x}_i', \varepsilon_i)$ is identically and independently distributed, and the disturbance ε_i has a density function with zero mean conditionally on the regressors, $f(\varepsilon_i \mid \mathbf{x}_i)$.

There exists a large literature on efficient estimation of β. It is well known that the most efficient estimator of β is the maximum likelihood estimator that requires that the functional form of f be known. In particular, if f is known to be normal, then the maximum likelihood estimator is the ordinary least squares (OLS) estimator. If the functional form of the error distribution is unknown, OLS is still consistent and asymptotically normal but may not be the most efficient one, and the exact maximum likelihood estimator is not available.

Bickel (1982) has shown that if the conditional density function of the disturbance is symmetric around zero, then the regression coefficients, β, can be estimated adaptively. This implies that the maximum attainable efficiency for the regression coefficients is that of the maximum likelihood estimator that could be obtained if the actual (unknown) functional form of the disturbance distribution were used in forming the likelihood. Newey (1988) has constructed adaptive estimators of β by a generalized method of moments when the foregoing is true. Therefore, it seems important to construct *consistent* tests of symmetry in linear regression models and to use the tests to verify the assumption of symmetry before applying the adaptive estimators. This is what this article attempts to accomplish. For an overview, the reader may refer to the work of Eubank, Hart, and LaRiccia (1993) for an extensive survey of the recent developments of testing the one-sample goodness of fit and the goodness of fit of a polynomial regression model.

In Section 2 we formalize the hypothesis and introduce the idea underlying the construction of our test as well as the estimators. In Section 3 we derive the main distribution result, construct the test statistic, and show that the test is consistent. We conduct simulation experiments to examine the performance of our test in finite samples in Section 4. Finally, in Section 5 we apply the test to a cross-country model of inflation and monetary growth and show that the null hypothesis of symmetry is retained.

2. THE DISCREPANCY MEASURE AND ESTIMATION

As motivated in Section 1, we are interested in testing H_0 : $f(\varepsilon \mid \mathbf{x}) = f(-\varepsilon \mid \mathbf{x})$ almost everywhere (a.e. hereafter) against the alternative $H_1 : f(\varepsilon \mid \mathbf{x}) \neq f(-\varepsilon \mid \mathbf{x})$ a.e. If H_0 is accepted, then Newey's adaptive estimators may be used to estimate β. For ease of exposition, we will assume that the regressors in (1) are fixed and construct a test for $H_0' : f(\varepsilon) = f(-\varepsilon)$ a.e. against the alternative $H_1' : f(\varepsilon) \neq f(-\varepsilon)$ a.e.

Suppose for a moment that ε is observable. The approach we take is to estimate $f(\varepsilon)$ by kernel density estimate $\hat{f}(\varepsilon)$ and then measure the closeness between $\hat{f}(\varepsilon)$ and $\hat{f}(-\varepsilon)$ by an appropriate discrepancy measure. This approach has already been used by Ahmad (1980) and Robinson (1991). There are several discrepancy measures, such as the integrated square difference and the integrated absolute difference, that can be used for this purpose. The one that we use was proposed by Ahmad and Van Belle (1974). It is defined for two probability distribution functions F and G with probability density functions f and g as

$$\lambda = \lambda(F, G) = \frac{2\delta}{\Delta(f) + \Delta(g)}, \tag{2}$$

where $\delta = \int f(\mathbf{x})g(\mathbf{x})\, d\mathbf{x}$, $\Delta(f) = \int f^2(x)\, d\mathbf{x}$, and $\Delta(g) = \int g^2(x)\, d\mathbf{x}$, where $g^2(\mathbf{x}) = (g(\mathbf{x}))^2$ and $f^2(\mathbf{x}) = (f(\mathbf{x}))^2$. The measure λ possesses the following properties that make it a desirable basis for hypothesis testing:

$$\lambda \le 1 \quad \text{and the equality holds iff} \quad f(\mathbf{x}) = g(\mathbf{x}) \text{ a.e.}$$

The measure λ is chosen so that the corresponding results of our earlier work (Fan and Gencay 1993) can be used, which will greatly simplify the derivation of the test statistic in this article. Note, however, that $(1 - \lambda)[\Delta(f) + \Delta(g)]$ equals the integrated square difference between f and g. Thus one can easily modify the test statistic based on λ to obtain a statistic based on the integrated square difference.

* Yanqin Fan is Associate Professor and Ramazan Gencay is Assistant Professor, Department of Economics, University of Windsor, Ontario N9B 3P4, Canada. The authors would like to thank the editor, the associate editor, and the two referees for their constructive comments and the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for financial support.

Journal of the American Statistical Association
June 1995, Vol. 90, No. 430, Theory and Methods

To measure the affinity between $f(\varepsilon)$ and $f(-\varepsilon)$, (2) can be utilized by setting $g(\varepsilon) = f(-\varepsilon)$. From the equality of $\int f^2(\varepsilon)\, d\varepsilon = \int f^2(-\varepsilon)\, d\varepsilon$, definition (2) simplifies to $\lambda^* = \delta^*/\Delta(f)$, where $\delta^* = \int f(\varepsilon) f(-\varepsilon)\, d\varepsilon$. Thus H_0' is equivalent to $H_0'': \lambda^* = 1$ and H_1' is equivalent to $H_1'': \lambda^* < 1$.

We earlier (Fan and Gencay 1993) proposed an estimator of λ^*, on which basis an asymptotically valid test for H_0' was constructed. In this article we extend these results to the case where ε is unobservable. In particular, we focus on model (1).

Let $K(\cdot)$ be a known symmetric probability density function (pdf) satisfying the following conditions:

K1. $\sup_u K(u) < \infty$ and $|u| K(u) \to 0$, as $|u| \to \infty$.
K2. $\int uK(u)\, du = 0$, and $\int u^2 K(u)\, du < \infty$.
K3. The kernel $K(\cdot)$ has derivatives up to $(m+1)$th order, where m is a positive integer and is greater than 2. In addition, $K^{(s)}(\cdot)$ is a continuous function of bounded variation for $s = 0, 1, \ldots, m$ and $K^{(m+1)}(\cdot)$ is bounded.

Furthermore, let $\{a_n\}$ be a sequence of nonnegative real numbers such that

A. $a_n \to 0$, $na_n^{(2+4/m)} \to \infty$, and $na_n^4 \to 0$, as $n \to \infty$.

The kernel estimate of $f(\varepsilon)$ is given by

$$\hat{f}(\varepsilon) = \frac{1}{na_n} \sum_{i=1}^{n} K\left(\frac{\varepsilon - \varepsilon_i}{a_n}\right). \quad (3)$$

Using ε_i and $\hat{f}(\varepsilon_i)$, Fan and Gencay (1993) proposed estimating λ^* by

$$\hat{\lambda}_\gamma^* = \frac{\hat{\delta}_\gamma^*}{\hat{\Delta}(f)}, \quad (4)$$

where $\hat{\Delta}(f) = \int \hat{f}^2(\varepsilon)\, d\varepsilon$, $\hat{\delta}_\gamma^* = n_\gamma^{-1} \sum_{i=1}^{n} C_i(\gamma) \hat{f}(-\varepsilon_i)$,

$$C_i(\gamma) = \begin{cases} 1+\gamma & \text{for } i \text{ odd} \\ 1-\gamma & \text{for } i \text{ even} \end{cases} \quad \text{and} \quad n_\gamma = \begin{cases} n & \text{for } n \text{ even} \\ n+\gamma & \text{for } n \text{ odd,} \end{cases}$$

and $0 < \gamma \le 1$.

If $\gamma = 0$, then the estimator $\hat{\lambda}_\gamma^*$ reduces to that of Ahmad (1980). But, as we showed in earlier work (Fan and Gencay 1993), $\hat{\lambda}_0^*$ is not a good candidate for constructing a test of H_0, because $\sqrt{n}(\hat{\lambda}_0^* - 1)$ is degenerate.

It is obvious that the estimators defined in (3) and (4) are not operational, because the disturbance in model (1) is not observable. But we can easily construct proxies for ε_i. For instance, the least squares residuals are suitable for this purpose. Let $\mathbf{b}$ be the OLS estimate of β in (1) and let e_i be the OLS residuals; that is, $e_i = y_i - \mathbf{x}_i'\mathbf{b}$. A well-defined estimate of λ^* is given by replacing ε_i in (3) and (4) by e_i. Let the resulting estimates be $\tilde{f}$, $\tilde{\lambda}_\gamma^*$, $\tilde{\Delta}(f)$, and $\tilde{\delta}_\gamma^*$. In the next section we study the asymptotic distribution of $\tilde{\lambda}_\gamma^*$ and construct a test statistic for H_0' on the basis of $\tilde{\lambda}_\gamma^*$.

3. THE TEST STATISTIC AND ITS PROPERTIES

Assume that the vector of regressors is uniformly bounded and that there exists a positive definite matrix $\mathbf{Q}_X$ such that $\lim_{n\to\infty}(\mathbf{X}'\mathbf{X}/n) = \mathbf{Q}_X$, where $\mathbf{X}$ is the $n \times d$ matrix of regressors. Then, we can show with the aid of Lemma 2.1 and Lemma 2.4 of Fan and Gencay (1993) that the following results hold.

Lemma 3.1. Assume that the null hypothesis of symmetry holds. *Then:*

a. If f and its first $(m+1)$ derivatives are bounded, then
$$\tilde{\delta}_\gamma^* - \delta^* = n^{-1} \sum_{i=1}^{n} C_i^*(\gamma)[f(-\varepsilon_i) - Ef(-\varepsilon_i)] + o_p(n^{-1/2}),$$
where $C_i^*(\gamma) = [nn_\gamma^{-1} C_i(\gamma) + 1]$. Consequently, $\sqrt{n}(\tilde{\delta}_\gamma^* - \delta^*)$ is asymptotically normal with mean zero and variance equal to $(4+\gamma^2) \int f(\varepsilon)[f(-\varepsilon) - \delta]^2\, d\varepsilon$.
b. If f and its first $(m+1)$ derivatives are bounded, then $\tilde{\Delta}(f) - \Delta(f) = 2n^{-1} \sum_{i=1}^{n} [f(\varepsilon_i) - Ef(\varepsilon_i)] + o_p(n^{-1/2})$. Consequently, $\sqrt{n}(\tilde{\Delta}(f) - \Delta(f))$ is asymptotically normal with mean zero and variance equal to $4 \int f(\varepsilon)[f(\varepsilon) - \Delta(f)]^2\, d\varepsilon$.

Based on Lemma 3.1, we get the main result of this section. All proofs are given in the Appendix.

Proposition 3.2. Under the assumptions of Lemma 3.1, $\sqrt{n}[\tilde{\lambda}_\gamma^* - \lambda^*]$ is asymptotically normal with mean zero and variance $\sigma_{\gamma,0}^{*2}$, where

$$\sigma_{\gamma,0}^{*2} = \frac{\gamma^2 \int f(\varepsilon)[f(-\varepsilon) - \delta^*]^2\, d\varepsilon}{\Delta^2(f)} = \frac{\gamma^2 \int f(\varepsilon)[f(\varepsilon) - \Delta(f)]^2\, d\varepsilon}{\Delta^2(f)},$$

provided that H_0' holds.

To base an asymptotic test of $H_0'': \lambda^* = 1$ against $H_1'': \lambda^* < 1$ on $\tilde{\lambda}_\gamma^*$, the variance of the limiting distribution of $\sqrt{n}(\tilde{\lambda}_\gamma^* - \lambda^*)$ (i.e., $\sigma_{\gamma,0}^{*2}$) must be estimated. The following estimator serves this purpose; $\hat{\sigma}_{\gamma,0}^{*2}$ is a consistent estimator of $\sigma_{\gamma,0}^{*2}$, where

$$\hat{\sigma}_{\gamma,0}^{*2} = \frac{\gamma^2 \int \tilde{f}(\varepsilon)[\tilde{f}(-\varepsilon) - \tilde{\delta}_\gamma^*]\, d\varepsilon}{\tilde{\Delta}^2(f)}. \quad (5)$$

Hence under H_0'', the test statistic

$$T_\gamma^* = \frac{\sqrt{n}(\tilde{\lambda}_\gamma^* - 1)}{\hat{\sigma}_{\gamma,0}^*} \quad (6)$$

follows the standard normal distribution asymptotically. The test then rejects H_0'' at a given significance level α, if $T_\gamma^* < -Z_\alpha$, where Z_α is such that $P(Z > Z_\alpha) = \alpha$ with $Z \sim N(0, 1)$.

It is important to note that although the test statistic T_γ^* is based on the OLS estimator $\mathbf{b}$ of β, which is not the most efficient estimator unless f is known to be normal, it gives a consistent test, because under the alternative $T_\gamma^*/\sqrt{n}$ converges to $(\lambda^* - 1)/\sigma_{\gamma,0}^* < 0$. So for a given α, the probability of rejecting the null hypothesis if it is not true approaches 1 as sample size n approaches ∞.

Proposition 3.3. The test given by T_γ^* and the left tail of the standard normal distribution is consistent.

4. THE MONTE CARLO EXPERIMENTS

In this section we investigate the size and the power characteristics of the symmetry test. Simulations were conducted for sample sizes of $n = 50$ and $n = 100$ and for six different error distributions. Three of these are symmetric distributions; the rest are asymmetric distributions. The symmetric distributions are standard normal, $N(0, 1)$; Student's t with 5 degrees of freedom, t_5, and double exponential. The asymmetric distributions are chi-squared with 2 degrees of freedom, χ_2^2; lognormal and standard exponential distributions. In each case the errors are standardized to have expectations zero and theoretical variance 1.

For each replication and distribution, n pseudorandom deviates were generated. All pseudorandom number deviates were obtained from the IMSL (International Mathematical and Statistical Libraries 1987). The standard normal, lognormal, χ^2, and standard exponential deviates were obtained from subroutines RNNOR, RNLNL, RNCHI, and RNEXP. The double exponential deviates were generated by

$$
\begin{aligned}
x_t &= (\alpha + \beta \ln 2) + \beta \ln z_t && \text{if } z_t < .5 \\
&= .5 && \text{if } z_t = .5 \\
&= (\alpha - \beta \ln 2) - \beta \ln(1 - z_t) && \text{otherwise,}
\end{aligned}
$$

where z_t is a uniform pseudorandom generate, $U(0, 1)$, and α and β are parameters that determine the mean and the variance of the distribution. The t_5 deviates are obtained by

$$x_t = \frac{y_t}{\sqrt{z_t/5}},$$

where $y_t \sim N(0, 1)$ and $z_t \sim \chi_5^2$.

The $n \times d$ regressor matrices $\mathbf{X}$ are constructed by first obtaining $(d - 1)$ number of $n \times 1$ vectors from $U(0, 1)$ distributed random generates, where $(d - 1)$ is the number of regressors excluding the intercept term and n is the number of observations. $U(0, 1)$ generates are obtained from subroutine RNUN. Each $n \times 1$ vector of regressors is then transformed to have mean zero and theoretical variance 1. By adjoining a $n \times 1$ vector of 1s, the basic $n \times d$ regressor matrix is formed. The regressor matrix is the same for each replication. The resulting $\mathbf{X}'\mathbf{X}/n$ (prime denotes transposition) matrices are displayed in Table 1 for sample sizes $n = 50$ and $n = 100$. The residual vector can be written as

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{M}\mathbf{y} = \mathbf{M}\varepsilon, \quad (7)$$

where $\mathbf{I}$ is a $n \times n$ identity matrix and $\mathbf{M}$ is an idempotent matrix. Because the value taken by residuals does not depend on the value taken by $\boldsymbol{\beta}$, there is no need to specify parameter values for the simulations. For each sample size and error distribution, 1,000 replications were performed.

The nonparametric kernel density estimation of the residuals requires a choice of smoothing parameter. There is extensive literature on the process of choosing optimal smoothing parameters (see Härdle 1990, Marron 1986, and Silverman 1986). Because the restrictions imposed on the smoothing parameter (see A) in this article require undersmoothing, the foregoing methods are not applicable. Instead, we choose the smoothing parameter according to the restrictions placed in assumption A, which permits the following specification:

$$a = \eta \sigma n^{-1/3} \quad (8)$$

where η is a constant and σ is the standard deviation of the residuals. (A Fortran 77 code for the test statistic is available from the authors on request.) With the sample sizes studied here, $\eta = 1.7$ provides a good empirical size which we adopted in the simulations.

Calculation of the test statistic requires a choice for the γ parameter. In small samples, γ should be chosen such that the empirical size of the test is within its nominal size. We experimented with various choices for γ and found that when $.55 < \gamma < .70$, the test statistic has a good empirical size for $n = 50$ and $n = 100$. Accordingly, we set $\gamma = .65$ in this study.

Table 2 presents the simulation results with two regressors (including an intercept term). The first part of Table 2, on the symmetric distributions, gives the empirical size of the test. The test does not overreject or underreject more than 2% of its nominal size. This implies that the test requires no adjustment to yield the desired size.

The results for the asymmetric alternatives are summarized in the second part of Table 2. The asymmetric alternatives determine the power of the test. For both 5% and 10% levels and for sample sizes $n = 50$ and $n = 100$, the test has a good power for all asymmetric alternatives considered in this article.

Tables 3 and 4 extend the size and power calculations of the test with the number of regressors $d = 4$ and $d = 6$. In

Table 1. Regressor Matrices $\mathbf{X}'\mathbf{X}/n$ for the Monte Carlo Simulations

			$n = 50$			
$d = 2$	1.000	−.0164				
	−.0164	1.0683				
$d = 4$	1.000	−.0164	−.0542	−.0506		
	−.0164	1.0683	.1883	.0594		
	−.0542	.1883	.9386	−.0514		
	−.0506	.0594	−.0514	.9953		
$d = 6$	1.0000	−.0164	−.0542	−.0506	.1152	.0300
	−.0164	1.0683	.1883	.0594	.0014	−.1213
	−.0542	.1883	.9386	−.0514	−.1060	−.0022
	−.0506	.0594	−.0514	.9953	.0500	−.0747
	.1152	.0014	−.1060	.0500	1.0226	.1235
	.0300	−.1213	−.0022	−.0747	.1235	1.0714
			$n = 100$			
$d = 2$	1.000	−.0353				
	−.0353	1.0035				
$d = 4$	1.000	−.0353	.0323	−.0548		
	−.0353	1.0035	−.0233	−.1380		
	.0323	−.0233	1.0090	.0071		
	−.0548	−.1380	.0071	1.1256		
$d = 6$	1.0000	−.0353	.0323	−.0548	−.0696	.0076
	−.0353	1.0035	−.0233	−.1380	.2129	.1852
	.0323	−.0233	1.0090	.0071	−.1248	.1072
	−.0548	−.1380	.0071	1.1256	−.1284	−.0577
	−.0696	.2129	−.1248	−.1284	.9208	.0455
	.0076	.1852	.1072	−.0577	.0455	1.0325

Table 2. The Size and the Power of the Symmetry Test With Two Regressors

No. of observations	n = 50		n = 100	
	Symmetric distributions			
Distribution	5%	10%	5%	10%
t_5	.042	.084	.052	.108
N(0, 1)	.045	.094	.062	.102
Double exponential	.064	.126	.047	.095
	Asymmetric distributions			
χ^2_2	.984	.992	1.000	1.000
Lognormal	.912	.932	.994	1.000
Exponential	.980	.994	1.000	1.000

NOTE: The size and power calculations reported in each cell are the ratio of the number of rejections to the number of replications. The critical value of the test statistic at the 5% level is −1.645; at the 10% level is −1.280. Number of replications: 1,000.

Table 4. The Size and the Power of the Symmetry Test with Six Regressors
Number of replications: 1,000

No. of observations	n = 50		n = 100	
	Symmetric distributions			
Distribution	5%	10%	5%	10%
t_5	.045	.095	.055	.093
N(0, 1)	.055	.093	.068	.103
Double exponential	.045	.092	.045	.092
	Asymmetric distributions			
χ^2_2	.814	.922	1.000	1.000
Lognormal	.776	.895	.970	.980
Exponential	.858	.923	1.000	1.000

NOTE: The size and power calculations reported in each cell are the ratio of the number of rejections to the number of replications. The critical value of the test statistic at the 5% level is −1.645 and 10% is −1.280.

both cases an intercept term is included. The reason for increasing the number of regressors is to observe the sensitiveness of the size and the power of the test to the additional regressors. When the number of observations is 100, the increase in the number of regressors does not alter the power of the test with χ^2_2 and exponential distributions. The power of the lognormal distribution falls by 3% to 97.0% at the 5% level and falls by 2% to 98.0% at the 10% level. For sample size 50, the reduction in the power of the test is greater. At the 5% level, the largest decline is 17.5% with the lognormal distribution to 77.6%.

The size of the test is rather stable to an increase in the number of regressors. The percentage of overrejection or underrejection of the empirical size are within 2% of its nominal size at both 5% and 10% levels.

5. AN EMPIRICAL EXAMPLE

To illustrate the use of the test statistic developed in this article, we apply it to a model with cross-country data on inflation and monetary growth. The model is

$$\Delta P_i = \beta_0 + \beta_1 \Delta M_i + \varepsilon_i, \qquad (9)$$

where ΔP_i is the average annual growth rate of consumer prices and ΔM_i is the growth rate of the stock of currency across 83 countries, gathered from the International Financial Statistics. The data set consists of the cross-country inflation and monetary growth of 83 countries during the post-World War II period. In total, there are 83 observations. All growth rates are annual averages for the sample period 1950–1986.

The symmetry test requires that the regression model be linear. One simple way to visualize the data is to plot ΔP_i against ΔM_i (see Fig. 1). Figure 1 brings out the nature of the association between inflation and monetary growth. The figure shows the positive correlation between inflation and growth rate of currency. Further, each one percentage point per year increase in the rate of monetary growth is associated with a roughly one percentage point per year increase in the rate of inflation, which indicates that the relationship is linear. In Table 5 we report the OLS regression of model (9). The coefficient of ΔM_i is statistically significant and is approximately equal to 1, confirming the positive relationship that Figure 1 revealed earlier. To test for the presence of heteroscedasticity in the residual distribution, the method of Cancer (1981) was used. This method is robust to the nonnormality

Table 3. The Size and the Power of the Symmetry Test with Four Regressors

No. of observations	n = 50		n = 100	
	Symmetric distributions			
Distribution	5%	10%	5%	10%
t_5	.044	.087	.054	.106
N(0, 1)	.046	.097	.066	.106
Double exponential	.040	.091	.045	.091
	Asymmetric distributions			
χ^2_2	.954	.962	1.000	1.000
Lognormal	.910	.930	.990	1.000
Exponential	.978	.990	1.000	1.000

NOTE: Number of replications: 1,000.

Table 5. Estimates of Equation (9) with OLS

Variable	Coefficient	Standard error	t-ratio
Constant	−3.7720	.54296	−6.9471
ΔM	1.0197	.02645	38.556
Number of Observations	83		
Standard error of the regression	3.27635		
Sum of squared residuals	869.490		
R-Squared	.948328		
T^*_γ	−.8491		
$\sqrt{b_1}$	−1.8342		
b_2	9.7559		
Heteroscedasticity	3.1456		

NOTE: The upper and lower 5 percentage points of $\sqrt{b_1}$ and b_2 statistics are (−.534, .534), (2.15, 3.99) for $n = 50$ and (−.389, .389), (2.35, 3.77) for $n = 100$. The values are taken from White and MacDonald (1980, p. 20). The heteroscedasticity test is due to Cancer (1981) and involves the regression of squared residuals on the regressor, its second and third powers. It is robust to nonnormality and calculated by the number of observations times the centered R^2 which is distributed χ^2_3. The $\chi^2_{0.05}(3)$ is 7.815.

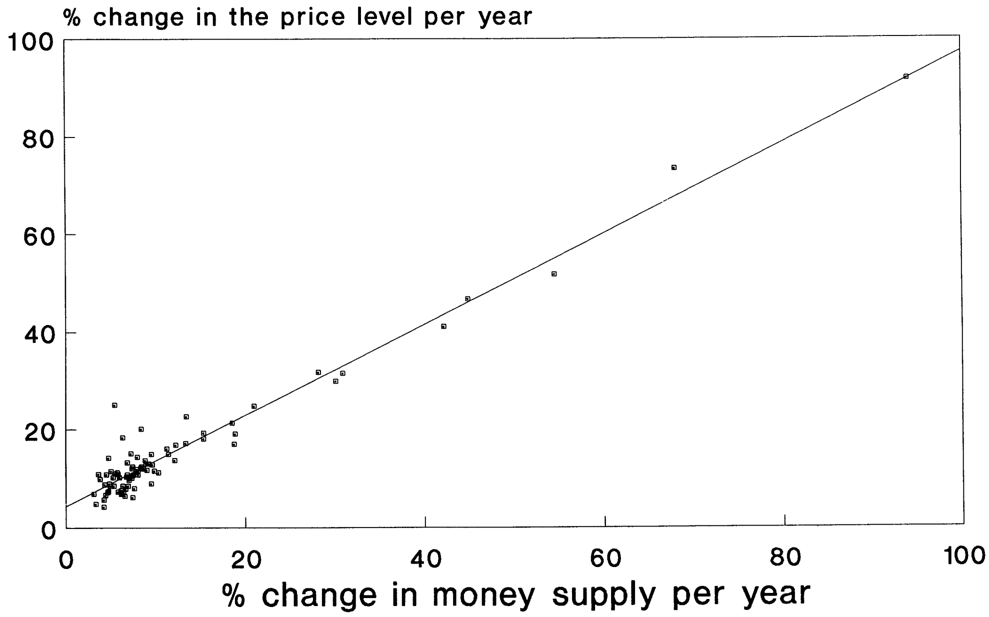

Figure 1. Inflation Rate Versus the Growth of Currency for 83 Countries.

of the error distribution. The test involves the regression of squared residuals on the regressors and their higher-order terms. We used the regressor and its second and the third powers in the test regression. The test statistic is calculated by the number of observations times the centered R^2 from the test regression. The calculated test statistic was 3.1456. Given that $\chi^2_{.05}$ is 7.815, the null hypothesis of homoscedasticity is retained.

The calculated symmetry test is equal to $-.8491$ and is reported in Table 5. The test statistic is distributed asymptotically normal. Given that the 5% critical value for the normal distribution is -1.645, the null hypothesis of symmetry is retained. Figure 2 plots the nonparametric kernel density estimator of the residual distribution, which reveals the symmetry of residual distribution. We also test whether the residual distribution is normal. If it is, then the maximum likelihood estimator is the OLS estimator. On the other hand, if the residual distribution is nonnormal, the OLS estimator may not be the most efficient one, and adaptive estimation procedures may be recommended to gain efficiency.

To test for normality of the residuals of a linear regression model, we use the $\sqrt{b_1}$ and b_2 statistics presented by White and MacDonald (1980). These two statistics are

$$\sqrt{b_1} = n^{-1} \sum_{i=1}^{n} e_i^3 \Big/ (s^2)^{3/2} \tag{10}$$

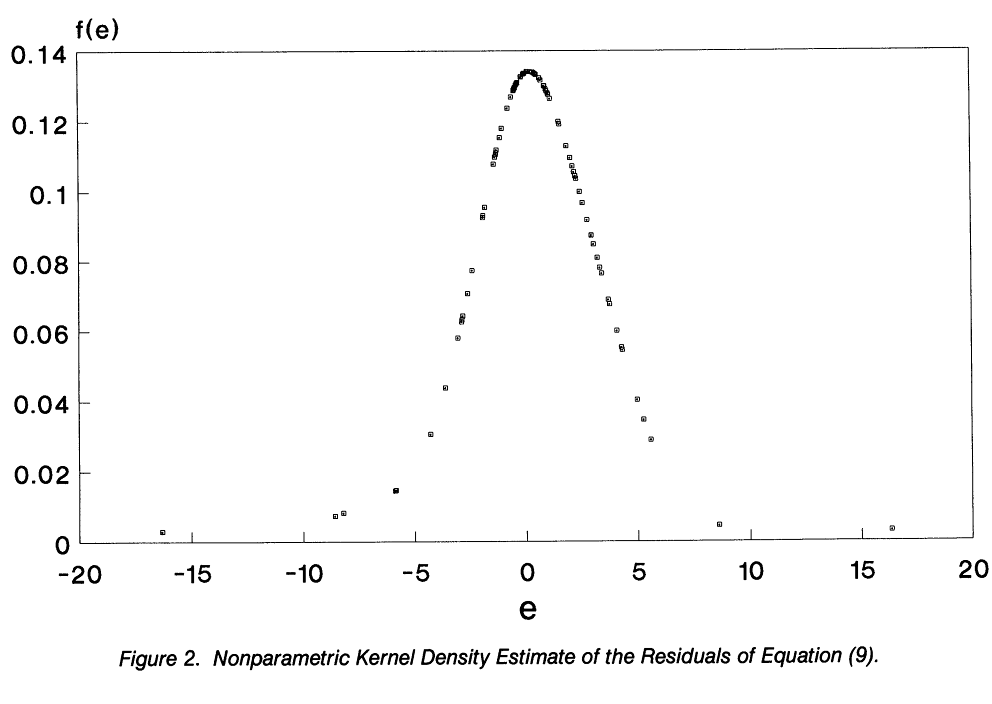

Figure 2. Nonparametric Kernel Density Estimate of the Residuals of Equation (9).

and

$$b_2 = n^{-1} \sum_{i=1}^{n} e_i^4 \Big/ (s^2)^2, \tag{11}$$

where

$$s^2 = n^{-1} \sum_{i=1}^{n} e_i^2 .$$

The calculated $\sqrt{b_1}$ and b_2 statistics are -1.8342 and 9.7559. The upper and lower 5% points (White and MacDonald 1980, p. 20) of $\sqrt{b_1}$ and b_2 statistics are ($-.534$, .534), (2.15, 3.99) for $n = 50$ and ($-.389$, .389), (2.35, 3.77) for $n = 100$. The normality of the residual distribution is strongly rejected at the 5% point for both $\sqrt{b_1}$ and b_2 statistics. Therefore, we may conclude that the OLS estimator is not the most efficient one, and a generalized method-of-moments estimator as that of Newey (1988) may be adopted to attain the maximum efficiency for the model (9).

6. CONCLUSIONS

Bickel (1982) has shown that if the conditional density function of the disturbance is symmetric around zero, then the regression coefficients, β, can be estimated adaptively. This implies that the maximum attainable efficiency for the regression coefficients is that of the maximum likelihood estimator that could be obtained if the actual (unknown) functional form of the disturbance distribution were used in forming the likelihood.

In this article we have proposed an asymptotically valid test for symmetric error distribution based on the nonparametric kernel estimation of the density function of the errors. The test is shown to be *consistent,* so it is able to detect any departure of the error distribution from symmetry for large samples. The test has good size and power properties in sample sizes as small as 50 observations and can be used to detect any departure from symmetry of the residual distribution of the linear regression models.

APPENDIX: PROOFS

This Appendix collects the proofs of the results stated in the article. Throughout, $a = a_n$, C denotes a generic constant, $\sum_i = \sum_{i=1}^n$, $\sum_{j \neq i} = \sum_i \sum_{j \neq i}$, and so on.

The following lemma is a restatement of lemma 2.3 of Schuster (1969). It will be used frequently in the proof of Lemma 3.1.

Lemma 7.1. Assume that the kernel K is chosen such that $\int |u| K(u)\, du$ is finite and such that $K^{(s)}$ is a continuous function of bounded variation for $s = 0, 1, \ldots, m$. If f and its first $(m + 1)$ derivatives are bounded, then for $s = 0, 1, \ldots, m$,

$$\lim_{n \to \infty} \int \frac{1}{a^{(s+1)}} K^{(s)}\left(\frac{x - y}{a}\right) f(y)\, dy = f^{(s)}(x), \quad \text{uniformly in } x.$$

Proof of Lemma 3.1

a. By lemma 2.4 of Fan and Gencay (1993), it is sufficient to show that $\sqrt{n}(\tilde{\delta}_\gamma^* - \hat{\delta}_\gamma^*) = o_p(1)$. Now, by Taylor series expansion,

$$\begin{aligned} \sqrt{n}(\tilde{\delta}_\gamma^* - \hat{\delta}_\gamma^*) &= \frac{1}{n^{3/2} a_n} \sum_i \sum_j \frac{n C_i(\gamma)}{n_\gamma} \left[K\left(\frac{e_i + e_j}{a_n}\right) - K\left(\frac{\varepsilon_i + \varepsilon_j}{a_n}\right) \right] \\ &= \frac{1}{n^{3/2} a} \sum_i \sum_j \frac{n C_i(\gamma)}{n_\gamma} \\ &\quad \times \left\{ \sum_{s=1}^{m} \frac{1}{s!} K^{(s)}\left(\frac{\varepsilon_i + \varepsilon_j}{a}\right) \left[\frac{(\mathbf{x}_i + \mathbf{x}_j)'(\mathbf{b} - \beta)}{a} \right]^s \right\} \\ &\quad + \frac{1}{n^{3/2} a} \sum_i \sum_j \frac{n C_i(\gamma)}{n_\gamma} \left\{ \frac{1}{(m+1)!} K^{(m+1)}(\psi_{ij}) \right. \\ &\quad \left. \times \left[\frac{(\mathbf{x}_i + \mathbf{x}_j)'(\mathbf{b} - \beta)}{a} \right]^{(m+1)} \right\}, \end{aligned} \tag{A.1}$$

where $\psi_{ij} = [\varepsilon_i + \varepsilon_j - \theta_{ij}(\mathbf{x}_i + \mathbf{x}_j)'(\mathbf{b} - \beta)]/a$ and $0 < \theta_{ij} < 1$.

Because $\mathbf{b} - \beta = O_p(n^{-1/2})$, (A.1) implies that it suffices to show that the following results hold:

$$\mathbf{I}_s \equiv \frac{1}{n^{(3+s)/2} a^{(1+s)}} \sum_i \sum_j \frac{n C_i(\gamma)}{n_\gamma} K^{(s)}\left(\frac{\varepsilon_i + \varepsilon_j}{a}\right) [(\mathbf{x}_i + \mathbf{x}_j)'\iota]^s = o_p(1) \tag{A.2}$$

and

$$\mathbf{II} \equiv \frac{1}{n^{(4+m)/2} a^{(2+m)}} \sum_i \sum_j \frac{n C_i(\gamma)}{n_\gamma} K^{(m+1)}(\psi_{ij}) [(\mathbf{x}_i + \mathbf{x}_j)'\iota]^{(m+1)} = o_p(1), \tag{A.3}$$

where $s = 1, 2, \ldots, m$ and ι is a $d \times 1$ vector of 1s.

Consider $\mathbf{I}_1$ first; it equals

$$\frac{1}{n^2} \sum_i \left[\frac{2n C_i(\gamma)}{n_\gamma a^2} \right] K'\left(\frac{2\varepsilon_i}{a}\right) [\mathbf{x}_i'\iota] + \frac{1}{n^2} \sum_i \sum_{j \neq i} \frac{n C_i(\gamma)}{n_\gamma a^2} K'\left(\frac{\varepsilon_i + \varepsilon_j}{a}\right) [(\mathbf{x}_i + \mathbf{x}_j)'\iota] \equiv \mathbf{I}_1 F + \mathbf{I}_1 S.$$

Under the stated assumptions (in particular, K3 and A), it is easy to see that $\mathbf{I}_1 F = o_p(1)$, because $E[|\mathbf{I}_1 F|] \leq C(n_\gamma a^2)^{-1} = o(1)$. With respect to $\mathbf{I}_1 S$, we have $E(\mathbf{I}_1 S) = 0$ under H_0, because $K'(\cdot)$ is antisymmetric and $f(\cdot)$ is symmetric under H_0. Thus we only need to show that $\text{var}(\mathbf{I}_1 S) = o(1)$. This is true under the stated assumptions and H_0, as shown next:

$$\begin{aligned} \text{var}(\mathbf{I}_1 S) &= \frac{1}{n^4} \sum_{i \neq j} \sum_{k \neq l} \left[\frac{n^2 C_i(\gamma) C_k(\gamma)}{n_\gamma^2 a^4} \right] [(\mathbf{x}_i + \mathbf{x}_j)'\iota][(\mathbf{x}_k + \mathbf{x}_l)'\iota] \\ &\quad \times E\left[K'\left(\frac{\varepsilon_i + \varepsilon_j}{a}\right) K'\left(\frac{\varepsilon_k + \varepsilon_l}{a}\right) \right] \\ &\leq \frac{C}{n^4 a^4} \sum_{i \neq j} \sum_{k \neq l} \left| E\left[K'\left(\frac{\varepsilon_i + \varepsilon_j}{a}\right) K'\left(\frac{\varepsilon_k + \varepsilon_l}{a}\right) \right] \right| \\ &= \frac{C}{n^4 a^4} \sum_{i \neq j} E\left[\left\{ K'\left(\frac{\varepsilon_i + \varepsilon_j}{a}\right) \right\}^2 \right] \\ &\quad + \frac{C}{n^4 a^4} \sum_{i \neq j \neq k} \left| E\left[K'\left(\frac{\varepsilon_i + \varepsilon_j}{a}\right) K'\left(\frac{\varepsilon_k + \varepsilon_i}{a}\right) \right] \right|, \end{aligned} \tag{A.4}$$

where the last equality is obtained from the fact that under H_0, $E[K'((\varepsilon_i + \varepsilon_j)/a)] = 0$. The first term on the right side of (A.4) equals $O((na^2)^{-2}) = o(1)$ by boundedness of K' and (A). The second term equals in order

$$\frac{1}{na^2}\left|\iiint\left[\frac{1}{a}K'\left(\frac{s+t}{a}\right)\right]\left[\frac{1}{a}K'\left(\frac{\tau+s}{a}\right)\right]f(s)f(t)f(\tau)\,ds\,dt\,d\tau\right|$$
$$=\frac{1}{n}\int\left\{\int\left[\frac{1}{a^2}K'\left(\frac{s+t}{a}\right)\right]f(t)\,dt\right\}^2 f(s)\,ds$$
$$\equiv\frac{1}{n}\int[g_n(s)]^2 f(s)\,ds,$$

where $g_n(s) \equiv \int a^{-2}K'((s+t)/a)f(t)\,dt$. Note that K2 implies $\int |u|K(u)\,du \le [\int u^2K(u)\,du]^{1/2} < \infty$. Thus, by Lemma 7.1, $\lim_{n\to\infty} g_n(s) = f'(-s)$ uniformly in s. By the Lebesgue dominated convergence, we have $\lim_{n\to\infty}[\int [g_n(s)]^2f(s)\,ds] = \int [f'(-s)]^2f(s)\,ds \le C\int f(s)\,ds < \infty$. Therefore, we have shown that (A.2) holds for $s = 1$.

Now consider $\mathbf{I}_2$; it equals

$$\mathbf{I}_2 = \frac{1}{n^{5/2}a^3}\sum_i\left[\frac{nC_i(\gamma)}{n_\gamma}\right]K''\left(\frac{2\varepsilon_i}{a}\right)[2\mathbf{x}_i'\iota]^2$$
$$+\frac{1}{n^{5/2}a^3}\sum_i\sum_{j\ne i}\frac{nC_i(\gamma)}{n_\gamma}K''\left(\frac{\varepsilon_i+\varepsilon_j}{a}\right)[(\mathbf{x}_i+\mathbf{x}_j)'\iota]^2$$
$$\equiv \mathbf{I}_2F + \mathbf{I}_2S. \qquad \text{(A.5)}$$

Because $E[|\mathbf{I}_2F|] \le C(n^{3/2}a^3)^{-1} = o(1)$ by boundedness of K'' and A, we get $\mathbf{I}_2F = o_p(1)$. For the second term, we have

$$|E[\mathbf{I}_2S]| \le \frac{C}{n^{1/2}a^3}\left|E\left[K''\left(\frac{\varepsilon_1+\varepsilon_2}{a}\right)\right]\right|$$
$$=\frac{C}{n^{1/2}}\left|\int\left[\int\frac{1}{a^3}K''\left(\frac{s+t}{a}\right)f(t)\,dt\right]f(s)\,ds\right|$$
$$=\frac{C}{n^{1/2}}\left|\int f''(-s)f(s)\,ds + o(1)\right| = o(1), \qquad \text{(A.6)}$$

where we have applied Lemma 7.1 and the Lebesgue-dominated convergence to (A.6). We now show that $\text{var}[\mathbf{I}_2S] = o(1)$. It follows from (A.5) that

$$\text{var}[\mathbf{I}_2S] \le \frac{C}{n^5a^6}\sum_i\sum_{j\ne i}\sum_k\sum_{l\ne k}E\left[K''\left(\frac{\varepsilon_i+\varepsilon_j}{a}\right)K''\left(\frac{\varepsilon_k+\varepsilon_l}{a}\right)\right]$$
$$=\frac{C}{n^5a^6}\sum_i\sum_{j\ne i}E\left[\left\{K''\left(\frac{\varepsilon_i+\varepsilon_j}{a}\right)\right\}^2\right]$$
$$+\frac{C}{n^5a^6}\sum_i\sum_{j\ne i}\sum_{k\ne j\ne i}E\left[K''\left(\frac{\varepsilon_i+\varepsilon_j}{a}\right)K''\left(\frac{\varepsilon_k+\varepsilon_i}{a}\right)\right]$$
$$+\frac{C}{n^5a^6}\sum_i\sum_{j\ne i}\sum_{k\ne j}\sum_{l\ne k}\left\{E\left[K''\left(\frac{\varepsilon_i+\varepsilon_j}{a}\right)\right]\right\}^2$$
$$\equiv C(A_1+A_2+A_3). \qquad \text{(A.7)}$$

It is easy to see that $A_1 = O((na^2)^{-3}) = o(1)$. Applying Lemma 7.1 and the Lebesgue dominated convergence to A_2, we have in order $A_2 = n^{-2}\int[\int a^{-3}K''((s+t)/a)f(s)\,ds]^2f(t)\,dt = O(n^{-2})$. Similarly, in order $A_3 = n^{-1}[\int\{\int a^{-3}K''((s+t)/a)f(s)\,ds\}f(t)\,dt]^2 = O(n^{-1})$. Therefore, $\mathbf{I}_2S = o_p(1)$ and $\mathbf{I}_2 = o_p(1)$ from (A.5).

Using similar arguments, one can show that $\mathbf{I}_s = o_p(1)$ for $s = 3, \ldots, m$. It remains to verify (A.3). For this, we note that $E[|II|] \le C(n^{m/2}a^{(2+m)})^{-1} = C[na^{(2+4/m)}]^{-m/2} = o(1)$ by boundedness of $K^{(m+1)}$ and assumption A. Hence, the proof of Part a is complete.

b. The proof is accomplished by using the same arguments as in a and by invoking Lemma 2.1 of Fan and Gencay (1993).

Proof of Proposition 3.2

Note that

$$\sqrt{n}(\tilde\lambda_\gamma^* - \lambda^*) = \frac{\sqrt{n}(\tilde\delta_\gamma^* - \delta^*)}{\tilde\Delta(f)} + \frac{\sqrt{n}(\Delta(f)-\tilde\Delta(f))\lambda^*}{\tilde\Delta(f)}.$$

The result then follows from Lemma 3.1 and the Liapounov Central Limit Theorem.

[Received April 1993. Revised June 1994.]

REFERENCES

Ahmad, I. A. (1980), "Nonparametric Estimation of an Affinity Measure Between Two Absolutely Continuous Distributions with Hypotheses Testing Applications," *Annals of the Institute of Statistics in Mathematics,* 32, Part A, 223–240.

Ahmad, I. A., and Van Belle, G. (1974), "Measuring Affinity of Distributions," in *Reliability and Biometry, Statistical Analysis of Life Testing,* eds. Proschan and Serfling, Philadelphia: SIAM, pp. 651–668.

Bickel, P. (1982), "On Adaptive Estimation," *The Annals of Statistics,* 10, 647–671.

Cancer, R. (1981), "A Note on Studentizing a Test for Heteroscedasticity," *Journal of Econometrics,* 17, 107–112.

Eubank, R. L., Hart, J. D., and LaRiccia, V. N. (1993), "Testing Goodness of Fit via Nonparametric Function Estimation Techniques," *Communications in Statistics, Part A—Theory and Methods,* 22, 3327–3354.

Fan, Y., and Gencay, R. (1993), "Hypotheses Testing Based on Modified Nonparametric Estimation of an Affinity Measure Between Two Distributions," *Journal of Nonparametric Statistics,* 2, 389–403.

Härdle, W. (1990), *Applied Nonparametric Regression,* Econometric Society Monographs No. 19, Cambridge, U.K.: Cambridge University Press.

International Mathematical and Statistical Libraries (1987), *The IMSL Library,* Houston: Author.

Marron, J. S. (1988), "Automatic Smoothing Parameter Selection: A Survey," *Empirical Economics,* 13, 187–208.

Newey, W. K. (1988), "Adaptive Estimation of Regression Models via Moment Restrictions," *Journal of Econometrics,* 38, 301–339.

Robinson, P. M. (1991), "Consistent Nonparametric Entropy-Based Testing," *The Review of Economic Studies,* 58, 437–453.

Schuster, E. F. (1969), "Estimation of a Probability Density Function and Its Derivatives," *The Annals of Mathematical Statistics,* 40, 1187–1195.

Silverman, B. W. (1986), *Density Estimation in Statistics and Data Analysis,* London: Chapman and Hall.

White, H., and MacDonald, G. M. (1980), "Some Large-Sample Tests for Nonnormality in the Linear Regression Model," *Journal of the American Statistical Association,* 75, 16–31.