THE INFLUENCE CURVE AND GOODNESS OF FIT

by

John R. Michael
William R. Schucany

Technical Report No. 137
Department of Statistics ONR Contract

May 9, 1980

Research Sponsored by the Office of Naval Research
Contract N00014-75-C-0439
Project NR 042-280

DIVISION OF MATHEMATICAL SCIENCES
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

THE INFLUENCE CURVE AND GOODNESS OF FIT

John R. Michael
Bell Laboratories
Holmdel, New Jersey 07733

William R. Schucany
Southern Methodist University
Dallas, Texas 75275

SUMMARY

The influence curve introduced by Hampel (1968) is applied to goodness-of-fit statistics. The efficacy curve is then defined to be the square of influence curve weighted by a constant which arises in the context of approximate Bahadur efficiency. For a number of goodness-of-fit statistics the ratios of these curves are shown to be equal to the asymptotic relative efficiency in the Pitman sense when testing for point contamination. These efficacy curves graphically demonstrate the sensitivities of certain goodness-of-fit statistics to minor perturbations in the assumed distribution.

Some key words: Bahadur efficiency; EDF statistics; Pitman efficiency; sensitivity curves.

Partially supported under ONR Contract N00014-75-C-0439.

1. THE INFLUENCE CURVE

The <u>influence curve</u> (IC) is introduced by Hampel (1968) as a tool in the study of the robustness of certain estimators of a location parameter, say θ. The basic idea is to first perturb the assumed distribution by mixing it with another distribution degenerate at the point $(\theta+c)$. This model then quantifies point contamination and, more importantly, approximates in a mathematically tractable fashion a minor irregularity in the assumed distribution or contamination by another distribution which has a relatively small variance. The IC then measures the asymptotic rate of change of the estimator as contamination is introduced, that is, it measures the influence on the estimator by a relatively small amount of contamination.

Let F be a cumulative distribution function (CDF) and let F_n denote the familiar empirical distribution function (EDF) which, for any argument x, is defined to be the proportion of a random sample that is less than or equal to x. Many statistics can be written as functionals of F_n, say $T(F_n)$. If we are sampling from F and we view $T(F_n)$ as an estimator for the parameter θ, then $T(F_n)$ is said to be Fisher consistent if $\theta = T(F)$. Since $F_n \to F$ uniformly with probability one then in many cases, though not all, $T(F_n) \to T(F)$ in probability.

Now let δ_c be the CDF which is degenerate at the point c and denote the mixture of F and δ_c by $F_\varepsilon = (1-\varepsilon)F + \varepsilon\,\delta_c$, where ε is the mixing proportion. The influence curve of T at F is defined pointwise by

$$IC_{T,F}(c) = \lim_{\varepsilon \downarrow 0} \left\{ \frac{T(F_\varepsilon) - T(F)}{\varepsilon} \right\},$$

if this limit is defined for every point c. We will omit the subscript F and write $IC_T(c)$ since the form of F will be understood from the context. In most cases $IC_T(c) = \lim_{\varepsilon \downarrow 0} \{T'(F_\varepsilon)\}$, where $T'(F_\varepsilon)$ is the first derivative of $T(F_\varepsilon)$ with respect to ε.

Hampel (1974) suggests that the IC for an estimator be sketched, examined, and considered along with other qualitative information about the estimator such as the form and variance of its asymptotic distribution. In the setting of robust estimation, it is undesirable for a small amount of contamination to have a large effect on the value of the estimator. In terms of the IC, large absolute values are undesirable. If goodness-of-fit statistics are considered, the interpretation of the IC must be reversed.

2. INFLUENCE CURVES FOR EDF STATISTICS

It is a natural step to construct ICs for goodness-of-fit statistics in hopes of shedding some light on the sensitivities of these statistics to perturbations at different points of the assumed distribution. In this setting it is desirable for a small amount of contamination to have a large effect on the value of the statistic. In terms of the IC, large values are desirable.

In this section we will consider the case of a simple null hypothesis, that is, the hypothesized distribution F is completely specified and contains no unknown parameters. If F is continuous, we can employ the probability integral transformation on the sample data and equivalently test the hypothesis of uniformity on the unit interval.

The so-called EDF statistics discussed by Stephens (1974) are formulated naturally as functionals of F_n. The five EDF statistics that will be considered here are:

1. The Kolmogorov-Smirnov statistic D:

$$D = \sup_x |F_n(x) - F(x)|.$$

2. The Kuiper statistic V:

$$V = \sup_x \{F_n(x) - F(x)\} + \sup_x \{F(x) - F_n(x)\}.$$

3. The Cramér-von Mises statistic W^2:

$$W^2 = n \int \{F_n(x) - F(x)\}^2 \, dF(x).$$

4. The Watson statistic U^2:

$$U^2 = n \int [F_n(x) - F(x) - \int \{F_n(s) - F(s)\} dF(s)]^2 \, dF(x).$$

5. The Anderson-Darling statistic A^2:

$$A^2 = n \int \frac{\{F_n(x) - F(x)\}^2}{F(x)\{1 - F(x)\}} \, dF(x).$$

The ICs for D and V can be derived in a straightforward manner. But W^2, U^2, and A^2 all have nondegenerate limiting distributions under the null hypothesis. In order to obtain ICs we are led to define the statistics $W^* = (W^2/n)^{\frac{1}{2}}$, $U^* = (U^2/n)^{\frac{1}{2}}$, and $A^* = (A^2/n)^{\frac{1}{2}}$ without dividing by n the original statistics are not functionals of F_n; without the square root the ICs are identically zero and the functionals are not standard sequences as defined by Bahadur (1960). The relevance of this latter point will become clear in the next section.

For all five of these EDF statistics $T(F) = 0$. Replacing F_n with F_ε in their definitions the ICs are found to be:

1. $IC_D(c) = \left|F(c) - \frac{1}{2}\right| + \frac{1}{2}.$

2. $IC_V(c) = 1.$

3. $IC_{W*}(c) = \{F^2(c) - F(c) + \frac{1}{3}\}^{\frac{1}{2}}.$

4. $IC_{U*}(c) = (\frac{1}{12})^{\frac{1}{2}}.$

5. $IC_{A*}(c) = (-\ \ell n[F(c)\{1 - F(c)\}] - 1)^{\frac{1}{2}}.$

Discussion of the shapes of these ICs will be postponed until Section 4; however, we will note here that the relative magnitudes of these ICs have little meaning. This becomes obvious when we consider what happens if we define, say, $D^* = 2 \cdot D$. The test based on D^* is equivalent to the test based on D, but $IC_{D*}(c) = 2 \cdot IC_D(c)$. This difficulty is not encountered for estimators when each is constrained to be consistent. Another complicating factor when making comparisons is the fact that different goodness-of-fit statistics often have asymptotic distributions of different forms. Both of these difficulties can be resolved by relating the ratios of ICs to Pitman efficiency.

3. RELATIONSHIP TO MEASURES OF EFFICIENCY

In this section it is demonstrated that the ratio of ICs can have a relationship to the Pitman efficiency (PE) of the corresponding statistics. See Kendall and Stuart (1979) for a general discussion of PE. If the asymptotic distributions of the two statistics to be compared are of different forms, then it is still possible to obtain the PE as the limit of the approximate Bahadur efficiency (ABE) which is introduced by Bahadur (1960). Emphasis will not be placed on the regularity conditions

under which the IC can be related to PE, rather the validity of this relationship will be argued on a case by case basis.

Both PE and ABE are asymptotic measures and are defined in the context of testing the hypothesis H_o: $\Theta = \Theta_o$ for some $\Theta_o \in \Omega_o$ against the alternative H_A: $\Theta \in \Omega_A$. We can fit the contamination alternative into this framework by assuming that $F_\varepsilon = (1-\varepsilon)F + \varepsilon\delta_c$ is the CDF of the underlying distribution, letting $\Theta = \varepsilon$, $\Omega_o = \{\Theta_o\} = \{0\}$, $\Omega_A = (0,1)$, and testing the null hypothesis H_o: $\varepsilon = 0$ against the alternative hypothesis H_A: $0 < \varepsilon < 1$. Here H_o corresponds to the hypothesis that $F_\varepsilon = F$ where F is some completely specified CDF.

In many cases the asymptotic distributions of two goodness-of-fit statistics to be compared are of different forms. Bahadur (1960) proposes a measure of efficiency which takes this difference into account. Bahadur defines $\{\sqrt{n}\, T_n\}$ to be a <u>standard sequence</u> if the following three conditions are met:

Condition I: There exists a continuous CDF G^* such that for each $\Theta \in \Omega_o$

$$\lim_{n\to\infty} P_\Theta(\sqrt{n}\, T_n < x) = G^*(x), \text{ for every } x.$$

Condition II: There exists a constant $a \in (0,\infty)$ such that, as $x \to \infty$,

$$\ln\,[1-G^*(x)] = -\frac{ax^2}{2}\,[1+o(1)].$$

Condition III: There exists a function b on Ω_A with $0 < b(\Theta) < \infty$ such that for each $\Theta \in \Omega_A$, $T_n \to b(\Theta)$, in probability.

The <u>approximate slope</u> of $\{\sqrt{n}\ T_n\}$ is defined to be $s(\Theta) = ab^2(\Theta)$ and the <u>approximate Bahadur efficiency</u> of $T_{1,n}$ relative to $T_{2,n}$ is defined to be

$$ABE_{12}(\Theta) = \frac{s_1(\Theta)}{s_2(\Theta)} = \frac{a_1 b_1^2(\Theta)}{a_2 b_2^2(\Theta)} ,$$

where the subscript convention should be obvious.

Whereas PE is a measure of local efficiency, i.e., as $\Theta \to \Theta_o$, ABE is more general in that it depends upon a particular non-null value of Θ. While PE yields a single number, ABE can vary over different values of $\Theta \in \Omega_A$. For this reason we will occasionally write $ABE(\Theta)$. Although more general, ABE has an approximate sample size interpretation while PE has an exact sample size interpretation.

It is reassuring that in most cases $\lim_{\Theta \to \Theta_o} ABE(\Theta) = PE$. Bahadur (1960) discusses the case in which both $\sqrt{n}\ T_{1,n}$ and $\sqrt{n}\ T_{2,n}$ have limiting normal distributions. Wieand (1976) does not place this restriction on the statistics but requires that each satisfy certain other conditions. The major condition, which Wieand terms Condition III*, is stronger than Bahadur's Condition III. Denote by B the interval $(\Theta_o, \Theta_o + \Theta^*)$ for one-sided alternatives and the interval $(\Theta_o - \Theta^*, \Theta_o + \Theta^*)$ with Θ_o deleted in the two-sided case. Wieand defines Condition III* as follows:

Condition III*: For the standard sequence $\{\sqrt{n}\ T_n\}$ there is a $\Theta^* > 0$ such that for every $\varepsilon > 0$ and $\delta \in (0,1)$ there is a γ such that for all $\Theta \in B$ and $n > [\gamma/b^2(\Theta)]$ we have $P_\Theta \{|T_n - b(\Theta)| < \varepsilon\, b(\Theta)\} > 1-\delta$.

Wieand's main theorem states that when two statistics satisfy Condition III*, and certain other minor conditions, then $\lim_{\theta \to \theta_o} ABE(\theta) = PE$.

Using this result, PEs can be obtained where never before possible. In particular Wieand shows that the goodness-of-fit statistics D, V, W*, and U* all satisfy his theorem when the underlying distribution is continuous. Since F_ε is discontinuous, we must argue that Wieand's theorem can be extended to the contamination alternative. When F is continuous the verification of Condition III* for each EDF statistic makes use of the fact that the CDF of $Z = \sqrt{n}\ \mathrm{Sup}\ |F_n - F|$, say $K_n(z)$, does not depend on F. When F is discontinuous this is no longer true and so for a particular F the CDF of Z, say $K_n'(z)$, depends on F. It is known, however, that $K_n'(z) \geq K_n(z)$ for every z (e.g. see Darling, 1957). This implies that F_n converges to F faster, and hence that T_n converges to $b(\theta)$ faster, when F is discontinuous. Condition III* requires only that T_n converge to $b(\theta)$ within a specified rate. Since D, V, W*, and U* all satisfy Condition III* when F is continuous, they must also satisfy Condition III* when F is discontinuous.

The statistic A* can also be included. The characteristic function for A^2 is derived by Anderson and Darling (1952). It then follows from Theorem 5.3 of Abrahamson (1965) that the constant "a" in Bahadur's Condition II is equal to 2 for A*. The verification of Condition III* for A* parallels that for W* in Wieand (1976).

With the contamination alternative an entire family of alternative distributions, parameterized by ε, is defined for each value of c. The PEs obtained below will depend in general on the path taken as $F_\varepsilon \to F$,

that is, on the particular value of c. To make clear this dependence on c we will write PE(c).

Using the relations $b(\varepsilon)$ and $T(F_\varepsilon)$ and $T(F) = 0$ which hold for the EDF statistics being considered, then for any pair of these statistics PEs can be obtained from ABEs as

$$PE(c) = \lim_{\varepsilon \downarrow 0} ABE(\varepsilon) = \frac{a_1}{a_2} \lim_{\varepsilon \downarrow 0} \frac{T_1^2(F_\varepsilon)}{T_2^2(F_\varepsilon)}$$

$$= \frac{a_1 \, IC_1^2(c)}{a_2 \, IC_2^2(c)} .$$

Thus we see that the ratios of the squares of ICs, when properly weighted, can have a PE interpretation. More specifically, PE here is the limit as $n \to \infty$ and $\varepsilon \to 0$ of the ratio of sample sizes required for two statistics to achieve the same power when testing the hypothesis of no contamination at the same significance level against the contamination alternative. For the statistics to be considered here, PEs are independent of the particular choices of significance level and power.

4. EFFICACY CURVES FOR EDF STATISTICS

We now define the efficacy curve, EC, as $EC = aIC^2$ where "a" is the constant in Condition II. The difficulties noted when comparing ICs for test statistics have now been overcome. The EC is invariant with respect to linear functions of a test statistic whereas the IC is not. Also through "a" the EC takes into account the fact that different statistics may have limiting distributions of different forms when the null hypothesis is true. For D, V, W*, U*, and A* the constant "a" has the value 4, 4, π^2, $4\pi^2$, and 2, respectively. The ECs for these statistics are:

1. $EC_D\ (c) = \{|2F(c) - 1| + 1\}^2$.

2. $EC_V\ (c) = 4$.

3. $EC_{W*}\ (c) = \pi^2\{F^2(c) - F(c) + \frac{1}{3}\}$.

4. $EC_{U*}\ (c) = \frac{\pi^2}{3}$.

5. $EC_{A*}\ (c) = -\ 2\ \ell n[F(c)\{1 - F(c)\}] - 2$.

Graphs of these curves are shown in Figure 1. Their shapes reveal in rather dramatic fashion the asymptotic sensitivity of each statistic to departures from the assumed distribution at the point F(c) as it varies in the interval. The ECs for V and U* are seen to be constants. This is not surprising considering that both statistics were orginally proposed as tests for uniformity on the circle and share the property that they are invariant with respect to the choice of origin. The EC for A* is unbounded and illustrates the well-known sensitivity of A^2 to perturbations in the tails of the assumed distribution. The statistics D and W* can also be seen to be more sensitive towards the tails of the distribution.

There are some surprising results when the relative heights of these ECs are examined. For example V dominates all other statistics except A*, and D and U* both dominate W*. To see just how accurate this representation is for finite samples, m = 10,000 samples were generated independently for samples of size n = 20, 40, and 80 with 11 choices for the point of contamination F(c) = .00(.05).50. The uniform distribution was then contaminated with probability ε = .10. The

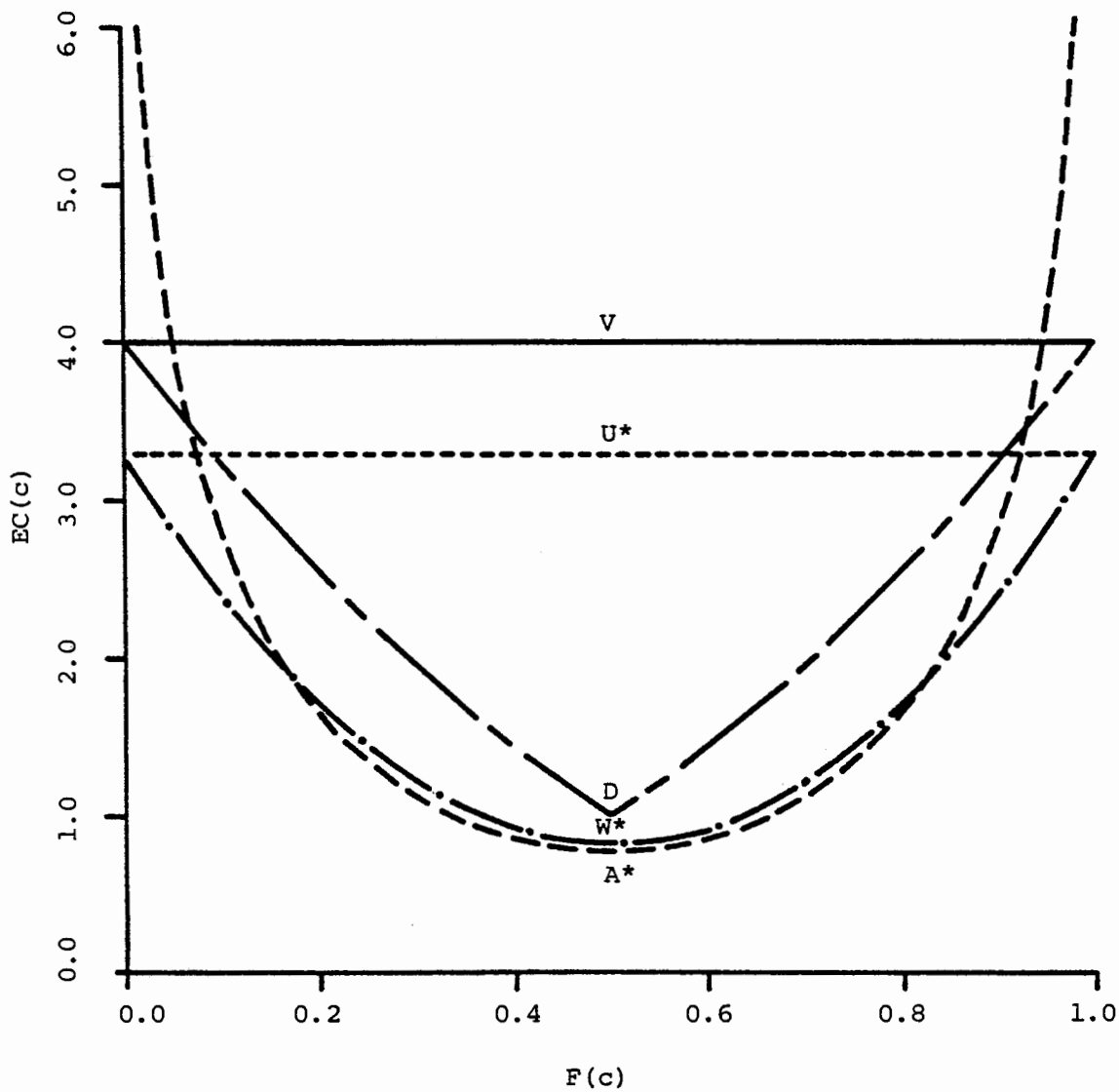

Figure 1. Efficacy curves for EDF statistics.

empirical powers were determined based on a .05-level test and an <u>empirical power curve</u> (EPC) was plotted for each statistic. Values of $c > .5$ were plotted using symmetry considerations. The EPCs for $n = 20$, 40, and 80 were computed; that for $n = 40$ is shown in Figure 2.

The overall appearances of the EPCs in Figure 2 are remarkably similiar to the ECs in Figure 1, both in individual shape and in relative position to one another. And it should be recalled that while relative heights of EPCs have a relative power interpretation, those for ECs have a relative sample size interpretation. Since there is a monotonic relationship between power and efficiency, we still expect the order of finish to be about the same for large samples. This is seen to be true in the center of the distribution, but not strictly true in the tails. When averaged over the 11 values of c, the EPC for V is significantly greater than that for U for all three sample sizes investigated and the observed significance level, $\hat{p}$, decreases as n increases ($\hat{p} \doteq .04, .001, < .0005$). This is in accordance with a PE of 1.216. One is struck by the fact that in no other case does one appear to dominate another, and that the order of finish in the tails is the reverse of that in the center of the distribution.

While the relative heights of ECs have some meaning, caution should be exercised when extrapolating to finite samples. The discrepancies noted between ECs and EPCs reflect the limitation of an asymptotic measure when attempting to describe small sample behavior. Regardless of this limitation, the notion of the EC provides an excellent method for determining meaningful scale factors when presenting and comparing ICs for test statistics. The visual inspection of ECs should be useful both

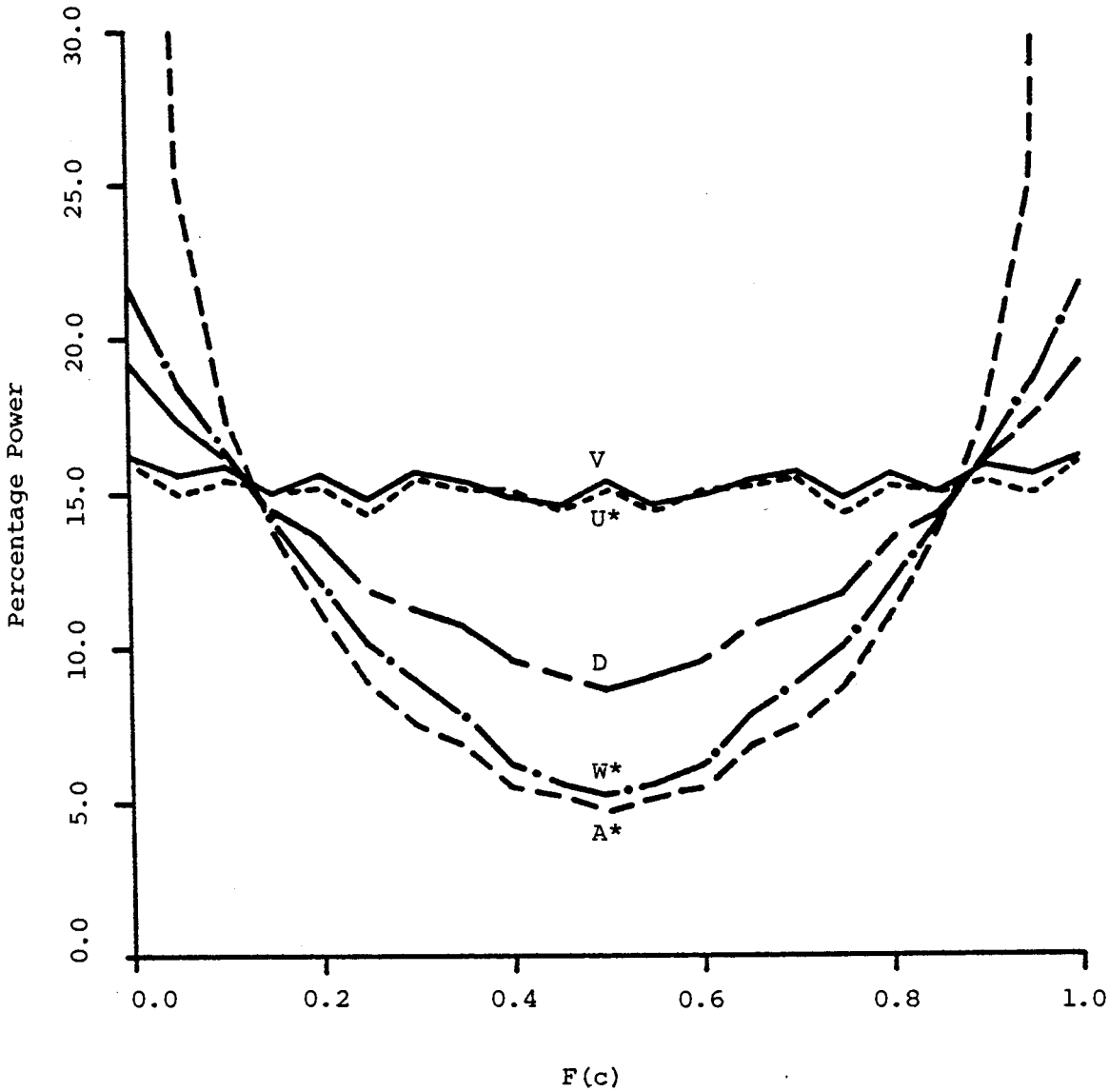

Figure 2. Empirical Power Curves for EDF statistics for the contamination alternative: n=40; m=10,000; ε=.10; level=.05.

in identifying statistics that are similar in performance against a variety of continuous altervatives and in predicting what this performance will be.

5. EXTENSIONS TO COMPOSITE HYPOTHESES

The above development can be extended to the case of a composite null hypothesis. Such a case commonly arises when distribution parameters are not specified in advance. Many statistics have been modified for testing a composite hypothesis by replacing nuisance parameters by efficient estimates. In other cases, goodness-of-fit statistics have been formulated specifically for testing certain families of distributions.

The null distributions of EDF statistics modified for testing composite hypotheses are complicated. Stephens (1976) gives some results for W^2, U^2, and A^2 when testing for normality and exponentiality. The constant "a" in Condition II has not been determined for any of these statistics and verification of Condition III* appears difficult. For the case of normality, Michael (1977) studies EDF statistics, the standardized third sample moment $\sqrt{b_1}$, the standardized fourth sample moment b_2, and statistics proposed by D'Agostino (1971) and van der Watt (1969). Although ICs are derived for W, U, $\sqrt{b_1}$, and b_2, only those for $\sqrt{b_1}$ and b_2 can as yet be related to PE. Details for $\sqrt{b_1}$ and b_2 are now given.

Denoting the sample values by $x_1, x_2, \ldots, x_n$, the standardized third and fourth sample moments can be written as

$$\sqrt{b_1} = \frac{\sum (x_i - \bar{x})^3}{ns^3} = \frac{\int (x - \bar{x})^3 \, dF_n}{s^3}$$

and

$$b_2 = \frac{\sum (x_i - \bar{x})^4}{ns^4} = \frac{\int (x - \bar{x})^4 \, dF_n}{s^4}$$

where

$$\bar{x} = \frac{\sum x_i}{n} = \int x dF_n(x)$$

and

$$s^2 = \frac{\sum (x - \bar{x})^2}{n} = \int (x - \bar{x})^2 \, dF_n(x) .$$

Without loss of generality we will let F represent the standard normal CDF contaminated at the point c. Replacing F_n with F in the above expressions the ICs are found to be:

$$IC_{\sqrt{b_1}}(c) = c^3 - 3c$$

and

$$IC_{b_2}(c) = c^4 - 6c^2 + 3 .$$

If the parent distribution is the standard normal, then it is well known that $\sqrt{nb_1}$ and $\sqrt{n}b_2$ have asymptotic normal distributions with variances 6 and 24 respectively. For asymptotically normal test statistics the constant "a" in Condition II is simply the reciprocal of the variance (Bahadur , 1960). Hence the ECs for $\sqrt{b_1}$ and b_2 are given by

$$EC_{\sqrt{b_1}}(c) = (c^3 - 3c)^2/6$$

and

$$EC_{b_2}(c) = (c^4 - 6c^2 + 3)^2/24 .$$

We can show that the ratio of these two ECs has a PE interpretation without having to verify Condition III*. Since the test statistics to be compared are both asymptotically normal, we can take the standard approach described in Chapter 25 of Kendall & Stuart (1979). The only additional requirement is that the statistics satisfy certain mild regularity conditions.

It is interesting to note that the derivative of the IC for b_2 is proportional to the IC for $\sqrt{b_1}$. Thus at the two points where the EC for b_2 attains local maxima, the EC for $\sqrt{b_1}$ achieves local minima. This observation supports the notion that the two statistics are sensitive to quite different types of departures from normality. This characteristic is rather dramatically illustrated by graphs of these ECs shown in Figure 3.

6. COMMENTS

A. Many statistics cannot be expressed as functionals of F_n. Yet it still may be possible to determine PEs for these statistics when testing for the contamination alternative. Thus it is unnecessarily restrictive to tie the definition of the EC to that of the IC. A more practical definition is $EC = a\{b'(0)\}^2$ where $b'(0) = \lim_{\varepsilon \downarrow 0} [\frac{d}{d\varepsilon}\{b(\varepsilon)\}]$.

B. One of the major goals of Hampel (1968) was to find that estimator in a certain class for which the <u>supremum</u> of the IC is <u>minimized</u>. A similar approach may be fruitful for goodness-of-fit statistics using the EC. Of course the optimality criteria would have to be quite different. We could find that statistic within a suitable class for which the <u>infimum</u> of the EC is <u>maximized</u>.

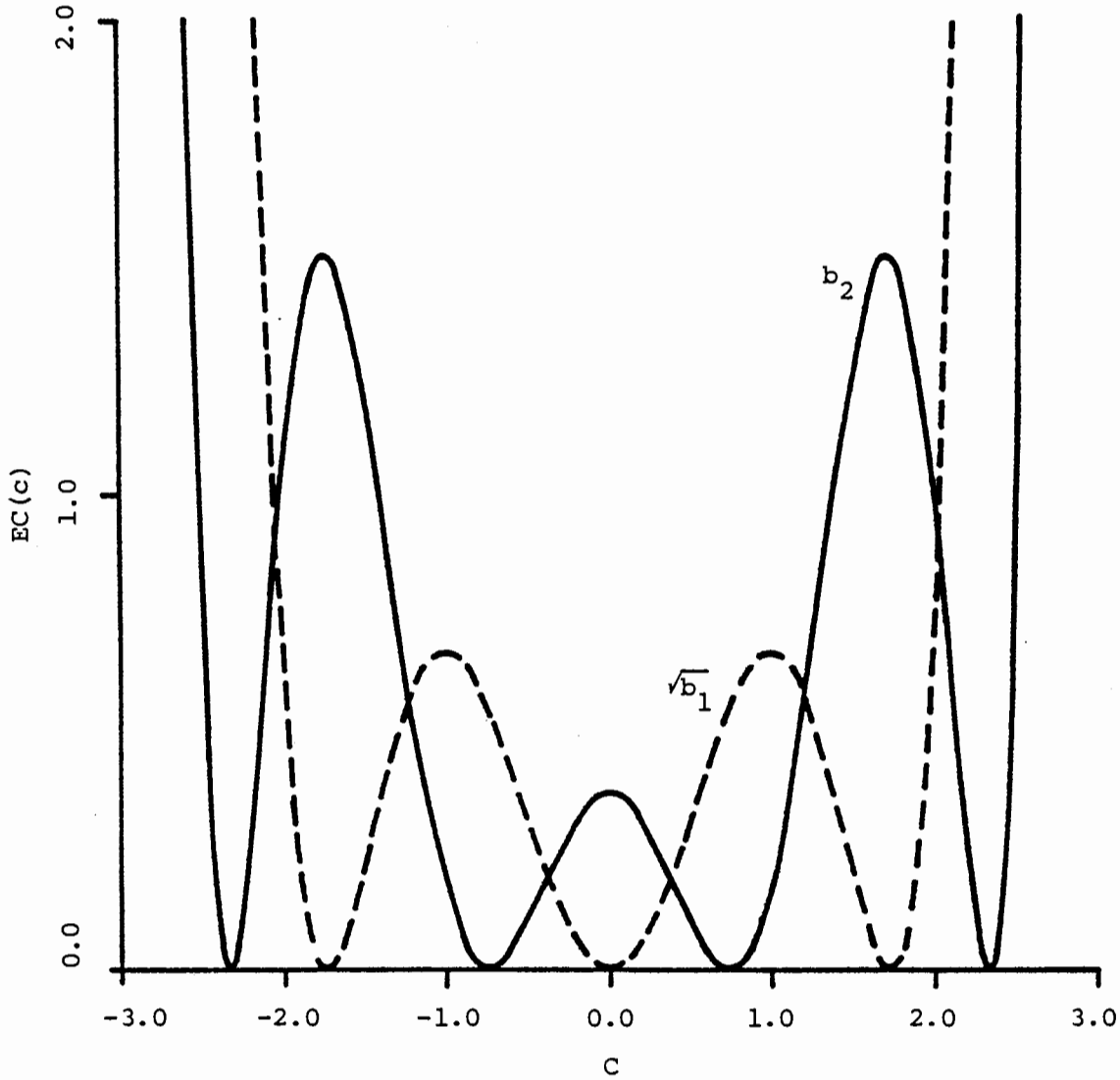

Figure 3. Efficacy curves for $\sqrt{b_1}$ and b_2.

C. The above development suggests an interesting use of the EC for tests of hypotheses. The EC for a test statistic can be interpreted as a large-sample measure of robustness of validity with respect to point contamination. That is, the greater the value of EC(c), the greater the probability of a Type I error when a small amount of contamination is present near the point c.

D. In related work Lambert (1979) defines the influence function for a test in terms of the instantaneous rate of change of the slope, either approximate as above or exact, as contamination is introduced at some fixed alternative distribution. Unlike the EC, which can be viewed as a measure of the robustness of power under the null hypothesis, Lambert's influence function for a test is a measure of the robustness of power under some alternative hypothesis.

E. A finite sample version of the IC termed the sensitivity curve is introduced by Tukey (1977). A stylized version in which sample values are replaced by expected order statistics is discussed by Andrews, Bickel, Hampel, Huber, Rogers, and Tukey (1972). Michael (1977) constructs stylized sensitivity curves for EDF statistics for testing a simple hypothesis and finds them very different in appearance from the ECs in Figure 1. For example, the stylized sensitivity curve for D is flat. Thus in this case the sensitivity curve is quite misleading.

F. Prescott (1976) introduces the stylized sensitivity surface by adding two arbitrary points to an idealized sample, and presents contours of such surfaces for a number of statistics for testing the composite hypothesis of normality. Michael (1977) introduces similar surfaces based on the IC by allowing equal amounts of contamination to be introduced at

two different points, and presents contours of surfaces for many of the same statistics. The results of these two approaches are not inconsistent, in contrast to the case of a simple null hypothesis discussed above in E.

REFERENCES

Abrahamson, I. G. (1965). On the stochastic comparison of tests of hypothesis. Unpublished Ph.D. dissertation, University of Chicago.

D'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. Biometrika, 58, 341-348.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes. Annals of Mathematical Statistics, 24, 193-212.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). Robust Estimates of Location. Princeton: Princeton University Press.

Bahadur, R. R. (1960). Stochastic comparison of tests. Annals of Mathematical Statistics, 31, 276-295.

Darling, D. A. (1957). The Kolmogorov-Smirnov, Cramér-von Mises tests. Annals of Mathematical Statistics, 28, 823-838.

Hampel, F. R. (1968). Contributions to the theory of robust estimation. Unpublished Ph.D. dissertation, University of California at Berkeley.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. Journal of the American Statistical Association, 69, 383-393.

Kendall, M. G. and Stuart, A. (1979). The Advanced Theory of Statistics. Vol. 2. 4th ed. New York: Macmillan.

Lambert, D. (1979). Influence functions for testing. Technical Report No. 152, Carnegie-Mellon University.

Michael, J. R. (1977). Goodness of fit: type II censoring; influence functions. Unpublished doctoral dissertation, Southern Methodist University.

Prescott, P. (1976). Comparison of tests for normality using stylized sensitivity surfaces. Biometrika, 63, 285-289.

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. Journal of the American Statistical Association, 69, 730-737.

Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. Annals of Statistics, 4, 357-369.

Tukey, J. W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley Pub. Co.

van der Watt, P. (1969). A comparison of certain tests of normality. South African Statistical Journal, 3, 69-82.

Wieand, H. S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. Annals of Statistics, 4, 1003-1011.