The Annals of Statistics
1979, Vol. 7, No. 1, 201–206

ON ESTIMATING THE SLOPE OF A STRAIGHT LINE WHEN BOTH VARIABLES ARE SUBJECT TO ERROR[1]

BY CLIFFORD SPIEGELMAN
Northwestern University

Let X_i and Y_i be random variables related to other random variables U_i, V_i, and W_i as follows: $X_i = U_i + W_i$, $Y_i = \alpha + \beta U_i + V_i$, $i = 1, \cdots, n$, where α and β are finite constants. Here X_i and Y_i are observable while U_i, V_i and W_i are not. This model is customarily referred to as the regression problem with errors in both variables and the central question is the estimation of β. We give a class of estimates for β which are asymptotically normal with mean β and variance proportional to $1/n^{\frac{1}{2}}$, under weak assumptions. We then show how to choose a good estimate of β from this class.

1. Introduction. Let X_i and Y_i be random variables related to other random variables U_i, V_i, and W_i as follows: $X_i = U_i + W_i$, $Y_i = \alpha + \beta U_i + V_i$, $i = 1, \cdots, n$, where α and β are finite constants. Here X_i and Y_i are observable while U_i, V_i and W_i are not. This model is customarily referred to as the regression problem with errors in both variables and the central question is the estimation of β. It is generally assumed that (U_i, V_i, W_i) $i = 1, \cdots, n$ are independent and identically distributed random vectors, and that U_1, V_1, and W_1 are mutually independent random variables.

An example that appears in economics is the following: U_i represents the true income of subject i. X_i his measured income, Y_i his measured consumption and W_i and V_i are measurement errors. β represents the marginal propensity to consume. References to this standard economic model can be found in Samuelson (1971).

We now discuss some previous contributions to the solution of this problem. Reiersöl (1950) proved that if U_1 is normal, β is identifiable if, and only if W_1 or V_1 have a distribution which is not divisible by a normal distribution, and if U_1 is not normal β is always identifiable. Neyman (1951) gave a consistent estimator for β when U_1 is not normal. Wolfowitz (1952, 1953, 1954a, 1954b, 1957) gave a method for estimating β when it is identifiable. This method is often reasonable if the distributions of W_1 and V_1 belong to a known small finite dimensional class, for example when W_1 and V_1 are normal. Kiefer and Wolfowitz (1956) show that the maximum likelihood estimates for this problem are consistent, if suitable regularity conditions attain. It is not clear whether the usual optimality properties of likelihood estimates apply to this problem. Rubin (1956) gives an estimate of β when the errors W_1 and V_1 are normal and U_1 is not. He uses the uniform convergence of

Received June 1974; revised November 1977.

[1]This research was supported in part by the Department of Mathematics at Northwestern University as well as the following grants: NIE-C-74-0115 and NSF-SOC-7103704. Now at the National Bureau of Standards, Washington, D. C.

AMS 1970 *subject classifications*. Primary 62J05; Secondary 62-02.

Key words and phrases. Errors in variables, regression, asymptotic distribution.

certain random functions to obtain his estimate. Other references may be found in a survey paper by Moran (1971).

Under the basic and customarily made assumption that U_1 is not normal we present an estimate b of β, which is easily obtained on a computer. When $E(X^4 + Y^4) < \infty$ this estimate is shown to be asymptotically normal with mean β and variance proportional to $n^{-\frac{1}{2}}$. This proof is based on Reiersöl's proof of identifiability and uses the technique of Rubin (1956).

2. The method. We begin by constructing an estimator β using a moment generating function approach.

Recall that our basic assumption is that U_1 is not normal. Suppose momentarily that $\beta \neq 0$. Since the arguments which we give do not depend upon the sign of β, let us take $\beta > 0$. Let us suppose that there are known positive numbers τ_1 and τ_2 such that $\tau_1 < \beta < \tau_2$. (If τ_1 and τ_2 are not known we shall show how one may deal with this situation later; see Comment 5 following Theorem 2′.)

Define $\phi_n(t_1, t_2) = \sum_{i=1}^{n} e^{t_1 X_i + t_2 Y_i}/n$ to be the joint sample moment generating function of X and Y and let $\phi_{nX}(t) = \phi_n(t, 0)$, and $\phi_{nY}(t) = \phi_n(0, t)$ be the sample moment generating functions of X and Y respectively. Define $\phi(t_1, t_2) = E\phi_n(t_1, t_2)$ to be the joint moment generating function of X and Y and let $\phi_X(t) = \phi(t, 0)$ and $\phi_Y(t) = \phi(0, t)$ be the moment generating functions of X and Y respectively. Define

$$\psi(b, t_1, t_2) = \frac{\phi(t_1, t_2)}{\phi_X(t_1)\phi_Y(t_2)} - \frac{\phi(bt_2, t_1/b)}{\phi_X(bt_2)\phi_Y(t_1/b)}.$$

This quantity is motivated by the identifiability proof in Reiersöl (1950) and has the property that $\psi(\beta, t_1, t_2) = 0$. If $\psi(b, t_1, t_2) = 0$ for t_1, t_2 in an open neighborhood of the origin it can be shown that $b = \beta$, by following the identifiability proof in Reiersöl (1950). We now define the sample analogue of ψ,

$$\psi_n(b, t_1, t_2) = \frac{\phi_n(t_1, t_2)}{\phi_{nX}(t_1)\phi_{nY}(t_2)} - \frac{\phi_n(bt_2, t_1/b)}{\phi_{nX}(bt_2)\phi_{nY}(t_1/b)}.$$

Choose $I \subset R^2$ to be a compact set containing the origin (we show how to choose it later; see Comment 4 after Theorem 2′). Define $F(b) = \int_I \psi^2(b, t_1, t_2)\, dt_1\, dt_2$. It is then clear from what we have said that $F(b) = 0$ iff $b = \beta$. Define the sample analogue of F, to be $F_n(b) = \int_I \psi_n^2(b, t_1, t_2)\, dt_1\, dt_2$.

We now define our estimate b_n of β. Define b_n as the minimizer of $F_n(b)$ for $\tau_1 \leqslant b \leqslant \tau_2$, i.e., $F_n(b_n) = \min_{\tau_1 \leqslant b \leqslant \tau_2} F_n(b)$, (if the minimizer is not unique we may take b_n to be any number between the smallest and largest minimizers). We now state our results.

THEOREM 1. *Let b_n be a minimizer of $F_n(b)$, $\tau_1 < b < \tau_2$, then b_n is a strongly consistent estimator for β*, i.e., $b_n \to \beta$ w.p. 1 *as $n \to \infty$.*

Under the same conditions as in Theorem 1 we have:

THEOREM 2. *b_n is asymptotically normal with mean β and variance proportional to $n^{-\frac{1}{2}}$ as $n \to \infty$.*

COMMENT 1. The limiting variance of b_n is complicated and is not given here. However, if the supremum of the m.g.f.'s in $F(\beta)$ are known then an upper bound on the variance can be obtained. This bound is useful for making a reasonable choice of I. A technique for deriving such a bound is shown in the proof of Theorem 2′.

COMMENT 2. The requirement that $\beta \neq 0$ is not restrictive. If we let

$$\bar{b}_n = b_n \quad \text{if } \Sigma_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})/n > n^{-\frac{1}{4}}$$
$$= 0 \quad \text{otherwise},$$

$\bar{b}_n$ satisfies the result of Theorem 2 for $\beta \in [0, \infty)$.

PROOF OF THEOREM 1. Step 1. Sample m.g.f.'s converge uniformly on compact intervals. Let G_n and G be the empirical and true distribution functions of (X, Y) respectively. For any finite constant, a, we have

$$\sup_{(t_1 t_2) \in I} |\phi_n(t_1, t_2) - \phi(t_1, t_2)|$$
$$\leqslant \sup_{(t_1, t_2) \in I} \left| \int_{|(x,y)| < a} \exp(t_1 x + t_2 y) d(G_n - G) \right|$$
$$+ \sup_{(t_1, t_2) \in I} \int_{|(x,y)| \geqslant a} \exp(t_1 x + t_2 y) d(G_n + G).$$

Since $\exp(t_1 x + t_2 y)$ is a uniformly continuous function on compact intervals and $\phi_n(t_1, t_2)$ converges uniformly w.p. 1 on finite sets $\{(t_{1i}, t_{2i}), i = 1, \cdots, k\}$ the first term on the right-hand side of this inequality converges to zero as $n \to \infty$, and for a and n large enough, the second term can be made arbitrarily small w.p. 1.

Step 2. $b_n \to \beta$ w.p. 1. Since sample m.g.f.'s converge uniformly on compact intervals w.p. 1 $\Psi_n(b, t_1, t_2) \to \Psi(b, t_1, t_2)$ as $n \to \infty$, uniformly in $[\tau_1, \tau_2] \times I$ w.p. 1. This implies that $F_n(b) \to F(b)$ as $n \to \infty$, uniformly in $[\tau_1, \tau_2]$. Therefore $b_n \to \beta$ w.p. 1 as $n \to \infty$. □

PROOF OF THEOREM 2. For $b_n \in (\tau_1, \tau_2)$, $0 = F_n'(b_n) = F_n'(\beta) + (b_n - \beta)F_n''(\xi_n)$ where ξ_n is between β and b_n. Since b_n is a consistent estimate of β it is sufficient to show that $n^{\frac{1}{2}}F_n'(\beta)/F_n''(\xi_n)$ has a limiting normal distribution with mean 0 and variance $\sigma^2(\beta) > 0$.

Let us define

$$f_n(t_1, t_2, b) = \frac{\phi_n(bt_2, t_1/b)}{\phi_{nX}(bt_2)\phi_{nY}(t_1/b)}$$

and $f(t_1, t_2, b)$ to be the corresponding quantity obtained when the subscripts n are deleted. Then

$$F_n'(\beta) = -2\int_I f_n'(t_1, t_2, \beta)\psi_n(t_1, t_2, \beta)\, dt_1\, dt_2.$$

We now rewrite the two terms of the integrand in $F_n'(\beta)$.

$$\psi_n(t_1, t_2, \beta) = \psi_n(t_1, t_2, \beta) - \psi(t_1, t_2, \beta),$$

which is the sum of continuous i.i.d. random variables $Q_i(t_1, t_2)$ with mean 0 and finite variance $+0_p(1/n)$ uniformly for $(t_1, t_2) \in I$. We also have $f_n(t_1, t_2, \beta) \to$

$f'(t_1, t_2, \beta)$ uniformly for $(t_1, t_2) \in I$ w.p. 1, and therefore

$$n^{\frac{1}{2}} F_n'(\beta) = -\int_I f'(t_1, t_2, \beta) \frac{\sum_{i=1}^n Q_i(t_1, t_2)}{n^{\frac{1}{2}}} dt_1\, dt_2 + 0_p(1/n^{\frac{1}{2}}).$$

It remains to show that $f'(t_1, t_2, \beta)$ is not zero in some neighborhood of the origin, for if it is we will see that both $F'(\beta)$ and $F''(\beta)$ are 0. If $f'(t_1, t_2, \beta) = 0$ then

$$(1) \qquad (\log f(t_1, t_2, \beta))' = \frac{\partial \log}{\partial b} \left[\frac{\phi_U\left(bt_2 + \frac{\beta}{b} t_1\right)}{\phi_U(bt_2)\phi_U(t_2/b)} \right] \Bigg|_{b=\beta} = 0.$$

Define $\tilde{\phi}(\cdot) = (\partial \log \phi_U(\cdot))/\partial t$. Expression (4) gives us

$$(2) \qquad \left(t_2 - \frac{t_1}{\beta}\right)\tilde{\phi}(t_1 + \beta t_2) = \tilde{\phi}(\beta t_2)t_2 - \tilde{\phi}(t_1)\left(\frac{t_1}{\beta}\right).$$

If we take the partial derivative of both sides of (5) with respect to t_1, and evaluate it along the line $t_1 = \beta t_2$ we obtain $\tilde{\phi}(t_1 + \beta t_2) = \tilde{\phi}(t_1) + \tilde{\phi}(t_1)t_1$, which implies that a one term Taylor expansion for $\tilde{\phi}$ is exact in the neighborhood of the origin. Whence $\tilde{\phi}''(t)$ is 0 in a neighborhood of the origin and log $\phi(t)$ is a quadratic contrary to the assumption that U_1 is not normal.

Finally we observe that $F_n''(\xi) \to_p F''(\beta) = \int_I (f'(t_1, t_2, \beta))^2\, dt_1\, dt_2 > 0$ as $n \to \infty$. □

If we replace the m.g.f.'s used to define b_n by characteristic functions, and redefine $\phi(t_1, t_2)$ and $f(t_1, t_2, b)$ we obtain a new estimate $b_n(C)$ of B as follows.

Choose $I_C \subset R^2$, to be a compact set such that for $b \in [\tau_1, \tau_2]$ and $(t_1, t_2) \in I_C$ both $|\phi_X(t_1)\phi_Y(t_2)| \geq C$ and $|\phi_X(bt_2)\phi_Y(t_2/b)| \geq C$, when $0 < C < 1$. Define $F_n(b, C) = \int_{I_C} \psi^2(b, t_1, t_2)\, dt_1\, dt_2$, and our new estimate $b_n(C)$ as the minimizer of $F_n(b, C)$, $\tau_1 \leq b \leq \tau_2$. Then we have the following theorems:

THEOREM 1′. *$b_n(C) \to \beta$ w.p. 1 as $n \to \infty$.*

THEOREM 2′. *If $E(X^4 + Y^4) < \infty$, $b_n(C)$ is asymptotically normal with mean β and variance $0(n^{-\frac{1}{2}})$, and in addition the limiting variance $\sigma^2(\beta, C)$ is bounded above by*

$$\frac{k_0\, \mu(I_C)}{C^2 \int_{I_C} |f'(t_1, t_2, \beta)|^2\, dt_1\, dt_2},$$

where k_0 is a positive constant which does not depend on C or β (we may take $k_0 = 36$), and μ is Lebesgue measure.

COMMENT 3. This bound shows the interaction between the choice of I_C and the distributions of U_1, V_1, W_1. It is not a sharp bound but it is somewhat intuitive. The actual performance of this estimate may be gleaned from simulations in Spiegelman (1976). Apparently $b_n(C)$ is a reasonable estimate.

For finite constants μ and σ a characteristic function which is equal to $\exp(i\mu t - \sigma^2 t^2)$ in the neighborhood of the origin must be equal to it everywhere (Kagan, Linnik and Rao (1973), page 21). Therefore the proofs of Theorem 1′ and 2′ follow by the same approach used to prove Theorems 1 and 2, except for the bound on $\sigma^2(\beta, C)$ which we now derive. By application of the Cauchy-Schwarz inequality we have

$$\sigma^2(\beta, C) \leqslant E\left[\frac{\int_{I_C}|f'(t_1, t_2, \beta)|\,|Q_1(t_1, t_2)|\,dt_1\,dt_2}{\int_{I_C}|f'(t_1, t_2, \beta)|^2\,dt_1\,dt_2}\right]^2$$

$$\leqslant \frac{\|Q_1(t_1, t_2)\|_\infty^2\,\mu(I_C)}{\int_{I_C}|f'(t_1, t_2, \beta)|^2\,dt_1\,dt_2}.$$

We also have the following inequality, $\|Q_1(t_1, t_2)\|_\infty^2 \leqslant k_0 C^{-2}$, because Q_1 is the sum of ratios of characteristic functions, on I_C. The denominators in Q_1 all have modulus greater than or equal to C, and the numerators are easily bounded by 1.

COMMENT 4. We suggest that C and I_C be chosen to minimize $[C^2\int_{I_C}|f'(t_1, t_2, \beta)|^2\,dt_1\,dt_2]^{-1}\mu(I_C)$. If f' is not known we suggest using f_n' in its place. This choice will make our bound on $\sigma^2(\beta, C)$ as small as possible.

COMMENT 5. If τ_1 and τ_2 are unknown and $E(X^4 + Y^4) < \infty$,

$$\tau_1 = \frac{\Sigma_{i=1}^n(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma_{i=1}^n(X_i - \bar{X})^2}$$

and

$$\tau_2 = \frac{\Sigma_{i=1}^n(Y_i - \bar{Y})^2}{\Sigma_{i=1}^n(X_i - \bar{X})(Y_i - \bar{Y})}$$

are useful choices for τ_1 and τ_2. They converge in probability to numbers bounding β from above and below, and our theorems remain valid for this choice. If $E(X^4 + Y^4) = \infty$ the variations of the Neyman and Wolfowitz estimates cited earlier may be used.

COMMENT 6. If the characteristic functions of X and Y are not known the sample characteristic functions of X and Y may be used in choosing I_C. Suppose we divide our sample into two independent groups. One group contains $\Delta(n)$ observations and the other the remainder. The group containing $\Delta(n)$ observations is used to estimate I_C, and the other group to obtain $b_n(C)$. Suppose that $\Delta(n) \to \infty$ as $n \to \infty$ and $\Delta(n) = o(n)$. Then conditioned on I_C Theorems 1′ and 2′ remain valid.

Acknowledgment. The author wishes to thank Professor Jerome Sacks, his advisor, for suggesting many improvements of earlier versions. He would also like to thank Rose Ray, Donald T. Campbell, Jacob Wolfowitz, Jayaram Sethuraman, an associate editor and the referees for many helpful suggestions.

REFERENCES

[1] CAMPBELL, D. and BORUCH, R. (1975). Making the case for randomized assignment to treatments by considering the alternatives: six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. To appear in *Central Issues in Social Program Evaluation* (C. Bennet and A. Lumsdaine, Eds.). Academic Press, New York.

[2] KAGAN, A., LINNIK, Y. and RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.

[3] KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

[4] MORAN, P. (1971). Estimating structural and functional relationships. *J. Multivariate Anal.* **1** 232–255.

[5] NEYMAN, J. (1951). Existence of consistent estimates of the directional parameter in a linear structural relation between two variables. *Ann. Math. Statist.* **22** 497–512.

[6] REIERSÖL, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica* **18** 375–389.

[7] RUBIN, H. (1956). Uniform convergence of random functions with applications to statistics. *Ann. Math. Statist.* **27** 200–203.

[8] SAMUELSON, P. A. (1971). *Foundations of Economic Analysis*. Atheneum, New York.

[9] SPIEGELMAN, C. (1976). Ph.D. thesis, Northwestern University.

[10] WOLFOWITZ, J. (1952). Consistent estimators of the parameters of a linear structural relation. *Skand. Aktuarietidskr.* **35** 132–151.

[11] WOLFOWITZ, J. (1953). Estimation by the minimum distance method. *Ann. Math. Statist.* **28** 75–88.

[12] WOLFOWITZ, J. (1954a). Estimation of the components of stochastic structures. *Proc. Nat. Acad. Sci.* **40** 602–606.

[13] WOLFOWITZ, J. (1954b). Estimation by the minimum distance method in non-parametric stochastic difference equations. *Ann. Math. Statist.* **25** 203–217.

[14] WOLFOWITZ, J. (1957). The minimum distance method. *Ann. Math. Statist.* **28** 75–88.

9910 WALKERS HOUSE RD, #1
GAITHERSBURG, MARYLAND 20760