The Annals of Statistics
1989, Vol. 17, No. 1, 282–292

EMPIRICAL PROCESSES BASED UPON RESIDUALS FROM ERRORS-IN-VARIABLES REGRESSIONS[1]

BY STEPHEN M. MILLER

U.S. Bureau of Labor Statistics

Multivariate errors-in-variables regression models with normal errors are considered and residuals, similar to those calculated from ordinary least squares regressions, are defined for these models. It is shown that under the assumption of a $n^{1/2}$-consistent estimator of the vector of regression coefficients, certain empirical processes based upon the residuals converge to the same Gaussian process as that of an infinite sequence of normal random variables standardized by their sample mean sample variance.

1. Introduction. There has been considerable interest during the last ten years in the behavior of empirical distribution functions based on regression residuals. Much of this interest stems from the desire to use residuals in place of unobservable experimental errors in goodness-of-fit tests, often with a specific interest in tests of normality. The early work of Durbin (1973), Rao and Sethuraman (1975), and Neuhaus (1976) set the stage by examining empirical processes when parameters of the underlying distribution function were estimated. Mugantseva (1977) and Pierce and Kopecky (1979) were the first to derive the limiting distribution of the sample empirical process based on least squares residuals. Shorack (1984) extended these results to the general class of residuals based on $n^{1/2}$-consistent estimators of the regression coefficients. [See also Section 4.6 in Shorack and Wellner (1986).] Wood (1984, 1981a) had earlier examined the special cases of ridge regression and multiple regression. Portnoy (1986) has examined the situation where the number of regression parameters is large and Loynes (1980) looked at the weak convergence of processes based on generalized residuals defined by Cox and Snell (1968). In all of these papers the authors have concentrated on regression models with only one dependent variable and where the independent variables are measured without error, though they did not always assume normally distributed equation errors as we will in the present, more complicated situation.

In this paper we further generalize the theory of empirical processes based on residuals to the case of multivariate errors-in-variables regression models. These models are defined by

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{x}_t\boldsymbol{\beta} + \mathbf{e}_t, \\ \mathbf{X}_t &= \mathbf{x}_t + \mathbf{u}_t, \qquad t = 1,\dots,n, \end{aligned} \tag{1.1}$$

Received June 1987; revised April 1988.

[1]Much of this research was conducted while at Iowa State University and was partly supported by the U.S. Bureau of the Census through Joint Statistical Agreement JSA 86-2.

AMS 1980 *subject classifications*. Primary 62G35; secondary 62E20, 62J99, 60F05.

Key words and phrases. Errors-in-variables, empirical processes, regression residuals, multivariate normality, measurement errors, limiting distribution.

where $\mathbf{Y}_t$ and $\mathbf{X}_t$ are observable random vectors of dimensions $1 \times r$ and $1 \times k$, respectively. The $k \times r$ matrix $\boldsymbol{\beta}$ contains the unknown regression coefficients. We assume that k and r are fixed. The sequence of $1 \times k$ true vectors $\mathbf{x}_t$ (the values of which may vary with the sample size) is not directly observable because of the $1 \times k$ vectors of unobservable measurement errors $\mathbf{u}_t$. The presence of the measurement errors $\mathbf{u}_t$ distinguishes model (1.1) from the usual multivariate regression model which assumes that the true $\mathbf{x}_t$ are observed directly. The $1 \times r$ vectors $\mathbf{e}_t$ are also unobservable and can arise from both measurement error and equation error. Equation error represents the fact that the linear relation between $\mathbf{Y}_t$ and $\mathbf{x}_t$ in (1.1) may not hold exactly.

We can write model (1.1) as

$$\text{(1.2)} \qquad \begin{aligned} \mathbf{Y}_t &= \mathbf{X}_t\boldsymbol{\beta} + \mathbf{v}_t, \\ \mathbf{v}_t &= \mathbf{e}_t - \mathbf{u}_t\boldsymbol{\beta}, \qquad t = 1, \dots, n. \end{aligned}$$

In this formulation the model looks like the usual multivariate regression model, except that $\mathbf{X}_t$ and $\mathbf{v}_t$ are correlated due to the measurement error in $\mathbf{X}_t$. As in the usual multivariate regression model if $\boldsymbol{\beta}$ is known, the vector $\mathbf{v}_t$ becomes observable since

$$\text{(1.3)} \qquad \mathbf{v}_t = \mathbf{Y}_t - \mathbf{X}_t\boldsymbol{\beta}, \qquad t = 1, \dots, n.$$

In practice we can only obtain an estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, so we define

$$\text{(1.4)} \qquad \hat{\mathbf{v}}_t = \mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}, \qquad t = 1, \dots, n,$$

and call these the residual vectors from models (1.1) and (1.2). These residuals have been suggested by Fuller (1987) for model checking, and the limiting behavior of several diagnostic procedures based on (1.4) have been examined by Miller (1986a). In this paper we examine the limiting behavior of empirical processes based upon the residuals of (1.4). Applications of our results to composite goodness-of-fit tests for normality are briefly discussed in Section 4 and have been previously discussed in Miller (1986b).

In the next section we present our assumptions, and define some additional notation. The main theorems are presented in Section 3 and additional technical results are given in the Appendix.

2. Assumptions and notation. We begin by stating our assumptions about the distribution of the errors in (1.1), the properties of the vectors $\mathbf{x}_t$ and the estimator $\hat{\boldsymbol{\beta}}$ used in defining (1.4).

ASSUMPTION 1. Let $\boldsymbol{\varepsilon}_t = (\mathbf{e}_t, \mathbf{u}_t)'$ and $p = k + r$. The $\boldsymbol{\varepsilon}_t$ are independent $N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon})$ random vectors.

Assumption 1 implies that the vectors $\mathbf{v}_t$ are independent $N_r(\mathbf{0}, \boldsymbol{\Sigma}_{vv})$ random vectors, where

$$\text{(2.1)} \qquad \begin{aligned} \boldsymbol{\Sigma}_{vv} &= \boldsymbol{\Sigma}_{ee} - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{ue} - \boldsymbol{\Sigma}_{eu}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_{uu}\boldsymbol{\beta}, \\ \boldsymbol{\Sigma}_{\varepsilon\varepsilon} &= \begin{bmatrix} \boldsymbol{\Sigma}_{ee} & \boldsymbol{\Sigma}_{eu} \\ \boldsymbol{\Sigma}_{ue} & \boldsymbol{\Sigma}_{uu} \end{bmatrix}. \end{aligned}$$

We allow $\Sigma_{\varepsilon\varepsilon}$ to be singular, but require Σ_{vv} to be positive definite. An example involving singular $\Sigma_{\varepsilon\varepsilon}$ is when the model contains an intercept, so Σ_{uu} contains a corresponding row and column of 0's.

ASSUMPTION 2. The $\mathbf{x}_t$ are independent of the ε_j for all t and j and the sequence of vectors $\mathbf{x}_t$ satisfies the condition

$$\mathbf{m}_{xx} \to \overline{\mathbf{m}}_{xx} \quad \text{a.s.,} \tag{2.2}$$

where $\overline{\mathbf{m}}_{xx}$ is a finite $k \times k$ nonnegative definite matrix and

$$\begin{aligned} \mathbf{m}_{xx} &= (n-1)^{-1} \sum_{t=1}^{n} (\mathbf{x}_t - \bar{\mathbf{x}})'(\mathbf{x}_t - \bar{\mathbf{x}}), \\ \bar{\mathbf{x}} &= n^{-1} \sum_{t=1}^{n} \mathbf{x}_t. \end{aligned} \tag{2.3}$$

Throughout this paper, $\mathbf{m}_{ab}$ denotes the corrected cross product matrix with divisor $(n-1)$ for any sequence of vectors $\mathbf{a}_t$ and $\mathbf{b}_t$, $t = 1, \ldots, n$.

Assumption 2 allows the $\mathbf{x}_t$ vectors to be quite general, with the possibility of having both fixed and random components as well as correlation between observations. The most common cases are the *functional model* (fixed $\mathbf{x}_t$), the *structural model* (random $\mathbf{x}_t$) and the *ultrastructural model* (random $\mathbf{x}_t$ with possibly different means). As examples, the usual multivariate regression model is a special case of the functional model with $\Sigma_{uu} = \mathbf{0}$ and the factor model is a special case of the structural model with $\Sigma_{\varepsilon\varepsilon}$ a diagonal matrix.

ASSUMPTION 3. $\hat{\boldsymbol{\beta}}$ is any estimator such that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(n^{-1/2})$.

Methods for producing $n^{1/2}$-consistent estimators can be found in Fuller (1987), Amemiya and Fuller (1984) and Gleser (1981).

It will be useful to define two additional random vectors, namely,

$$\begin{aligned} \ddot{\mathbf{x}}_t &= \mathbf{X}_t - \mathbf{v}_t \Sigma_{vv}^{-1} \Sigma_{vu}, \qquad & t &= 1, \ldots, n, \\ \ddot{\mathbf{x}}_{t(i)} &= \mathbf{X}_t - v_{ti} \sigma_{vvii}^{-1} \Sigma_{v(i)u}, \qquad & t &= 1, \ldots, n,\ i = 1, \ldots, r, \end{aligned} \tag{2.4}$$

where $\Sigma_{vu} = E\{\mathbf{v}_t' \mathbf{u}_t\}$, $\sigma_{vvii} = V\{v_{ti}\}$ and $\Sigma_{v(i)u} = E\{v_{ti} \mathbf{u}_t\}$. Both random vectors are predictors of the true vector $\mathbf{x}_t$. The vector $\ddot{\mathbf{x}}_t$ is constructed by subtracting from $\mathbf{X}_t$ the best predictor, under normality, of $\mathbf{u}_t$ given the entire vector $\mathbf{v}_t$, while $\ddot{\mathbf{x}}_{t(i)}$ is constructed by subtracting the best predictor of $\mathbf{u}_t$ given only v_{ti}. From Assumptions 1 and 2 it follows that $\ddot{\mathbf{x}}_t$ and $\mathbf{v}_j$ are independent for all t and j, while $\ddot{\mathbf{x}}_{t(i)}$ and v_{ji} are independent for all t and j given $i = 1, \ldots, r$.

Finally we define the sample covariance matrix of the residual vectors for $n > k$ by

$$\begin{aligned} \mathbf{S}_{vv} &= (n-k)^{-1} \sum_{t=1}^{n} (\hat{\mathbf{v}}_t - \bar{\hat{\mathbf{v}}})'(\hat{\mathbf{v}}_t - \bar{\hat{\mathbf{v}}}), \\ \bar{\hat{\mathbf{v}}} &= n^{-1} \sum_{t=1}^{n} \hat{\mathbf{v}}_t \end{aligned} \tag{2.5}$$

and we define the *marginally standardized* residual vectors by

$$s_{vvii}^{-1/2}\left(\hat{v}_{ti} - \bar{\hat{v}}_i\right), \qquad t = 1, \dots, n,\ i = 1, \dots, r, \tag{2.6}$$

where $s_{vvii} = (\mathbf{S}_{vv})_{ii}$. We also define the *jointly standardized* residual vectors by

$$\left(\hat{\mathbf{v}}_t - \bar{\hat{\mathbf{v}}}\right)\mathbf{S}_{vv}^{-1/2}, \qquad t = 1, \dots, n. \tag{2.7}$$

Throughout this paper, $\mathbf{M}^{1/2}$ denotes the symmetric square root of the matrix $\mathbf{M}$. Finally, we define diag($\mathbf{M}$) to be a matrix with diagonal elements equal to those of $\mathbf{M}$ but with off-diagonal elements equal to 0.

3. Main results. We first present two lemmas which give representations of the marginally standardized and jointly standardized residual vectors. While interesting in their own right, these lemmas will later form the basis for the proofs of our main theorems.

LEMMA 3.1. *The marginally standardized residuals can be represented as*

$$s_{vvii}^{-1/2}\left(\hat{v}_{ti} - \bar{\hat{v}}_i\right) = \left(v_{ti} - \bar{v}_i\right)m_{vvii}^{-1/2}(1 + a_{ni}) + \left(\ddot{\mathbf{x}}_{t(i)} - \bar{\ddot{\mathbf{x}}}_{(i)}\right)\mathbf{B}_{ni},$$

$$a_{ni} = m_{vvii}^{1/2} s_{vvii}^{-1/2}\left[1 - \sigma_{vvii}^{-1}\boldsymbol{\Sigma}_{v(i)u}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\right)\right] - 1,$$

$$\mathbf{B}_{ni} = -s_{vvii}^{-1/2}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\right),$$

where $\mathbf{B}_{ni} = O_p(n^{-1/2})$, $a_{ni} = o_p(n^{-1/2})$, $m_{vvii} = (\mathbf{m}_{vv})_{ii}$ *and* $\hat{\boldsymbol{\beta}}_i, \boldsymbol{\beta}_i$ *are the ith columns of* $\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}$, *respectively.*

PROOF. We can write

$$\begin{aligned} \hat{v}_{ti} &= Y_{ti} - \mathbf{X}_t\hat{\boldsymbol{\beta}}_i \\ &= v_{ti} - \mathbf{X}_t\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\right) \\ &= v_{ti}\left[1 - \sigma_{vvii}^{-1}\boldsymbol{\Sigma}_{v(i)u}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\right)\right] - \ddot{\mathbf{x}}_{t(i)}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\right). \end{aligned} \tag{3.1}$$

The representation in terms of a_{ni} and $\mathbf{B}_{ni}$ now follows by algebra. The order result for $\mathbf{B}_{ni}$ follows from Assumption 3. Using (3.1) we can write

$$s_{vvii} = m_{vvii}\left[1 - \sigma_{vvii}^{-1}\boldsymbol{\Sigma}_{v(i)u}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\right)\right]^2 + o_p\left(n^{-1/2}\right), \tag{3.2}$$

where we used the fact that the sample covariance between $\ddot{\mathbf{x}}_{t(i)}$ and v_{ti} is $o_p(1)$, in determining the order of the remainder. The order result for a_{ni} now follows from (3.2). □

LEMMA 3.2. *The jointly standardized residuals can be represented as*

$$\left(\hat{\mathbf{v}}_t - \bar{\hat{\mathbf{v}}}\right)\mathbf{S}_{vv}^{-1/2} = \left(\mathbf{v}_t - \bar{\mathbf{v}}\right)\mathbf{m}_{vv}^{-1/2}(\mathbf{I}_r + \mathbf{A}_n) + \left(\ddot{\mathbf{x}}_t - \bar{\ddot{\mathbf{x}}}\right)\mathbf{B}_n,$$

$$\mathbf{A}_n = \mathbf{m}_{vv}^{1/2}\left[\mathbf{I}_r - \boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vu}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right]\mathbf{S}_{vv}^{-1/2} - \mathbf{I}_r,$$

$$\mathbf{B}_n = -\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\mathbf{S}_{vv}^{-1/2},$$

where $\mathbf{B}_n = O_p(n^{-1/2})$ *and* $\mathbf{A}_n$ *has off-diagonal elements which are* $O_p(n^{-1/2})$ *and diagonal elements which are* $o_p(n^{-1/2})$.

PROOF. We can write the residual vectors as

$$\begin{aligned}\hat{\mathbf{v}}_t &= \mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol\beta}\\ &= \mathbf{v}_t - \mathbf{X}_t(\hat{\boldsymbol\beta} - \boldsymbol\beta)\\ &= \mathbf{v}_t[\mathbf{I}_r - \boldsymbol\Sigma_{vv}^{-1}\boldsymbol\Sigma_{vu}(\hat{\boldsymbol\beta} - \boldsymbol\beta)] - \ddot{\mathbf{x}}_t(\hat{\boldsymbol\beta} - \boldsymbol\beta).\end{aligned} \tag{3.3}$$

The representation in terms of $\mathbf{A}_n$ and $\mathbf{B}_n$ now follows by algebra. The result $\mathbf{B}_n = O_p(n^{-1/2})$ follows by Assumption 3. By (3.3) we can write

$$\mathbf{S}_{vv} = [\mathbf{I}_r - \boldsymbol\Sigma_{vv}^{-1}\boldsymbol\Sigma_{vu}(\hat{\boldsymbol\beta} - \boldsymbol\beta)]'\mathbf{m}_{vv}[\mathbf{I}_r - \boldsymbol\Sigma_{vv}^{-1}\boldsymbol\Sigma_{vu}(\hat{\boldsymbol\beta} - \boldsymbol\beta)] + o_p(n^{-1/2}), \tag{3.4}$$

where we used the fact that the sample covariance between $\ddot{\mathbf{x}}_t$ and $\mathbf{v}_t$ is $o_p(1)$, in determining the order of the remainder. Now define

$$\mathbf{Q}_n = \mathbf{m}_{vv}^{1/2}[\mathbf{I}_r - \boldsymbol\Sigma_{vv}^{-1}\boldsymbol\Sigma_{vu}(\hat{\boldsymbol\beta} - \boldsymbol\beta)]\mathbf{S}_{vv}^{-1/2}. \tag{3.5}$$

It follows that $\mathbf{Q}_n = \mathbf{I}_r + O_p(n^{-1/2})$ and by (3.4), $\mathbf{Q}_n'\mathbf{Q}_n = \mathbf{I}_r + o_p(n^{-1/2})$. By Lemma A in the Appendix, it follows that $\operatorname{diag}(\mathbf{Q}_n - \mathbf{I}_r) = o_p(n^{-1/2})$. Since $\mathbf{A}_n = \mathbf{Q}_n - \mathbf{I}_r$, the order results for $\mathbf{A}_n$ are established. □

The key result of Lemma 3.2 is that the stochastic order of the diagonal elements of the matrix $\mathbf{A}_n$ is $o_p(n^{-1/2})$, which is of smaller order than the rest of the matrix. This is a generalization of the $o_p(n^{-1/2})$ result for a_{ni} from Lemma 3.1.

Before presenting our main theorems we need to introduce the Gaussian process W. Let $R_1, \dots, R_n$ be independent $N(\mu, \sigma^2)$ random variables with $\sigma^2 > 0$. Define for $\omega \in [0,1]$,

$$\begin{aligned} B_n(\omega) &= n^{-1/2}\sum_{t=1}^{n}\{1[\Phi(Z_t) \le \omega] - \omega\},\\ Z_t &= s^{-1}(R_t - \overline{R}), \qquad t = 1, \dots, n,\end{aligned} \tag{3.6}$$

where $1(\cdot)$ is the indicator function, Φ is the standard normal distribution function and $s^2 = (n-1)^{-1}\sum_{t=1}^{n}(R_t - \overline{R})^2$. Then $B_n \to_{\mathscr{D}} W$ as $n \to \infty$, where $\to_{\mathscr{D}}$ denotes weak convergence in the Skorohod space $D[0,1]$ and W is a Gaussian process on $[0,1]$ with mean and covariance function given by

$$\begin{aligned} E\{W(\omega)\} &= 0 \quad \text{for } \omega \in [0,1],\\ E\{W(\omega_1)W(\omega_2)\} &= \min(\omega_1, \omega_2) - \omega_1\omega_2\\ &\quad - [1 + \tfrac{1}{2}\Phi^{-1}(\omega_1)\Phi^{-1}(\omega_2)]\\ &\quad \times\phi[\Phi^{-1}(\omega_1)]\phi[\Phi^{-1}(\omega_2)]\end{aligned} \tag{3.7}$$

for $\omega_1, \omega_2 \in [0,1]$, where ϕ is the standard normal density function. This result can be found in Kac, Kiefer and Wolfowitz (1955).

We are now ready to present our main theorems. Theorem 3.1 describes the limiting behavior of empirical processes based on the components of the marginally standardized residual vectors.

THEOREM 3.1. *Denote the marginally standardized residuals by*

$$Z_{ti} = s_{vvii}^{-1/2}\left(\hat{v}_{ti} - \bar{\hat{v}}_i\right), \qquad t = 1, \dots, n, \; i = 1, \dots, r,$$

and for $\omega \in [0,1]$ *define*

$$W_{ni}(\omega) = n^{-1/2} \sum_{t=1}^{n} \{1[\Phi(Z_{ti}) \le \omega] - \omega\}, \qquad i = 1, \dots, r.$$

Then $W_{ni} \to_{\mathscr{D}} W$ *as* $n \to \infty$ *for* $i = 1, \dots, r$.

PROOF. By Lemma 3.1 the marginally standardized residuals for $i = 1, \dots, r$ have a representation which satisfies the conditions of Theorem A in the Appendix, so the result follows by that theorem in conjunction with the result of Kac, Kiefer and Wolfowitz (1955) mentioned above. □

Note that Theorem 3.1 deals with the convergence of each of the process W_{ni} separately and does not describe their joint behavior. While each process individually has the same limiting distribution, when treated jointly the limiting processes are correlated. This joint correlation disappears in our next Theorem 3.2 where we consider the limiting behavior of processes based on the components of the jointly standardized residual vectors.

THEOREM 3.2. *Denote the jointly standardized residuals by*

$$\mathbf{Z}_t = \left(\hat{\mathbf{v}}_t - \bar{\hat{\mathbf{v}}}\right)\mathbf{S}_{vv}^{-1/2}, \qquad t = 1, \dots, n,$$

where $\mathbf{Z}_t = (Z_{t1}, \dots, Z_{tr})$ *and for* $\omega \in [0,1]$ *define*

$$W_{ni}(\omega) = n^{-1/2} \sum_{t=1}^{n} \{1[\Phi(Z_{ti}) \le \omega] - \omega\}, \qquad i = 1, \dots, r.$$

Then $W_{ni} \to_{\mathscr{D}} W$ *as* $n \to \infty$ *for* $i = 1, \dots, r$ *and the limiting processes are independent.*

PROOF. Define the random vector

$$\boldsymbol{\xi}_t = \mathbf{v}_t \boldsymbol{\Sigma}_{vv}^{-1/2}, \qquad t = 1, \dots, n, \tag{3.8}$$

so that $\boldsymbol{\xi}_t$ are independent $N_r(\mathbf{0}, \mathbf{I}_r)$ random vectors. Using Lemma 3.2 we can write

$$\begin{aligned} \mathbf{Z}_t &= \left(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}\right)\mathbf{m}_{\xi\xi}^{-1/2}\mathbf{Q}_n(\mathbf{I}_r + \mathbf{A}_n) + \left(\ddot{\mathbf{x}}_t - \bar{\ddot{\mathbf{x}}}\right)\mathbf{B}_n, \\ \mathbf{Q}_n &= \mathbf{m}_{\xi\xi}^{1/2}\boldsymbol{\Sigma}_{vv}^{1/2}\mathbf{m}_{vv}^{-1/2} = \left(\boldsymbol{\Sigma}_{vv}^{-1/2}\mathbf{m}_{vv}\boldsymbol{\Sigma}_{vv}^{-1/2}\right)^{1/2}\left(\boldsymbol{\Sigma}_{vv}^{1/2}\mathbf{m}_{vv}^{-1/2}\right). \end{aligned} \tag{3.9}$$

Notice that $\mathbf{Q}_n$ is an orthogonal matrix converging in probability to $\mathbf{I}_r$ at the rate $O_p(n^{-1/2})$. For notational convenience we write the (i, l)th element of the matrix $\mathbf{m}_{\xi\xi}^{-1/2}\mathbf{Q}_n(\mathbf{I}_r + \mathbf{A}_n)$ as

$$\left[\mathbf{m}_{\xi\xi}^{-1/2}\mathbf{Q}_n(\mathbf{I}_r + \mathbf{A}_n)\right]_{il} = \sum_{j=1}^{r}\sum_{k=1}^{r} m_{ij}^{*} q_{jk}(\delta_{kl} + a_{kl}), \tag{3.10}$$

where a_{kl}, q_{jk} and m_{ij}^{*} are elements of the matrices $\mathbf{A}_n$, $\mathbf{Q}_n$ and $\mathbf{m}_{\xi\xi}^{-1/2}$, respectively, and $\delta_{kl} = 1$ if $k = l$ and 0 otherwise. First examine the case where $i = l$. Then

$$\left[\mathbf{m}_{\xi\xi}^{-1/2}\mathbf{Q}_n(\mathbf{I}_r + \mathbf{A}_n)\right]_{ii} = \sum_{j=1}^{r}(1 + a_{ii}) m_{ij}^{*} q_{ji} + \sum_{j=1}^{r}\sum_{k\neq i}^{r} m_{ij}^{*} q_{jk} a_{ki}. \tag{3.11}$$

From Lemma 3.2 $a_{ii} = o_p(n^{-1/2})$ and $a_{ki} = O_p(n^{-1/2})$ when $k \neq i$. From the definition of $\mathbf{m}_{\xi\xi}^{-1/2}$ we have $m_{ij}^{*} = O_p(n^{-1/2})$ when $i \neq j$ and by applying Lemma A to the matrix $\mathbf{m}_{\xi\xi}^{-1/2}\{\operatorname{diag}(\mathbf{m}_{\xi\xi})\}^{1/2}$ we obtain $m_{ii}^{*} = m_{\xi\xi ii}^{-1/2} + o_p(n^{-1/2})$, where $m_{\xi\xi ii}^{-1/2}$ is the reciprocal of the square root of the sample variance of ξ_{ti}, $t = 1, \dots, n$. From the definition of $\mathbf{Q}_n$ and by Lemma A we have $q_{ii} = 1 + o_p(n^{-1/2})$ and $q_{ij} = O_p(n^{-1/2})$ when $i \neq j$. Thus there exist random variables c_{ni} such that

$$\left[\mathbf{m}_{\xi\xi}^{-1/2}\mathbf{Q}_n(\mathbf{I}_r + \mathbf{A}_n)\right]_{ii} = (1 + c_{ni}) m_{\xi\xi ii}^{-1/2}, \qquad i = 1, \dots, r, \tag{3.12}$$

where $c_{ni} = o_p(n^{-1/2})$. Returning to (3.9) we define the random variables d_{nil} by

$$d_{nil} = \left[\mathbf{m}_{\xi\xi}^{-1/2}\mathbf{Q}_n(\mathbf{I}_r + \mathbf{A}_n)\right]_{il}, \qquad i \neq l, \tag{3.13}$$

and note that $d_{nil} = O_p(n^{-1/2})$ for $i \neq l$. Letting $\mathbf{b}_{ni}$ denote the ith column of $\mathbf{B}_n$, we can represent the components of $\mathbf{Z}_t$ as

$$Z_{ti} = (1 + c_{ni}) m_{\xi\xi ii}^{-1/2}(\xi_{ti} - \bar{\xi}_i) + \sum_{j\neq i}^{r} d_{nji}(\xi_{tj} - \bar{\xi}_j) - (\ddot{\mathbf{x}}_t - \bar{\ddot{\mathbf{x}}})\mathbf{b}_{ni}, \tag{3.14}$$

for $t = 1, \dots, n$ and $i = 1, \dots, r$.

Since ξ_{ti} is independent of $\ddot{\mathbf{x}}_t$ and ξ_{tj} for $j \neq i$ the convergence result $W_{ni} \to_{\mathscr{D}} W$ for $i = 1, \dots, r$ follows from Theorem A in the Appendix (using the order properties established for c_{ni}, d_{nij} and $\mathbf{b}_{ni}$) in conjunction with the Kac, Kiefer and Wolfowitz (1955) result mentioned prior to Theorem 3.1. The independence of the limiting processes follows from the independence of $\xi_{t1}, \dots, \xi_{tr}$. □

Theorem 3.2 has an interesting corollary. Let $\mathscr{S}$ be a fixed subset of $\{1, \dots, r\}$ and let m be the number of elements of $\mathscr{S}$. Using the notation of Theorem 3.2, define for $\omega \in [1, 0]$

$$W_{n\mathscr{S}}(\omega) = m^{-1/2} \sum_{i\in\mathscr{S}} W_{ni}(\omega). \tag{3.15}$$

Then $W_{n\mathscr{S}} \to_{\mathscr{D}} W$ as $n \to \infty$. In the special case $\mathscr{S} = \{1, \dots, r\}$ we get that

$$(rn)^{-1/2} \sum_{i=1}^{r}\sum_{t=1}^{n} \{1[\Phi(Z_{ti}) \leq \omega] - \omega\} \to_{\mathscr{D}} W. \tag{3.16}$$

This is similar to a result by Wood (1981b) who proved (3.16) for the case when Z_{ti} is the ith component of

$$\mathbf{Z}_t = \mathbf{m}_{YY}^{-1/2}(\mathbf{Y}_t - \overline{\mathbf{Y}}), \qquad t = 1, \dots, n, \tag{3.17}$$

and $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are independent $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vectors with positive definite $\boldsymbol{\Sigma}$.

4. Applications to tests of normality. A consequence of Theorem 3.1 and Theorem 3.2 is that we can easily construct weak convergence results for functions of $D[0,1]$ which are continuous with respect to the Skorohod topology (except at possibly a set of probability 0 with respect to the distribution of the limiting process W). In particular, if T is such a function and W_n is one of the sample processes in either Theorem 3.1 or Theorem 3.2, then $T(W_n) \to_{\mathscr{D}} T(W)$. Many tests of normality, based on residuals, can be written as functions of sample processes and their theoretical limiting distributions can be expressed in this way. Fortunately, tables have already been constructed which contain percentage points for the limiting distributions of the most commonly used tests.

In practice we have used the marginally standardized residuals to construct tests of marginal normality for components of the vectors $\mathbf{v}_t$ and sometimes combined these tests by using Bonferroni's inequality to test for multivariate normality. We have also used the jointly standardized residuals to test for multivariate normality of the vectors $\mathbf{v}_t$ by combining components into groups as described in (3.15) and (3.16). The most common choice of groupings has been to either combine all the components together into a single group or form r groups based upon each component separately. Our overall recommendation is that if someone is comfortable with using a particular large sample test in the usual regression setting, then they should also feel comfortable with the same test when applied to errors-in-variables residuals.

APPENDIX

LEMMA A. *Let $\mathbf{Q}_n$ be a random $r \times r$ matrix such that $\mathbf{Q}_n = \mathbf{I}_r + O_p(a_n)$ and* $\operatorname{diag}(\mathbf{Q}_n'\mathbf{Q}_n) = \mathbf{I}_r + o_p(a_n)$ *for a sequence of positive real numbers $\{a_n\}_{n=1}^{\infty}$ decreasing to* 0. *Then* $\operatorname{diag}(\mathbf{Q}_n - \mathbf{I}_r) = o_p(a_n)$.

PROOF. Let $\boldsymbol{\Delta}_n = \mathbf{Q}_n - \mathbf{I}_r$. Then

$$\mathbf{Q}_n'\mathbf{Q}_n = \mathbf{I}_r + \boldsymbol{\Delta}_n + \boldsymbol{\Delta}_n' + \boldsymbol{\Delta}_n'\boldsymbol{\Delta}_n, \tag{A.1}$$

so $\operatorname{diag}(\boldsymbol{\Delta}_n' + \boldsymbol{\Delta}_n) = o_p(a_n)$ since $\operatorname{diag}(\boldsymbol{\Delta}_n'\boldsymbol{\Delta}_n) = O_p(a_n^2)$. Thus

$$2\operatorname{diag}(\mathbf{Q}_n - \mathbf{I}_r) = \operatorname{diag}(\boldsymbol{\Delta}_n' + \boldsymbol{\Delta}_n) = o_p(a_n). \qquad \square$$

THEOREM A. *Let $\xi_1, \dots, \xi_n$ be independent $N(\mu, \sigma^2)$ random variables with $\sigma^2 > 0$ and independent of the random $1 \times k$ vectors $(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n)$ where $\mathbf{m}_{\eta\eta} \to \overline{\mathbf{m}}_{\eta\eta}$ a.s., and $\overline{\mathbf{m}}_{\eta\eta}$ is a nonnegative definite matrix. Let $\mathbf{B}_n$ be a random $k \times 1$*

vector such that $\mathbf{B}_n = O_p(n^{-1/2})$ *and* a_n *a random variable such that* $a_n = o_p(n^{-1/2})$. *Define for* $t = 1, \dots, n$,

$$Z_t = (1 + a_n) m_{\xi\xi}^{-1/2}(\xi_t - \bar{\xi}) + (\eta_t - \bar{\eta})\mathbf{B}_n,$$
$$Z_t^* = m_{\xi\xi}^{-1/2}(\xi_t - \bar{\xi})$$

and define for $\omega \in [0,1]$

$$W_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \{1[\Phi(Z_t) \le \omega] - \omega\},$$
$$W_n^*(\omega) = n^{-1/2} \sum_{t=1}^{n} \{1[\Phi(Z_t^*) \le \omega] - \omega\}.$$

Then

$$\sup_{\omega \in [0,1]} |W_n^*(\omega) - W_n(\omega)| = o_p(1).$$

PROOF. We sketch the proof which is given in detail in Miller (1986a). For $\omega \in [0,1]$ we can write

$$W_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \Big\{1\Big[m_{\xi\xi}^{-1/2}(\xi_t - \bar{\xi}) \le (1 + c_n)\Phi^{-1}(\omega) + (\eta_t - \bar{\eta})\mathbf{D}_n\Big] - \omega\Big\} + o_p(1), \tag{A.2}$$

where $c_n \equiv 0$ and $\mathbf{D}_n \equiv \mathbf{0}$ when $a_n = -1$ and $c_n = -(1 + a_n)^{-1} a_n = o_p(n^{-1/2})$, $\mathbf{D}_n = -(1 + a_n)^{-1}\mathbf{B}_n = O_p(n^{-1/2})$ when $a_n \neq -1$. The remainder is $o_p(1)$ uniformly over $\omega \in [0,1]$ since $P\{(1 + a_n) \le 0\} \to 0$ as $n \to \infty$. After some algebra we can write

$$W_n(\omega) = W_n^*(\omega) + W_{1n}(\omega) + W_{2n}(\omega) - W_{3n}(\omega) + o_p(1), \tag{A.3}$$

where the remainder is $o_p(1)$ uniformly over $\omega \in [0,1]$ and the other terms are given by

$$W_{1n}(\omega) = n^{-1/2} \sum_{t=1}^{n} \{\Phi[z_{tn}(\omega)] - \Phi[y_n(\omega)]\},$$
$$W_{2n}(\omega) = n^{-1/2} \sum_{t=1}^{n} \big[1\{U_t \le \Phi[z_{tn}(\omega)]\} - 1\{U_t \le \omega\} - \Phi[z_{tn}(\omega)] + \omega\big],$$
$$W_{3n}(\omega) = n^{-1/2} \sum_{t=1}^{n} \big[1\{U_t \le \Phi[y_n(\omega)]\} - 1\{U_t \le \omega\} - \Phi[y_n(\omega)] + \omega\big], \tag{A.4}$$
$$y_n(\omega) = \sigma^{-1}(\bar{\xi} - \mu) + \left(\sigma^{-2} m_{\xi\xi}\right)^{1/2} \Phi^{-1}(\omega),$$
$$z_{tn}(\omega) = y_n(\omega) + \left[\Phi^{-1}(\omega) c_n + (\eta_t - \bar{\eta})\mathbf{D}_n\right]\left(\sigma^{-2} m_{\xi\xi}\right)^{1/2},$$
$$U_t = \Phi\left[\sigma^{-1}(\xi_t - \mu)\right].$$

We will next show that $\sup_{\omega \in [0,1]} |W_{in}(\omega)| = o_p(1)$ for $i = 1, 2, 3$ which will complete the proof.

By Taylor's theorem we can write

$$
\text{(A.5)}\qquad \begin{aligned} W_{1n}(\omega) &= n^{1/2}c_n\phi[y_n(\omega)]\Phi^{-1}(\omega)\big(\sigma^{-2}m_{\xi\xi}\big)^{1/2} \\ &\quad + n^{-1/2}\sum_{t=1}^{n}\tfrac{1}{2}\phi'[\gamma_{tn}^{*}(\omega)]\big\{\Phi^{-1}(\omega)c_n + (\eta_t - \bar{\eta})\mathbf{D}_n\big\}^2\big(\sigma^{-2}m_{\xi\xi}\big), \end{aligned}
$$

where $\gamma_{tn}^{*}(\omega)$ is a random variable on the line segment joining $z_{tn}(\omega)$ and $y_n(\omega)$. The result $\sup_{\omega\in[0,1]}|W_{1n}(\omega)| = o_p(1)$ follows after tedious algebra by repeatedly using the bound

$$
\text{(A.6)}\qquad \sup_{x\in\Re}|x^p\phi(a+bx)| < |b|^{-p}(|a|+4p)^p
$$

(for positive integer p and real numbers a and $b \neq 0$) and the fact that

$$
\text{(A.7)}\qquad \begin{aligned} n^{-1}\sum_{t=1}^{n}|(\eta_t - \bar{\eta})\mathbf{D}_n|^q &= O_p(n^{-1/2}) \quad \text{for } q = 1 \\ &= O_p(n^{-1}) \quad \text{for } q = 2, 3, \ldots . \end{aligned}
$$

The bound in (A.6) is from Miller (1986a) and can be derived by differentiating and maximizing the left-hand side.

We next examine $W_{2n}(\omega)$ and $W_{3n}(\omega)$. Note that $W_{3n}(\omega)$ is a special case of $W_{2n}(\omega)$ since $W_{2n}(\omega)$ reduces to $W_{3n}(\omega)$ if c_n and $\mathbf{D}_n$ are uniquely 0. Thus we are done if we can show that $\sup_{\omega\in[0,1]}|W_{2n}(\omega)| = o_p(1)$, since this would imply the same result for $W_{3n}(\omega)$. Our method of proof will involve only a slight modification to the approach of Rao and Sethuraman (1975). First define

$$
\text{(A.8)}\qquad z_{tn}(\omega;\zeta) = \zeta_1 + (1+\zeta_2)\Phi^{-1}(\omega) + (\eta_t - \bar{\eta})\zeta_3,
$$

where $\zeta = (\zeta_1, \zeta_2, \zeta_3')$ and $\zeta \in \Re^{k+2}$. By analogy also define $W_{2n}(\omega;\zeta)$ in terms of $z_{tn}(\omega;\zeta)$. Note that $z_{tn}(\omega) = z_{tn}(\omega;\hat{\zeta})$ and $W_{2n}(\omega) = W_{2n}(\omega;\hat{\zeta})$, where

$$
\text{(A.9)}\qquad \begin{aligned} \hat{\zeta}_1 &= \sigma^{-1}(\bar{\xi} - \mu), \\ \hat{\zeta}_2 &= (1+c_n)\big(\sigma^{-2}m_{\xi\xi}\big)^{1/2} - 1, \\ \hat{\zeta}_3 &= \mathbf{D}_n\big(\sigma^{-2}m_{\xi\xi}\big)^{1/2}. \end{aligned}
$$

For $\varepsilon > 0$ we can find an $L_\varepsilon > 0$ such that for all n, $P(\hat{\zeta} \in C_{ne}) > 1 - \varepsilon$, where $C_{n\varepsilon}$ is a hypercube in $\Re^{k+2}$ centered at the origin with sides of length $2L_\varepsilon n^{-1/2}$. Now note that

$$
\text{(A.10)}\qquad P\Big\{\sup_{\omega\in[0,1]}|W_{2n}(\omega)| > \varepsilon\Big\} < P\Big\{\sup_{\omega\in[0,1]}\sup_{\zeta\in C_{n\varepsilon}}|W_{2n}(\omega;\zeta)| > \varepsilon\Big\} + \varepsilon.
$$

We will be done if we can show that the right-hand side of (A.10) converges to ε as $n \to \infty$. Further, we are done if we can show that

$$
\text{(A.11)}\qquad P\Big[\sup_{\omega\in[0,1]}\sup_{\zeta\in C_{n\varepsilon}}|W_{2n}(\omega;\zeta)| > \varepsilon \,|\, \{\eta_t\}_{t=1}^{\infty}\Big] \to 0 \quad \text{as } n \to \infty
$$

for almost every sequence $\{\eta_t\}_{t=1}^{\infty}$. In fact the proof (A.11) only requires the almost sure convergence $\mathbf{m}_{\eta\eta} \to \overline{\mathbf{m}}_{\eta\eta}$ which was part of the statement of

Theorem A. Now (A.11) follows by exactly the same method used to prove Lemma 1 of Loynes (1980), so we do not repeat it here.

Acknowledgments. I would like to thank two referees and an Associate Editor for their helpful comments. I would also like to thank Wayne Fuller for suggesting the topic to me and for his help while conducting the research.

REFERENCES

AMEMIYA, Y. and FULLER, W. A. (1984). Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. *Ann. Statist.* **12** 497–509.

COX, D. R. and SNELL, E. J. (1968). A general definition of residuals (with discussion). *J. Roy. Statist. Soc. Ser. B* **30** 248–275.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

FULLER, W. A. (1987). *Measurement Error Models*. Wiley, New York.

GLESER, L. J. (1981). Estimation in a multivariate errors-in-variables regression model: Large sample results. *Ann. Statist.* **9** 24–44.

KAC, M., KIEFER, J. and WOLFOWITZ, J. (1955). On tests of normality and other tests of goodness-of-fit based on distance methods. *Ann. Math. Statist.* **26** 189–211.

LOYNES, R. M. (1980). The empirical distribution function of residuals from generalised regression. *Ann. Statist.* **8** 285–298.

MILLER, S. M. (1986a). The limiting behavior of residuals from measurement error regressions. Ph.D. dissertation, Iowa State Univ.

MILLER, S. M. (1986b). Tests for normality in errors-in-variables regression models. *Proc. Bus. Econ. Statist. Sec.* 567–571. Amer. Statist. Assoc., Washington.

MUGANTSEVA, L. A. (1977). Testing normality in one dimensional and multi-dimensional linear regression. *Theory Probab. Appl.* **22** 591–602.

NEUHAUS, G. (1976). Weak convergence under contiguous alternatives of the empirical process when parameters are estimated: The D_k approach. *Empirical Distributions and Processes. Lecture Notes in Math.* **566** 68–82. Springer, New York.

PIERCE, D. and KOPECKY, K. (1979). Testing goodness-of-fit for the distribution of errors in regression models. *Biometrika* **66** 1–6.

PORTNOY, S. (1986). Asymptotic behavior of the empiric distribution of M-estimated residuals from a regression model with many parameters. *Ann. Statist.* **14** 1152–1170.

RAO, J. S. and SETHURAMAN, J. (1975). Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. *Ann. Statist.* **3** 299–313.

SHORACK, G. R. (1984). Empirical and rank processes of observations and residuals. *Canad. J. Statist.* **12** 319–332.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

WOOD, C. L. (1981a). On tests for normality of experimental error in multiple regression. Technical Report 185, Dept. Statistics, Univ. Kentucky.

WOOD, C. L. (1981b). Goodness-of-fit for multivariate normality. Technical Report 182, Dept. Statistics, Univ. Kentucky.

WOOD, C. L. (1984). On test of normality of experimental error in ridge regression. *J. Statist. Plann. Inference* **9** 367–374.

OFFICE OF RESEARCH AND EVALUATION
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR
441 G STREET N.W.
WASHINGTON, D.C. 20212