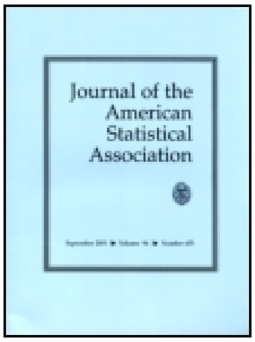

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Inferences for the Linear Errors-in-Variables with Changepoint Models

Yi-Ping Chang & Wen-Tao Huang

Inferences for the Linear Errors-in-Variables With Changepoint Models

Yi-Ping CHANG and Wen-Tao HUANG

A new linear structural errors-in-variables regression with changepoint model is considered. In this model we consider the likelihood ratio test based on the maximum Hotelling T^2 for the test of no change against the alternative of exactly one change. If there is a change, either known a priori or by testing, then we estimate the unknown changepoint parameter and some other related parameters by maximum likelihood. The limiting distribution of the changepoint estimator is investigated and it is shown that the asymptotic efficiency increases as the absolute regression slope coefficient increases. A Monte Carlo study shows that the proposed estimator performs satisfactorily.

KEY WORDS: Asymptotic distribution; Asymptotic efficiency; Consistency; Likelihood ratio test.

1. INTRODUCTION

The classical linear errors-in-variables regression (EIVR) model with one independent variable can be expressed by

$$Y_i = \alpha + \beta\xi_i + \varepsilon_i, \qquad (i = 1, \ldots, n)$$

and

$$X_i = \xi_i + \delta_i, \qquad (i = 1, \ldots, n), \tag{1}$$

where α and β are unknown constants. Here the ξ's are unobservable and we observe only pairs of values $(X_i, Y_i), i = 1, \ldots, n$. The errors of measurement $(\varepsilon_i, \delta_i)$ are assumed to be independent and identically distributed with finite second moments. If the ξ's are random, then model (1) is called a *structural EIVR model;* if the ξ's are deterministic, then the model is called a *functional EIVR model.* In this article we assume the structural EIVR model.

In the structural EIVR model, usually it is assumed that $\xi_i \sim N(\mu_\xi, \sigma_\xi^2), \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, and $\delta_i \sim N(0, \sigma_\delta^2)$ and that they are all independent. It is well known that this model lacks identifiability (Madansky 1959). The situation is widely discussed in the literature (see, e.g., Fuller 1987 and Moran 1971).

However, in many practical situations, the assumption that the ξ_i's are identical is violated. Consider an example in economics that can be stated as follows. Let ξ_i denote some family's true income at time i, let X_i denote the family's measured income, let Y_i denote its measured consumption, and let ε_i and δ_i denote measurement errors. During the observations of (X_i, Y_i), some new impact on the financial system in the society may occur—for instance, a new economic policy may be announced. The family's true income structure may start to change some time after the announcement; however, the relation between income and consumption remains unchanged. Under this situation, we may consider the structural EIVR model (1), but with the independent variables replaced by a changepoint model. Let $\{\xi_i, \varepsilon_i, \delta_i\}_{i=1}^n$ be a sequence of independent random vectors satisfying

$$\xi_i \sim N(\mu_1, \sigma_\xi^2) \qquad (i = 1, \ldots, [n\lambda]),$$

$$\xi_i \sim N(\mu_2, \sigma_\xi^2) \qquad (i = [n\lambda] + 1, \ldots, n),$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2) \qquad (i = 1, \ldots, n),$$

and

$$\delta_i \sim N(0, \sigma_\delta^2) \qquad (i = 1, \ldots, n), \tag{2}$$

where $\lambda, \mu_1, \mu_2, \sigma_\xi^2, \sigma_\varepsilon^2$, and σ_δ^2 are all unknown parameters. As can be seen, this model occurs when changepoints occur in a linear error-in-variables model. We call this model a *linear errors-in-variables regression with changepoint* model. In this article we focus on the situation that has at most one changepoint, i.e. EIVR model (1) with assumption (2).

This article provides an alternative approach to modeling the EIVR model. In Section 2 and 3, based on observations $\{(X_i, Y_i), i = 1, 2, \ldots, n\}$, we are interested in testing whether a change occurs in ξ_i and estimating the changepoint parameter λ, some related parameters α and β, and also other nuisance parameters $\mu_1, \mu_2, \sigma_\xi^2, \sigma_\varepsilon^2$, and σ_δ^2, if there is a change in ξ_i. In Section 4 we study the limiting distribution and asymptotic efficiency of the proposed changepoint estimator. We also give some Monte Carlo results. Finally, in Section 5 we analyze some real data on stock market sales volumes using the proposed method.

2. LIKELIHOOD RATIO TEST

In this section we discuss the likelihood ratio test for testing H_0: $\mu_1 = \mu_2 = \mu$ (no change) against H_1: $\mu_1 \neq \mu_2$, where μ is an unknown constant. Under the EIVR model (1) and assumption (2), it is easy to obtain the marginal distribution of (X_i, Y_i), which is given by $BN(\mu_1, \alpha + \beta\mu_1, \sigma_\delta^2 + \sigma_\xi^2, \sigma_\varepsilon^2 + \beta^2\sigma_\xi^2, \beta\sigma_\xi^2), i = 1, \ldots, [n\lambda]$ and $BN(\mu_2, \alpha + \beta\mu_2, \sigma_\delta^2 + \sigma_\xi^2, \sigma_\varepsilon^2 + \beta^2\sigma_\xi^2, \beta\sigma_\xi^2), i = [n\lambda] + 1, \ldots, n$, with (X_i, Y_i) independent, where $BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\sigma_1\sigma_2)$ de-

Yi-Ping Chang is Associate Professor, Department of Business Mathematics, Soochow University, Taipei, Taiwan, R.O.C. Wen-Tao Huang is a Research Fellow, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C. The authors are grateful to the associate editor and two referees for their helpful comments and valuable suggestions that improved the presentation of this article. This research was partially supported by grant NSC 83-0208-M-001-050 from the National Science Council of Taiwan, R.O.C.

Journal of the American Statistical Association
March 1997, Vol. 92, No. 437, Theory and Methods

notes the bivariate normal with means μ_1 and μ_2, variances σ_1^2, and σ_2^2, and correlation ρ.

We note that under H_0: $\mu_1 = \mu_2 = \mu$, this model is unidentifiable for the parameters $(\mu, \alpha, \beta, \sigma_\delta^2, \sigma_\varepsilon^2, \sigma_\xi^2)$; however, it is identifiable for the parameters $(\mu, \alpha + \beta\mu, \sigma_\delta^2 + \sigma_\xi^2, \sigma_\varepsilon^2 + \beta^2\sigma_\xi^2, \beta\sigma_\xi^2)$. Therefore, let $\mu_x = \mu, \mu_y = \alpha + \beta\mu, \sigma_x^2 = \sigma_\delta^2 + \sigma_\xi^2, \sigma_y^2 = \sigma_\varepsilon^2 + \beta^2\sigma_\xi^2$, and $\rho_{xy} = \beta\sigma_\xi^2/(\sigma_x\sigma_y)$, and consider $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy})$ as a reparameterization. This means that the original parameter space (our present model) is partitioned into equivalent class by the reparameterization. Each parameter vector in the equivalent class corresponds to some common distribution.

The likelihood function under H_0 in the reparameterized space (i.e., the classical model) is the same as that in the original parameter space. The maximum likelihood function stays the same for each parameter vector in the corresponding equivalent class. With the reparameterized space, the marginal distributions of $(X_i, Y_i), i = 1, \ldots, n$ are independent and commonly distributed according to $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho\sigma_x\sigma_y)$, and the likelihood function can be obtained. Under H_1, the EIVR with changepoint model is identifiable for the parameters $(\mu, \alpha, \beta, \sigma_\delta^2, \sigma_\varepsilon^2, \sigma_\xi^2)$. Thus it is a classical changepoint in mean vectors for a bivariate normal model while the covariance remains unchanged. Accordingly, the likelihood ratio test is not affected by this reparameterization under H_0, and the present model is a classical bivariate normal with changepoint model under H_1. Srivastava and Worsley (1986) dealt with such testing problems; James, James and Siegmund (1992) considered the likelihood ratio test. Following Srivastava and Worsley (1986), the likelihood ratio statistic can be defined as follows. Denote the pooled sample variance matrix by

$$\mathbf{W}_\lambda = \frac{1}{n-2}\left\{\sum_{i=1}^{[n\lambda]}\begin{bmatrix} X_i - \bar{X}(0,[n\lambda]) \\ Y_i - \bar{Y}(0,[n\lambda]) \end{bmatrix} \times \begin{bmatrix} X_i - \bar{X}(0,[n\lambda]) \\ Y_i - \bar{Y}(0,[n\lambda]) \end{bmatrix}' + \sum_{i=[n\lambda]+1}^{n}\begin{bmatrix} X_i - \bar{X}([n\lambda],n) \\ Y_i - \bar{Y}([n\lambda],n) \end{bmatrix} \times \begin{bmatrix} X_i - \bar{X}([n\lambda],n) \\ Y_i - \bar{Y}([n\lambda],n) \end{bmatrix}'\right\},$$

where $\bar{X}(a,b) = 1/(b-a)\sum_{i=a+1}^{b} X_i$. The standardized difference vector between the observations before and after the changepoint is defined by

$$\mathbf{Z}_\lambda = \left\{\frac{[n\lambda](n-[n\lambda])}{n}\right\}^{1/2} \times \left\{\begin{bmatrix} \bar{X}(0,[n\lambda]) \\ \bar{Y}(0,[n\lambda]) \end{bmatrix} - \begin{bmatrix} \bar{X}([n\lambda],n) \\ \bar{Y}([n\lambda],n) \end{bmatrix}\right\},$$

and the Hotelling T^2 for testing this difference is

$$T_\lambda^2 = \mathbf{Z}_\lambda' \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda.$$

The likelihood ratio test for unknown λ can be based on the maximum Hotelling T^2,

$$T_{\hat{\lambda}}^2 = \sup_{0<\lambda<1} \mathbf{Z}_\lambda' \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda;$$

a large value of $T_{\hat{\lambda}}^2$ indicates that it is in favor of H_1. Srivastava and Worsley (1986) derived a conservative approximation for the null distribution of testing statistics based on an improved Bonferroni inequality; James et al. (1992) also obtained a tail approximation for the significance level of the modified likelihood ratio test.

3. ESTIMATORS FOR THE PARAMETERS IF A CHANGE IS INDICATED

If a change is indicated by the test, or if it is known a priori that a change has indeed taken place, then the unknown changepoint parameter λ can be estimated by $\hat{\lambda} = j/n$, where j maximizes

$$\mathbf{Z}_{j/n}' \mathbf{W}_{j/n}^{-1} \mathbf{Z}_{j/n},$$

in the range $n_0 \le j \le n - n_0$, where $1 \le n_0 \le n/2$. The other parameters can be estimated as follows.

For convenience, we denote the true value of λ by λ_0. If λ_0 is known, then the model can be seen as a special EIVR with two groups model. For this model, Wald (1940) gave a consistent estimator for the slope β. For two or more groups, Richardson and Wu (1970) proposed an estimator of the slope β under the condition that each group have equal sample size, and Cox (1976) considered the testing and estimation problems by using the maximum likelihood method. Hence when there are only two groups, for estimating β, the estimators of Richardson and Wu (1970) reduced to Wald's result and Cox (1976) showed the estimator to be $(\bar{Y}(0,[n\lambda_0]) - \bar{Y}([n\lambda_0],n))/(\bar{X}(0,[n\lambda_0]) - \bar{X}([n\lambda_0],n))$. This is in fact the slope of the line joining two means of the two groups.

When λ_0 is known, the maximum likelihood estimators (MLE's) for the other parameters are given by

$$\hat{\mu}_1(\lambda_0) = \bar{X}(0,[n\lambda_0]), \qquad \hat{\mu}_2(\lambda_0) = \bar{X}([n\lambda_0],n),$$

$$\hat{\beta}(\lambda_0) = \frac{\bar{Y}(0,[n\lambda_0]) - \bar{Y}([n\lambda_0],n)}{\bar{X}(0,[n\lambda_0]) - \bar{X}([n\lambda_0],n)},$$

$$\begin{aligned}\hat{\alpha}(\lambda_0) &= \frac{1}{2}\left[\{\bar{Y}(0,[n\lambda_0]) + \bar{Y}([n\lambda_0],n)\} - \hat{\beta}(\lambda_0) \times \{\bar{X}(0,[n\lambda_0]) + \bar{X}([n\lambda_0],n)\}\right] \\ &= \bar{Y}(0,n) - \hat{\beta}(\lambda_0)\bar{X}(0,n) \\ &= \frac{\bar{Y}([n\lambda_0],n)\bar{X}(0,[n\lambda_0]) - \bar{Y}(0,[n\lambda_0])\bar{X}([n\lambda_0],n)}{\bar{X}(0,[n\lambda_0]) - \bar{X}([n\lambda_0],n)},\end{aligned}$$

$$\hat{\sigma}_\xi^2(\lambda_0) = \left[\frac{1}{n\hat{\beta}(\lambda_0)}\{S_{XY}(0,[n\lambda_0]) + S_{XY}([n\lambda_0],n)\}\right] \vee 0,$$

$$\hat{\sigma}_\varepsilon^2(\lambda_0) = \left[\frac{1}{n}\{S_{YY}(0,[n\lambda_0]) + S_{YY}([n\lambda_0],n)\} - \hat{\beta}^2(\lambda_0)\hat{\sigma}_\xi^2(\lambda_0)\right] \vee 0,$$

and

$$\hat{\sigma}_\delta^2(\lambda_0) = \left[\frac{1}{n}\{S_{XX}(0,[n\lambda_0]) + S_{XX}([n\lambda_0],n)\} - \hat{\sigma}_\xi^2(\lambda_0)\right] \vee 0, \quad (3)$$

where $S_{XY}(a,b) = \sum_{i=a+1}^{b}\{X_i - \bar{X}(a,b)\}\{Y_i - \bar{Y}(a,b)\}$ and $x \vee y = \max(x,y)$.

For any value of λ, let $\{(X_i,Y_i), i = 1,\ldots,[n\lambda]\}$ be the first group and let $\{(X_i,Y_i), i = [n\lambda]+1,\ldots,n\}$ be the second group. Consider the slope of the line joining the point $(\bar{X}(0,[n\lambda]), \bar{Y}(0,[n\lambda]))$ and the point $(\bar{X}([n\lambda],n), \bar{Y}([n\lambda],n))$ as our estimator for β. We define a class of estimators of β as

$$\mathcal{W} = \left\{\hat{\beta}(\lambda)\colon \hat{\beta}(\lambda) = \frac{\bar{Y}(0,[n\lambda]) - \bar{Y}([n\lambda],n)}{\bar{X}(0,[n\lambda]) - \bar{X}([n\lambda],n)}, 0 < \lambda < 1\right\}.$$

It is quite interesting to find in the following that each estimator in $\mathcal{W}$ is strongly consistent for β even if the partition of the data into two groups by λ is incorrect.

Theorem 3.1. For any $0 < \lambda < 1$ and $\hat{\beta}(\lambda) \in \mathcal{W}$,

a. $\beta(\lambda) \to \beta$ with probability 1 as $n \to \infty$;

b. $\sqrt{n}\{\hat{\beta}(\lambda) - \beta\} \xrightarrow{\mathcal{L}} N(0, \sigma^2(\lambda))$, where

$$\sigma^2(\lambda) = \frac{\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2}{\lambda(1-\lambda)(\mu_1-\mu_2)^2} \times \left\{\left(\frac{1-\lambda}{1-\lambda_0}\right)^2 I(\lambda \le \lambda_0) + \left(\frac{\lambda}{\lambda_0}\right)^2 I(\lambda > \lambda_0)\right\};$$

and

c. $\sigma^2(\lambda_0) < \sigma^2(\lambda)$ for all $\lambda \neq \lambda_0$.

Proof. See the Appendix.

Remarks.

1. Note that the consistency of $\hat{\beta}$ by (a) is also guaranteed under some general regularity conditions without assuming normality of X_i and Y_i. When λ_0 is known, the minimum asymptotic variance in this class is equal to the asymptotic variance of MLE under normality conditions, which is generally the minimum in the class of all estimators.

2. Because the asymptotic variance of $\hat{\beta}(\lambda)$ is independent of σ_ξ^2, this theorem also holds when the true model is a functional EIVR with changepoint model.

3. If $\lambda, \beta, \sigma_\varepsilon^2$, and σ_δ^2 are all fixed, then the asymptotic variance of $\hat{\beta}(\lambda)$ is decreasing in $|\mu_1 - \mu_2|$.

When λ_0 is unknown, we propose estimators for the parameters $\mu_1, \mu_2, \alpha, \beta, \sigma_\xi^2, \sigma_\varepsilon^2$, and σ_δ^2 to be plug-in estimators of (3); that is,

$$\hat{\mu}_1 = \hat{\mu}_1(\hat{\lambda}), \quad \hat{\mu}_2 = \hat{\mu}_2(\hat{\lambda}), \quad \hat{\alpha} = \hat{\alpha}(\hat{\lambda}),$$
$$\hat{\beta} = \hat{\beta}(\hat{\lambda}), \quad \hat{\sigma}_\xi^2 = \hat{\sigma}_\xi^2(\hat{\lambda}), \quad \hat{\sigma}_\varepsilon^2 = \hat{\sigma}_\varepsilon^2(\hat{\lambda}),$$

and

$$\hat{\sigma}_\delta^2 = \hat{\sigma}_\delta^2(\hat{\lambda}).$$

We investigate some properties of the proposed estimators in next section.

4. THE LIMITING DISTRIBUTION AND ASYMPTOTIC EFFICIENCY

4.1 The Limiting Distribution

In this section we study the limiting distributions of the proposed estimators when the sample size increases to infinity. The limiting distribution can be used to construct approximate confidence interval for the changepoint. If $|\mu_1 - \mu_2|$ is large (relative to the variance) and σ_ξ^2 is moderate, then detection of change should be easier and estimation of the changepoint should be more precise. Thus in application it might be more desirable to consider the situation of small changes. Furthermore, a confidence interval for the changepoint parameter based on the limiting distribution for small $|\mu_1 - \mu_2|$ is expected to cover the corresponding interval when $|\mu_1 - \mu_2|$ is actually large. Now suppose that $|\mu_2 - \mu_1|$ depends on n and converges to zero at a rate v_n. The following theorem describes the asymptotic distribution of the changepoint estimator under local shifts in mean.

Theorem 4.1. Under the conditions $n_0 = n^{1/2}v_n, \mu_1 - \mu_2 = Cv_n \to 0$, and $nv_n^2 \to \infty$, as $n \to \infty$,

$$nv_n^2(\hat{\lambda} - \lambda_0) \xrightarrow{\mathcal{L}} C^{-2}\sigma^2 T, \quad (4)$$

where $\sigma^2 = (\sigma_\delta^2\sigma_\xi^2 + \sigma_\xi^2\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2\sigma_\xi^2)/(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2), C$ is a real number, and T takes value t at which $B(t) + |t|/2$ attains its unique minimum with probability 1. Here $B(t)$ is a two-sided Brownian motion.

Proof. See the Appendix.

To find the confidence intervals, it would be useful to have the distribution of T. Bhattacharya and Brockwell (1976) showed that the probability density function $g(t)$ of T is symmetric about 0 and

$$g(t) = \frac{3}{2}\exp(t)\left\{1 - \Phi\left(\frac{3}{2}\sqrt{t}\right)\right\} - \frac{1}{2}\left\{1 - \Phi\left(\frac{1}{2}\sqrt{t}\right)\right\}, \quad t > 0,$$

where Φ is the standard normal cumulative distribution function.

Next, we state the limiting distribution of the estimators for the parameters α and β, which are also in our main concern.

Theorem 4.2. Under the conditions of Theorem 4.1, as $n \to \infty$, the random vector

$$\sqrt{n} v_n(\hat{\alpha} - \alpha, \hat{\beta} - \beta)' \xrightarrow{\mathcal{L}} BN\left(0, 0, \frac{\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2}{C^2}\left(\frac{\mu_1^2}{\lambda_0} + \frac{\mu_2^2}{1-\lambda_0}\right), \frac{\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2}{C^2\lambda_0(1-\lambda_0)}, 0\right).$$

Proof. See the Appendix.

To construct the approximate confidence intervals for λ_0, α, and β, we need the following theorem.

Theorem 4.3. Under the conditions of Theorem 4.1, $\hat{\alpha} \to \alpha, \hat{\beta} \to \beta, \hat{\mu}_1 \to \mu_1, \hat{\mu}_2 \to \mu_2, \hat{\sigma}_\xi^2 \to \sigma_\xi^2, \hat{\sigma}_\varepsilon^2 \to \sigma_\varepsilon^2$, and $\hat{\sigma}_\delta^2 \to \sigma_\delta^2$, where all converge in probability as $n \to \infty$.

Proof. See the Appendix.

4.2 Efficiency Using the Additional Information of Y_i

If we do not have information for the concomitant variables Y_i, then we have a mean change problem, which has been discussed by, for example, Sen and Srivastava (1975) and Worsley (1979) for the detection of change. If additional observations Y_i are obtained, how can we utilize this information for estimation of the changepoint? Information concerning the changepoint is implicitly contained in Y_i through the relationship (1). This could significantly improve the estimation over that based only on observations X_i. We thus consider the asymptotic gain due to the additional information of Y_i. If we omit the information of Y_i, then the MLE of λ_0 is given by $\hat{\lambda}^* = j/n$, where j maximizes

$$\frac{j(n-j)}{n}\left(\frac{\sum_{i=1}^{j} X_i}{j} - \frac{\sum_{i=j+1}^{n} X_i}{n-j}\right)^2$$

in the range $n_0 \le j \le n - n_0$. Then (by Bhattacharya and Brockwell 1976), the limiting distribution of $\hat{\lambda}^*$ can be described by the following.

Theorem 4.4. Under the conditions of Theorem 4.1, as $n \to \infty$,

$$n v_n^2(\hat{\lambda}^* - \lambda_0) \xrightarrow{\mathcal{L}} C^{-2}(\sigma^*)^2 T,$$

where $(\sigma^*)^2 = \sigma_\delta^2 + \sigma_\xi^2$.

Yao (1987) also independently obtained this result for a sequence of independent random variables, and his numerical results showed that the limiting distribution provided a good approximation in the normal case.

Intuitively, it is clear that the information concerning the changepoint parameter in (X_i, Y_i) is better than that of X_i alone. To confirm this point and quantify the gain due to additional information of Y_i, we consider some relative efficiency based on the sample sizes.

Under local shifts in mean, and considering $\mu_1 - \mu_2 = Cv_n \to 0$ as $n \to \infty$, let $n' = n'(n)$ denote the sample size needed for the estimator $\hat{\lambda}^*$ to attain the same asymptotic variance of $\hat{\lambda}$. Then a plausible definition of asymptotic relative efficiency (ARE) of $\hat{\lambda}^*$ with respect to $\hat{\lambda}$ which utilizing the additional information of Y_i is $\text{eff}(\hat{\lambda}^*, \hat{\lambda}) = \lim_{n\to\infty} n'/n$. When there is no confusion, we denote $\text{eff}(\hat{\lambda}^*, \hat{\lambda})$ by eff for simplicity. Then, it follows from Theorem 4.1 and Theorem 4.4 that we have

$$\text{eff} = \lim_{n\to\infty} \frac{n'}{n} = \frac{\text{sd}\{(\sigma^*)^2 T\}}{\text{sd}\{\sigma^2 T\}} = \frac{(\sigma^*)^2}{\sigma^2}. \tag{5}$$

It is easy to see that $1 \le \text{eff} < \infty$. In the following the behavior of eff is more clearly described. By Theorem 4.1 and Theorem 4.4, we have the following.

Theorem 4.5. Under the conditions of Theorem 4.1,

a. $\text{eff} = 1 + \beta^2\sigma_\delta^4/(\sigma_\delta^2\sigma_\varepsilon^2 + \sigma_\xi^2\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2\sigma_\xi^2)$, which increases in $|\beta|$;
b. $1 \le \text{eff} \le 1 + \sigma_\delta^2/\sigma_\xi^2$;
c. eff depends on $\sigma_\xi^2, \sigma_\varepsilon^2$, and σ_δ^2 only through $\sigma_\xi^2/\sigma_\delta^2$ and $\sigma_\varepsilon^2/\sigma_\delta^2$; and
d. if $\sigma_\xi^2 = \sigma_\varepsilon^2 = \sigma_\delta^2$, then $\text{eff} = 1 + \beta^2/(2+\beta^2)$.

Note that the ARE eff $= 1$ if and only if $\beta = 0$ or $\sigma_\delta^2 = 0$. When $\beta = 0$, it is clear that any information concerning the changepoint in X is not contained in Y. When $\sigma_\delta^2 = 0$, there is no measurement error in X, and in this situation it is quite interesting to find that the changepoint estimator gains nothing at all by taking additional information from Y, because $\hat{\lambda}^*$ and $\hat{\lambda}$ are both asymptotically unbiased and attain same asymptotic variance. Intuitively, this is because all information concerning the changepoint in Y is completely included in that of X.

On the other hand, from (a) and (b) of Theorem 4.5, with other parameters being fixed, the loss of efficiency incurred by using only X_i instead of both X_i and Y_i is increasing in $|\beta|$ and is bounded by $\sigma_\delta^2/(\sigma_\xi^2 + \sigma_\delta^2)$. This accords with the intuition that the correlation between X_i and Y_i is $\beta\sigma_\xi^2/\{(\sigma_\delta^2 + \delta_\xi^2)(\sigma_\varepsilon^2 + \beta^2\sigma_\xi^2)\}^{1/2}$, which is increasing in $|\beta|$. Thus $\hat{\lambda}$ turns to be more efficient than $\hat{\lambda}^*$ when the correlation of X_i and Y_i becomes larger. The ARE eff, plotted in Figure 1, is a function of $|\beta|$ when $\sigma_\xi^2 = \sigma_\varepsilon^2 = \sigma_\delta^2$. Note that in this case, eff increases to the limit 2. This indicates that under the structural EIVR with changepoint model, the loss of efficiency incurred by using only X_i instead of both X_i and Y_i is increasing in $|\beta|$ but is at most 50%.

Table 1 gives some results of a simulation study performed to examine the efficiency of $\hat{\lambda}$ for the changepoint parameter. The entries tabulate the bias and root mean squared error (RMSE) of $\hat{\lambda}$ and $\hat{\lambda}^*$ and the relative RMSE of $\hat{\lambda}$ with respect to $\hat{\lambda}^*$, which is defined by $\text{RMSE}(\hat{\lambda}^*)/\text{RMSE}(\hat{\lambda})$. For each simulation, 10,000 replications of sample sizes $n = 50, 100$, or 200 were generated using the IMSL (International Mathematical and Statistical Libraries 1991) routine RNNOA with true underlying parameters $\lambda = .4, \mu_1 = 0, \mu_2 = 1$ or $2, \alpha = 0, \beta = 1$ or 2, and $\sigma_\xi^2 = \sigma_\varepsilon^2 = \sigma_\delta^2 = 1$.

Table 1 shows that when $\beta = 1$ or 2, the estimator $\hat{\lambda}$ behaves much better than $\hat{\lambda}^*$ in the sense of bias and RMSE.

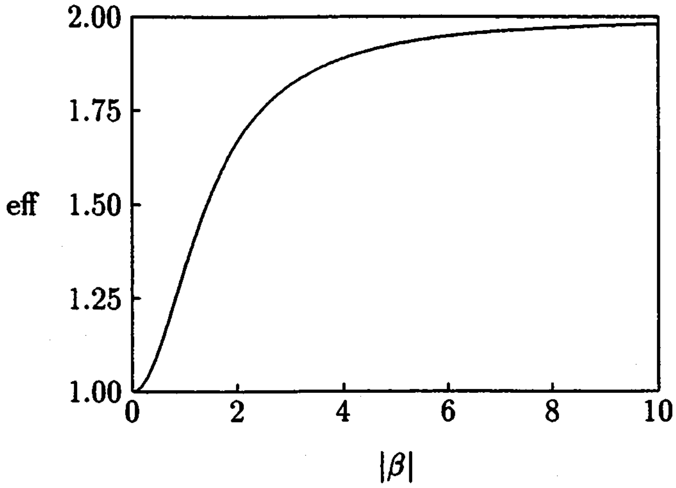

Figure 1. Asymptotic Relative Efficiency of the Estimator by Using the Additional Information of Y_i as a Function of $|\beta|$ With $\sigma_\xi^2 = \sigma_\varepsilon^2 = \sigma_\delta^2$.

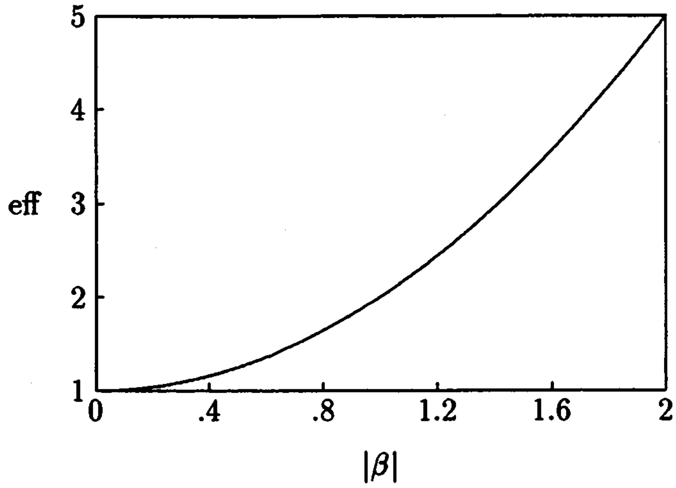

Figure 2. Asymptotic Relative Efficiency of the Estimator by Using the Additional Information of Y_i as a Function of $|\beta|$ With $\sigma_\varepsilon^2 = \sigma_\delta^2$ and $\sigma_\xi^2 = 0$.

The relative RMSE of $\hat{\lambda}$ with respect to $\hat{\lambda}^*$ increases as $|\beta|$ increases, which agrees with Theorem 4.5.

If the true model is a functional EIVR with changepoint model (i.e., $\sigma_\xi^2 = 0$ in model (1) with (2)), then the ARE eff $= 1 + \beta^2\sigma_\delta^2/\sigma_\varepsilon^2$, which is also an increasing function of $|\beta|$ and eff $= 1$, if and only if $\beta = 0$ or $\sigma_\delta^2 = 0$. In this case it is obvious that $1 \le \text{eff} < \infty$. The ARE eff is also plotted in Figure 2 as a function of $|\beta|$ when $\sigma_\varepsilon^2 = \sigma_\delta^2$.

5. EXAMPLES

5.1 Stock Market Sales Volumes

It was reported in the January 3, 1970 issue of *Business Week* that regional stock exchanges were hurt by abolition of give-ups (commission splitting) on December 5, 1968. The relationship between the monthly dollar volume of sales on the Boston Stock Exchange (Y) and the combined monthly dollar volumes for the New York and American Stock Exchanges (X) can be analyzed. McGee and Carleton (1970) provided the data and analyzed it using the piecewise regression method. Holbert (1982) also analyzed the data from the Bayesian viewpoint. The results of the structural EIVR with changepoint model are shown in Table 2. The approximate p value is evaluated by using the improved Bonferroni upper bound of Srivastava and Worsley (1986). The estimated changepoint is the ninth observation (September 1967) with the maximum Hotelling T^2 statistic $T_{9/25}^2 = 9.1656$, which is significant at level $\alpha = .05$. For this level, we may find further changes in a sequence that is split into two subsequences. The subsequence January 1967–September 1967 has no significant change (p value = .3599); however, the remaining subsequence (October 1967–November 1969) has a significant split after November 1968 with the maximum $T^2 = 10.8985$ and p value = .0149. Furthermore, the subsequence October 1967–November 1968 has a significant split after July 1968 with the maximum $T^2 = 10.3473$ and p value = .0413; however, the subsequences December 1968–November 1969 and October 1967–July 1968 contain no further significant

Table 1. Bias and RMSE of $\hat{\lambda}$ and $\hat{\lambda}^*$ in the Structural EIVR With Changepoint Model

			$\hat{\lambda}$		$\hat{\lambda}^*$		Relative
β	μ_2	n	Bias	RMSE	Bias	RMSE	RMSE
1.0	1.0	50	.0380	.1950	.0403	.2007	1.0293
		100	.0136	.1166	.0179	.1337	1.1465
		200	.0027	.0523	.0046	.0675	1.2915
	2.0	50	.0029	.0534	.0043	.0653	1.2224
		100	.0004	.0211	.0008	.0286	1.3548
		200	.0001	.0098	.0002	.0132	1.3540
2.0	1.0	50	.0287	.1694	.0405	.2014	1.1891
		100	.0080	.0897	.0158	.1257	1.4011
		200	.0015	.0390	.0046	.0678	1.7385
	2.0	50	.0015	.0389	.0044	.0664	1.7093
		100	.0003	.0173	.0009	.0307	1.7726
		200	.0001	.0078	.0002	.0138	1.7730

Table 2. Monthly Dollar Volume of Sales (in Millions) on the Boston Stock Exchange and Combined New York and American Stock Exchanges

Sequence	Changepoint	$\sup_\lambda T_\lambda^2$	p value	Estimates of other parameters
1–35	9	9.1656*	.0210	$\hat{\mu}_1 = 12002.03$, $\hat{\mu}_2 = 14122.18$, $\hat{\alpha} = -233.86$, $\hat{\beta} = .026$
1–9	2	3.9620	.3599	
10–35	23	10.8985*	.0149	$\hat{\mu}_1 = 14482.86$, $\hat{\mu}_2 = 13701.38$, $\hat{\alpha} = -953.50$, $\hat{\beta} = .077$
10–23	19	10.3473*	.0413	$\hat{\mu}_1 = 14611.60$, $\hat{\mu}_2 = 14161.02$, $\hat{\alpha} = 2349.37$, $\hat{\beta} = -1.151$
24–35	33	8.0140	.0926	
10–19	15	5.2972	.2294	

* Significant at level $\alpha = .05$.

Table 3. Changepoint for the Imports Data

Sequence	Changepoint	$\sup_\lambda T_\lambda^2$	p value	Estimates of other parameters
1–18	14	13.7726*	.0113	$\hat{\mu}_1 = 211.19$, $\hat{\mu}_2 = 329.65$, $\hat{\alpha} = -20.76$, $\hat{\beta} = .214$
1–14	9	9.3914	.0521	

* Significant at level $\alpha = .05$.

changes. Accordingly, we can conclude that three changes take place, one after September 1967, another one after June 1968, and the last one after November 1968. McGee and Carleton (1970) concluded that changes occurred at the 10th, 19th, and 23rd observations. Our analysis of changes almost agrees with that of McGee and Carleton (1970) except for the first change at the ninth observation. However, this difference is quite small.

5.2 Imports and Domestic Production Data

The second example is based on aggregate data concerning import activity in the French economy. Malinvaud (1968) provided the data of imports (Y), gross domestic product (X), and other variables in France, all measured in billions of French francs from 1949–1966. Chatterjee and Price (1991) analyzed this data by the principal component method, which essentially followed that of Malinvaud (1968). Chatterjee and Price (1991) found two patterns in the data: the residuals declining until 1960 and then rising. They argued that the models before and after 1960 should be different due to the fact that the European Common Market began operations in 1960. So we believe that some changepoint may exist in the data. Maddala (1992) considered a functional EIVR model; however, he ignored the possibility that some changes in the data may arise, as Chatterjee and Price (1991) pointed out. In this example we assume the structural EIVR with changepoint model; the results are shown in Table 3. The only changepoint is estimated to be at the 14th observation (1962) with the maximum Hotelling T^2 statistic $T^2_{14/18} = 13.7726$, which is significant at level $\alpha = .05$. Furthermore, at the same level, the subsequence 1949–1962 contains no further significant changes. Thus we can conclude that only one change occurs in 1962.

Comparing our results to the estimates of Maddala (1992), we see that they are not so close, mainly because Maddala (1992) considered a functional EIVR that is in fact doubtful, because $|\hat{\mu}_1 - \hat{\mu}_2|$ is large.

6. DISCUSSION

We have investigated the asymptotic efficiency of the changepoint estimator proposed in Section 3 for the structural EIVR with changepoint model. If $\sigma_\xi^2 = \sigma_\varepsilon^2 = \sigma_\delta^2$, then the loss of asymptotic efficiency incurred by using only information of X_i instead of both X_i and Y_i could be as much as 50%, too great to be ignored.

In Examples 5.1 and 5.2 we considered possible multiple changepoints, and tested them sequentially. Note, however, that it is possible to have a wonderful fit with, say, four segments, but for the three-segment fit to not be statistically better than the two-segment fit. Moreover, the optimal breakpoints for the four-segment fit are not necessarily a superset of those for the three-segment fit. Further research into this problem is needed.

It needs to be mentioned that we have not extended the problem to the situation of segmented regression. This kind of model covers many important phenomena, and further research into this topic would be worthwhile. It may be noted that under the situation of segmented regression with errors in variables, the problem would be reduced to a bivariate normal model with changes in both mean vector and covariance matrix, which is more complicated than the model in this study. The changepoint model involving changes in both mean and covariance matrix in multivariate normal has not yet been investigated and is worth studying. However, when the errors-in-variables aspect is not taken into consideration, the pure segmented regression problem has been considered in the literature (see, e.g., Hinkley 1969, Kim and Siegmund 1989, Quandt 1958, and Van de Geer 1988).

APPENDIX: PROOFS

Proof of Theorem 3.1

To verify (a), note that with probability 1 for each $\lambda \le \lambda_0$,

$$\hat{\beta}(\lambda) \to \left\{ (\alpha+\beta\mu_1) - \frac{(1-\lambda_0)(\alpha+\beta\mu_2) + (\lambda_0-\lambda)(\alpha+\beta\mu_1)}{1-\lambda} \right\} \div \left\{ \mu_1 - \frac{(1-\lambda_0)\mu_2 + (\lambda_0-\lambda)\mu_1}{1-\lambda} \right\} = \beta,$$

and with probability 1 for each $\lambda > \lambda_0$,

$$\hat{\beta}(\lambda) \to \left\{ \frac{\lambda_0(\alpha+\beta\mu_1) + (\lambda-\lambda_0)(\alpha+\beta\mu_2)}{\lambda} - (\alpha+\beta\mu_1) \right\} \div \left\{ \frac{\lambda_0\mu_1 + (\lambda-\lambda_0)\mu_2}{\lambda} - \mu_2 \right\} = \beta.$$

Note that Wald (1940) has shown that under fairly general conditions, the Wald estimators are consistent in the EIVR with two groups model.

To verify (b), for convenience, let $\bar{X}^*(a,b) = 1/(b-a)\sum_{i=a+1}^{b}(X_i - EX_i)$ and $\bar{Y}^*(a,b) = 1/(b-a)\sum_{i=a+1}^{b}(Y_i - EY_i)$. By the multivariate central limit theorem, we have the random vector

$$\sqrt{n}(\bar{X}^*(0,[n\lambda]), \bar{X}^*([n\lambda],n), \bar{Y}^*(0,[n\lambda]), \bar{Y}^*([n\lambda],n))' \xrightarrow{\mathcal{L}} \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma}(\lambda)),$$

where

$$\boldsymbol{\Sigma}(\lambda) = \begin{pmatrix} \frac{1}{\lambda}\,(\sigma_\delta^2+\sigma_\xi^2) & 0 & \frac{1}{\lambda}\,\beta\sigma_\xi^2 & 0 \\ 0 & \frac{1}{1-\lambda}\,(\sigma_\delta^2+\sigma_\xi^2) & 0 & \frac{1}{1-\lambda}\,\beta\sigma_\xi^2 \\ \frac{1}{\lambda}\,\beta\sigma_\xi^2 & 0 & \frac{1}{\lambda}\,(\sigma_\varepsilon^2+\beta^2\sigma_\xi^2) & 0 \\ 0 & \frac{1}{1-\lambda}\,\beta\sigma_\xi^2 & 0 & \frac{1}{1-\lambda}\,(\sigma_\varepsilon^2+\beta^2\sigma_\xi^2) \end{pmatrix}. \tag{A.1}$$

Rewrite

$$\sqrt{n}(\hat{\beta}(\lambda)-\beta)$$

$$= \frac{\sqrt{n}\{\bar{Y}^*(0,[n\lambda]) - \bar{Y}^*([n\lambda],n) - \beta\bar{X}^*(0,[n\lambda]) + \beta\bar{X}^*([n\lambda],n)\}}{\bar{X}(0,[n\lambda]) - \bar{X}([n\lambda],n)}. \tag{A.2}$$

By the δ method (see, e.g., Serfling 1980, p. 122), the numerator of (A.2) converges in distribution to $N(0,\{(1/\lambda) + [1/(1-\lambda)]\}(\sigma_\varepsilon^2+\beta^2\sigma_\delta^2))$, and the denominator of (A.2) converges in probability to $\{(1-\lambda_0)/(1-\lambda)I(\lambda \le \lambda_0) + (\lambda_0/\lambda)I(\lambda > \lambda_0)\}(\mu_1-\mu_2)$. Furthermore, by Slutsky's theorem, (b) is proved.

By a straightforward calculation, (c) can also be verified.

Proof of Theorem 4.1

Srivastava and Worsley (1986) examined

$$S_\lambda = T_\lambda^2/(n-2+T_\lambda^2) = \mathbf{Z}_\lambda'\mathbf{V}\mathbf{Z}_\lambda,$$

where

$$\mathbf{V} = \sum_{i=1}^{n}\begin{bmatrix}\mathbf{X}_i - \bar{X}(0,n)\\ \mathbf{Y}_i - \bar{Y}(0,n)\end{bmatrix}\begin{bmatrix}\mathbf{X}_i - \bar{X}(0,n)\\ \mathbf{Y}_i - \bar{Y}(0,n)\end{bmatrix}'.$$

Note that each of the first $[n\lambda_0]$ independent random vectors $(\mathbf{X}_i,\mathbf{Y}_i)', 1 \le i \le n$, has mean vector $(\mu_1, \alpha+\beta\mu_1)'$ and covariance matrix $\mathbf{H}$, whereas each of the last $n-[n\lambda_0]$ has mean vector $(\mu_1,\alpha+\beta\mu_1)' + v_n^{-1}\mathbf{d}$ and covariance matrix $\mathbf{H}$, where $\mathbf{d} = C(1,\beta)'$ and

$$\mathbf{H} = \begin{bmatrix}\sigma_\delta^2+\sigma_\xi^2 & \beta\sigma_\varepsilon^2\\ \beta\sigma_\varepsilon^2 & \sigma_\varepsilon^2+\beta^2\sigma_\xi^2\end{bmatrix}.$$

Accordingly, λ_0 is a changepoint for the sequence $(X_i,Y_i)'$. Considering $(X_i,Y_i)'$ as the sequence $Y_i(\phi)$ of Bhattacharya (1987), because S_λ is increasing in T_λ^2, the estimator $\hat{\lambda}$ can be seen as a proposed estimator of Bhattacharya (1987). From the results of Bhattacharya (1987),

$$nv_n^2(\hat{\lambda}-\lambda_0) \xrightarrow{\mathcal{L}} T_{(\mathbf{d}'\mathbf{H}^{-1}\mathbf{d})^{1/2}}. \tag{A.3}$$

By a simple calculation,

$$\mathbf{d}'\mathbf{H}^{-1}\mathbf{d} = \frac{C^2(\sigma_\varepsilon^2+\beta^2\sigma_\delta^2)}{\sigma_\delta^2\sigma_\varepsilon^2+\sigma_\xi^2\sigma_\varepsilon^2+\beta^2\sigma_\delta^2\sigma_\xi^2}.$$

Here, for any $\mu > 0, T_u$ takes value t at which $B(t) + u|t|/2$ attains its unique minimum with probability 1, and the probability density function $g_u(t)$ of T_u is symmetric about 0 and

$$g_u(t) = \frac{3}{2}\,u^2\exp(u^2t)\left\{1-\Phi\left(\frac{3}{2}\,\mu\sqrt{t}\right)\right\} - \frac{1}{2}\,u^2\left\{1-\Phi\left(\frac{1}{2}\,u\sqrt{t}\right)\right\}, \qquad t>0.$$

By a straightforward calculation, it can be shown that the distributions of T_u and T_1/u^2 are the same. Hence, Relation (4) holds.

Proof of Theorem 4.2

By the multivariate central limit theorem, and random vector

$$\sqrt{n}(\bar{X}^*(0,[n\lambda_0]),\bar{X}^*([n\lambda_0],n),\bar{Y}^*(0,[n\lambda_0]),\bar{Y}^*([n\lambda_0],n))'$$

converges in distribution to $\mathrm{MVN}(\mathbf{0},\boldsymbol{\Sigma}(\lambda_0))$, where $\boldsymbol{\Sigma}(\lambda_0)$ is given in (A.1). By Theorem 4.1, $\hat{\lambda}-\lambda_0 \to 0$ in probability as $n\to\infty$, applying the same technique in the proof of the Anscombe's theorem (see, e.g., Chow and Teicher 1988), we can show that the random vector

$$\sqrt{n}(\bar{X}^*(0,[n\hat{\lambda}]),\bar{X}^*([n\hat{\lambda}],n),\bar{Y}^*(0,[n\hat{\lambda}]),\bar{Y}^*([n\hat{\lambda}],n))' \tag{A.4}$$

converges in distribution to $\mathrm{MVN}(\mathbf{0},\boldsymbol{\Sigma}(\lambda_0))$. Straightforward calculation shows that

$$\begin{aligned}
&\sqrt{n}(\hat{\alpha}-\alpha)\\
&= \sqrt{n}\left\{\frac{\bar{Y}([n\hat{\lambda}],n)\bar{X}(0,[n\hat{\lambda}]) - \bar{Y}(0,[n\hat{\lambda}])\bar{X}([n\hat{\lambda}],n)}{\bar{X}(0,[n\hat{\lambda}]) - \bar{X}([n\hat{\lambda}],n)} - \alpha\right\}\\
&= \sqrt{n}[\bar{Y}([n\hat{\lambda}],n)\bar{X}(0,[n\hat{\lambda}]) - \bar{Y}(0,[n\hat{\lambda}])\bar{X}([n\hat{\lambda}],n)\\
&\qquad - \alpha\{\bar{X}(0,[n\hat{\lambda}]) - \bar{X}([n\hat{\lambda}],n)\}]\\
&\qquad \div \{\bar{X}(0,[n\hat{\lambda}]) - \bar{X}([n\hat{\lambda}],n)\}.
\end{aligned} \tag{A.5}$$

For convenience, let $E^*\bar{X}(a,b) = 1/(b-a)\sum_{i=a+1}^{b} EX_i$ and $E^*\bar{Y}(a,b) = 1/(b-a)\sum_{i=a+1}^{b} EY_i$. Note that a and b may be random variables, and hence $E^*\bar{X}(a,b)$ and $E^*\bar{Y}(a,b)$ also may be random variables. By Theorem 4.1, $n(\hat{\lambda}-\lambda_0) = O_p(v_n^{-2})$ as $n\to\infty$, it can be shown that $E^*\bar{X}(0,[n\hat{\lambda}]) - \mu_1 \to 0$ and $E^*\bar{X}([n\hat{\lambda}],n) - \mu_2 \to 0$ in probability as $n\to\infty$. The numerator in (A.4) is equal to

$$\begin{aligned}
&\sqrt{n}\{\bar{Y}^*([n\hat{\lambda}],n)\bar{X}^*(0,[n\hat{\lambda}]) - \bar{Y}^*(0,[n\hat{\lambda}])\bar{X}^*([n\hat{\lambda}],n)\\
&\quad + \bar{X}^*(0,[n\hat{\lambda}])E^*\bar{Y}([n\hat{\lambda}],n) + \bar{Y}^*([n\hat{\lambda}],n)E^*\bar{X}(0,[n\hat{\lambda}])\\
&\quad + E^*\bar{X}(0,[n\hat{\lambda}])E^*\bar{Y}([n\hat{\lambda}],n) - \bar{X}^*([n\hat{\lambda}],n)E^*\bar{Y}(0,[n\hat{\lambda}])\\
&\quad - \bar{Y}^*(0,[n\hat{\lambda}])E^*\bar{X}([n\hat{\lambda}],n) - E^*\bar{X}([n\hat{\lambda}],n)E^*\bar{Y}(0,[n\hat{\lambda}])\\
&\quad - \alpha\bar{X}^*(0,[n\hat{\lambda}]) + \alpha\bar{X}^*([n\hat{\lambda}],n)\\
&\quad - \alpha E^*\bar{X}(0,[n\hat{\lambda}]) + \alpha E^*\bar{X}([n\hat{\lambda}],n)\}\\
&= \sqrt{n}\{(\alpha+\beta\mu_2)\bar{X}^*(0,[n\hat{\lambda}]) + \mu_1\bar{Y}^*([n\hat{\lambda}],n)\\
&\qquad - (\alpha+\beta\mu_1)\bar{X}^*([n\hat{\lambda}],n)\\
&\qquad - \mu_2\bar{Y}^*(0,[n\hat{\lambda}]) - \alpha\bar{X}^*(0,[n\hat{\lambda}])\\
&\qquad + \alpha\bar{X}^*([n\hat{\lambda}],n)\} + o_p(1)
\end{aligned}$$

$$= \sqrt{n}\{\beta\mu_2\bar{X}^*(0,[n\hat{\lambda}]) - \beta\mu_1\bar{X}^*([n\hat{\lambda}],n) - \mu_2\bar{Y}^*(0,[n\hat{\lambda}]) + \mu_1\bar{Y}^*([n\hat{\lambda}],n)\} + o_p(1).$$

On the other hand, because $v_n \to 0$ and $nv_n^2 \to \infty$ as $n \to \infty$, it can be shown that $v_n^{-1}\{\bar{X}(0,[n\hat{\lambda}]) - \mu_1\} \xrightarrow{p} 0$ and $v_n^{-1}\{\bar{X}([n\hat{\lambda}],n) - \mu_2\} \xrightarrow{p} 0$. Hence as $n \to \infty$, the denominator of (A.4),

$$v_n^{-1}\{\bar{X}([0,n\hat{\lambda}]) - \bar{X}([n\hat{\lambda}],n)\} \xrightarrow{p} C.$$

Therefore, by Slutsky's theorem,

$$\sqrt{n}v_n(\hat{\alpha} - \alpha) = C^{-1}\sqrt{n}\{\beta\mu_2\bar{X}^*(0,[n\hat{\lambda}]) - \beta\mu_1\bar{X}^*([n\hat{\lambda}],n) - \mu_2\bar{Y}^*(0,[n\hat{\lambda}]) + \mu_1\bar{Y}^*([n\hat{\lambda}],n)\} + o_p(1) \quad \text{(A.6)}$$

and

$$\begin{aligned}\sqrt{n}v_n(\hat{\beta} - \beta) &= C^{-1}\sqrt{n}[\{\bar{Y}(0,[n\hat{\lambda}]) - \bar{Y}([n\hat{\lambda}],n)\} - \beta\{\bar{X}(0,[n\hat{\lambda}]) - \bar{X}([n\hat{\lambda}],n)\}] \\ &= C^{-1}\sqrt{n}\{\bar{Y}^*(0,[n\hat{\lambda}]) - \bar{Y}^*([n\hat{\lambda}],n) - \beta\bar{X}^*(0,[n\hat{\lambda}]) \\ &\quad + \beta\bar{X}^*([n\hat{\lambda}],n) + E^*\bar{Y}(0,[n\hat{\lambda}]) - E^*\bar{Y}([n\hat{\lambda}],n) \\ &\quad - \beta E^*\bar{X}(0,[n\hat{\lambda}]) + \beta E^*\bar{X}([n\hat{\lambda}],n)\} \\ &= C^{-1}\sqrt{n}\{\bar{Y}^*(0,[n\hat{\lambda}]) - \bar{Y}^*([n\hat{\lambda}],n) \\ &\quad - \beta\bar{X}^*(0,[n\hat{\lambda}]) + \beta\bar{X}^*([n\hat{\lambda}],n)\}. \qquad \text{(A.7)}\end{aligned}$$

From (A.3), (A.5), and (A.6), the asymptotic joint distribution of $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by the δ method and Slutsky's theorem.

Proof of Theorem 4.3

We show only $\hat{\beta} \xrightarrow{p} \beta$ as $n \to \infty$; the remaining parts are analogous and thus are omitted. Because $v_n^{-1}\{\bar{X}(0,[n\hat{\lambda}]) - \mu_1\} \xrightarrow{p} 0, v_n^{-1}\{\bar{X}([n\hat{\lambda}],n) - \mu_2\} \xrightarrow{p} 0, v_n^{-1}\{\bar{Y}(0,[n\hat{\lambda}]) - (\alpha + \beta\mu_1)\} \xrightarrow{p} 0$, and $v_n^{-1}\{\bar{Y}([n\hat{\lambda}],n) - (\alpha + \beta\mu_2)\} \xrightarrow{p} 0$, as $n \to \infty$. Hence as $n \to \infty, v_n^{-1}\{\bar{X}([0,n\hat{\lambda}]) - \bar{X}([n\hat{\lambda}],n)\} \xrightarrow{p} C$ and $v_n^{-1}\{\bar{Y}([0,n\hat{\lambda}]) - \bar{Y}([n\hat{\lambda}],n)\} \xrightarrow{p} C\beta$. Therefore, as $n \to \infty$,

$$\hat{\beta} = \frac{\bar{Y}(0,[n\hat{\lambda}]) - \bar{Y}([n\hat{\lambda}],n)}{\bar{X}(0,[n\hat{\lambda}]) - \bar{X}([n\hat{\lambda}],n)} \xrightarrow{p} \beta.$$

This completes the proof.

[Received February 1994. Revised June 1996.]

REFERENCES

Bhattacharya, G. K. (1987), "Maximum Likelihood Estimation of a Change-Point in the Distribution of Independent Random Variables: General Multiparameter Case," *Journal of Multivariate Analysis*, 23, 183–208.

Bhattacharya, G. K., and Brockwell, P. J. (1976), "The Minimum of an Additive Process With Applications to Signal Estimation and Storage Theory," *Z. Wahrsch. Verw. Gebiete*, 37, 51–75.

Chatterjee, E., and Price, B. (1991), *Regression Analysis by Example*, New York: Wiley.

Chow, Y. S., and Teicher, H. (1988), *Probability Theory: Independence, Interchangeability, Martingales*, Heidelberg: Springer-Verlag.

Cox, N. R. (1976), "The Linear Structural Relation for Several Groups of Data," *Biometrika*, 63, 231–237.

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

Hinkley, D. V. (1969), "Inference About the Intersection in Two-Phase Regression," *Biometrika*, 56, 495–504.

Holbert, D. (1982), "A Bayesian Analysis of a Switching Linear Model," *Journal of Econometrics*, 19, 77–87.

International Mathematical and Statistical Libraries, Inc. (1991), *IMSL User's Manual*, Houston: Author.

James, B., James, K. L., and Siegmund, D. (1992), "Asymptotic Approximations for Likelihood Ratio Tests and Confidence Regions for a Change-Point in the Mean of a Multivariate Normal Distribution," *Statistica Sinica*, 2, 69–90.

Kim, H. J., and Siegmund, A. (1989), "The Likelihood Ratio Test for a Changepoint in Simple Linear Regression," *Biometrika*, 76, 409–423.

Madansky, A. (1959), "The Fitting of Straight Lines When Both Variables are Subjected to Error," *Journal of the American Statistical Association*, 54, 173–205.

Maddala, G. S. (1992), *Introduction to Econometrics*, New York: Macmillan.

Malinvaud, E. (1968), *Statistical Methods of Econometrics*, Chicago: Rand McNally.

McGee, V. E., and Carleton, W. T. (1970), "Piecewise Regression," *Journal of the American Statistical Association*, 65, 1109–1124.

Moran, P. (1971), "Estimating Structural and Functional Relationships," *Journal of Multivariate Analysis*, 1, 232–255.

Quandt, R. E. (1958), "The Estimation of the Parameter of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53, 873–880.

Richardson, D. H., and Wu, D. M. (1970), "Least Squares and Grouping Method Estimators in the Errors-in-Variables Model," *Journal of the American Statistical Association*, 65, 724–748.

Sen, A. K., and Srivastava, M. S. (1975), "On Tests for Detecting Change in Mean When Variance is Unknown," *Annals of the Institute of Statistical Mathematics*, 27, 479–486.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Srivastava, M. S., and Worsley, K. J. (1986), "Likelihood Ratio Tests for a Change in the Multivariate Normal Mean," *Journal of the American Statistical Association*, 81, 199–204.

Van de Geer, S. A. (1988), *Regression Analysis and Empirical Processes*, Centrum voor Wiskunde Informatica.

Wald, A. (1940), "The Fitting of Straight Line if Both Variables are Subject to Error," *Annals of Mathematical Statistics*, 11, 284–300.

Worsley, K. J. (1979), "On the Likelihood Ratio Test for a Shift in Location of Normal Populations," *Journal of the American Statistical Association*, 74, 365–367.

Yao, Y. C. (1987), "Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables," *The Annals of Statistics*, 3, 1321–1328.