CONSISTENT MOMENT ESTIMATORS OF REGRESSION COEFFICIENTS IN THE PRESENCE OF ERRORS IN VARIABLES

Manoranjan PAL*

Indian Statistical Institute, Calcutta 700 035, India

Received August 1979, final version August 1980

This paper examines the possibilities of moment estimators of regression coefficients in the errors-in-variables problem suggested by Geary (1942) and others [Scott (1950) and Drion (1951)]. This approach yields consistent estimators of regression coefficients based on uni- and bi-variate moments (or cumulants) of third or higher order. These are computationally simple and need milder assumptions than the standard techniques, viz., ML and IV estimation. After a review of past investigations, this paper proposes new moment estimators and compares the asymptotic efficiencies of six estimators proposed earlier or here and of the OLS estimator. The case where the true regressor is lognormally distributed receives considerable attention in this communication.

1. Introduction

In many econometric investigations, the errors in variables (EIV) are not negligible [Morgenstern (1963)] and vitiate LS estimation of regression coefficients [Johnston (1972)]. Thus, examination of 25 series relating to national accounts by Langaskens and Rijckeghem (1974) showed that the standard deviations of the errors ranged from 5 to 77 percent of the average value of the corresponding variable.

The well-known methods proposed for handling the classical EIV model (EVM) in regression analysis suffer from serious limitations:

(a) ML estimation requires strong assumptions about the distribution of the errors and also some knowledge of the covariance matrix of the error terms.

(b) The technique of IV estimation is not always handy because suitable instruments may not be available, and in any case, one can never check the assumptions that the instrument is uncorrelated in the limit with each of the error terms. The Wald–Bartlett grouping methods as well as the method due to Durbin (1954) tacitly assume that the errors affecting the regressor values are too small to alter their grouping or ranking.

*I am indebted to Professor N. Bhattacharya of Indian Statistical Institute for his valuable guidance in preparing this paper. I am also thankful to anonymous referees for their helpful suggestions.
This paper examines the possibilities of an approach made by Geary (1942) and others apparently neglected by later researchers. The approach yields consistent and reasonably efficient estimators of regression coefficients based on uni- and bi-variate moments of third or higher order which are computationally simple and need milder assumptions than those mentioned in (a) and (b) above. It is assumed that the errors in the variables are independent of their true variables respectively. The case where they are dependent will be dealt with in a later communication.

Section 2 specifies the two-variable regression model under investigation and reviews the work done by Geary and others. Section 3 compares the asymptotic variances of six moment based estimators mentioned in section 2. Section 4 compares the asymptotic variance of the moment based estimator with the least asymptotic variance with that of OLS assuming that the regressor is error free, under specific distributional assumptions. Section 5 compares the asymptotic efficiencies of the six estimators relative to OLS estimator assuming that the regressor is lognormally distributed. Section 6 deals with estimators based on higher moments which would be useful if the estimators based on third-order moments fail, because the distribution of the 'true regressor' is asymmetric. Section 7 extends these ideas to the case of $m > 1$ regressors and to the case where the error terms are correlated. Section 8 makes some concluding observations on the limitations of the results reached and on further work undertaken by the author.

The motivation for the present study was the need for allowing for transitory (seasonal) elements in both regressor and regressand in Engel curve analysis based on National Sample Survey household budget data where data collected from any individual sample household relates to moving reference period of last 30 days preceding the date of interview. Liviatan (1961) considered the same problem, but his solutions, e.g., the use of recorded income as the IV, are not applicable to our situation.

2. The model and available moment estimators

Consider the following set of relations:

$$Y_i = \alpha + \beta X_i + \epsilon_i, \quad i = 1, 2, \ldots, n, \quad (1)$$

where $X$ and $Y$ are true but non-observable magnitudes of the regressor and the regressand respectively; $\alpha$ and $\beta$ are unknown parameters; and $\epsilon$ is the disturbance which is normally distributed. The assumptions of the Classical Linear Regression Model [Goldberger (1964)] hold excepting that $X$ is stochastic and fully independent of $\epsilon$. The observed values of regressor and regressand are

$$x_i = X_i + u_i \quad \text{and} \quad y_i = Y_i + v_i, \quad (2)$$
where \( u_i \) and \( v_i \) are EIV's assumed to be independent of true values and between themselves. \( \epsilon, u, v \) are assumed to be serially i.i.d. with

\[
E(u_i) = E(v_i) = 0, \quad V(u_i) = \sigma_u^2, \quad V(v_i) = \sigma_v^2, \quad \forall i,
\]
and

\[
E(\epsilon_i) = 0, \quad V(\epsilon_i) = \sigma_\epsilon^2, \quad \forall i.
\]

Let us write for sample moments

\[
m_{ij}(x, y) = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})^j,
\]
and

\[
m_{ij}'(x, y) = \frac{1}{n} \sum_i x_i y_i^j,
\]

where

\[
\bar{x} = \frac{1}{n} \sum x_i \quad \text{and} \quad \bar{y} = \frac{1}{n} \sum y_i.
\]

We may also write for simplicity

\[
m_{ij_0}(x, y) = m_i(x), \quad m_{ij_0}(x, y) = m_j(y).
\]

Correspondingly true moments will be denoted \( \mu_{ij}, \mu_{ij}', \mu_i(x) \) or \( \mu_j(y) \), as the case may be.

It is well-known that under certain conditions the sample moments are consistent estimators of corresponding true moments which are functions of \( \alpha, \beta \), the variances \( \sigma_u^2, \sigma_v^2, \sigma_\epsilon^2 \) and the true moments of \( X \). If only the moments of the first and the second order are considered, five relations are obtained for seven unknown parameters, viz. \( \alpha, \beta, \sigma_u^2, \sigma_v^2, \sigma_\epsilon^2, \mu_i'(X) \) and \( \mu_j'(X) \) [or \( \mu_j'(X) \)]. In fact \( \sigma_v^2 \) and \( \sigma_\epsilon^2 \) always appear in the form of \( \sigma_v^2 + \sigma_\epsilon^2 \), so that in effect we have five relations for six unknown parameters. The first five equations considered by Drion (1951) are

1. \( m_1'(x) = \mu_1'(X) \),
2. \( m_1'(y) = \alpha + \beta \mu_1'(X) \),
3. \( m_2'(x) = \mu_2'(X) + \sigma_u^2 \),
4. \( m_2'(y) = \alpha^2 + 2\alpha \beta \mu_1'(X) + \beta^2 \mu_2'(X) + (\sigma_v^2 + \sigma_\epsilon^2) \),
5. \( m_{11}'(x, y) = \alpha \mu_1'(X) + \beta \mu_2'(X) \).

Drion assumed a functional relationship between \( X \) and \( Y \) so that \( X \) is non-stochastic and \( \sigma_\epsilon^2 = 0 \). However, the introduction of \( \sigma_\epsilon^2 \) does not alter the picture.
One may, then, include similar equations based on third-order moments, if $u$ and $v$ are further assumed to be symmetrically distributed or rather having zero first- and third-order moments. One can choose from the four equations given below:

\begin{itemize}
  \item[(vi)] $m_3(x) = \mu_3(X)$
  \item[(vii)] $m_3(y) = \beta^2 \mu_3(X)$
  \item[(viii)] $m_{21}(x, y) = \beta \mu_3(X)$
  \item[(ix)] $m_{12}(x, y) = \beta^2 \mu_3(X)$
\end{itemize}

Inclusion of any two of these equations introduces only one new unknown parameter, namely $\mu_3(X)$. Drion used the equations for $m_3(x)$ and $m_3(y)$ and solved the system of seven equations for the seven unknown parameters to get $\beta = \sqrt[3]{m_3(y)/m_3(x)}$ as an estimator of $\beta$ which is consistent under mild conditions if $\lim_{n \to \infty} m_3(x) \neq 0$ [or simply $\mu_3(X) \neq 0$].

For each pair of equations from the set (vi) to (ix) we get a separate set of estimators. Thus, for estimation of $\beta$ we have six choices:

\begin{align*}
  \hat{\beta}_1 &= m_{03}/m_{12}, \quad \hat{\beta}_2 = m_{12}/m_{21}, \quad \hat{\beta}_3 = m_{21}/m_{30} \quad \text{[Durbin (1954)]},
  \\
  \hat{\beta}_4 &= \sqrt[3]{m_{03}/m_{30}} \quad \text{[Drion]}, \quad \hat{\beta}_5 = \pm \sqrt{m_{03}/m_{21}}, \quad \hat{\beta}_6 = \pm \sqrt{m_{12}/m_{30}}
\end{align*}

The choice of signs for $\hat{\beta}_4$ and $\hat{\beta}_5$ can be based on the sign of any one of the other four estimates. Obviously from eqs. (vi) to (ix) it follows that each of the estimators is consistent if $\mu_3(X) \neq 0$. Of course for the estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ we have to assume that $\beta \neq 0$.

The first three estimators may be regarded as basic estimators. All other estimators, based only on moments up to third order, must be functions of these three estimators. Thus,

\begin{align*}
  \hat{\beta}_4 &= \sqrt[3]{\hat{\beta}_1 \cdot \hat{\beta}_2 \cdot \hat{\beta}_3}, \quad \hat{\beta}_5 = \pm \sqrt{\hat{\beta}_1 \cdot \hat{\beta}_2}, \quad \hat{\beta}_6 = \pm \sqrt{\hat{\beta}_2 \cdot \hat{\beta}_3}.
\end{align*}

In fact we can find infinitely many consistent estimates forming weighted arithmetic or geometric means of these estimators. More generally, suppose

\[ \hat{\beta}_n = f(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3), \]

such that

\[ f(c\hat{\beta}_1, c\hat{\beta}_2, c\hat{\beta}_3) = cf(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) \quad \text{for all} \quad c \neq 0 \]

\footnote{If $\beta = 0$, OLS estimate $\hat{\beta}_0$ is a consistent estimate of $\beta$ and the asymptotic variance is $O(1/n)$ like that of $\hat{\beta}_4$ — and this can be used to test $H_0: \beta = 0$ in large samples. However, $\beta = 0 \Rightarrow x$ and $y$ are fully independent (in the present model) and one can test $H_0: \beta = 0$ in finite samples using rank correlation methods, or through the $t$-test for sample correlation coefficient assuming conditional distributions of $y$ given $x$ to be normal.}

and

\[ f(1, 1, 1) = 1, \]

then \( \hat{\beta}_s \) is consistent since \( \hat{\beta}_1, \hat{\beta}_2 \) and \( \hat{\beta}_3 \) are consistent. Note that Scott's (1950) estimate\(^3\) can be shown to belong to this class.

It may be mentioned that only \( \hat{\beta}_2 \) out of the estimators mentioned above is a member of the class of estimators proposed by Geary (1942).

3. Comparative asymptotic variances in the general case

Each of the above six estimators ‘under mild conditions’ have asymptotic normal distribution. Asymptotic variances of the six estimators can easily be obtained. If in addition \( u \) and \( v \) are normally distributed the expressions for the asymptotic variances reduce to the following:

\[
V_1 = R \cdot \{(C_u + C_v)(b_2 - 1) + 4C_uC_v + 8C_v^2 \\
+ 2C_uC_v^2 + 6C_v^3\},
\]

\[
V_2 = R \cdot \{(C_u + C_v)(b_2 - 1) + 2C_u^2 + 2C_uC_v^2 + 2C_v^2 \\
+ 2C_uC_v^3\},
\]

\[
V_3 = R \cdot \{(C_u + C_v)(b_2 - 1) + 8C_u^2 + 6C_u^3 + 4C_uC_v^2 \\
+ 2C_uC_v^3\},
\]

\[
V_4 = R \cdot \{(C_u + C_v)(b_2 - 1) + 2C_u^2 + \frac{3}{2}C_u^3 + 2C_v^2 \\
+ \frac{3}{2}C_v^3\},
\]

\[
V_5 = R \cdot \{(C_u + C_v)(b_2 - 1) + 0.5C_u^2 + C_uC_v + 0.5C_vC_u^2 \\
+ 4.5C_v^2 + 1.5C_v^3\},
\]

\[
V_6 = R \cdot \{(C_u + C_v)(b_2 - 1) + 4.5C_u^2 + 1.5C_v^3 + C_uC_v^2 \\
0.5C_uC_v + 0.5C_v^3\},
\]

and for OLS estimator we have

\[
V_0 = R' \cdot [C_u(b_2C_u + 1 - C_u + C_v^2) + C_v(1 + C_u)^3],
\]

\(^3\)Scott proved that \( m_{01} - 3bm_{12} + 3b^2m_{11} - b^3m_{30} = 0 \) has a root which will be a consistent estimator of \( \beta \). But she does not give any method to find out the particular root which will be consistent. Both Scott and Drion assumed that \( \sigma^2 = 0 \). But obviously if \( \sigma^2 > 0 \) the approaches remain valid; only the estimate \( \hat{\sigma}^2 \) now estimates \( \sigma^2 + \sigma^2 \). However, Scott also assumed that \( u \) and \( v \) are normally distributed which is not necessary for the consistency property. Symmetry of \( u \) and \( v \) serves the purpose.
where
\[ R = \beta^2 / (nb_1), \quad C_u = \sigma_u^2 / \sigma_X^2, \quad C_c = \sigma_c^2 / (\beta^2 \sigma_X^2) = (\sigma_c^2 + \sigma_u^2) / (\beta^2 \sigma_X^2), \]
and
\[ b_1 = \mu_3(X) / \sigma_X^2, \quad b_2 = \mu_4(X) / \sigma_X^4 \quad \text{and} \quad R' = \beta^2 / \{n(1 + C_u)^4\}. \]

Three interesting specific cases may be investigated here.

**Case 1:** \( C_u = 0. \) Here OLS estimation is optimal. In this case,
\[ V_0 \leq V_3 \leq V_6 \leq V_2 \leq V_4 \leq V_5 \leq V_1. \]
The equality between \( V_0 \) and \( V_3 \) holds iff \( b_2 - b_1 - 1 = 0 \), i.e., if the variable \( X \) takes only two distinct values. Equality between any two of \( V_1 \) to \( V_6 \) holds iff \( C_c = 0. \)

**Case 2:** \( C_c = 0. \) Here also we get straightforward inequalities between \( V_1 \) to \( V_6 \) as follows:
\[ V_1 \leq V_5 \leq V_2 \leq V_4 \leq V_6 \leq V_3. \]

In general we cannot say which one of \( V_1 \) and \( V_0 \) is larger. But there are cases for which we can say something such as the following theorem:

**Theorem 1.** \( V_0 < V_1 \) if \( b_1 < 4 + 5C_u + 4C_u^2 + C_u^3. \)

**Proof.** Writing \( g(b_1, b_2) = V_0 / V_1 \) we can easily see that
\[ \max_{b_2} g(b_1, b_2) < 1 \iff b_1 < 4 + 5C_u + 4C_u^2 + C_u^3. \quad \text{Q.E.D.} \]

\( C_c = 0 \) is the structural case where the regressand is free from error. Hence reverse least squares yields MLE
\[ \hat{\beta}_1 = m_{02} / m_{11}, \]

asymptotic variance of which, in general,
\[ V_7 = (\beta^2 / n)[C_c \{ C_c b_2 + 1 - C_c + C_c^2 \} + C_u (1 + C_c)^3]. \]
The efficiency of \( \hat{\beta}_1 \) relative to \( \hat{\beta}_7 \), in the present case, is
\[ E(\hat{\beta}_1 | C_c = 0) = b_1 / (b_2 - 1). \]
Observe that this is the efficiency of $\hat{\beta}_3$ relative to $\hat{\beta}_0$ where $C_u = 0$ (section 4).

**Case 3:** $C_u = C_v = C$ (say). If $C_u = C_v$, the relative magnitudes of the variances can be shown diagrammatically as under

![Diagram](image)

where ‘$\rightarrow$’ means ‘$\leq$’.

Here the MLE is

$$\hat{\beta}_8 = \pm \sqrt{\frac{m_{02}}{m_{20}}}$$

where the sign depends on the sign of $m_{11}$. The asymptotic variance of $\hat{\beta}_8$ can be shown to be, in general,

$$V_8 = \frac{\beta^2}{4u(1+C_v)(1+C_u)} [ (b_2 - 1)(C_u - C_v)^2 + 2C_v(2+C_u)(1+C_v)^2 + 2C_u(2+C_v)(1+C_u)^2 ]$$

which reduces, in the present case, to

$$V_8 = \beta^2 C(2+C)/n(1+C)^2.$$

So, efficiency of $\hat{\beta}_4$ relative to $\hat{\beta}_8$ is

$$E(\hat{\beta}_4 | C_u = C_v = C) = \frac{(2+C)b_1}{2(1+C)^2 \{ (b_2 - 1) + 2C + \frac{3}{2}C_2 \}}$$

$$\approx \frac{b_1}{(b_2 - 1)}, \text{ for small } C.$$

**General Conclusion:** If error in $X$ is zero then use the estimator which has most influence of $x$, i.e., $\hat{\beta}_3 (= m_{21}/m_{30})$. If error in $Y$ is zero then by the same argument use $\hat{\beta}_1 (= m_{03}/m_{12})$. If the two relative errors are equal then use $\hat{\beta}_4 (= 3\sqrt{m_{03}/m_{30}})$ which have equal influences of $x$ and $y$.

**General Case:** $(C_u \geq 0, C_v \geq 0)$. It is clear that $V_1$ and $V_3$ are symmetric in the sense that

$$V_1(C_u, C_v) = V_3(C_v, C_u),$$
i.e., from $V_i$ we get $V_4$ simply by interchanging roles of $C_u$ and $C_r$. This is also true for $V_5$ and $V_6$. Moreover

(i) $V_i < V_3$ if $C_r < C_u$ and  
(ii) $V_5 < V_6$ if $C_r < C_u$.

We also note that if $C_u > C_r$, then $V_3 = \max(V_1, V_2, \ldots, V_6)$; and if $C_u < C_r$, then $V_i = \max(V_1, V_2, \ldots, V_6)$.

4. Efficiency of $\hat{\beta}_3$ where OLS is valid

We have seen that if $C_u = 0$, $\hat{\beta}_3$ is the best among six moment estimators. In order to study the efficiency of $\hat{\beta}_3$ relative to the OLS estimator, we assume plausible forms for the distribution of $X$.

Asymptotic efficiency of $\hat{\beta}_3$ relative to $\hat{\beta}_0$, if $\sigma_u^2 = 0$, is

$$E(\hat{\beta}_3 \mid \sigma_u^2 = 0) = b_1 / (b_2 - 1).$$

Observe that the efficiency does not depend on $C_r$, and is a function of $b_1$ and $b_2$ only of the distribution of the true regressor.

Lognormal Distribution: Let $X \sim \mathcal{N}(\mu, \sigma^2)$. For simplicity take $\mu = 1$, since efficiency remains unaffected by change of $\mu$,

$$E(\hat{\beta}_3 \mid \sigma_u^2 = 0) = \frac{(w^2 - 1)(w^2 + 2)^2}{w^8 + 2w^6 + 3w^4 - 4},$$

where $w = \exp(\sigma^2/2)$. Denoting $E(\hat{\beta}_3 \mid \sigma_u^2 = 0)$ by $E(\sigma^2)$ we may present the following values:

$$\lim_{\sigma^2 \to 0} E(\sigma^2) = 0, \quad E(0.01) = 0.042, \quad E(0.1) = 0.263,$$

$$E(0.5) = 0.421, \quad E(1) = 0.339, \quad E(2) = 0.143.$$

Efficiency of $\hat{\beta}_3$ increases from zero to slightly over 0.42 reaching peak between $\sigma^2 = 0.5$ and 0.6; and then slowly decreases to zero.

It may be mentioned here that for empirical size distributions of population by per capita household consumption expenditure estimated for rural and urban India from different rounds of NSS the fitted LN distributions have $\sigma^2$ in the region of $(0.25, 0.5)$ corresponding to Lorenz radio $(0.28, 0.38)$ [Roy and Dhar (1961)].
Gamma Distribution: Let \( X \sim G(\alpha, \beta) \). Putting \( \alpha = 1 \) to simplify calculations we find the efficiency of \( \hat{\beta}_3 \) as

\[ E(\beta) = \frac{2}{(p + 3)}. \]

Efficiency is maximum, \( 2/3 \), when \( p = 0 \) and decreases to zero as \( p \to \infty \).

Salem and Mount (1974) fitted Gamma distribution to personal income data for the United States for the years 1960–69, and found that (the particular) \( p \) lay in the interval (1.94, 2.51) so that the asymptotic efficiency of \( \hat{\beta}_3 \) falls in the interval (0.365, 0.405).

5. Comparative asymptotic efficiencies where \( X \sim \Lambda(\mu, \sigma^2) \)

We now examine the comparative asymptotic efficiencies of \( \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_6 \) in the general case where neither \( C_u \) not \( C_\epsilon \) is necessarily zero. We assume that \( X \) is lognormally distributed, which is realistic in engel curve analysis in many countries [vide Aitchison and Brown (1957), Bhattacharya and Iyengar (1961), Roy and Dhar (1960), Iyengar (1967)].

Symbolically, let \( X \sim \Lambda(\mu, \sigma^2) \), and

\[ E_i = V_i/V_0 = \text{the efficiency of } \hat{\beta}_i \text{ relative to } \hat{\beta}_0. \]  

(3)

Obviously \( E_i \) is a function of \( \sigma^2, C_u \) and \( C_\epsilon \). We calculated the asymptotic efficiency of each estimator as defined in (3) for each combination of values of \( \sigma^2, C_u \) and \( C_\epsilon \).

\[ C_u : 0, 0.01, 0.02, 0.05, 0.07, 0.1, 0.15, 0.2, 0.5, 1.0. \]
\[ C_\epsilon : 0, 0.01, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0. \]
\[ \sigma^2 : 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. \]
\[ 1.0, 1.5, 2.0, 5.0. \]

The results are presented briefly through table 1 and fig. 1.

(a) As could be expected the ranking of estimators is independent of \( \sigma^2 \), though the actual asymptotic efficiency is influenced by \( \sigma^2 \).

(b) When \( C_\epsilon \) is zero, \( \beta_3 \) ranks first, as it increases (\( C_u \) remaining constant) \( \beta_3, \beta_4, \beta_6 \) and then \( \beta_3 \) take the first rank sequentially. Fig. 1 gives a broad picture of their relative positions for different combinations of \( C_u \) and \( C_\epsilon \).

(c) We must note that as \( \sigma^2 \) increases asymptotic efficiencies of the six estimators approach equality and for \( \sigma^2 > 0.7 \) the estimators are practically equally efficient.

These values of \( \sigma^2 \) correspond to LR's ranging from 0.06 to 0.89.
Table 1

Asymptotic efficiencies of $b_i$ (i = 1, 2, ... 6) w.r.t. $b_i$ for different values of $\sigma^2$, $C_r$ and $C_c$.

<table>
<thead>
<tr>
<th>$C_r$</th>
<th>$\sigma^2$</th>
<th>$0.1$</th>
<th>$0.5$</th>
<th>$1$</th>
<th>$0.1$</th>
<th>$0.5$</th>
<th>$1$</th>
<th>$0.1$</th>
<th>$0.5$</th>
<th>$1$</th>
<th>$0.1$</th>
<th>$0.5$</th>
<th>$1$</th>
<th>$0.1$</th>
<th>$0.5$</th>
<th>$1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.61</td>
<td>0.40</td>
<td>0.34</td>
<td>0.37</td>
<td>0.42</td>
<td>0.34</td>
<td>0.37</td>
<td>0.42</td>
<td>0.34</td>
<td>0.37</td>
<td>0.42</td>
<td>0.34</td>
<td>0.37</td>
<td>0.42</td>
<td>0.34</td>
<td>0.37</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>0.5</td>
<td>0.21</td>
<td>0.61</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>0.5</td>
<td>0.01</td>
<td>0.11</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.5</td>
<td>0.05</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>0.5</td>
<td>0.17</td>
<td>0.06</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>0.5</td>
<td>0.15</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>0.5</td>
<td>0.04</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>0.5</td>
<td>0.10</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>0.5</td>
<td>0.08</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>0.5</td>
<td>0.03</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
</tr>
</tbody>
</table>

M. Pal, Moment estimators of regression coefficients
It is seen from the table that when $C_v$ and $\sigma^2$ are held constant, the efficiency of the various $\beta_i$ relative to OLS does not always improve as $C_u$ increases. The picture would of course be different if MSE were considered.

In table 2 we present some typical situations in Engel curve analysis.

<table>
<thead>
<tr>
<th>$C_u$</th>
<th>$\sigma^2$</th>
<th>$C_v$</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\hat{\beta}_1$</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td>0.306</td>
</tr>
<tr>
<td>0.4</td>
<td>1</td>
<td></td>
<td>0.209</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td></td>
<td>0.108</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td></td>
<td>0.250</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td></td>
<td>0.140</td>
</tr>
</tbody>
</table>

Fig. 1. Regions of best estimator for different combinations of $C_u$ and $C_v$.
6. Estimation via cumulants

We first give some results on bivariate cumulants which throw up a series of estimators of $\beta$ including some of these proposed earlier.

Let $K(m, n)$ or $K_{m,n}(X, Y)$ denote bivariate cumulant of order $(m, n)$ of the joint distribution of $(X, Y)$; $K'(m, n)$ denotes the same quantity for $(x, y)$.\(^{5}\)

The following theorem was proved by Geary (1942):

Theorem 2. Suppose $(u, v)$ is jointly independent of $(X, Y)$. Also suppose that $u$ is independent of $v$. Then

$$
\hat{\beta} = \frac{K'(c_1, c_2 + 1)}{K'(c_1 + 1, c_2)}, \quad c_1 \text{ and } c_2 > 0,
$$

is a consistent estimator of $\beta$ if $K(c_1 + 1, c_2) \neq 0$.

The only estimator of this type based on cumulants of order three is $\hat{\beta}_1 = K'(0, r)/K'(1, r - 1)$. If in addition to above, $u$ and $v$ are symmetrically distributed then there exist two more estimators via cumulants of order three, namely,

$$
\hat{\beta}_2 = \frac{K'(0, r)}{K'(1, r - 1)} \quad \text{and} \quad \hat{\beta}_3 = \frac{K'(r, 0)}{K'(1, r - 1)}
$$

which are consistent under the same condition. Observe that $K_{ij} = \mu_{ij}$ for $i \neq j = 3$. Hence these estimators are nothing but our $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ considered in section 2.

Evidently, our method of estimation via third-order cumulants fails when $X$ is symmetrically distributed. For a symmetric distribution of $X$, provided $K_4(X) \neq 0$, estimation via fourth-order cumulants is possible. In general the following results follow from properties of cumulants listed in appendix:

(1) For $r > 2$,

$$
K'(0, r) = \beta K'(1, r - 1) + K_r(v).
$$

Now if $v$ is symmetric and $r$ is odd ($\geq 3$) then $K_r(v) = 0$. Hence $\hat{\beta} = \frac{K'(0, r)}{K'(1, r - 1)}$ is consistent if $K(1, r - 1) \neq 0$.

(2) Similarly, if $u$ is symmetric and $r$ is odd ($\geq 3$) then $\hat{\beta} = \frac{K'(r - 1, 1)}{K'(r, 0)}$ is consistent if $K(r, 0) \neq 0$.

(3) If $u$ is normally distributed then $K_r(u) = 0$ for $r \geq 3$. Hence $\hat{\beta} = \frac{K'(r - 1, 1)}{K'(r, 0)}$ is consistent for any $r \geq 3$ if $K(r, 0) \neq 0$.

(4) If $v$ is normally distributed then $K_r(v) = 0$ for $r \geq 3$. Hence $\hat{\beta} = \frac{K'(0, r)}{K'(1, r - 1)}$ is consistent for any $r \geq 3$ if $K(1, r - 1) \neq 0$.

\(^{5}\)Properties of cumulants are stated in the appendix.
In particular, in addition to Geary's estimators via fourth-order cumulants, we may have two more consistent estimators if we assume that \( u \) and \( v \) are normally distributed,

\[
\hat{K}'(3,1)/\hat{K}'(4,0) \quad \text{and} \quad \hat{K}'(0,4)/\hat{K}'(1,3),
\]

both estimators being consistent if \( K(4,0) \neq 0 \). For the second estimator we must have \( \beta \neq 0 \).

7. Some extensions and comments

We may consider the case of \( m \) regressors each subject to error besides the regressand, making the same assumptions as in the previous sections. Our model is

\[
Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m + \varepsilon, \quad (4)
\]

\[
x_i = X_i + u_i, \quad i = 1, 2, \ldots, m, \quad (5)
\]

\[
y = Y + v. \quad (6)
\]

Suppose

\[
Y' = Y - \bar{Y} \quad \text{and} \quad X'_i = X_i - \bar{X}_i, \quad i = 1, 2, \ldots, m,
\]

and similarly

\[
y' = y - \bar{y} \quad \text{and} \quad x'_i = x_i - \bar{x}_i, \quad i = 1, 2, \ldots, m;
\]

\( v \) and \( u_i \)'s are assumed to be symmetrically distributed.

Then

\[
E(y'_i x'_i^2) = \sum_{j=i}^m \beta_j E(x'_j x'_i^2), \quad i = 1, 2, \ldots, m,
\]

Hence

\[
A\beta = B,
\]

where

\[
A = ((a_{ij})) \quad \text{with} \quad a_{ij} = E(x'_i x'_j),
\]

\[
\beta' = (\beta_1, \beta_2, \ldots, \beta_m),
\]

and

\[
B' = (E(y'_1 x'_1^2), \ldots, E(y'_m x'_m^2)).
\]

\[
\hat{\beta} = \hat{A}^{-1}\hat{B} \quad \text{is consistent provided that} \quad |A| \neq 0, \quad \text{where} \quad \hat{A} \quad \text{and} \quad \hat{B} \quad \text{are sample estimates of} \quad A \quad \text{and} \quad B.
\]
In many empirical studies we must relax the assumption that $u$ and $v$ are independent. We may take $(u,v)$ to be bivariate normally distributed with unknown correlation coefficient $\rho$. Interestingly enough our estimation procedure does not differ at all in either way. Both sets of assumptions give us same estimate of $\beta$. This is due to the peculiarity of bivariate normal distribution which says that $K_{ij}(u,v)$, for $i+j>2$, is zero. The multivariate extension of the problem of correlated errors is also similar. This is due to the fact that marginally each $u_i$ is normally distributed and each pair $(u_i,u_j)$ is bivariate normal.

The method of estimation via cumulants originally introduced by Geary may be looked upon as application of the IV method; instruments being taken from $x$ and $y$ itself. As for example, the three third-order cumulant estimators may be viewed to have instruments $z_1(=y^2)$, $z_2(=x'y')$ and $z_3(=x^2)$ respectively.

8. Conclusion

This paper examines the possibilities of moment/cumulant based estimators of the kind first proposed by Geary. It proposes some new estimators of that class which have smaller asymptotic variances in some specific situations. It also compares the asymptotic efficiencies of various estimators based on third-order cumulants and finds the best estimator in different regions of the parametric space assuming lognormality of the regressor which is realistic for some economic data, e.g., in the Engel curve analysis. Efficiencies of these estimators relative to OLS have been investigated.

As soon as errors affect observations on the regressor, comparison of the variances of these six estimators ($\hat{\beta}_1, \ldots, \hat{\beta}_6$) with the variance of the OLS estimator ($\hat{\beta}_0$) does not seem to be justifiable on the ground that the OLS estimator is biased and inconsistent. So the two MSE's should be compared and since the $\hat{\beta}_i's$ ($i=1,2,\ldots,6$) are consistent, the efficiency of the $\hat{\beta}_i's$ relative to $\hat{\beta}_0$ based on the MSE criterion goes to infinity as $n\to\infty$.

The choice of one out of the class of estimators mentioned in this paper may be difficult in many situations, because this requires estimation of the variance of every estimator. It is well-known that standard errors of estimators of cumulants generally increase with their order [Madansky (1959), Geary (1942)]. Hence one should not take higher-order cumulants when estimation is possible by taking lower-order cumulants. So if $X$ is asymmetric ($\mu_3(X)\neq0$) one should base the estimate on cumulants of order three. The assumption that $\mu_3(X)\neq0$ is very important in this case. If $X$ is symmetric but non-normal [so that $K_4(X)\neq0$] then one should use fourth-order cumulants.

Geary admitted that this method is inapplicable if $X$ is normally
distributed. In fact, if all the cumulants of order three or more vanish, then one can conclude that either (i) the variates \( X \) and \( Y \) are independent or (ii) they are normally distributed. It may be recalled that if \((X, Y, u, v, e)\) are normally distributed, then the parameter, \( \beta \), of this model is not identifiable [Reiersol (1950)].

The assumption of independence of true and error components may sometimes be inappropriate. Many economic variables like income or consumer expenditure are seasonally affected and if the reference period of the enquiry is short (say, a month preceding date of interview) then the error components are likely to be dependent on the true components. This problem will be taken up in a later communication.

In general, methods of tackling econometric problems with errors in variables should depend heavily on what is known about these errors. To reiterate Morgenstern (1963): 'As long as theory has not been sufficiently developed to cover the complicated cases of many simultaneous sources of error and their shifting nature of interdependence, one must proceed on an heuristic and common sense basis.'

**Appendix: Properties of cumulants**

*Property 1:* Cumulants are invariant under changes of origin, except the first \( (K_1 = \mu_1) \).

*Property 2:* If the variate values are multiplied by a constant \( C \), \( K_r \) is multiplied by \( C^r \).

*Property 3:* The cumulant of a sum of independent variables is the sum of the cumulants of the variables.

*Property 4:* \( K_r(r \geq 3) = 0 \) for normal distribution.

*Property 5:* The bivariate cumulant \( K_{c_1, c_2} (c_1 > 0, c_2 > 0) \) of independent random variables is zero.

*Property 6:* If \((u, v)\) is independent of \((X, Y)\), then \( K_{c_1, c_2} (X + u, Y + v) = K_{c_1, c_2} (X, Y) + K_{c_1, c_2} (u, v) \).

*Property 7:* \( K_{c_1, c_2} (X_1, a + bX_2) \) equals \( a + b K_{c_1, c_2} (X_1, X_2) \) if \( c_1 = 0 \) and \( c_2 = 1 \), and \( b^2 K_{c_1, c_2} (X_1, X_2) \) otherwise.

*Property 8:* Any odd \((\geq 3)\) order cumulant of a symmetric distribution is zero.

*Property 9:* For a bivariate normal distribution, the bivariate cumulant \( K_{c_1, c_2} = 0 \) if \( c_1 + c_2 > 2 \).
References

Bhattacharya, N. and N.S. Iyengar, 1961, Change in the pattern of distribution of persons by monthly per capita total consumer expenditure over NSS rounds, Working paper no. 218 (Indian Statistical Institute, Planning Division, Calcutta).
Drion, E.F., 1951, Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error, Indagationes Mathematicae 13, 256-260.
Reiersol, O., 1950, Identifiability of a linear relation between variables which are subject to error, Econometrica 18, 375–389.