Chapter 16
Measurement Error Models

A fundamental assumption in all the statistical analysis is that all the observations are correctly measured. In the context of multiple regression model, it is assumed that the observations on the study and explanatory variables are observed without any error. In many situations, this basic assumption is violated. There can be several reasons for such a violation.

- For example, the variables may not be measurable, e.g., taste, climatic conditions, intelligence, education, ability etc. In such cases, the dummy variables are used, and the observations can be recorded in terms of values of dummy variables.
- Sometimes the variables are clearly defined, but it is hard to take correct observations. For example, the age is generally reported in complete years or in multiple of five.
- Sometimes the variable is conceptually well defined, but it is not possible to take a correct observation on it. Instead, the observations are obtained on closely related proxy variables, e.g., the level of education is measured by the number of years of schooling.
- Sometimes the variable is well understood, but it is qualitative in nature. For example, intelligence is measured by intelligence quotient (IQ) scores.

In all such cases, the true value of the variable can not be recorded. Instead, it is observed with some error. The difference between the observed and true values of the variable is called as **measurement error** or **errors-in-variables.**

Difference between disturbances and measurement errors:

The disturbances in the linear regression model arise due to factors like the unpredictable element of randomness, lack of deterministic relationship, measurement error in study variable etc. The disturbance term is generally thought of as representing the influence of various explanatory variables that have not actually been included in the relation. The measurement errors arise due to the use of an imperfect measure of true variables.

Large and small measurement errors

If the magnitude of measurement errors is small, then they can be assumed to be merged in the disturbance term, and they will not affect the statistical inferences much. On the other hand, if they are large in magnitude, then they will lead to incorrect and invalid statistical inferences. For example, in the context of linear regression model, the ordinary least squares estimator (OLSE) is the best linear unbiased estimator of the regression coefficient when measurement errors are absent. When the measurement errors are present in the data, the same OLSE becomes biased as well as inconsistent estimator of regression coefficients.

Consequences of measurement errors:

We first describe the measurement error model. Let the true relationship between correctly observed study and explanatory variables be

$$\tilde{y} = \tilde{X}\beta$$

where $\tilde{y}$ is a $(n \times 1)$ vector of true observation on study variable, $\tilde{X}$ is a $(n \times k)$ matrix of true observations on explanatory variables and β is a $(k \times 1)$ vector of regression coefficients. The value $\tilde{y}$ and $\tilde{X}$ are not observable due to the presence of measurement errors. Instead, the values of $\tilde{y}$ and $\tilde{X}$ are observed with additive measurement errors as

$$y = \tilde{y} + u$$
$$X = \tilde{X} + V$$

where y is a $(n \times 1)$ vector of observed values of study variables which are observed with $(n \times 1)$ measurement error vector u. Similarly, X is a $(n \times k)$ matrix of observed values of explanatory variables which are observed with $(n \times k)$ matrix V of measurement errors in X. In such a case, the usual disturbance term can be assumed to be subsumed in u without loss of generality. Since our aim is to see the impact of measurement errors, so it is not considered separately in the present case.

Alternatively, the same setup can be expressed as

$$y = \tilde{X}\beta + u$$
$$X = \tilde{X} + V$$

where it can be assumed that only X is measured with measurement errors V and u can be considered as the usual disturbance term in the model.

In case, some of the explanatory variables are measured without any measurement error then the corresponding values in V will be set to zero.

We assume that

$$E(u) = 0,\ E(uu') = \sigma^2 I$$
$$E(V) = 0, E(V'V) = \Omega,\ E(V'u) = 0.$$

The following set of equations describes the measurement error model

$$\tilde{y} = \tilde{X}\beta$$
$$y = \tilde{y} + u$$
$$X = \tilde{X} + V$$

which can be re-expressed as

$$\begin{aligned} y &= \tilde{y} + u \\ &= \tilde{X}\beta + u \\ &= \left(X - V\right)\beta + u \\ &= X\beta + \left(u - V\beta\right) \\ &= X\beta + \omega \end{aligned}$$

where $\omega = u - V\beta$ is called as the **composite disturbance term**. This model resemble like a usual linear regression model. A basic assumption in linear regression model is that the explanatory variables and disturbances are uncorrelated. Let us verify this assumption in the model $y = X\beta + w$ as follows:

$$\begin{aligned} E\left[\{X - E(X)\}'\{\omega - E(\omega)\}\right] &= E\left[V'(u - V\beta)\right] \\ &= E\left[V'u\right] - E\left[V'V\right]\beta \\ &= 0 - \Omega\beta \\ &= -\Omega\beta \\ &\neq 0. \end{aligned}$$

Thus X and ω are correlated. So OLS will not provide efficient result.

Suppose we ignore the measurement errors and obtain the OLSE. Note that ignoring the measurement errors in the data does not mean that they are not present. We now observe the properties of such an OLSE under the setup of measurement error model.

The OLSE is

$$b=(X'X)^{-1}X'y$$
$$b-\beta=(X'X)^{-1}X'(X\beta+\omega)-\beta$$
$$=(X'X)^{-1}X'\omega$$
$$E(b-\beta)=E\left[(X'X)^{-1}X'\omega\right]$$
$$\neq (X'X)^{-1}X'E(\omega)$$
$$\neq 0$$

as X is a random matrix which is correlated with ω. So b becomes a biased estimator of β.

Now we check the consistency property of OLSE. Assume

$$\text{plim}\left(\frac{1}{n}\tilde{X}'\tilde{X}\right)=\Sigma_{xx}$$
$$\text{plim}\left(\frac{1}{n}V'V\right)=\Sigma_{vv}$$
$$\text{plim}\left(\frac{1}{n}\tilde{X}'V\right)=0$$
$$\text{plim}\left(\frac{1}{n}V'u\right)=0.$$

Then

$$\text{plim}(b-\beta)=\text{plim}\left[\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\omega}{n}\right)\right]$$
$$\frac{1}{n}X'X=\frac{1}{n}(\tilde{X}+V)'(\tilde{X}+V)$$
$$=\frac{1}{n}\tilde{X}'\tilde{X}+\frac{1}{n}\tilde{X}'V+\frac{1}{n}V'\tilde{X}+\frac{1}{n}V'V$$
$$\text{plim}\left(\frac{1}{n}X'X\right)=\text{plim}\left(\frac{1}{n}\tilde{X}'\tilde{X}\right)+\text{plim}\left(\frac{1}{n}\tilde{X}'V\right)+\text{plim}\left(\frac{1}{n}V'\tilde{X}\right)+\text{plim}\left(\frac{1}{n}V'V\right)$$
$$=\Sigma_{xx}+0+0+\Sigma_{vv}$$
$$=\Sigma_{xx}+\Sigma_{vv}$$

$$\frac{1}{n}X'\omega=\frac{1}{n}\tilde{X}'\omega+\frac{1}{n}V'\omega$$
$$=\frac{1}{n}\tilde{X}'(u-V\beta)+\frac{1}{n}V'(u-V\beta)$$

$$\text{plim}\left(\frac{1}{n}X'\omega\right)=\text{plim}\left(\frac{1}{n}\tilde{X}'u\right)-\text{plim}\left(\frac{1}{n}\tilde{X}'V\right)\beta+\text{plim}\left(\frac{1}{n}V'u\right)-\text{plim}\left(\frac{1}{n}V'V\right)\beta$$
$$=0-0+0-\Sigma_{vv}\beta$$
$$\text{plim}\left(b-\beta\right)=\text{plim}\left(\frac{X'X}{n}\right)^{-1}\text{plim}\left(\frac{X'\omega}{n}\right)$$
$$=-\left(\Sigma_{xx}+\Sigma_{vv}\right)^{-1}\Sigma_{vv}\beta$$
$$\neq 0.$$

Thus b is an inconsistent estimator of β. Such inconsistency arises essentially due to correlation between X and ω.

Note: It should not be misunderstood that the OLSE $b=\left(X'X\right)^{-1}X'y$ is obtained by minimizing $S=\omega'\omega=\left(y-X\beta\right)'\left(y-X\beta\right)$ in the model $y=X\beta+\omega$. In fact $\omega'\omega$ cannot be minimized as in the case of usual linear regression, because the composite error $\omega=u-V\beta$ is itself a function of β.

To see the nature of consistency, consider the simple linear regression model with measurement error as

$$\tilde{y}_i=\beta_0+\beta_1\tilde{x}_i,\qquad i=1,2,...,n$$
$$y_i=\tilde{y}_i+u_i$$
$$x_i=\tilde{x}_i+v_i\ .$$

Now

$$X=\begin{pmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix},\qquad \tilde{X}=\begin{pmatrix}1 & \tilde{x}_1\\ 1 & \tilde{x}_2\\ \vdots & \vdots\\ 1 & \tilde{x}_n\end{pmatrix},\qquad V=\begin{pmatrix}0 & v_1\\ 0 & v_2\\ \vdots & \vdots\\ 0 & v_n\end{pmatrix}$$

and assuming that

$$\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\right)=\mu$$
$$\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}(\tilde{x}_i-\mu)^2\right)=\sigma_x^2,$$

we have

$$\Sigma_{xx} = \text{plim}\left(\frac{1}{n}\tilde{X}'\tilde{X}\right)$$

$$= \text{plim}\begin{pmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i \\ \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i & \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \mu \\ \mu & \sigma_x^2 + \mu^2 \end{pmatrix}.$$

Also,

$$\Sigma_{vv} = \text{plim}\left(\frac{1}{n}V'V\right)$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}.$$

Now

$$\text{plim}(b-\beta) = -\left(\Sigma_{xx} + \Sigma_{vv}\right)^{-1}\Sigma_{vv}\beta$$

$$\text{plim}\begin{pmatrix} b_0 - \beta_0 \\ b_1 - \beta_1 \end{pmatrix} = -\begin{pmatrix} 1 & \mu \\ \mu & \sigma_x^2 + \mu^2 + \sigma_v^2 \end{pmatrix}^{-1}\begin{pmatrix} 0 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$= -\frac{1}{\left(\sigma_x^2 + \sigma_v^2\right)}\begin{pmatrix} \sigma_x^2 + \mu^2 + \sigma_v^2 & -\mu \\ -\mu & 1 \end{pmatrix}\begin{pmatrix} 0 \\ \beta\sigma_v^2 \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{\sigma_v^2}{\sigma_v^2 + \sigma_x^2}\mu\beta_1 \\ -\dfrac{\sigma_v^2}{\sigma_x^2 + \sigma_v^2}\beta_1 \end{pmatrix}.$$

Thus we find that the OLSEs of β_0 and β_1 are biased and inconsistent. So if a variable is subjected to measurement errors, it not only affects its own parameter estimate but also affect other estimator of parameter that are associated with those variable which are measured without any error. So the presence of measurement errors in even a single variable not only makes the OLSE of its own parameter inconsistent but also makes the estimates of other regression coefficients inconsistent which are measured without any error.

Forms of measurement error model:

Based on the assumption about the true values of the explanatory variable, there are three forms of measurement error model .

Consider the model

$$\tilde{y}_i = \beta_0 + \beta_1 \tilde{x}_i, \qquad i = 1, 2, ..., n$$
$$y_i = \tilde{y}_i + u_i$$
$$x_i = \tilde{x}_i + v_i \ .$$

1. **Functional form:** When the $\tilde{x}_i$'s are unknown constants (fixed), then the measurement error model is said to be in its functional form.

2. **Structural form:** When the $\tilde{x}_i$'s are identically and independently distributed random variables, say, with mean μ and variance $\sigma^2 \left(\sigma^2 > 0\right)$, the measurement error model is said to be in the structural form.

Note that in case of functional form, $\sigma^2 = 0$.

3. **Ultrastructural form**: When the $\tilde{x}_i$'s are independently distributed random variables with different means, say μ_i and variance $\sigma^2 \left(\sigma^2 > 0\right)$, then the model is said to be in the ultrastructural form. This form is a synthesis of function and structural forms in the sense that both the forms are particular cases of ultrastructural form.

Methods for consistent estimation of β:

The OLSE of β which is the best linear unbiased estimator becomes biased and inconsistent in the presence of measurement errors. An important objective in measurement error models is how to obtain the consistent estimators of regression coefficients. The instrumental variable estimation and method of maximum likelihood (or method of moments) are utilized to obtain the consistent estimates of the parameters.

Instrumental variable estimation:

The instrumental variable method provides the consistent estimate of regression coefficients in linear regression model when the explanatory variables and disturbance terms are correlated. Since in measurement error model, the explanatory variables and disturbance are correlated, so this method helps. The instrumental variable method consists of finding a set of variables which are correlated with the explanatory variables in the model but uncorrelated with the composite disturbances, at least asymptotically, to ensure consistency.

Let $Z_1, Z_2, ..., Z_k$ be the k instrumental variables. In the context of the model

$$y = X\beta + \omega, \ \omega = u - V\beta,$$

let Z be the $n \times k$ matrix of k instrumental variables $Z_1, Z_2, ..., Z_k$, each having n observations such that

- Z and $\tilde{X}$ are correlated, atleast asymptotically and
- Z and ω are uncorrelated, at least asymptotically.

So we have

$$\text{plim}\left(\frac{1}{n}Z'X\right) = \Sigma_{ZX}$$
$$\text{plim}\left(\frac{1}{n}Z'\omega\right) = 0.$$

The instrumental variable estimator of β is given by

$$\begin{aligned}
\hat{\beta}_{IV} &= \left(Z'X\right)^{-1} Z'y \\
&= \left(Z'X\right)^{-1} Z'\left(X\beta + \omega\right) \\
\hat{\beta}_{IV} - \beta &= \left(Z'X\right)^{-1} Z'\omega \\
\text{plim}\left(\hat{\beta}_{IV} - \beta\right) &= \text{plim}\left(\frac{1}{n}Z'X\right)^{-1} \text{plim}\left(\frac{1}{n}Z'\omega\right) \\
&= \Sigma_{ZX}^{-1}.0 \\
&= 0.
\end{aligned}$$

So $\hat{\beta}_{IV}$ is consistent estimator of β.

Any instrument that fulfils the requirement of being uncorrelated with the composite disturbance term and correlated with explanatory variables will result in a consistent estimate of parameter. However, there can be various sets of variables which satisfy these conditions to become instrumental variables. Different choices of instruments give different consistent estimators. It is difficult to assert that which choice of instruments will give an instrumental variable estimator having minimum asymptotic variance. Moreover, it is also difficult to decide that which choice of the instrumental variable is better and more appropriate in comparison to other. An additional difficulty is to check whether the chosen instruments are indeed uncorrelated with the disturbance term or not.

Choice of instrument:

We discuss some popular choices of instruments in a univariate measurement error model. Consider the model

$$y_i = \beta_0 + \beta_1 x_i + \omega_i, \quad \omega_i = u_i - \beta_1 v_i, \; i = 1, 2, ..., n.$$

A variable that is likely to satisfy the two requirements of an instrumental variable is the discrete grouping variable. The Wald's, Bartlett's and Durbin's methods are based on different choices of discrete grouping variables.

1. Wald's method

Find the median of the given observations $x_1, x_2, ..., x_n$. Now classify the observations by defining an instrumental variable Z such that

$$Z_i = \begin{cases} 1 & \text{if } x_i > \text{median } (x_1, x_2, ..., x_n) \\ -1 & \text{if } x_i < \text{median } (x_1, x_2, ..., x_n). \end{cases}$$

In this case,

$$Z = \begin{pmatrix} 1 & Z_1 \\ 1 & Z_2 \\ \vdots & \vdots \\ 1 & Z_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

Now form two groups of observations as follows.

- One group with those x_i's below the median of $x_1, x_2, ..., x_n$. Find the means of y_i's and x_i's, say $\bar{y}_1$ and $\bar{x}_1$, respectively in this group..
- Another group with those x_i's above the median of $x_1, x_2, ..., x_n$. Find the means of y_i's and x_i's, say $\bar{y}_2$ and $\bar{x}_2$, respectively in this group.

Now we find the instrumental variable estimator under this set up as follows. Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

$$\hat{\beta}_{IV} = \left(Z'X\right)^{-1} Z'y$$

$$Z'X = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} Z_i & \sum_{i=1}^{n} Z_i x_i \end{pmatrix} = \begin{pmatrix} n & n\bar{x} \\ 0 & \frac{n}{2}\left(\bar{x}_2 - \bar{x}_1\right) \end{pmatrix}$$

$$Z'y = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} Z_i y_i \end{pmatrix} = \begin{pmatrix} n\bar{y} \\ \frac{n}{2}\left(\bar{y}_2 - \bar{y}_1\right) \end{pmatrix}$$

$$\begin{aligned}
\begin{pmatrix} \hat{\beta}_{0IV} \\ \hat{\beta}_{1IV} \end{pmatrix} &= \begin{pmatrix} n & n\bar{x} \\ 0 & \frac{n}{2}\left(\bar{x}_2 - \bar{x}_1\right) \end{pmatrix}^{-1} \begin{pmatrix} n\bar{y} \\ \frac{n}{2}\left(\bar{y}_2 - \bar{y}_1\right) \end{pmatrix} \\
&= \frac{2}{\left(\bar{x}_2 - \bar{x}_1\right)} \begin{pmatrix} \frac{\bar{x}_2 - \bar{x}_1}{2} & -\bar{x} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{y} \\ \frac{\bar{y}_2 - \bar{y}_1}{2} \end{pmatrix} \\
&= \begin{pmatrix} \bar{y} - \left(\frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}\right)\bar{x} \\ \left(\frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}\right) \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \hat{\beta}_{1IV} &= \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1} \\
\hat{\beta}_{0IV} &= \bar{y} - \left(\frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}}\right)\bar{x} = \bar{y} - \hat{\beta}_{1IV}\bar{x}.
\end{aligned}$$

If n is odd, then the middle observations can be deleted. Under fairly general conditions, the estimators are consistent but are likely to have large sampling variance. This is the limitation of this method.

2. Bartlett's method:

Let $x_1, x_2, ..., x_n$ be the *n* observations. Rank these observation and order them in an increasing or decreasing order. Now three groups can be formed, each containing $n/3$ observations. Define the instrumental variable as

$$Z_i = \begin{cases} 1 & \text{if observation is in the top group} \\ 0 & \text{if observation is in the middle group} \\ -1 & \text{if observation is in the bottom group.} \end{cases}$$

Now discard the observations in the middle group and compute the means of y_i's and x_i's in

- bottom group, say $\bar{y}_1$ and $\bar{x}_1$ and
- top group, say $\bar{y}_3$ and $\bar{x}_3$.

Substituting the values of X and Z in $\hat{\beta}_{IV} = (Z'X)^{-1} Z'y$ and on solving, we get

$$\hat{\beta}_{1IV} = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1}$$
$$\hat{\beta}_{0IV} = \bar{y} - \hat{\beta}_{1IV}\bar{x}.$$

These estimators are consistent. No conclusive pieces of evidence are available to compare Bartlett's method and Wald's method but three grouping method generally provides more efficient estimates than two grouping method is many cases.

3. Durbin's method

Let $x_1, x_2, ..., x_n$ be the observations. Arrange these observations in an ascending order. Define the instrumental variable Z_i as the rank of x_i. Then substituting the suitable values of Z and X in $\hat{\beta}_{IV} = (Z'X)^{-1} Z'y$, we get the instrumental variable estimators

$$\hat{\beta}_{1IV} = \frac{\sum_{i=1}^{n} Z_i (y_i - \bar{y})}{\sum_{i=1}^{n} Z_i (x_i - \bar{x})}.$$

When there are more than one explanatory variables, one may choose the instrument as the rank of that particular variable.

Since the estimator uses more information, it is believed to be superior in efficiency to other grouping methods. However, nothing definite is known about the efficiency of this method.

In general, the instrumental variable estimators may have fairly large standard errors in comparison to ordinary least square estimators which is the price paid for inconsistency. However, inconsistent estimators have little appeal.

Maximum likelihood estimation in structural form

Consider the maximum likelihood estimation of parameters in the simple measurement error model given by

$$\begin{aligned}
\tilde{y}_i &= \beta_0 + \beta_1 \tilde{x}_i,\ i = 1,2,...,n \\
y_i &= \tilde{y}_i + u_i \\
x_i &= \tilde{x}_i + v_i.
\end{aligned}$$

Here $(\tilde{x}_i, \tilde{y}_i)$ are unobservable and (x_i, y_i) are observable.

Assume

$$\begin{aligned}
&E(u_i) = 0,\ E(u_i u_j) = \begin{cases} \sigma_u^2 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases} \\
&E(v_i) = 0,\ E(v_i v_j) = \begin{cases} \sigma_v^2 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases} \\
&E(u_i V_j) = 0 \text{ for all } i = 1,2,...,n;\ j = 1,2,...,n.
\end{aligned}$$

For the application of the method of maximum likelihood, we assume the normal distribution for u_i and v_i.

We consider the estimation of parameters in the structural form of the model in which $\tilde{x}_i$'s are stochastic. So assume

$$\tilde{x}_i \sim N(\mu, \sigma^2)$$

and $\tilde{x}_i$'s are independent of u_i and v_i.

Thus

$$E(\tilde{x}_i) = \mu$$
$$Var(\tilde{x}_i) = \sigma^2$$
$$E(x_i) = \mu$$
$$\begin{aligned} Var\left(x_i\right) &= E\left[x_i - E(x_i)\right]^2 \\ &= E\left[\tilde{x}_i + v_i - \mu\right]^2 \\ &= E\left(\tilde{x}_i - \mu\right)^2 + E\left(v_i^2\right) - 2\left(\tilde{x}_i - \mu\right)v_i \\ &= \sigma^2 + \sigma_v^2 \end{aligned}$$
$$\begin{aligned} E\left(y_i\right) &= \beta_0 + \beta_1 E\left(\tilde{x}_i\right) \\ &= \beta_0 + \beta_1\mu. \end{aligned}$$

$$\begin{aligned} Var\left(y_i\right) &= E\left[y_i - E(y_i)\right]^2 \\ &= E\left[\beta_0 + \beta_1\tilde{x}_i + u_i - \beta_0 - \beta_1\mu\right]^2 \\ &= \beta_1^2 E\left(\tilde{x}_i - \mu\right)^2 + E\left(u_i^2\right) - 2\beta_1 E\left(\tilde{x}_i - \mu\right)u_i \\ &= \beta_1^2\sigma^2 + \sigma_u^2 \end{aligned}$$

$$\begin{aligned} Cov\left(x_i, y_i\right) &= E\left[\left\{x_i - E(x_i)\right\}\left\{y_i - E(y_i)\right\}\right] \\ &= E\left[\left\{\tilde{x}_i + v_i - \mu\right\}\left\{\beta_0 + \beta_1\tilde{x}_i + u_i - \beta_0 - \beta_1\mu\right\}\right] \\ &= \beta_1 E\left(\tilde{x}_i - \mu\right)^2 + E\left(\tilde{x}_i - \mu\right)u_i + \beta_1 E\left(\tilde{x}_i - \mu\right)v_i + E\left(u_i v_i\right) \\ &= \beta_1\sigma^2 + 0 + 0 + 0 \\ &= \beta_1\sigma^2. \end{aligned}$$

So

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} \sim N\left[\begin{pmatrix} \beta_0 + \beta_1\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \beta_1^2\sigma^2 + \sigma_u^2 & \beta_1\sigma^2 \\ \beta_1\sigma^2 & \sigma^2 + \sigma_v^2 \end{pmatrix}\right].$$

The likelihood function is the joint probability density function of u_i and v_i, $i = 1, 2, ..., n$ as

$$\begin{aligned} L &= f\left(u_1 u_2, ..., u_n, v_1, v_2, ..., v_n\right) \\ &= \left(\frac{1}{2\pi\sigma_u^2\sigma_v^2}\right)^{n/2} \exp\left[-\frac{\sum_{i=1}^n u_i^2}{2\sigma_u^2}\right] \exp\left[-\frac{\sum_{i=1}^n v_i^2}{2\sigma_v^2}\right] \\ &= \left(\frac{1}{2\pi\sigma_u^2\sigma_v^2}\right)^{n/2} \exp\left[-\frac{\sum_{i=1}^n \left(x_i - \tilde{x}_i\right)^2}{2\sigma_u^2}\right] \exp\left[-\frac{\sum_{i=1}^n \left(y_i - \beta_0 - \beta_1\tilde{x}_i\right)^2}{2\sigma_v^2}\right]. \end{aligned}$$

The log-likelihood is

$$L^* = \ln L = \text{constant} - \frac{n}{2}\left(\ln\sigma_u^2 + \ln\sigma_v^2\right) - \frac{\sum_{i=1}^{n}\left(x_i - \tilde{x}_i\right)^2}{2\sigma_u^2} - \frac{\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right)^2}{2\sigma_v^2}.$$

The normal equations are obtained by equating the partial differentiations equals to zero as

$$(1) \quad \frac{\partial L^*}{\partial\beta_0} = \frac{1}{\sigma_v^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right) = 0$$

$$(2) \quad \frac{\partial L^*}{\partial\beta_1} = \frac{1}{\sigma_v^2}\sum_{i=1}^{n}\tilde{x}_i\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right) = 0$$

$$(3) \quad \frac{\partial L^*}{\partial\tilde{x}_i} = \frac{1}{\sigma_u^2}\left(x_i - \tilde{x}_i\right) + \frac{\beta}{\sigma_v^2}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right) = 0,\ i = 1,2,...,n$$

$$(4) \quad \frac{\partial L^*}{\partial\sigma_u^2} = -\frac{n}{2\sigma_u^2} + \frac{1}{2\sigma_v^4}\sum_{i=1}^{n}\left(x_i - \tilde{x}_i\right)^2$$

$$(5) \quad \frac{\partial L^*}{\partial\sigma_v^2} = -\frac{n}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right)^2.$$

These are $(n+4)$ equations in $(n+4)$ parameters but summing equation (3) over $i = 1,2,...,n$ and using equation (4), we get

$$\sigma_v^2 = \beta_1^2\sigma_u^2$$

which is undesirable.

These equations can be used to estimate the two means $(\mu$ and $\beta_0 + \beta_1\mu)$, two variances and one covariance. The six parameters $\mu, \beta_0, \beta_1, \sigma_u^2, \sigma_v^2$ and σ^2 can be estimated from the following five structural relations derived from these normal equations

(i) $\quad \bar{x} = \mu$

(ii) $\quad \bar{y} = \beta_0 + \beta_1\mu$

(iii) $\quad m_{xx} = \sigma^2 + \sigma_v^2$

(iv) $\quad m_{yy} = \beta_1^2\sigma^2 + \sigma_u^2$

(v) $\quad m_{xy} = \beta_1\sigma^2$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i$, $m_{xx} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$, $m_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$ and $m_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$.

These equations can be derived directly using the sufficiency property of the parameters in bivariate normal distribution using the definition of structural relationship as

$$
\begin{aligned}
E(x) &= \mu \\
E(y) &= \beta_0 + \beta_1 \mu \\
Var(x) &= \sigma^2 + \sigma_v^2 \\
Var(y) &= \beta_1^2 \sigma^2 + \sigma_u^2 \\
Cov(x, y) &= \beta_1 \sigma^2.
\end{aligned}
$$

We observe that there are six parameters $\beta_0, \beta_1, \mu, \sigma^2, \sigma_u^2$ and σ_v^2 to be estimated based on five structural equations (i)-(v). So no unique solution exists. Only μ can be uniquely determined while remaining parameters can not be uniquely determined. So only μ is identifiable and remaining parameters are unidentifiable. This is called the **problem of identification**. One relation is short to obtain a unique solution, so additional *a priori* restrictions relating any of the six parameters is required.

Note: The same equations (i)-(v) can also be derived using the **method of moments.** The structural equations are derived by equating the sample and population moments. The assumption of normal distribution for u_i, v_i and $\tilde{x}_i$ is not needed in case of method of moments.

Additional information for the consistent estimation of parameters:

The parameters in the model can be consistently estimated only when some additional information about the model is available.

From equations (i) and (ii), we have

$$\hat{\mu} = \overline{x}$$

and so μ is clearly estimated. Further

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

is estimated if $\hat{\beta}_1$ is uniquely determined. So we consider the estimation of $\beta_1, \sigma^2, \sigma_u^2$ and σ_v^2 only. Some additional information is required for the unique determination of these parameters. We consider now various type of additional information which are used for estimating the parameters uniquely.

1. σ_v^2 **is known:**

Suppose σ_v^2 is known a *priori*. Now the remaining parameters can be estimated as follows:

$$
\begin{aligned}
m_{xx} &= \sigma^2 + \sigma_v^2 \Rightarrow \hat{\sigma}^2 = m_{xx} - \sigma_v^2 \\
m_{xy} &= \beta_1 \sigma^2 \Rightarrow \hat{\beta}_1 = \frac{m_{xy}}{m_{xx} - \sigma_v^2} \\
m_{yy} &= \beta_1 \sigma^2 + \sigma_u^2 \Rightarrow \hat{\sigma}_u^2 = m_{yy} - \hat{\beta}_1^2 \hat{\sigma}^2 \\
&\qquad\qquad\qquad\qquad = m_{yy} - \frac{m_{xy}^2}{m_{xx} - \sigma_v^2}.
\end{aligned}
$$

Note that $\hat{\sigma}^2 = m_{xx} - \sigma_v^2$ can be negative because σ_v^2 is known and m_{xx} is based upon sample. So we assume that $\hat{\sigma}^2 > 0$ and redefine

$$
\hat{\beta}_1 = \frac{m_{xy}}{m_{xx} - \sigma_v^2};\ m_{xx} > \sigma_v^2.
$$

Similarly, $\hat{\sigma}_u^2$ is also assumed to be positive under suitable condition. All the estimators $\hat{\beta}_1, \hat{\sigma}^2$ and $\hat{\sigma}_u^2$ are the consistent estimators of β, σ^2 and σ_u^2 respectively. Note that $\hat{\beta}_1$ looks like as if the direct regression estimator of β_1 has been adjusted by σ_v^2 for its inconsistency. So it is termed as adjusted estimator also.

2. σ_u^2 **is known**

Suppose σ_u^2 is known *a priori*. Then using $m_{xy} = \beta_1 \sigma^2$, we can rewrite

$$
\begin{aligned}
m_{yy} &= \beta_1^2 \sigma^2 + \sigma_u^2 \\
&= m_{xy} \beta_1 + \sigma_u^2 \\
\Rightarrow \hat{\beta}_1 &= \frac{m_{yy} - \sigma_u^2}{m_{xy}};\ m_{yy} > \sigma_u^2 \\
\hat{\sigma}^2 &= \frac{m_{xy}}{\hat{\beta}_1} \\
\hat{\sigma}_v^2 &= m_{xx} - \hat{\sigma}^2.
\end{aligned}
$$

The estimators $\hat{\beta}_1, \hat{\sigma}^2$ and $\hat{\sigma}_v^2$ are the consistent estimators of β_1, σ^2 and σ_v^2 respectively. Note that $\hat{\beta}_1$ looks like as if the reverse regression estimator of β_1 is adjusted by σ_u^2 for its inconsistency. So it is termed as adjusted estimator also.

3. $\lambda = \frac{\sigma_u^2}{\sigma_v^2}$ is known

Suppose the ratio of the measurement error variances is known, so let

$$\lambda = \frac{\sigma_u^2}{\sigma_v^2}$$

is known.

Consider

$$\begin{aligned} m_{yy} &= \beta_1^2\sigma^2 + \sigma_u^2 \\ &= \beta_1 m_{xy} + \lambda\sigma_v^2 && \text{(using (iv))} \\ &= \beta_1 m_{xy} + \lambda\left(m_{xx} - \sigma^2\right) && \text{(using (iii))} \\ &= \beta_1 m_{xy} + \lambda\left(m_{xx} - \frac{m_{xy}}{\beta_1}\right) && \text{(using iv)} \end{aligned}$$

$$\beta_1^2 m_{xy} + \lambda\left(\beta_1 m_{xx} - m_{xy}\right) - \beta_1 m_{yy} = 0 \quad \left(\beta_1 \neq 0\right)$$

$$\beta_1^2 m_{xy} + \beta\left(\lambda m_{xx} - m_{yy}\right) - \lambda m_{xy} = 0.$$

Solving this quadratic equation

$$\begin{aligned} \hat{\beta}_1 &= \frac{\left(m_{yy} - \lambda m_{xx}\right) \pm \sqrt{\left(m_{yy} - \lambda m_{xx}\right)^2 + 4\lambda m_{xy}^2}}{2m_{xy}} \\ &= \frac{U}{2m_{xy}}, \text{ say.} \end{aligned}$$

Since $m_{xy} = \beta_1\sigma^2$ and

$$\begin{aligned} &\hat{\sigma}^2 \geq 0 \\ \Rightarrow &\frac{m_{xy}}{\hat{\beta}_1} \geq 0 \\ \Rightarrow &\frac{2m_{xy}^2}{U} \geq 0 \\ \Rightarrow &\text{ since } m_{xy}^2 \geq 0, \text{ so } U \text{ must be nonnegative.} \end{aligned}$$

This implies that the positive sign in U has to be considered and so

$$\hat{\beta}_1 = \frac{\left(m_{yy} - \lambda m_{xx}\right) + \sqrt{\left(m_{yy} - \lambda m_{xx}\right)^2 + 4\lambda m_{xy}^2}}{2m_{xy}}.$$

Other estimates are

$$\hat{\sigma}_v^2 = \frac{m_{yy} - 2\hat{\beta}_1 m_{xy} + \hat{\beta}_1^2 s_{xx}}{\lambda + \hat{\beta}_1^2}$$

$$\hat{\sigma}^2 = \frac{m_{xy}}{\hat{\beta}_1}.$$

Note that the same estimator $\hat{\beta}_1$ of β_1 can be obtained by orthogonal regression. This amounts to transform x_i by x_i / σ_u and y_i by y_i / σ_v and use the orthogonal regression estimation with transformed variables.

4. Reliability ratio is known

The reliability ratio associated with the explanatory variable is defined as the ratio of variances of true and observed values of explanatory variables, so

$$K_x = \frac{Var(\tilde{x})}{Var(x)} = \frac{\sigma^2}{\sigma^2 + \sigma_v^2};\ 0 \le K_x \le 1$$

is the reliability ratio. Note that $K_x = 1,$ when $\sigma_v^2 = 0$ which means that there is no measurement error in the explanatory variable and $K_x = 0,$ means $\sigma^2 = 0$ which means the explanatory variable is fixed. A higher value of K_x is obtained when σ_v^2 is small, i.e., the impact of measurement errors is small. The reliability ratio is a popular measure in psychometrics.

Let K_x be known *a priori*. Then

$$m_{xx} = \sigma^2 + \sigma_v^2$$

$$m_{xy} = \beta_1 \sigma^2$$

$$\Rightarrow \frac{m_{xy}}{m_{xx}} = \frac{\beta_1 \sigma^2}{\sigma^2 + \sigma_v^2}$$

$$= \beta_1 K_x$$

$$\Rightarrow \hat{\beta}_1 = \frac{m_{xy}}{K_x m_{xx}}$$

$$\sigma^2 = \frac{m_{xy}}{\beta_1}$$

$$\Rightarrow \hat{\sigma}^2 = K_x m_{xx}$$

$$m_{xx} = \sigma^2 + \sigma_v^2$$

$$\Rightarrow \hat{\sigma}_v^2 = \left(1 - K_x\right) m_{xx}.$$

Note that $\hat{\beta}_1 = K_x^{-1} b$

where b is the ordinary least squares estimator $b = \dfrac{m_{xy}}{m_{xx}}$.

5. β_0 is known

Suppose β_0 is known *a priori* and $E(x) \neq 0.$ Then

$$
\begin{aligned}
\bar{y} &= \beta_0 + \beta_1 \mu \\
\Rightarrow \hat{\beta}_1 &= \frac{\bar{y} - \beta_0}{\hat{\mu}} \\
&= \frac{\bar{y} - \beta_0}{\bar{x}} \\
\hat{\sigma}^2 &= \frac{m_{xy}}{\hat{\beta}_1} \\
\hat{\sigma}_u^2 &= m_{yy} - \hat{\beta}_1 m_{xy} \\
\hat{\sigma}_v^2 &= m_{xx} - \frac{m_{xy}}{\hat{\beta}_1}.
\end{aligned}
$$

6. Both σ_u^2 and σ_v^2 are known

This case leads to over-identification in the sense that the number of parameters to be estimated is smaller than the number of structural relationships binding them. So no unique solutions are obtained in this case.

Note: In each of the cases 1-6, note that the form of the estimate depends on the type of available information which is needed for the consistent estimator of the parameters. Such information can be available from various sources, e.g., long association of the experimenter with the experiment, similar type of studies conducted in the part, some external source etc.

Estimation of parameters in function form:

In the functional form of the measurement error model, $\tilde{x}_i 's$ are assumed to be fixed. This assumption is unrealistic in the sense that when $\tilde{x}_i 's$ are unobservable and unknown, it is difficult to know if they are fixed or not. This can not be ensured even in repeated sampling that the same value is repeated. All that can be said in this case is that the information, in this case, is conditional upon $\tilde{x}_i 's$. So assume that $\tilde{x}_i 's$ are conditionally known. So the model is

$$
\begin{aligned}
\tilde{y}_i &= \beta_0 + \beta_1 \tilde{x}_i \\
x_i &= \tilde{x}_i + v_i \\
y_i &= \tilde{y}_i + u_i
\end{aligned}
$$

then

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} \sim N\left[\begin{pmatrix} \beta_0 + \beta_1 \tilde{x}_i \\ \tilde{x}_i \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}\right].$$

The likelihood function is

$$L = \left(\frac{1}{2\pi\sigma_u^2}\right)^{\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2}{2\sigma_u^2}\right]\left(\frac{1}{2\pi\sigma_v^2}\right)^{\frac{n}{2}} \exp\left[-\frac{\sum\left(x_i - \tilde{x}_i\right)^2}{2\sigma_v^2}\right].$$

The log-likelihood is

$$L^* = \ln L = \text{constant} - \frac{n}{2}\ln\sigma_u^2 - \frac{\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right)^2}{2\sigma_u^2} - \frac{n}{2}\ln\sigma_v^2 - \frac{1}{2\sigma_v^2}\sum_{i=1}^{n}\left(x_i - \tilde{x}_i\right)^2.$$

The normal equations are obtained by partially differentiating L^* and equating to zero as

$$(I) \qquad \frac{\partial L^*}{\partial \beta_0} = 0 \Rightarrow \frac{1}{\sigma_u^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right) = 0$$

$$(II) \qquad \frac{\partial L^*}{\partial \beta_1} = 0 \Rightarrow \frac{1}{\sigma_v^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right)\tilde{x}_i = 0$$

$$(III) \qquad \frac{\partial L^*}{\partial \sigma_u^2} = 0 \Rightarrow -\frac{n}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right)^2 = 0$$

$$(IV) \qquad \frac{\partial L^*}{\partial \sigma_v^2} = 0 \Rightarrow -\frac{n}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}\sum_{i=1}^{n}\left(x_i - \tilde{x}_i\right)^2 = 0$$

$$(V) \qquad \frac{\partial L^*}{\partial \tilde{x}_i} = 0 \Rightarrow \frac{\beta}{\sigma_u^2}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right) + \frac{1}{\sigma_v^2}(x_i - \tilde{x}_i) = 0.$$

Squaring and summing equation (*V*), we get

$$\sum_i\left[\frac{\beta_1}{\sigma_u^2}\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right)\right]^2 = \sum_i\left[-\frac{1}{\sigma_v^2}\left(x_i - \tilde{x}_i\right)\right]^2$$

or $\displaystyle \frac{\beta_1^2}{\sigma_u^2}\sum_i\left(y_i - \beta_0 - \beta_1\tilde{x}_i\right)^2 = \frac{1}{\sigma_v^4}\sum_i\left(x_i - \tilde{x}_i\right)^2.$

Using the left-hand side of equation (*III*) and right-hand side of equation (*IV*), we get

$$\frac{n\beta_1^2}{\sigma_u^2} = \frac{n}{\sigma_v^2}$$

$$\Rightarrow \beta_1 = \frac{\sigma_u}{\sigma_v}$$

which is unacceptable because β can be negative also. In the present case, as $\sigma_u > 0$ and $\sigma_v > 0$, so β will always be positive. Thus the maximum likelihood breaks down because of insufficient information in the model. Increasing the sample size n does not solve the purpose. If the restrictions like σ_u^2 known, σ_v^2 known or $\frac{\sigma_u^2}{\sigma_v^2}$ known are incorporated, then the maximum likelihood estimation is similar to as in the case of structural form and the similar estimates may be obtained. For example, if $\lambda = \sigma_u^2 / \sigma_v^2$ is known, then substitute it in the likelihood function and maximize it. The same solution as in the case of structural form are obtained.