Measurement error models:

jointly works with Prof. Heleno Bolfarine

Alexandre Galvão Patriota

Departamento de Estatística
Instituto de Matemática e Estatística
Universidade de São Paulo

YoutTube: “A Ciência da Estatística”

Research fields

Master (2005-2006) and Doctoral (2006-2010) degrees:

- **Measurement error models** under Bolfarine's supervision.

Independent research (after 2010):

- Regression models with general parameterization.
- Foundations of Probability and Statistics.
- Asymptotic theory.

Measurement error models

In this presentation, I discuss 5 works published with Professor Heleno Bolfarine:

1. A heteroscedastic structural errors-in-variables model with equation error (2009). Statistical Methodology.
2. A heteroscedastic polynomial regression with measurement error in both axes (2008). Sankhya.
3. Measurement error models with a general class of error distribution (2010). Statistics.
4. A multivariate ultrastructural errors-in-variables model with equation error (2011). JMA.
5. Improved maximum likelihood estimators in a heteroskedastic errors-in-variables model (2011). Statistical papers.

Heteroscedastic errors-in-variables: linear model

Heteroscedastic errors-in-variables: linear model

The model was proposed by Kulathinal et al. (2002):

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + q_i \\ Y_i &= y_i + e_i \\ X_i &= x_i + u_i \end{aligned}$$

The random quantities are distributed as

$$\begin{pmatrix} u_i \\ e_i \\ q_i \\ x_i \end{pmatrix} \overset{ind}{\sim} \mathcal{N}_4 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ \mu_x \end{pmatrix}, \begin{pmatrix} \tau_{ui} & 0 & 0 & 0 \\ 0 & \tau_{e_i} & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma_x^2 \end{pmatrix} \right),$$

where τ_{ui} and τ_{ei} are known for $i = 1, \ldots, n$.

Results: linear model

This model is applied in

- **epidemiology** (y=cardiovascular mortality index; x=risk factors index)
- **astrophysics** (y=density of Black Holes; x=Accretion disk luminosity)

Results:

1. we computed the asymptotic variance of the MM and ML estimators.
2. we compared Wald test statistics via MC simulations under both approaches.
3. we concluded that both methods are robust against the distribution of covariate x_i.

Heteroscedastic errors-in-variables: polynomial model

Heteroscedastic errors-in-variables: polynomial model

The model is given by (Zavala et al., 2007, with no equation error):

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \ldots + \beta_p x_i^p + q_i \\ Y_i &= y_i + e_i \\ X_i &= x_i + u_i \end{aligned}$$

where x_i, $i = 1, \ldots, n$, are incidental parameters. The random quantities are distributed as

$$\begin{pmatrix} u_i \\ e_i \\ q_i \end{pmatrix} \stackrel{ind}{\sim} \mathcal{N}_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{ui} & 0 & 0 \\ 0 & \tau_{e_i} & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \right),$$

where τ_{ui} and τ_{ei} are known for $i = 1, \ldots, n$.

Results: polynomial model

This model is also applied in

- **epidemiology** (y=cardiovascular mortality index; x=risk factors index)
- **astrophysics** (y=density of Black Holes; x=Accretion disk luminosity)

Results:

1. we computed the consistent estimators based on the corrected score approach.
2. we study Wald test statistics via MC simulations.
3. we apply the methods to a epidemiology (Kulathinal) data and astrophysics data (quadratic and cubic regressions).

Errors-in-variables with a general class of error distribution

Errors-in-variables: general class of error distributions

Let $(Y_i, \boldsymbol{W}_i, \boldsymbol{X}_i)$ be observable vector related by

$$\begin{aligned} Y_i &= \boldsymbol{\beta}^\top \boldsymbol{W}_i + \boldsymbol{\gamma}^\top \boldsymbol{x}_i + e_i, \\ \boldsymbol{X}_i | \boldsymbol{x}_i &\overset{ind}{\sim} F_{\boldsymbol{X}_i|\boldsymbol{x}_i} \in \mathcal{C}(\boldsymbol{x}_i, g_1, g_2), \end{aligned}$$

where $\boldsymbol{x}_i$, $i = 1, \ldots, n$, are incidental parameters and the functions $g_1(.)$ and $g_2(.)$ are known and must satisfy the following two conditions

$$E[g_1(\boldsymbol{X}_i)|\boldsymbol{x}_i] = \boldsymbol{x}_i \quad \text{and} \quad E[g_2(\boldsymbol{X}_i)|\boldsymbol{x}_i] = \boldsymbol{x}_i \boldsymbol{x}_i^\top$$

Remarks

Application: sleep study, where
y = systolic blood pressure;
x = apnea-hypopnea index,
W = body mass index.

- We use the corrected score method proposed by Nakamura (1990) to conduct inferences about the parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and σ^2.
- It is not necessary to know the shape of $F_{\boldsymbol{X}_i|\boldsymbol{x}_i}$,
- It is only required to know the shape of g_1 and g_2 to employ this methodology.

Next we present same examples of g_1 and g_2.

Particular cases

Normal distribution

Assume that $X_i|x_i \sim \mathsf{N}(x_i, \phi)$, where $\phi > 0$ is known. Then,

- $E(X_i|x_i) = x_i$ and $E(X_i^2|x_i) = \phi + x_i^2$
- $g_1(X_i) = X_i$ and $g_2(X_i) = X_i^2 - \phi$.

This structure is the same as the one attained from: $X_i = x_i + u_i$, where $u_i \sim N(0, \phi)$.

Notice that any distribution $F_{X_i|x_i}$ that yields the same g_1 and g_2 as above is such that $F_{X_i|x_i} \in \mathcal{C}(x_i, g_1, g_2)$.

Poisson distribution

Assume that $X_i|x_i \sim \text{Poisson}(x_i)$. Then,

- $E(X_i|x_i) = x_i$ and $E(X_i^2 - X_i|x_i) = x_i^2 \qquad Var(X_i|x_i) = x_i$
- $g_1(X_i) = X_i$ and $g_2(X_i) = X_i^2 - X_i$.

Notice that any distribution $F_{X_i|x_i}$ such that

$$E(X_i|x_i) = Var(X_i|x_i) = x_i$$

produces the same g_1 and g_2 as above is such that $F_{X_i|x_i} \in \mathcal{C}(x_i, g_1, g_2)$.

Multiplicative normal model or Gamma distribution

Assume $X_i|x_i \sim \mathcal{N}(x_i, x_i^2\phi)$, with $\phi > 0$ known, then

- $E(X_i|x_i) = x_i$ and $E(X_i^2|x_i) = (\phi + 1)x_i^2$
- $g_1(X_i) = X_i$ and $g_2(X_i) = X_i^2/(\phi + 1)$.

It is the multiplicative model: $X_i = x_i u_i$, where $u_i \sim N(1, \phi)$.

Notice that $X_i|x_i \sim \text{Gamma}(x_i, \phi)$, where $E(X_i|x_i) = x_i$ and $Var(X_i|x_i) = x_i^2\phi$ also yields the same functions above. This gamma distribution is a reparameterization of the usual version.

That is, $\mathcal{N}(x_i, x_i^2\phi), \text{Gamma}(x_i, \phi) \in \mathcal{C}(x_i, g_1, g_2)$

A multivariate ultrastructural errors-in-variables

A multivariate ultrastructural errors-in-variables

Let $(\boldsymbol{Y}_i, \boldsymbol{X}_i)$ random vectors related by

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{q}_i, \\ \boldsymbol{Y}_i &= \boldsymbol{y}_i + \boldsymbol{e}_i, \\ \boldsymbol{X}_i &= \boldsymbol{x}_i + \boldsymbol{u}_i, \end{aligned}$$

The errors $\boldsymbol{q}_i$, $\boldsymbol{e}_i$ and $\boldsymbol{u}_i$ are simetrically distributed around zero.

$$\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{q}_i \\ \boldsymbol{e}_i \\ \boldsymbol{u}_i \end{pmatrix} \sim \left(\begin{pmatrix} \boldsymbol{\xi}_i \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{\boldsymbol{x}} & 0 & 0 & 0 \\ 0 & \Sigma_{\boldsymbol{q}} & 0 & 0 \\ 0 & 0 & \Sigma_{\boldsymbol{e}} & 0 \\ 0 & 0 & 0 & \Sigma_{\boldsymbol{u}} \end{pmatrix} \right)$$

where $\boldsymbol{\xi}_i$, $i = 1, \ldots, n$, are incidental parameters.

Some theoretical results

We consider that $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_u$ are known matrices.

Results:

1. we established some regular conditions to attain consistent estimators for $\boldsymbol{a}$ and $\boldsymbol{B}$.
2. we computed the asymptotic normality of the proposed estimators of $\boldsymbol{a}$ and $\boldsymbol{B}$.
3. we specialized the results to the eliptical class of distributions and univariate regressions
4. we showed that previous results (Cheng and Van Ness, 1991, Arellano-Valle et al, 1996) are particular instances of ours.

Improved maximum likelihood estimators

The model is given by

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_i + \boldsymbol{q}_i \\ \boldsymbol{Y}_i &= \boldsymbol{y}_i + \boldsymbol{\eta}_{\boldsymbol{y}_i} \\ \boldsymbol{X}_i &= \boldsymbol{x}_i + \boldsymbol{\eta}_{\boldsymbol{x}_i} \end{aligned}$$

where the errors are normally distributed as

$$\begin{pmatrix} \boldsymbol{\eta}_{\boldsymbol{y}_i} \\ \boldsymbol{\eta}_{\boldsymbol{x}_i} \end{pmatrix} \overset{ind}{\sim} \mathcal{N}_{v+m} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\tau}_{\boldsymbol{y}_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\tau}_{\boldsymbol{x}_i} \end{pmatrix} \right],$$

The variances matrices $\boldsymbol{\tau}_{\boldsymbol{x}_i}$ and $\boldsymbol{\tau}_{\boldsymbol{x}_i}$ are assumed to be known.

Results:

1. We computed the second order biases of the MLE,
2. We proposed bias-corrected estimators ,
3. We conducted MC simulations to verify if the corrected estimators have smaller biases.

References:

1. Patriota, AG; **BOLFARINE**, H. A heteroscedastic polynomial regression with measurement error in both axes. Sankhya. Series B, 70, 267-282, 2008.
2. Patriota, AG; **BOLFARINE**, H. ; de Castro, M. A heteroscedastic structural errors-in-variables model with equation error. Statistical Methodology, 6, 408–423, 2009.
3. Patriota, AG; **BOLFARINE**, H. Measurement error models with a general class of error distribution. Statistics, 44, 119–127, 2010.
4. Patriota, AG; Lemonte, AJ ; **BOLFARINE**, H. Improved maximum likelihood estimators in a heteroskedastic errors-in-variables model. Statistical Papers, 52, 455–467, 2011.
5. Patriota, AG; **BOLFARINE**, H; Arellano-Valle, RB. A multivariate ultrastructural errors-in-variables model with equation error. Journal of Multivariate Analysis, 102, 386–392, 2011.

Thank You