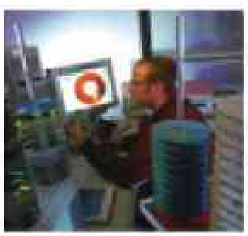
Laboratoire national de métrologie et d’essais


LNE

Le progrès, une passion à partager

Uncertainty of standard concentrations,

Calibration Line and Predictions

Catherine YARDIN
Sacko KABA

LNE
Le progrès, une passion à partager

24 June 2008

Ordinary Least Squares (OLS)

- Model $$Y_t = a + b.X_t + v_t$$

  Y_t = signal X_t = concentration v_t = error

- Main hypothesis :

  X_t fixed variables,

  v_t random variables with zero mean value, non correlated and constant standard deviation σ.

  Y_t random variables with $(a + b X_t)$ mean value and standard deviation σ .

- Estimator
  intercept a, slope b, standard deviation σ (experimental error s)
  prediction at $X_t = X_0$: signal Y_0, $u(Y_0)$

Ordinary Least Squares (OLS)

- **Properties**

  Best Linear Unbiased Estimator

- **Inconvenients**

  1) The results differ when the axes are switched :

  Regression $$Y_t = a + b.X_t + v_t$$

  Inverse Regression $$X_t = a' + b'.Y_t + \mu_t$$

  $$a' \neq \frac{-a}{b} \qquad b' \neq \frac{1}{b}$$

  2) The OLS estimator does not take into account the uncertainty of the standard.

  In fact, OLS is biased : the slope b is under – estimated.

Ordinary Least Squares (OLS)

3) The OLS estimator is not the minimum variance estimator when the errors v_t are correlated or have different variances

The Generalised Least Squares Estimator (GLS) is the best estimator.
(ex : Weighted Least squares estimator is GLS estimator)

- Two methods are presented after :
  - Method from « Error in variables model »
  - Pratical method : Augmented – u

Error in variables model

- Unobservable true variables

$$y_t = a + b.x_t \qquad (1)$$

Observed variables

$$Y_t = a + b.X_t + v_t \qquad (2)$$

$$\text{where} \quad Y_t = y_t + e_t$$

$$X_t = x_t + u_t \qquad v_t = e_t - b.u_t$$

- Hypothesis :
  - true variables x_t
  - errors e_t and u_t
  - relation (1)

Error in variables model

- True variables x_t : random or fixed value

- Error e_t or u_t :

  zero mean

  non autocorrelated

  homoscedasticity or heteroscedasticity

- Correlation between variables :

  no correlation between x_t and u_t , y_t and e_t, e_t and u_t

- model (1) with or without equation error q_t

Metrological context

- Type of variables x_t : fixed variables
  the model is functionnal (instead of structural).

- Hypothesis on the errors :
  - homoscedasticity or heteroscedasticity
  - errors e_t and u_t are not correlated

- Errors u_t associated with the standards are autocorrelated.

Method of estimation

- Many methods :
method of maximum likelihood, method of moments, instrumental variables, …

- Some methods are best for some type of model, hypothesis and available information.

- Two methods are presented here :
  . BLS (for error in variables model)
  . Augmented – u (pratical method)

Bivariate Least Squares (BLS)

- **Weighted regression**

  residual : $e_t = Y_t - (a + bX_t)$

  variance : $w_t = s_{et}^2 = s_{Yt}^2 + b^2 s_{Xt}^2 - 2b * \mathrm{cov}(X_t, Y_t)$

BLS minimize this weighted sum of squares residuals :

$$\sum_t \frac{e_t^2}{w_t}$$

Bivariate Least Squares (BLS)

- **Estimates of the parameters are solutions of the non linear system :**

$$RB = g \Leftrightarrow \begin{vmatrix} \sum \dfrac{1}{s_{et}^2} & \sum \dfrac{x_t}{s_{et}^2} \\ \sum \dfrac{x_t}{s_{et}^2} & \sum \dfrac{x_t^2}{s_{et}^2} \end{vmatrix} \times \begin{vmatrix} a \\ b \end{vmatrix} = \begin{vmatrix} \sum \left(\dfrac{y_t}{s_{et}^2} + \dfrac{1}{2}\left(\dfrac{e_t}{s_{et}^2}\right)^2 \dfrac{\partial s_{et}^2}{\partial a} \right) \\ \sum \left[\dfrac{x_t * y_t}{s_{et}^2} + \dfrac{1}{2}\left(\dfrac{e_t}{s_{et}^2}\right)^2 \dfrac{\partial s_{et}^2}{\partial b} \right] \end{vmatrix}$$

- **After simplification**

$$\begin{vmatrix} \sum \frac{1}{s_{et}^2} & \sum \frac{x_t}{s_{et}^2} \\ \sum \frac{x_t}{s_{et}^2} & \sum \frac{x^2{}_t}{s_{et}^2} \end{vmatrix} \times \begin{vmatrix} a \\ b \end{vmatrix} = \begin{vmatrix} \sum \left(\frac{y_t}{s_{et}^2} \right) \\ \sum \left[\frac{x_t * y_t}{s_{et}^2} + \left(\frac{e_t}{s_{et}^2} \right)^2 * b * s_{X_t}^2 \right] \end{vmatrix}$$

Bivariate Least Squares (BLS)

- **Advantages :**
  - no sense of regression
  - heteroscedascticity

- **Properties : equivalent to the estimator of the likelihood**

- **Inconvenients :**
  - information about the errors
  - depends on the quality of this information

« Augmented – u » Method

- The line is estimated by the OLS, but the uncertainty of the line is augmented by the uncertainty of the standard
- Prediction :

$$Y_0 = a + b.X_0$$

$$u^2(Y_0) = u^2(a) + (X_0)^2 u^2(b) + 2X_0 u(a,b) + b^2.u^2(x_0)$$

The last term represents the uncertainty due to the standard.

« Augmented – u » Method

- <u>Principle :</u>

  . Augment the variance to compensate the bias of the OLS method.

  . Consequently, the confidence interval around the calibration line is more high : prediction Y_o and calibration X_o have a larger calculated uncertainty.

- <u>Inconvenient</u>

  . Biased like OLS

  . Perhaps, large variance

- <u>Advantages</u>

  Simple method

Ex 1 : calibration of a radiometer

- Radiometer : What is the tension at the heat flux 2,5 W/cm² ?
- Calibration with a calorimeter.
- Calibration process : 4 points below and beside the flux aim 2,5 Wcm²
- Calibration line of the radiometer :

$$U\ (mv) = a + b * Q$$

Ex 1 : calibration of a radiometer

Data for calibration

- 4 flux encadrant la valeur cible : Q = 2,5 Wcm²

Flux Q	u(Q)	Tension U	u(U)
2.209859	0.039	7.505937	0.005
2.389225	0.042	8.020660	0.005
2.558037	0.045	8.433618	0.005
2.749269	0.049	8.902940	0.005

- Standard : $Q_i = a_i * Z$

  Z is the coefficient of the calorimeter

Ex 1 : calibration of a radiometer

- **Parameters of the calibration line**

500 000 simulations

Calibration line	OLS Augmented-	BLS u(Y) MCO	BLS u(Y)	MC without $u(Q_i,Q_j)$	MC with $u(Q_i,Q_j)$	Fuller reg inverse
intercept	1.837	1.79593	1.7931555	1.91	1.84	1.8271
u(intercept)	0.182	0.18659	0.18684	0.7	0.01	
slope	2.576	2.59237	2.5934979	2.55	2.58	2.5935
u(slope)	0.073	0.0761	0.0762	0.29	0.05	

Ex 1 : calibration of a radiometer

Predictions and calibrations

500 000 simulations

Prediction	OLS	Augmented-u	BLS u(Y) OLS	BLS u(Y)	MC without $u(Q_i,Q_j)$	MC with $u(Q_i,Q_j)$
U0	8.28	8.28	8.28	8.28	8.28	8.28
u(U0)	0.015	0.115	0.114	0.115	0.06	0.11
Q0	2.5000	2.5000	2.5012	2.50	2.5	2.5
u(Q0)	0.028	0.045	0.044	0.044	0.02	0.04

. All methods give same predictive values for U_0 and Q_0

. But the uncertainties depend on the method used :
OLS and Monte carlo without correlation under-estimate uncertainties.

Ex 2 : measuring mass of Benzene

Parameters of the calibration line

Calibration line	OLS Augmented-u	BLS u(Y) OLS	BLS u(Y) = 0.98%	500 000 itérations MC without $u(Q_i,Q_j)$	500 000 itérations MC with $u(Q_i,Q_j)$	Fuller reg inverse
intercept	16582.17	17355.04	19556.93	20347	16564.27	12865.65
u(intercept)	22223	22004	21729	29260	23989	
slope	1783.88	1782.38	1777.96	1776.51	1784.16	1791.20
u(slope)	43.18	43.78	44.31	59.61	52.40	

Ex 2 : measuring mass of Benzene

Predictions and calibrations

Prediction	OLS	Augmented-u	BLS u(Y) OLS	BLS u(Y) = 0.98%	500 000 itérations MC without $u(X_i,X_j)$	500 000 itérations MC with $u(X_i,X_j)$
y0 u(Y0)	916278 3584	916278 12080	916278 12234	916278 12226	916325 4871.7	916399 11319.5
x0 u(X0)	504.3 2.4	504.3 6.9	504.3373 6.86	504.3529 6.88	504.3 2.7	504.4 6.3

Prediction :

x = 504.4 ng - u(x) = 6.9 ng

Real value :

x = 501.0 ng - u(x) = 1.3 ng

Conclusion

- We have used methods which permit to take into account uncertainty of the standards : during the estimation of the calibration function and in the predictions.
- OLS very often give very low uncertainties (of the predictions).
- the best method are methods which include all information about the variables : heteroscedasticity of the errors, covariances of the standards.

Conclusion

- In the future, we will study methods for taking into account covariances of the standards.
  Perhaps, these methods will be like a mixture of GLS and BLS.

- The auhors have also studied non linear model. The second order polynomial is currently used in some analysis methods.

Bibliographic

- C.L. Cheng and J.W. Van Ness, "Statistical regression with measurement error", London, Oxford university press (1999).

- F. Javier del Rio, Jordi Riu and F. Xavier Rius, Journal of chemometrics, 15, 773 - 788, "Prediction intervals in linear regression taking into account errors of both axes" (2001).

- C.L. Cheng, J. Riu, Technometrics, VOL. 48,NO 4, "On estimating linear relationships when both variables are subject to heteroscedastic measurement errors" (2006).