

# Estimer la droite d'étalonnage avec les moindres carrés généralisés et évaluer le résultat de mesure

## *Estimating the straight-line calibration function with generalized least squares and evaluating the measurement result*

Catherine YARDIN

Laboratoire national de métrologie et d'essais (LNE), 1 rue Gaston Boissier, 75724 Paris Cedex 15, France, catherine.yardin@lne.fr.

### Résumé

La méthode des moindres carrés généralisés (GLS) estime la droite d'étalonnage en tenant compte de l'incertitude associée aux variables, de l'hétéroscédasticité et des corrélations. Mais l'efficacité des GLS requiert une bonne connaissance de la méthode elle-même et du processus d'étalonnage. Cet article précise les caractéristiques de la fonction d'étalonnage et introduit le « modèle à erreurs sur les variables » plus représentatif du contexte. La méthode GLS est présentée selon différents scénarios d'incertitude en considérant les trois phases de l'estimation : modélisation, évaluation et validation. La propagation de l'incertitude des variables aux coefficients de la droite est discutée. Puis, le résultat de mesure est calculé avec la droite estimée, en mode direct et en mode inverse. Son incertitude est donnée en distinguant la valeur vraie ou moyenne sans incertitude et une valeur mesurée ou individuelle avec incertitude. Deux exemples sont analysés. Les incertitudes des variables influencent peu la valeur des coefficients de la droite et du résultat. En revanche, les incertitudes associées sont très affectées.

**MOTS CLÉS** : MOINDRES CARRÉS GÉNÉRALISÉS, MODÈLE À ERREURS SUR LES VARIABLES, DROITE D'ÉTALONNAGE, DROITE AVEC INCERTITUDE SUR LES DEUX VARIABLES.

### Abstract

*The Generalized Least Squares (GLS) method estimates the straight-line calibration function by taking into account uncertainties in variables, heteroscedasticity and correlations. However, the efficiency of the method depends on a good knowledge of the method itself and of the calibration process. This paper details the characteristics of the*

*calibration function and introduces the “model with errors in variables” which is more representative of the context. The GLS method is analysed in the three steps of the estimation process: assumption, computation and validation. Propagation of the uncertainty in input data to coefficients of the straight-line calibration function is also discussed. The measurement result is then computed with the estimated curve, directly and by inverse evaluation. The associated uncertainty is given by distinguishing the mean value without uncertainty and an individual value with its associated uncertainty. Two examples are analysed. Uncertainties in variables have little influence on the value of the measurement result and of coefficients of the straight-line calibration function. On the contrary, associated uncertainties are highly affected by them.*

**KEY WORDS**: GENERALISED LEAST SQUARES, MEASUREMENT ERRORS MODEL, CALIBRATION CURVE, STRAIGHT-LINE WITH UNCERTAINTY ON VARIABLES.

### 1. Introduction

L'étalonnage permet de convertir une indication en un résultat de mesure. Dans le cas d'un instrument de mesure utilisé sur une plage de valeurs, cette conversion est réalisée par une fonction d'étalonnage. L'indication peut être conservée telle quelle, mais en général elle est transformée ou corrigée.

Cette opération transfère au résultat de mesure une incertitude de justesse et de traçabilité (voir [1] pour la relation entre incertitude et traçabilité) qui ne figurait pas

dans l'incertitude associée à l'indication initiale. Ainsi, le rôle fondamental de l'étalonnage est de déterminer cette incertitude de justesse et de traçabilité qui, dans cet article, est nommée « incertitude d'étalonnage ».

Cette incertitude sera complétée lors de l'utilisation de l'instrument par l'incertitude issue des conditions de la mesure.

La fonction d'étalonnage est souvent estimée par la méthode des « moindres carrés ordinaires » (OLS pour *Ordinary Least Squares*), présente dans la plupart des logiciels (par ex. : Droitereg dans Excel), mais elle n'est pas adaptée à ces cas.

Soit une fonction à estimer  $Y = f(X)$ , où  $X$  est la variable d'entrée et  $Y$  la variable de sortie ; les OLS supposent que seule la variable  $Y$  est connue avec une incertitude, constante et sans corrélation. D'autres méthodes d'estimation existent. Cet article traite des méthodes de régression pondérée regroupées sous le terme unique de méthode des « moindres carrés généralisés » (GLS pour *Generalized Least Squares*).

Les principales caractéristiques des GLS sont :

- l'application du principe des OLS, en minimisant une fonction des carrés des écarts en  $X$  et en  $Y$  à la droite
- l'utilisation de poids pour augmenter ou diminuer l'influence relative d'une valeur  $x$  de  $X$  ou  $y$  de  $Y$  dans le calcul. En métrologie, ces poids dépendent des incertitudes.

Les GLS englobent par exemple, les moindres carrés pondérés, la droite de régression orthogonale, les moindres carrés totaux ou encore *Orthogonal Distance Regression*... En fait, l'estimateur des GLS dépend du modèle statistique considéré.

Malgré l'ancienneté de la méthode, le traitement d'exemples du laboratoire s'est révélé plutôt ardu. Les connaissances et les compétences utilisées étaient insuffisantes pour estimer correctement la droite d'étalonnage.

En effet, la méthode est plus complexe que les OLS et elle est moins documentée. Elle nécessite l'utilisation d'un logiciel pour effectuer les calculs. En outre, passer de la fonction d'étalonnage au modèle statistique sous-jacent n'est pas immédiat. Peu d'exemples d'utilisation des GLS en métrologie ont été publiés.

Le présent article commence par l'analyse des données de l'étalonnage et discute de leur utilisation pour estimer la fonction (partie 2). Puis le « modèle de régression standard » avec  $X$  sans incertitude est abordé. Il est estimé par les OLS (partie 3) ou les GLS (partie 4). Lorsque  $X$  et  $Y$  sont évalués avec une incertitude, les GLS sont associés au « modèle à erreurs sur les variables » (ou modèle à erreurs de mesure) (partie 5). Deux estimateurs des GLS sont présentés : les GLS-simples inspirés du modèle de régression standard (partie 6) et les GLS-GGMR, estimateur complet mais plus complexe (partie 7).

Le calcul du résultat de mesure est traité dans la partie 8 en distinguant l'évaluation directe et l'évaluation inverse. La validation des estimations, droite et résultats de mesure, est présentée dans la partie 9 pour toutes les méthodes. Deux exemples sont analysés dans la partie 10.

L'article reprend les notations du GUM [6] : les lettres majuscules représentent des variables aléatoires et les lettres minuscules, des variables certaines dont la valeur est connue ou inconnue. L'expression des estimateurs et des statistiques calculées sur les observations est donnée sous forme matricielle. Cette écriture matricielle est explicitée dans les ouvrages de statistiques [2, 14] et quelques éléments sont rappelés dans l'Annexe 1.

## 2. Droite d'étalonnage

La fonction d'étalonnage détermine le résultat de mesure correspondant à une indication. Dans le cas d'une droite, elle s'écrit :

$$z_t = \beta_0 + \beta_1 \cdot y_t, \quad (1)$$

où :

- les coefficients  $\beta_0$  et  $\beta_1$  sont inconnus ;
- $y$  est l'indication ou la valeur de la grandeur de sortie  $Y$  de l'instrument de mesure ;
- $z$  est la valeur du résultat  $Z$  ;
- $x$  est la valeur de la grandeur étalon  $X$  ;
- $t$  est un indice qui représente les variations de la grandeur dans le domaine d'étalonnage ( $x_t, y_t, z_t, X_t, Y_t$  et  $Z_t$ ).

La fonction d'étalonnage est déterminée à partir de la relation entre les valeurs  $x$  fournies par des étalons et les indications  $y$  correspondantes de l'instrument de mesure. Les valeurs étalons et les indications étant complétées par leur incertitude de mesure. Cette relation est établie à l'Étape 1 de l'étalonnage sur  $n$  couples de valeurs ( $x, y$ ) (§ 2.39 de [7]). Dans le présent article, elle est nommée « Relation 1 ».

Ces données et le contexte d'estimation sont analysés ci-après.

### 2.1. Incertitude des données d'étalonnage

La définition de l'étalonnage suppose que les incertitudes des étalons et des indications soient déterminées complètement et distinctement. Or, cette situation ne correspond pas à la pratique générale des laboratoires d'étalonnage.

Souvent, les incertitudes sont communiquées ainsi :

- la valeur étalon et son incertitude d'étalonnage ;
- l'indication avec une incertitude due à la fidélité de l'instrument à étalonner combinée à d'autres composantes comme par exemple, l'influence des variations de température sur l'étalon, l'instrument et les résultats de l'étalonnage.

Affecter une incertitude non prépondérante à une indication ou à une valeur étalon a peu de conséquences sous réserve d'utiliser toutes les incertitudes dans le calcul. Néanmoins, les covariances qui indiquent les liaisons existantes entre les sources d'incertitude sont rarement évaluées alors qu'elles influencent fortement l'incertitude des résultats.

## 2.2. Relation 1

Cette relation n'est pas formulée dans la définition de l'étalonnage. Elle peut être représentée par un tableau ou un graphique regroupant les  $n$  couples  $(x, y)$ . Lorsqu'une fonction d'étalonnage est évaluée, la Relation 1 est définie par une fonction mathématique. La Relation 1 est *bi-univoque* et la fonction mathématique peut s'écrire dans les deux sens :  $Y = h(X)$  ou  $X = g(Y)$ . En effet :

- du point de vue statistique, la Relation 1 associe deux variables aléatoires dans un contexte de corrélation. La valeur étalon et l'indication correspondante varient ensemble sans que l'une explique l'autre. Ce cas est représenté par le « modèle à erreurs sur les variables ». Dans le contexte de régression, représenté par le « modèle de régression standard », une variable connue sans incertitude explique la variable aléatoire.
- du point de vue métrologique, le modèle  $Y = h(X)$  est souvent postulé mais il ne correspond guère aux situations suivantes : le laboratoire d'étalonnage génère des grandeurs qui sont mesurées simultanément par l'instrument de mesure étalon et l'instrument à étalonner ; les incertitudes associées aux valeurs étalons sont plus élevées que les incertitudes associées aux indications.

En fait, la valeur étalon et l'indication correspondante sont corrélées car elles dépendent de la même grandeur (la grandeur mesurée et sa « valeur vraie »).

## 2.3. Fonction d'étalonnage

Elle est déterminée, à l'Étape 2 de l'étalonnage, en utilisant la Relation 1. Les deux relations diffèrent sur les points suivants :

- l'une associe des indications et des valeurs étalons ; l'autre associe des indications et des résultats de mesure ;
- l'une est bi-univoque, l'autre est univoque. La fonction d'étalonnage convertit une indication en un résultat de mesure et l'inverse n'est pas possible.

La confusion provient du fait que la fonction d'étalonnage est estimée par la Relation 1. La fonction d'étalonnage est la fonction  $g$  connue ( $z = g(y)$ ) ou la fonction  $h$  connue et utilisée de manière inverse (il s'agit de déterminer  $z$  tel que  $y = h(z)$  ; cette utilisation inverse de la fonction doit être distinguée de l'inversion de la fonction).

Dans de nombreuses publications, la fonction d'étalonnage est assimilée à la fonction  $h$ . Cette fonction représente généralement le principe de mesure de l'instrument. Néanmoins, beaucoup d'utilisateurs travaillent avec la fonction  $g$  pour obtenir directement le résultat de mesure.

Il n'existe pas *a priori* un sens d'estimation privilégié. En outre, cette question ne se pose plus lorsque la fonction estimée est symétrique.

## 2.4. Symétrie et évaluation inverse

La symétrie est la possibilité d'interchanger les variables dans la fonction :  $Y = h(X)$  et  $X = h^{-1}(Y)$ . Cette propriété dépend du modèle statistique et de la méthode d'estimation utilisée. La méthode des OLS n'est pas symétrique (régresser  $Y$  sur  $X$  n'est pas équivalent à régresser  $X$  sur  $Y$ ). La droite de régression orthogonale qui est un cas particulier des GLS préserve la symétrie.

Pour prévoir le résultat  $Z$  à partir d'une indication  $y$ , il est plus simple d'estimer directement la droite  $g$ .

Dans le cas symétrique, les évaluations  $z = g(y)$  et  $z = h^{-1}(y)$  concordent. Dans le cas contraire, il est nécessaire de considérer l'écart entre ces deux évaluations. Aussi, l'opération « évaluation inverse » associée à l'utilisation de la fonction  $h$  devrait inclure une composante d'incertitude représentant la non-symétrie (partie 8 de l'article).

Les parties 3 à 7 traitent de l'estimation d'une droite en général et elles s'appliquent aux fonctions  $g$  et  $h$ . Néanmoins, la transposition au contexte de l'étalonnage est facilitée en prenant l'exemple de la fonction  $h$ .

## 3. Le modèle de régression standard et les OLS

### 3.1. Présentation des OLS

#### 3.1.1. Le modèle de régression standard

Dans ce modèle, la variable  $y_t$  est mesurée avec une erreur aléatoire  $E_t$ . La variable  $x_t$  est supposée connue sans incertitude. Il s'écrit dans le cas d'une droite :

$$Y_t = \beta_0 + \beta_1 \cdot x_t + E_t \quad (2)$$

avec les hypothèses principales suivantes :

- Hyp. 1, le modèle est bien spécifié ;
- Hyp. 2, la variable explicative  $x_t$  est certaine ;
- Hyp. 3, l'erreur  $E_t$  a une moyenne nulle ;
- Hyp. 4, la variable  $Y_t$  est aléatoire avec une incertitude type  $\sigma_y$  constante (cas de l'homoscédasticité). Les différentes observations (valeurs)  $Y_t$  ne sont pas corrélées ;
- Hyp. 5, l'incertitude type  $\sigma_y$  est inconnue. L'erreur  $E_t$  représente la partie aléatoire de  $Y_t$  :

$$Y_t = y_t + E_t.$$

Les espérances mathématiques sont  $E(Y_t) = \beta_0 + \beta_1 \cdot x_t$  et  $E(E_t) = 0$ ; la matrice de variances-covariances identique pour  $Y$  et  $E$  est :

$$V(Y) = \begin{pmatrix} v(y_1) & \text{cov}(y_1, y_n) \\ \vdots & \vdots \\ \text{cov}(y_1, y_n) & v(y_n) \end{pmatrix} = \begin{pmatrix} \sigma_y^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_y^2 \end{pmatrix} = \sigma_y^2 \mathbf{I} \quad (3)$$

où  $\mathbf{I}$  est la matrice identité.

### 3.1.2. Les OLS

Dans ce contexte, la méthode des OLS détermine les estimateurs des coefficients, notés  $B_0$  et  $B_1$ , qui minimisent la somme des carrés des écarts entre les observations  $y_t$  et le modèle. Ces écarts étant regroupés dans un vecteur  $\mathbf{D}$ , les OLS minimisent :

$$\mathbf{D}^T \mathbf{D}, \quad (4)$$

où l'exposant « T » représente la transposée d'une matrice ou d'un vecteur.

Les OLS estiment les deux coefficients de la droite et la variance des erreurs.

$\boldsymbol{\beta} = (\beta_0, \beta_1)^T$  est estimé par  $\mathbf{B}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$  et  $\sigma_y^2$  par  $S_{OLS}^2 = (\mathbf{D}^T \mathbf{D}) / (n - 2)$ .

La matrice de variances-covariances de  $\mathbf{B}_{OLS}$  dépend de la variance des erreurs. Elle est estimée par :

$$V(\mathbf{B}_{OLS}) = S_{OLS}^2 (\mathbf{X}^T \mathbf{X})^{-1}. \quad (5)$$

Avec les hypothèses précédentes, les OLS donnent les meilleurs estimateurs linéaires (sans biais et à variance minimale). De plus, les OLS sont aisés à mettre en œuvre. Aussi, les OLS sont-ils couramment utilisés en métrologie et dans d'autres disciplines.

### 3.2. Limites des OLS

Les hypothèses des OLS sont très restrictives pour représenter de nombreuses situations réelles.

Dans le cas de l'étalonnage, les écarts peuvent porter sur les hypothèses Hyp. 2, Hyp. 4 et Hyp. 5. Les valeurs  $x_t$  ne sont pas connues sans incertitude. Les incertitudes varient (cas de l'hétéroscédasticité) généralement avec la valeur mesurée et il existe des corrélations. La matrice de variances-covariances de  $Y$  (comme celle de  $X$ ) est plus proche de (6) que de (3). Ces incertitudes sont en général connues, à l'exception des covariances qui ne sont pas systématiquement évaluées.

Les propriétés de l'estimateur sont modifiées [2-4] ainsi :

- l'hétéroscédasticité ou la corrélation des variables  $Y$  augmente son incertitude. En fait,  $V(\mathbf{B}_{OLS})$  ne correspond pas à (5) mais à  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T V(\mathbf{Y})^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$  qui est plus élevée que  $V(\mathbf{B}_{GLS})$  (9). Par contre, l'estimateur  $\mathbf{B}_{OLS}$  reste sans biais ;

- les OLS qui supposent l'incertitude de  $Y$  inconnue ne propagent pas cette incertitude aux coefficients de la droite et aux valeurs calculées avec la droite ;
- l'incertitude associée à la variable  $X$  biaise l'estimateur  $\mathbf{B}_{OLS}$ . En général, il sous-estime la valeur de la pente  $\beta_1$ , l'importance du biais dépendant de la variance de  $X$ .

La méthode des OLS distingue une variable explicative et une variable expliquée qui n'ont pas le même rôle dans le modèle. Elle n'est pas symétrique : les coefficients des deux droites diffèrent mais ils sont équivalents aux incertitudes près. Ce résultat peut être constaté sur l'exemple du paragraphe 10.2.

En conclusion, la méthode des OLS ne devrait pas être utilisée pour estimer la fonction d'étalonnage.

## 4. Le modèle de régression standard et les GLS

Dans ce modèle, les GLS sont substitués aux OLS lorsque la variable  $Y$  est hétéroscédastique ou autocorrélée.

### 4.1. Présentation des GLS

#### 4.1.1. Hypothèses

- Hyp. 1, Hyp. 2 et Hyp. 3 des OLS sont conservées ;
- Hyp. 4 est remplacée par : Hyp. 6, la variable  $Y$  est aléatoire avec une incertitude type variable ou elle est autocorrélée ;
- Hyp. 5 est remplacée par : Hyp. 7, la matrice de variances-covariances de  $Y$  notée  $V(\mathbf{Y})$  est connue et inversible.

Cette matrice s'écrit :

$$V(\mathbf{Y}) = \begin{pmatrix} v(y_1) & \dots & \text{cov}(y_1, y_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(y_1, y_n) & \dots & v(y_n) \end{pmatrix} = \begin{pmatrix} \sigma_{y_1}^2 & \dots & \sigma_{y_1, y_n} \\ \vdots & \dots & \vdots \\ \sigma_{y_1, y_n} & \dots & \sigma_{y_n}^2 \end{pmatrix}. \quad (6)$$

Des valeurs de  $Y$  ont de grandes variances et d'autres des variances faibles. Par ailleurs, certaines de ces valeurs ne sont pas indépendantes.

La méthode des GLS consiste à associer des poids aux valeurs  $y_t$  pour atténuer ou augmenter leur influence relative dans le processus d'estimation. Dans le cas où ces poids sont les inverse des variances, les valeurs dont l'incertitude est élevée ont moins d'influence dans le processus d'estimation que les valeurs avec une incertitude faible.

#### 4.1.2. Transformation du modèle

Les GLS sont une transformation des données pour pouvoir appliquer les OLS : la pondération des

observations avec des poids issus des variances et des covariances donne de nouvelles valeurs, non corrélées et de variance constante égale à 1.

Mathématiquement, la matrice de variances-covariances des erreurs qui est symétrique est décomposée en un produit de deux matrices  $\mathbf{V}(\mathbf{Y}) = \mathbf{P}\mathbf{P}^T$ .

La transformation des observations en  $\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\mathbf{B} + \mathbf{P}^{-1}\mathbf{E}$  permet d'obtenir de nouvelles erreurs dont la matrice de variances-covariances est :  $\mathbf{V}(\mathbf{P}^{-1}\mathbf{E}) = \mathbf{I}$  comme dans le modèle des OLS, à une constante près.

Alors que les OLS minimisent  $\mathbf{D}^T\mathbf{D}$ , les GLS minimisent :

$$\mathbf{D}^T\mathbf{V}(\mathbf{Y})^{-1}\mathbf{D} \quad (7)$$

où  $\mathbf{V}(\mathbf{Y})^{-1}$  représente la matrice des poids.

Lorsque  $\mathbf{V}(\mathbf{Y})$  est diagonale, les GLS minimisent la somme des carrés des écarts pondérés par l'inverse de leur écart type.

#### 4.1.3. Expression des estimateurs

Les GLS estiment les coefficients  $\beta$  de la droite par :

$$\mathbf{B}_{\text{GLS}} = (\mathbf{X}^T\mathbf{V}(\mathbf{Y})^{-1}\mathbf{X})^{-1} \mathbf{X}^T\mathbf{V}(\mathbf{Y})^{-1}\mathbf{Y}. \quad (8)$$

La matrice de variances-covariances de  $\mathbf{B}_{\text{GLS}}$  est :

$$\mathbf{V}(\mathbf{B}_{\text{GLS}}) = (\mathbf{X}^T\mathbf{V}(\mathbf{Y})^{-1}\mathbf{X})^{-1}. \quad (9)$$

Il n'est pas nécessaire d'estimer les variances des erreurs supposées connues avant d'utiliser la méthode.

L'estimateur GLS est sans biais avec une matrice de variances-covariances minimale (meilleur estimateur linéaire) lorsque  $\mathbf{V}(\mathbf{Y}) \neq \sigma^2\mathbf{I}$ . Ce résultat est établi par le théorème de Gauss-Markov, d'où le nom « méthode de Gauss-Markov » également attribué aux GLS.

#### 4.1.4. Limites des GLS

L'Hyp. 7 conditionne l'utilisation de la méthode.

La connaissance de  $\mathbf{V}(\mathbf{Y})$  est une hypothèse théorique ; dans la pratique, les variances et les covariances des erreurs ont été estimées avec les données de l'étalonnage. Ces estimations ne sont pas exemptes d'erreurs. Ce point est discuté lors de la validation de la droite estimée.

Il existe d'autres présentations de la méthode : les incertitudes sont connues à une constante près ( $\mathbf{V}(\mathbf{Y}) = \sigma_y^2 \cdot \mathbf{M}$  où  $\sigma_y^2$  est inconnue et  $\mathbf{M}$  est connue) ; ou la méthode est réitérée jusqu'à trouver deux estimations consécutives de  $\mathbf{V}(\mathbf{Y})$  très proches. Ces applications au caractère plus mathématique que métrologique ne sont pas retenues ici.

Par ailleurs, l'estimateur n'est pas (est difficilement) calculable lorsque les valeurs de la variable  $Y$  sont parfaitement (très) corrélées.

Comme les OLS, ces GLS ne sont pas symétriques.

## 4.2. Propagation de l'incertitude de $Y$

La variance de l'estimateur (9) dépend directement de l'incertitude des valeurs de  $Y$ . Lorsque :

- $\mathbf{V}(\mathbf{Y})$  est modifiée par un facteur multiplicatif  $\mu$ , la matrice  $\mathbf{V}(\mathbf{B}_{\text{GLS}})$  est multipliée par  $\mu$  :

$$\mathbf{V}(\mathbf{B}_{\text{GLS}}) = (\mathbf{X}^T\mu^{-1}\mathbf{V}(\mathbf{Y})^{-1}\mathbf{X})^{-1} = \mu(\mathbf{X}^T\mathbf{V}(\mathbf{Y})^{-1}\mathbf{X})^{-1},$$

l'estimateur  $\mathbf{B}_{\text{GLS}}$  restant inchangé :

$$\begin{aligned} \mathbf{B}_{\text{GLS}} &= (\mathbf{X}^T\mu^{-1}\mathbf{V}(\mathbf{Y})^{-1}\mathbf{X})^{-1} \mathbf{X}^T\mu^{-1}\mathbf{V}(\mathbf{Y})^{-1}\mathbf{Y} \\ &= (\mathbf{X}^T\mathbf{V}(\mathbf{Y})^{-1}\mathbf{X})^{-1} \mathbf{X}^T\mathbf{V}(\mathbf{Y})^{-1}\mathbf{Y}. \end{aligned}$$

- $\mathbf{V}(\mathbf{Y}) = \sigma_y^2\mathbf{I}$  avec  $\sigma_y^2$  connue.  $\mathbf{V}(\mathbf{B}_{\text{GLS}}) = \sigma_y^2(\mathbf{X}^T\mathbf{X})^{-1}$  est simple comme la variance de l'estimateur OLS (4) ; sauf que la méthode des OLS ré-estime plus ou moins correctement  $\sigma_y^2$  par  $S_{\text{OLS}}^2$ . Les estimateurs  $\mathbf{B}_{\text{GLS}}$  et  $\mathbf{B}_{\text{OLS}}$  sont identiques.

L'estimateur  $\mathbf{B}_{\text{GLS}}$  est plus influencé par la structure des incertitudes (poids relatif des variables les unes par rapport aux autres) que par leur ordre de grandeur. Au contraire, sa matrice de variances-covariances dépend des deux.

L'influence des incertitudes est discutée dans la partie 9 de l'article consacrée à la validation.

La propagation de l'incertitude de  $Y$  aux résultats de mesure est explicitée dans la partie 8.

En conclusion, les GLS appliqués au modèle de régression standard sont une méthode relativement simple. Ils peuvent être programmés dans Excel avec les outils matriciels.

Ces GLS sont adaptées à la droite d'étalonnage lorsqu'une des variables  $X$  ou  $Y$  est connue avec une incertitude négligeable.

## 5. Le modèle à erreurs sur les variables (MEV)

Ce modèle permet de considérer l'incertitude sur les deux variables  $X$  et  $Y$ .

### 5.1. Présentation

Le modèle MEV est décrit dans les références [3, 4]. Dans ce modèle, les variables  $x_t$  et  $y_t$  liées par l'équation  $y_t = h(x_t)$ , ne sont pas observées. Elles sont mesurées avec une erreur aléatoire, additive et de moyenne nulle.

Dans le cas d'une droite, le modèle s'écrit :

$$y_t = \beta_0 + \beta_1 \cdot x_t \quad (10)$$

$$X_t = x_t + L_t \quad (11)$$

$$Y_t = y_t + E_t \quad (12)$$

où  $x_t$  et  $y_t$  sont les variables vraies ou latentes,  $X_t$  et  $Y_t$  sont les variables observées ou mesurées,  $E_t$  et  $L_t$  les erreurs.

Le modèle comporte différentes hypothèses sur les variables vraies, les erreurs et l'équation (10). Le choix des hypothèses est très large : par exemple, les variables vraies  $x_t$  et  $y_t$  sont certaines ou aléatoires ; l'équation (10) peut comporter une erreur  $C_t$  aléatoire,

$$y_t = \beta_0 + \beta_1 \cdot x_t + C_t. \quad (10\text{bis})$$

Ainsi, il existe un grand nombre de modèles à erreurs sur les variables (MEV) susceptibles de représenter différents contextes de régression. Le MEV associé à la fonction d'étalonnage est présenté ci-après.

## 5.2. Modèle de la fonction d'étalonnage

La droite d'étalonnage est modélisée par les équations (10), (11) et (12) avec les hypothèses suivantes :

- Hyp. 1 des OLS ;
- Hyp. 3 des OLS appliquée aux erreurs  $E_t$  et  $L_t$  ;
- Hyp. 6 et Hyp. 7 des GLS appliquées aux erreurs  $E_t$  et  $L_t$  ;
- Hyp. 8, les erreurs  $E_t$  et  $L_t$  sont indépendantes. L'erreur qui affecte  $X_t$  est sans lien avec l'erreur qui affecte  $Y_t$  ;
- Hyp. 9, les variables vraies  $x_t$  et  $y_t$  sont certaines ;
- Hyp. 10, l'équation (10) représente bien la droite d'étalonnage ; il n'est pas nécessaire d'ajouter un terme supplémentaire comme dans l'équation (10bis).

La matrice de variances-covariances de  $X$  a la même structure que  $V(\mathbf{Y})$  (6) :

$$V(\mathbf{X}) = \begin{pmatrix} \sigma_{x1}^2 & \dots & \sigma_{x1,xn} \\ & \dots & \\ \sigma_{x1,xn} & & \sigma_{xn}^2 \end{pmatrix}. \quad (13)$$

### 5.2.1. Interprétation

Les variables vraies  $x_t$  sont considérées comme certaines car l'étalonnage est réalisé en des points choisis du domaine de mesure. L'équation (10) traduit le principe de mesure de l'instrument étalonné et en général, celui-ci est bien connu. Il en résulte que les variables  $y_t$  sont certaines. Ces hypothèses, plutôt simples, ont été choisies pour focaliser le problème de l'estimation sur l'hétéroscédasticité et la corrélation des valeurs.

Il est clair qu'une fonction d'étalonnage mal connue conduira à inclure le terme erreur  $C_t$ , qu'une valeur  $x_t$  d'un étalon variant entre le moment de sa détermination et l'étalonnage en cours sera considérée comme une variable aléatoire... L'étude de ces cas dépasse le cadre de cet article.

### 5.2.2. Caractéristiques du modèle

Avec les hypothèses précédentes :

- la droite d'étalonnage est symétrique dans les variables vraies ( $x_t, y_t$ ) et observées ( $X_t, Y_t$ ) ; il n'y a pas de sens de régression privilégié ;
- l'équation (10) n'est pas linéaire dans les paramètres  $x_t$  et  $\beta$  inconnus ;
- le nombre de valeurs vraies  $x_t$  à déterminer dépend du nombre de points utilisés pour estimer la fonction. Du point de vue métrologique, on obtient une nouvelle évaluation de la valeur des étalons.

Ces variables vraies ( $x_t, y_t$ ) correspondent aux valeurs moyennes (ou espérances mathématiques) des variables mesurées et aléatoires ( $X_t, Y_t$ ). Une valeur individuelle de ces variables mesurées sera notée ( $x_{t\text{-ind}}, y_{t\text{-ind}}$ ) pour la distinguer de la valeur moyenne. La distinction entre valeur moyenne et valeur individuelle intervient lors du calcul du résultat de mesure.

## 5.3. Méthodes d'estimation

Les méthodes d'estimation proposées pour le MEV dépendent des hypothèses du modèle. Il peut s'agir des méthodes classiques d'estimation comme la méthode du maximum de vraisemblance, la méthode des moments... généralement dans le cas simple où les variables sont connues avec une incertitude constante et les covariances nulles. De nombreux articles existent sur le sujet : ratio des variances constant, connu ou inconnu, droite de régression orthogonale. Dans certains cas, ces méthodes donnent des résultats équivalents [8].

Les GLS appliqués au modèle du paragraphe 5.2 sont caractérisés par :

- des poids évalués avec les incertitudes associées aux variables  $X$  et  $Y$  ;
- des estimateurs définis par un système d'équations non linéaires et calculés avec une méthode d'analyse numérique.

Une version simplifiée des GLS a été proposée ; elle reprend l'expression de l'estimateur  $\mathbf{B}_{\text{GLS}}$  du modèle de régression standard. Cette méthode GLS-simples est présentée dans la partie 6 de l'article. La méthode complète appelée « GLS-GGMR » est présentée dans la partie 7.

## 6. Le modèle à erreurs sur les variables et les GLS-simples

Les GLS-simples [9] sont issus de la « méthode des *Bivariate Least Squares* (BLS) » [10] et des GLS du modèle de régression standard.

## 6.1. Modèle

Le modèle considéré est l'équation (10) écrite avec les variables observées. Soit :

$$Y_t = \beta_0 + \beta_1 \cdot X_t + Q_t \quad (14)$$

$$\text{avec } Q_t = E_t - \beta_1 \cdot L_t. \quad (15)$$

Ces équations et la méthode ignorent les variables vraies  $x_t$  et  $y_t$ .

L'équation (14) ressemble à l'équation (2) du modèle de régression standard. Mais, l'erreur  $Q_t$  qui intègre le terme  $L_t$  est corrélée avec la variable explicative  $X_t$ . Donc, les OLS ne sont pas adaptés.

Comme les erreurs  $E_t$  et  $L_t$  ne sont pas corrélées (Hyp. 8), la variance d'une erreur  $Q_t$  est :

$$v(Q_t) = \sigma^2 y_t + \beta_1^2 \cdot \sigma^2 x_t \quad (16)$$

et la covariance de deux erreurs  $Q_t$  et  $Q_1$  est :

$$\text{cov}(Q_t, Q_1) = \sigma_{y_t, y_1} + \beta_1^2 \cdot \sigma_{x_t, x_1}. \quad (17)$$

La matrice de variances-covariances des erreurs s'écrit :

$$\mathbf{V}(\mathbf{Q}) = \begin{pmatrix} \sigma_{y_1}^2 + \beta_1^2 \sigma_{x_1}^2 & \dots & \sigma_{y_1, y_n} + \beta_1^2 \sigma_{x_1, x_n} \\ \dots & \dots & \dots \\ \sigma_{y_1, y_n} + \beta_1^2 \sigma_{x_1, x_n} & \dots & \sigma_{y_n}^2 + \beta_1^2 \sigma_{x_n}^2 \end{pmatrix}. \quad (18)$$

$\mathbf{V}(\mathbf{Q})$  est une combinaison linéaire des matrices de variances-covariances des erreurs de  $X$  et de  $Y$  :

$$\mathbf{V}(\mathbf{Q}) = [\mathbf{V}(\mathbf{Y}) + \beta_1^2 \cdot \mathbf{V}(\mathbf{X})].$$

Elle dépend du coefficient  $\beta_1$  inconnu.

## 6.2. Estimateur

Le principe est d'utiliser l'estimateur GLS du modèle de régression standard en remplaçant  $\mathbf{V}(\mathbf{Y})$  par  $\mathbf{V}(\mathbf{Q})$ .

Ainsi, l'estimateur (8) s'écrit ici :

$$\mathbf{B}_{\text{GLS-simples}} = (\mathbf{X}^T \mathbf{V}(\mathbf{Q})^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}(\mathbf{Q})^{-1} \mathbf{Y} \quad (19)$$

$$\text{avec } \mathbf{V}(\mathbf{Q}) = [\mathbf{V}(\mathbf{Y}) + \mathbf{B}_{1\text{-GLS-simples}}^2 \cdot \mathbf{V}(\mathbf{X})]$$

et sa matrice de variances-covariances (9) :

$$\mathbf{V}(\mathbf{B}_{\text{GLS-simples}}) = (\mathbf{X}^T \mathbf{V}(\mathbf{Q})^{-1} \mathbf{X})^{-1}. \quad (20)$$

Le système d'équations (19) est non linéaire dans les inconnues  $B_{0\text{-GLS-simples}}$  et  $B_{1\text{-GLS-simples}}$ , puisque  $B_{1\text{-GLS-simples}}$  figure dans la matrice  $\mathbf{V}(\mathbf{Q})^{-1}$ . Il est résolu de manière itérative :

- une solution initiale est par exemple, l'estimation  $b_1$  obtenue avec les GLS du modèle de régression standard (8) ;

- à l'Étape 1, la matrice  $\mathbf{V}(\mathbf{Q})$  est recalculée ;
- à l'Étape 2, de nouvelles estimations  $\mathbf{b}_{\text{GLS-simples}}$  et  $\mathbf{V}(\mathbf{b}_{\text{GLS-simples}})$  sont calculées et la matrice  $\mathbf{V}(\mathbf{Q})$  est actualisée avec  $b_{1\text{-GLS-simples}}$  ;
- puis l'Étape 2 est réitérée jusqu'à la convergence de deux solutions  $b_{1\text{-GLS-simples}}$  consécutives.

## 6.3. Avantages et inconvénients

Cette méthode propose une formule simple de calcul de l'estimateur, facile à implémenter dans un tableur. Elle permet également de mieux appréhender la propagation des incertitudes dans le système.

En retour, elle n'estime pas les variables vraies et il est possible que la valeur  $\mathbf{b}_{\text{GLS-simples}}$  diffère de la valeur obtenue avec la méthode complète  $\mathbf{b}_{\text{GLS-GGMR}}$ . De plus, la méthode n'est pas symétrique.

## 7. Le modèle à erreurs sur les variables et les GLS-GGMR

### 7.1. Présentation

Les GLSD-GGMR sont décrits dans [12 et 5] où GGMR signifie « *Generalised Gauss-Markov Regression* ».

Le modèle considéré est représenté par les équations (10), (11) et (12) associées aux hypothèses du paragraphe 5.2. En particulier, les matrices de variances-covariances des variables  $\mathbf{V}(\mathbf{X})$  et  $\mathbf{V}(\mathbf{Y})$  sont connues.

Les paramètres à estimer sont les coefficients de la droite d'étalonnage et les variables vraies  $x_t$ . Ils sont regroupés dans le vecteur  $\mathbf{B}_{\text{GLS-GGMR}}$  :

$$\mathbf{B}_{\text{GLS-GGMR}} = (x_1, x_2, \dots, x_n, b_0, b_1)^T. \quad (21)$$

Le nombre de variables  $x_t$  correspond au nombre  $n$  de points utilisés pour estimer le modèle.

La méthode des GLS consiste à minimiser une forme quadratique composée des écarts en  $X$  et en  $Y$  :

$$\mathbf{S}(\mathbf{B}) = \mathbf{D}^T \mathbf{V}(\mathbf{D})^{-1} \mathbf{D} \quad (22)$$

où  $\mathbf{D}$  est le vecteur des écarts :

$$\mathbf{D} = \begin{pmatrix} \mathbf{L} \\ \mathbf{E} \end{pmatrix} = \begin{pmatrix} \mathbf{X} - \mathbf{x} \\ \mathbf{Y} - (\beta_0 + \beta_1 \cdot \mathbf{x}) \end{pmatrix}. \quad (23)$$

$\mathbf{V}(\mathbf{D})$  est la matrice de variances-covariances des variables ou des écarts :

$$\mathbf{V}(\mathbf{D}) = \begin{pmatrix} \mathbf{V}(\mathbf{X}) & \mathbf{V}_{\mathbf{X},\mathbf{Y}} \\ \mathbf{V}_{\mathbf{X},\mathbf{Y}} & \mathbf{V}(\mathbf{Y}) \end{pmatrix} = \begin{pmatrix} \mathbf{V}(\mathbf{X}) & \mathbf{0} \\ \mathbf{0} & \mathbf{V}(\mathbf{Y}) \end{pmatrix}. \quad (24)$$

Par rapport à la méthode GLS-simples, la structure de la matrice des poids issue de  $\mathbf{V}(\mathbf{D})$  est plus directe.

## 7.2. Estimateur

Ce problème de moindres carrés non linéaires est résolu avec la méthode de Gauss-Newton puisque  $S(B)$  est le produit d'un vecteur par son transposé :

$$\mathbf{S}(\mathbf{B}) = \mathbf{D}^T \mathbf{V}(\mathbf{D})^{-1} \mathbf{D} = \mathbf{D}^T \mathbf{P}^{-1T} \mathbf{P}^{-1} \mathbf{D}, \quad (25)$$

où  $\mathbf{V}(\mathbf{D}) = \mathbf{P}\mathbf{P}^T$  est décomposée avec l'algorithme de Cholevski [5] (l'algorithme QR peut également être utilisé).

L'algorithme de calcul (issu de [12]) se déroule comme suit :

- à la première étape, une estimation initiale du vecteur  $\mathbf{B}_{\text{GLS-GGMR}}$  est calculée – les variables observées  $X$  et l'estimation  $\mathbf{b}_{\text{OLS}}$  peuvent être utilisées :  $\mathbf{b}_{\text{GLS-GGMR}}^* = (x_1^*, x_2^*, \dots, x_n^*, b_0^*, b_1^*)^T$ . On obtient l'estimation correspondante  $\mathbf{S}(\mathbf{b}^*)$  ;
- le pas de variation nommé «  $\delta$  » qui permet de diminuer la solution  $\mathbf{S}(\mathbf{b}^*)$  est donné par :

$$\mathbf{E} = (\mathbf{P}^{-1T} \cdot \mathbf{J} \cdot \delta) + (\mathbf{P}^{-1T} \mathbf{D}), \quad (26)$$

où  $\mathbf{J}$  est la matrice des dérivées partielles de  $\mathbf{D}$  par rapport à  $\mathbf{B}$  et évaluée en  $\mathbf{b}_{\text{GLS-GGMR}}^*$ . (26) définit un système d'équations surdéterminé qui peut être résolu par la méthode des OLS ;

- lorsque  $\delta$  est estimé, le vecteur  $\mathbf{b}_{\text{GLS-GGMR}}^*$  est actualisé :

$$\mathbf{b}_{\text{GLS-GGMR}}^* = \mathbf{b}_{\text{GLS-GGMR}}^* + \delta$$

et l'algorithme est réitéré jusqu'à l'obtention d'un  $\delta$  très faible ou lorsque le nombre maximal d'itérations fixé est atteint.

La matrice de variances-covariances de l'estimation est donnée par :

$$\mathbf{V}(\mathbf{b}_{\text{GLS-GGMR}}^*) = (\mathbf{J}^* \mathbf{P}^{-1} \mathbf{P}^{-1T} \mathbf{J}^{*T}). \quad (27)$$

Il s'agit d'une matrice partitionnée :

$$\mathbf{V}(\mathbf{b}_{\text{GLS-GGMR}}^*) = \begin{pmatrix} \mathbf{V}_X & \mathbf{V}_{X,B} \\ \mathbf{V}_{X,B} & \mathbf{V}(\mathbf{b}_0^*, \mathbf{b}_1^*) \end{pmatrix}$$

dont la sous-matrice correspondant aux coefficients  $b_0$  et  $b_1$  de la droite est donnée par :

$$\mathbf{V}(\mathbf{b}_0^*, \mathbf{b}_1^*) = (\mathbf{x}^{*T} \mathbf{P}^{-1} \mathbf{x}^*)^{-1} \quad (28)$$

où  $\mathbf{x}^* = (\mathbf{1} \ \mathbf{x}^*)$  est la matrice comportant le vecteur unité et la variable  $x$  évaluée à la solution.

L'expression (28) a la même structure que la variance (9) de l'estimateur GLS du modèle de régression standard. La propagation des incertitudes des variables  $X$  et  $Y$  apparaît ainsi plus clairement.

## 7.3. Avantages et inconvénients

La méthode GLS-GGMR utilise toute l'information sur les variables. Ses estimations seront donc plus fiables que celles obtenues avec les autres méthodes. L'estimateur GLS-GGMR est convergent : le biais diminue avec l'augmentation de  $n$ . Dans le contexte du MEV, l'estimateur des OLS est biaisé et non convergent.

GLS-GGMR est la seule méthode qui préserve la symétrie de la droite d'étalonnage : les droites estimées  $y = f(x)$  et  $x = g(y)$  sont équivalentes. Cette propriété est utile lorsqu'il est prévu de calculer le résultat de mesure par évaluation inverse.

Le critère minimisé (25) se simplifie en fonction de la structure de la matrice  $\mathbf{V}(\mathbf{D})$ . Ainsi, lorsque :

- les valeurs de  $X$  et  $Y$  ont une variance égale à 1 et ne sont pas corrélées ( $\mathbf{V}(\mathbf{D}) = \mathbf{I}$ ),  $\mathbf{S}(\mathbf{B})$  se réécrit :

$$\mathbf{D}^T \mathbf{D} = \sum_{t=1}^n (X_t - x_t)^2 + \sum_{t=1}^n (Y_t - y_t)^2. \quad (29)$$

La droite obtenue correspond à la droite de régression orthogonale. Elle passe par le point moyen des  $n$  couples  $(x, y)$  et l'expression de la pente est donnée dans [16].

- les valeurs de  $X$  et de  $Y$  ne sont pas corrélées ( $\mathbf{V}(\mathbf{X})$  et  $\mathbf{V}(\mathbf{Y})$  sont diagonales),  $\mathbf{S}(\mathbf{B})$  se réécrit :

$$\mathbf{D}^T \mathbf{V}(\mathbf{D})^{-1} \mathbf{D} = \sum_{t=1}^n \sigma_{xt}^{-2} (X_t - x_t)^2 + \sum_{t=1}^n \sigma_{yt}^{-2} (Y_t - y_t)^2 \quad (30)$$

et on retrouve la droite de régression orthogonale pondérée appelée aussi régression de la distance généralisée ( $\mathbf{S}(\mathbf{B})$  correspond au critère utilisé dans la régression de Deming et dans la méthode de Williamson).

La méthode généralise différentes techniques de régression, avec l'avantage de calculer également toutes les statistiques utiles pour valider et utiliser le modèle.

## 7.4. Conclusion sur le MEV

Lorsque les deux variables ont une incertitude, la méthode GLS-GGMR est à privilégier. Dans la pratique, l'utilisateur vérifie l'adéquation de son modèle d'étalonnage avec les hypothèses du paragraphe 5.2. Il doit connaître ou estimer avec soin les variances et les covariances des variables. En cas de doute sur les corrélations, il est préférable de tester des valeurs *a priori* plutôt que de les ignorer.

Le calcul des estimations nécessite l'utilisation d'un logiciel spécialisé ou la programmation de l'algorithme de résolution. D'où l'intérêt des GLS-simples qui donnent souvent les mêmes valeurs que les GLS-GGMR



et qui sont plus faciles à calculer. Les GLS-simples peuvent être utilisés lors des premières estimations de la droite pour vérifier l'adéquation droite/incertitude des variables en entrée et valider les hypothèses du modèle.

## 8. Calcul du résultat de mesure

### 8.1. Sources d'incertitude

Le résultat de mesure est déduit d'une indication en utilisant la droite d'étalonnage estimée (il s'agit de la fonction  $g$  ou  $h$  du §.2.2).

L'incertitude due à la droite est la première source d'incertitude associée au résultat de mesure. Elle est évaluée, dans le cas général, au paragraphe 8.2.

Une deuxième source d'incertitude peut provenir de l'utilisation d'une valeur individuelle.

Une valeur individuelle est une réalisation d'une variable mesurée. Elle est formée d'une valeur moyenne sans incertitude et d'une erreur aléatoire. Par exemple, une indication individuelle s'écrit :  $y_{0-ind} = y_0 + e_0$  conformément à l'équation (10). Les points dont les coordonnées sont les valeurs moyennes forment la droite d'étalonnage estimée alors que les autres points, dont une coordonnée au-moins est une valeur individuelle, se répartissent autour de la droite.

Dans la suite, le résultat est calculé en considérant les différentes options : indication moyenne ou individuelle, résultat moyen ou individuel. Pour simplifier :

- les notations précédentes  $e_t$  et  $l_t$  des erreurs sont conservées (les erreurs sur les variables  $x_0$  et  $y_0$  sont indépendantes des erreurs sur les variables  $x_t$  et  $y_t$ ) ;
- les composantes d'incertitude sont supposées indépendantes et sommées quadratiquement.

### 8.2. Incertitude associée à une droite

La droite estimée est constituée de l'ensemble des points du domaine (d'étalonnage) qui vérifient son équation. Soit :

$$y_0 = b_0 + b_1 \cdot x_0, \quad (31)$$

avec  $b_0$  et  $b_1$  les coefficients estimés,  $y_0$  et  $x_0$  les valeurs moyennes. L'intervalle d'incertitude de la droite dépend de l'incertitude associée à chaque point.

#### 8.2.1. Incertitude d'un point moyen

L'incertitude type est obtenue en appliquant la formule de propagation des incertitudes à (31) :

$$u(y_{0-droite}) = \sqrt{u^2(b_0) + x_0^2 u^2(b_1) + 2x_0 u(b_0, b_1)}. \quad (32)$$

L'incertitude élargie correspondante est :

$$U(y_{0-droite}) = k \cdot u(y_{0-droite}). \quad (33)$$

Dans le cas où les variables suivent une loi normale, le facteur d'élargissement  $k$  est issu de la loi de Student avec  $(n-2)$  degrés de liberté.

Cette incertitude dépend de la valeur du prédicteur  $x_0$ . Elle a une allure parabolique dans le cas des OLS : elle est plus faible au centre du domaine et plus large aux extrémités. Ce comportement existe généralement dans le cas des GLS du modèle de régression standard. Dans le cas du MEV, avec de l'incertitude sur  $X$ , cette incertitude croît linéairement avec  $x_0$ .

#### 8.2.2. Incertitude de la droite

L'intervalle d'incertitude de la droite est obtenu en joignant les limites des intervalles d'incertitude ponctuels [ $y_0 \pm U(y_{0-droite})$ ]. La correction du facteur  $k$  de l'équation (33) est négligée (cet intervalle d'incertitude issu de l'ensemble des points dépend de la loi de Fisher au lieu de la loi de Student utilisée dans (28)). Ce point est traité dans Willink [13]).

Cet intervalle est présenté dans les deux exemples de la partie 10. L'allure parabolique de l'incertitude type se traduit par un intervalle d'incertitude de forme hyperbolique autour de la droite. Cette évolution est nette dans l'Exemple 1 (partie 10). Dans le cas du MEV (méthodes GLS-simples et GLS-GGMR), l'intervalle représente une bande linéaire autour de la droite qui s'élargit avec l'augmentation de  $x_0$  (voir Exemple 2 en partie 10).

### 8.3. Évaluation directe du résultat de mesure

Le résultat de mesure est calculé avec la fonction  $g$  estimée (§.2.2). Dans cette fonction :

- la variable d'entrée est l'indication, notée  $y$ , associée à la variable  $Y$  ;
- la variable de sortie est le résultat de mesure noté  $Z$ .

Ainsi, le résultat de mesure se déduit de :

$$z_0 = b_0 + b_1 \cdot y_0, \quad (34)$$

qu'il s'agisse du résultat moyen  $z_0$  ou individuel  $z_{0-ind}$ , évalué avec l'indication moyenne  $y_0$  ou individuelle  $y_{0-ind}$ . Par contre, les incertitudes diffèrent.

#### 8.3.1. Incertitude

L'incertitude type associée au résultat moyen  $y_0$  est donnée par (32) en remplaçant  $x_0$  par  $y_0$  et  $y_{0-droite}$  par  $z_{0-droite}$ .

Cette incertitude est augmentée des composantes issues des valeurs individuelles :

- $u(e_0)$  pour un résultat individuel  $z_{0-ind}$  :

$$u^2(z_{0-ind}) = u^2(z_{0-droite}) + u^2(e_0) \quad (35)$$

–  $u(l_0)$  pour une indication individuelle :

$$u^2(z_0) = u^2(z_{0\text{-droite}}) + b_1^2 u^2(l_0). \quad (36)$$

L'incertitude maximale est obtenue pour un résultat individuel évalué à une indication individuelle :

$$u^2(z_{0\text{-ind}}) = u^2(z_{0\text{-droite}}) + u^2(e_0) + b_1^2 u^2(l_0). \quad (37)$$

En général, l'incertitude élargie est calculée avec le coefficient d'élargissement  $k = 2$ , par exemple,  $U(z_0) = 2u(z_0)$ .

### 8.3.2. Quelle incertitude associer à son résultat ?

Dans « l'incertitude d'étalonnage » évoquée dans la partie 1, l'indication est une valeur « vraie » sans incertitude. Sans connaissance des futures mesures réalisées avec l'instrument étalonné, le laboratoire d'étalonnage considèrera plutôt une indication « vraie ». Ensuite, l'utilisateur augmentera cette incertitude d'étalonnage avec l'incertitude sur l'indication individuelle.

En général, le résultat évalué est une valeur moyenne. Son incertitude est déterminée par les équations (32) et (36).

Néanmoins, il est important d'indiquer qu'il s'agit d'un résultat moyen. En effet, le terme  $u(e_0)$  associé au résultat individuel est souvent plus élevé que l'incertitude associée à la droite (32). Dans ce cas, l'incertitude sur le résultat, obtenue avec l'expression (36), est plus faible que l'incertitude sur la valeur étalon correspondante  $x_i$  utilisée pour estimer la droite.

### 8.4. Évaluation inverse du résultat de mesure

Le résultat de mesure est calculé avec la fonction  $h$  estimée (§.2.2). Dans cette fonction :

- la variable d'entrée est le résultat de mesure noté  $Z$  ;
- la variable de sortie est l'indication, notée  $y$ , associée à la variable  $Y$ .

La fonction est utilisée dans le sens opposé au sens de régression. Soit déterminer la valeur  $z_0$  qui correspond à l'indication  $y_0$ . Dans le cas d'une droite,  $z_0$  est obtenu en inversant l'équation :

$$z_0 = (y_0 - b_0) / b_1. \quad (38)$$

Lorsque la fonction est polynomiale, cette évaluation inverse donne plusieurs racines  $z_0$  qui correspondent à la même valeur  $y_0$ .

L'équation (38) s'applique au résultat moyen  $z_0$  ou individuel  $z_{0\text{-ind}}$  évalué à l'indication moyenne  $y_0$  ou individuelle  $y_{0\text{-ind}}$ .

L'incertitude associée est plus difficile à calculer que dans le cas de l'évaluation directe. La difficulté provient de l'incertitude associée à la droite. Cette incertitude est évaluée ci-dessous dans le cas général.

#### 8.4.1. Incertitude associée à l'utilisation inverse de la droite

L'équation (31) est utilisée dans le sens inverse, soit :

$$y_0 = (x_0 - b_0) / b_1. \quad (39)$$

L'incertitude associée est généralement évaluée en utilisant cette équation (39). En simplifiant, on obtient :

$$u(y_{0\text{-inv-droite}}) = \sqrt{\frac{u^2(b_0) + y_0^2 u^2(b_1) + 2y_0 u(b_0, b_1)}{b_1^2}} \quad (40)$$

Lorsque la droite estimée est symétrique, ce calcul convient ; l'expression (40) est équivalente à (32). C'est le cas avec la méthode GLS-GGMR.

Lorsque la droite estimée dépend du sens de la régression, la valeur  $y_0$  calculée suivant (39) diffère de celle calculée suivant (31). Pour tenir compte de l'écart entre ces deux valeurs, l'expression (39) devrait intégrer un terme correctif qui induit une composante d'incertitude supplémentaire.

Par exemple,  $u(y_{0\text{-inv-droite}})$  pourrait être déterminée par renvoi de la valeur  $y_0$  sur l'intervalle d'incertitude de la droite (Le calcul est explicité dans [14]). Cette méthode présente les inconvénients suivants :

- la solution est graphique ou elle est obtenue à l'issue d'un calcul compliqué ;
- lorsque l'intervalle d'incertitude de la droite a l'allure d'une hyperbole (cas des OLS),  $u(y_{0\text{-inv-droite}})$  n'est pas symétrique ;
- comment combiner cette incertitude avec les autres composantes  $u(l_0)$  et  $u(e_0)$  ?

#### 8.4.2. Incertitude

L'incertitude type associée au résultat moyen  $z_0$  est donnée par (40) en remplaçant  $y_0$  par  $z_0$ , ou par une méthode adéquate.

Cette incertitude est augmentée des composantes issue des valeurs individuelles :

- soit  $u(l_0)$  pour un résultat individuel  $z_{0\text{-ind}}$  :

$$u^2(z_{0\text{-ind}}) = u^2(z_{0\text{-inv-droite}}) + u^2(l_0) \quad (41)$$

- ou  $u(u_0)$  pour une indication individuelle :

$$u^2(z_0) = u^2(z_{0\text{-inv-droite}}) + u^2(e_0) / b_1^2. \quad (42)$$

L'incertitude maximale est obtenue en considérant les valeurs individuelle :

$$u^2(z_{0\text{-ind}}) = u^2(z_{0\text{-inv-droite}}) + u^2(l_0) + u^2(e_0) / b_1^2. \quad (43)$$

## 8.5. Conclusion

Le résultat de mesure est généralement évalué par la seule valeur moyenne  $z_0$ .

Par contre, l'incertitude associée varie avec les sources d'incertitude considérées. Il est donc important de mentionner ces sources d'incertitude pour faciliter la comparaison de résultats.

Ces sources d'incertitudes peuvent être :

- la méthode d'estimation utilisée : les OLS qui sous-estiment l'incertitude, le calcul direct ou l'évaluation inverse ;
- le mesurande ciblé : résultat moyen ou résultat individuel affecté d'une composante d'incertitude supplémentaire ;
- l'indication correspondante qui peut être moyenne ou individuelle.

Les calculs précédents supposent des composantes d'incertitude indépendantes. Par exemple, l'indication  $y_0$  est sans lien avec les indications antérieures  $y_i$  utilisées pour estimer la droite d'étalonnage ; et donc elle est indépendante des coefficients de la droite estimée. Dans certains cas, cette hypothèse n'est pas justifiée et les covariances sont à déterminer. En cas de doute, il est possible de maximiser l'incertitude en sommant linéairement les termes d'incertitude.

## 9. Validation des estimations

La droite d'étalonnage estimée et des résultats de mesure calculés avec la droite sont validés avant de clore le processus d'estimation. La validation comprend deux volets complémentaires : la validation statistique et la validation métrologique.

### 9.1. Validation statistique

Elle concerne plutôt la droite estimée. Il est vérifié que les hypothèses du modèle sont cohérentes avec les résultats obtenus.

L'outil principal est un test de validation globale qui porte sur les écarts des  $n$  points à la droite. Il s'agit du test de Fisher pour les OLS et du test du  $\chi^2$  pour les GLS. Ces tests supposent que les variables sont distribuées selon une loi normale. Puis, les outils de validation partielle habituellement présentés dans le cadre des OLS sont appliqués.

#### 9.1.1. Le test de Fisher

Les OLS minimisent les écarts entre les points et la droite quantifiés par la variance résiduelle  $S_{OLS}^2 = (\mathbf{D}^T \mathbf{D}) / (n-2)$  (§.3). Si la droite représente bien les données, la variance expliquée par la droite (notée  $S_{\text{expliquée}}^2$ ) sera très élevée par rapport à cette variance résiduelle  $S_{OLS}^2$ .

Le rapport  $F = S_{\text{expliquée}}^2 / S_{OLS}^2$  sera alors très grand. Sous l'hypothèse de normalité des variables, ce rapport suit une loi de Fisher à  $(n-2, 1)$  degrés de liberté et on détermine la valeur limite au niveau de signification de 5 % que  $F$  doit atteindre pour conclure que la droite explique les données (l'hypothèse testée est « la droite n'explique pas les données » et le test est unilatéral à 5 % ;  $S_{\text{expliquée}}^2$  ne doit pas être trop élevée).

En général, le test est concluant et il ne suffit pas pour valider le modèle. Il est complété par les tests et outils présentés au paragraphe 9.1.3.

#### 9.1.2. Le test du $\chi^2$ et le ratio de Birge

Dans la méthode GLS, l'hypothèse de départ est différente : les incertitudes associées aux variables expliquent les écarts des points à la droite. Le critère minimisé,  $\mathbf{D}^T (\mathbf{Y})^{-1} \mathbf{D}$  (7) ou  $\mathbf{S}(\mathbf{B})$  (22), représente le rapport de ces deux quantités, avec la matrice d'incertitude connue au dénominateur. Si ces quantités sont cohérentes, le critère n'est ni trop élevé, ni trop faible. Dans l'hypothèse de normalité des variables, ce rapport est un  $\chi^2$  à  $(n-2)$  degrés de liberté et on détermine un intervalle de confiance bilatéral à 95 % (Dans [12] le test est unilatéral et il porte sur la limite supérieure) pour les valeurs possibles de  $\mathbf{S}(\mathbf{B})$ . Si  $\mathbf{S}(\mathbf{B})$  calculé sort de l'intervalle, par valeur inférieure ou par valeur supérieure, le modèle n'est pas validé (la minimisation de  $\mathbf{S}(\mathbf{B})$  peut sembler incompatible avec la recherche d'une valeur ni trop élevée, ni trop faible). Lorsque le nombre de points d'étalonnage est faible, l'intervalle de confiance du  $\chi^2$  est large et le modèle sera souvent validé.

En pratique, il est plus simple de calculer le ratio de Birge [15], d'expression  $\mathbf{D}^T \mathbf{V}(\mathbf{Y})^{-1} \mathbf{D} / (n-2)$  ou  $\mathbf{S}(\mathbf{B}) / (n-2)$  ; dont l'espérance mathématique est 1, quel que soit le nombre  $n$  de données d'étalonnage.

Alors :

- ratio de Birge  $> 1$  suggère que la droite estimée n'est pas validée car elle n'explique pas suffisamment les données. On peut chercher à changer de modèle et par exemple à ajuster un polynôme de degré plus élevé. Mais la cause de l'inadéquation peut provenir des incertitudes des variables en entrée qui sont « trop faibles » comparées aux écarts au modèle.
- au contraire, ratio de Birge  $< 1$  signifie que les incertitudes en entrée sont « plus élevées » que les écarts au modèle. Dans ce cas, l'analyse doit porter sur les incertitudes associées aux variables.

Dans les exemples testés, les covariances des variables avaient été omises. Or, ajouter des covariances positives peut être équivalent à diminuer l'incertitude en entrée. Ces covariances proviendraient d'erreurs systématiques qui n'ont pas influencé les données dans l'unique expérience d'étalonnage considérée.

### 9.1.3. Tests de validation partielle et autres outils statistiques

Les écarts  $\mathbf{D}$  ou transformés  $\mathbf{P}^{-1}\mathbf{D}$  sont examinés : vérification de la répartition aléatoire, de l'absence de corrélation et d'une moyenne proche de 0. Lorsque aucune tendance ou valeur aberrante n'est constatée, les hypothèses sur les erreurs aléatoires sont validées.

Puis, le caractère significatif des coefficients du modèle est testé. En général, la droite d'étalonnage est écrite avec un terme constant sans raison précise.

Une droite d'étalonnage validée globalement par un des tests précédents a une pente significative. C'est donc la constante qui sera testée avec un test de Student et retirée du modèle si elle n'est pas significative. Comme les valeurs de la pente et de la constante covariant, la variation de valeur de la pente sera nette.

### 9.2. Validation métrologique

Cette validation concerne les valeurs calculées et les incertitudes associées. Dans de nombreux cas, le responsable de l'étalonnage a une bonne connaissance de la fonction d'étalonnage et il est plutôt intéressé par l'incertitude associée à cette fonction et sa propagation. Ce métrologue connaît à l'avance ou cible le niveau d'incertitude des résultats de mesure calculés avec la fonction. Comme les méthodes étudiées donnent des estimations très proches, l'efficacité d'une méthode se traduit dans ces incertitudes finales.

Les outils de la validation métrologique sont notamment :

- la comparaison des résultats obtenus avec d'autres méthodes comme les simulations de Monte-Carlo ;
- la détermination avec la fonction d'étalonnage de valeurs étalons connus ;
- l'analyse des résultats en considérant la connaissance du processus ; par exemple, comparer la droite d'étalonnage estimée au temps  $t$  avec celles obtenues lors des étalonnages précédents.

### 9.3. Conclusion

Il n'existe pas de procédure particulière exposant la démarche à suivre pour la validation. Les outils statistiques et métrologiques présentés seront utilisés en fonction des besoins, des compétences vis-à-vis du contexte de l'étalonnage et de l'aisance à les manipuler.

Une procédure rapide pour valider la fonction d'étalonnage serait d'utiliser le test statistique global et l'estimation de quelques résultats de mesure déjà connus. Si les résultats obtenus sont corrects et l'incertitude associée proche de l'incertitude ciblée, le modèle d'étalonnage est correct. Il est alors possible de le confirmer en utilisant les autres outils de validation.

Tableau 1

Données  $X$  et  $Y$  : valeurs et incertitudes types associées.

$x$	$u(x)$	$y$	$u(y)$
1,0	0	3,3	0,5
2,0	0	5,6	0,5
3,0	0	7,1	0,5
4,0	0	9,3	0,5
5,0	0	10,7	0,5
6,0	0	12,1	0,5

Tableau 2

Estimations de  $Y = b_0 + b_1 \cdot X$  : OLS et GLS.

	OLS	OLS variance corrigée	GLS $u(y) = 0,5$
$b_0$	1,867		1,867
$b_1$	1,757		1,757
$u(b_0)$	0,3	0,47	0,47
$u(b_1)$	0,077	0,12	0,12
$u(b_0, b_1)$	-0,020 8	-0,05	-0,05
$u(y)$ -estimée	0,323		
ratio de Birge			0,645
Validation	oui		oui

Lorsque les estimations ne sont pas validées, les données d'entrée de la méthode sont modifiées et la droite est ré-estimée. L'estimation de la droite d'étalonnage est un processus itératif à trois étapes : hypothèses, estimation, validation. Le processus est achevé lorsque les résultats confirment les hypothèses.

À l'issue de la validation, le laboratoire d'étalonnage a amélioré sa connaissance des valeurs utilisées dans la fonction d'étalonnage.

## 10. Exemples

Deux exemples sont présentés dans cette partie. Le premier exemple est un cas simple issu de la norme ISO 28037 [12] dans lequel une incertitude constante est associée aux valeurs de  $Y$ . Cet exemple, plus général que la droite d'étalonnage, permet de comparer les méthodes OLS et GLS sous différentes hypothèses. Le deuxième exemple traite de l'analyse de la masse de benzène contenue dans une cartouche avec des incertitudes variables sur  $X$  et  $Y$ . Les calculs ont été effectués avec le logiciel Regpoly développé au LNE.

### 10.1. Exemple 1 : Equal weights de la norme ISO 28037

Le tableau 1 reprend les données de la Table 4 du paragraphe 6.4 de [12]. Les grandeurs représentées par  $X$  et  $Y$  ne sont pas spécifiées.

Ce cas correspond au modèle de régression standard avec  $u(y)$  connue et constante. La méthode adaptée est celle des GLS avec des poids  $u^{-1}(y) = 1/0,5$ . La méthode OLS suppose que  $u(y)$  est inconnue.

L'estimation de la droite est présentée au tableau 2.

Tableau 3

Matrice de variances-covariances des valeurs  $y$ .

0,25	0,175	0,175	0,175	0,175	0,175
0,175	0,25	0,175	0,175	0,175	0,175
0,175	0,175	0,25	0,175	0,175	0,175
0,175	0,175	0,175	0,25	0,175	0,175
0,175	0,175	0,175	0,175	0,25	0,175
0,175	0,175	0,175	0,175	0,175	0,25

Tableau 4

Estimations GLS obtenues en modifiant les incertitudes des valeurs  $y$ .

	GLS $u(y) = 0,25$	GLS $y$ corrélé
$b_0$	1,867	
$b_1$	1,757	
$u(b_0)$	0,233	0,490
$u(b_1)$	0,060	0,065
$u(b_0, b_1)$	-0,012 5	-0,015
ratio de Birge	1,290	1,178
Validation	oui	oui

### 10.1.1. Commentaires

Conformément aux résultats du paragraphe 4.2, les deux méthodes estiment la même droite ( $\mathbf{B}_{OLS} = \mathbf{B}_{GLS}$ ).

Les OLS sont validés globalement avec le test de Fisher. Mais, le test ne porte pas sur les incertitudes. Or, la méthode sous-estime  $u(y)$  (0,322 au lieu de 0,5) et  $\mathbf{V}(\mathbf{B}_{OLS})$  ( $S_{OLS}^2 (\mathbf{X}'\mathbf{X})^{-1}$  est inadéquate) correspond à  $\mathbf{V}(\mathbf{B}_{GLS})$  lorsqu'elle est corrigée.

Les GLS sont validés par le test du  $\chi^2$ . Le ratio de Birge avec une valeur inférieure à 1 indique que les incertitudes associées aux  $y$  sont supérieures aux écarts à la droite. Dans la suite, ces incertitudes sont modifiées pour améliorer l'accord entre la droite et les données.

### 10.1.2. Modifier l'incertitude associée aux valeurs $y$

Une première solution consiste à diminuer  $u(y)$ . Par exemple, poser  $u(y) = 0,25$  pour diviser par 2 l'incertitude considérée lors de l'estimation.

On peut noter que  $u(y)$ -estimée des OLS correspond à  $u(y)$  multipliée par le ratio de Birge, et que l'utilisation des GLS avec  $u(y)$ -estimée donne un ajustement parfait avec un ratio de Birge proche de 1 [15].

Une autre solution consiste à ajouter des covariances pour étudier leur influence. Par exemple, une corrélation constante  $r = 0,7$  est supposée pour toutes les valeurs  $y$ . 30 covariances  $\text{cov}(y_i, y_j) = 0,7 \times 0,5 \times 0,5 = 0,175$  sont introduites dans le système (tab. 3).

Le tableau 4 présente les estimations GLS obtenues avec ces modifications.

Les valeurs des coefficients sont inchangées. Les incertitudes associées ont été divisées par 2 lorsque  $u(y) = 0,25$  et elles sont inférieures à celles des OLS.

Dans le cas  $y$  corrélé, la variation des incertitudes des coefficients est inégale : l'incertitude de  $b_0$  augmente

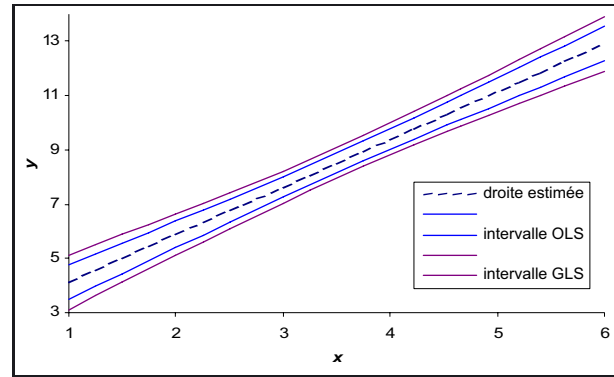


Fig. 1. – Intervalles d'incertitude de la droite obtenus avec les OLS et les GLS sans corrélation.

légèrement; l'incertitude de  $b_1$  est divisée par quasiment 2 et la covariance par 3.

L'ajustement est très satisfaisant avec un ratio de Birge proche de 1. Les estimations des coefficients ont été validées avec des simulations de Monte-Carlo.

Le ratio de Birge supérieur à 1 dans les deux cas indique que les incertitudes des valeurs  $y$  sont maintenant plus faibles que les écarts au modèle.

### 10.1.3. Prévisions

L'intervalle d'incertitude associé à la droite est représenté à la figure 1. pour les deux méthodes. Les deux intervalles ont la même forme; mais l'intervalle des GLS est plus large.

La droite estimée est utilisée dans les deux sens pour prévoir une nouvelle valeur. L'incertitude est évaluée en considérant les deux cas : prédicteur avec et sans incertitude.

L'exemple de la partie 11 de [12] comprend les prévisions suivantes :  $x_0 = 4,913$  à  $y_0 = 10,5$ ;  $y_0 = 8,017$  à  $x_0 = 3,5$ . Les incertitudes associées sont présentées dans les tableaux 5 et 6.

L'incertitude de ces prévisions varie avec les méthodes et l'incertitude  $u(y)$ .

Comme attendu, les incertitudes ont été divisées par deux avec  $u(y) = 0,25$ . Elles sont inférieures aux incertitudes obtenues avec les OLS.

Par contre, les incertitudes ont augmenté dans le cas  $y$  corrélé; alors que l'incertitude de la pente  $b_1$  avait diminué. Cela s'explique par la covariance négative entre les 2 coefficients qui a été beaucoup plus réduite.

Cet exemple montre l'influence des covariances qui actuellement sont trop souvent négligées lors de l'estimation. Supposer une autre configurations des corrélations est intéressant, mais l'étude de l'influence des covariances dépasse le cadre de cet article. L'objectif est d'alerter sur ce thème.

Comme le modèle prévoit une valeur moyenne, l'incertitude  $u(y_{0\text{-droite}})$  due au modèle est plus faible que

Tableau 5

Incertitude de  $x_0 = 4,913$  prévue à  $y_0 = 10,5$  évaluée dans différents contextes.

composantes d'incertitude	OLS	GLS $u(y) = 0,5$	GLS $u(y) = 0,25$	GLS $y$ corrélé
$u(x_{0\text{-inv-droite}})$	0,097	0,15	0,075	0,25
$u(x_{0\text{-inv-droite}}), u(y_0) = 0,5$	0,30	0,32	0,29	0,38
$u(x_{0\text{-inv-droite}})$ droite inversée	0,106	0,177	pas évalué	pas évalué

Tableau 6

Incertitude de  $y_0 = 8,017$  prévue à  $x_0 = 3,5$  évaluée dans différents contextes.

composantes d'incertitude	OLS	GLS $u(y) = 0,5$	GLS $u(y) = 0,25$	GLS $y$ corrélé
$u(y_{0\text{-droite}})$	0,13	0,20	0,10	0,43
$u(y_{0\text{-droite}}), u(x_0) = 0,2$	0,38	0,41	0,37	0,56

l'incertitude  $u(y)$  initiale (tab. 6). Par ailleurs, la prise en compte de la non symétrie de la droite estimée induit une incertitude  $u(x_{0\text{-inv-droite}})$  légèrement plus élevée pour  $x_0$  (tab. 5).

Hormis les GLS corrélés, l'écart entre les méthodes devient négligeable lorsque l'incertitude du prédicteur est prépondérante. Dans ce cas, il devient possible d'utiliser une méthode plus simple quoiqu'approximative.

En conclusion, les GLS propagent les incertitudes des variables  $y$  aux valeurs calculées avec le modèle, lorsque l'information utilisée est correcte. Des résultats non validés indiquent une incohérence dans le modèle qui devrait donner lieu à une analyse des données de l'étalonnage.

## 10.2. Exemple 2 : Étalonnage pour le dosage du benzène

La masse de benzène contenue dans un tube est dosée par chromatographie en phase gazeuse couplée à un détecteur à ionisation de flamme. La gamme de mesure s'étend de 300 ng à 2 000 ng. L'étalonnage est réalisé en 5 points répétés 5 fois ; le point milieu de la gamme est repassé en fin de cycle. Comme la mesure est destructive, 26 tubes étalons sont analysés.

L'incertitude associée à une masse étalon provient principalement de la stabilité, de la reproductibilité et du blanc. Ces facteurs corréleront les valeurs des masses étalons ; la corrélation de deux valeurs est environ  $r = 0,98$ .

L'incertitude sur une aire de pic provient de la reproductibilité de la mesure. À chaque point d'étalonnage, le coefficient de variation (CV) estimé sur les 5 ou 6 mesures est utilisé pour estimer l'incertitude associée à l'aire de pic :  $u(\text{Aire-de-pic}) = CV \cdot \text{Aire-de-pic}$ .

Les données de l'étalonnage figurent dans le tableau 7. La matrice de variances-covariances des masses étalons de format (26 lignes, 26 colonnes) n'est pas présentée dans cet article.

Le modèle estimé est  $\text{Masse} = b_0 + b_1 \text{Aire-de-pic}$  ; il est écrit dans le sens d'utilisation (fonction  $g$  de la partie 2).

Dans ce cas où les deux variables sont connues avec une incertitude, la méthode complète GLS-GGMR est

Tableau 7

Données de l'étalonnage.

Masse (ng)	$u(\text{Masse})$ (ng)	Aire-de-pic	$u(\text{Aire-de-pic})$
300,26	5,49	491 628	2 066
300,18	5,49	486 838	2 046
300,19	5,49	491 277	2 065
300,10	5,49	491 483	2 065
300,21	5,49	491 362	2 065
399,39	7,28	648 633	2 452
399,38	7,28	653 781	2 472
399,40	7,28	652 774	2 468
399,32	7,28	653 749	2 472
399,32	7,28	649 421	2 455
500,60	9,11	811 400	4 544
501,65	9,13	818 169	4 582
500,53	9,11	824 575	4 618
500,59	9,11	813 778	4 558
500,50	9,10	819 287	4 589
500,62	9,11	817 205	4 577
599,76	10,90	982 352	3 439
599,64	10,90	983 872	3 444
599,74	10,90	979 059	3 428
599,65	10,90	988 175	3 460
599,65	10,90	985 733	3 451
701,02	12,73	1 141 976	5 557
701,07	12,74	1 155 644	5 623
701,10	12,74	1 151 403	5 603
700,99	12,73	1 155 300	5 622
700,98	12,73	1 153 303	5 612

préconisée. Elle est comparée aux OLS et aux GLS-simples, en intégrant ou en négligeant la corrélation des masses étalons.

### 10.2.1. Estimation de la droite d'étalonnage

Les résultats figurent dans le tableau 8.

Les GLS avec corrélation donnent des résultats très satisfaisants avec un ratio de Birge proche de 0,9. Les covariances ont une grande influence sur la qualité de l'estimation puisque le modèle n'est pas validé lorsqu'elles sont négligées. Les GLS-simples fournissent une bonne approximation des résultats GLS-GGMR.

Tableau 8  
Estimations de  $Masse = b_0 + b_1 Aire-de-pic$  : OLS et GLS.

	OLS	GLS-GGMR $r = 0$	GLS-GGMR $r = 0,98$	GLS-simples $r = 0,98$
$b_0$	4,3	2,8	2,8	2,8
$b_1$	$6,056 \times 10^{-4}$	$6,07 \times 10^{-4}$	$6,065 \times 10^{-4}$	$6,04 \times 10^{-4}$
$u(b_0)$	1,8	5,5	1,45	1,45
$u(b_1)$	$2,07 \times 10^{-5}$	$7,6 \times 10^{-6}$	$1,1 \times 10^{-5}$	$1,1 \times 10^{-5}$
$u(b_0, b_1)$	$-3,5 \times 10^{-6}$	$-1,0 \times 10^{-5}$	$-1,6 \times 10^{-6}$	$-1,6 \times 10^{-6}$
$u(Masse)$	2,452 7			
ratio de Birge		0,251	0,899	0,900
Validation	oui	non	oui	oui

Tableau 9  
Masses calculées avec le modèle.

Aire-de-pic	OLS		GLS-GGMR $r = 0,98$		GLS-simples $r = 0,98$	
	$y_0$ (ng)	$u(y_0)$ (ng)	$y_0$ (ng)	$u(y_0)$ (ng)	$y_0$ (ng)	$u(y_0)$ (ng)
750 000,00	458,48	0,50	457,7	8,3	455,5	8,3
800 000,00	488,76	0,48	488,0	8,8	485,7	8,8
910 000,00	555,37	0,51	555,00	10,00	552,00	10,00
1 000 000,00	609,88	0,61	609,00	11,00	606,00	11,00
1 100 000,00	670,44	0,75	670,00	12,00	667,00	12,00

Les OLS et les GLS-GGMR avec corrélation estiment différemment les coefficients de la droite :

- les incertitudes associées aux valeurs de la constante sont élevées et dans le cas GLS-GGMR, elle pourrait être retirée du modèle, puisqu'elle n'est pas significative;
- l'incertitude de la pente est beaucoup plus élevée avec les GLS qu'avec les OLS. Les deux estimations sont différentes même en tenant compte des incertitudes.

### 10.2.2. Prévisions

La droite estimée par les GLS-GGMR avec corrélation est représentée avec son intervalle d'incertitude à la figure 2.

La droite des OLS n'est pas représentée; l'écart entre les deux droites (y compris l'intervalle d'incertitude OLS) est négligeable comparé à l'incertitude des GLS. Cette incertitude augmente linéairement avec  $x$ .

La masse de benzène a été évaluée avec les deux droites dans le domaine 450 ng–670 ng. Il s'agit d'une prévision en valeur moyenne sans incertitude sur le prédicteur; donc l'incertitude associée provient uniquement de la droite (32). Les résultats sont présentés dans le tableau 9.

Les incertitudes évaluées par les GLS-GGMR représentent environ 1,8 % de la valeur prévue. Cette incertitude correspond à l'incertitude des masses étalons

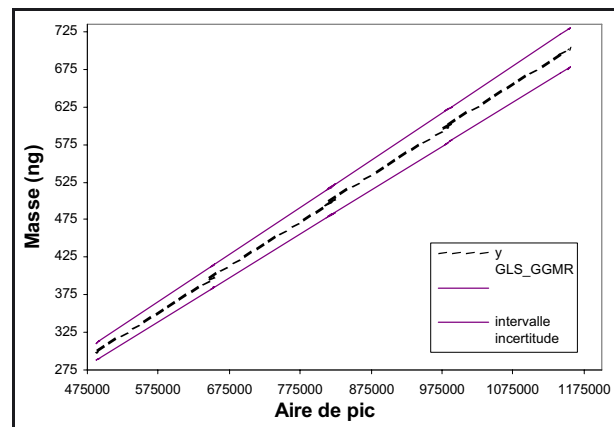


Fig. 2. – Droite estimée par les GLS-GGMR avec corrélation.

spécifiée en entrée du modèle sans toutefois la dépasser. La prévision calculée étant une valeur moyenne.

Les GLS-simples donnent les mêmes incertitudes que les GLS-GGMR quoiqu'elles soient systématiquement plus faibles avant d'arrondir les valeurs. L'incertitude obtenue avec les OLS est dérisoire. Comme attendu, les valeurs des masses sont très proches et elles sont équivalentes compte tenu de l'incertitude des GLS.

### 10.2.3. Modèle inverse

Dans une deuxième étape, le modèle inverse est évalué :  $Aire-de-pic = b'_0 + b'_1 \cdot Masse$ . Il s'agit d'étudier la symétrie de la droite estimée et de comparer les deux modes

Tableau 10  
Estimations de « Aire-de-pic =  $b'_0 + b'_1 \cdot \text{Masse}$  » : OLS et GLS.

	OLS	GLS-GGMR $r = 0$	GLS-GGMR $r = 0,98$	GLS-simples $r = 0,98$
$b'_0$	-686 4	-4 593	-4 611	-4581
$b'_1$	1 650,80	1 646,00	1 649,00	1 646,00
$u(b'_0)$	2 934,00	9 093,00	2 406,00	2 405,00
$u(b'_1)$	5,7	21,00	30,00	30,00
$u(b'_0, b'_1)$	-16 000	-180 000	-9 800,00	-9 820,00
$u(\text{Aire-de-pic})$	4 005,00			
ratio de Birge		0,251	0,899	0,899
Validation	oui	non	oui	oui

Tableau 11  
Détermination des coefficients  $b_0$  et  $b_1$  : avec le modèle inverse.

	OLS	GLS-GGMR $r = 0$	GLS-GGMR $r = 0,98$	GLS-simples $r = 0,98$
$b_0 = -b'_0/b'_1$	4,157	2,790 1	2,796 3	2,782 87
$u(b_0)$	6 1,77	5,54	1,46	1,46
$b_1 = 1/b'_1$	$6,057 \times 10^{-4}$	$6,074 7 \times 10^{-4}$	$6,065 \times 10^{-4}$	$6,075 \times 10^{-4}$
$u(b_1)$	$2,1 \times 10^{-6}$	$7,64 \times 10^{-6}$	$1,1 \times 10^{-5}$	$1,1 \times 10^{-5}$

de calcul des résultats. L'inversion de cette deuxième droite fournit une nouvelle estimation du modèle initial. Les résultats sont regroupés dans les tableaux 10 et 11.

Comme la méthode GLS-GGMR est symétrique, les résultats du tableau 11 correspondent aux estimations initiales du tableau 8 ; y compris le ratio de Birge. La méthode estime donc les mêmes valeurs de masse par évaluation directe et par évaluation inverse.

L'Annexe 2 présente l'écran de calcul du logiciel Regpoly, où les résultats du modèle inverse sont affichés.

Les autres méthodes donnent des résultats équivalents aux incertitudes près.

## 11. Conclusions et perspectives

La méthode des GLS est adaptée à différentes structures d'incertitude sur les variables (hétéroscédasticité, autocorrélation voire corrélation de  $X$  et  $Y$ ). Dans tous les cas, la procédure appliquée est : calcul des poids avec les incertitudes sur les variables et validation des estimations en comparant ces incertitudes avec les écarts des points à la droite d'étalonnage.

Cet exercice conduit généralement à une meilleure connaissance du processus d'étalonnage. Dans la plupart des exemples étudiés, les incertitudes sur les variables ont été revues et des covariances ont été prises en compte.

Evidemment, plus l'incertitude sur les variables présente une structure complexe, plus les calculs inhérents à la méthode deviennent compliqués. Aujourd'hui, des algorithmes sont proposés dans des publications et le LNE prévoit de communiquer sur le logiciel Regpoly.

Les GLS permettent de propager les incertitudes sur les variables aux résultats de mesure calculés avec la droite. La question du sens de modélisation n'intervient plus puisque la méthode est symétrique. Dans les exemples étudiés, les OLS donnent des estimations très proches de celles obtenues avec les GLS mais les incertitudes associées sont très inférieures.

Le calcul du résultat de mesure a été présenté en distinguant la valeur moyenne sans incertitude et une valeur individuelle avec incertitude. En général, les laboratoires d'étalonnage utilisent l'indication moyenne et les utilisateurs de l'instrument une indication individuelle avec son incertitude associée. Le résultat de mesure calculé est pour l'instant un résultat moyen vérifiant exactement la droite d'étalonnage. Il est nécessaire de l'indiquer puisque l'incertitude associée correspond aux incertitudes en entrée quoiqu'elle soit légèrement inférieure. Pour retrouver l'incertitude des valeurs en entrée, il faudrait considérer le résultat individuel.

Cet article concerne le cas simple où la fonction d'étalonnage est une droite. L'étude des fonctions polynomiales est en cours. Les algorithmes utilisés seront les mêmes mais les calculs matriciels seront plus complexes.

La difficulté concerne surtout l'évaluation inverse du résultat de mesure, tant pour la valeur du résultat que pour l'incertitude associée.

Enfin, il est prévu d'améliorer l'utilisation des simulations de Monte-Carlo qui actuellement valide les résultats obtenus et d'appliquer les techniques Bayésiennes au cas de la droite d'étalonnage.



## Annexe 1

### La droite de régression sous forme matricielle : quelques éléments

Vecteurs des valeurs de  $y$  et de  $x$  :

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix}$$

Modèle :

$$Y = X\beta + E$$

(n, 1)    (n, 2)(2, 1)    (n, 1)

Estimateur des OLS :

$$\mathbf{B}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \begin{pmatrix} n & \sum x_t \\ \sum x_t & \sum x_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum x_t y_t \end{pmatrix}$$

Il correspond au système d'équations suivant :

$$\begin{pmatrix} n & \sum x_t \\ \sum x_t & \sum x_t^2 \end{pmatrix} \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} = \begin{pmatrix} \sum y_t \\ \sum x_t y_t \end{pmatrix}$$

$$\mathbf{B}_{OLS} = \begin{pmatrix} \bar{y} - B_1 \bar{x} \\ \frac{\sum (x_t - \bar{x})(x_t - \bar{y})}{\sum (x_t - \bar{x})^2} \end{pmatrix}$$

Estimateur des GLS quand  $\mathbf{V}(\mathbf{Y})$  est diagonale :

$$\mathbf{B}_{GLS} = (\mathbf{X}^T \mathbf{V}(\mathbf{Y})^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \begin{pmatrix} \sum \frac{1}{\sigma_{y_t}^2} & \sum \frac{x_t}{\sigma_{y_t}^2} \\ \sum \frac{x_t}{\sigma_{y_t}^2} & \sum \frac{x_t^2}{\sigma_{y_t}^2} \end{pmatrix} \begin{pmatrix} \sum \frac{y_t}{\sigma_{y_t}^2} \\ \sum \frac{x_t y_t}{\sigma_{y_t}^2} \end{pmatrix}$$

Il correspond au système d'équations suivant :

$$\begin{pmatrix} \sum \frac{1}{\sigma_{y_t}^2} & \sum \frac{x_t}{\sigma_{y_t}^2} \\ \sum \frac{x_t}{\sigma_{y_t}^2} & \sum \frac{x_t^2}{\sigma_{y_t}^2} \end{pmatrix} \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} = \begin{pmatrix} \sum \frac{y_t}{\sigma_{y_t}^2} \\ \sum \frac{x_t y_t}{\sigma_{y_t}^2} \end{pmatrix}$$

$$\mathbf{B}_{GLS} = \begin{pmatrix} \bar{y}_w - B_1 \bar{x}_w \\ \frac{\sum w_t (x_t - \bar{x}_w)(x_t - \bar{y}_w)}{\sum w_t (x_t - \bar{x}_w)^2} \end{pmatrix}$$

où

$$\bar{y}_w = \frac{\sum y_t / \sigma_{y_t}^2}{\sum 1 / \sigma_{y_t}^2} \quad \text{et} \quad \bar{x}_w = \frac{\sum x_t / \sigma_{y_t}^2}{\sum 1 / \sigma_{y_t}^2}$$

$$w_t = \frac{1 / \sigma_{y_t}^2}{\sum 1 / \sigma_{y_t}^2}$$

Les sommes symbolisées par  $\Sigma$  portent sur l'indice  $t$  ( $t = 1, \dots, n$ ).

## Annexe 2

### Résultats obtenus avec le modèle inverse

#### Ecran de calcul et d'affichage des résultats du logiciel Regpoly

The screenshot shows the 'REGPOLY : Estimation - Prévission' window. It is divided into several sections:

- Données d'entrée (I):** 'Données x, y' set to 'BenzeneClaire2012cov', 'Nb points' is 26. Buttons for 'Afficher' and 'variances' are present.
- Degré du polynôme (II):** Set to 1.
- Méthode d'estimation (III):** Set to '4 - GLS\_GGMR'.
- Hypothèses de la méthode:** Text explaining the assumptions for x and y values.
- Estimations:** A table showing polynomial coefficients (b, ub, s) and a list of 'x' values with their corresponding 's' values.
- Graphique des données:** A plot of 'Aire' vs 'Masse' showing observed data points (red circles), a modified y curve (blue line), and a modified U(y) curve (green line).
- Test du Khi-deux:** 'Chi2 = -19.376', 'Ratio de Birge = 0.89852'. A message states 'le modèle est validé avec le test du Khi-deux'.
- Prévission (IV):** Sections for 'x0 connues' and 'y0 connues' with input fields and 'Afficher' buttons.

Estimations des valeurs des masses X

	valeur x	incertitude-type u(x)
1	306.1343	5.4762
2	299.2623	5.4707
3	306.0231	5.4747
4	299.9977	5.4735
5	306.0524	5.4751
6	397.9514	7.2489
7	398.9535	7.2548
8	398.7718	7.2539
9	398.9120	7.2539
10	398.0654	7.2488
11	498.9718	9.0861
12	506.5779	9.1672
13	506.3382	9.0899
14	499.2248	9.0869
15	499.7529	9.0875
16	499.6228	9.0887
17	598.7750	10.8626
18	599.0197	10.8628
19	598.0525	10.8576
20	599.5398	10.8688
21	599.4277	10.8655
22	698.9653	12.7010
23	706.8323	12.7095
24	706.2970	12.7077
25	706.7304	12.7081
26	706.4594	12.7068

Prévissions des valeurs des masses avec la droite estimée

	1	2
1	457.6573	8.2709
2	487.9814	8.8157
3	554.6943	10.0171
4	609.2777	11.0025
5	669.9250	12.0991

## Références

- [1] BICH W., “Interdependence between measurement uncertainty and metrological traceability”, *Accred. Qual. Assur.*, **14**, 2009, 581–586.
- [2] JOHNSTON J., “Econometric methods”, *International student edition*, 1972.
- [3] FULLER W.A., “Measurement error models”, *Wiley*, 1987.
- [4] CHENG C.L. et VAN NESS J.W., “Statistical regression with measurement error”, London, *Oxford university press*, 1999.
- [5] MILTON M.J.T., HARRIS M., SMITH I.M., BROWN A.S. et Goody B.A., “Implementation of a generalized least-squares method for determining calibration curves from data with general uncertainty structures”, *Metrologia*, **43**, 2006, S291–S298.
- [6] ISO, NF ENV 13005, août 1999, « Guide pour l’expression de l’incertitude de mesure ».
- [7] ISO, ISO/CEI GUIDE 99:2007, décembre 2007, « Vocabulaire international de métrologie – Concepts fondamentaux et généraux et termes associés (VIM) ».
- [8] CHENG C.L. et RIU J., “On estimating linear relationships when both variables are subject to heteroscedastic measurement errors”, *Technometrics*, **48**, 4, 2006.
- [9] YARDIN C. et KABA S., “Uncertainty of standard concentrations, calibration line and predictions”, *Congrès AMCTM*, 2008, livre VIII.
- [10] RIU J. et RIUS F.X., “Univariate regression models with errors in both axes”, *Journal of chemometrics*, **9**, 1995, 343–362.
- [11] COX M.G., FORBES A.B. et HARRIS P.M., “The classification and solution of regression problems for calibration”, *NPL Report*, CMSC 24/03, 2004.
- [12] ISO/DTS TS 28037 2009–01–21, “The determination and use of straight-line calibration functions”.
- [13] WILLINK R., “Estimation and uncertainty in fitting straight lines to data : different techniques”, *Metrologia*, **45**, 2008, 290–298.
- [14] DRAPER N. et SMITH H., “Applied regression analysis”, *Wiley*, second edition, 1981.
- [15] LIRA I., “Evaluating the measurement uncertainty”, *Institute of Physics Publishing*, 2002.
- [16] DAGNELIE P., « Statistique théorique et appliquée », Tome 1, *Bibliothèque des universités*, 1998.

---

Article reçu le 28 septembre 2012, version révisée reçue le 24 avril 2013.