

SIMONE HUYBERECHTS

## **Estimation des paramètres dans le modèle linéaire général dynamique**

*RAIRO. Recherche opérationnelle*, tome 13, n°2 (1979),  
p. 143-149

[http://www.numdam.org/item?id=RO\\_1979\\_\\_13\\_2\\_143\\_0](http://www.numdam.org/item?id=RO_1979__13_2_143_0)

© AFCET, 1979, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## ESTIMATION DES PARAMÈTRES DANS LE MODÈLE LINÉAIRE GÉNÉRAL DYNAMIQUE (\*)

par Simone HUYBERECHTS (1)

Résumé. — *Les résultats classiques du modèle linéaire général sont étendus au cas du modèle linéaire général dynamique introduit récemment par Harrison et Stevens; l'approche, ici considérée, est non bayésienne. On détermine, d'abord, les estimateurs du maximum de vraisemblance des paramètres; on considère ensuite les estimateurs récurrents linéaires, non biaisés et de variance minimale et l'on démontre un théorème de Gauss-Markov généralisé. Dans les deux cas, les estimateurs obtenus sont identiques à ceux issus du filtre de Kalman.*

Abstract. — *The classical results of the general linear model are extended to the dynamic linear model introduced recently by Harrison and Stevens; we consider here a non Bayesian approach. First we find maximum likelihood estimators of the parameters; then we consider minimum variance unbiased recursive linear estimators and we prove an extended Gauss-Markov theorem. In both cases, the estimators obtained are identical with those which are issued from the recursive method called Kalman filter.*

### 1. INTRODUCTION

Le modèle linéaire général dynamique (M.L.G.D.), introduit récemment par Harrison et Stevens [3], présente un intérêt double : d'une part, il généralise le modèle linéaire général classique (ou statique), d'autre part, il s'avère utile dans l'étude des séries chronologiques. Nous traitons, dans le cadre de ce modèle, le problème de l'estimation des paramètres, d'abord dans le cas gaussien, encore appelé cas 1, en faisant appel à la méthode du maximum de vraisemblance; dans le cas 2, lorsque les distributions des termes erreurs sont inconnues, nous présentons un théorème généralisant celui de Gauss-Markov. Dans les deux cas, l'estimateur optimal est identique à celui issu du filtre de Kalman.

### 2. LE MODÈLE LINÉAIRE GÉNÉRAL DYNAMIQUE

1.1. DÉFINITION : Le M.L.G.D. peut être défini par le système d'équations suivant :

$$\text{équation liée aux observations : } y_t = F_t \theta_t + v_t;$$

$$\text{équation liée au système : } \theta_t = G \theta_{t-1} + w_t;$$

---

(\*) Reçu février 1978.

(1) Université Libre de Bruxelles, Belgique.

les différents symboles introduits sont définis comme suit :

$t$  est l'indice du temps (ici  $t=1, 2, \dots$ );  
 $y_t$  est le vecteur ( $m \times 1$ ) du processus des observations faites au temps  $t$ ;  
 $\theta_t$  est le vecteur ( $p \times 1$ ) du processus des paramètres au temps  $t$ ;  
 $F_t$  est la matrice ( $m \times p$ ), supposée connue, des variables indépendantes au temps  $t$ ;

$G$  est la matrice ( $p \times p$ ), supposée connue, du système;  
 $v_t, w_t$  sont des vecteurs aléatoires ( $m \times 1$ ) et ( $p \times 1$ ) respectivement de moyennes nulles et de variances connues, au temps  $t$ , données par

$$V_t = \mathcal{E}(v_t v_t^T), \quad W_t = \mathcal{E}(w_t w_t^T);$$

en outre  $\mathcal{E}(v_t v_{t'}^T) = \mathcal{E}(w_t w_{t'}^T) = 0$ , pour tout  $t \neq t'$ , et  $\mathcal{E}(v_t w_{t'}^T) = 0$  pour tout  $t, t'$ .  
 Ces hypothèses définissent le cas 2 du M.L.G.D.

Dans le cas 1 (cas gaussien), on suppose  $v_t \sim N(0; V_t)$  et  $w_t \sim N(0; W_t)$ .

**1.2. ESTIMATION :** Nous dirons que  $\hat{\theta}_t$  est un estimateur récurrent de  $\theta_t$ , s'il peut s'écrire sous la forme :

$$\hat{\theta}_t = f_t(\hat{\theta}_{t-1}, y_t),$$

où  $f_t$  est une fonction quelconque; nous dirons que  $\hat{\theta}_t$  est un estimateur récurrent linéaire de  $\theta_t$ , s'il peut s'écrire sous la forme :

$$\hat{\theta}_t = B_t \hat{\theta}_{t-1} + A_t y_t,$$

où  $B_t$  et  $A_t$  sont des matrices de constantes de dimensions respectives ( $p \times p$ ) et ( $p \times m$ ).

Notons  $\tilde{\theta}_t = \hat{\theta}_t - \theta_t$ , l'erreur d'estimation au temps  $t$ . Nous dirons que  $\hat{\theta}_t$  est non biaisé si l'espérance mathématique (conditionnelle à  $y_1, \dots, y_{t-1}$ ) de  $\tilde{\theta}_t$  est nulle.  $\hat{\theta}_t$ , supposé non biaisé, sera appelé le meilleur estimateur de  $\theta_t$  s'il minimise la variance (conditionnelle à  $y_1, \dots, y_{t-1}$ ) de l'erreur d'estimation.

### 3. ESTIMATION DE $\theta_t$ : CAS 1

Posons  $Y = y_1, y_2, \dots, y_{t-1}$ . La fonction de vraisemblance à maximiser, pour tout  $t$ , est la fonction de fréquence combinée

$$L(\theta_t) = f(\theta_t, y_t | Y) = f(\theta_t, e_t | Y) = f(e_t | \theta_t, Y) f(\theta_t | Y),$$

où

$$e_t = y_t - F_t G \hat{\theta}_{t-1}$$

représente l'erreur de prévision (conditionnelle à  $F_t$ ) d'horizon 1 faite à l'instant  $t-1$ .

Posons

$$C_t = \mathcal{E}(\tilde{\theta}_t \tilde{\theta}_t^T).$$

Il résulte des hypothèses faites et des propriétés classiques de la distribution multidimensionnelle normale que

$$(\theta_t, e_t \mid Y) \sim N(G \hat{\theta}_{t-1}, 0; \Sigma)$$

avec

$$\Sigma = \begin{bmatrix} R_t & R_t F_t^T \\ F_t R_t & \hat{Y}_t \end{bmatrix}$$

où

$$R_t = G C_{t-1} G^T + W_t \quad \text{et} \quad \hat{Y}_t = V_t + F_t R_t F_t^T,$$

sont les matrices de variances-covariances (conditionnelles) de  $\theta_t$  et  $y_t$ , respectivement.

Maximiser  $L(\theta_t)$  revient à minimiser l'exposant de l'exponentielle figurant dans l'expression de  $f(\theta_t, y_t \mid Y)$ , c'est-à-dire

$$((\theta_t - G \hat{\theta}_{t-1})^T, e_t^T) \Sigma^{-1} \begin{pmatrix} \theta_t - G \hat{\theta}_{t-1} \\ e_t \end{pmatrix}.$$

En se rappelant que l'inverse d'une matrice partitionnée non singulière

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

telle que  $|B_{11}| \neq 0$  et  $|B_{22}| \neq 0$ , peut s'écrire sous la forme :

$$B^{-1} = \begin{bmatrix} [B_{11} - B_{12} B_{22}^{-1} B_{21}]^{-1} & -B_{11}^{-1} B_{12} [B_{22} - B_{21} B_{11}^{-1} B_{12}]^{-1} \\ -B_{22}^{-1} B_{21} [B_{11} - B_{12} B_{22}^{-1} B_{21}]^{-1} & [B_{22} - B_{21} B_{11}^{-1} B_{12}]^{-1} \end{bmatrix},$$

il vient

$$\Sigma^{-1} = \begin{bmatrix} [R_t - R_t F_t^T \hat{Y}_t^{-1} F_t R_t]^{-1} & -F_t^T [\hat{Y}_t - F_t R_t F_t^T]^{-1} \\ -\hat{Y}_t^{-1} F_t R_t [R_t - R_t F_t^T \hat{Y}_t^{-1} F_t R_t]^{-1} & [\hat{Y}_t - F_t R_t F_t^T]^{-1} \end{bmatrix}.$$

En remplaçant  $\Sigma^{-1}$  par sa valeur et en annulant la dérivée première par rapport à  $\theta_t$  de l'expression à minimiser ci-dessus, on obtient l'équation de vraisemblance suivante :

$$2 [R_t - R_t F_t^T \hat{Y}_t^{-1} F_t R_t]^{-1} (\theta_t - G \hat{\theta}_{t-1}) - F_t^T [\hat{Y}_t - F_t R_t F_t^T]^{-1} e_t - e_t^T \hat{Y}_t^{-1} F_t R_t [R_t - R_t F_t^T \hat{Y}_t^{-1} F_t R_t]^{-1} = 0,$$

qui admet comme unique solution

$$\hat{\theta}_t = G \hat{\theta}_{t-1} + R_t F_t^T \hat{Y}_t^{-1} e_t.$$

La matrice

$$A_t = R_t F_t^T \hat{Y}_t^{-1},$$

est appelée filtre de Kalman.

Les calculs qui précèdent peuvent être résumés dans le théorème suivant :

**THÉORÈME 1** : *Dans le M.L.G.D. cas 1 (ou Gaussien) l'estimateur du maximum de vraisemblance de  $\theta_t$  est donné pour tout  $t$  par*

$$\hat{\theta}_t = G \hat{\theta}_{t-1} + A_t e_t,$$

où  $A_t = R_t F_t^T \hat{Y}_t^{-1}$  est la matrice de Kalman, et  $e_t = y_t - F_t G \hat{\theta}_{t-1}$  est l'erreur de prévision (conditionnelle à  $F_t$ ) d'horizon 1 faite en  $t-1$ .

Cet estimateur est récurrent linéaire et non biaisé.

#### 4. ESTIMATION DE $\theta_t$ : CAS 2

**4.1. THÉORÈME 2 OU THÉORÈME DE GAUSS-MARKOV GÉNÉRALISÉ** : *Dans le M.L.G.D. cas 2, le meilleur estimateur récurrent linéaire non biaisé du paramètre  $\theta_t$  est donné, pour tout  $t$ , par l'estimateur issu du filtre de Kalman.*

*Démonstration* : Soit

$$\hat{\theta}_t = B_t \hat{\theta}_{t-1} + A_t y_t,$$

un estimateur linéaire de  $\theta_t$ .

Imposons à  $\hat{\theta}_t$  d'être non biaisé, pour tout  $t$ ; il vient

$$\begin{aligned} \mathcal{E}(\hat{\theta}_t | Y) &= \mathcal{E}[B_t \hat{\theta}_{t-1} + (A_t F_t - I)\theta_t + A_t v_t | Y] \\ &= \mathcal{E}[B_t \hat{\theta}_{t-1} + (A_t F_t - I)G\theta_{t-1} | Y] = 0. \end{aligned}$$

Par conséquent,

$$B_t = (I - A_t F_t) G$$

et

$$\hat{\theta}_t = (I - A_t F_t) G \hat{\theta}_{t-1} + A_t y_t = G \hat{\theta}_{t-1} + A_t e_t,$$

où

$$e_t = y_t - F_t G \hat{\theta}_{t-1},$$

représente, comme ci-dessus, l'erreur de prévision (conditionnelle à  $F_t$ ) d'horizon 1 faite à l'instant  $t-1$ .

Imposons, à présent, à  $A_t$  de minimiser la variance conditionnelle de l'erreur d'estimation. Il vient, à partir des équations de définition,

$$\begin{aligned}\bar{\theta}_t &= (I - A_t F_t) G \hat{\theta}_{t-1} + (A_t F_t - I) (G \theta_{t-1} + w_t) + A_t v_t \\ &= (I - A_t F_t) G \bar{\theta}_{t-1} - (I - A_t F_t) w_t + A_t v_t.\end{aligned}$$

Posons

$$C_t = \mathcal{E} [\bar{\theta}_t \bar{\theta}_t^T] = (I - A_t F_t) [G C_{t-1} G^T + W^T] (I - A_t F_t)^T + A_t V_t A_t^T;$$

les autres termes s'annulant si l'on suppose

$$\mathcal{E} [w_t v_{t'}^T] = 0 \quad \text{pour tout } t, t'.$$

Minimiser la variance ( $\bar{\theta}_t | Y$ ) revient à minimiser la trace de la matrice de variances-covariances  $C_t$ .

La condition nécessaire s'écrit :

$$\frac{\partial}{\partial A_t} (\text{tr } C_t) = 0.$$

En appliquant les règles classiques de dérivation matricielle [1], on obtient :

$$-(I - A_t F_t) [G C_{t-1} G^T + W_t] F_t^T + A_t V_t = 0;$$

en résolvant par rapport à  $A_t$ , il vient

$$A_t = R_t F_t^T \hat{Y}_t^{-1},$$

où

$$R_t = G C_{t-1} G^T + W_t \quad \text{et} \quad \hat{Y}_t = V_t + F_t R_t F_t^T,$$

sont, comme précédemment, les matrices de variances-covariances (conditionnelles) de  $\theta_t$  et  $y_t$ , respectivement.

On vérifie que l'expression de  $A_t$ , appelée matrice de Kalman, obtenue ci-dessus, réalise bien le minimum de la trace de  $C_t$ . [On considère, par exemple, un autre estimateur linéaire  $\bar{\theta}_t$  de  $\theta_t$ , de la forme

$$\bar{\theta}_t = (B_t + B'_t) \bar{\theta}_{t-1} + (A_t + A'_t) y_t,$$

où  $B'_t$  et  $A'_t$  sont également des matrices de constantes de dimensions respectives  $(p \times p)$  et  $(p \times m)$ .

On impose à  $\bar{\theta}_t$  d'être non biaisé et de minimiser la variance de l'erreur d'estimation; on obtient :  $B'_t = A'_t = 0$ ].

**4.2. Remarques :** (i) La méthode récurrente d'estimation qui vient d'être développée, et qui porte le nom de filtre de Kalman, suppose que l'on dispose d'une valeur initiale  $\theta_0$  de  $\theta$ .

(ii) Diverses manipulations algébriques permettent d'exprimer simplement  $C_t$  en fonction de  $C_{t-1}$  (il suffit de remplacer  $A_t$  par son expression); on a notamment,

$$C_t = (I - A_t F_t) R_t$$

et

$$C_t = R_t - A_t \hat{Y}_t A_t^T.$$

(iii) La démonstration du théorème 2 s'applique *mutatis mutandis* dans le cas du modèle linéaire général classique (avec  $\theta_t = \theta_{t-1} = \theta$  et  $V_t = V$ ).

**THÉORÈME 2' OU THÉORÈME DE GAUSS-MARKOV :** *Dans le modèle linéaire général,  $y = F\theta + v$ , où  $v \sim (0; V = \sigma^2 I)$ , le meilleur estimateur linéaire, non biaisé de  $\theta$ , est l'estimateur des moindres carrés.*

*Démonstration :* Soit

$$\hat{\theta} = A y = AF\theta + A v,$$

un estimateur linéaire de  $\theta$ .

Imposons à  $\hat{\theta}$  d'être non biaisé; il vient

$$\mathcal{E}(\hat{\theta}) = AF\theta = \theta.$$

Par conséquent

$$A = F^- + B,$$

où  $F^- = (F^T F)^{-1} F^T$  est l'inverse généralisée de  $F$  et  $B$  est une matrice telle que

$$BF = 0.$$

Tenant compte de cette condition, imposons à  $A$  de minimiser la variance de l'erreur d'estimation.

Posons

$$\begin{aligned} C &= \mathcal{E}(\hat{\theta} \hat{\theta}^T) = \mathcal{E} [((F^- + B)y - \theta)((F^- + B)y - \theta)^T] \\ &= \mathcal{E} [(F^- + B) v v^T (F^- + B)^T] \\ &= \sigma^2 [F^- (F^-)^T + B B^T] = \sigma^2 [(F^T F)^{-1} + B B^T]. \end{aligned}$$

Minimiser la variance de  $\hat{\theta}$  revient à minimiser  $\text{tr } C$ . La condition nécessaire s'écrit :

$$\frac{\partial}{\partial B} (\text{tr } C) = 0,$$

c'est-à-dire

$$B=0;$$

cette condition est compatible avec  $BF=0$ .

Il vient

$$\hat{\theta} = Ay = F^{-1}y = (F^T F)^{-1} F^T y$$

qui est l'estimateur classique des moindres carrés.

(iv) Si l'on adopte, au lieu des notations de Harrison et Stevens les notations classiques du modèle linéaire général, on a le système

$$y_t = X_t \beta_t + v_t,$$

$$\beta_t = G \beta_{t-1} + w_t.$$

Les estimateurs précédents s'écrivent alors sous la forme :

$$\hat{\beta}_t = G \hat{\beta}_{t-1} + A_t e_t,$$

avec

$$e_t = y_t - X_t G \hat{\beta}_{t-1}$$

et

$$A_t = (GC_{t-1} G^T + W_t) X_t^T [V_t + X_t (GC_{t-1} G^T + W_t) X_t^T]^{-1}.$$

Dans le cas du modèle statique, on retrouve

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

## 5. DISTRIBUTION DE $\tilde{\theta}_t$

Il découle immédiatement des développements ci-dessus que

$$\tilde{\theta}_t \sim (0, C_t),$$

avec

$$C_t = (I - A_t F_t) R_t,$$

et

$$G \hat{\theta}_{t-1} - \theta_t \sim (0, R_t).$$

## BIBLIOGRAPHIE

1. P. S. DWYER et M. S. MACPHAIL, *Symbolic Matrix Derivatives*, Ann. Math. Stat., vol. 19, 1948, p. 517-534.
2. F. A. GRAYBILL, *An Introduction to Linear Statistical Models*, vol. I, New York, McGraw-Hill Book Co., Inc., 1961.
3. P. J. HARRISON et C. F. STEVENS, *Bayesian Forecasting*, J. Roy. Stat. Soc. B, vol. 38, 1976, p. 205-228.