L'Actualité économique

Modèle de régression avec variables d'écart : un commentaire

Marcel G. Dagenais

Volume 50, Number 3, juillet–septembre 1974

Montréal : problèmes de croissance et éléments d'une stratégie de développement

URI: https://id.erudit.org/iderudit/803060ar
DOI: https://doi.org/10.7202/803060ar

See table of contents

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)
1710-3991 (digital)

Explore this journal

Cite this note

Dagenais, M. G. (1974). Modèle de régression avec variables d'écart : un commentaire. *L'Actualité économique*, *50*(3), 465–467. https://doi.org/10.7202/803060ar

Modèle de régression avec variables d'écart : un commentaire

Le but de cette note est de faire quelques remarques sur un article paru dans le numéro précédent de cette *Revue* [1].

L'article en question semble contenir plusieurs points obscurs. Mon but n'est pas, ici, d'en faire procès à l'auteur car chacun sait combien il est difficile, lorsque l'on fait un travail de recherche original, d'en exprimer les résultats de façon claire et convaincante. Je déplore cependant qu'un certain nombre de points n'aient pas été éclaircis.

J'énumère ci-dessous quelques-uns de ces points :

1) Les équations (2.3) et (2.4) devraient être remplacées par :

$$E(y_i \mid X_i) = X_i\beta = \text{Prob}\ (y_i = 1 \mid X_i) \qquad i \in n \qquad (1)$$

En effet, il n'est pas exact d'écrire :

$$E(y_i \mid X_i) = \text{Prob}\ (y_i = 0 \mid X_i) \qquad i \in n_1 \qquad (2)$$

comme on l'a fait dans l'équation (2.3).

2) Dans l'équation (2.18), il vaudrait mieux écrire, par exemple :

$$L(y) = \prod_{i \in n_1} \text{Prob}\ (y_i = 0 \mid X_i) \prod_{i \in n_2} \text{Prob}\ (y^* < y_i < y^* + dy^* \text{ et } y^* > 0 \mid X_i) \qquad (3)$$

au lieu de :

$$L(y) = \prod_{i \in n_1} \text{Prob}\ (y_i = 0 \mid X_i) \prod_{i \in n_2} \text{Prob}\ (y_i > 0 \mid X_i)$$

puisque $\prod_{i \in n_2}$ Prob $(y_i > 0 \mid X_i)$ serait un produit d'intégrales et ne correspondrait sûrement pas à l'expression de la dernière ligne de la page 180.

---

1. Jean-Guy Loranger, « Modèle de régression avec variables d'écart », n° 2, avril-juin 1974, pp. 177-190.

3) A la page 181, à la note 4, l'auteur écrit : « Dagenais se limite à différencier β et β^* par la différence dans le terme constant. Mais il n'y a aucune raison qui puisse justifier à priori cette limitation. On peut très bien concevoir une pente pour la droite de régression pour $i \in n_1$ et une autre pour $i \in n_2$ puisque, de toute manière, la pente est 0 pour $i \in n_3$ ». Effectivement, il y a une raison qui justifie cette limitation. En effet, si on utilisait la suggestion de l'auteur telle que représentée au graphique 1 de la page 182, on pourrait aisément se trouver dans une situation où, pour des valeurs de X fortement négatives, l'on aurait *à la fois* $y_{1i} - u_i < L - v_i$ *et* $y_{2i} - u_i > L + s_i$. C'est-à-dire que l'observation se trouverait *à la fois* au-dessous et au-dessus de la valeur limite L [2].

4) A la page 184, on affirme que le vecteur X_i est distribué indépendamment de (u_i, η_i, s_i, W_i). Est-il raisonnable de penser que le vecteur aléatoire X_i et la variable W_i sont distribués indépendamment, alors que $W_i = X_i\beta - u_i$, selon les équations (3.2) et (3.3) ? De la même façon, on affirme que le vecteur Z_i est indépendant de (u_i, η_i, X_i, W_i). Est-ce vraisemblable, si l'on considère que W_i est égal à L ou à $Y_i - u_i$ selon que $W_i \leq L + Z_i'\theta + \eta_i$ ou $W_i > L + Z_i\theta + \eta_i$, d'après (3.1), (3.2), (3.4) et (3.5) ?

5) A la page 185, équation (3.6), l'auteur écrit :

$$W_i = L + s_i - \varepsilon_{1i} \qquad i \in n_1 \tag{4}$$

Si l'on note qu'à la page précédente, équation (3.1), l'auteur mentionne que $W_i = L \;\; \forall \;\; i \in n_1$, ceci implique $\varepsilon_{1i} = s_i$, donc $E(\varepsilon_{1i}) = E(s_i) = E(Z_i'\theta) + E(\eta_i)$. Comme, d'après l'auteur, $E(\eta_i) = 0$ et $E(\varepsilon_{1i}) = \mu$, on a nécessairement $E(Z_i'\theta) = \mu$. Puisque l'on écrit en (3.18) que $E(Z_i\theta) = \mu + \rho$, on doit en conclure que $\rho = 0$. Si, à priori, on a : $\rho = 0$, il est inutile, plus loin, d'essayer d'estimer ce paramètre.

6) A la page 188, l'équation (3.28) devrait se lire :

$$E(\omega_i) = -\rho - E\{(1 - D_i)\varepsilon_{1i}\} + E(D_i\varepsilon_{2i}) \;. \tag{5}$$

Comme D_i dépend de X_i et que X_i est une variable aléatoire, il est inexact de supposer que $E\{(1 - D_i)\varepsilon_{1i}\} = (1 - D_i)\mu$ et $E\{D_i\varepsilon_{2i}\} = D_i\nu$, comme l'écrit l'auteur. Tout au plus, en supposant, comme le fait implicitement l'auteur, que D_i est indépendant de ε_{1i} et ε_{2i}, on peut écrire :

$$E(\omega_i) = -\rho - [1 - E(D)]\mu + E(D_i)\nu. \tag{6}$$

2. Le modèle dont il est question ici est présenté dans Dagenais (1975).

Puisque $\rho = 0$ comme nous l'avons indiqué ci-dessus, on a donc

$$E(\omega_i) = -[1 - E(D_i)]\mu + E(D_i)\nu. \tag{7}$$

Si on se rapporte maintenant à la définition de D_i telle qu'expliquée aux pages 186-187 de l'article, on voit que D_i est une certaine fonction du vecteur aléatoire X_i. Si l'on désigne cette fonction par Φ, on peut écrire : $E(D_i) = E\{\Phi(X_i)\}$. Comme les X_i sont aléatoires par hypothèse, $E\{\Phi(X_i)\}$ sera le même pour tout i. On peut donc écrire : $E(D_i) = E\{\Phi(X_i)\} = \gamma$, où γ est un terme constant. L'espérance mathématique de ω_i peut donc s'exprimer comme :

$$E(\omega_i) = -(1 - \gamma)\mu + \gamma\nu. \tag{8}$$

Le même genre de raisonnement nous force à conclure que Var (ω_i) est égale à une constante Var (ω), pour tout i. La fonction de vraisemblance $L(\omega)$ exprimée en (3.31) devient alors :

$$L(\omega) = (2\pi)^{-n/2} [\text{Var } (\omega)]^{-n/2}$$

$$\exp \left\{ -\tfrac{1}{2} \sum_{i=1}^{n} \left[\frac{[w_i - Z_i' \theta + (1 - \gamma)\mu - \gamma\nu]^2}{\text{Var } (\omega)} \right] \right\} \tag{9}$$

On s'aperçoit alors que cette fonction de vraisemblance a la même structure que celle que l'on obtiendrait si l'on appliquait un simple modèle de régression multiple avec w_i comme variable dépendante et les éléments de Z_i comme variables indépendantes, sans *aucunement* tenir compte du problème qui est à l'origine du modèle suggéré !

Tel que présenté dans l'article qui fait l'objet de ce commentaire, il semble donc, après examen, que le modèle suggéré ne soit pas très utile.

Marcel G. DAGENAIS,
Université de Montréal

RÉFÉRENCES

DAGENAIS, MARCEL G., « Application of a Threshold Regression Model to Household Purchases of Automobiles », *Review of Economics and Statistics,* 1975, à paraître.

LORANGER, JEAN-GUY, « Modèle de régression avec variables d'écart », *L'Actualité Economique,* avril-juin 1974, pp. 177-190.