

La prévision par la Régression Dynamique

Le cours suivant sur **La prévision par la Régression Dynamique** qui suit est issu du support de la formation « **La prévision des Ventes** » que propose PREDICONSULT.

Il utilise des ressources utilisées dans le logiciel de prévision « **Forecast Pro** » que commercialise PREDICONSULT.

La régression dynamique de Forecast Pro permet le développement de modèles qui combinent la dynamique des modèles type séries chronologiques et qui permettent de prendre en compte des variables explicatives ou des indicateurs avancés. Le modèle classique de régression est amélioré par la possibilité de prise en compte d'un terme d'erreur autorégressif (type Cochrane-Orcutt), et par l'utilisation de variables explicatives décalées. Forecast Pro ne permet pas le développement d'équations simultanées.

La régression dynamique peut être utilisée quand (1) les données sont suffisamment longues et stables pour supporter un modèle de corrélation (2) l'introduction de variables explicatives se traduit par une amélioration significative de la précision du modèle et (3) des prévisions valides sont disponibles pour les variables explicatives. Rappelez vous que les modèles complexes produisent souvent des prévisions qui sont moins bonnes que des modèles plus simples, même s'ils ajustent mieux les données observées.

Description du modèle de régression dynamique

Le modèle usuel de régression dynamique par la méthode des moindres carrés s'écrit sous la forme :

$$P(B)Y_t = \beta X_t + e_t \quad (1)$$

dans lequel les erreurs e_t sont distribuées de façon normale et indépendante. Les notations utilisées sont les suivantes :

N	Nombre de points de l'historique
m	Horizon de prévision
s	Nombre de périodes par an
Y_t	Valeur observée au temps t
X_{it}	Valeur observée de la i ème variable explicative au temps t
B	Opérateur de différenciation
ϕ_i	Coefficient d'autorégression de Y_{t-i}
ρ_i	Coefficient d'autorégression de e_{t-i}
β_i	Coefficient de X_i
$\hat{Y}_t(m)$	Prévision pour le temps t+m faite au temps t
e_t	Erreur de prévision $Y_t - Y_{t-1}$

Les valeurs décalées de la variable dépendante sont contenues dans le polynôme P(B), de la même façon que dans le modèle de Box-Jenkins. Le modèle de régression dynamique diffère de Box-Jenkins sur deux points importants :

Il inclut une ou plusieurs variables indépendantes qui détermine la variable dépendante. Par exemple, la publicité et les promotions déterminent les ventes.

L'équation (1) ne contient pas de termes moyennes mobiles qui sont souvent très utiles dans Box-Jenkins. Les modèles de régression seront, par conséquent, parfois moins précis que Box-Jenkins dans certains cas particuliers.

La régression dynamique sera, selon les cas, meilleure ou non, par rapport à Box-Jenkins.

On trouvera assez souvent que les résidus obtenus avec l'équation (1) sont autocorrélés, ce qui est en contradiction avec l'hypothèse de départ. On peut s'en rendre compte par examen des diagnostics dans le module de dynamique. Ceci est susceptible d'indiquer que des termes complémentaires doivent être introduits : de nouvelles variables explicatives et/ou des valeurs décalées des variables déjà introduites.

Le modèle généralisé de Cochrane-Orcutt est une alternative pour améliorer le modèle qui demande, en général, peu de nouveaux termes. Dans le modèle de Cochrane-Orcutt, l'équation (1) est remplacée par la paire d'équation suivantes :

$$P(B)Y_t = \beta X_t + v_t \quad (2)$$

$$R(B)v_t = e_t \quad (3)$$

dans lesquelles les résidus bruts sont corrélés par un processus autorégressif décrit par le polynôme $R(B)$. Les équations (2) et (3) peuvent s'écrire sous la forme unique:

$$R(B)P(B)Y_t = R(B)\beta X_t + v_t \quad (4)$$

Les diagnostics dans la régression dynamique

Un modèle de régression est plus difficile à ajuster qu'un modèle de Box-Jenkins pour différentes raisons. Tout d'abord la partie dynamique du modèle (variable dépendante décalée et termes de Cochrane-Orcutt) doit être déterminée terme par terme sur la base de tests et non automatiquement. Ensuite il n'y a pas de termes autorégressifs dans le modèle de régression dynamique; s'ils sont nécessaires, ils doivent être approchés par une complexité complémentaire du modèle. Enfin, la partie explicative apporte une couche complémentaire de complexité par rapport au modèle endogène. Il faut également considérer le décalage des variables explicatives. Ce qui fait qu'il est possible que des centaines de termes aient à être considérés dans ce type de modèle.

Cette situation complexe nécessite une stratégie ordonnée et systématique. Les diagnostics de Forecast Pro se présentent dans trois batteries de tests pouvant être utilisés à deux phases du processus de développement du modèle. Ces phases sont :

développement du modèle dynamique

développement du modèle explicatif

La batterie de tests de la régression dynamique fournit les tests spécifiques pour le modèle courant. La plupart des diagnostics sont du type χ^2 basés sur des tests sur les multiplicateurs de Lagrange (Engle [1984]). Les tests sur les multiplicateurs de Lagrange sont asymptotiquement équivalents aux tests plus fréquemment utilisés de Wald ou aux tests de vraisemblance. Les paragraphes suivants décrivent ces tests en détail.

Chaque test se rapporte à un défaut précis du modèle. Néanmoins, ils ne sont pas indépendants les uns des autres. Une déficience sur un point précis peut entraîner la significativité de plusieurs tests. A cause de ceci, il est préférable de chercher le test qui est rejeté avec la plus haute significativité et de faire ensuite la modification correspondante. Puis on réexamine les diagnostics sur le nouveau modèle.

Spécification de la dynamique

Le premier groupe de tests est relatif à l'introduction de termes d'erreurs type Cochrane-Orcutt dans le modèle.

_AUTO[-n]. L'hypothèse alternative est qu'un terme d'autocorrélation de l'erreur d'ordre n doit être ajouté au modèle. Forecast Pro effectue un test pour chacun des douze premiers termes et les deux premiers termes de saisonnalité. Le test n'est pas effectué si le terme figure déjà dans le modèle.

Les autres tests de dynamique sont relatifs à l'inclusion de termes décalés de la variable dépendante.

test Y[-n]. L'hypothèse alternative est que le n ième terme décalé de la variable dépendante doit être ajouté au modèle. Forecast Pro effectue un test pour chacun des douze premiers termes et les deux premiers termes de saisonnalité. Le test n'est pas effectué si le terme figure déjà dans le modèle.

Le programme recommande l'addition de nouveaux termes sauf si tous les tests ne sont pas significatifs au niveau 0.01.

Spécification des variables explicatives

Ces tests sont relatifs à la significativité des variables explicatives.

Variables exclues. Un test multiplicateur de Lagrange est calculé pour chaque variable inactive du tableau de données.

Tendance. L'hypothèse alternative est qu'un terme de tendance linéaire améliore le modèle. Un test significatif n'indique pas nécessairement qu'un terme de tendance linéaire doit être ajouté. La difficulté réside souvent dans la dynamique du modèle ou la non prise en compte d'une autre variable.

Terme constant. L'hypothèse alternative est qu'un terme constant améliore le modèle.

Variables explicatives décalées. Un test est effectué pour chaque variable introduite dans le modèle. L'hypothèse alternative est que le premier terme décalé améliore le modèle.

Tests spécifiques sur les variables exclues Cette possibilité permet de tester toute combinaison de variables non introduites.

Diagnosics standards

Les diagnostics standards sont calculés chaque fois qu'un modèle est ajusté aux données. Pour la plupart ils sont très usuels comme la moyenne et l'écart type.

Erreur de prévision standard (Standard forecast error) C'est la racine carrée de la somme des carrés des résidus. Elle est utilisée pour calculer les limites de confiance de la prévision. En fait, c'est en général une évaluation optimiste de l'erreur réelle de prévision.

R-2 C'est la fraction de variance expliquée par le modèle.

BIC Le AIC (Akaike Information Criterion) et le BIC (Bayesian Information Criterion) sont les deux critères les plus utilisés. On sélectionne parmi un ensemble de modèles possibles celui qui minimise le AIC ou le BIC.

Chacun de ces tests mesure la qualité de l'ajustement, représentée par l'écart quadratique moyen s^2 et pénalisée par la complexité du modèle, à savoir le nombre de paramètres. Koehler et Murphree [1986] ont montré que, pour des séries issues de la M-compétition, le BIC conduisait à de meilleures performances de prévision sur les échantillons. Pour cette raison Forecast Pro utilise le BIC.

Il existe plusieurs expressions équivalentes du BIC. Dans Forecast Pro c'est l'expression suivante qui est utilisée, dans laquelle T représente la taille de l'échantillon :

$$BIC = sT^{n/2T}$$

Cette expression du BIC s'exprime dans la même échelle que l'écart type de la prévision. On peut l'interpréter comme une estimation de l'erreur de prévision sur l'échantillon.

Le BIC permet de comparer différents modèles d'une même famille ajustés sur les mêmes données. Comme il a la dimension d'un écart type, sa valeur absolue n'a pas d'importance particulière.

Le test de Durbin-Watson

La statistique de Durbin-Watson est utilisée pour tester la corrélation des résidus successifs du modèle (à savoir l'autocorrélation d'ordre 1). Si T est la taille de l'échantillon et e_t est le résidu au point t , alors la statistique a pour expression :

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

Si la statistique d est facile à calculer, son interprétation est plus délicate.

L'hypothèse nulle est que l'autocorrélation d'ordre 1 est nulle. On trouve les limites d_L et d_U pour un échantillon de taille T au risque α dans une table. L'hypothèse nulle est acceptée si $d < d_L$ et rejetée si $d > d_U$. Si $d_L < d < d_U$, le test ne permet pas de conclure. Notre recommandation, avec laquelle un certain nombre ne sont pas d'accord, est de rejeter l'hypothèse nulle uniquement quand le test est significatif.

Un autre problème vient du fait que la statistique d n'est pas parfaitement valide pour des modèles avec des variables retards. Dans ce cas un certain nombre de statisticiens utilisent le test h au lieu de d .

Nous vous recommandons de rester sur le test de Durbin-Watson et sur l'examen visuel de la fonction d'autocorrélation des erreurs.

Le test de Ljung-Box

Le test de Ljung-Box Q qui est utilisé pour tester l'autocorrélation globale des résidus d'un modèle est une amélioration du test de Box-Pierce. Si T est le nombre de points, r_i le i ème coefficient d'autocorrélation et L le nombre de coefficients d'autocorrélation, alors on calcule Q par :

$$Q = T(T+2) \sum_{i=1}^L \frac{r_i^2}{(T-i)}$$

Le test est la somme pondérée des carrés des coefficients d'autocorrélation, de sorte qu'elle ne vaut zéro que lorsque chaque autocorrélation est nulle. Plus les résidus sont autocorrélés et plus forte est la valeur de Q . Les poids sont choisis de telle façon que Q soit approximativement un χ^2 à $L-n$ degré de liberté.

Les transformations de Box-Cox

On fait l'hypothèse dans le modèle de Box-Jenkins et dans la régression dynamique que les résidus sont indépendants et normalement distribués. L'hétéroscédasticité du processus, à savoir une variance non constante, contredit cette hypothèse. L'effet de l'hétéroscédasticité est moins importante que l'autocorrélation des résidus car elle ne biaise pas l'estimation des coefficients. Son principale effet est de réduire l'efficacité statistique de sorte que l'échantillonnage devient important.

On peut parfois éliminer l'hétéroscédasticité ou la réduire en transformant la variable à étudier. On effectue la prévision sur la variable transformée puis on revient ensuite à la variable d'origine. La transformation la plus importante est celle de analysée par Box et Cox [1964]. Les équations suivantes sont utilisées pour transformer les données d'origine Y_t en $Y_t(\lambda)$:

$$Y_t(\lambda) = \frac{Y_t^\lambda - 1}{\lambda} (\lambda \neq 0)$$

$$Y_t(0) = \ln(Y_t)$$

Le paramètre λ indique la puissance à laquelle les données sont élevées, sauf pour zéro. Dans ce cas on prend le log des données.

Les transformations de Box-Cox ne s'appliquent qu'aux données positives.