CONSTRUCTION DU « MODELE LINEAIRE GENERAL » AVEC PLUSIEURS VARIABLES EXOGENES ET TEMPORELLES DANS LE MARCHE IMMOBILIER A MADAGASCAR

Henipanala Mampionona, Rambinintsoa Tahina, Rafidimanantsoa Tojoandry Tiana

Université d'Antananarivo – Ecole Supérieure Polytechnique d'Antananarivo- Mention Urbanisme Architecture et Génie Civil, hmampi@yahoo.fr

Résumé

Le « Modèle Linéaire Général » est pratique pour simuler le marché immobilier à Madagascar. Un modèle est linéaire si ses paramètres et ses variables sont linéaires, ainsi sa forme fonctionnelle sera linéaire. Pour une variable, elle sera linéaire si la sensibilité d'une autre variable à la suite de sa variation est constante, c'est-à-dire indépendante de la variable à la base du changement ou d'une autre variable quelconque dans le modèle.

La linéarité est l'une des hypothèses fondamentales de la méthode des Moindres Carrés Ordinaires (MCO) au point que sa violation rend inefficace l'utilisation de cette méthode d'estimation (erreurs de spécification, procédures usuelles de test non valides).

Mots-clés : modèle linéaire, Aménagement, immobilier, variables exogènes.

Abstract

The "General Linear Model" is useful for simulating the real estate market in Madagascar. A model is linear if its parameters and variables are linear, so its functional form will be linear. For a variable, it will be linear if the sensitivity of another variable following its variation is constant, i.e. independent of the variable at the basis of the change or of any other variable in the model.

Linearity is one of the fundamental assumptions of the Ordinary Least Squares (OLS) method, and its violation renders the use of this estimation method ineffective (specification errors, invalid standard test procedures).

Keywords: linear model, development, real estate, exogenous variables.

1- INTRODUCTION

La valeur vénale est déterminée : soit par une approche par comparaison directe, soit par capitalisation d'un revenu théorique. La méthode par le coût de remplacement n'est utilisée que de façon marginale, à titre consultatif ou par défaillance des autres méthodes. La méthode théorique de l'évaluation immobilière necessite la construction du modèle linéaire générale. Et notons qu'une estimation est d'autant plus précise que la base d'informations initiales est importante : le relevé et le contrôle des données jouent donc également un rôle essentiel.

2-METHODES

21- Relation fonctionnelle mathématique

Le modèle mathématique le plus utilisé est celui de la régression linéaire qui se présente sous la forme :

$$\mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 x_{1t} + \mathbf{a}_2 x_{2t} + ... + \mathbf{a}_k x_{kt} + \varepsilon_t \text{ pour } t = 1, \ldots n$$

Avec :

$\mathbf{y}_t$: la valeur du bien à la date t ou variable expliquée à la date t.

x_{it}= variable explicative i (i = 1, ..., k) à la date t

$\mathbf{a}_i$ = paramètres estimés du modèle pour la variable i (i = 1, ..., k). Donc, a_0 est une constante.

ε_t = erreur de spécification (différence entre le modèle vrai et le modèle spécifié) ou résidu. Elle est et restera inconnue.

n = nombre d'observations.

La technique de la régression linéaire permet d'obtenir des valeurs de a_0 et des a_i qui réduisent les écarts ε_t au minimum. La constante a_0 est la valeur dépendante lorsque toutes les variables explicatives ont une valeur de zéro. Les coefficients a_i mesurent les variations du prix y_t qui correspondent à une variation de chaque $\boldsymbol{x}_{it}$ à la date t. Les coefficients représentent donc la contribution marginale à la variable dépendante de chaque variable explicative. Le terme d'erreur ε_t est la portion résiduelle du prix y_t d'une observation qui n'est pas représentée ni expliquée par le modèle de régression calculé.
Dans les modèles hédonistes, on connaît les différentes valeurs de y_t correspondant à un échantillon ainsi que les variables explicatives $\boldsymbol{x}_{it}$ (surface, présence d'un garage, etc.) et l'on cherche à déterminer le poids de chaque caractéristique a_i. Celui-ci indique l'augmentation ou la diminution de la valeur marchande expliquée par un changement de la valeur d'une caractéristique particulière, en supposant que toutes les autres variables sont maintenues constantes. Par exemple, l'augmentation de la surface d'un appartement amènera une augmentation de sa valeur indépendamment de sa situation ou de son confort proportionnellement au coefficient a_i déterminé par la régression. Autre exemple, en introduisant la date de l'observation dans les variables explicatives on pourra construire un indice de l'évolution des prix pour les immeubles d'un ou plusieurs quartiers ou en comparant les prix des maisons et des appartements.

22- Relation matricielle

L'écriture précédente du modèle est d'un maniement peu pratique. Afin d'en alléger l'écriture et de faciliter l'expression de certains résultats, on a habituellement recours aux notations matricielles. En écrivant le modèle, observation par observation, nous obtenons :

$Y_1 = a_0 + a_1 x_{11} + a_2 x_{21} + ... + a_k x_{kl} + \varepsilon_1$

$Y_2 = a_0 + a_1 x_{12} + a_2 x_{22} + ... + a_k x_{k2} + \varepsilon_2$

....

$Y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + ... + a_k x_{kt} + \varepsilon_t$

...

$Y_n = a_0 + a_1 x_{1n} + a_2 x_{2n} + ... + a_k x_{kn} + \varepsilon_n$

Soit, sous forme matricielle :

$$\underset{(n,1)}{Y} = \underset{(n,k+1)}{X} \; \underset{(k+1,1)}{a} + \underset{(n,1)}{\varepsilon}$$

Avec

$$Y = \begin{pmatrix} y_1 \\ y_1 \\ \vdots \\ y_1 \\ \vdots \\ y_n \end{pmatrix} : X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{kt} \\ 1 & x_{11} & \cdots & x_{k2} \\ \vdots & \vdots & \cdots & \\ 1 & x_{1t} & x_{2r} & x_{kt} \\ \vdots & \vdots & \vdots & \\ 1 & y_{1n} & y_{2n} & k_n \end{pmatrix} : a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} : = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Nous remarquons la première colonne de la matrice X, composée de 1, qui correspond au coefficient a_0 (coefficient du terme constant).
La dimension de la matrice X est donc de n lignes et k + I colonnes (k étant le nombre de variables explicatives réelles, c'est-à-dire constante exclue).
L'écriture sous forme matricielle rend plus aisée la manipulation du modèle linéaire général.

23- Les coefficients de régression

231- Estimation des coefficients de régression

Soit le modèle sous forme matricielle à k variables explicatives et n observations :

$$\mathbf{Y = Xa + \varepsilon}$$

Afin d'estimer le vecteur **a** composé des coefficients a0, a1 …ak, nous appliquons la méthode des Moindres Canés Ordinaires (MCO) qui consiste à minimiser la somme des carrés des erreurs, soit :

$$\mathbf{Min}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t^2 = \mathbf{Min}\ \boldsymbol{\varepsilon}'\ \boldsymbol{\varepsilon} = \mathbf{Min}\ \mathbf{(y - Xa)'\ (Y - Xa) = Min\ S}$$

avec ε' transposé du vecteur ε.

Pour minimiser cette fonction par rapport à a, nous différencions S par rapport à a :

$$\frac{\partial S}{\partial a} = -\,2X'Y + 2\,X'X\hat{a} = 0 \quad \Rightarrow \quad \hat{a} = \boxed{(X'\,X)^{-1}X'Y}$$

Cette solution est réalisable si la matrice carrée X'X de dimension (k + I, k + I) est inversible. La matrice X'X est la matrice des produits croisés des variables explicatives ; en cas de colinéarité parfaite entre deux variables explicatives, la matrice X'X est singulière et la méthode des MCO défaillante.

On appelle équations normales les équations issues de la relation : $X'X\hat{a} = X'Y$

Soit, sous forme matricielle :

$$\begin{pmatrix} n & \sum x_{1t} & \sum x_{2t} & \dots & \sum x_{kt} \\ \sum x_{1t} & \sum x_{1t}^2 & \sum x_{1t}\,x_{2t} & \dots & \sum x_{1t}\,x_{kt} \\ \sum x_{2t} & \sum x_{2t}\,x_{1t} & \sum x_{2t}^2 & \dots & \sum x_{2t}\,x_{kt} \\ \vdots & \dots & \dots & \dots & \dots \\ \sum x_{kt} & \sum x_{kt}\,x_{1t} & \sum x_{kt}\,x_{2t} & \dots & \sum x_{kt}^2 \end{pmatrix} \begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \widehat{a}_2 \\ \dots \\ \widehat{a}_k \end{pmatrix} = \begin{pmatrix} \sum y_t \\ \sum x_{1t}\,y_t \\ \sum x_{2t}\,y_t \\ \dots \\ \sum x_{kt}\,y_t \end{pmatrix}$$

Le modèle estimé s'écrit : $y_t = \hat{a}_0 + \hat{a}_1 + \hat{a}_2 x_{2t} + \ldots + \hat{a}_2\, x_{kt} + e_1$
avec $e_t = y_t - \hat{a}_t$, où et est le résidu, c'est-à-dire l'écart entre la valeur observée de la variable à expliquer et sa valeur estimée (ajustée).

Attention il convient de bien distinguer entre l'erreur de spécification du modèle (noté ε_t) qui est et restera inconnue et le résidu (e_t) qui, lui, est connu.

a) Cas particulier

Si nous raisonnons sur des données centrées, l'estimateur de a peut s'écrire en fonction des matrices des variances et covariances empiriques :

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \dots \\ \hat{a}_k \end{pmatrix} = \begin{pmatrix} Var(x_1) & Cov\,(x_1, x_2) & \dots.\ Cov\,(x_1, x_k) \\ Cov\,(x_2, x_1) & Var(x_2) & \dots\ Cov\,(x_2, x_k) \\ Cov\,(x_3, x_1) & Cov\,(x_3, x_2) & \dots\ Cov\,(x_3, x_k) \\ \dots & \dots & \dots \\ Cov\,(x_k, x_1) & Cov\,(x_k, x_2) & \dots\ Var(x_k) \end{pmatrix}^{-1} \times \begin{pmatrix} Cov\,(x_1, y) \\ Cov\,(x_2, y) \\ Cov\,(x_3, y) \\ \dots \\ Cov\,(x_k, y) \end{pmatrix}$$

Avec $\hat{a}_0 = \hat{y}_1\ \bar{a}_1\ -\ \hat{y}_2\ \bar{a}_2\ -\ \ldots\ \hat{y}_k\ \bar{a}_k$

Que sont des données centrées sur la moyenne ? soit x_1 une variable connue sur n observation et $\bar{x}$ sa moyenne, nous pouvons calculer une nouvelle variable :

($X_1 = x_1 - x$) dont la somme, nous est par construction nulle $\sum_{t=1}^{n}(x_t - \bar{x}) = \sum_{t=1}^{n} X_t = 0$
Nous avons donc $\bar{X} = 0$.

b) effet de la variation d'une seule des variables explicatives

Soit le modèle estimé $y_1 = \hat{a}_0 + \hat{a}_1\hat{x}_{1t} + \hat{a}_2\hat{x}_{2t} + \ldots + \hat{a}_k\hat{x}_{kt} + e_t$

Si la variable x_1 passe de la valeur x_{2t} à ($x_{2t} + \Delta\, x_{2t}$), toute chose étant égale par ailleurs (les k–1 autres variables restant constantes), alors la variable à expliquer varié de $\hat{a}_2\, \Delta\, x_2$;

$\Delta\, \hat{y}_1 = \hat{a}_2\, \Delta\, x_2$

Les coefficients s'interprètent donc directement en termes de propension marginale.

232- Hypothèses stochastiques et les propriétés des estimateurs

Par construction, le modèle est linéaire en X (ou sur ces coefficients) et nous distinguons les hypothèses stochastique (liées à l'erreur ε) des hypothèses structurelles.

a) Les Hypothèses stochastique

- H1 : les valeurs x_{it} sont observées sans erreur
- H2 : $E(\varepsilon_t) = 0$, l'Esperance mathématique de l'erreur est nulle
- H3 : $E(\varepsilon_t{}^2) = \sigma_t^2$, la variance de l'erreur est constante ($\forall t$) (homoscédasticité)
- H4 : $E(\varepsilon_{t}, \varepsilon_{t'}) = 0$ si $t \neq t'$, les erreurs sont non corrélées (ou encore indépendante)
- H5 : Cov (x_{1t} , ε_t) = 0, l'erreur est indépendante des variables explicatives

b) Les Propriétés des estimateurs

Considérons les propriétés de l'estimateur.
Les modèles sous forme matricielle peuvent s'écrire, comme pour le modèle de régression simple, de différentes manières :

$Y = X_a + \mathrm{F}$

$Y = X\hat{a} + \mathrm{F} \quad \rightarrow e = \mathrm{Y} - \hat{Y} \qquad (e = \text{résidu})$

$\hat{Y} = X\hat{a}$

Nous obtenons :

$\hat{a} = (X'X)^{-1}X'Y = (X'X)^{-1}X(Xa + \varepsilon)$

$\hat{a} = (X'X)^{-1}X'(Xa) + (X'X)^{-1}X'\varepsilon$

$\hat{a} = \mathrm{a} + (X'X)^{-1}X'\varepsilon$

D'où $\mathrm{E}(\hat{a}) = \mathrm{a} + (X'X)^{-1}X'E(\epsilon) = a \quad car\ E(\varepsilon) = 0$

L'estimateur est donc sans biais : $\boxed{\boldsymbol{E(\hat{a}) = a}}$

Calculons maintenant la matrice des variances et covariances des coefficients de régression Ω_k :

$$\Omega_k = \mathrm{E}\,\{(\hat{a} - \mathrm{a})\,(\hat{a} - \mathrm{a})\}$$

Or, d'après [4], $(\hat{a} - \mathrm{a}) = (X'X)^{-1}E(\varepsilon)$ et $(\hat{a} - \mathrm{a}) = \varepsilon\,\mathrm{X}'(X'X)^{-1}$ puisque $(X'X)^{-1}$ est symétrique :

$$(\hat{a} - \mathrm{a})\,(\hat{a} - \mathrm{a}) = (X'X)^{-1}X\,\varepsilon\,\varepsilon'\,X(X'X)^{-1} \qquad d'où$$

Ω_k= (X' X) $^{-1}$X E(ε ε')X(X' X) $^{-1}$ avec E (ε ε') = Ω_k = σ_1^2I est la matrice des variances et covariance de l'erreur ε.

En effet d'après les hypothèses H3 et H4 nous avons :

$$\Omega_k = E(\varepsilon\,\varepsilon') = \begin{pmatrix} E(\varepsilon_1\,\varepsilon_1) & E(\varepsilon_1\,\varepsilon_2) & \ldots\ldots\ E(\varepsilon_1\,\varepsilon_n) \\ E(\varepsilon_2\,\varepsilon_1) & E(\varepsilon_2\,\varepsilon_2) & \ldots\ldots\ E(\varepsilon_2\,\varepsilon_n) \\ E(\varepsilon_n\,\varepsilon_1) & E(\varepsilon_n\,\varepsilon_1) & \ldots\ldots\ E(\varepsilon_n\,\varepsilon_n) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 & 0 & \ldots\ldots 0 \\ 0 & \sigma_1^2 & \ldots\ldots 0 \\ 0 & 0 & \ldots\ldots \sigma_n^2 \end{pmatrix}$$

Soit Ω_k= σ_k^2 *(X' X)* $^{-1}$ *X' X (X' X)* $^{-1}$ ➔ $\boxed{\Omega_k = \boldsymbol{\sigma_1^2}\ \textbf{(X' X)}^{-1}}$

$\Omega_0 = \frac{\sigma_1^2}{n}(\frac{X\,X}{n})$ ⇒ lim Ω = 0 si $n \to \infty$ (d'après l'hypothèse H3) L'estimateur est donc convergent.

Théorème de Gauss-Markov : L'estimateur [3] des moindres carrés est qualifié de BLUE (Best Linear Unbiaised Estimator), car il s'agit du meilleur estimateur linéaire sans biais (au sens qu'il fournit les variances les plus faibles pour les estimateurs).
II est â noter que l'estimateur du maximum de vraisemblable des paramètres fournit des résultats identiques à ceux des l'estimateur des MCO si l'hypothèse de normalité des erreurs est vérifiée.

Apres un calcul matriciel, il apparaît que nous pouvoir estimer sans biais σ_1^2 par :

$$\boxed{\boldsymbol{\hat{\sigma}_k^2} = \frac{\boldsymbol{e'\,e}}{\boldsymbol{n-k-1}}}$$

En remplaçant la variance de l'erreur par son estimateur, nous obtenons :

$$\boxed{\widehat{\Omega_k} = \boldsymbol{\hat{\sigma}_k^2}\ \textbf{(X' X)}^{-1}}$$

233- Calcul de la variance et ajustement

Nous avons, dans toute régression :

a) $\sum_t\ y_t = \sum_t\ \hat{Y}_t \to \hat{Y} = \hat{Y}$
b) $\sum_t\ e_t = 0$

De ces deux relation, nous en déduisions fondamentale d'analyse de la variance :

$$\boxed{\begin{array}{c} \sum_t\ (\boldsymbol{y_t} - \boldsymbol{\hat{y}})^2 = \sum_t\ (\boldsymbol{\hat{y}_t} - \boldsymbol{\hat{y}})^2 + \sum_t\ \boldsymbol{e_t^2} \\ \textbf{SCT} = \textbf{SCE} + \textbf{SCR} \end{array}}$$.

La variabilité totale (SCT) est égale à la variabilité expliquée (SCE) + la variabilité des résidus (SCR). Cette équation va nous permettre de juger de la qualité de l'ajustement d'un modèle en effet, plus la variance expliquée est « proche » de la variance totale, meilleur est l'ajustement global du modèle. C'est pourquoi nous calculons le rapport SCE sur SCT:

$$\mathbf{R^2} = \frac{\sum_t\ (\hat{y}_t - \hat{y})^2}{\sum_t\ (y_t - \hat{y})^2} = \mathbf{1} - \frac{\sum e_1^2}{\sum_t\ (y_t - \hat{y})^2}$$

R^2 est appelé le coefficient de détermination, et R le coefficient de corrélation multiple. R^2 mesure la proportion de la variance de Y expliquée par la régression de Y sur X.

Dans le cas de données centrées (moyenne nulle) et seulement dans ce cas, le coefficient de détermination est égal à :

$$\mathbf{R^2 = \frac{\hat{Y}'\hat{Y}}{Y'Y} = 1 - \frac{e'e}{Y'Y}}$$

24- La Prévision et la régression récursive

241- La Prédiction conditionnelle

Le problème consiste à déterminer quelle valeur doit être attribuée à la variable endogène lorsque nous connaissons les valeurs des variables exogènes.
Le modèle général estimé est le suivant :

$$Y_t = \hat{a}_o + \hat{a}_1 x_{1t} + \hat{a}_2\, x_{2t} + \ldots\ldots + \hat{a}_k\, x_{kt} + e_t$$

La prévision pour la donnée t + h est la suivante :

$$\hat{y}_{t+h} = \hat{a}_o + \hat{a}_1\, x_{1t+h} + \hat{a}_2\, x_{2t+h} + \ldots\ldots + \hat{a}_k\, x_{kt+h}$$

L'erreur de prévision est donnée par : $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$

Considérant que les hypothèses du modèle linéaire général sont vérifiées, la prévision $\hat{y}_{t+h}$ est sans biais.

Nous avons postulé que nous connaissons sans erreur les valeurs en t + h des variables explicatives. Si pour un modèle en coupe instantanée cette hypothèse est réaliste, il n'en est pas de même dans les modèles en séries temporelles pour lesquelles les valeurs futures des variables explicatives sont estimées avec imprécision et donc introduisent un élément supplémentaire d'incertitude dans le calcul de la prévision.
Ce problème est exclu du champ de notre analyse, puisque nous traitons de prédiction conditionnelle.

242- Fiabilité de la prévision et intervalle de prévision

L'erreur de prévision calculée en t à l'horizon h peut s'écrire aussi :

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h} = X'_{t+h}(a - \hat{a}) + \varepsilon_{t+4}$$

Calculons la variance de cette expression: $V(e_{t+h}) = V(X'_{t+h}(a - \hat{a}) + \varepsilon_{t+h})$.
$X_{t+h}(a - \hat{a})$ est une combinaison linéaire des y_t et puisque ε_{t+h} est sans corrélation avec les y_t, nous avons $Cov(X'_{t+h}(a - \hat{a}), \varepsilon_{t+h}) = 0$

On a alors : $V(e_{t+h}) = V(X_{t+h}(a - \hat{a}) + V(\varepsilon_{t+h})$
$V(X_{t+h}(a - \hat{a}) = X_{t+h} V(a - \hat{a})\, X_{t+h} = \sigma^2_\varepsilon\ [X_{t+h}(X'X)^{-1} X_{t+h}]$
(puisque $V(\hat{a}) = \sigma^2_\varepsilon (X'X)^{-1}$)
On connaît : $V(\varepsilon_{t+h}) = \sigma^2_\varepsilon$
La variance de l'erreur de prévision est donc égale à :

$$\boxed{\sigma^2_{et+h} = \sigma^2_\varepsilon [X'_{t+h}(X'X)^{-1} X_t + 1_h + 1]}$$

Avec $X_{t+h} = \begin{pmatrix} 1 \\ x_{1t+h} \\ x_{2t+h} \\ \ldots.. \\ x_{kt+h} \end{pmatrix}$ vecteur des valeurs prévues des variables explicatives.

Or l'erreur de prévision $(e_{t+h} = Y_{t+h} - \hat{y}_{t+h})$ est distribuée suivant une loi normale $N(0, \sigma^2_{et+h})$' en remplaçant la variance théorique c par la variance empirique dans l'expression [5.12], nous pouvons en déduire que:

$$\frac{\hat{y}_{t+h} - y_{t+h}}{\hat{\sigma}_\varepsilon [X'_{t+h} \ (X'X)^{1/2}}$$ suit une loi de student à n – k – 1 degrés de liberté.

Nous remarquons que, comme pour le modèle de régression simple, la variance de l'erreur de prévision est d'autant plus faible lorsque:

- la variance résiduelle est faible;
- les valeurs prévues des variables explicatives se rapprochent de leur moyenne.

L'intervalle au seuil (1 — α) de la prévision est alors :

$$y_{t+h} = \hat{y}_{t+h} \pm t^{a/2}_{n-K-1} \ \sqrt{\hat{\sigma}^2_\varepsilon \ [X'_{t+h}(X'X)^{-1} \ X_{t+h} + 1]}$$

3- RESULTAT : Application du « Modèle Linéaire Général » avec plusieurs valeurs endogènes avec le temps – cas des « villas haut standing » d'Ivandry 2011 à 2020

Nous étudions le prix de villas de haut standing sise à Ivandry, avec piscine ou Spa. Prenons un modèle à trois variables explicatives, qui sont :

- x_{1t} : le nombre de chambres (hors cuisine, salle d'eau et pièces annexes)
- x_{2t} : SHON (Surface Hors Œuvre Nette) en m²
- x_{3t} : la surface du terrain en m² (bâti compris)
- y_t : le prix de la villa (en millier d'euro)

Nous avons alors la relation suivante : $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + e_t$
(avec ε_t résidu ou erreur)

Le tableau 1 suivant nous donne les valeurs observées de y, x_1 , x_2 et x_3.

Tableau 1: Valeurs observées de y, x_1 , x_2 et x_3

Année	y	x_1	x_2	x_3
2011	175	5	150	400
2012	130	3	100	300
2013	160	4	110	350
2014	195	6	180	400
2015	155	4	100	320
2016	210	5	170	450
2017	165	3	115	270
2018	245	7	180	410
2019	205	4	145	360
2020	230	5	140	375

(NB : les valeurs prises sont les prix moyens des villas. Par ailleurs, les prix ont été pris chez le même promotteur immobilier et les villas de haut standing sur Antananarivo sont très peu différents).

31- Mise sous forme matricielle du modèle

La relation matricielle est de: Y = Xa +ε

$$Y = \begin{pmatrix} 175 \\ 130 \\ 160 \\ 195 \\ 155 \\ 210 \\ 165 \\ 245 \\ 205 \\ 230 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 5 & 150 & 400 \\ 1 & 3 & 100 & 310 \\ 1 & 4 & 110 & 350 \\ 1 & 6 & 180 & 400 \\ 1 & 4 & 100 & 320 \\ 1 & 5 & 170 & 450 \\ 1 & 3 & 115 & 270 \\ 1 & 7 & 180 & 410 \\ 1 & 4 & 145 & 360 \\ 1 & 5 & 140 & 375 \end{pmatrix} \qquad a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \end{pmatrix}$$

3-2 Estimation des coefficients de régression

Nous avons: $\boldsymbol{\hat{a}} = \mathbf{(X'\,X)^{-1}\,X'Y}$

Nous commençons par calculer X'X et $(X'X)^{-1}$

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 3 & 4 & 6 & 4 & 5 & 3 & 7 & 4 & 5 \\ 150 & 100 & 110 & 180 & 100 & 170 & 115 & 180 & 145 & 140 \\ 400 & 310 & 350 & 400 & 320 & 450 & 270 & 410 & 360 & 375 \end{pmatrix}$$

En effectuant le produit matriciel, nous obtenons :

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 3 & 4 & 6 & 4 & 5 & 3 & 7 & 4 & 5 \\ 150 & 100 & 110 & 180 & 100 & 170 & 115 & 180 & 145 & 140 \\ 400 & 310 & 350 & 400 & 320 & 450 & 270 & 410 & 360 & 375 \end{pmatrix} . \begin{pmatrix} 1 & 5 & 150 & 400 \\ 1 & 3 & 100 & 310 \\ 1 & 4 & 110 & 350 \\ 1 & 6 & 180 & 400 \\ 1 & 4 & 100 & 320 \\ 1 & 5 & 170 & 450 \\ 1 & 3 & 115 & 270 \\ 1 & 7 & 180 & 410 \\ 1 & 4 & 145 & 360 \\ 1 & 5 & 140 & 375 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 10 & 46 & 1390 & 3635 \\ 46 & 226 & 6705 & 17225 \\ 1390 & 6705 & 202150 & 518550 \\ 3635 & 17225 & 518550 & 1348625 \end{pmatrix}$$. Nous voulons calculer $(X'X)^{-1}$

Le déterminant de la matrice est égal à 2264009375

La comatrice est : $\begin{pmatrix} 13319025625 & 746318750 & 14811125 & -51126375 \\ 746318750 & 675931250 & -17954375 & -3741250 \\ 14811125 & -17954375 & 1391400 & -345600 \\ -51126375 & -3741250 & -345600 & 320150 \end{pmatrix}$.

La transformée de la comatrice est : $\begin{pmatrix} 13319025625 & 746318750 & 14811125 & -51126375 \\ 746318750 & 675931250 & -17954375 & -3741250 \\ 14811125 & -17954375 & 1391400 & -345600 \\ -51126375 & -3741250 & -345600 & 320150 \end{pmatrix}$.

La matrice inverse est égale à :

$$(X'X)^{-1} = \begin{pmatrix} 5{,}8829 & 0{,}3296 & 0{,}006542 & -0{,}0226 \\ 0{,}3296 & 0{,}2986 & -0{,}00793 & -0{,}00165 \\ 0{,}006542 & -0{,}00793 & 0{,}00061 & -0{,}00015 \\ -0{,}0226 & -0{,}00165 & -0{,}00015 & 0{,}00014 \end{pmatrix}.$$

Calculons maintenant X'Y :

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 3 & 4 & 6 & 4 & 5 & 3 & 7 & 4 & 5 \\ 150 & 100 & 110 & 180 & 100 & 170 & 115 & 180 & 145 & 140 \\ 400 & 310 & 350 & 400 & 320 & 450 & 270 & 410 & 360 & 375 \end{pmatrix} . \begin{pmatrix} 175 \\ 130 \\ 160 \\ 195 \\ 155 \\ 210 \\ 165 \\ 245 \\ 205 \\ 230 \end{pmatrix} = \begin{pmatrix} 1870 \\ 8925 \\ 268150 \\ 692150 \end{pmatrix}.$$

Appliquons maintenant la formule (5.3) : $\boldsymbol{\hat{a}} = \mathbf{(X'\,X)^{-1}\,X'Y}$

$$\hat{a} = \begin{pmatrix} 5,8829 & 0,3296 & 0,006542 & -0,0226 \\ 0,3296 & 0,2986 & -0,00793 & -0,00165 \\ 0,006542 & -0,00793 & 0,00061 & -0,00015 \\ -0,0226 & -0,00165 & -0,00015 & 0,00014 \end{pmatrix} \cdot \begin{pmatrix} 1870 \\ 8925 \\ 268150 \\ 692150 \end{pmatrix} = \begin{pmatrix} 67,118 \\ 10,747 \\ 0,597 \\ -0,0344 \end{pmatrix}$$

Donc, $\widehat{a_0}$ = 67, 118 $\widehat{a_1}$ = 10,747 $\widehat{a_2}$ = 0,597 $\widehat{a_3}$ = -0,0344 sont les valeurs estimées des coefficients.
Et la relation s'écrit : $\mathbf{y_t = 67,118 + 10,747\, x_{1t} + 0,597\, x_{2T} - 0,}0344\, x_{3t} + e_t$

3-3 Estimation de la variance de l'erreur et les écart-types de chacune des coefficients estimés

D'après la formule 5.6 : $\boldsymbol{\hat{\sigma}_k^2} = \frac{\boldsymbol{e' e}}{\boldsymbol{n-k-1}}$

Nous rappelons que e_t est composé par (e_1 , e_2 , e_3, e_4 , …, e_{10}) ou (e_{2011} , e_{2012}, …., e_{2020}).
Donc, pour e_1 , nous avons : $e_1 = y_1 - (67,118 + 10,747\, x_{1t} + 0,597\, x_{2T} - 0,0344\, x_{3t})$

$e_{2011} = e_1 = 175 - (67,118 + 10,747 .5 + 0,597 .150 - 0,0344 .400) = -21,643$

En faisant le même calcul, nous avons les valeurs des résidus e_t suivants (Tableau 2) :

Tableau 2 : Valeurs des résidus

Année	y_t	a_0	a_1	x_1	a_2	x_2	a_3	x_3	e_t
2011	175	67,118	10,747	5	0,597	150	-0,0344	400	-21,643
2012	130	67,118	10,747	3	0,597	100	-0,0344	300	-18,739
2013	160	67,118	10,747	4	0,597	110	-0,0344	350	-3,736
2014	195	67,118	10,747	6	0,597	180	-0,0344	400	-30,3
2015	155	67,118	10,747	4	0,597	100	-0,0344	320	-3,798
2016	210	67,118	10,747	5	0,597	170	-0,0344	450	3,137
2017	165	67,118	10,747	3	0,597	115	-0,0344	270	6,274
2018	245	67,118	10,747	7	0,597	180	-0,0344	410	9,297
2019	205	67,118	10,747	4	0,597	145	-0,0344	360	20,713
2020	230	67,118	10,747	5	0,597	140	-0,0344	375	38,467

Après calcul, nous avons $\sum_{t=0}^{10} e_t{}^2$ = 3810,42 et $\sum_{t=0}^{10} e_t \approx 0$.

Ainsi, la formule (5.6) devient : $\hat{\sigma}_k^2 = \frac{e' e}{n-k-1} = \frac{\sum_{t=0}^{10} e_t{}^2}{n-k-1} = \frac{3810,42}{10-3-1}$ = 635,07

La matrice de variance de l'erreur est donnée par la formule (5.7) : $\widehat{\Omega_k} = \hat{\sigma}_k^2\, (X'\, X)^{-1}$

$$\widehat{\boldsymbol{\Omega_k}} = 635,07 \begin{pmatrix} 5,8829 & 0,3296 & 0,006542 & -0,0226 \\ 0,3296 & 0,2986 & -0,00793 & -0,00165 \\ 0,006542 & -0,00793 & 0,00061 & -0,00015 \\ -0,0226 & -0,00165 & -0,00015 & 0,00014 \end{pmatrix}$$

La variance des coefficients de régression estimée est constituée par la diagonale de la matrice de variance $\widehat{\boldsymbol{\Omega_k}}$. Nous savons aussi que l'ecart-type est la racine carrée de la variance. Ce qui donne le résultat suivant :

Tableau 3: Variance et Ecart-type des coefficients de régression estimée

Variance des coefficients de régression estimée	Ecart-type des coefficients de régression estimée
$\hat{\sigma}_{a0}^2$ = 635,07 . 5,8829 = 3736,05	$\hat{\sigma}_{a0}$ = 61,12

$\hat{\sigma}^2_{a1}$ = 635,07 . 0,2986 = 189,63	$\hat{\sigma}_{a1}$= 13,77
$\hat{\sigma}^2_{a2}$ = 635,07 . 0,00061 = 0,387	$\hat{\sigma}_{a2}$ = 0,62
$\hat{\sigma}^2_{a3}$ = 635,07 . 0,00014 = 0,0889	$\hat{\sigma}_{a3}$ = 0,29

3-4 Calcul du coefficient de détermination R^2 et du $\vec{R}^2$ corrigé

Le Tableau 4 concerne les étapes pour le calcul du coefficient de détermination R^2

R^2 est calculé par la formule (5.9) : $R^2 = 1 - \frac{\sum e_t^2}{\sum_t (y_t - \hat{y})^2}$

Nous avons déjà calculé $\sum e_t^2$ = 635,07

Tableau 5: Calcul du coefficient de détermination R^2

Année	y	$\widehat{a_0}$	$\widehat{a_1}$	x_1	$\widehat{a_2}$	x_2	$\widehat{a_3}$	x_3	**y_t - $\hat{y}$**	**$(y_t - \hat{y})^2$**
2011	175	67, 118	10,747	5	0,597	150	-0,0344	400	-21,643	468,419449
2012	130			3		100		300	-18,739	351,150121
2013	160			4		110		350	-3,736	13,957696
2014	195			6		180		400	-30,3	918,09
2015	155			4		100		320	-3,798	14,424804
2016	210			5		170		450	3,137	9,840769
2017	165			3		115		270	6,274	39,363076
2018	245			7		180		410	9,297	86,434209
2019	205			4		145		360	20,713	429,028369
2020	230			5		140		375	38,467	1479,71009

Donc, $\sum_t (y_t - \hat{y})^2$ = 3810,41858

Ainsi, $R^2 = 1 - \frac{635{,}07}{3810{,}41858}$ = 0,83333

Rappelons que plus R^2 est proche de 1, plus, nous avons une meilleure estimation dans le calcul de Régression.

La formule (5.11) nous donne le $\vec{R}^2$ corrigé : $\vec{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2)$ [5.11]

$\vec{R}^2 = 1 - \frac{14-1}{14-3-1}(1 - 0{,}833333) = 0{,}783333$

Il est normal que $R^2 > \vec{R}^2$.

Remarque par rapport à la valeur de $\widehat{a_3}$:

Après calcul, nous constatons que la valeur de $\widehat{a_3}$ est presque nulle. Cela signifie que la surface du terrain ne tient pas une place importante dans l'évaluation du prix des villà haut standing d'Ivandry.

3-5 Evaluation du prix des villas pour les années futures et l'importance de la variance dans la Prévision

Nous allons évaluer le prix d'un villa haut standing à Ivandry pour l'année 2021, en partant des données de 2010 à 2020.

Prenons les données suivantes comme caractéristique :

année	x_1	x_2	x_3
2021	5	150	420

Il faut appliquer le résultat de notre estimation :

y_t = 67,118+10,747 x_{1t} +0,597 x_{2t} - 0,0344 x_{3t}

Après calcul, nous obtenons :

y_{2021} = 67,118 + 10,747 .5 + 0,587.150 – 0,0344.420 = 194,455
La question suivante se pose : « Quel est l'erreur de cette prévision ? ». La réponse à cette question se trouve dans la valeur de la « variance de l'erreur de prévision » (ou écart-type de l'erreur de prévision) qui équivaut à une marge d'erreur de prévision.
La variance de l'erreur de prévision est donnée par la formule (5.12) :

$$\sigma^2_{et+h} = \sigma^2_\varepsilon[X'_{t+h}(X'X)^{-1}X_{t+h}+1] \qquad (5.12)$$

Comme application numérique de cette formule, nous avons pour les données 2021 :
σ^2_ε= 635,07 $\qquad$ X'_{2021} = (1 5 150 420)

$$(X'X)^{-1} = \begin{pmatrix} 5{,}8829 & 0{,}3296 & 0{,}006542 & -0{,}0226 \\ 0{,}3296 & 0{,}2986 & -0{,}00793 & -0{,}00165 \\ 0{,}006542 & -0{,}00793 & 0{,}00061 & -0{,}00015 \\ -0{,}0226 & -0{,}00165 & -0{,}00015 & 0{,}00014 \end{pmatrix}.$$

$$X_{2021} = \begin{pmatrix} 1 \\ 5 \\ 150 \\ 420 \end{pmatrix}$$

Après calcul, σ^2_{2021}= 816,78, Donc σ_{2021} = 28,579
Ainsi, dans notre prévision, la villa est estimée à 194455 euros avec une marge d'erreur de 28579 euros, soit une marge de 14%.

4- DISCUSSIONS

Cette discussion fait le point des avantages et des inconvénients que présente la Méthode de régressions hédoniques pour construire les paramètres des prix de l'immobilier résidentiel. Ses principaux avantages sont les suivants :

- Si la liste disponible des caractéristiques des biens immobiliers est assez détaillée, les méthodes hédoniques peuvent, en principe, permettre d'éliminer l'incidence des changements survenus dans la composition de l'échantillon ainsi que dans la qualité des divers biens immobiliers.
- Des indices de prix peuvent être construits pour différents types de logements et différents emplacements si l'échantillon est stratifié comme il convient.
- La méthode hédonique est probablement celle qui permet d'exploiter au mieux les données disponibles.
- Des indices de prix peuvent être construits pour différents types de logements et d'emplacements par le biais de la stratification et l'application de techniques hédoniques à chaque strate individuelle.

Les principaux inconvénients des régressions hédoniques sont les suivants :

- Il peut s'avérer difficile de neutraliser suffisamment l'effet de l'emplacement si les prix des biens immobiliers et leur évolution varient au sein des régions. L'adoption d'une approche stratifiée des régressions hédoniques permettra toutefois de surmonter en partie ce problème.
- L'approche hédonique nécessite beaucoup de données puisque, pour l'appliquer, il faut disposer de données sur toutes les caractéristiques pertinentes des biens immobiliers, ce qui rend son utilisation assez coûteuse.
- L'idée générale de la méthode hédonique est facilement compréhensible, mais certains de ses aspects techniques risquent d'être difficiles à expliquer aux utilisateurs.

5- CONCLUSION

Nous avons défini dans cette publication la notion de la *méthode hédoniste* comme considérant qu'un bien est essentiellement un ensemble de caractéristiques de performance et que sa valeur résulte du plaisir ou du déplaisir que ses caractéristiques procurent à son propriétaire ou usager. Ces attributs influencent significativement le prix du bien immobilier car en additionnant la valeur des différentes caractéristiques, on obtient une marge d'erreur assez faible. Cet aspect d'évaluation minimise la discordance sur le prix du marché du bien immobilier suite à sa performance par rapport aux autres outils d'évaluation que nous avons pu montrer dans ce travail de recherche.

La fiabilité des méthodes d'évaluation immobilière est intimement liée à l'information dont on dispose. L'expertise ne se heurte pas à des obstacles conceptuels mais à des difficultés d'information. En effet, la manque d'information handicape l'exercice de l'évaluation tout en justifiant son existence : rente de situation et rente d'information vont de paire. Nous comprenons alors que la diffusion de l'information, dans une large mesure, reste cantonnée à la sphère professionnelle. Toutefois, le client de l'immobilier est l'usufruitier de cette démarche. L'amélioration de la lecture du marché immobilier passe par l'élaboration d'outil d'information. Dans le cadre de l'approche traditionnelle, on a déjà signalé l'importance de la constitution d'une base de données. Cette base sera d'autant plus significative qu'elle s'inscrira dans le temps.

REFERENCES BIBLIOGRAPHIQUES

- Bover, O., and M. Izquierdo (2003), "Quality-adjusted Prices: Hedonic Methods and Implications for National Accounts," Investigaciones Económicas 27, 199–238.
- Butler, J.S., Y. Chang and A. Crews Cutts (2005), "Revision Bias in Repeat-Sales Hone Price Indices," Freddie Mac Working Paper No. 05–03.
- Francke, M.K. (2010), "Repeat Sales Index for Thin Markets: A Structural Time Series Approach," Journal of Real Estate Finance and Economics 41(1), 24–52.
- Francke, M.K., and G.A. Vos (2004), "The Hierarchical Trend Model for Property Valuation and Local Price
- Nations Unies (2009), Guide pratique pour l'établissement d'indices des prix à la consommation, New York et Genève: Nations Unies.
- Quigley, J.M. (1995), "A Simple Hybrid Model for Estimating Real Estate Price Indexes,"
- Schéma Simplifié d'Urbanisme de la commune Itaosy, Ministère de l'Aménagement du Territoire, 2008
- Schéma Simplifié d'Urbanisme de la commune Andranonahoatra, Ministère de l'Aménagement du Territoire, 2008