



# L'ANALYSE PAR REGRESSION LINEAIRE MULTIPLE ET LE MODELE A PLUSIEURS EQUATIONS : UNE ETUDE THEORIQUE

Par

**MOKEMBA MABATE Fidèle<sup>1(+)</sup>**

Chercheur à la Faculté des Sciences Economiques et de gestion de l'Université de Kinshasa en République Démocratique du Congo.

Email: [mokembafidel28@gmail.com](mailto:mokembafidel28@gmail.com)

**Ibrahim YERO KANDE<sup>2(+)</sup>**

Chercheur à la Faculté des Sciences Economiques et de gestion de l'Université de Kinshasa en République Démocratique du Congo.

Email: [kandeyero2@gmail.com](mailto:kandeyero2@gmail.com)

## Résumé

La régression linéaire multiple (RLM) est une technique statistique importante en économie et gestion consistant à utiliser deux variables indépendantes ou plus pour prédire ou expliquer le résultat d'une seule variable dépendante en modélisant la relation linéaire entre elles. L'analyse descriptive des données repose sur une démarche en plusieurs étapes. Il s'agit tout d'abord de définir les caractéristiques des variables prises une à une (analyse univariée ou tri à plat) puis une observation des liens qui les caractérisent deux par deux (Analyse bivariée ou trivariée) pour finir par l'observation des structures multiples liant plusieurs variables (analyse multivariée) on a distingué alors deux familles principales, la première a consisté à observer les liens unissant une variable avec plusieurs autres ( $1 \rightarrow n$ ) la seconde a considéré simultanément les structures multiples liant différentes variables ( $n \rightarrow$ ). Les résultats ont montré que la statistique de F calculée par le logiciel Eviews est  $F = 234.457$  et la probabilité associée est inférieure à 5% ( $0,000 < 0,05$ ). Or la statistique lue dans la table de Fisher à 2 degrés de liberté au seuil (de 5% est de 5.991 donc l'hypothèse nulle ( $H_0$ ) est rejetée et le modèle est globalement ajusté (0.868) qui renseigne aussi sur la qualité du modèle économétrique. Les résultats de l'estimation montrent que ces deux variables indépendantes ou explicatives sont statistiquement significatives au vu de la probabilité qui leur est attribuée : La variable IPR est significative au seuil de 1% car la probabilité (prob.) est inférieure à 1% ; La variable TVA est significative au seuil de 1% car la probabilité (prob.) est inférieure à 1%

**Mots-clés:** Analyse, Regression linéaire multiple, Modèle, Equation.

## I. Introduction

Il est délicat de fournir une définition unique de la notion de modèle. Dans le cadre de l'économétrie, nous pouvons considérer qu'un modèle consiste en une présentation formalisée d'un phénomène sous forme d'équations dont les variables sont des grandeurs économiques. L'objectif du modèle est de représenter les traits les plus marquants d'une réalité qu'il cherche à styliser. Le modèle est donc l'outil que le modélisateur utilise lorsqu'il cherche à comprendre et à expliquer des phénomènes. Pour ce faire, il émet des hypothèses et explicite des relations (Bourbonnais, 2019). Le modèle est donc une présentation schématique et partielle d'une réalité naturellement plus complexe. Toute la difficulté de la modélisation consiste à ne retenir que la ou les représentations intéressantes pour le problème que le modélisateur cherche à expliciter. Ce choix dépend de la nature du problème, du type de décision ou de l'étude à effectuer. La même réalité peut ainsi être formalisée de diverses manières en fonction des objectifs (Bourbonnais, 2019).

Dans les sciences sociales, et particulièrement en économie, les phénomènes étudiés concernent le plus souvent des comportements afin de mieux comprendre la nature et le fonctionnement des systèmes économiques. L'objectif du modélisateur est, dans le cadre de l'économétrie et au travers d'une mesure statistique, de permettre aux agents économiques (ménages, entreprises, État...) d'intervenir de manière plus efficace. La construction d'un modèle comporte un certain nombre d'étapes qui sont toutes importantes. En effet, en cas de faiblesse d'un des « maillons », le modèle peut se trouver invalidé pour cause d'hypothèses manquantes, de données non représentatives ou observées avec des erreurs, etc.

Le modèle étant spécifié, il convient de collecter les variables représentatives des phénomènes économiques. Ce choix n'est pas neutre et peut conduire à des résultats différents, les questions qu'il convient de se poser sont par exemple : *Faut-il raisonner en dollars constants ou en dollars courants ? Les données sont-elles brutes ou corrigées des variables saisonnières ? Quel taux d'intérêt faut-il retenir (taux au jour le jour, taux directeur de la Banque centrale...) ?* Etc. Dans le cadre de modèle spécifié en séries temporelles, les relations entre les variables ne sont pas toujours synchrones mais peuvent être décalées dans le temps.

L'économétrie est un outil à la disposition de l'économiste qui lui permet d'infirmer ou de confirmer les théories qu'il construit. Le théoricien postule des relations ; l'application de méthodes économétriques fournit des estimations sur la valeur des coefficients ainsi que la précision attendue (Bourbonnais, 2019). Une question se pose alors : pourquoi estimer ces relations, et les tester statistiquement ? Plusieurs raisons incitent à cette démarche : tout d'abord cela force l'individu à établir clairement et à estimer les interrelations sous-jacentes. Ensuite, la confiance aveugle dans l'intuition peut mener à l'ignorance de liaisons importantes ou à leur mauvaise utilisation. De plus, des relations marginales mais néanmoins explicatives, qui ne sont qu'un élément d'un modèle global, doivent être testées et validées afin de les mettre à leur véritable place. Enfin, il est nécessaire de fournir, en même temps que l'estimation des relations, une mesure de la confiance que l'économiste peut avoir en celles-ci, c'est-à-dire la précision que l'on peut en attendre. Là encore, l'utilisation de méthodes purement qualitatives exclut toute mesure quantitative de la fiabilité d'une relation.

Ce papier porte sur la régression linéaire multiple ainsi que les modèles à plusieurs équations. Cette évaluation qui porte, à la fois, sur la régression linéaire multiple ainsi que les modèles à plusieurs équations. Veut effectuer une analyse en deux temps. D'abord une analyse sur les étapes de la régression linéaire multiple ; ensuite une analyse sur les modèles à plusieurs équations.

## II. Analyse sur la régression linéaire multiple.

D'abord signalons que la régression linéaire simple est une approche statistique qui permet d'évaluer la liaison linéaire entre deux variables. Plus précisément, elle permet de quantifier la relation et d'évaluer sa significativité.

La régression linéaire multiple est une généralisation de la régression linéaire simple, dans le sens où cette approche permet d'évaluer les relations linéaires entre une variable réponse et plusieurs variables explicatives (de type numérique ou catégorielle).

### 2.1. Etablissement du modèle et estimation des paramètres

Le point de départ est l'estimation des paramètres d'une régression mettant en jeu une variable endogène  $Y$  et  $p$  variables exogènes  $X_j$ . Nous disposons de  $n$  observations. L'équation de régression s'écrit :

$$Y_t = a_0 + a_1x_{1t} + a_2x_{2t} + \dots + a_kx_{kt} + \mu_t(1);$$

$$t = 1, 2, 3, \dots, n.$$

En variant :

$$\begin{cases} Y_1 = a_0 + a_1x_{11} + a_2x_{21} + \dots + a_kx_{k1} + \mu_1 \\ Y_2 = a_0 + a_1x_{12} + a_2x_{22} + \dots + a_kx_{k2} + \mu_2 \\ \cdot \\ \cdot \\ Y_n = a_0 + a_1x_{1n} + a_2x_{2n} + \dots + a_kx_{kn} + \mu_n \end{cases}$$

Sous forme matricielle,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} \dots X_{k1} \\ 1 & X_{12} \dots X_{k2} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_{1n} \dots X_{kn} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ a_k \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \mu_n \end{pmatrix} \quad (3).$$

Sous forme condensée

$$\vec{Y} = \underline{X}\vec{A} + \vec{\varepsilon} \quad (4)$$

Le modèle prévisionnel est  $\hat{y} = \underline{X}\hat{A}$  (5)

$$\hat{A} = (\underline{X}'\underline{X})^{-1} \cdot \underline{X}'\vec{Y} = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \cdot \\ \cdot \\ \hat{a}_n \end{pmatrix} \text{Modèle estimé des paramètres par moindres carrés ordinaires.}$$

En outre, l'estimateur doit être non-biaisé, d'où  $V(\hat{A}) = \sigma^2 (\underline{X}' \underline{X})^{-1} = \overline{\hat{A}}$  où  $\sigma^2 = \gamma^2 = \frac{\sum ei^2}{n-k}$  avec  $n$  = nombre d'observations ;  $k$  = nombre des paramètres estimés.

$$\overline{\hat{A}} = \begin{pmatrix} \sigma_{\hat{a}_0}^2 & \text{cov}(\hat{a}_0, \hat{a}_1) & \text{cov}(\hat{a}_0, \hat{a}_k) \\ \text{cov}(\hat{a}_1, \hat{a}_0) & \sigma_{\hat{a}_1}^2 & \text{cov}(\hat{a}_1, \hat{a}_k) \\ \text{cov}(\hat{a}_k, \hat{a}_0) & \text{cov}(\hat{a}_k, \hat{a}_1) & \sigma_{\hat{a}_k}^2 \end{pmatrix}$$

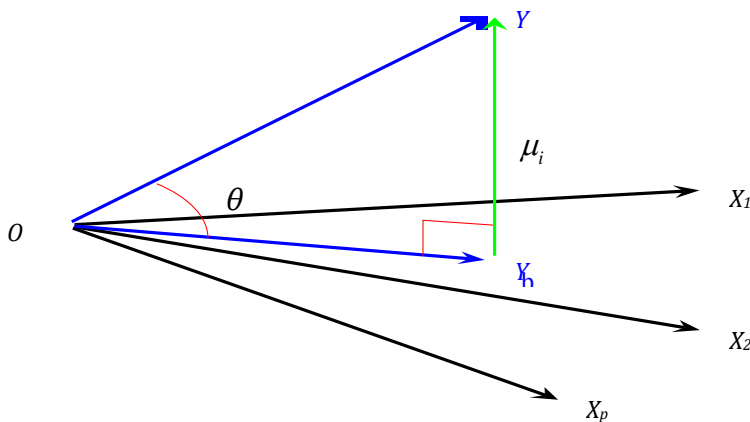
En utilisant les variables centrées réduites, nous avons :

$$\overline{\hat{A}} = (\underline{x}' \underline{x})^{-1} \cdot \underline{x}' \vec{y} = \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \cdot \\ \hat{a}_n \end{pmatrix} \text{ où } \hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{x}_1 - \hat{a}_2 \bar{x}_2 + \dots + \hat{a}_k \bar{x}_k .$$

Et

$$\underline{x}' \underline{x} = \begin{pmatrix} \sum x_1^2 & \sum x_1 x_2 & \dots & \sum x_1 x_k \\ \sum x_2 x_1 & \sum x_2^2 & \dots & \sum x_2 x_k \\ \dots & \dots & \dots & \dots \\ \sum x_k x_1 & \sum x_k x_2 & \dots & \sum x_k^2 \end{pmatrix} \text{ et } \underline{x}' \vec{y} = \begin{pmatrix} \sum x_i y \\ \sum x_2 y \\ \dots \\ \sum x_k y \end{pmatrix}$$

**Graphiquement : Démonstration graphique de l'erreur.**



**a. Les hypothèses stochastiques<sup>1</sup>**

H1 : X est une matrice non aléatoire, c'est-à-dire les réalisations de X sont observées sans erreur, notamment de mesure.

<sup>1</sup> Tirées de J.P Bosonga (Avril, 2021).

H2 : L'espérance mathématique de l'erreur est nulle. Donc, le modèle empirique est un bon modèle, au sens où il représente une bonne approximation de la réalité. Formellement, on a :

$$E(U)=0, \text{ alors } E(Y) = E(XA) + E(U)=XA$$

H3 :  $Var(U) = E[UU'] = \sigma_u^2 I = \Omega_u$  (6). Avec  $\Omega_u$  est la matrice des variances et de covariances de l'erreur U. Et I représente la matrice identité de dimension (nxn). En principe, on fait ici deux hypothèses.

- Les lois de probabilité suivies par les erreurs U ont la même variance :  $E(u_i^2) = \sigma^2, \forall t$  (variance constante et finie).  
Donc, quelle que soit la partie considérée de l'échantillon, la variance du terme d'erreur est constante. Cette hypothèse est généralement appelée l'homoscédasticité, c'est-à-dire que la dispersion ou variance du terme d'erreur est la même du début à la fin de la période. Lorsque cette hypothèse n'est pas vérifiée, on dit qu'il y a hétéroscédasticité (dispersion ou variance inégale).
- Les erreurs, prises deux à deux, ne sont pas corrélées :  $E(u_i u_j) = 0, \forall_i \neq j$

Cette hypothèse implique de supposer l'absence d'autocorrélation ou l'absence de corrélation sérielle. Si tel n'est pas le cas, on dit que les erreurs sont autocorrélées.

H4 : Le vecteur de perturbation suit une loi normale à n dimension. Donc :  $U \sim N(0, \sigma_u^2 I)$ .

H5 : Covariance nulle entre U et X, c'est-à-dire :

$$cov(u_i, X_{ij}) = E(u_i X_{ij}) = 0 \quad (7).$$

Cette hypothèse suppose que l'erreur n'est pas corrélée avec les variables explicatives. C'est l'hypothèse d'endogénéité de la variable  $X_{ij}$ .

H6 : Absence de colinéarité entre les variables explicatives ; cela implique que la matrice  $(XX')$  est régulière (non singulière) et que la matrice  $(XX')^{-1}$  existe. Donc le rang de matrice des variables explicatives doit être égal à k :  $\rho(X) = K$ .

H7. Le nombre d'observations doit être supérieur au nombre de paramètre à estimer ( $n > k$ ).

En termes de propriétés, les estimateurs de MCO doivent être BLUE (Best Linear Unbiased Estimator). C'est-à-dire qu'il satisfait le théorème de Gauss-Markov. En effet, un bon estimateur doit être linéaire, non biaisé et convergent.

## b. Qualité d'ajustement

L'équation d'analyse de la variance permet d'évaluer la qualité de l'ajustement d'un modèle. En effet, plus la variance expliquée est proche de la variance totale, meilleur est l'ajustement du modèle.

Ainsi, le coefficient de détermination  $R^2 = \frac{\text{Variance expliquée}}{\text{Variance Totale}} = \frac{SCE}{SCT} = 1 - \frac{SCE}{SCT}$  (8), ou encore dans le cas des variables

centrées réduite.

$R^2 = 1 - \frac{e'e}{y'y} = 1 - \frac{\sum_t e_t^2}{\sum_t y_t^2}$  (9). Pour une meilleure utilisation, le coefficient de détermination peut être ajusté, par :

$$\bar{R} = 1 - \frac{e'e/(n-K)}{y'y/(n-1)} \quad (10).$$

### c. Estimation des modèles sous contraintes linéaires

Il s'agit dans ce contexte des restrictions linéaires qu'on peut porter sur les coefficients des modèles linéaires. En effet, l'énoncée est donnée par :

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \mu_t \quad (11).$$

Les contraintes linéaires peuvent prendre l'une des formes suivantes :

- a)  $\beta_2 = 0$  : une condition ou contrainte
- b)  $\beta_2 = \beta_3$  : une condition ou contrainte
- c)  $\beta_2 = -1$  et  $\beta_3 = 0$  : deux conditions ou contraintes
- d)  $\beta_2 + \beta_3 = 1$  et  $\beta_1 = 0$  : deux conditions ou contraintes
- e)  $\beta_2 = \beta_3 = \beta_4 = 0$  : trois conditions ou contraintes.

Ainsi, la prise en compte de ces contraintes dans l'estimation peut être effectuée de deux façons :

- En intégrant les contraintes dans la spécification du modèle à estimer ;
- En utilisant directement la méthode de moindres carrés sous-contraintes.

### d. Tests de stabilité

Les tests de stabilité des paramètres estimés du modèle de régression ou test de robustesse du modèle estimé, visent à vérifier si les valeurs des coefficients estimés sont stables sur l'ensemble de la période de l'étude considérée. Généralement en régression multiple, on expose deux tests, à savoir le test de Chow, et celui de Cusum et Cusum carré.

#### a) Test de Chow.

Il est également appelé test de changement structurel ou changement de régime, il permet de mesurer si les coefficients du modèle se sont modifiés au cours du temps, avec une période de rupture bien précise.

Ce test respecte les étapes suivantes :

- 1) Estimation du modèle général, et détermination de SCR (n-K, ddl).
- 2) Estimation du modèle de la première partie et détermination de la SCR<sub>1</sub> (n<sub>1</sub>-K, ddl).
- 3) Estimation du modèle de la deuxième partie et détermination de la (n<sub>2</sub>-K, ddl).
- 4) Calculer SCR<sub>3</sub>= SCR<sub>1</sub>+ SCR<sub>2</sub>, avec (n<sub>1</sub>-K) + (n<sub>2</sub>-K) = (n-2K) ddl.

5) Calcul de la Statistique :  $F_{cal} = \frac{SCR - (SCR_1 + SCR_2)}{SCR_1 + SCR_2} \cdot \frac{n - 2K}{K}$

Règle de décision :  $F_{cal} > F_{th}$ , on rejette H<sub>0</sub> : c'est-à-dire les deux régressions sont différentes, il n'y a pas stabilité des paramètres.

#### b) Tests de Cusum et Cusum carré.

Les deux tests sont basés sur les résidus et la somme de résidus, ils sont utilisés sous forme graphique (généralement, à travers des Logiciels) lorsque la période de rupture n'est pas identifiée.

**e. Exemple d'illustration (Sur 72 Observations mensuelles)**

RECETTES	IPR	TVA	RECETTES	IPR	TVA
642100886.2	401250098.2	140942724	1158989662	767647350.1	224701477.7
606178672.1	378204368.1	121748307	991680282.7	711542986.4	216390279.5
802489653.2	424903523.4	283971009.4	1005944949	680417700.9	241373806.2
823503486.6	368261918.2	146188550.5	1674995885	776855506.1	221073671.6
511632829.6	350700998.2	105094151.5	1125564384	766480460.7	303173775.4
573789572.8	423126218.4	123784478	1133427284	839343887.1	238308230.7
835958253.2	415971017	119208716.2	1428509618	815712253.9	245499463.8
614235133.4	361269330.3	179692123	1396249997	757165413.1	563014117.7
541991709.9	328697059.2	164011295.4	1226728114	736248224.7	437016755.1
621016606.9	366410645.1	136397830	1075641404	749208371.2	251639396.4
853208110.8	346547284.2	160799527	1974509251	933708549	521151009.4
796029995.7	480491826.7	231999318	1286389298	1023424542	215944277.4
681349023.5	344848756.6	200776360	1202462469	764380227.9	239496368.2
748430525.7	367537211.8	147771780.9	753385142	528269021	173680360.7
713922779.1	485543576.8	152384407	1129840099	778224553	279682795.2
912057812	357953094.2	137780805	1283752099	627741504.2	150830512.8
673647674.5	395166633.3	197313292.5	1259734795	889841352.2	294177722.9
641325631.8	396975027.9	164157507	1221852024	894669061.7	259047596.2
890836774	431590672.3	130476000	1536758295	1029612014	173497677.1
702271489.8	363724288.3	203406147	1598758347	827917655.7	604285028.5
686917007.6	337423858	178330609	1252890933	645648596.7	517911744.8
667890939	414285067.9	151147961	1828433159	901809832.1	795535093.9
1047423317	424227723	245868762.7	2728515583	1485394576	737038002.3
793471243.8	568228735.2	102301659	900554764.1	598530939.1	214112659.7
697219624.3	422469123.3	157828106	945086878.6	549279134.2	195433476.8
605645009.6	393075794.9	115112314.9	853896297.4	540791606.9	239669761.7
807146865.9	563750333.4	175112204.3	1577226966	790265845.9	677721117
988207929.5	559217680.8	190617942.1	1620979996	680995869.3	491963492.1
964643660.7	582935766.2	271913665.4	1173624473	857659553.2	123195899.3
948263795.6	558592561.9	231301294.7	916835416.7	695750744.4	181372910
1222005137	608908272.2	215313307.6	1433626467	800291240.7	284185155.5
1016500736	611138507.5	235317786.3	1033600789	737670130.6	244196580.5
871777271.2	601012102.4	233119767	1098611607	628766572.3	309812349.1
918191903.8	639016051.1	219840407.3	1471912972	1059519033	295397055.8
1160416269	672088662.4	191921496.9	1367561547	804259223.6	349919114.5
1149870265	847020936.6	230305125.3	1612814350	984219900.3	424967799.6

Source : DGI/Agence du Sud-Kivu.

**Tableau 1 : Résultats de la régression avec deux variables**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
IPR	1.110077	0.093437	11.88048	0.0000
TVA	1.018460	0.143231	7.110586	0.0000
C	98775704	49453194	1.997357	0.0497
R-squared	0.871727	Mean dependent var		1.06E+09

Adjusted R-squared	0.868009	S.D. dependent var	3.89E+08
S.E. of regression	1.41E+08	Akaike info criterion	40.41347
Sum squared resid	1.38E+18	Schwarz criterion	40.50834
Log likelihood	-1451.885	F-statistic	234.4570
Durbin-Watson stat	2.706243	Prob (F-statistic)	0.000000

Source : Traitement des données via Eviews 3.11.

La statistique de F calculée par le logiciel Eviews est  $F = 234.457$  et la probabilité associée est inférieure à 5% ( $0,000 < 0,05$ ). Or la statistique lue dans la table de Fisher à 2 degrés de liberté au seuil (de 5% est de 5.991 donc l'hypothèse nulle ( $H_0$ ) est rejetée et le modèle est globalement ajustée (0.868) qui renseigne aussi sur la qualité du modèle économétrique. Les résultats de l'estimation montrent que ces deux variables indépendantes ou explicatives sont statistiquement significatives au vue la probabilité qui leur attribuée :

- La variable IPR est significative au seuil de 1% car la probabilité (prob.) est inférieure à 1% ;
- La variable TVA est significative au seuil de 1% car la probabilité (prob.) est inférieure à 1% ;

## II. Analyse sur les modèles à plusieurs équations<sup>2</sup>.

Deux types de modèles sont présentés, à savoir le modèle à équations apparemment indépendantes, et les modèles à équations simultanées.

### a. Modèle à équation apparemment indépendantes.

Il peut être présenté par la forme :

$$\begin{cases} Y_1 = X_1\beta_1 + \mu_1 \\ Y_2 = X_2\beta_2 + \mu_2 \\ \dots \\ Y_m = X_m\beta_m + \mu_m \end{cases} \quad (12) \text{ Il existe donc dans ce type de modèle une absence de simultanités.}$$

La méthode d'estimation dépend de la structure des erreurs :

Soit le modèle  $Y = X\beta + U$  (13).

$E(U_i) = 0$

$\text{Var}(U_i) = E(U_i^2) = \sigma_u^2$

$\text{Cov}(U_{it}, U_{is}) = 0$

$\text{Cov}(U_{it}, U_{jt}) = 0$

$\text{Cov}(U_{it}, U_{js}) = 0$

$E(UU) = \sigma_u^2 I$

Ainsi, comme toutes ces conditions sont réunies, utilisations des moindres carrés ordinaires (MCO), équation par équation.

En outre, lorsque les hypothèses des variances des équations ne sont pas plus respectées. On utilise la méthode **SUR (Seeming Unrelated Regression) de Zellner.**

#### a) Estimateurs

1) Si les valeurs de  $\sigma_i^2$  et  $\sigma_{ij}^2$  de la matrice de variances et covariances, sont connus, la méthode de Zellner consistant à appliquer les moindres carrés Généralisés (MCG) sur le système peut être utilisée. Ainsi :

$$\hat{\beta}_{SUR} = [X\Omega^{-1}X']^{-1} X\Omega^{-1}Y = [X'(\Sigma^{-1} \otimes I)X]^{-1} X'(\Sigma^{-1} \otimes I)Y \quad (14).$$

$$E(\hat{\beta}_{SUR}) = \beta \quad \text{et} \quad \text{Var}(\hat{\beta}_{SUR}) = (X\Omega^{-1}X')^{-1}.$$

<sup>2</sup> Tirées de JP Bosonga (Avril, 2021).



2) Si les valeurs de  $\sigma_i^2$  et  $\sigma_{ij}^2$  de la matrice de variances et covariances, sont inconnus, Zellner a proposé une méthode itérative.

### b) La méthode de Zellner et l'autocorrélation des erreurs.

Lorsque cette situation surgit, il faut d'abord transformer les variables avant d'appliquer la méthode de Zellner.

#### b. Modèles à équations simultanées.

Parfois, les modèles à une équation sont insuffisants, d'où la nécessité de recourir aux modèles à plusieurs équations. C'est-à-dire aux systèmes d'équation. Mais dans les équations simultanées, les équations ne sont pas indépendantes les unes envers les autres. Et cette interdépendance simultanée des différentes équations peut avoir des conséquences importantes au niveau de l'estimation.

#### a) Présentation

C'est juste un exemple qui montre les modèles à équations simultanées, ainsi :

$$\begin{cases} Q_{dt} = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t + \mu_{dt} & (15) \\ Q_{ot} = \beta_0 + \beta_1 P_t + \beta_1 P_{ft} + \mu_{ot} & (16) \\ Q_{dt} = Q_{ot} = Q_t & (17) \end{cases}$$

Autant d'équation qu'on peut avoir (par exemple d'autres modèles macroéconomiques) pour déboucher à différentes équations simultanées.

Les équations issues de la théorie économique sont appelées des équations structurelles ou de comportement, car elles décrivent soit la structure d'une économie ou le comportement d'un agent économique.

#### b) Problème d'identification

- Si on ne peut pas obtenir les estimations des paramètres de la forme structurelle à partir des coefficients de la forme réduite, le modèle est dit non identifié ou sous-identifié. Un modèle est sous-identifié si au moins une équation du modèle est sous-identifiable. Cela signifie que le nombre d'équations est inférieure au nombre des paramètres à identifier dans la forme structurelle.
- Si les valeurs numériques des paramètres structurels peuvent être obtenues à partir des coefficients de la forme réduite, on dit que le modèle est identifié.
- On dira qu'une équation est exactement ou pleinement ou strictement ou juste, identifiée ; s'il existe une solution unique pour les coefficients structurels obtenus à partir des coefficients de la forme réduite du modèle.
- Elle est dite « sur-identifiée » si plusieurs valeurs numériques des coefficients de la forme réduite peuvent correspondre aux paramètres structurels. Le modèle est sur-identifié si toutes les équations sont sur-identifiées ou certaines équations sont sur-identifiées et d'autres exactement identifiées.

#### c) Condition d'ordre d'identification

Dans un modèle où :  $G$ =Nombre de variable endogènes dans le modèle,  $g$ =nombre des variables endogènes dans une équation donnée,  $K$ =nombre de variables prédéterminées dans le modèle y compris le vecteur, unité de la constante (intercept) et  $k$ =nombre de variables prédéterminées présentes dans une équation donnée y compris le vecteur, unité du terme constant. Ainsi, les conditions sont telles que :  $K-k \geq g-1$

- Si  $K-k = g-1$ , l'équation est exactement identifiée ;
- Si  $K-k > g-1$ , elle est sur-identifiée ;
- Si  $K-k < g-1$ , elle est sous-identifiée.

#### d) Conditions de rang d'identification

Les conditions d'ordre et de rang d'identification conduisent aux principes généraux pour identifier une équation structurelle dans le

système de  $G$  équations simultanées

0. Si  $K-k > g-1$  et si le rang de la matrice  $P$  d'ordre  $(G-1)$  est maximum, alors l'équation considérée est sur-identifiée ;
1. Si  $K-k = g-1$  et si le rang de la matrice  $P$  d'ordre  $(G-1)$  est maximum, alors l'équation considérée est exactement identifiée ;
2. Si  $K-k \geq g-1$  et si le rang de la matrice  $P$  d'ordre  $(G-1)$  n'est pas maximum, alors l'équation considérée est sous-identifiée ;
3. Si  $K-k < g-1$  et si l'équation structurelle n'est pas identifiée. Dans ce cas, le rang de la matrice  $P$  est inférieur à  $(G-1)$ .

**e) Conditions d'identification en cas de restrictions linéaires**

Lorsque le modèle contient des restrictions linéaires sur les paramètres, la condition d'ordre devient :

$(K-k) + r \geq g-1$ , or  $r$  désigne le nombre des restrictions linéaires.

- Si  $(K-k) + r > g-1$ , alors l'équation est sur-identifiée ;
- Si  $(K-k) + r = g-1$ , alors l'équation est juste identifiée ;
- Si  $(K-k) + r < g-1$ , alors l'équation est sous-identifiée ;

Quant aux conditions nécessaire et suffisante de l'identification on a :

- Si  $(K-k) + r < g-1$ , ou si la condition de rang n'est pas vérifiée, l'équation est sous-identifiée ;
- Si  $(K-k) + r = g-1$ , et que la condition de rang est vérifiée, l'équation est exactement identifiée ;
- Si  $(K-k) + r > g-1$ , et que la condition de rang est vérifiée, l'équation est sur-identifiée.

**f) Méthode d'estimation**

Dans le cas d'un modèle exactement identifié ou sur-identifié, on peut adopter deux approches à l'estimation des équations structurelles, en l'occurrence les méthodes à information limitée et les méthodes à informations complète. Signalons tout d'abord que lorsque le modèle est sous-identifié, il n'y a pas d'estimation possible. En effet,

- Dans le cas des méthodes à information limitée, on estime équation par équation sous l'hypothèse qu'il n'existe pas de corrélations entre termes d'erreurs des différentes équations. En adoptant cette procédure, on néglige les informations contenues dans d'autres équations (information limitée) : Méthode de moindres carrés indirects, méthode de doubles moindres carrés et méthode de maximum de vraisemblance à information limitée.
- Dans les méthodes à information complète, on considère le modèle dans sa globalité en estimant les paramètres sous-hypothèses qu'il existe des corrélations entre erreurs. On utilise les méthodes suivantes : <sup>3</sup>les triples moindres carrés, le maximum de vraisemblance à information complète et les moments généralisés des systèmes d'équations.

**REFERENCES**

1. Bosonga Bofeki Lounga, J-P, Manuel d'Econométrie, Unikin/FASEG, DEA Economie-Gestion, Avril-2021.
2. Bourbonnais, R., Econométrie cours et exercice corrigés, 11<sup>ème</sup> Edition, Dunod, Paris, 2019.
3. DGI/Direction Provinciale du Sud-Kivu, Recettes mensuelles de 2015-2020, Div. Assiettes, 2021.
4. WWW.statistics.net.

<sup>3</sup> Qui sont lourdes en termes de mise en application, de la sensibilité aux erreurs de spécification et à l'existence de solutions non linéaires.