The Complete Guide to Ordinary Least Squares (OLS) Regression Using EViews

Josué, ANDRIANADY

2024

Manuel Complet de la Régression par les Moindres Carrés Ordinaires (OLS) sous EViews

ANDRIANADY Ravahiny Josué

Abstract

This manual offers a detailed and accessible approach to Ordinary Least Squares (OLS) regression using EViews software. Aimed at students and practitioners in economics and the social sciences, it covers the fundamental concepts of OLS regression, practical implementation in EViews, as well as advanced techniques for result analysis. Each section includes concrete examples and EViews commands to facilitate understanding and application of statistical methods. Additionally, diagnostic tests and strategies to assess model validity are presented to ensure the robustness of results. This manual serves as a practical guide for those looking to deepen their knowledge and skills in econometrics.

Avant-propos

Dans un monde où les données jouent un rôle essentiel dans la prise de décision, la maîtrise des outils statistiques devient indispensable. La régression par les moindres carrés ordinaires (OLS) est l'une des méthodes les plus utilisées pour analyser et interpréter les relations entre variables. Ce manuel a été conçu pour accompagner les étudiants, chercheurs et professionnels dans l'apprentissage et l'application de cette technique à l'aide du logiciel EViews.

L'objectif de ce manuel est de fournir une ressource complète, combinant théorie et pratique. Chaque section s'efforce de décomposer des concepts complexes en éléments simples, permettant ainsi une meilleure assimilation des notions. Les commandes EViews sont présentées de manière claire, accompagnées d'exemples concrets pour illustrer leur application.

Comme l'a si bien dit Isaac Newton : "Ce que nous savons est une goutte d'eau ; ce que nous ignorons est un océan." Ce manuel vise à explorer cet océan de connaissances en économétrie, en offrant des outils et des perspectives qui permettront d'enrichir votre compréhension des méthodes statistiques.

J'espère que ce manuel sera une aide précieuse pour ceux qui souhaitent explorer les merveilles de l'économétrie et des méthodes statistiques.

ANDRIANADY Ravahiny Josué

Contents

1 Introduction

La régression par les moindres carrés ordinaires (OLS) constitue l'une des techniques les plus couramment utilisées en économétrie pour estimer les relations entre une variable dépendante et une ou plusieurs variables explicatives. Elle repose sur des hypothèses de linéarité, d'homoscédasticité des erreurs, et d'absence de corrélation des résidus. Lorsqu'elles sont respectées, l'OLS permet de produire des estimations fiables et impartiales, facilitant ainsi l'analyse et l'interprétation des résultats économiques.

L'un des principaux avantages de l'OLS est sa simplicité d'implémentation et son caractère intuitif, ce qui en fait une méthode accessible pour de nombreux chercheurs appliqués. Toutefois, elle présente également des limites.

Ce manuel vise à guider les utilisateurs dans l'application pratique de l'OLS à travers le logiciel EViews, un outil puissant et convivial pour mener des analyses économétriques. Nous y aborderons les différentes étapes de la régression OLS, de l'importation des données à l'interprétation des résultats, en passant par les diagnostics des modèles et les tests de robustesse. Grâce à une série d'exemples concrets, ce guide propose une approche pédagogique pour rendre l'OLS accessible et compréhensible.

2 Revue de la littérature

La régression par les moindres carrés ordinaires (OLS) est une méthode utilisée pour estimer les relations linéaires entre une variable dépendante et une ou plusieurs variables explicatives [2]. Le but est de minimiser la somme des carrés des résidus, qui sont les différences entre les valeurs observées et les valeurs prédites par le modèle.

Le modèle OLS est généralement exprimé comme suit :

$$Y_t = \alpha + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_n X_{n,t} + \epsilon_t \tag{1}$$

Où :

- Y_t est la variable dépendante,
- $X_{1,t}, X_{2,t}, \ldots, X_{n,t}$ sont les variables explicatives,
- α est l'ordonnée à l'origine,
- $\beta_1, \beta_2, \ldots, \beta_n$ sont les coefficients de régression,
- ϵ_t est le terme d'erreur (résidu).

L'objectif est de trouver les valeurs des coefficients qui minimisent la somme des carrés des résidus.

A titre d'illustration,dans Figure 1 l'OLS est utilisé pour modéliser la relation entre la taille des élèves (en centimètres) et leurs notes, en traçant une ligne de régression qui résume cette relation. Les points bleus sur le graphique représentent les données réelles, chaque point indiquant la taille et la note d'un élève. La ligne rouge symbolise la tendance générale, suggérant qu'en moyenne, plus un élève est grand, plus sa note tend à être élevée, même si cela ne s'applique pas à tous. Les flèches vertes illustrent les erreurs de prédiction, montrant la différence entre les notes réelles et celles prédites par la ligne de régression. L'OLS cherche à minimiser ces écarts pour fournir la meilleure approximation de la relation entre la taille et les notes.

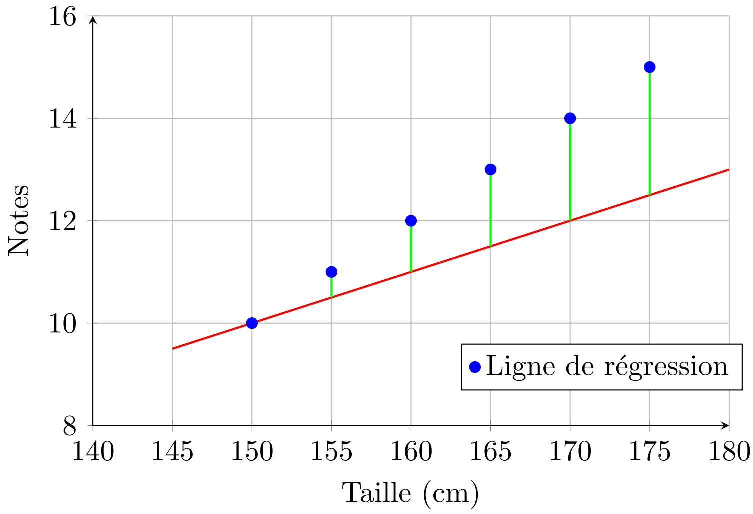

Figure 1: Relation entre la taille et les notes avec la ligne de régression et les erreurs (en vert)

2.1 Avantages de l'OLS

L'un des principaux avantages de l'OLS,[1], est sa simplicité. Elle est facile à mettre en œuvre, ce qui en fait un choix populaire pour les économistes appliqués et les chercheurs en sciences sociales. Le cadre linéaire qu'elle propose permet une interprétation intuitive des relations entre les variables, et la méthode produit des résultats facilement compréhensibles.

Murray [6] complète cette analyse en notant que, sous les hypothèses de linéarité, d'homoscédasticité (variance constante des erreurs), et de non-corrélation des erreurs, l'OLS fournit ce que l'on appelle le meilleur estimateur linéaire non biaisé (BLEU). Cela signifie que l'estimation par OLS est la plus précise (avec la variance la plus faible) parmi les estimateurs linéaires impartiaux. Cette caractéristique fait de l'OLS un outil particulièrement puissant lorsque les hypothèses sous-jacentes sont respectées.

2.2 Limites et critiques de l'OLS

Cependant, malgré ces avantages, l'OLS n'est pas sans inconvénients.Certains auteurs comme Phillips [7] critique la méthode pour sa sensibilité aux violations des hypothèses sous-jacentes. En particulier, l'OLS suppose une relation linéaire entre les variables et une homoscédasticité. Lorsque ces hypothèses sont violées, par exemple en présence d'hétéroscédasticité (variance inégale des erreurs) ou de non-linéarité, les estimations peuvent devenir biaisées et inefficaces.

Une autre faiblesse notoire de l'OLS est sa sensibilité aux valeurs aberrantes (outliers). Comme le soulignait Klein [4], la présence de points de données aberrants peut avoir un impact disproportionné sur les résultats des estimations. Ces valeurs extrêmes peuvent déformer les résultats et rendre les conclusions trompeuses ou peu fiables.

2.3 Alternatives à l'OLS

Face à ces critiques, plusieurs chercheurs ont exploré des méthodes alternatives. Par exemple, les méthodes de régression robustes, telles que la régression quantile ou les moindres carrés généralisés (GLS), sont souvent recommandées lorsque les hypothèses de l'OLS ne sont pas respectées [5]. Ces techniques permettent de prendre en compte l'hétéroscédasticité ou les structures de dépendance plus complexes, tout en atténuant l'impact des valeurs aberrantes.

De plus, avec le développement des modèles économétriques plus avancés, des méthodes comme les modèles de variables instrumentales (IV) ou les régressions non linéaires sont de plus en plus adoptées pour répondre aux limites de l'OLS dans des contextes où les relations sont plus complexes [8].

3 Estimation du modèle OLS dans EViews

A titre d'information, dans ce manuel nous allons utiliser les commandes et non les boutons dans Eviews. Les commandes sont à insérés comme dans la Figure 2.

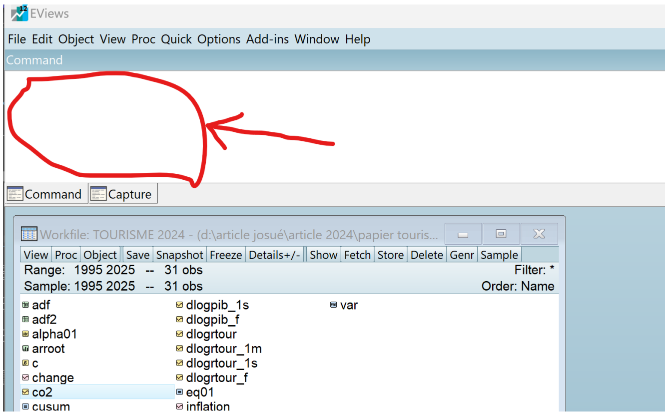

Figure 2: Indication où on doit écrire les commandes

3.1 Préparation des données

La préparation des données est une étape cruciale avant l'estimation d'un modèle. Elle garantit que vos analyses reposent sur des données complètes, fiables et cohérentes. Une mauvaise préparation des données peut entraîner des estimations biaisées, des résultats erronés ou non interprétables. Cette section détaille les étapes essentielles pour bien préparer les données dans EViews.

- **Importer les données :** La première étape consiste à charger vos données dans EViews. Assurez-vous que le fichier source (CSV, Excel, etc.) est correctement formaté, avec des colonnes bien définies correspondant aux différentes variables. Dans EViews, allez dans le menu *File* et sélectionnez *Import - Import from File*. Suivez les instructions pour choisir votre fichier, puis configurez correctement les options d'importation (par exemple, spécifiez si la première ligne contient les noms des variables). Une fois les données importées, EViews créera une base de données contenant toutes vos séries temporelles ou variables observées.

- **Vérifier la complétude des données :** Avant d'utiliser les données pour une estimation, il est essentiel de vérifier s'il existe des valeurs manquantes ou aberrantes dans vos séries. Pour afficher un aperçu des données, utilisez la commande suivante dans la fenêtre de commande d'EViews :

```
show y x1 x2
```

Ici, `y` représente la variable dépendante, et `x1`, `x2` sont les variables explicatives. Cette commande ouvrira une feuille de calcul vous permettant de visualiser les données et de détecter d'éventuelles valeurs manquantes (NA) ou incohérentes (comme des valeurs extrêmes ou erronées). Il est important d'examiner la distribution des données pour s'assurer qu'elles sont conformes aux attentes théoriques et empiriques.

- **Traiter les valeurs manquantes :** Si vous identifiez des valeurs manquantes, vous devez décider comment les traiter. Plusieurs options sont possibles :
  - *Suppression des observations :* Si les valeurs manquantes sont peu nombreuses, vous pouvez envisager de supprimer les lignes correspondantes. Cependant, cette approche peut entraîner une perte d'information, surtout si les données manquantes sont nombreuses.
  - *Interpolation :* EViews permet de remplir les valeurs manquantes à l'aide de techniques d'interpolation (linéaire, polynomiale, spline, etc.). Pour ce faire, sélectionnez la série concernée, allez dans le menu *Proc - Interpolate* et choisissez la méthode appropriée.
  - *Imputation statistique :* Une autre option consiste à estimer les valeurs manquantes à partir de méthodes statistiques comme la moyenne, la médiane, ou des modèles plus sophistiqués (comme les régressions multiples pour prédire les valeurs manquantes).
- **Vérifier la cohérence des données :** Outre les valeurs manquantes, il est également important de vérifier la cohérence et la plausibilité des données. Pour cela, vous pouvez :
  - Examiner les statistiques descriptives en utilisant la commande suivante :

    ```
    stats y x1 x2
    ```

    Cela affichera des informations comme la moyenne, l'écart-type, les minimums et maximums, qui permettent d'identifier rapidement des valeurs aberrantes.
  - Visualiser les séries temporelles ou les histogrammes des données avec :

    ```
    line y x1 x2
    ```

    Ces graphiques peuvent révéler des anomalies visuelles, comme des sauts brusques ou des tendances inhabituelles.
- **Transformer les données si nécessaire :** Si vos données présentent des caractéristiques indésirables, comme une forte asymétrie ou une hétéroscédasticité, vous pouvez envisager de les transformer. Les transformations les plus courantes incluent :
  - La prise de logarithmes pour stabiliser la variance.
  - La différence première pour rendre une série stationnaire.

  Ces transformations peuvent être réalisées directement dans EViews en appliquant des fonctions aux séries ou en créant de nouvelles séries transformées à partir des séries existantes.

3.2 Estimation du modèle OLS

3.2.1 Modèle de régression simple

Pour une régression simple (une variable explicative X_1) :

```
ls y c x1
```

Où :

- y est la variable dépendante,
- c est la constante du modèle,
- $x1$ est la variable explicative.

3.2.2 Modèle de régression multiple

Pour une régression multiple (plusieurs variables explicatives) :

```
ls y c x1 x2 x3
```

Cette commande estime la relation entre la variable Y et les variables explicatives X_1, X_2, X_3.

3.3 Interprétation des résultats

Après avoir estimé le modèle, EViews affiche une fenêtre contenant les coefficients, les statistiques t et les p-values. Voici comment interpréter ces résultats :

- **Coefficients (β) :** Ces valeurs représentent l'effet marginal de chaque variable explicative sur la variable dépendante. Par exemple si on reprend l'exemple cité précédemment sur les notes et la taille, si le coefficient $\beta_1 = 0.5$ pour la variable X_1 (taille), cela signifie qu'une augmentation de 1 cm de X_1 entraîne une augmentation de 0.5 point de Y (note). Inversement, si $\beta_2 = -0.3$ pour la variable X_2 (absences), cela signifie qu'une augmentation d'une absence entraîne une diminution de 0.3 point de Y.
- **Prob :** Les p-values associées aux coefficients indiquent la signification statistique de chaque coefficient. Par exemple, si la valeur p pour β_1 est 0.02, cela signifie que le coefficient est significatif au seuil de 5%, indiquant une relation statistiquement significative entre X_1 et Y. En revanche, si la valeur p pour β_2 est 0.15, cela indique que ce coefficient n'est pas significatif au seuil de 5%.
- **R-squared (R^2) :** Cette statistique mesure la proportion de la variance de la variable dépendante qui est expliquée par le modèle. Par exemple, si $R^2 = 0.75$, cela signifie que 75% de la variance des notes peut être expliquée par le modèle, ce qui indique un bon ajustement.
- **Statistique F :** Cette statistique teste l'hypothèse selon laquelle tous les coefficients du modèle sont simultanément égaux à zéro. Par exemple, si la statistique F est 10 avec une p-value de 0.004, cela indique que le modèle est globalement significatif, ce qui signifie qu'il existe une relation entre au moins une des variables explicatives (comme la taille ou les absences) et la variable dépendante (les notes).

Dependent Variable: PIB
Method: Least Squares
Date: 09/27/24 Time: 20:59
Sample: 1990 2020
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PPT	129.5220	37.91260	3.416332	0.0020
TCH	0.014247	0.015283	0.932234	0.3592
C	186.9442	26.03551	7.180353	0.0000

R-squared	0.561343	Mean dependent var	329.1394
Adjusted R-squared	0.530010	S.D. dependent var	85.99966
S.E. of regression	58.95776	Akaike info criterion	11.08329
Sum squared resid	97328.48	Schwarz criterion	11.22206
Log likelihood	-168.7909	Hannan-Quinn criter.	11.12852
F-statistic	17.91559	Durbin-Watson stat	0.811513
Prob(F-statistic)	0.000010		

Figure 3: Illustration fenêtre de résultat OLS

Exemple cas fictif: PIB en fonction taux d'alphabétisation et dépenses investissement

Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
C	5000.000	1000.000	5.000	0.000
Taux d'alphabétisation	300.000	50.000	6.000	0.000
Dépenses d'investissement	2.000	0.500	4.000	0.001
R-squared		0.750		
F-statistic		25.000		

Table 1: Résultats de l'estimation OLS

Les résultats de l'estimation montrent que :

- Le coefficient du *taux d'alphabétisation* est significatif, indiquant que son augmentation influence positivement le PIB.
- Les *dépenses d'investissement* ont également un effet positif sur le PIB, ce qui souligne l'importance de l'investissement public.
- La valeur de R^2 suggère que le modèle explique une grande partie de la variance du PIB.

4 Tests de diagnostic pour vérifier la validité du modèle

Après avoir effectuer la régression, appui sur le bouton "name" pour nommer l'équation.

4.1 Test de normalité des résidus (Test de Jarque-Bera)

Un test de normalité des résidus permet de vérifier si les erreurs de prévision de notre modèle de régression se comportent comme on s'y attend. Si ces erreurs ne suivent pas une distribution normale, cela peut signaler des problèmes, comme des choix incorrects dans le modèle ou des éléments importants oubliés. Lorsque les erreurs ne sont pas normales, les résultats que nous obtenons, comme les estimations des coefficients, peuvent être biaisés, ce qui remet en question la fiabilité de nos conclusions.

Le test de Jarque-Bera permet de vérifier si les résidus suivent une distribution normale, ce qui est une hypothèse clé pour la validité des tests t et F.

Commande :

```
nom_de_l'equation.hist
```

Interprétation :

- Si la p-value du test est supérieure à 0.05, l'hypothèse de normalité des résidus est acceptée (les résidus sont normalement distribués).
- Si la p-value est inférieure à 0.05, cela signifie que les résidus ne sont pas normalement distribués.

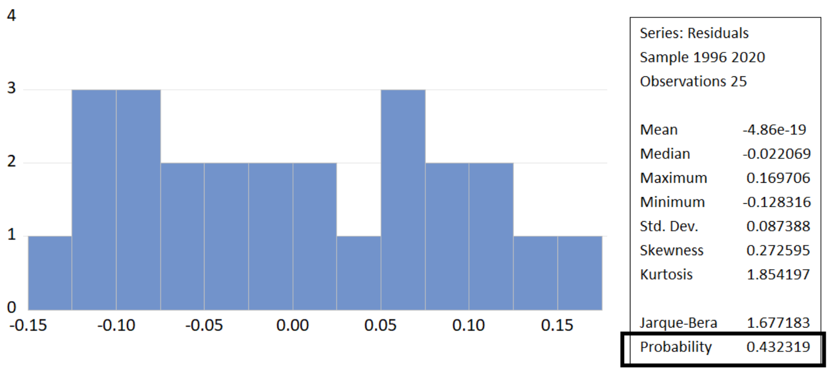

Figure 4: Illustration du test de Jarque-Bera

4.2 Test d'hétéroscédasticité

L'hétéroscédasticité se produit lorsque les erreurs d'un modèle de régression (les écarts entre les valeurs prédites et réelles) varient en fonction des variables utilisées. En d'autres termes, les erreurs peuvent être plus grandes ou plus petites selon les valeurs des variables, ce qui peut rendre les résultats moins précis et moins fiables.

La commande est :

```
nom_de_l_equation.hettest
```

Après avoir entré cette commande, vous pouvez choisir le type de test que vous préférez, Recommandation: test de Breusch-Pagan-Godfrey (n'oubliez pas d'insérer les regressors), on va se concentrer sur la partie supérieure de la fenêtre, notamment "Prob. Chi Square 3". Figure 5

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	0.890344	Prob. F(7,21)	0.5313
Obs*R-squared	6.636939	Prob. Chi-Square(7)	0.4676
Scaled explained SS	3.580957	Prob. Chi-Square(7)	0.8266

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 09/27/24 Time: 21:04
Sample: 1991 2020
Included observations: 29

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	369.8004	296.1936	1.248509	0.2256
IPC(-1)	0.433875	0.747592	0.580363	0.5678
LIMP	8.396269	34.02666	0.246756	0.8075
LPIB	-104.7017	94.87473	-1.103578	0.2823
LPIB(-1)	32.49427	85.09113	0.381876	0.7064
LPPT	38.66163	24.69150	1.565787	0.1323
LTCH	14.57912	80.28697	0.181588	0.8576
LTCH(-1)	-29.76041	74.13430	-0.401439	0.6922

R-squared	0.228860	Mean dependent var	20.82196
Adjusted R-squared	-0.028187	S.D. dependent var	29.01667
S.E. of regression	29.42277	Akaike info criterion	9.830366
Sum squared resid	18179.69	Schwarz criterion	10.20755
Log likelihood	-134.5403	Hannan-Quinn criter.	9.948495
F-statistic	0.890344	Durbin-Watson stat	1.229338
Prob(F-statistic)	0.531317		

Figure 5: Illustration test d'heteroscedasticité

- Si la p-value est supérieure à 0,05 : L'hypothèse d'homoscédasticité est acceptée, donc il n'y a pas de problème d'hétéroscédasticité.
- Si la p-value est inférieure à 0,05 : Cela signifie qu'il y a de l'hétéroscédasticité, ce qui pourrait affecter la fiabilité des résultats.

4.3 Test d'autocorrélation des résidus

Le test de Breusch-Godfrey vérifie la présence d'autocorrélation dans les résidus.

Commande :

```
nom_de_l'equation.auto
```

En laissant le retard à deux (2) par défaut.

Breusch-Godfrey Serial Correlation LM Test:
Null hypothesis: No serial correlation at up to 2 lags

F-statistic	0.160378	Prob. F(2,20)	0.8529
Obs*R-squared	0.457754	Prob. Chi-Square(2)	0.7954

Test Equation:
Dependent Variable: RESID
Method: ARDL
Date: 09/27/24 Time: 21:07
Sample: 1991 2020
Included observations: 29
Presample and interior missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
IPC(-1)	-0.010640	0.161741	-0.065786	0.9482
LIMP	0.743086	4.706409	0.157888	0.8761
LPIB	-3.046524	20.01519	-0.152211	0.8805
LPIB(-1)	1.151782	17.92969	0.064239	0.9494
LPPT	0.592062	4.583314	0.129178	0.8985
LTCH	0.111905	16.36660	0.006837	0.9946
LTCH(-1)	-0.751894	15.08784	-0.049834	0.9607
RESID(-1)	0.113822	0.289380	0.393329	0.6982
RESID(-2)	-0.126451	0.266804	-0.473947	0.6407

R-squared	0.015783	Mean dependent var	-0.005870
Adjusted R-squared	-0.377904	S.D. dependent var	4.643874
S.E. of regression	5.451172	Akaike info criterion	6.478664
Sum squared resid	594.3055	Schwarz criterion	6.902998
Log likelihood	-84.94063	Hannan-Quinn criter.	6.611560
Durbin-Watson stat	2.101387		

Figure 6: Illustartion test d'autocorrélation

- Si la p-value est supérieure à 0.05, il n'y a pas d'autocorrélation des résidus.
- Si la p-value est inférieure à 0.05, cela signifie qu'il y a une autocorrélation des résidus.

4.4 Test de multicolinéarité (VIF - Variance Inflation Factor)

La multicolinéarité se produit lorsque deux ou plusieurs variables explicatives dans un modèle de régression sont fortement corrélées. Il devient difficile de déterminer l'effet de chaque variable sur le résultat. Les résultats peuvent être peu fiables car les coefficients changent avec de petites variations dans les données. Par exemple, Si vous étudiez l'impact de la taille et de la taille de chaussures sur le poids, et que ces deux variables sont corrélées, il est compliqué de savoir laquelle influence le poids.

Commande :

```
nom_de_l'equation.varinf
```

Interprétation :

- Un VIF centré (centered vif) supérieur à 5 suggère une forte multicolinéarité entre les variables explicatives. De ce fait il faut supprimé la variable explicative qui est corrélée.

Variance Inflation Factors
Date: 09/27/24 Time: 21:10
Sample: 1990 2020
Included observations: 29

Variable	Coefficient Variance	Uncentered VIF
IPC(-1)	0.016938	4.251959
LIMP	19.04016	9055.363
LPIB	285.3255	10029.73
LPIB(-1)	208.7351	7326.776
LPPT	18.37575	5.856226
LTCH	204.0365	11603.41
LTCH(-1)	173.0725	9621.590

Figure 7: Illustration du VIF

4.5 Analyse de la stabilité : Test CUSUM

Le test CUSUM (Cumulative Sum of Recursive Residuals) est utilisé pour vérifier la stabilité des coefficients du modèle.

Commande pour exécuter le test CUSUM :

```
nom_de_l'equation.rls
```

Choisir CUSUM TEST et CUSUM SQUARE TEST. Interprétation du graphique CUSUM :

- Si la courbe CUSUM reste à l'intérieur des bornes, cela suggère que le modèle est stable dans le temps.
- Si la courbe dépasse les bornes, cela indique une instabilité des coefficients.

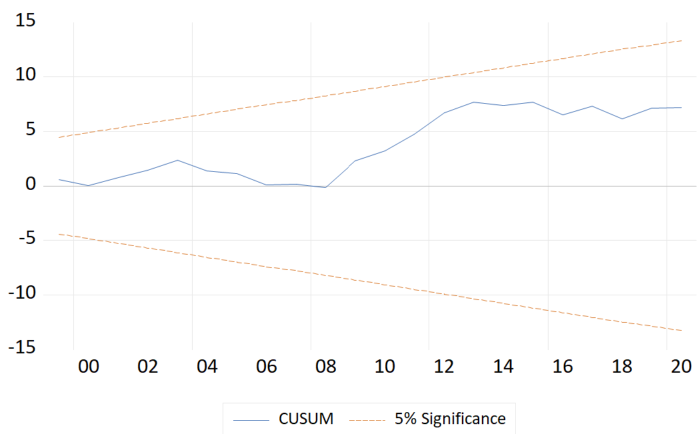

Figure 8: Illustration cusum test

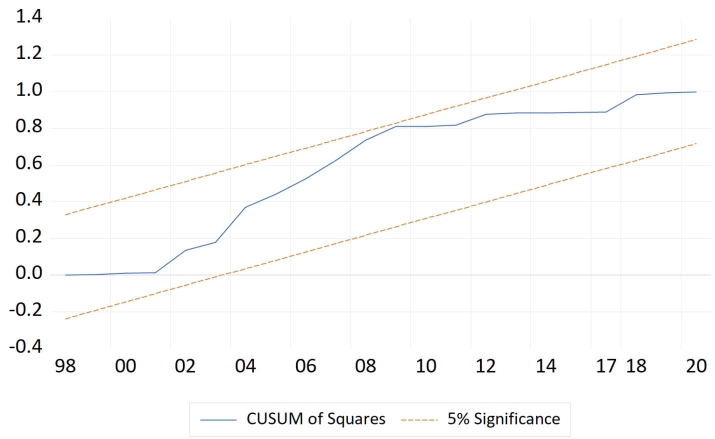

Figure 9: Illustration cusum square

5 Prédiction avec le modèle OLS

Une fois que vous avez validé votre modèle, vous pouvez l'utiliser pour faire des prévisions.

Commande pour prédire des valeurs futures :

```
nom_de_l'equation.forecast y_f
```

Cela génère une nouvelle série de données y_f, qui contient les valeurs prédites par le modèle OLS.

6 Résumé des commandes EViews

- **Estimation du modèle OLS**
  - Modèle de régression simple :

    `ls y c x1`
  - Modèle de régression multiple :

    `ls y c x1 x2 x3`
- **Tests de diagnostic**
  - Test de normalité des résidus (Test de Jarque-Bera) :

    `nom_de_l'equation.hist`
  - Test d'hétéroscédasticité :

    `nom_de_l_equation.hettest`
  - Test d'autocorrélation des résidus :

    `nom_de_l'equation.auto`
  - Test de multicolinéarité (VIF - Variance Inflation Factor) :

    `nom_de_l'equation.varinf`
- **Analyse de la stabilité : Test CUSUM**

  `nom_de_l'equation.rls`

References

[1] Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

[2] Sara, van, de Geer. (2014). *Least Squares Estimation.* doi: 10.1002/0470013192.BSA199

[3] Gujarati, D. N. and Porter, D. C. (2009). *Basic Econometrics.* 5th Edition, McGraw Hill Inc., New York.

[4] Klein, L. R. (1947). The Use of Econometric Models as a Guide to Economic Policy. *Econometrica*, 15(2), 111-151.

[5] Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33-50.

[6] Murray, M. P. (2005). *Econometrics: A Modern Introduction*. Pearson Addison-Wesley.

[7] Phillips, P. C. B. (1988). Reflections on Econometric Methodology. *The Economic Record*, 64(4), 344-359.

[8] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.