
**Analyse temporelle du taux de mortalité
des moins de cinq ans en fonction du PIB
par habitant et du taux de fertilité au
Népal pendant la période 1960-2020**

**Une approche par la méthodologie ARIMA de
Box-Jenkins**

Master Econométrie, Statistiques
Parcours Econométrie Appliquée

Avril 2024

Yava Vilar Valera

Dorota Girard

SOMMAIRE

1.	Introduction	2
2.	Analyse economique du sujet.....	3
3.	Présentation de la méthodologie ARIMA	10
4.	Choix du modèle ARIMA.....	17
5.	Cointégration d'Engel-Granger.....	31
6.	Conclusion	37
7.	Discussion	38
8.	Bibliographie	40
9.	Annexes.....	42
10	Table des matieres	49

1. INTRODUCTION

La mortalité infantile, en particulier chez les enfants de moins de cinq ans, est un indicateur crucial de la qualité de santé et du développement d'un territoire. Dans le contexte du Népal, pays qui a connu des transformations significatives dans sa structure socio-économique et démographique au cours des dernières décennies, comprendre les tendances et les facteurs associés à la mortalité infantile est d'une grande importance pour aider à mettre en place des politiques et des programmes de santé efficaces.

Ce dossier se concentre sur l'analyse des séries chronologiques de mortalité des moins de cinq ans au Népal sur les années 1960 à 2020, en explorant sa relation avec deux variables clés : le taux de fertilité et le produit intérieur brut (PIB), en utilisant les données fournies par la Banque Mondiale. Le taux de fertilité, qui a connu des changements remarquables en raison de divers facteurs tels que des politiques de planification familiale et les changements d'attitudes socioculturelles, est essentiel pour comprendre la dynamique démographique du pays. D'autre part, le PIB par habitant reflète la croissance économique et le niveau de richesse moyen des habitants d'un pays, influençant la disponibilité des ressources et des services de santé qui ont un impact direct sur la mortalité infantile.

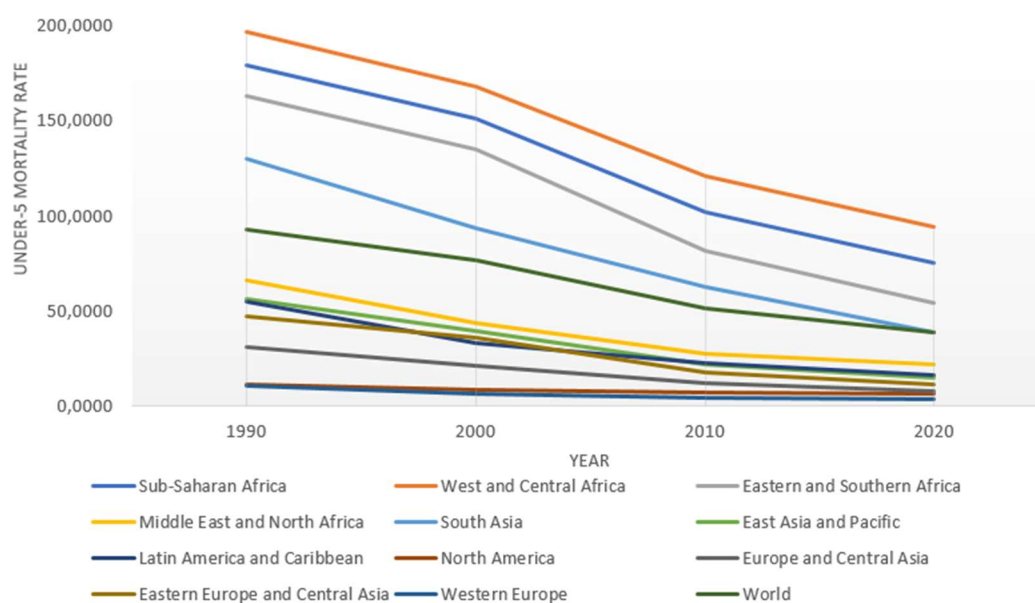
Le travail se déroule de la manière suivante : dans un premier temps, les tendances des trois séries mentionnées ainsi que leur relation économique seront étudiées. Deuxièmement, la méthodologie ARIMA de Box-Jenkins sera présentée pour ensuite être appliquée au cas particulier de cette étude. Nous réaliserons des prévisions sur une période avec différents modèles optimaux, en comparant la valeur prédite à la réelle ayant l'objectif de mesurer la qualité des modèles. Finalement, le test de cointégration d'Engel Granger sera appliqué pour évaluer la présence des relations stables à long-terme entre les séries analysées.

2. ANALYSE ECONOMIQUE DU SUJET

1.1 MORTALITE DES ENFANTS DE MOINS DE 5 ANS

Même si des améliorations considérables ont été réalisées au cours des dernières décennies, le taux de mortalité des moins de cinq ans reste un problème social mondial. Le terme fait référence à la probabilité pour un enfant de décéder avant l'âge de cinq ans, généralement mesurée sur un taux de 1 000 habitants. En 2020, le monde entier a enregistré un taux de 42,8% et un nombre total de 5,186 millions de décès d'enfants en dessous de cet âge, selon les données fournies par l'UNICEF¹. Ces chiffres mettent en évidence l'ampleur du problème et les efforts qu'il reste à fournir. Cependant, tous les pays et régions géographiques ne sont pas confrontés au problème avec la même intensité. Des facteurs tels que la maladie des enfants, l'accès limité aux soins de santé, le manque de vaccination, l'état de santé de la mère et les soins dispensés lors de l'accouchement, qui sont étroitement liés à la mortalité infantile, sont plus courants dans les pays à faible revenu que dans les pays à revenu élevé.

Figure 1: Taux de mortalité des moins de 5 ans de 1990-2020 par région du monde

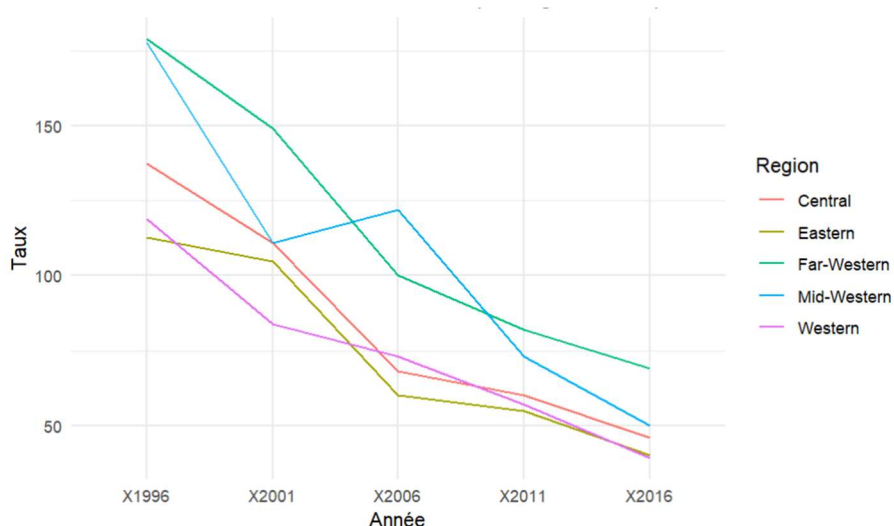


Source : Propre élaboration, données de l'UNICEF

¹ <https://data.unicef.org/topic/child-survival/under-five-mortality/> Consulté le 20/03/2024 à 00 :55

Dans la figure 1, des données ont été collectées pour analyser l'évolution du taux de mortalité des moins de cinq ans dans toutes les régions existantes du monde, sur la période de 1990 à 2020. Nous pouvons facilement comprendre que les groupes à revenus faibles et moyens se situent au-dessus de la courbe mondiale, tandis que les groupes plus avancés restent en-dessous. En particulier, l'Afrique occidentale et centrale, l'Afrique subsaharienne, orientale et australe, suivies par l'Asie du Sud, présentent les taux les plus élevés, contribuant ainsi à un taux moyen mondial plus élevé. Au contraire, l'ensemble de l'Europe, l'Asie centrale et l'Amérique du Nord affichent des valeurs bien inférieures, sans atteindre 50 % sur aucune des périodes étudiées. Par conséquent, on peut interpréter qu'une raison principale expliquant cet écart repose sur les différences de niveau de pauvreté et de développement. C'est la raison pour laquelle le Népal semble être un pays intéressant à analyser, étant donné que c'est est l'un des pays les plus pauvres au monde. Toutefois, ce pays a réalisé des progrès remarquables au cours des dernières décennies. La figure 2 montre comment les chiffres du taux de mortalité ont constamment diminué et se sont améliorés dans l'ensemble des cinq régions qui conformement le pays de 1996 à 2016, à l'exception de la région du Centre-Ouest, où le taux a augmenté pendant les années 2001-2006.

Figure 2 : Taux de mortalité des moins de 5 an par région au Népal, 1996-2016



Source : Propre élaboration, données DHS², R

² <https://dhsprogram.com/Countries/> Consulté le 21/03/2024 à 15 :30

1.2 PRODUIT INTERIEUR BRUT (PIB) PAR HABITANT

Le PIB (Produit Intérieur Brut) par habitant est un indicateur qui permet d'évaluer le niveau de richesse économique moyenne des habitants d'une société. L'inclusion de cette mesure dans l'étude se justifie par l'importance de son poids dans l'explication du phénomène auquel on s'intéresse. L'article « Determinants of under-5 mortality among the poor and the rich: a cross-national analysis of 43 developing countries » (Tanja AJ Houweling, 2005) ³ corrobore à travers de recherches empiriques l'hypothèse selon laquelle le niveau de développement économique des nations et de richesse des ménages impacte la mortalité des enfants de moins de cinq ans dans les pays en voie de développement. Similairement, l'étude nommée « Income and child mortality in developing countries : a systematic review and meta-analysis » (Bernadette O'Hare, 2013)⁴ a trouvé que l'augmentation du revenu national de 10% réduit la mortalité à laquelle on fait référence de 45%, ce chiffre diminuant à 28% après avoir rajouté des variables de contrôle.

En effet, plus les habitants d'un pays profitent d'une capacité économique élevée, plus on s'attend à ce que les agents économiques puissent investir dans l'amélioration de différents domaines qui ont un impact direct sur la mortalité des enfants. Par exemple, le domaine sanitaire, via le financement de vaccins ou d'innovations visant la prévention et guérison de maladies ; la construction de centres de santé ou encore la formation de professionnels qualifiés. Il est estimé que 15,8% des morts des moins des cinq ans dans les pays à revenu faible et intermédiaire proviennent des naissances prématurées, souvent liées à des problèmes de santé de la mère, tandis qu'une bonne partie du 84,2% restant est expliqué par des maladies de l'enfant telles que la pneumonie (13%), diarrhée (8,7%) ou le malaria (5,7%) (Robert E.Black, 2016) ⁵. De même, l'insécurité alimentaire semble être un facteur déterminant, tandis que la quantité et la qualité de l'alimentation peut s'améliorer lorsque les ménages ont des ressources financières majeures et le pays une capacité de production plus grande. Cet

³ <https://academic.oup.com/ije/article/34/6/1257/707470> Consulté le 21/03/2024 à 16:05

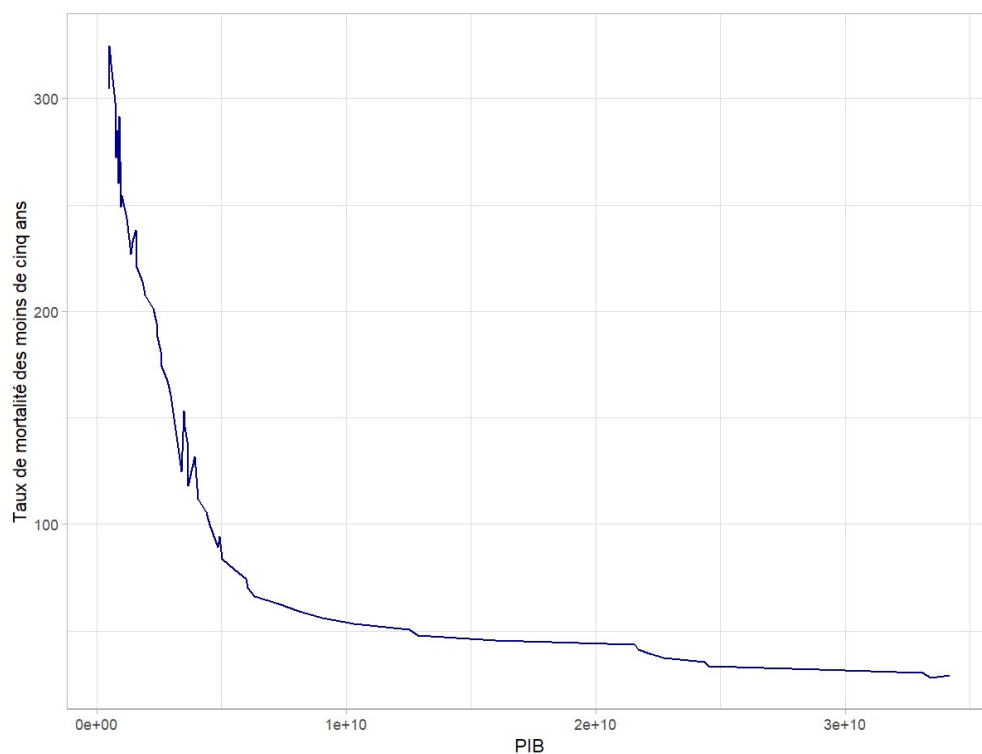
⁴ <https://journals.sagepub.com/doi/full/10.1177/0141076813489680>. Consulté le 21/03/2024 à 16 :30

⁵ <https://pubmed.ncbi.nlm.nih.gov/27227235/>. Consulté le 21/03/2024 à 17 :04

enchaînement de liens a été étudié par l'économiste R. Easterlin à travers des analyses historiques (Easterlin, 1996)⁶.

Le Népal a connu une forte augmentation continue de son Produit Intérieur Brut par habitant au cours des dernières six décennies, fait qui explique qu'on trouve une corrélation négative très importante entre cette variable et la variable d'intérêt, puisque comme il a déjà été indiqué, la mortalité infantile n'a cessé de diminuer au niveau national pendant cette même période. La figure 3 permet d'illustrer cette corrélation.

Figure 3 : Taux de mortalité des moins de 5 ans en fonction du PIBpc, Népal



Source : Propre élaboration, données de la Banque Mondiale⁷, R

⁶ <https://www.jstor.org/stable/10.3998/mpub.14163>. Consulté le 21/03/2024 à 17 :13

⁷ <https://datos.bancomundial.org/pais/nepal?view=chart>. Consulté le 23/03/2024 à 18 :58

1.3 TAUX DE FERTILITE

Le taux de fertilité se définit comme la proportion de naissances vivantes par rapport au total de la population féminine en âge de procréer, qui se situe entre les 15 et 60 ans, sur une année particulière (INSEE, 2021)⁸. Il s'agit donc d'un indicateur démographique permettant de calculer et mesurer des tendances des naissances d'une population donnée.

En accord avec le livre « Population in the Human Sciences: Concepts, Models, Evidence » (Philip Kreager, 2015)⁹, un taux de fertilité important réduit le temps et l'énergie qui peut se consacrer à chaque progéniture individuelle, augmentant ainsi leur probabilité de décès. Le rapport « Factors associated with trends in infant and child mortality in developing countries during the 1990s » (Rutstein, 2000)¹⁰, étudie les causes relatives à la diminution de la mortalité infantile dans la décennie des années 1990s dans les pays en développement à partir des données DHS (Demographic and Health Surveys), et trouve un lien significatif entre la variable référée et la dynamique de la fertilité. Un résultat similaire est obtenu dans l'article « Socio-economic, macroeconomic, demographic, and environmental variables as determinants of child mortality in South Asia » (Husnain, 2019)¹¹ où les auteurs affirment que la diminution continue du taux de fécondité en Asie du Sud est la cause principale de la réduction que la région a connue en termes de décès des enfants à partir des années 1973.

Au Népal, il existe également de l'évidence qui montre la corrélation entre ces deux variables. L'article « Effect of birth spacing on infant and child mortality in rural Nepal » (Gubhaju, 1986)¹² confirme que le taux de fertilité affecte principalement le nombre de décès pro génitaux, et que cette mortalité est plus fréquente chez les femmes qui ont leur premier enfant. De même, il existe une relation entre l'âge de la mère et le nombre de nouveau-nés décédés.

⁸ <https://www.insee.fr/fr/metadonnees/definition/c1872>. Consulté le 23/03/2024 à 19 :20

⁹ <https://academic.oup.com/book/8256>. Consulté le 23/03/2024 à 19:35

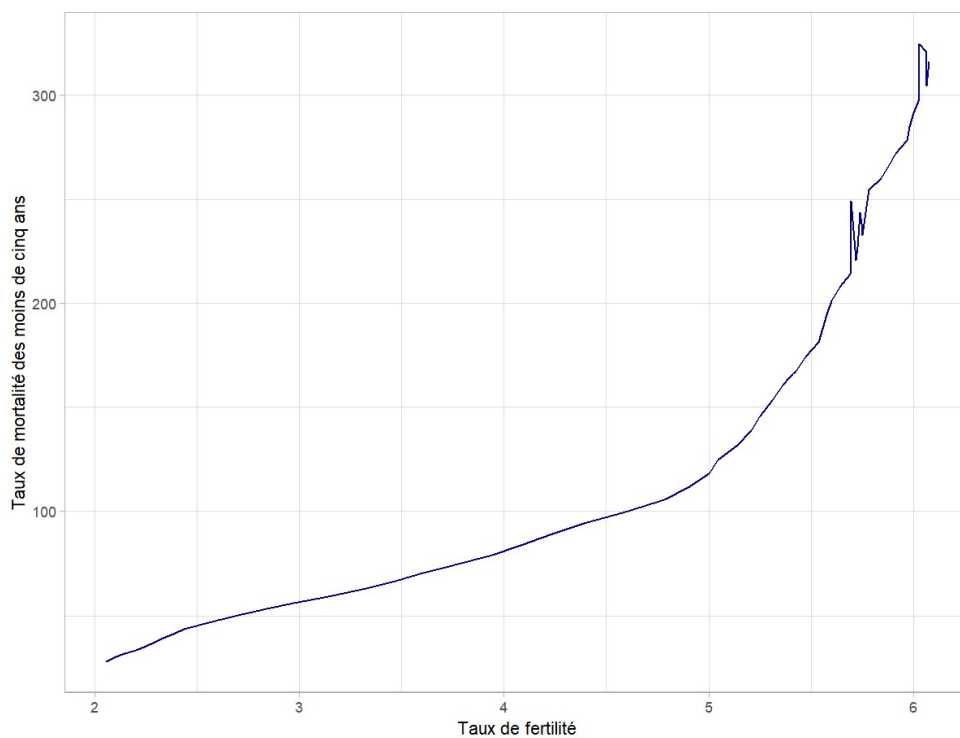
¹⁰ <https://pubmed.ncbi.nlm.nih.gov/11100620/>. Consulté le 23/03/2024 à 19:42

¹¹ <https://pubmed.ncbi.nlm.nih.gov/31820247/>. Consulté le 23/03/2024 à 19:55

¹² <https://pubmed.ncbi.nlm.nih.gov/3782194/>. Consulté le 23/03/2024 à 20 :10

La figure 4 exhibe le taux de décès des moins de cinq ans en fonction du taux de fertilité au Népal. Le graphique a été effectué en utilisant des données concernant la période examinée dans cette étude, du 1960 au 2020. Nous pouvons observer, en effet, une corrélation positive et forte qui renforce et confirme la littérature économique.

Figure 4 : Mortalité des moins de cinq ans en fonction du taux de fertilité, Népal



Source : Propre élaboration, données de la Banque Mondiale¹³, R

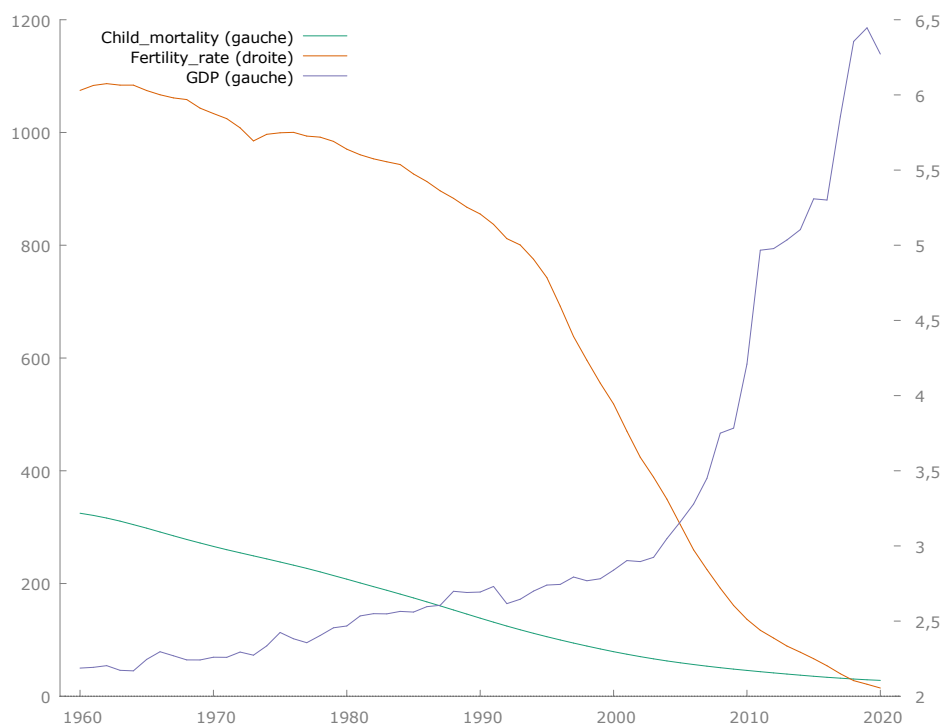
1.4 COMPARAISON DES TROIS SERIES

Sur la figure 5, nous pouvons observer la tendance des trois séries présentées pendant la période étudiée dans ce dossier de 1960 à 2020. L'augmentation graduelle du PIB par habitant jusqu'aux années 2005 peut s'expliquer par la mise en place du régime du Panchayat (1960-1990), où des réformes sociales et économiques ont été menées, telles que la modernisation du code juridique et les

¹³ <https://datos.bancomundial.org/pais/nepal?view=chart>. Consulté le 23/03/2024 à 18 :58

réformes agraires qui ont contribué à améliorer l'efficacité et la productivité. Viennent ensuite le rétablissement de la démocratie et la libéralisation économique depuis 1991. La forte croissance à partir des années 2005 coïncide avec la fin du conflit armé avec les maoïstes en 2006 et la stabilisation politique (Osmani, 2007)¹⁴. En ce qui concerne la dynamique décroissante du taux de fertilité, depuis la fin des années 1970 et le début des années 1980, des efforts intensifs ont été déployés pour élargir la disponibilité des services de planification familiale au Népal, à mesure que la sensibilisation à la planification familiale a augmenté et que la taille souhaitée de la famille a diminué (Dangol, 1997)¹⁵. L'évolution de ces deux variables a pu contribuer à la décroissance constante de la mortalité infantile à travers les mécanismes expliqués précédemment. Il convient de préciser, toutefois, qu'une corrélation n'implique pas une causalité, étant donné que d'autres facteurs peuvent influencer la relation observée entre les séries.

Figure 5 : Comparaisons des trois séries



Source : Propre élaboration, Gretl

¹⁴ <https://www.studocu.com/row/document/st-xaviers-college-nepal/english/the-economic-development-of-nepal-a-long-term-perspective/51946754>. Consulté le 23/03/2024 à 20 :30

¹⁵ <https://pubmed.ncbi.nlm.nih.gov/12292422/>. Consulté le 23/03/2024 à 20 :33

3. PRESENTATION DE LA METHODOLOGIE ARIMA

La méthodologie ARIMA fait référence à un modèle statistique permettant d'analyser des séries temporelles et de prévoir des valeurs futures en fonction des valeurs passées de la série concernée. Les modèles ARIMA ont été formalisés par Box et Jenkins en 1976, raison pour laquelle ces modèles sont également appelés « modèles Box-Jenkins ». Cette approche suppose que la série temporelle est générée par un processus stochastique (un ensemble de variables aléatoires indexées dans le temps), dont sa nature peut être caractérisée par un modèle mathématique. Il s'agit d'une méthode de modélisation des séries non stationnaires, dont leur comportement est expliqué à partir de valeurs passées de cette même série et/ou des erreurs de prévision passées.

3.1. STATIONNARITE

La stationnarité, dans le contexte d'un processus stochastique, est une propriété qui suppose certaines hypothèses sous-jacentes des séries temporelles, qui doivent être respectées pour pouvoir les modéliser. Lorsque les séries brutes ont des caractéristiques qui suggèrent la présence de stationnarité, il s'agira des modèles ARMA. En revanche, lorsque des transformations de la série originale, telles que des différenciations, nécessitent d'être effectuées pour vérifier la condition de stationnarité, on parlera des modèles ARIMA, où I, avec son paramètre d , capture le nombre de différenciations nécessaires pour satisfaire l'hypothèse de stationnarité. Les séries stationnaires sont dotées des propriétés suivantes :

- 1) Espérance des moyennes constante :

Cela indique que la moyenne de la série est invariante au temps, et ne dépend que d'une seule constante. Ainsi, la série reviendra vers la même valeur tout au long de l'intervalle de temps concerné. Mathématiquement, cette hypothèse peut être représentée de la forme :

$$H_0 : \mu_t = \mu, \forall_t$$

Où μ représente la moyenne de la série temporelle.

Lorsque l'on se retrouve avec des séries non stationnaires au niveau de la moyenne, on apprécie généralement une tendance à long terme qui indique une direction générale ou une évolution croissante ou décroissante des données au fil du temps. Des différenciations premières, qui consistent à soustraire chaque observation de la précédente, peut contribuer à rendre la série stationnaire en éliminant la tendance à long terme.

2) Variance constante :

La variance, ou autrement dit, les écarts par rapport à la moyenne sera constante et indépendante du temps. Cette condition s'exprime mathématiquement :

$$H_0: \sigma_T = \sigma, \forall t$$

Où σ indique la variance de la série.

Lorsque cette condition n'est pas vérifiée, des transformations logarithmiques ou exponentielles sur les données peuvent être visées afin de stabiliser la variance.

3) Autocovariances constantes :

Cela se traduit par l'hypothèse selon laquelle les autocovariances entre des couples des variables ou d'observations ne dépendent pas du temps, mais du nombre de périodes qui séparent les observations. Mathématiquement :

$$\text{Cov}(Y_T, Y_{T+K}) = \text{Cov}(Y_T + Y_{T+K+K'}) = \gamma_t$$

Où T représente un temps spécifique dans la série et $T + K$ ou $T + K + K'$ indique le même point de la série décalé de T périodes.

3.2. CORRELOGRAMME

Les corrélogrammes sont une représentation graphique des autocorrélations des observations en différents retards (k) de la série temporelle. Ils fournissent des informations sur la façon dont des valeurs à un moment donné se rapportent à d'autres valeurs d'instantanés passés. L'axe x du corrélogramme indique les retards temporels, tandis que l'axe y montre les valeurs d'autocorrélation, qui renseignent sur la force et la direction de la relation entre les observations à différents moments. A l'intérieur du corrélogramme, chaque barre représente l'autocorrélation de la

série à un retard donné. A titre d'exemple, pour le niveau de retard trois, le corrélogramme montre les relations entre l'observation actuelle (3) et celles qui ont eu lieu trois unités de temps avant. Plus la barre est élevée, plus l'autocorrélation est importante. D'autre part, les autocorrélations significatives, celles qui visuellement dépassent l'intervalle de confiance, servent de référence pour identifier la structure des modèles ARIMA et peuvent également indiquer des patrons saisonniers dans la série.

Le corrélogramme montre deux parties essentielles des modèles ARIMA, la FAC (Fonction d'Autocorrélation) et FACP (Fonction d'Autocorrélation Partielle). La structure de ces deux fonctions aide à déterminer si une série est stationnaire et à définir le nombre de paramètres dans AR et MA convenants.

3.2.1. Composante AR

La composante AR (Autoregressive en anglais) indique une dépendance des observations actuelles par rapport aux valeurs passées. C'est la partie Autorégressive du modèle. Le nombre d'observations précédentes qui influence la valeur actuelle est capté par l'ordre de la composante, noté sous le paramètre p. L'équation mathématique s'exprime de la façon suivante :

$$Z_t = \delta + a_t + \sum_{j=1}^p \varphi_j Z_{t-j}$$

Où :

z_t = l'observation de la série différenciée à la période t

δ = la constante

a_t = le terme d'erreur à la période t, qui représente la différence entre l'observation réelle et celle prédite par le modèle.

p_i = l'ordre de la composante AR

φ = coefficient associé aux terme AR

Le modèle Autorégressif doit vérifier la condition de stationnarité pour garantir que le processus reste stable au fil du temps. Les racines du polynôme caractéristique associées au nombre de paramètres doivent être en dehors du cercle unité pour que l'hypothèse de stationnarité soit respectée.

- Pour un nombre de paramètres égal à 1 ($p=1$), l'équation doit être vérifiée :

$$|\varphi_1| < 1$$

- Pour un nombre de paramètres égal à 2 ($p=2$), ces trois équations doivent tenir :

$$\varphi_2 < 1$$

$$\varphi_1 + \varphi_2 < 1$$

$$\varphi_2 - \varphi_1 < 1$$

3.2.2. Composante AM

La partie « MA » (Moyenne Mobile ou Moving Average en anglais) représente la composante moyenne mobile. Cette partie du modèle ARIMA capture la relation linéaire entre une observation dans une série chronologique et les valeurs passées du terme d'erreur dans la même série chronologique. Le terme d'erreur est supposé être un bruit blanc, un processus stochastique caractérisé par avoir une espérance mathématique nulle et une variance constante, contenant des observations indépendamment et identiquement distribuées. Cette condition implique donc une absence de corrélation au niveau des résidus. L'équation suivante reflète le processus de Moyenne Mobile :

$$z_t = u + a_t + \sum_{i=1}^q \theta_i - a_{t-i}$$

Où:

u = constante

a_t = le bruit blanc

θ_i = le coefficient à estimer qui représente la mesure dans laquelle l'erreur au temps $t-i$ influence la valeur observée actuelle

q = le paramètre qui détermine combien de termes d'erreur passés sont inclus dans le modèle pour prédire l'observation actuelle au fil du temps

Afin que le modèle ARIMA soit correctement spécifié, la condition d'inversibilité au sein du processus MA doit tenir, qui implique le respect des caractéristiques suivantes :

- Pour un nombre de paramètres égal à 1 ($q=1$) :

$$\theta_1 < 1$$

- Pour un nombre de paramètres égal à 2 ($p=2$) :

$$\theta_2 < 1$$

$$\theta_1 + \theta_2 < 1; \theta_2 - \theta_1 < 1$$

3.3. MODELE BOX-JEKINS

La méthodologie du modèle Box-Jenkins est une méthode largement utilisée dans le domaine de l'économétrie et des statistiques qui a été développée originellement par George Box et Gwilym Jenkins dans les années 1970. La méthodologie comprend trois étapes principales :

1) Identification du modèle :

À cette étape, le type de modèle ARIMA adapté à la série est identifié. Il s'agit de déterminer si la série est stationnaire ou non, et si une ou plusieurs différenciations sont nécessaires pour la rendre stationnaire. De plus, les paramètres optimaux d'AutoRégression (AR) et de Moyenne Mobile (MA) sont induits en analysant la

fonction d'autocorrélation (FAC) et la fonction d'autocorrélation partielle (FACP) de la série temporelle.

2) Estimation du modèle :

Une fois le modèle optimal identifié, nous estimons le modèle. Pour ce faire, il faut vérifier la significativité des coefficients obtenus ainsi que la condition de stationnarité pour le modèle d'AutoRégression et d'inversibilité pour le processus de Moyenne Mobile. Si l'une de ces trois conditions n'est pas respectée, il faudra revenir à la première étape d'identification jusqu'à ce que le bon modèle soit estimé.

3) Vérification du modèle :

Dans la dernière étape, l'adéquation du modèle ajusté aux données de séries chronologiques est évaluée. Les résidus du modèle sont examinés pour vérifier s'ils répondent aux hypothèses du modèle ARIMA, telles que la normalité, la stationnarité et l'absence de corrélation en série. Le corrélogramme d'autocorrélation (FAC) et d'autocorrélation partielle (FACP) des résidus est utilisé via le test de Bartlett pour confirmer la normalité en vérifiant que les résidus se situent à l'intérieur de l'intervalle de confiance, tandis que le test de Parmenteau est utilisé pour tester la condition de bruit blanc des résidus expliquée précédemment. Similairement à la deuxième étape, si toutes ces suppositions ne sont pas remplies, nous devons revenir au début jusqu'à ce qu'un modèle approprié soit identifié.

Enfin, si les trois étapes ont bien été surmontées, nous pouvons procéder à la prévision de valeurs futures en utilisant la structure du modèle ARIMA ainsi que les paramètres estimés.

3.4. MODELES MULTIPLICATIFS

Les modèles ARIMA multiplicatifs sont une extension des modèles ARIMA traditionnels qui permettent de capturer à la fois les tendances et la variabilité non saisonnières ainsi que les modèles saisonniers dans les données. Lorsqu'une série chronologique comporte des composantes saisonnières, cela signifie qu'il existe des modèles qui se répètent à intervalles de temps réguliers, tels que les saisons, les mois, les semaines ou les jours.

Ainsi, les modèles ARIMA sont composés d'une partie classique, représentée par ARIMA(p,q,d), et d'une partie saisonnière, ARIMA(P,Q,D)s où s indique le nombre de périodes dans la saisonnalité. Ensemble, ils forment le modèle ARIMA(p,q,d) (P,Q,D)s qui s'exprime mathématiquement comme suit :

$$Y_t = 1 - \varphi_1 B - \varphi_2 B^2 (1 - B^s) X_t = (1 - \varphi_1 B - \varphi_2 B^2) X_t - (1 - B^s) X_t$$

Dans cette équation, Y_t représente la valeur de la série temporelle au fil du temps que nous essayons de modéliser ou de prédire. $\varphi_1 B$ et $\varphi_2 B^2$ indiquent les composantes autorégressives d'ordre 1 et 2 respectivement, où φ est le paramètre d'autorégression pour le premier (φ_1) ou deuxième (φ_2) décalage, et B est l'opérateur de retard. Le terme $(1 - B^s)$ représente la composante saisonnière, où s est le nombre de périodes dans la saisonnalité. X_t est la valeur observée au fil du temps t. En ce qui concerne la deuxième équation, $(1 - \varphi_1 B - \varphi_2 B^2) X_t$ représente la partie non saisonnière du modèle ARIMA, qui capture la tendance et la variabilité non saisonnière de la série chronologique. En revanche, le terme $(1 - B^s) X_t$ montre la partie saisonnière du modèle ARIMA. C'est la partie qui modélise les fluctuations saisonnières des données.

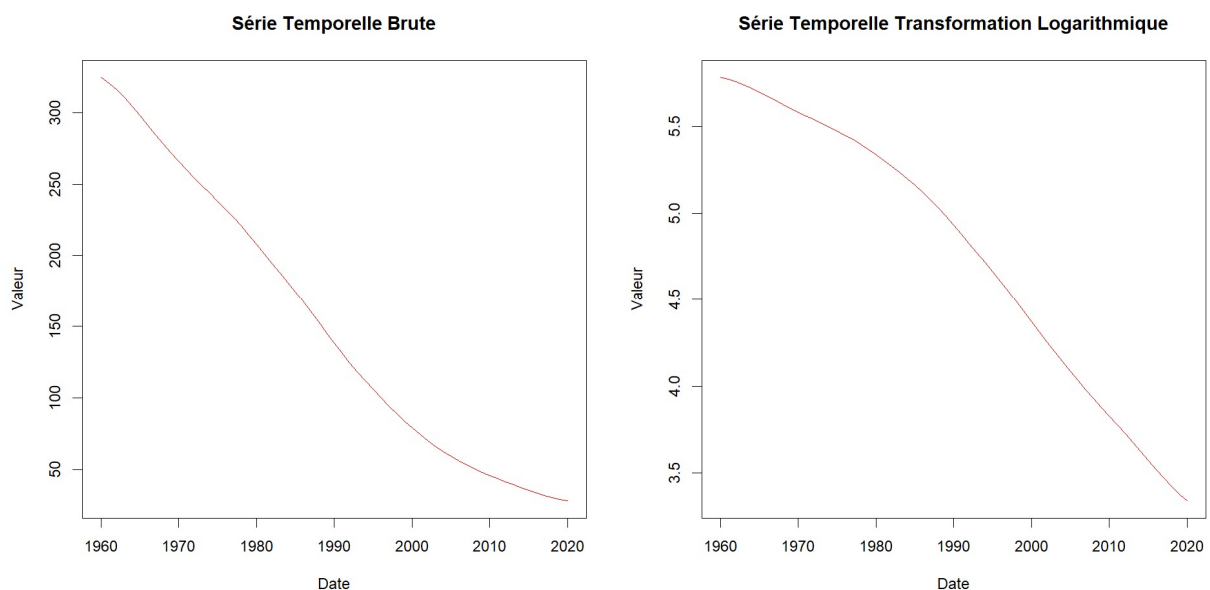
En résumé, l'équation complète représente la relation entre la valeur observée au fil du temps et les composantes autorégressives, saisonnières et tendanciennes de la série temporelle.

4. CHOIX DU MODELE ARIMA

4.1. SERIE, CORRELOGRAMME

Cette partie vise à choisir un modèle ARIMA optimal pour la série temporelle Y_t , taux de mortalité des moins de cinq ans, en appliquant la méthodologie de Box-Jenkins. A préciser que tous les modèles proposés dans ce dossier ne comportent pas de constante lorsque celle-ci n'est pas significative. La figure 5 montre, à gauche, les données de la série brute, dans laquelle une tendance de diminution est observée tout au long de la période concernée. Ce constat suggère un non-respect de la condition de stationnarité, étant donné que la moyenne n'a donc pas une valeur constante. A droite, une transformation logarithmique a été appliquée afin de comparer l'évolution de la série. Nous observons que cela a semblé d'apporter des légères modifications qui pourraient être considérées pour trouver des modèles de meilleure qualité. Toutefois, ils ne semblent pas être nécessaires pour la stationnarisation, car les logarithmes agissent sur la variance.

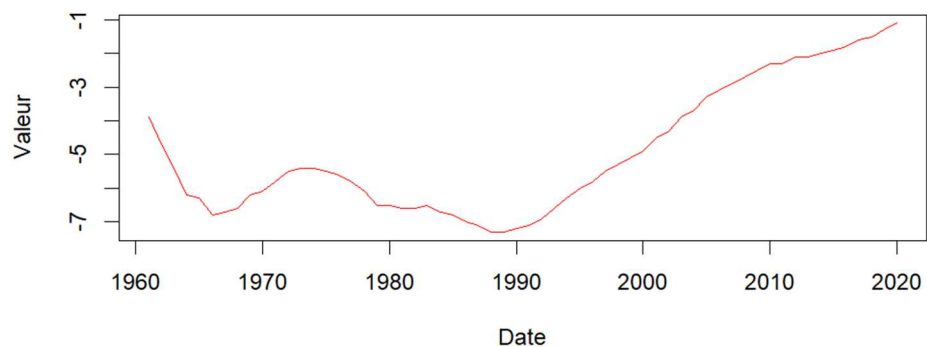
Figure 5 : Séries brute et logarithmique du taux de mortalité des moins de cinq ans



Source : Propre élaboration, R

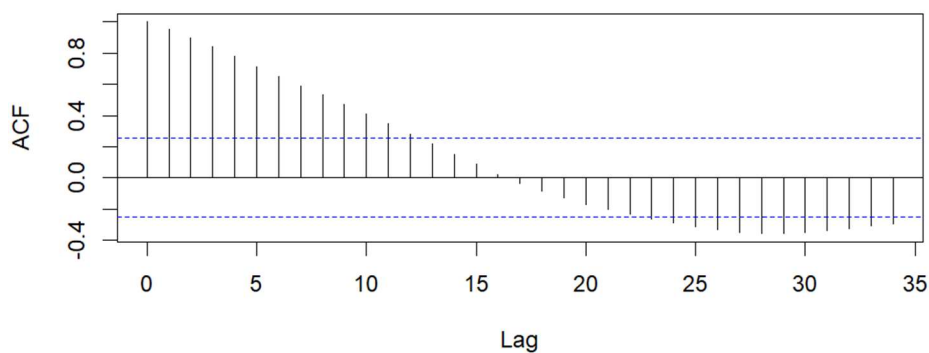
Afin d'agir sur la stationnarisation de la moyenne, nous pouvons appliquer une différenciation. Les figures 6, 7 et 8 montrent l'évolution graphique des données après avoir été différenciées une fois, ainsi que le corrélogramme de la FAC et FACP correspondant.

Figure 6 : Série de Yt différenciée d'ordre 1



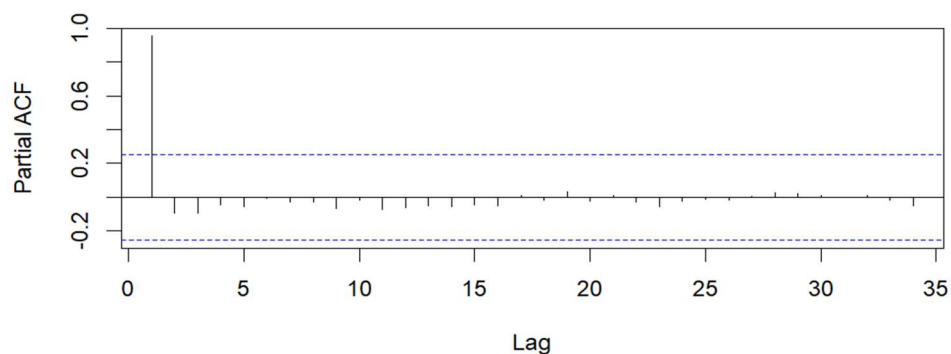
Source : Propre élaboration, R

Figure 7 : FAC de Yt différenciée d'ordre 1



Source : Propre élaboration, R

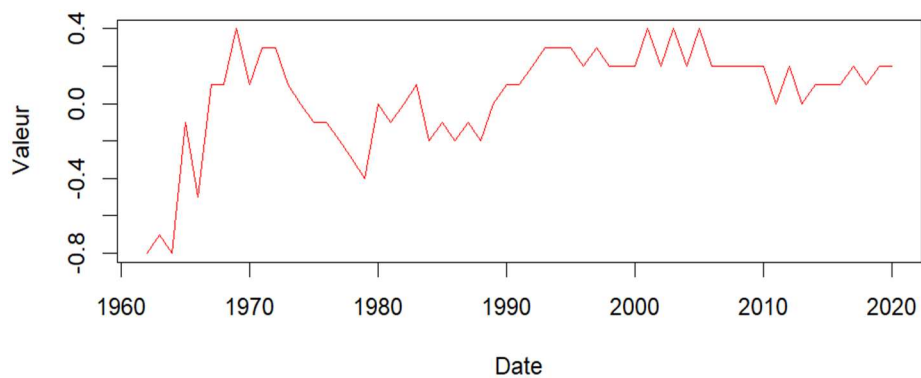
Figure 8 : FACP de Yt différenciée d'ordre 1



Source : Propre élaboration, R

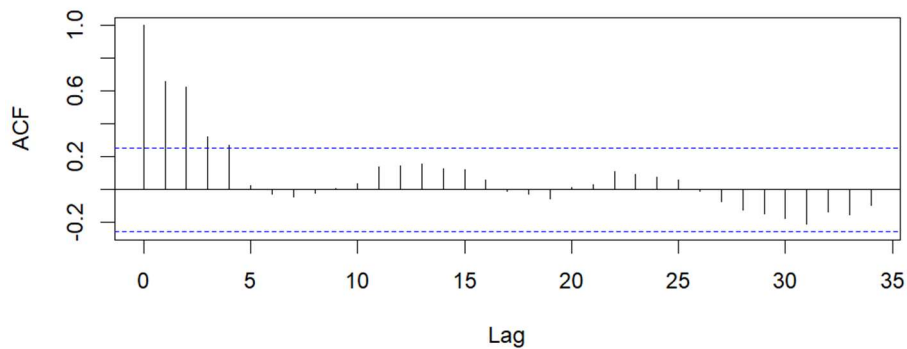
Nous pouvons apprécier que la série, avec ces caractéristiques, ne semble pas être stationnaire. En effet, son évolution ne met pas en preuve une moyenne et une variance constantes, de même que la FAC du corrélogramme exhibe des retards significatifs jusqu'au retard 14. Par conséquent, une deuxième différenciation a été menée, dont les nouvelles propriétés sont affichées sur les figures 9,10 et 11.

Figure 9 : Série de Yt différenciée d'ordre 2



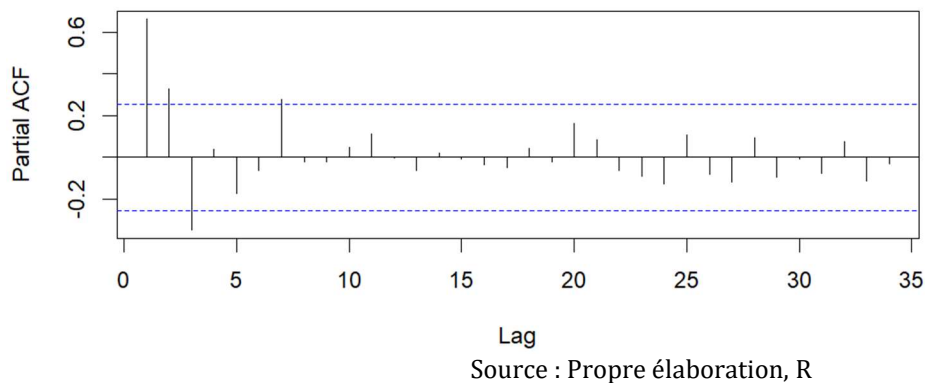
Source : Propre élaboration, R

Figure 10 : Série de Yt différenciée d'ordre 2



Source : Propre élaboration, R

Figure 11 : Série de Yt différenciée d'ordre 2



Malgré le fait que l'évolution graphique de tendance de la série montre encore des fluctuations au niveau de la moyenne et de la variance, le corrélogramme a une structure de décroissance rapide vers la non-significativité et propose des modèles ARIMA. Par conséquent, nous considérons cette série comme stationnaire.

4.2. IDENTIFICATION, ESTIMATION

Comme expliqué dans la méthodologie de Box-Jenkins, l'étape d'identification et estimation consiste à trouver une série stationnaire ainsi que le nombre de paramètres associés. Ayant identifié un nombre optimal de deux différenciations, il semblerait que la série se caractérise par avoir une partie classique, et non saisonnière. La partie classique s'explique par le fait que le corrélogramme a un premier retard significatif, tandis qu'il n'y a pas d'évidence de saisonnalité parce qu'on n'aperçoit pas de nouveaux pics significatifs au sein des intervalles de temps donnés. Le corrélogramme se caractérise par avoir quatre retards significatifs et positifs dans la FAC, ce qui peut suggérer la présence d'une partie MA. D'autre part, la FACP indiquant également une partie AR. Par conséquent, nous pourrions être face à des modèles simples ou mixtes. En commençant par la construction de modèles simples, $ARIMA(2,2,0)$ est proposé. Même s'il y a plus de pics significatifs au niveau de la FACP, on s'intéresse à un définir un modèle parcimonieux qui ne dépasse pas les deux paramètres au sein de chaque composante ARIMA. Les coefficients du modèle proposé respectent bien la condition de stationnarité. $ARIMA(1,2,0)$ est un deuxième candidat qui vérifie la stationnarité au niveau du

coefficient phi et qui est significatif. Le même arrive avec ARIMA (0,2,1) et ARIMA(0,2,2).

Le corrélogramme suggère également un potentiel modèle mixte, composé autant de paramètres AR que MA. Ainsi, nous proposons ARIMA (1,2,1) qui est parvenu à atteindre la condition de stationnarité et d'inversibilité, tandis que ARIMA (2,2, 1), et ARIMA (2,2,2) comportent des coefficients non significatifs. Finalement, ARIMA(1,2,2) ne vérifie pas la condition d'inversibilité. Ces caractéristiques emmènent à rejeter ces trois derniers modèles.

D'autre part, étant donné qu'un modèle simple et ayant subi peu de différenciations est privilégié, des modèles avec une seule différenciation ont été testés. En particulier, ARIMA (1,1,0) est un possible candidat, en raison du fait que le corrélogramme montre une rupture au premier retard de la FAC Partielle, tandis qu'un patron de croissance et décroissance continue au niveau de la FAC. Il a été modélisé autant avec une transformation logarithmique que sans elle. Le coefficient des deux modèles s'est avéré statistiquement significatifs et inférieurs à 1, même si très proches de ce chiffre. ARIMA (1,1,1) sans et avec des logs est également un modèle qui respecte toutes les hypothèses requises au niveau des coefficients. Par la suite, nous allons vérifier la condition de stationnarité au niveau des résidus des modèles qui ont satisfait les hypothèses des coefficients.

4.3. VERIFICATION

Après deux différenciations, il en résulte que le modèle ARIMA (2,2,0) comporte des résidus qui ne sont pas corrélés entre eux, de même que ARIMA(0,2,2). Avec une différenciation, ARIMA (1,1,1) sans des logarithmes ne confirmaient pas une absence de corrélation au niveau des résidus. Avec les logs, en revanche, les résidus au retard 24 dépassent légèrement le seuil de significativité de 5%. Les modèles restants proposés auparavant ne surmontent pas les tests de Portmanteau ni de Bartelet. Les figures de 12 à 17 et montrent les résultats des modèles retenus et le corrélogramme de leurs résidus pour une et deux différenciations respectivement. Des essais supplémentaires ont été réalisés avec deux différenciations et des logarithmes, dont le corrélogramme de cette série est disponible en annexes, figure

1. Seuls les modèles avec les mêmes paramètres que ceux sans des logarithmes montrent une structure stationnaire. Le Qstat au niveau du 24^{ème} retard est supérieur pour le modèle logarithmique ARIMA(2,2,0) mais pas pour ARIMA(0,2,2). Les modèles résultants et le corrélogramme des résidus sont également disponibles en annexes, figures 2 à 5.

Figure 12 : Modèle ARIMA (2,2,0)

Modèle 20: ARMA, utilisant les observations 1962-2020 (T = 59)
Estimation par AS 197 (MV exact)
Variable dépendante: d_d_Child_mortality
Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique	
phi_1	0,423056	0,109919	3,849	0,0001	***
phi_2	0,513246	0,114387	4,487	7,23e-06	***
Moyenne var. dép.	0,047458	Éc. type var. dép.	0,271875		
Moyenne innovations	0,021800	Éc. type innovations	0,159612		
R2	0,669759	R2 ajusté	0,663965		
Log de vraisemblance	23,53837	Critère d'Akaike	-41,07673		
Critère de Schwarz	-34,84412	Hannan-Quinn	-38,64377		

Source : Propre élaboration, Gretl

Figure 14 : Modèle ARIMA (0,2,2)

Modèle 21: ARMA, utilisant les observations 1962-2020 (T = 59)
Estimation par AS 197 (MV exact)
Variable dépendante: d_d_Child_mortality
Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique	
theta_1	0,603681	0,103577	5,828	5,60e-09	***
theta_2	0,640395	0,105453	6,073	1,26e-09	***
Moyenne var. dép.	0,047458	Éc. type var. dép.	0,271875		
Moyenne innovations	0,024200	Éc. type innovations	0,181284		
R2	0,581863	R2 ajusté	0,574527		
Log de vraisemblance	16,43586	Critère d'Akaike	-26,87171		
Critère de Schwarz	-20,63910	Hannan-Quinn	-24,43875		

Source : Propre élaboration, Gretl

Figure 16 : Modèle log_ARIMA (1,1,1)

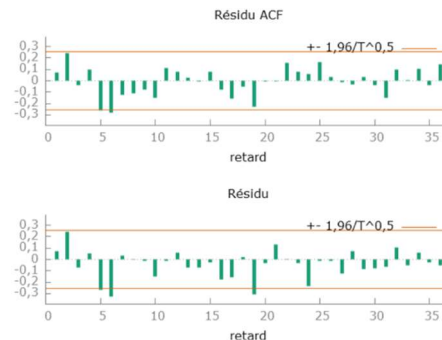
Modèle 2: ARMA, utilisant les observations 1961-2020 (T = 60)
Estimation par AS 197 (MV exact)
Variable dépendante: d_l_Child_mortality
Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique	
const	-0,0280264	0,0143024	-1,960	0,0500	*
phi_1	0,990362	0,0103594	95,60	0,0000	***
theta_1	0,277536	0,0920362	3,016	0,0026	***
Moyenne var. dép.	-0,040726	Éc. type var. dép.	0,014702		
Moyenne innovations	-0,000393	Éc. type innovations	0,001718		
R2	0,987291	R2 ajusté	0,987072		
Log de vraisemblance	294,5943	Critère d'Akaike	-581,1886		
Critère de Schwarz	-572,8112	Hannan-Quinn	-577,9118		

Source : Propre élaboration, Gretl

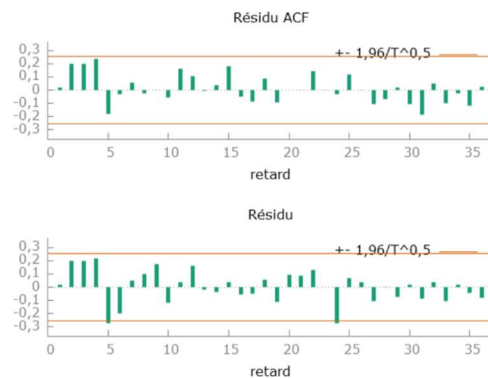
Figure 13 : corrélogramme des résidus

ARIMA(2,2,0)



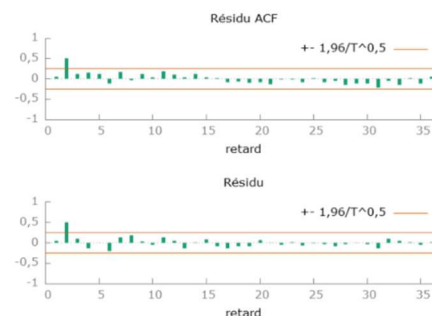
Source : Propre élaboration, Gretl

Figure 15 : Corrélogramme résidus ARIMA(0,2,2)



Source : Propre élaboration, Gretl

Figure 17 : Corrélogramme des résidus log_ARIMA(1,1,1)



Source : Propre élaboration, Gretl

Table 1: Comparaison des modèles pour Y_t

	Nombre de paramètres	Q stat(24) p-value	Ecart-type	AIC	R^2 ajusté	Erreur de Prévision
ARIMA (0,2,2)	2	0.51	0.1813	-26.87	0.58	0.15%
ARIMA(2,2,0)	2	0.0777	0.1596	-41.07	0.67	0.35%

La table 1 permet de comparer les deux modèles retenus sans logarithmes. A préciser que le taux d'erreur est calculé en comparant la valeur prédite et la vraie valeur observée pour 2021 qui est égale à 27.2%. Nous observons que malgré un R^2 ajusté plus faible et un AIC plus élevé, le modèle ARIMA(0,2,2) semble être le plus performant car il présente un Qstat bien plus important et une erreur de prévision plus faible. C'est donc le modèle qui sera utilisé pour calculer la prévision à la main.

Table 2 : Comparaison des modèles pour $\log(Y_t)$

	Nombre de paramètres	Q stat(24) p-value	Ecart-type	AIC	R^2 ajusté	Erreur de Prévision
logARIMA (2,2,0)	2	0.273	0.001812	-614.6657	0.525609	0.20%
logARIMA (1,1,1)	2	0.051	0.001237	-581,1886	0.987072	0.10%
logARIMA (0,2,2)	2	0,171	0.001413	-599,5591	0,385522	0.18%

En ce qui concerne la comparaison des modèles logarithmiques, il semblerait que le modèle qui donne l'erreur de prévision la plus faible est celui qui présente le Qstat plus petit (logARIMA(1,1,1)). De même, il a un R^2 ajusté assez élevé et le plus petit écart-type, mais un AIC moins faible comparé aux modèles restants. Malgré le fait que les modèles logarithmiques ne semblent pas être de meilleure qualité, un parmi les trois modèles gardés est le seul à avoir des résidus qui suivent la loi normale entre tout l'ensemble des modèles présentés dans les tables 1 et 2. C'est le logARIMA(0,2,2). Ce fait le rend préférable. La figure 6 en annexes met à disposition le test de normalité pour le modèle en question.

4.4. PREVISION A UNE PERIODE Y_T

Nous allons à présent nous intéresser au calcul de la prévision à la main pour le modèle ARIMA (0,2,2) et son taux d'erreur afin de comparer et vérifier les résultats fournis par le logiciel Gretl.

La prévision pour z_{t+1} s'exprime : $z_{t+1} = \theta_1 a_t + \theta_2 a_{t-1} + a_{t+1}$

Où z est la valeur au temps t de la série différenciée, θ est la composante moyenne mobile et a le terme d'erreur, où a_{t+1} est supposé être égale à 0.

Ainsi, nous pouvons remplacer l'équation théorique par les vraies valeurs :

$$z_{t+1} = 0.603681 * 0.110537 + 0.640395 * 0.114727$$

$$z_{t+1} = 0.1402$$

La valeur résultante correspond effectivement à celle prédite par Gretl. Ensuite, pour calculer l'erreur de prévision, on débute par calculer Y_{t+1} , qui est obtenue à partir de la suivante chaîne d'équations :

$$Y_t \Delta^2 = \Delta Y_t - \Delta_{t-1}$$

$$Y_t \Delta^2 = Y_t - Y_{t-1} + Y_{t-2}$$

$$Z_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

$$Z_t(1) = Y_{t+1} - 2Y_t + Y_{t-1}$$

$$Y_{t+1} = Z_t(1) + 2(Y_t) - Y_{t-1}$$

Tout ce processus a pour objectif la transformation et l'inversion d'une série temporelle qui a été différenciée deux fois pour calculer la prévision à l'instant Y_{t+1} . Le paramètre Δ représente le fait de différencier la série. Si on remplace l'équation finale par les vraies valeurs :

$$27.2402 = 0.1402 + 2 * 28.2 - 29.3$$

Et on compare à la vraie valeur de 2021 (27.2) pour obtenir l'erreur de prévision :

$$\frac{27.24702}{27.2} - 1 * 100 = 0.15\%$$

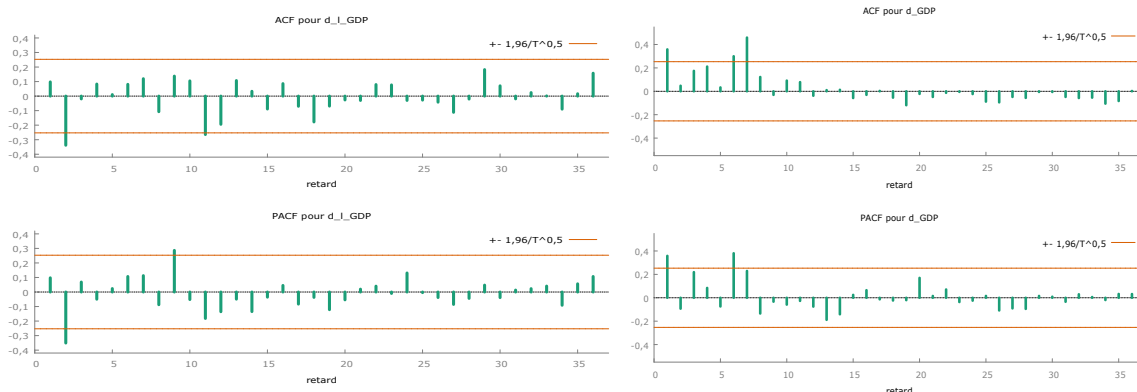
Nous pouvons conclure que l'erreur de prévision est faible et que le modèle est parvenu à prédire une valeur proche de la valeur réelle, ce qui confirme une bonne qualité du modèle.

4.5. MODELE ARIMA POUR LES SERIES XT

4.5.1. PIB par habitant

Il convient à présent d'étudier la série chronologique du PIB par habitant et d'estimer un modèle ARIMA qui s'ajuste aux données. Sur la partie analyse économique du sujet, et en particulier, le graphique 5, nous avons pu constater la présence d'une tendance ainsi que des fluctuations de la variance au sein de cette variable. Cela suggère le non-respect de la stationnarité. Par conséquent, une différenciation a été appliquée. Il en résulte qu'après cette procédure, le corrélogramme montre une décroissance rapide vers la non-significativité, ce qui laisse penser que la série est devenue stationnaire. Une transformation logarithmique a été également réalisée avec l'intention de comparer et de définir des modèles qui pourraient être plus performants. Le graphique de l'annexe 6 laisse penser que l'ajout des logarithmes parvient à dessiner une série temporelle qui se ressemble plus à une série stationnaire. Les figures 10 et 11 affichent les corrélogrammes après une différenciation avec et sans des valeurs logarithmiques respectivement.

Figure 10 : Corrélogramme log_PIBpc, ordre 1 **Figure 11 : Corrélogramme PIBpc, ordre 1**



Source : Propre élaboration, Gretl

Source : Propre élaboration, Gretl

La figure 10 emmène à proposer un modèle classique, sans aucune composante saisonnière, mixte, car on retrouve deux retards significatifs autant

dans la FAC que dans la FACP. En particulier, des modèles ARIMA (1,1,1) ou (2,1,2) pourraient être adaptés à la structure des données. Le premier modèle cité comporte des coefficients significatifs et satisfait les hypothèses de stationnarité et d'inversibilité. De même, tous les résidus se situent au sein de l'intervalle de confiance et la Qstat au niveau du 24^{ème} retard est égal à 0.544. Le deuxième modèle, cependant, comptait avec des coefficients non significatifs statistiquement, ce qui nous emmène à arrêter l'analyse. D'autres modèles ont été testés, autant mixtes en variant le nombre de paramètres des processus AR et MA, que simples avec un ou deux paramètres dans chaque composante, mais tous ces modèles se caractérisent par avoir au moins un coefficient non significatif. Les caractéristiques du modèle retenu sont affichées dans la table 3, où une valeur Qstat élevée et un écart-type faible sont observés. Cependant, le R^2 ajusté pourrait s'améliorer, de même que l'AIC et l'erreur de prévision.

En ce qui concerne la figure 11 sans des logarithmes, les propriétés du corrélogramme suggèrent à nouveau un modèle classique soit simple, soit mixte car on trouve des pics significatifs au niveau de la FAC et FACP. En ce qui respecte les modèles mixtes, seul le modèle ARIMA (1,1,2) vérifie toutes les hypothèses. D'autre part, ARIMA (1,1,0) et ARIMA (0,1,1) parviennent à surmonter également tous les tests, au contraire qu'un modèle avec deux paramètres AR ou deux paramètres MA. Ainsi, nous parvenons à faire la prévision pour les trois modèles retenus afin de comparer l'erreur de prévision et d'autres caractéristiques à prendre en compte, en sachant que la vraie valeur pour 2021 est de 1229.39. Ces informations sont disponibles dans la table 4.

Table 3 : Modèle pour log_PIBpc

	Nombre de paramètres	Q stat(24) p-value	Ecart-type	AIC	R^2 ajusté	Erreur de Prévision
ARIMA (1,1,1)	2	0,544	0,094959	-103,8229	0,090478	6.20%

Table 4 : Comparaison des modèles pour PIBpc

	Nombre de paramètres	Q stat(24) p-value	Ecart-type	AIC	Erreur de Prédiction
ARIMA (0,1,1)	1	0,626	37,00933	609,8045	7.84%
ARIMA (1,1,0)	1	0.382	37,41762	611,0764	7.843%
ARIMA (1,1,2)	3	0,344	36,77475	611,5967	6.74%

En fonction du critère qu'on privilégie, on préfère un modèle ou autre. Les deux premiers modèles ARIMA(0,1,1) et (1,1,0) ont un nombre de paramètre plus souhaitable si on s'intéresse à la parcimonie. En revanche, le troisième modèle composé de trois paramètres a réalisé une prévision plus précise que les autres. En ce qui concerne le Qstat, ARIMA(0,1,1) est le meilleur candidat. Au niveau de l'écart-type et l'AIC, on n'observe pas de grandes différences entre les trois modèles, même si le troisième est légèrement préféré en termes de d'écart-type et le premier en termes d'AIC. En tout cas, le modèle ARIMA(1,1,0) serait celui à retirer. Parmi les modèles restants, ARIMA(0,1,1), même s'il présente un taux d'erreur légèrement supérieur, ses conditions de stationnarité semblent être plus convaincantes. En effet, ARIMA(1,1,2) possède un coefficient très proche de la racine unitaire.

Après avoir réalisé des tests de normalité des résidus, aucun de ces trois derniers modèles respecte cette hypothèse. Par ailleurs, le modèle logarithmique ARIMA(1,1,1) le fait. Par conséquent, ce dernier est avantageux. Le test de normalité pour ce dernier est disponible sur la figure 7 en annexes.

Figure 12 : Modèle PIBpcARIMA(0,1,1)

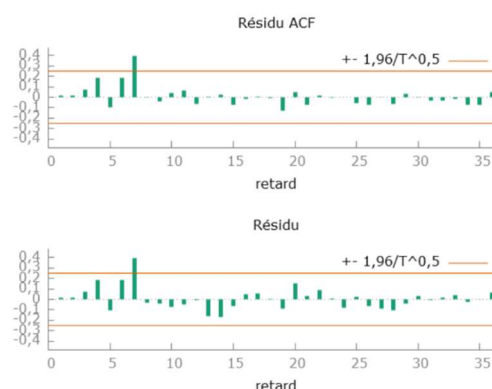
Modèle 50: ARMA, utilisant les observations 1961-2020 (T = 60)
 Estimation par AS 197 (MV exact)
 Variable dépendante: d_GDP
 Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique	
const	17,6646	6,74025	2,621	0,0088	***
theta_1	0,417400	0,111860	3,731	0,0002	***
Moyenne var. dép.	18,15327	Éc. type var. dép.	40,56402		
Moyenne innovations	0,074016	Éc. type innovations	37,00933		
R2	0,153734	R2 ajusté	0,153734		
Log de vraisemblance	-301,9022	Critère d'Akaike	609,8045		
Critère de Schwarz	616,0875	Hannan-Quinn	612,2621		

Source : Propre élaboration, Gretl

Figure 13 : Corrélogramme

résidus PIBpc ARIMA(10,1,1)



Source : Propre élaboration, Gretl

4.5.2. Taux de fertilité

De manière similaire au taux de mortalité infantile, l'évolution du taux de fertilité se caractérise par une descente continue depuis les années 1960 jusqu'en 2020. Cette tendance empêche de considérer la série comme stationnaire. Après une différenciation, le corrélogramme de la figure 14 suggère la présence d'une composante AR classique car le seul pic significatif du premier retard de la FACP indique une forte corrélation entre l'observation actuelle et l'observation immédiatement précédente. En revanche, la FAC montre une décroissance continue qui renforce l'existence potentielle d'un processus AR. D'autre part, nous n'observons pas de récurrences des fluctuations, rejetant la possibilité de modéliser une partie saisonnière.

Une version logarithmique a été également réalisée après une différenciation, dont son corrélogramme exhibe une structure très similaire à celle du graphique 14, duquel on déduit un même modèle ARIMA (1,1,0). Cependant, le modèle sans les logarithmes est plus performant au niveau des résidus, raison pour laquelle on privilège ce dernier.

Les figures 15 et 16 montrent le modèle en question ARIMA (1,1,0) ainsi que le corrélogramme des résidus. Nous constatons un respect de la significativité du

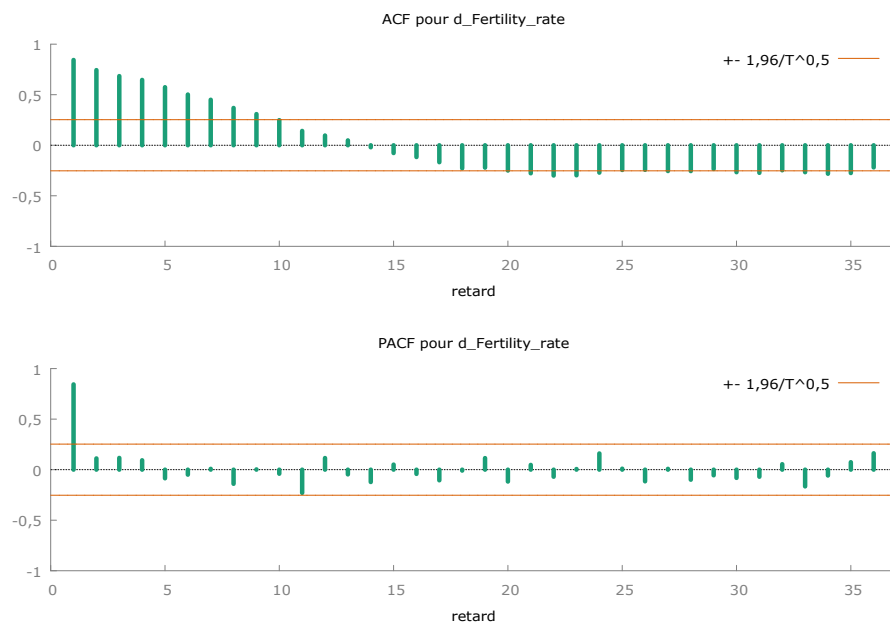
coefficient ainsi que de la condition de stationnarité, en même temps que tous les résidus se trouvent à l'intérieur de l'intervalle de confiance et que la valeur critique au niveau du 24^{ème} retard est bien supérieure à 5% (0.593).

La table 5 présente les différents critères pour évaluer la qualité du modèle retenu. La vraie valeur utilisée pour calculer l'erreur de prévision est égal à 2 pour l'année 2021. Le modèle se caractérise par avoir un Qstat satisfaisant, ainsi qu'un écart-type et AIC faible et un R ajusté élevé. La qualité de la prévision est bonne, même si elle pourrait être encore meilleure. Le modèle n'est composé que d'un paramètre, ce qui est une bonne indication. Cependant, l'hypothèse de normalité des résidus n'est pas validée, même après avoir appliqué des logarithmes, disponible en annexes, figures 8 et 9.

Table 5 : Modèle pour Taux de Fertilité

	Nombre de paramètres	Q stat(24) p-value	Ecart-type	AIC	R^2 ajusté	Erreur de Prévision
ARIMA (1,1,0)	1	0,593	0,022872	-247,64	0,76	1.35%

Figure 14 : Corrélogramme Taux de Fertilité, ordre 1



Source : Propre élaboration, Gretl

Figure 15 : Modèle X2

ARIMA (1,1,1)

Modèle 30: ARMA, utilisant les observations 1961-2020 (T = 60)
 Estimation par AS 197 (MV exact)
 Variable dépendante: d_Fertility_rate
 Écart-types basés sur la matrice hessienne

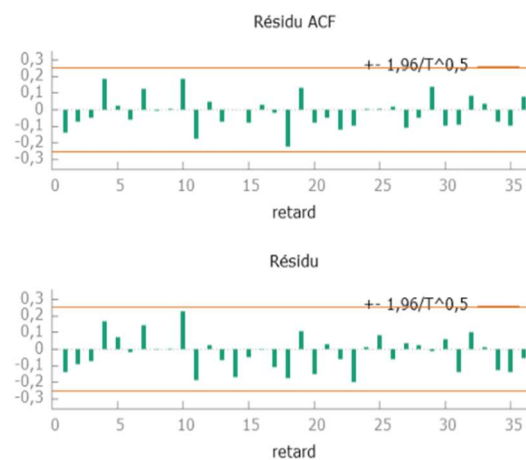
	coefficient	éc. type	z	p. critique	
const	-0,0526761	0,0281049	-1,874	0,0609	*
phi_1	0,877825	0,0611205	14,36	8,93e-047	***

Moyenne var. dép.	-0,066250	Éc. type var. dép.	0,058821
Moyenne innovations	-0,001997	Éc. type innovations	0,028872
R2	0,756336	R2 ajusté	0,756336
Log de vraisemblance	126,8201	Critère d'Akaike	-247,6402
Critère de Schwarz	-241,3572	Hannan-Quinn	-245,1826

Source : Propre élaboration, Gretl

Figure 16 : Corrélogramme des

résidus X2 ARIMA(1,1,1)



Source : Propre élaboration, Gretl

5. COINTEGRATION D'ENGEL-GRANGER

La cointégration d'Engel-Granger fait référence à une combinaison linéaire de variables non stationnaires, qui a été développée par Robert F. Engel et Clive W.J. Granger dans la décade des 1980. Cette approche s'appuie sur l'idée que même si les séries ne sont pas stationnaires individuellement, il peut y exister une relation stable à long-terme. Cela implique une combinaison linéaire stationnaire ainsi qu'une corrélation économique entre les variables.

Pour être en mesure de cointégrer deux séries avec cette approche, il est essentiel que ces dernières soient cointégrées du même ordre, qui doit être égal à 1. Ainsi, nous allons à présent vérifier la stationnarité des variables explicatives avec le test de Dickey-Fuller Augmenté (ADF) et le test Kwiatkowski-Phillips-Schmidt-Shin (KPSS) après une différenciation.

Table 6 : Probabilité critique des tests de stationnarité

Variable	PIB	Taux de fertilité	Taux de mortalité
Test ADF	0,8859	0,1639	0,9999
Test KPSS	0,041	< .01	>.10

Note : Les tests ont été réalisés avec le nombre de retards suggéré par le critère AIC

L'hypothèse nulle du test ADF est que la série chronologique a une racine unitaire, c'est-à-dire qu'elle n'est pas stationnaire. Si la valeur de la probabilité critique obtenue à partir du test est inférieure au niveau de signification de 5%, alors l'hypothèse nulle est rejetée, indiquant une preuve de stationnarité dans la série temporelle. En revanche, l'hypothèse nulle du test KPSS est que la série est stationnaire autour d'une tendance déterministe. Si la valeur critique du test est inférieure à 5%, l'hypothèse nulle est rejetée et la série n'est donc pas stationnaire. En appliquant ces règles, les résultats de la table suggèrent des séries non stationnaires après une seule différenciation pour les variables PIB et Taux de Fertilité selon le critère ADF et KPSS. Toutefois, étant donné que nous avons trouvé des modèles satisfaisants avec une seule différenciation à partir du corrélogramme, nous pouvons identifier ces séries en tant que stationnaires. En ce qui concerne la variable Taux de Mortalité, le test ADF ne permet pas de considérer la série comme

stationnaire, contraire au résultat fourni par le test KPSS. Par conséquent, nous pouvons faire la cointégration de Y avec chaque X

Pour ce faire, nous débutons par réaliser une régression par MCO entre les séries brutes Y et chacune des X. Les modèles 1 et 2 montrent les résultats de ces deux modèles. Nous observons un effet statistiquement significatif et négatif de la variable GDP sur le taux de mortalité de moins de cinq ans sur le modèle 1, ainsi qu'un effet positif de la variable taux de fertilité sur le modèle 2, comme il pouvait être attendu par rapport à la littérature. De plus, le R^2 ajusté des deux modèles est plutôt favorable et supérieur à 0.5 dans les deux cas, ce qui indique une bonne représentation des données.

Modèle 1: MCO, utilisant les observations 1960-2020 (T = 61)

Variable dépendante: Child_mortality

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
const	220,874	11,5976	19,04	<0,0001	***
GDP	-0,228471	0,0267622	-8,537	<0,0001	***
Moyenne var. dép.	152,4000	Éc. type var. dép.		96,99872	
Somme carrés résidus	252551,8	Éc. type régression		65,42583	
R2	0,552630	R2 ajusté		0,545047	
F(1, 59)	72,88179	P. critique (F)		6,80e-12	
Log de vraisemblance	-340,5744	Critère d'Akaike		685,1489	
Critère de Schwarz	689,3706	Hannan-Quinn		686,8034	
rho	0,969672	Durbin-Watson		0,025059	

Modèle 2: MCO, utilisant les observations 1960-2020 (T = 61)

Variable dépendante: Child_mortality

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
const	-128,296	16,7973	-7,638	<0,0001	***
Fertility_rate	61,8065	3,52886	17,51	<0,0001	***
Moyenne var. dép.	152,4000	Éc. type var. dép.		96,99872	
Somme carrés résidus	91062,44	Éc. type régression		39,28653	
R2	0,838692	R2 ajusté		0,835958	
F(1, 59)	306,7598	P. critique (F)		4,76e-25	
Log de vraisemblance	-309,4623	Critère d'Akaike		622,9245	
Critère de Schwarz	627,1463	Hannan-Quinn		624,5791	
rho	0,962539	Durbin-Watson		0,012940	

En deuxième lieu, il faut que les résidus de ces deux régressions soient stationnaires. Afin de vérifier cela, nous appliquons le test ADF-augmenté, dont son hypothèse nulle a été spécifié précédemment, sur les résidus. Les résultats des tests n'ont pas permis de définir les résidus comme stationnaires car la probabilité critique asymptotique est supérieure à 5%. Par conséquent, des régressions au sens inverse ont été effectuées. Autrement dit, une régression du PIB et du taux de fertilité a été réalisé en fonction du taux de mortalité des moins de cinq ans. Cependant, même si la variable indépendante avait un effet significatif sur les différentes variables dépendantes, les résidus sont à nouveau non stationnaires. La table 7 montre la probabilité critique résultante du test mentionné pour chacune des régressions.

Table 7 : Probabilité critique du test ADF pour les résidus des différents modèles MCO

	Mortalité des enfants - PIBpc	Mortalité des enfants - Fertilité	PIBpc - Mortalité des enfants	Fertilité - Mortalité des enfants
Test ADF-augmenté	0,09493	0,1619	0,06528	0,4147

Note : Les tests ont été réalisés avec le nombre de retards suggéré par le critère AIC

Etant donné que les séries PIB par habitant et taux de fertilité ont une relation et un sens économique, nous pouvons également appliquer la cointégration entre elles. D'une part, il est estimé qu'une réduction de la fertilité dans un territoire entraîne une perte du PIB par habitant¹⁶ (Quamrul H. Ashraf, 2013), tandis qu'une augmentation du PIB par habitant peut réduire le taux de fertilité, en particulier chez les pays en développement, cas applicable au Népal¹⁷ (Luci-Greulich, 2014). Toutefois, il convient de préciser que cette relation peut être complexe et le signe de la corrélation varier en fonction du contexte spécifique de chaque pays. Après avoir réalisé des régressions par MCO entre ce couple de séries, le modèle 3 montre un

¹⁶ https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1728-4457.2013.00575.x?casa_token=YI3OKf_fLPiAAAAA:QNWclPLkxSJtQUAeEFp3EPmqEJEKincj4M2VOZa6benwQ6QCVkkj0O7ldV_A7nieSkvicB7Q2vJJQZl Consulté le 01/04/2024 à 12 :46

¹⁷ https://link.springer.com/article/10.1007/s10680-013-9309-2?sa_campaign=email/event/articleAuthor/onlineFirst. Consulté le 01/04/2024 à 12:53

effet significatif et négatif du taux de fertilité sur le PIB par habitant au Népal. Ce résultat diffère de celui trouvé par l'article référencé. Néanmoins, nous avons déjà souligné l'importance du contexte de l'échantillon analysé. La relation entre ces variables est sujette au moment historique, au pays et à tout autre facteur pouvant l'influencer. En ce qui concerne le modèle 4, on observe la manière dont le PIB par habitant influence significative et négativement le taux de fertilité au Népal, comme il était attendu par rapport à la littérature retrouvée. Le R^2 ajusté est à nouveau élevé dans les modèles 3 et 4.

Modèle 3: MCO, utilisant les observations 1960-2020 (T = 61)
Variable dépendante: GDPpc

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
const	1184,69	62,7338	18,88	<0,0001	***
Fertility_rate	-194,866	13,1795	-14,79	<0,0001	***
Moyenne var. dép.	299,7041	Éc. type var. dép.		315,6104	
Somme carrés résidus	1270180	Éc. type régression		146,7259	
R2	0,787474	R2 ajusté		0,783872	
F(1, 59)	218,6135	P. critique (F)		1,67e-21	
Log de vraisemblance	-389,8410	Critère d'Akaike		783,6820	
Critère de Schwarz	787,9038	Hannan-Quinn		785,3366	
rho	0,960193	Durbin-Watson		0,079355	

Modèle 4: MCO, utilisant les observations 1960-2020 (T = 61)
Variable dépendante: Fertility_rate

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
const	5,75266	0,118443	48,57	<0,0001	***
GDPpc	-0,00404110	0,000273314	-14,79	<0,0001	***
Moyenne var. dép.	4,541525	Éc. type var. dép.		1,437252	
Somme carrés résidus	26,34078	Éc. type régression		0,668172	
R2	0,787474	R2 ajusté		0,783872	
F(1, 59)	218,6135	P. critique (F)		1,67e-21	
Log de vraisemblance	-60,94271	Critère d'Akaike		125,8854	
Critère de Schwarz	130,1072	Hannan-Quinn		127,5400	
rho	0,978768	Durbin-Watson		0,063575	

De plus, les résidus des deux modèles se sont avérés stationnaires au seuil de 5%. Les probabilités critiques asymptotiques retrouvées peuvent être observées dans la table 8.

Table 8 : Probabilité critique du test ADF pour les résidus des modèles MCO entre les X

	PIB pc- Taux de fertilité	Taux de fertilité- PIB pc
Test ADF- augmenté	0,002615	0,008872

Note : Les tests ont été réalisés avec le nombre de retards suggéré par le critère AIC

Nous pouvons, par conséquent, appliquer le modèle MCE (Modèle à Correction d'Erreur), qui modélise à la fois la relation à long terme entre les variables et la manière dont les écarts à court terme par rapport à cette relation sont corrigés au fil du temps. Pour que la relation observée ne soit pas fallacieuse, il faut que le coefficient des résidus décalés d'une période du modèle mentionné soit significativement négatif. Néanmoins, comme nous pouvons l'observer dans les modèles 3 et 4, pour le MCE du PIB par habitant en fonction du taux de fertilité, les résidus ne sont pas significatifs. En revanche, les résidus retardés d'une période pour le MCE du taux de fertilité en fonction du PIB par habitant, sont significatifs, mais positifs.

Modèle 3: MCO, utilisant les observations 1961-2020 (T = 60)

Variable dépendante: d_GDPpc

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
const	18,1249	9,62184	1,884	0,0647	*
d_Fertility_rate	3,68112	124,776	0,02950	0,9766	
uhat1_1	-0,0460190	0,0527522	-0,8724	0,3867	
Moyenne var. dép.	18,15327	Éc. type var. dép.		40,56402	
Somme carrés résidus	94772,50	Éc. type régression		40,77592	
R2	0,023779	R2 ajusté		-0,010475	
F(2, 57)	0,694202	P. critique (F)		0,503646	
Log de vraisemblance	-306,0830	Critère d'Akaike		618,1660	
Critère de Schwarz	624,4491	Hannan-Quinn		620,6237	
rho	0,391046	Durbin-Watson		1,210685	

Modèle 4: MCO, utilisant les observations 1961-2020 (T = 60)

Variable dépendante: d_Fertility_rate

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
const	-0,0703524	0,00523466	-13,44	<0,0001	***
d_GDPpc	0,000288527	0,000125982	2,290	0,0257	**
uhat3_1	0,0751891	0,00777234	9,674	<0,0001	***
Moyenne var. dép.	-0,066250	Éc. type var. dép.		0,058821	
Somme carrés résidus	0,076440	Éc. type régression		0,036620	
R2	0,625543	R2 ajusté		0,612405	
F(2, 57)	47,61030	P. critique (F)		6,95e-13	
Log de vraisemblance	114,8316	Critère d'Akaike		-223,6632	
Critère de Schwarz	-217,3801	Hannan-Quinn		-221,2055	
rho	0,583860	Durbin-Watson		0,775095	

Dans tous les cas, ces résultats indiquent la présence d'une relation fallacieuse, ou autrement dit, l'apparente association entre les variables est trompeuse ou fausse, qui n'a pas de sens économique. Ce qui nous empêche d'interpréter les coefficients associés.

6. CONCLUSION

En conclusion, ce travail a eu pour objectif l'étude des séries temporelles du taux de mortalité des enfants de moins de cinq ans, du PIB par habitant et du taux de fertilité au Népal pendant la période 1960-2020. Une corrélation économique entre ces trois variables a semblé évidente, où la croissance économique et la réduction de la dynamique de la fertilité sont vinculées à la grande amélioration que le pays a expérimenté en termes de santé infantile au cours des dernières décennies. Des modèles ARIMA ont pu être modélisés et comparés après avoir différencié les différentes séries une ou deux fois, ce qui a aidé à comprendre les relations de dépendance entre la valeur à un moment donné et les valeurs et/ou erreurs à instants passés. Cela a été crucial pour pouvoir réaliser des prévisions qui ont été efficaces et précises. En particulier, le modèle ARIMA (0,2,2) établi pour le taux de mortalité des moins de cinq ans présente des caractéristiques satisfaisantes avec un taux d'erreur de 0.15%. Cependant, afin de vérifier l'hypothèse de normalité des résidus, ce même modèle ayant subi des transformations logarithmiques a été nécessaire, avec un taux d'erreur de 0.2%. La cointégration entre différents couples de séries, en revanche, n'a pas permis d'établir une relation non fallacieuse et stable à long-terme entre ces mêmes. Ceci suggère que, bien qu'il existe une corrélation économique entre les variables analysées, cette relation n'est pas nécessairement stable à long terme et peut être soumise à des changements ou à des influences externes non prises en compte dans cette étude. L'importance de prendre en compte d'autres facteurs et de mener des analyses supplémentaires pour bien comprendre la dynamique sous-jacente est soulignée.

7. DISCUSSION

En tant que discussion, il est important de préciser, dans un premier temps, que le fondement théorique de relation entre les séries analysées dans ce dossier a été validé par l'étude quantitative. En effet, les graphiques réalisés qui montrent l'évolution conjointe des variables a permis d'observer et confirmer la corrélation attendue, qui a été très élevée et de la direction correspondante. La même conclusion peut être tirée à partir des résultats des MCO effectués sur la partie de cointégration. De plus, des modèles ARIMA ont été trouvés qui ont respecté toutes les hypothèses sous-jacentes pour les séries taux de mortalité des moins de cinq ans et PIB par habitant, en particulier après application des logarithmes pour valider l'hypothèse de normalité des résidus. Le modèle retenu pour le taux de fertilité, cependant, n'a pas réussi à valider cette dernière hypothèse d'aucune manière.

Des différenciations ont dû être appliquées afin de pouvoir proposer des modèles ARIMA. Dans une situation idéale, compte tenu du critère de parcimonie, de la plus grande robustesse des résultats et de la moindre perte d'information, il est préférable que la série brute soit stationnaire sans besoin de la transformer. De plus, en ce qui concerne le PIB par habitant et le taux de fertilité, même si le corrélogramme des FAC et FACP proposaient des modèles ARIMA après une seule différenciation, d'après les tests ADF-augmenté et KPSS, les séries en question ne sont stationnaires qu'après une deuxième différenciation. Malgré ce fait, les résidus des modèles MCO entre ce couple de variables brutes sont stationnaires. En ce qui respecte le taux de mortalité des moins de cinq ans, le modèle proposé $\text{logARIMA}(1,1,1)$ pour faire la cointégration présente une qualité très limitée. En effet, le coefficient ϕ est très proche du bruit blanc, de même que le Qstat au 24^{ème} retard est égal à 0.051, seulement surmontant d'une millièème le seuil de 5%. De plus, le corrélogramme de la série ne semblait pas suggérer une série stationnaire. Même si le test KPSS a réussi à considérer la série comme stationnaire après une différenciation, ceci a été possible avec un nombre de retards élevé (19). C'est pour cette raison que le modèle différencié deux fois $\text{ARIMA}(0,2,2)$ a été privilégié pour réaliser la prévision, qui a été remarquable.

Une contrainte à laquelle nous avons fait face c'est la taille réduite de l'échantillon, qui a été limitée en raison de la disponibilité des données. Ceci peut porter préjudice sur la précision des tests de stationnarité et des intervalles de confiance, parmi d'autres. Lorsque la taille de l'échantillon analysé augmente, les intervalles de confiance deviennent plus étroits, indiquant une plus grande précision dans les estimations des coefficients d'autocorrélation partielle et d'autocorrélation simple, ce qui fournit des meilleures conditions pour sélectionner le nombre de paramètres des modèles ARIMA.

En ce qui concerne la cointégration, aucun des couples de séries est arrivé jusqu'au bout de l'analyse et les coefficients associés n'ont pas pu être interprétés. Entre le PIB par habitant et le taux de fertilité, cela pourrait suggérer la présence d'autres facteurs intervenants qui influencent la relation entre les variables étudiées, qui seraient utiles d'explorer dans des recherches futures afin d'acquérir une compréhension plus complète de la dynamique socioéconomique et démographique. D'autre part, le fait que les résidus des MCO impliquant le taux de mortalité infantile ne soient pas stationnaires peut s'expliquer par les doutes de la stationnarité de cette série après une seule différenciation, comme il a été discuté auparavant.

En résumé, malgré les défis rencontrés, cette étude fournit des informations précieuses sur les relations entre les variables étudiées, soulignant l'importance de la méthodologie et d'une interprétation prudente des résultats dans le contexte des limites de la taille de l'échantillon et des modèles utilisés.

8. BIBLIOGRAPHIE

Sources des données

Taux de mortalité des moins de cinq ans : Banque Mondiale

<https://datos.bancomundial.org/indicador/SH.DYN.MORT?locations=NP>

PIB par habitant : Banque Mondiale

<https://datos.bancomundial.org/indicador/NY.GDP.PCAP.CD?locations=NP>

Taux de fertilité : Banque Mondiale

<https://datos.bancomundial.org/indicador/SP.DYN.TFRT.IN?locations=NP>

Sources littéraires

Bernadette O'Hare, I. M.-Z. (2013). Income and child mortality in developing countries: a systematic review and meta-analysis. *Journal of the Royal Society of Medicine*.

Dangol, B. D. (1997). Declining fertility in Nepal. *Asia-Pacific Population Journal*, 33-54.

Easterlin, R. A. (1996). *Growth Triumphant: The Twenty-first Century in Historical Perspective*. Michigan.

Gubhaju, B. B. (1986). Effect of birth spacing on infant and child mortality in rural Nepal. *Journal of Biosocial Science, Volume 18, Issue 4*, 435 - 447.
doi:10.1017/S002193200001645X

Husnain, M. Z. (2019). Socio-economic, macroeconomic, demographic, and environmental. *Environmental Science and Pollution Research*.

INSEE. (27 de 1 de 2021). Fertility rate. France. Obtenido de
<https://www.insee.fr/en/metadonnees/definition/c1872>

Luci-Greulich, A. T. (2014). Does Economic Advancement 'Cause' a Re-increase in Fertility? An Empirical Analysis for OECD Countries (1960–2007). *Eur J Population*, 187–221.

Osmani, S. R. (2007). The economic development of Nepal: A long-term perspective. 2-4.

Philip Kreager, B. W. (2015). *Population in the Human Sciences: Concepts, Models, Evidence*. Oxford University Press. Obtenido de
https://books.google.fr/books?id=_PazBgAAQBAJ&dq=relationship+between+child+mortality+rate+and+fertility+rate&lr=&hl=es&source=gbs_navlinks_s

Quamrul H. Ashraf, D. N. (2013). The Effect of Fertility Reduction on Economic Growth. *Population and Development Review*, 97-130.

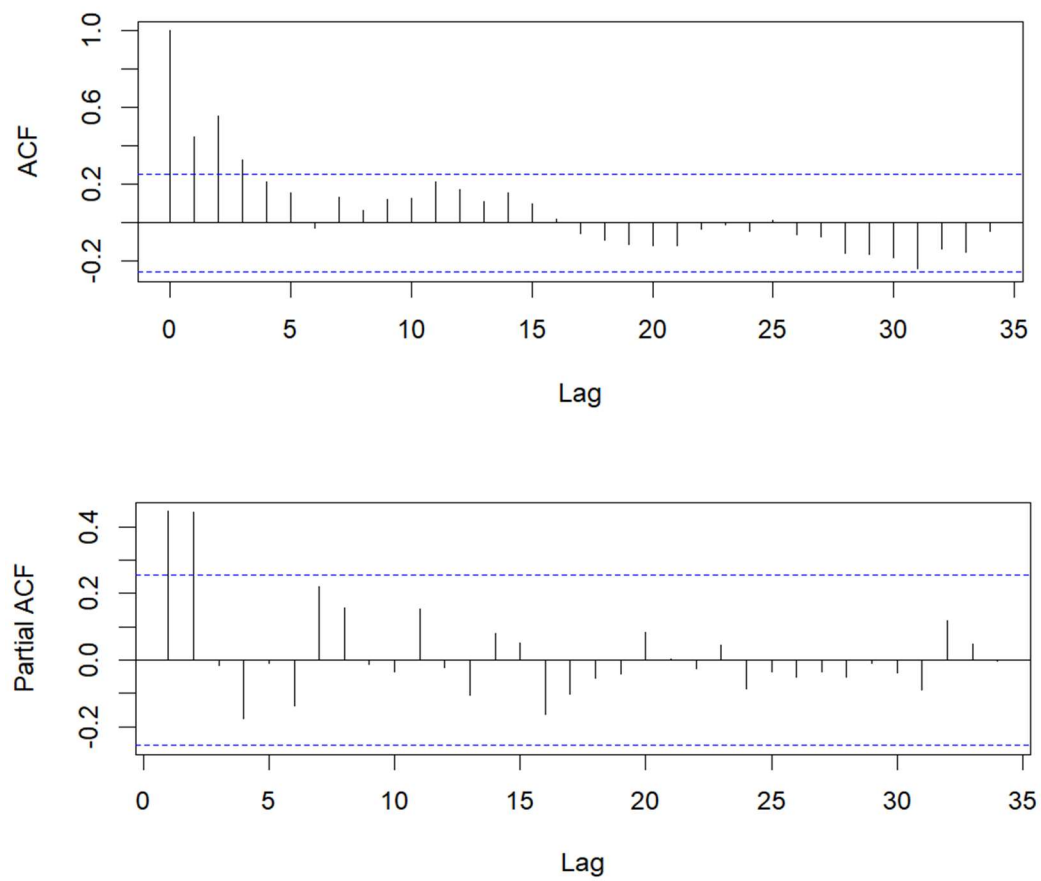
Robert E. Black, R. L. (2016). *Reproductive, Maternal, Newborn and Child Health* (Vol. 2). World Bank Group.

Rutstein, S. O. (2000). *Factors associated with trends in infant and*. Bulletin of the World Health Organization.

Tanja AJ Houweling, A. E. (13 de Septembre de 2005). Determinants of under-5 mortality among the poor and the rich: a cross-national analysis of 43 developing countries. *International Journal of Epidemiology*, 1257–1265.

9. ANNEXES

Figure 1 : Corrélogramme log Y_t différencié d'ordre 2



Source : Propre élaboration, R

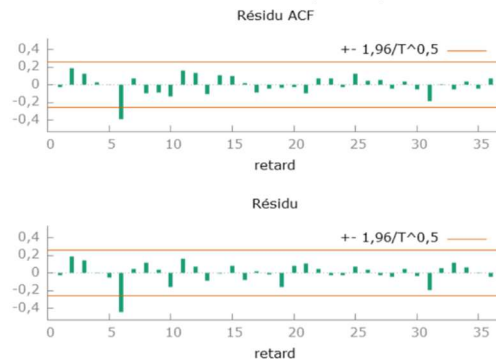
Figure 2 : Modèle log Yt ARIMA(2,2,0)

Modèle 28: ARMA, utilisant les observations 1962-2020 (T = 59)
Estimation par AS 197 (MV exact)
Variable dépendante: d_d_l_Child_mortality
Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique	
phi_1	0,272986	0,0983346	2,776	0,0055	***
phi_2	0,642931	0,103273	6,226	4,80e-010	***
Moyenne var. dép.	-0,000444	Éc. type var. dép.		0,001812	
Moyenne innovations	0,000138	Éc. type innovations		0,001237	
R2	0,533788	R2 ajusté		0,525609	
Log de vraisemblance	310,3329	Critère d'Akaike		-614,6657	
Critère de Schwarz	-608,4331	Hannan-Quinn		-612,2328	

Source : Propre élaboration, Gretl

Figure 3 : Corrélogramme résidus log Yt ARIMA(2,2,0)



Source : Propre élaboration, Gretl

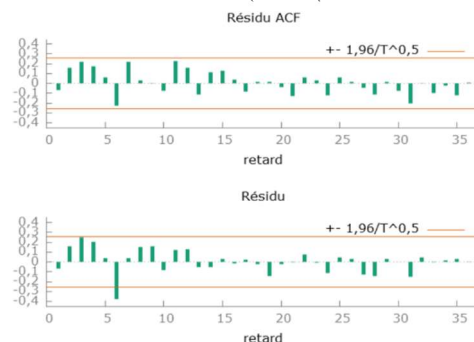
Figure 4 : Modèle log Yt ARIMA(0,2,2)

Modèle 30: ARMA, utilisant les observations 1962-2020 (T = 59)
Estimation par AS 197 (MV exact)
Variable dépendante: d_d_l_Child_mortality
Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique	
theta_1	0,529563	0,170598	3,104	0,0019	***
theta_2	0,674760	0,188338	3,583	0,0003	***
Moyenne var. dép.	-0,000444	Éc. type var. dép.		0,001812	
Moyenne innovations	-0,000147	Éc. type innovations		0,001413	
R2	0,396117	R2 ajusté		0,385522	
Log de vraisemblance	302,7795	Critère d'Akaike		-599,5591	
Critère de Schwarz	-593,3265	Hannan-Quinn		-597,1261	

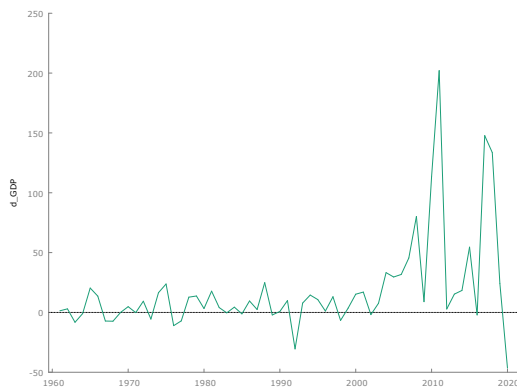
Source : Propre élaboration, Gretl

Figure 5 : Corrélogramme résidus log Yt ARIMA(0,2,2)

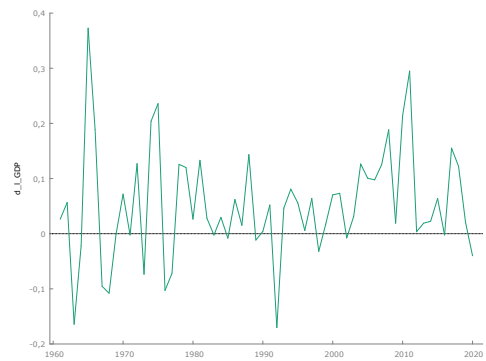


Source : Propre élaboration, Gretl

Figure 4: PIBpc différencié ordre 1 sans log **Figure 5:** PIBpc différencié ordre 1 log

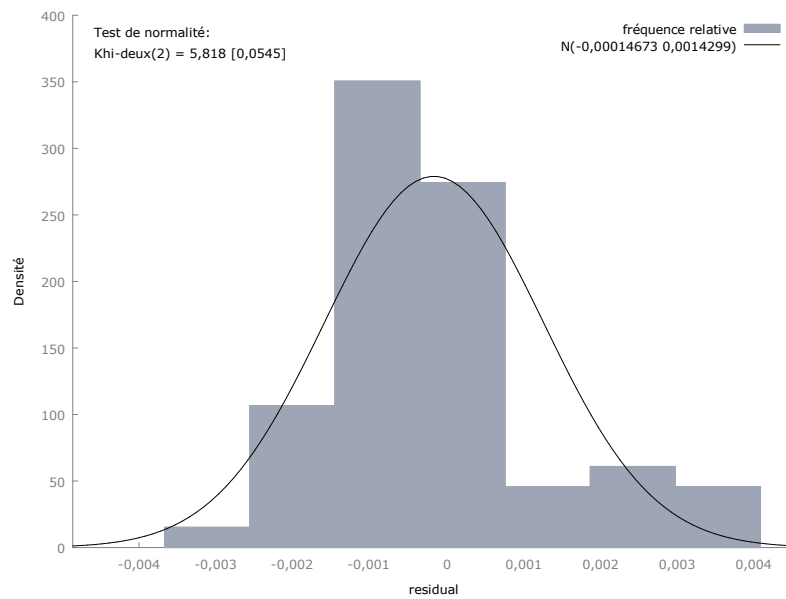


Source : Propre élaboration, Gretl



Source : Propre élaboration, Gretl

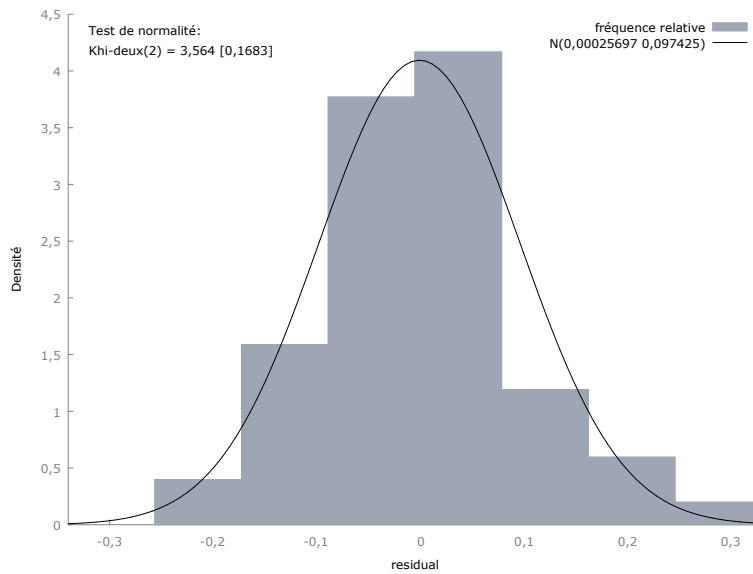
Figure 6: Test de normalité des résidus pour log YtARIMA(0,0,2)



Source : Propre élaboration, Gretl

Compte tenu de la valeur critique supérieure à 5%, l'hypothèse de normalité des résidus est acceptée.

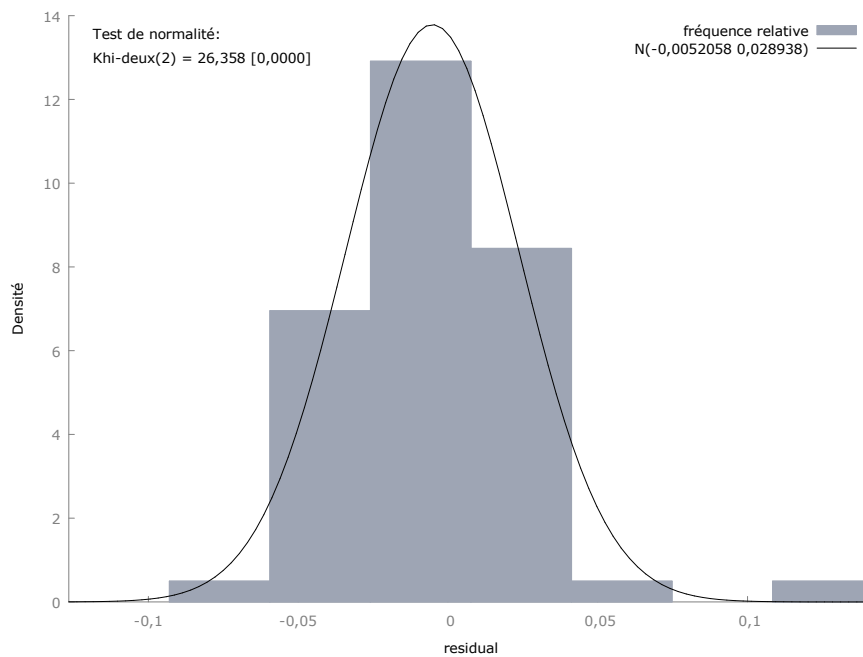
Figure 7: Test de normalité des résidus pour logPIBpc ARIMA(1,1,1)



Source : Propre élaboration, Gretl

Compte tenu de la valeur critique supérieure à 5%, l'hypothèse de normalité des résidus est acceptée.

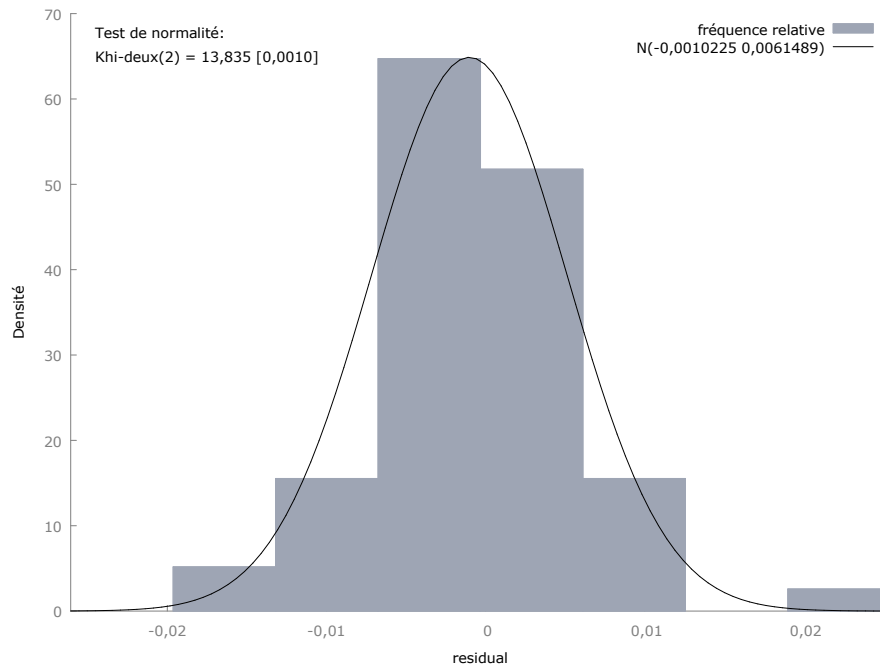
Figure 8: Test de normalité des résidus pour Taux de Fertilité ARIMA(1,1,1)



Source : Propre élaboration, Gretl

Compte tenu de la valeur critique inférieure à 5%, l'hypothèse de normalité des résidus est rejetée.

Figure 9: Test de normalité des résidus pour log_Taux de Fertilité ARIMA(1,1,1)



Source : Propre élaboration, Gretl

Compte tenu de la valeur critique inférieure à 5%, l'hypothèse de normalité des résidus est rejetée.

Tableau des données

Year	Child_mortality	GDPpc	Fertility_rate
1960	324,7	49,9938399	6,03
1961	320,8	51,3219653	6,063
1962	316,1	54,3095757	6,075
1963	310,7	46,0874476	6,065
1964	304,5	45,0883338	6,065
1965	298,2	65,4562707	6,029
1966	291,4	79,0580829	6,001
1967	284,7	71,8678971	5,98
1968	278,1	64,5116875	5,969
1969	271,9	64,4743369	5,912
1970	265,8	69,2709036	5,876
1971	260	69,1017185	5,842
1972	254,5	78,4616255	5,781
1973	249,1	72,893253	5,694
1974	243,7	89,3825512	5,738
1975	238,2	113,158207	5,748
1976	232,6	102,088096	5,751
1977	226,8	95,0075015	5,726
1978	220,7	107,743118	5,719
1979	214,2	121,484545	5,691
1980	207,7	124,734708	5,639
1981	201,1	142,492984	5,602
1982	194,5	146,534869	5,575
1983	188	146,181466	5,555
1984	181,3	150,581386	5,537
1985	174,5	149,363094	5,474
1986	167,5	158,933704	5,424
1987	160,4	161,36759	5,362
1988	153,1	186,264475	5,312
1989	145,8	184,132236	5,253
1990	138,6	184,923648	5,208
1991	131,5	194,800017	5,139
1992	124,6	164,292809	5,044
1993	118	172,09667	5,002
1994	111,7	186,594263	4,907
1995	105,7	197,30965	4,785
1996	99,9	198,454465	4,595
1997	94,4	211,561947	4,393
1998	89,1	204,876507	4,235
1999	84	208,491474	4,083
2000	79,1	223,711892	3,944

2001	74,6	240,705159	3,763
2002	70,3	238,861254	3,59
2003	66,4	246,485838	3,457
2004	62,7	279,7242	3,311
2005	59,4	309,31042	3,142
2006	56,3	341,02814	2,974
2007	53,4	386,529586	2,843
2008	50,7	466,693379	2,719
2009	48,2	475,635976	2,604
2010	45,9	589,165435	2,512
2011	43,6	791,225577	2,439
2012	41,5	794,092559	2,387
2013	39,4	809,384458	2,334
2014	37,4	827,744705	2,293
2015	35,5	882,307663	2,25
2016	33,7	880,224894	2,203
2017	32,1	1027,96547	2,15
2018	30,6	1161,53435	2,104
2019	29,3	1185,68232	2,08
2020	28,2	1139,18989	2,055

10 TABLE DES MATIERES

1.	Introduction	2
2.	Analyse economique du sujet.....	3
1.1	Mortalité des enfants de moins de 5 ans	3
1.2	Produit Intérieur Brut (PIB)	5
1.3	Taux de fertilité	7
1.4	Comparaison des trois séries.....	8
3.	Présentation de la méthodologie ARIMA	10
3.1.	Stationnarité.....	10
3.2.	Corrélogramme	11
3.3.	Modèle Box-Jekins	14
3.4.	Modèles Multiplicatifs	15
4.	Choix du modèle ARIMA.....	17
4.1.	Série, corrélogramme	17
4.2.	Identification, estimation.....	20
4.3.	Vérification	21
4.4.	Prévision à une période Y_t	24
4.5.	Modèle ARIMA pour les séries X_t	25
4.5.1.	PIB par habitant	25
4.5.2.	Taux de fertilité	28
5.	Cointégration d'Engel-Granger.....	31
6.	Conclusion	37
7.	Discussion	38
8.	Bibliographie	40
9.	Annexes.....	42
10	Table des matieres	49