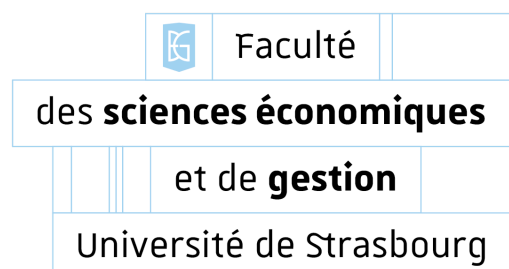


FOGUE Carl-Dimitri

LIGIER Célestin

DOSSIER : Econométrie des séries temporelles

**L'impact des variables démographiques sur le PIB
d'un pays en pleine transition démographique**



SOMMAIRE

1) Présentation de la problématique et le contexte théorique dans lequel a elle été développée.....	3
2) Les principales idées dégagées de nos différentes lectures sur la problématique.....	7
3) Elaboration de trois à quatre hypothèses émanant des idées dégagées.....	9
4) Présentation des concepts économétriques utilisés pour mener l'étude empirique.....	10
5) Présentation, de la base des données et des estimations, obligatoirement sous R (les codes doivent êtres fournis ainsi que la base des données).....	14
6) Interprétation des résultats et recommandations.....	18

1) Présentation de la problématique et le contexte théorique dans lequel a elle été développée

À la suite de la Première Révolution Industrielle dans la seconde moitié du XIXème siècle, de nombreux pays Occidentaux ont connu leur transition démographique. A l'époque, la loi de Malthus a également marqué les esprits en économie et a pu renforcer ce lien entre faible croissance démographique et forte croissance économique.

Depuis, à raison ou à tort, nous avons tendance à associer faible taux de natalité et forte croissance économique. Et ces dernières décennies, les choses ne sont pas allées en changeant puisque certains pays moins développés ont aussi connu leur transition démographique et ont alors enregistré des taux de croissance bien supérieurs à ceux des pays Occidentaux.

Toutefois, la question démographique revient sur le devant de la scène dernièrement du fait du vieillissement important des pays les plus développés et de leur natalité en berne, pour ne pas dire quasiment au point mort (0,82 enfant par femme en Corée du Sud par exemple). Seulement cette fois-ci, l'affaiblissement de la croissance démographique s'accompagne également d'un affaiblissement très clair des taux de croissances de ces mêmes pays.

Notre étude portera donc sur la croissance économique en elle-même, dans quelle mesure les taux de croissance passés permettent de déterminer les taux de croissance futurs, mais avec un accent sur le contenu prédictif de facteurs démographiques (fécondité, structure par âge, espérance de vie, vieillissement,...) sur la croissance économique.

Situation théorique du problème :

Alfred Sauvy disait « Le vieillissement de la population a touché plusieurs populations dans l'histoire : la Grèce, Rome et Venise en sont les exemples les plus fameux. A chaque fois, ce fut la mort de la société en question ». Si l'affirmation est quelque peu catégorique et ne dit rien de ce qu'il en sera dans notre cas, elle

témoigne tout de même de l'importance de la question démographique pour une société et son économie.

Peut-être l'argument le plus évident en faveur de la croissance démographique comme facteur favorisant la croissance économique est celui de la main-d'œuvre abondante. Cela s'est notamment vu lors de la seconde moitié du XX^{ème} siècle où de nombreux PED ont bénéficié de l'avantage comparatif qu'est une main-d'œuvre abondante et bon marché dans leur processus d'industrialisation. W. Arthur Lewis (1954) a montré que le secteur agricole, caractérisé par un sous-emploi important à l'époque, a constitué une abondante réserve de main-d'œuvre à bas coûts pour le secteur industriel. Cela a facilité l'embauche et l'investissement, et donc la croissance. Les exemples les plus parlants sont bien évidemment la Chine et l'Inde. Mais même auparavant, W. W. Rostow (1960) argumentait que la croissance soutenue de la population a favorisé l'industrialisation des pays Européens au XIX^{ème} siècle et a été à l'origine du *take-off* des pays développés actuels en assurant une main d'œuvre bon marché aux industries de ces pays.

Néanmoins, d'autres auteurs ont associé croissance démographique et difficultés économiques. Comme nous l'avons mentionné plus haut, l'un des premiers à avoir souligné l'importance de la question démographique a été Thomas Malthus avec sa loi de population (1798). Celle-ci stipulait qu'il y aurait un arrêt de la croissance à terme car il finirait par y avoir un trop grand écart entre la population, qui croît de façon géométrique, et les ressources, qui croissent de façon arithmétique. Sa solution étant la réduction de la natalité chez les classes pauvres. Si cette option a été laissée de côté car jugée moralement douteuse par la suite, cette logique a parfois pris d'autres formes.

En effet, cette réflexion a par exemple été renouvelée par certains auteurs qui valorisent la question écologique. Paul Ehrlich a par exemple publié *La bombe P* (1968) où il considère que la crise écologique est principalement due à la « prolifération humaine » et il cherche en conséquence à décourager la natalité. Le rapport de Meadows de 1972 mentionnait également que la résolution de la crise écologique passait par une nécessaire réduction de la croissance démographique.

Toutefois, ces approches écologiques ne nous disent pas grand-chose du lien unissant démographie et croissance du PIB en tant que tel, et surtout depuis quelques temps il s'avère que le problème n'est plus la surpopulation mais bien l'inverse dans la majorité des pays Occidentaux. Le seuil de renouvellement de la population (2,1) enfants par femme n'est plus atteint, et le vieillissement de la population pose de nombreuses questions. Un vieillissement de la population qui s'explique par un recul de la natalité et de la fécondité qui abaisse la part des jeunes dans la population, et d'autre part par une augmentation de l'espérance de vie à la naissance.

Avant d'aller plus loin quant à la question du rôle de la démographie dans la prévision de la croissance économique, nous rappelons qu'il faut bien distinguer population totale et groupe d'âge des actifs pour ne pas ignorer les effets de la structure par âge de la population (Bloom, Canning, Sevilla, 2001). Pour faire large nous pouvons simplement distinguer deux classes d'âge : la population active et les retraités.

D'abord, rappelons que l'accumulation du capital est directement liée au taux d'épargne et il importe donc de s'interroger sur les liens entre les structures démographiques et l'évolution de l'épargne. L'une des théories les plus célèbres en lien avec ceci est celle du cycle de vie de Modigliani (1963). Si l'épargne obéit à ce modèle du cycle de vie, les effets d'un vieillissement de la population devraient être très importants. En effet, ce modèle suppose que les individus âgés, après avoir épargné pendant leur période d'activité, adoptent une attitude de désépargne. Dans ce cadre, le vieillissement de la population aboutirait donc à une baisse de l'épargne et de l'accumulation du capital tandis que l'augmentation de la population active aurait l'effet inverse. Ce qui irait dans le sens d'une diminution de la croissance si la population est trop âgée.

Cependant, la validité du modèle du cycle de vie a fait l'objet de nombreuses contestations. Ainsi, pour Flipo (1999), le modèle du cycle de vie n'est pas vérifié empiriquement tandis que selon Blanchet (1990 et 1994), l'origine du vieillissement de la population influence le comportement d'épargne. Ainsi, si le modèle du cycle de vie est vérifié lorsque le vieillissement s'explique par un ralentissement de la croissance démographique, il ne l'est pas lorsque l'augmentation de l'espérance de

vie est la cause principale du vieillissement. De plus, les modalités de financement des études d'une part, la nature et la fiabilité des systèmes de retraite d'autre part, interfèrent nécessairement avec le cycle de l'épargne, alors que la logique de Modigliani impliquait une absence de système de retraite effectif.

D'autres auteurs ont quant à eux souligné un lien entre croissance démographique et innovation. La première à l'avoir fait est Esther Boserup (1965). Selon elle, le progrès technique issu de l'agriculture résulterait d'une « pression créatrice ». L'expansion de la population qui entraîne une augmentation de la demande de biens agricoles a contraint les sociétés à adapter leurs méthodes de production afin d'accroître les rendements agricoles et nourrir l'ensemble de la population (en améliorant la fertilité des sols par exemple). On a ici en quelque sorte une réponse à l'impasse que prétendait mettre en lumière Malthus. Plus tard, la thèse de Boserup a été développée par Darity (1980) et Pryor et Maurer (1981). On retrouve donc bien un lien entre croissance démographique et croissance économique même si le lien est cette fois-ci indirect.

Enfin, le vieillissement de la population risque d'engendrer une réduction de la création d'entreprises comme l'a montré Duhautois (2000) en indiquant que la propension à créer une entreprise est deux fois plus élevée pour les moins de 45 ans, alors même que l'on sait que les processus de création et destruction des entreprises peuvent être générateurs de gains de productivité.

2) Les principales idées dégagées de nos différentes lectures sur la problématique

Pour la lecture d'articles, nous nous sommes concentrés sur le rôle de la transition démographique dans la croissance économique, d'abord dans les pays de la Révolution Industrielle, et autant que la littérature nous le permettait dans des pays peu développés ayant connu leur transition démographique tardivement.

Bloom, Canning et Sevilla (2001) est un article essentiel pour comprendre cela. Ils se concentrent sur la question de la structure par âge de la population. Selon eux, c'est un élément qui a été sous-estimé dans le débat du rôle de la croissance démographique sur la croissance économique.

Ils avancent que la transition démographique entraîne une génération « boom », c'est-à-dire plus nombreuse que les précédentes, grâce à laquelle la majeure partie de la population se situe dans les tranches d'âge actives. Ainsi, la productivité accrue de ce groupe peut générer ce qu'ils appellent un « dividende démographique » dont le pays pourra tirer avantage à condition d'avoir mis en place les politiques nécessaires.

Dans la continuité, mais centré sur un pays Africain (en l'occurrence le Nigéria) Karra, Canning, Wilde (2017) montrent l'importance de la transition démographique y compris dans ces pays. Ils ne se limitent donc pas à la transition démographique qui a suivi la Révolution Industrielle pour les pays Européens, mais montrent également son importance pour des pays peu développés ou en développement. Ainsi, selon eux un taux de fertilité sur le déclin (tout en partant d'un niveau très haut donc puisque c'est au début de la transition démographique du pays), implique une hausse de la croissance économique. Ils estiment que le revenu par personne pourrait doubler au cours de ce processus. Ils rappellent bien évidemment qu'il y a de nombreux déterminants de la croissance économique et que prendre en compte la question démographique seule ne suffit pas. C'est d'ailleurs pour cela que la croissance économique permise par le déclin de la fertilité ne suffira évidemment pas à faire subitement passer un pays émergent au stade de pays développé.

Nos lectures nous ont montré que de nombreux auteurs ont pu avancer au fil du temps qu'un vieillissement trop important de la population. Comme nous l'avons mentionné plus haut, avec sa théorie du cycle de vie, Modigliani a peut-être été l'un des premiers à souligner que les personnes plus âgées tendent à réduire leur épargne. Ainsi, la conclusion à en tirer est qu'une population au sein de laquelle les personnes âgées sont surreprésentées risquerait de voir sa croissance diminuer du fait d'une baisse de l'épargne et d'un ralentissement ou une diminution de l'accumulation de capital. Il s'avère que la conclusion pourrait être bonne, mais pas forcément la raison pour laquelle la croissance ralentit. Un article récent de Maestas, Mullen et Powell (2023) indique qu'entre 1998 et 2010, une augmentation de 10% de la fraction de la population âgée de 60 ans et plus aurait entraîné une réduction du PIB par habitant de 5,5%. Un tiers de ces 5,5% serait dû au ralentissement de la croissance de l'emploi, et les deux tiers restants seraient la conséquence d'un ralentissement de la productivité du travail. La rémunération du travail et les salaires ont également diminué en conséquence. Finalement, ils estiment que le vieillissement de la population a réduit le taux de croissance du PIB par habitant de 0,3 point de pourcentage par an sur la période concernée.

3) Elaboration de trois à quatre hypothèses émanant des idées dégagées

Nous avons pu élaborer trois hypothèses différentes découlant logiquement des idées dégagées au préalable grâce à nos lectures.

La première hypothèse est qu'un pays avec une forte population en âge de travailler (population active) connaîtra certainement une croissance plus importante que certains pays avec une population très jeune ou très vieillissante.

Notre seconde hypothèse est la suivante : une baisse de la fertilité dans un pays peu développé (donc partant avec un taux de fertilité très élevé à l'origine) au cours du processus de transition démographique devrait entraîner une hausse de la croissance.

Enfin, nous supposons qu'un vieillissement trop important d'une population permettrait de prédire en partie la croissance du PIB future, et ce sûrement car elle la ralentit.

4) Présentation des concepts économétriques utilisés pour mener l'étude empirique

Durant cette partie on s'attelle à définir les différents concepts économétriques que nous avons utilisés pour réaliser notre étude, on donnera des exemples pour chacun des concepts afin de mieux saisir leur intérêt dans notre analyse.

ARIMA : Le processus ARIMA (AutoRegressive Integrated Moving Average) est un modèle statistique utilisé pour analyser et prédire les séries temporelles il est composé de différents processus plus simples que l'on va expliciter. Il y a d'abord la partie AR (AutoRegressive) qui prend en compte les valeurs passées de la série temporelle pour prédire les valeurs futures avec p comme ordre du terme autorégressif. Il faut ici préciser que nous aurons recours au critère de Bayes (BIC) et à celui d'Akaike afin de déterminer le nombre optimal de retards. Ajoutons également qu'un retard, ou *lag* en anglais fait référence au décalage d'une variable dans le passé par rapport à son emplacement temporel actuel. Ensuite le I (Integrated) qui permet de corriger la non-stationnarité de la série temporelle en la différenciant, avec d comme ordre de la différenciation. Et enfin le MA (Moving Average) qui utilise la moyenne des erreurs passées pour prédire les erreurs futures, avec q comme ordre du terme de la moyenne mobile

$$AR(p) : y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t$$

$$MA(q) : y_t = \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2} - \dots - \alpha_q \varepsilon_{t-q}$$

ARMA(p,q)

:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Néanmoins afin de pouvoir réaliser une prédiction à l'aide du modèle ARIMA il faut que la série temporelle soit stationnaire. Une série temporelle est dite stationnaire si elle conserve ses caractéristiques statistiques à tous les moments. Autrement dit, on peut dire que la distribution des probabilités ne change pas au cours du temps. C'est un concept clé pour la validité externe d'un modèle. La distribution jointe $(Y_{s+1}, Y_{s+2}, \dots, Y_{s+T})$ ne doit pas dépendre de s , sinon on dit que Y_t est

non-stationnaire. Toute série chronologique est composée de deux effets stationnaire et tendancielle. Cette tendance peut être décomposée en un effet stochastique et un effet déterministe. Selon l'effet de cette tendance, la non-stationnarité est de type TS (Trend Stationary) et DS (Difference Stationary). Cette non-stationnarité comporte plusieurs risques, dont celle d'avoir une régression factice. Afin d'éviter ce problème, nous allons transformer une série non-stationnaire, en utilisant la première différence de la variable endogène. Si elle devient stationnaire alors on peut estimer les paramètres à partir d'ARIMA. Sinon, on calcule la seconde différence, et ainsi de suite jusqu'à obtenir une série stationnaire (dans la majorité des cas, la première différence est suffisante). C'est ce qu'on appelle l'ordre d'intégration d'une série chronologique noté d , que nous avons mentionné plus haut.

Nous précisons ici qu'il existe plusieurs tests que l'on appelle "test de racine unitaire" afin de déterminer si oui ou non la série est stationnaire. Les deux tests les plus célèbres et courants sont sans doute ceux de Dickey-Fuller (DF) et Dickey-Fuller augmenté (ADF). L'hypothèse nulle (H_0) de ces tests est que la série temporelle possède une racine unitaire, c'est-à-dire qu'elle est non-stationnaire, et H_1 est donc "la série est stationnaire". Le test de Dickey-Fuller compare la valeur de p à une valeur critique pour déterminer si la série temporelle est stationnaire ou non. Si p est significativement différent de zéro, alors nous pouvons rejeter H_0 , et la série est stationnaire. Le test de Dickey-Fuller est fondé sur les régressions suivantes :

$$\Delta y_t = (\rho_1 - 1) y_{t-1} + u_t \text{ (sans constante, sans tendance)}$$

$$\Delta y_t = \beta_0 + (\rho_1 - 1) y_{t-1} + u_t \text{ (Avec constante, sans tendance)}$$

$$\Delta y_t = \beta_0 + \beta_1 t + (\rho_1 - 1) y_{t-1} + u_t \text{ (Avec constante, avec tendance)}$$

Un autre test de racine unitaire très courant est celui de Phillips Perron (PP). Il faut savoir qu'il est moins sensible aux problèmes d'hétéroscédasticité que le test ADF. La procédure consiste à tester l'hypothèse de racine unitaire $H_0 : \rho = 0$ dans les modèles suivants :

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \beta t + \varepsilon_t$$

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \varepsilon_t$$

$$\Delta Y_t = \rho Y_{t-1} + \varepsilon_t$$

En résumé, c'est une statistique de student corrigée de la présence d'autocorrélation par la prise en compte d'une estimation de la variance de long terme de ε_t (calculée

par la densité spectrale de ϵ_t et à la fréquence zéro), robuste à la présence d'autocorrélation et d'hétéroscédasticité.

Mentionnons enfin le test KPSS (Kwiatkowski-Phillips-Schmidt-Shin). Contrairement aux tests de Dickey-Fuller (ADF) et de Phillips-Perron (PP) qui testent l'hypothèse de présence d'une racine unitaire, le test KPSS teste l'hypothèse opposée, c'est-à-dire l'hypothèse de présence de stationnarité. Le test KPSS évalue si une série temporelle est stationnaire autour d'une tendance déterministe, comme une tendance linéaire ou quadratique, ou stationnaire autour d'une moyenne constante. Ainsi, il peut détecter différents types de non-stationnarité, notamment les tendances et les cycles. Il est basé sur une régression linéaire, et décompose une série en trois parties : une tendance déterministe (βt), une marche aléatoire (r_t) et une erreur stationnaire (ϵ_t), avec l'équation de régression : $x_t = r_t + \beta t + \epsilon_t$.

Revenons-en à notre modèle ARIMA. Nous avons vu de quelle façon on pouvait calculer un des paramètres du modèle ARIMA, mais qu'en est-il des deux autres ? Nous pouvons utiliser des corrélogrammes afin d'estimer chacun de ces paramètres. Pour le p : le corrélogramme simple d'un processus AR(p) est caractérisé par une

décroissance géométrique de ses termes de type : $\rho_k = \rho^k$ Le corrélogramme partiel à ses seuls p premiers termes différents de 0. Pour q : un corrélogramme de

$$\rho_k = \frac{\sum_{i=0}^{i=q-k} \alpha_i \alpha_{i+k}}{\sum_{i=0}^{i=q} \alpha_i^2} \text{ pour } k = 0, 1, \dots, q \text{ et } \rho_k = 0 \text{ pour } k > q.$$

type MA se présente comme suit :

C'est-à-dire que seuls les q premiers termes du corrélogramme simple sont significativement différents de 0. Le corrélogramme partiel est caractérisé par une décroissance géométrique des retards. Nous avons utilisé les corrélogrammes ARMA dans le cadre de notre étude ils calculent les deux paramètres en même temps.

Voici leur propriété :

ARMA(1, 1)	Décroissance géométrique à partir du premier retard, le signe est déterminé par $\theta_1 - \alpha_1$	Décroissance exponentielle ($\alpha_1 > 0$) ou sinusoïdale amortie ($\alpha_1 < 0$)
ARMA(p , q)	Décroissance exponentielle ou sinusoïdale amortie tronquée après $(q - p)$ retards	Décroissance exponentielle ou sinusoïdale amortie tronquée après $p - q$ retards

Nous avons mentionné plus haut la possibilité d'utiliser la différenciation pour contourner le problème de la non-stationnarité, mais il est aussi possible de voir si les séries sont cointégrées. Élaborée par Engle et Granger en 1987, la théorie de la cointégration stipule que si deux variables sont intégrées d'ordre 1, noté $I(1)$, et si leur combinaison linéaire est intégrée d'ordre 0, noté $I(0)$, alors ces deux variables sont dites cointégrées. Nous rappelons que l'ordre d'intégration traduit le nombre de filtres appliquées à la série pour qu'elle devienne stationnaire. $I(d)$ signifie que la série est stationnaire après d différenciations.

x et y deux variables $I(1)$ sont cointégrées s'il existe un vecteur telle que l'erreur d'équilibre v_t :

$$\begin{bmatrix} \beta_1 & -\beta_2 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = v_t$$

Les deux mêmes auteurs ont également proposé un test de cointégration qui repose sur une méthode à deux étapes : la première consiste à estimer la relation de long terme, et la seconde à substituer les paramètres estimés au cours de la première étape dans la spécification incorporant les variations de court terme et les ajustements de long terme. C'est une méthode de régression spécifiquement adaptée à l'analyse des séries temporelles et qui s'attache à la fois aux dynamiques de court et de long terme : le modèle à correction d'erreur (ECM).

RMSFE : Mesure qui permet d'évaluer la précision des prévisions en mesurant l'erreur moyenne des prévisions.

$$RMSFE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_{t|t-1})^2}$$

5) Présentation, de la base des données et des estimations, obligatoirement sous R (les codes doivent être fournis ainsi que la base des données)

La base de données qui a été utilisée provient du site databank.worldbank.org, plus précisément de la partie “World Development indicator”. Le WDI est la principale collection de la Banque mondiale d'indicateurs de développement, compilée à partir de sources internationales officiellement reconnues, on peut citer l'ONU, l'OMS ou encore le FMI. Elle présente les données mondiales les plus actuelles et précises. Nous avons choisi d'inclure dans notre base de données 5 pays différents afin de pouvoir comparer facilement les résultats que nous aurons obtenu de chacun d'entre eux. Ces pays sont le Sénégal, la Chine, la France, les Etats-Unis et le Royaume-Unis. Les deux premiers venaient de commencer la transition démographique lors du début de notre série temporelle, en effet nous avons agrégé des données pour ces 5 pays sur 50 ans, de 1973 à 2022, les données disponibles les plus récentes. Les données sont prises à une fréquence annuelle. Enfin, pour conclure cette présentation, notre base de données est composée de 28 variables. Dont 3 sont des variantes de la variable endogène, le PIB. On a la variation du PIB en %, le PIB total du pays et le PIB par tête. On a ensuite choisi quelques variables démographiques afin de regarder l'impact qu'elles auront sur le taux de croissance. On a : le taux de fertilité (nombre d'enfant par femme), l'espérance de vie, la pyramide des âges, les migrations net, le pourcentage de population habitant en ville, et des variables concernant le chômage (le pourcentage de chômeur ayant une éducation avancée, intermédiaire et basique). Ensuite, quelques variables économiques pour contrôler, isoler l'effet de la/les variables démographiques dont l'inflation, la balance externe des biens et des services, l'épargne intérieure brute (se calcule en** retranchant la consommation finale des dépenses (total de la consommation) du PIB), les intérêts payés et la taxe sur les exportations.

Ensuite concernant le traitement des données manquantes ou aberrantes qui se doit d'être fait quand on crée de toute pièce une base de données, il s'est effectué selon

quelques principes simples. Quand il y avait des valeurs manquantes, on faisait quelques recherches pour voir si l'information était librement disponible sur le net. Sinon, on faisait une moyenne 20 valeurs les plus proches (passé et futur) en mettant un poids aux années les plus proches afin d'obtenir une approximation la plus juste possible. Toutefois, certaines variables n'ont malheureusement pas assez d'information, et sont majoritairement vides, dans ce genre de cas on a tout simplement évité d'utiliser la variable. Cela est particulièrement le cas pour la Chine et le Sénégal qui de par leur statut de pays en développement et en voie de développement durant une partie de la période de cette base de données n'avaient pas récupéré les données. On a pas trouvé de valeurs particulièrement aberrantes, de ce côté-là, on a rien changé.

Nous allons désormais nous pencher sur les étapes que nous avons prises concernant l'analyse. D'abord, il a fallu séparer les variables par pays.

Ensuite on a voulu plot les différentes variables pour voir s'il y a une quelconque corrélation entre elles. On a surtout regardé nos variables endogènes pour voir si elles avaient des tendances observables, de la saisonnalité.

On s'est aussi servi de graphiques à visualisation interactive pour plus de précision dans l'étude de nos variables. Et on a bien évidemment réalisé les statistiques descriptives de nos variables.

On a créé la série temporelle à partir de notre base de donnée, qui est plutôt sous forme de donnée de panel afin de pouvoir utiliser différent package dessus. On a créé deux sortes séries temporelles : les séries temporelles univarié (une seule variable) et les série temporelles multivariée (n variables). Elles avaient encore une fréquence donc une fréquence de un (pour 1 année) avec un début en 1973 et la fin en 2022 ce qui nous donne 50 périodes.

Lag / lead ? Estimation à l'aide de lag ? Combien de lag nécessaire ? Automatisation

Ensuite décomposition en tendance / saisonnalité / random(bruit) (montrer graphique). ne marche plus

On a réalisé des tests ACF(AutoCorrelation Function) pour connaître l'ordre d'intégration de nos séries. Pour la variation du PIB, les valeurs de la série semblent être indépendantes, en effet, les autocorrélations sont pour la grande majorité proche de 0. De plus, il présente une décroissance lente des coefficients. Donc cela signifierait, qu'il n'a ni tendance, saisonnalité ou motifs particuliers. Néanmoins, des pics importants à des lags spécifiques indiquent une certaine dépendance saisonnière. On a donc décidé de différencier la série une fois ($d=1$).

Concernant le logarithme du PIB, il présente aussi une décroissance des coefficients jusqu'à un certains point, le lag 22 précisément, puis les coefficients deviennent négatifs et se rapprochent de un. Il y a aussi des pics ici, suggérant une certaine saisonnalité. Pour supprimer cette saisonnalité on va aussi différencier la série une fois ($d=1$). En résumé, l'analyse des corrélogrammes des série différenciées et des confirme que la série est intégrée d'ordre un. En d'autres termes, la série est non stationnaire dans sa forme originale, mais devient stationnaire après une seule différenciation. On peut quand même tester l'ordre d'intégration de notre série en utilisant les tests de racine unitaire.

Les tests de Dickey-Fuller augmentés (ADF) sont couramment utilisés pour tester la présence de racines unitaires dans les séries temporelles. Une racine unitaire indique que la série temporelle est non stationnaire. Les tests ADF comparent la série originale à une version transformée qui est régressée sur ses valeurs retardées et des tendances, permettant ainsi de déterminer si la série est stationnaire ou non. Pour le log du PIB, le test statistique (-2.8889) est plus grand que la valeur critique à 5% (-3.50). Nous ne pouvons pas rejeter l'hypothèse nulle d'une racine unitaire au seuil de signification de 5%. Cela suggère que la série log du PIB pourrait avoir une racine unitaire, ce qui signifie qu'elle présente une tendance non stationnaire. Concernant le taux de croissance, la régression du test indique que la tendance et la valeur retardée de la série ont un effet significatif sur la série différenciée. Le coefficient pour la tendance (tt) n'est pas significatif au niveau de 5%, mais il l'est au niveau de 10%, ce qui indique une possible tendance dans les données. On estime donc d'après ceci qu'une différenciation est bien nécessaire.

On ne peut pas rejeter l'hypothèse nulle (H_0) de stationnarité en niveau pour la série de croissance du PIB. En d'autres termes, selon le test de KPSS, la série de

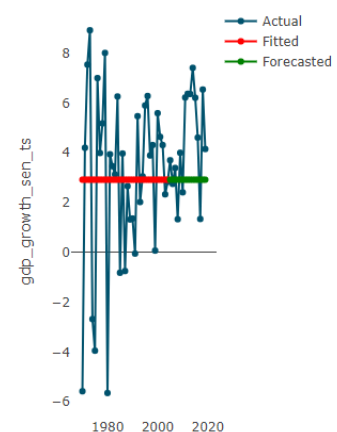
croissance du PIB **pourrait** être stationnaire en niveau, ce qui signifie que sa moyenne, sa variance et sa structure d'autocorrélation sont constantes au fil du temps.

6) Interprétation des résultats et recommandations

On a d'abord séparé notre base de donnée en deux, une partie pour l'entraînement du Arima et une autre pour tester la qualité du modèle. On s'est d'abord reposé sur le auto.arima qui est censé calculer les paramètres idéaux de notre modèle. Cependant les résultats obtenus ne convenaient pas du tout à ce qu'ils auraient pu.

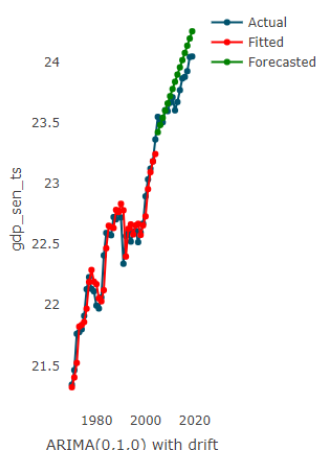
Pour le l'auto arima de la croissance du PIB, il nous indiquait un modèle avec 0 lag, 0 différenciation et 0 q. Les résultats données n'étaient absolument pas concluants. Avec un BIC et un AIC très élevé, le modèle ne pouvait pas être pris au sérieux. Quant à celui du PIB, l'auto ARIMA nous donne "ARIMA(0,1,0) with drift". Une

wth_sen_ts - Actual vs Forecasted ar



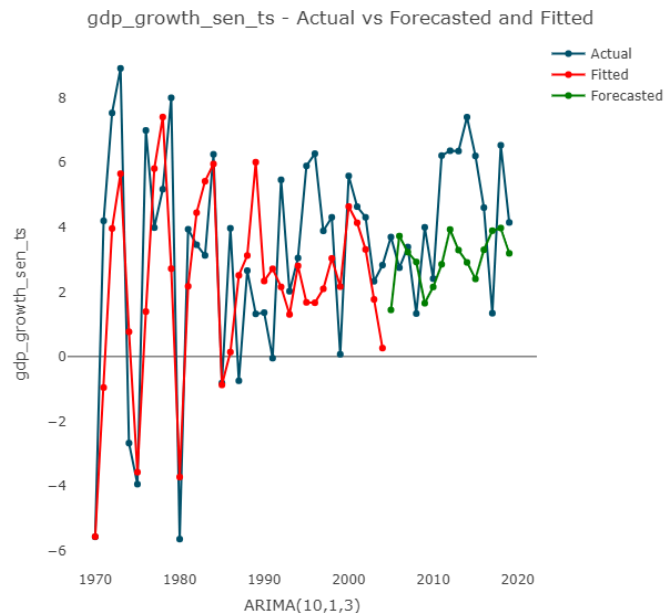
différenciation mais pas de lag ni de moving average. ARIMA(0,0,0) with non-zero mean

_sen_ts - Actual vs Forecasted and Fi

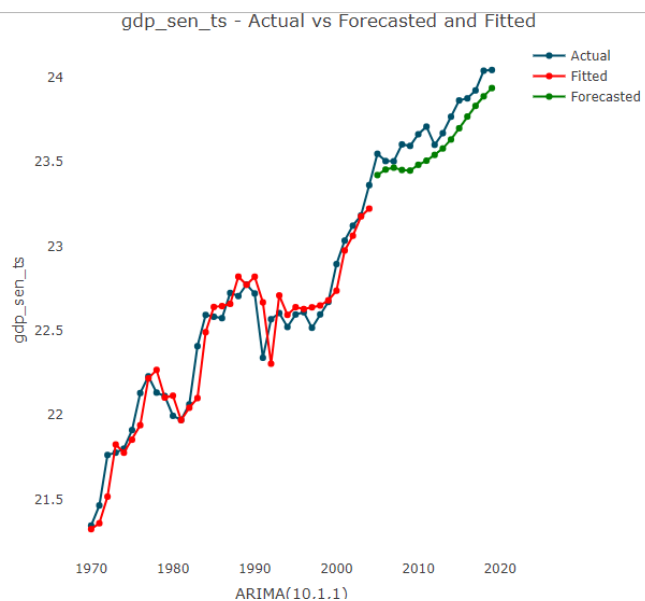


Les prédictions des données de test dans les deux cas par rapport aux données réelles étaient très décevantes par rapport à ce qu'on attendait.

En concordance avec nos tests effectués plus tôt, on a effectué cette fois-ci un ARIMA avec des ordres que nous avons déduit des tests. Principalement de la fonction d'autocorrélation, et de la fonction d'autocorrélation partielle. Ceux-là ont donné des AIC et BIC plus faibles mais surtout des données de prédiction bien plus



proches de la réalité.

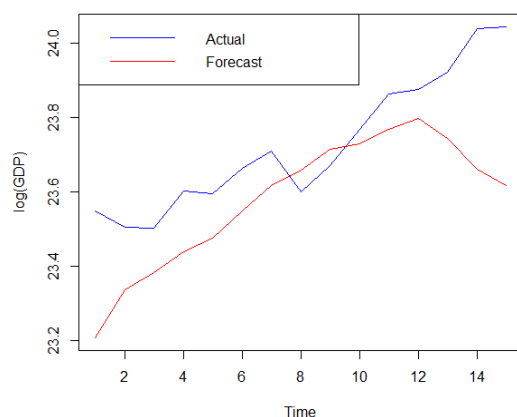


Mais peut-on être encore plus précis ? Les données démographiques que nous avons récoltées sont, nous le croyons, vecteurs de croissance ou de décroissance selon leur type. Peut-on obtenir une régression plus précise grâce à ces données, autrement nous pouvons nous démontrer une corrélation entre la croissance et l'âge de la population ? Le taux de fertilité ? D'autres variables démographiques ? On a

essayé de prévoir les données des derniers PIB du Sénégal avec d'autres variables à l'aide d'un modèle à correction d'erreur et d'une régression linéaire.

On a utilisé les variables démographiques que l'on a cité plus pour agrémenter notre modèle dont l'espérance de vie la pyramide des âges et... On a aussi utilisé

l'inflation pour contrôler tout cela.



Les valeurs pour l'EM, le RMSE, le MAE, le MPE et le MAPE sont tous plus bas par rapport aux valeurs de l'ensemble de test dans la première régression. Cela suggère un meilleur ajustement entre les valeurs prédites et réelles pour les données de test. La première métrique régression met en évidence une divergence significative entre

les performances d'entraînement et de test, soulevant des inquiétudes concernant la capacité du modèle à généraliser. La deuxième métrique montre des valeurs d'erreur plus faibles, suggérant un meilleur ajustement sur les données de test non observées.

On a quand même réalisé un Breusch-Pagan test pour voir s'il y a présence d'hétéroscédasticité. Étant donné que la valeur p (0.004141) est inférieure au niveau de significativité, on rejette l'hypothèse nulle d'homoscédasticité dans les résidus.. Cela signifie que la variance des erreurs n'est pas constante à travers toutes les observations, ce qui peut compromettre certaines des conclusions du modèle linéaire.

Puisque la valeur p est inférieure à 5%, on rejette l'hypothèse nulle selon laquelle il n'y a pas d'autocorrélation dans les résidus. Cela indique qu'il y a des preuves statistiques d'autocorrélation dans les résidus du modèle de régression.

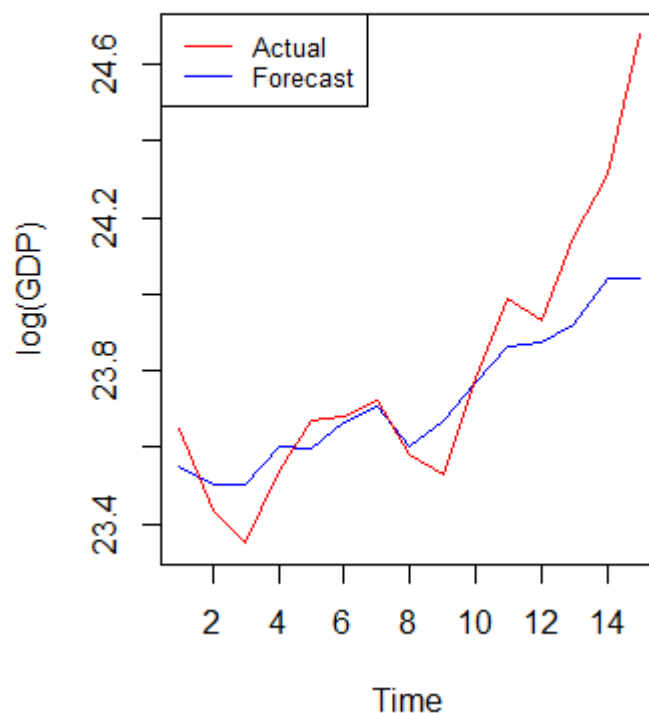
On a donc choisi de réessayer avec un modèle à correction d'erreur afin de capter un effet différent. L'accuracy de ce modèle est bien meilleur que celui des autres

comme on peut le voir dans le tableau ci-dessous.

Model	ME	RMSE	MAE	MPE	MAPE	Rang
ECM	-0.0727	0.2033	0.1331	-0.3018%	0.5579%	1
LM	0.1476	0.2002	0.1609	0.6201%	0.6767%	2
ARIMA	-193.84	19.47	19.38	-6100.54%	610.05%	3

Le modèle à correction d'erreur présente la plus faible erreur moyenne (ME) et un MAPE relativement bas, indiquant une bonne précision globale. Le premier donne la moyenne des erreurs de prédiction. Dans ce cas, la moyenne des erreurs est légèrement négative, ce qui indique que, en moyenne, les prédictions sont sous-estimées. Le second représente la moyenne des pourcentages d'erreur absolue par rapport aux valeurs réelles.

Les valeurs ici sont bien mieux calculées comme l'indique ce graphique.



Le modèle à correction d'erreur est donc bien le meilleur modèle pour analyser l'effet des variables démographiques sur la croissance.

Ce modèle possède un R carré ajusté très élevé ce qui confirme la bonne performance du modèle car il nous permet de prédire une bonne partie de la variable endogène.

Le test de Fisher est quant à lui très significatif, à moins de 1% ce qui signifie que le modèle globalement est statistiquement significatif pour expliquer notre variable endogène.

On va regarder quelques variables. L'intercept, la migration nette et l'espérance de vie sont significatifs à 5%, ils ont donc un réel effet sur notre variable endogène. Tandis que l'inflation est significative à 1%, elle joue donc bien son rôle qu'on avait prévu pour isoler l'impact des variables démographiques. On peut noter que ni les variables concernant la pyramide des âges, le taux de fertilité ou le lag diff du log du PIB ne sont significatives. On comprend donc qu'avec nos données on ne peut tirer de conclusion de l'effet de la pyramide d'âge sur la croissance. Néanmoins, on peut quand même regarder les signes de ces variables démographiques. Évidemment, l'augmentation de la masse de personnes en âge de travailler augmente le PIB, de même que pour l'espérance de vie ou la migration nette. Néanmoins on s'attendait à ce qu'une hausse du taux de fertilité par femme fasse diminuer le PIB, ce qui n'est visiblement pas le cas puisque le signe est positif. L'intercept est positif aussi.

Comme le montrent nos régressions, un pays avec une forte population en âge de travailler (population active) connaîtra certainement une croissance plus importante que certains pays avec une population très jeune ou très vieillissante. Une baisse de fertilité augmente aussi considérablement la croissance. Enfin, le vieillissement de la population ne permet pas de prédire la croissance du PIB, ici il est non significatif, malgré cela on voit bien qu'une hausse de la population de retraités diminue le PIB.

Au-delà de cela, l'analyse aurait pu être plus précise avec des données trimestrielles du PIB et de toutes les autres variables par exemple.

On ne peut pas conseiller éthiquement de faire moins d'enfant, néanmoins on peut amener ces pays à se poser des questions sur le système de santé car, une durée de vie plus longue va augmenter exponentiellement le PIB au final.