

« Une analyse économétrique des déterminants des hospitalisations dues à la Covid-19 »

Auteur

Francis Bismans

Document de Travail n° 2021 – 42

Octobre 2021

Bureau d'Économie
Théorique et Appliquée
BETA

www.beta-umr7522.fr

 @beta_economics

Contact :
jaoulgrammare@beta-cnrs.unistra.fr

Une analyse économétrique des déterminants des hospitalisations dues à la Covid-19

Francis Bismans

BETA, Université de Lorraine

COEF, Université Nelson Mandela, South Africa

Résumé

Ce document de travail s'appuie sur des données journalières relatives aux personnes vaccinées, à celles infectées par le coronavirus et aux hospitalisations pour tenter d'établir une relation de cointégration entre ces différentes variables. Les séries utilisées concernent la Belgique et vont du 1^{er} février 2021 au 20 septembre 2021.

Cette relation s'obtient à l'aide d'un modèle économétrique dynamique ARDL à correction d'erreur. Les principaux résultats de l'analyse sont les suivants : lorsque le nombre de vaccinés s'accroît de 1%, le nombre de personnes hospitalisées diminue, lui, de 0,24% ; quand celui des contaminations s'accroît de 1 pourcent, les hospitalisations augmentent de 1,05%.

Abstract

This working document is based on daily data on people vaccinated, those infected with the coronavirus and hospitalizations in an attempt to establish a cointegrating relationship between these different variables. The series used relate to Belgium and go from February 1, 2021 to September 20, 2021.

This relationship is obtained using an ARDL dynamic econometric model. The main results of the analysis are as follows: when the number of vaccinated increases by 1%, the number of people hospitalized decreases by 0.24%; when the infection number increases by 1 percent, hospitalizations increase by 1.05%.

Mots clés

Cointégration, Coronavirus, mécanisme à correction d'erreur, méthodologie GETS, modèle ARDL, pandémie, tests de racine unitaire

Codes *JEL*

C12, C13, C22, C51, C52

1. Introduction

L'existence des coronavirus constitue une donnée bien connue en infectiologie. Certains d'entre eux sont tout à fait inoffensifs ou causent de simples rhumes. D'autres, par contre, se traduisent par de graves pneumonies, parfois mortelles : en termes médicaux par un « syndrome respiratoire aigu sévère » (*Severe Acute Respiratory Syndrome*, SARS). En 2003, on avait d'ailleurs pu observer une véritable épidémie reliée à un tel coronavirus, ce dernier baptisé pour cette raison SARS-CoV.

Le 11 février 2020, un groupe d'experts mandatés par l'OMS avait établi que le coronavirus responsable du foyer survenu le 31 décembre 2019 à Wuhan, en Chine, appartenait à la famille des virus apparentés au SARS. D'où la dénomination de SARS-CoV-2. Le même jour, l'OMS annonçait que la COVID-19 serait le nom de la maladie causée par ce nouveau coronavirus.

L'expansion de l'épidémie à l'échelle mondiale fut foudroyante, au point que l'OMS décréta le 11 mars 2020 que la COVID-19 pouvait être qualifiée de véritable pandémie. Sa sévérité et sa létalité étaient déjà alors bien établies. Une étude réalisée par l'*Imperial College* de Londres en février de cette même année 2020 – voir Ferguson et al., 2020 ainsi que la synthèse de l'étude réalisée par Bismans et al., 202, pp. 19 et sq. – avait montré, sur l'exemple des Etats-Unis et de la Grande Bretagne, que le laissez-faire face à la propagation du virus conduirait à 2.200.000 morts aux USA et à 510.000 au Royaume Uni, sans parler de l'engorgement des hôpitaux et des unités de soins critiques.

Dès lors, en l'absence de traitements efficaces et aussi à l'époque, de vaccins, la seule manière de contenir la pandémie et d'éviter la saturation des hôpitaux, consista à introduire des formes plus ou moins rigoureuses de confinement dès le mois de mars 2020. Ces *lockdowns* ont plongé l'ensemble des économies dans une récession profonde (sur ce point, consulter Bismans et Majetti, 2021) et provoqué également un recul généralisé des cours boursiers, bien analysé par Bagchi et al. (2020).

Le relâchement progressif des mesures de confinement a cependant conduit à une deuxième vague pandémique dès l'automne 2020, vague qui sera contenue par la réintroduction de formes partielles de confinement. Pour autant, un élément nouveau survint lorsque les premiers vaccins furent disponibles au tournant des années 2020 et 2021. Suivirent alors, du moins dans tous les pays économiquement avancés, des campagnes de vaccination massive de la population.

L'objet de ce document de travail est d'analyser économétriquement la relation entre les admissions à l'hôpital consécutivement à la pandémie, le nombre cumulé de personnes vaccinées et le nombre de cas d'infection par le virus effectivement constatés. L'étude se mène en estimant un modèle ARDL (*autoregressive distributed lags*), qui fournit comme sous-produit la relation en question.

Voici le plan retenu. La section 2 présente les séries utilisées et teste leur non-stationnarité. La section suivante traite du modèle ARDL mis en œuvre, qui incorpore à la fois une relation de cointégration et un mécanisme à correction d'erreur. Elle spécifie ensuite ce modèle pour l'étude de la variable « nombre d'hospitalisations ».

La section 4 donne tous les résultats fournis par l'estimation du modèle final obtenu à l'issue de l'application de la méthodologie dite « du général au spécifique », tandis que la dernière

section conclut et souligne les implications empiriques des estimations économétriques réalisées.

2. Les séries utilisées et leurs caractéristiques

Nous commencerons par décrire les séries utilisées, après quoi nous examinerons leurs différentes propriétés, notamment en matière de stationnarité ou de non-stationnarité (au second ordre).

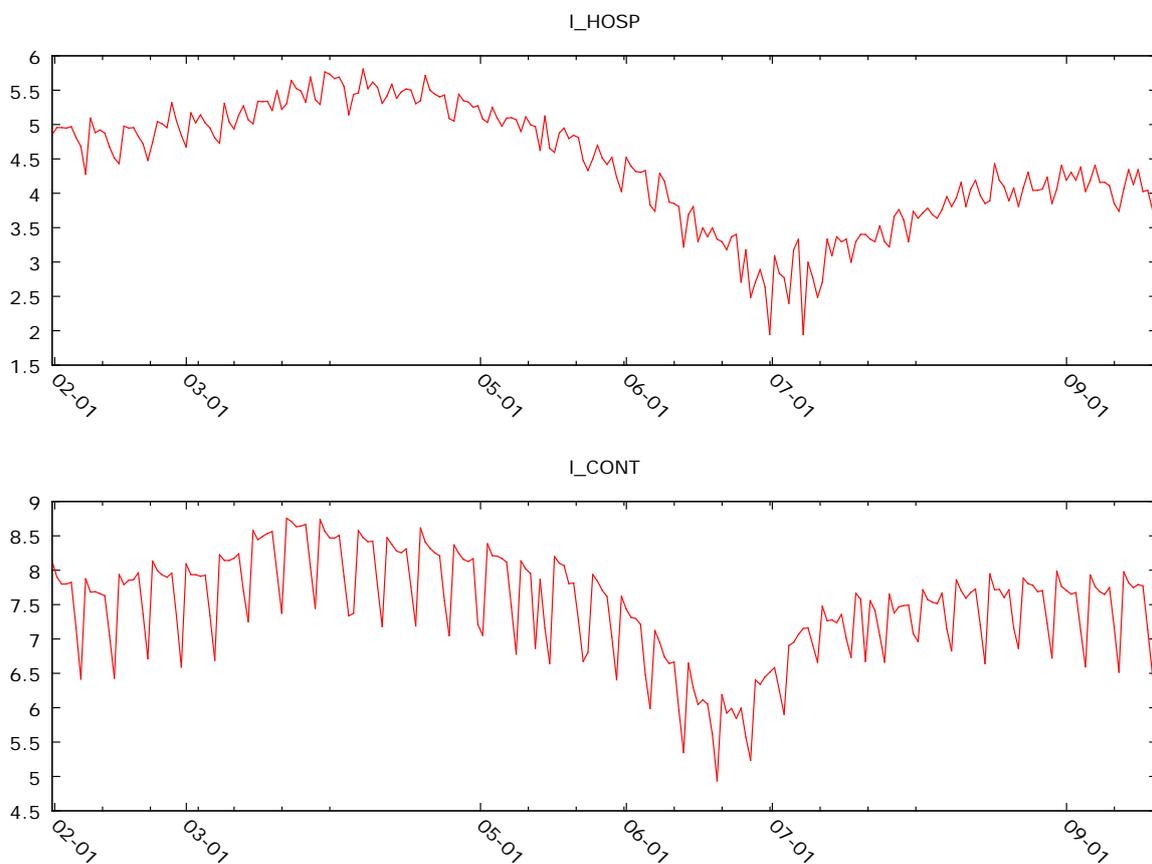
2.1. Trois séries chronologiques

Les chroniques sont celles du nombre d'hospitalisations pour cause de Covid (notée LHOSP, le « L » dans cette désignation signifiant le logarithme base e), du nombre total de personnes complètement vaccinées (donc deux doses) contre la pandémie, cette variable étant notée LVAC, et enfin, du nombre de contaminations, désigné par LCONT.

Ajoutons que ces données, qui concernent la Belgique exclusivement, sont journalières (les 7 jours de la semaine) et couvrent la période allant du 1^{er} février 2021 jusqu'au 20 septembre de la même année. Un tel intervalle temporel a été retenu, parce que le nombre de personnes vaccinées complètes ne devient significatif qu'à partir du mois de février.

Voici un graphe comportant les données brutes d'une part ainsi que des différences premières des séries précédentes.

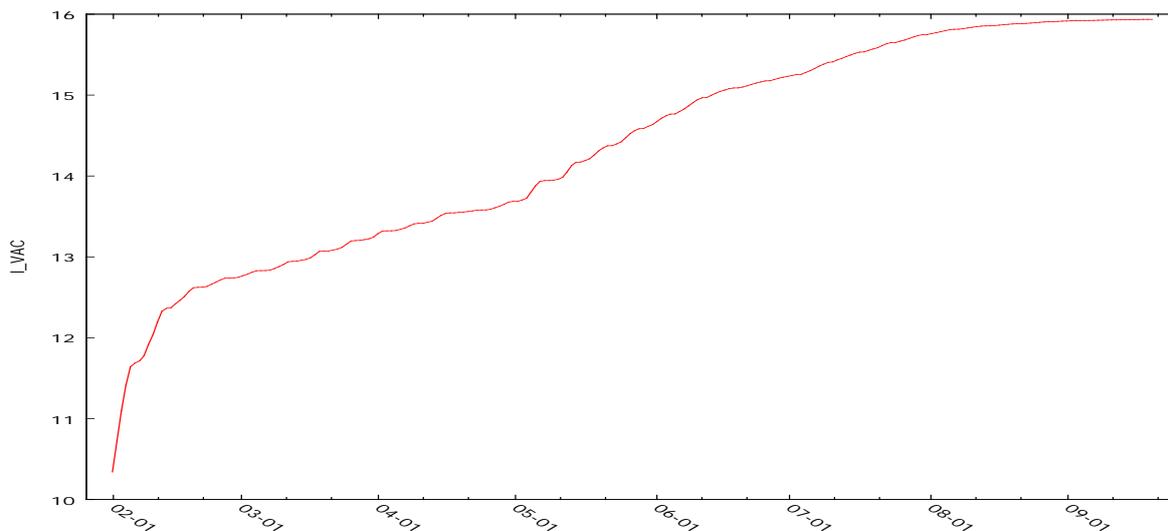
Figure 1. Représentation graphique des séries LHOSP et LCONT



Le profil temporel de ces deux variables est en définitive fort semblable. On observe d’abord une augmentation (assez faible) des hospitalisations et des contaminations jusqu’en avril, après quoi les deux séries décroissent lentement jusqu’en juin. Dès le mois de juillet, avec le déconfinement et les vacances, contaminations et hospitalisations repartent à la hausse, mais sans jamais revenir aux maxima observés durant le premier trimestre de 2021.

La chronique du nombre cumulé de personnes complètement vaccinées présente évidemment un trend, une tendance croissante comme l’indique clairement le graphique 2. C’est ainsi que dans la série brute, ce nombre est passé de 30.972 le 1^{er} février à 8.332.836 le 20 septembre 2021.

Graphique 2. Graphe de LVAC



2.2. Tester la non-stationnarité

Les tests dits de racine unitaire sont nombreux. Nous nous contenterons ici d’utiliser deux d’entre eux tout en renvoyant à Patterson (2010) pour les développements les plus techniques sur lesquels ces tests reposent. En l’occurrence, il s’agit :

- du test de Dickey-Fuller Augmenté (ADF), augmenté parce qu’il consiste à ajouter des termes autorégressifs au Dickey-Fuller original ;
- du test d’Elliot-Rothenberg-Stock (ERS), qui réalise une transformation des données initiales via une régression par moindres carrés généralisés.

Toutefois, eu égard à la présence de ruptures, spécialement dans la série LHOSP, nous réaliserons également un test de racine unité avec rupture structurelle. Ces tests sont nombreux, la première contribution majeure étant celle de Perron (1989). Autre contribution importante, celle de Zivot-Andrews (1992) qui ne préjuge pas de la date à laquelle survient une rupture. Perron (2007) fournit une synthèse commode des écrits sur le sujet.

Sans entrer dans trop de détails, on utilisera ici exclusivement le test associé à Perron (1997). L’idée de base est simple. D’abord, il faut distinguer deux modélisations possibles selon que les valeurs aberrantes (*outliers*) se présentent sous la forme additive ou innovationnelle. Ensuite la détermination des dates inconnues se réalise à l’aide du critère de la minimisation d’une statistique t associée au test de la présence d’une racine unité.

Enfin, pour en terminer, il faut également ajouter que les trois tests mis en œuvre ont tous comme hypothèse nulle la non-stationnarité de la série analysée, l'hypothèse alternative étant celle de stationnarité ou encore dans le dernier cas, celle de stationnarité avec rupture structurelle.

Voici un tableau qui regroupe les différents tests de racine unité mis en œuvre. Précisons que LHOSP et LCONT incorporent tous deux une constante, tandis qu'on a ajouté un trend pour la troisième variable. Enfin, les statistiques de tests sont mentionnées et les p-valeurs asymptotiques correspondantes sont données entre parenthèses sous chacune de ces statistiques, tandis que pour les tests avec rupture structurelle, la date associée à cette rupture est indiquée.

Tableau 1. Tests de racine unitaire

Variables	Test ADF	Test ERS	Perron
LHOSP	-1,3 (0,63)	-1,05 (0,27)	2,72 (21/05/2021) (0,82)
LVAC	0,94 (0,99)	-1,62 (> 0,1)	-1,05 (29/07/2021) (0,98)
LCONT	-1,97 (0,3)	-1,46 (0,14)	-3,43 (20/05/2021) (0,2)

La conclusion à tirer de cet ensemble de tests est sans ambiguïté : les p-valeurs asymptotiques sont toutes supérieures à 10%, de sorte que l'hypothèse nulle d'une racine unité est acceptée. En conséquence, toutes les variables retenues sont *intégrées du même ordre*, soit un, et l'on note donc que

$$\text{LHOSP, LVAC, LCONT} \sim I(1).$$

3. Le modèle économétrique

Le modèle à estimer est un modèle autorégressif à retards échelonnés (ARDL), qui fournit la relation de cointégration recherchée tout en incorporant explicitement un mécanisme à correction d'erreur.

3.1. Relation de cointégration et correction d'erreur

On sait depuis Engle et Granger (1987) qu'une combinaison de variables $I(1)$ est en général non stationnaire, mais qu'il peut se faire que cette combinaison soit stationnaire, c'est-à-dire intégrée d'ordre zéro, ce que l'on nomme une relation de cointégration. De plus, ces mêmes auteurs ont démontré qu'une telle relation était nécessairement associée à une représentation à correction d'erreur.

Au départ, c'est Denis Sargan (1964) qui a été le premier à utiliser un mécanisme à correction d'erreur (MCE). L'école groupée autour de David Hendry s'en est ensuite servi à de multiples reprises – voir en particulier les précisions dans Hendry (2000).

Quoi qu'il en soit, le modèle ARDL (p, k_1, \dots, k_K) peut s'écrire sous la forme générale

$$\Delta y_t = \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \sum_{j=0}^{k_1} \beta_{1j} \Delta x_{1,t-j} + \dots + \sum_{j=0}^{k_k} \beta_{Kj} \Delta x_{K,t-j} - (1 - \alpha_1 - \dots - \alpha_p) \left[y_{t-1} - a - \sum_{r=1}^K \gamma_r x_{r,t-1} \right] + \varepsilon_t, \quad (1)$$

où les deltas représentent des différences premières de la variable considérée et ε_t est un bruit blanc gaussien, de moyenne nulle et de variance constante.

L'expression entre crochets dans (1) représente la relation de cointégration entre les variables exprimées en niveau, par ailleurs toutes intégrées d'ordre un, tandis que les autres termes donnent la dynamique de court terme du modèle et constituent ainsi le mécanisme à correction d'erreur.

Il existe bien sûr d'autres techniques que l'utilisation d'un modèle ARDL pour déterminer une relation de cointégration. La première d'entre elles a été préconisée par Engle et Granger (1987) ; elle n'est toutefois plus guère utilisée de nos jours. La deuxième est due à Pesaran et al. (2001). Elle exige également l'estimation d'un modèle ARDL, mais entre des variables qui sont exprimées en niveau et dont on ne sait si elles sont intégrées d'ordre zéro ou d'ordre un. (Nous ne nous en servons pas, dans la mesure où nous avons vérifié que nos propres variables étaient toutes $I(1)$.)

Enfin, il existe une dernière approche, qui est associée aux travaux de Johansen (1988). Elle est de nature essentiellement multidimensionnelle et conduit, en conséquence, à l'estimation d'un *vecteur à correction d'erreur*. En pratique, elle part du fait que lorsque le modèle comporte plus de deux variables, il peut exister plus d'une relation de cointégration. L'approche de Johansen revient alors à déterminer le nombre de ces relations, puis ensuite à les estimer. Nous utiliserons cette méthode à des fins de comparaison avec nos propres résultats dans la sous-section 4.2 de ce document de travail.

3.2. Spécification du modèle ARDL

Partons d'abord d'un modèle autorégressif à retards échelonnés (ARDL) en niveau, donc dynamique, dans lequel LHOP est la variable endogène, LVAC et LCONT les variables explicatives, tout en y ajoutant une constante m , soit

$$LHOP_t = m + \sum_{i=1}^p \alpha_i LHOP_{t-i} + \sum_{j=0}^{k_1} \beta_{1j} LVAC_{1,t-j} + \sum_{j=0}^{k_2} \beta_{2j} LCONT_{2,t-j} + \varepsilon_t, \quad \varepsilon_t \sim bb(0, \sigma^2). \quad (2)$$

On peut, dans un deuxième temps, respécifier ce modèle de manière à obtenir l'équation suivante :

$$\Delta LHOP_t = \sum_{i=1}^p \alpha_i \Delta LHOP_{t-i} + \sum_{j=0}^{k_1} \beta_{1j} \Delta LVAC_{1,t-j} + \sum_{j=0}^{k_2} \beta_{2j} \Delta LCONT_{2,t-j} - (1 - \alpha_1 - \dots - \alpha_p) \left[LHOP_{t-1} - a - \gamma_1 LVAC_{1,t-1} - \gamma_2 LCONT_{2,t-1} \right] + \varepsilon_t, \quad (3)$$

où les variables initiales sont exprimées en différences premières. Cette dernière équation a la structure d'un modèle ARDL (p, k_1, k_2) .

Pour estimer le modèle (3), qui comporte donc à la fois une relation de cointégration entre les trois variables et un mécanisme à correction d'erreur, nous mettrons en œuvre la méthodologie

dite du « Général au Spécifique » (GETS, acronyme de *General To Specific*). Une telle approche est préconisée par David Hendry (voir par exemple Hendry et Doornik, 2014). Elle peut être entièrement automatisée.

Brièvement dit, la méthodologie GETS consiste à partir d'un modèle très général avec de nombreux décalages pour chacune des variables retenues, puis à réduire ce nombre de variables sur base de trois critères tout en observant la significativité des coefficients estimés :

- significativité du test F après élimination de chaque ensemble de variables ;
- amélioration des trois critères d'information (Akaike, Schwartz bayésien, Hannan-Quinn) ;
- augmentation du R^2 corrigé lors de chaque phase de la réduction séquentielle du modèle initial.

Le modèle final est bien spécifié lorsqu'il n'est plus possible d'opérer une réduction additionnelle sans détériorer les critères retenus.

4. Résultats d'estimation

Nous commencerons par donner les estimateurs des différents paramètres du modèle terminal, après quoi nous nous occuperons plus spécialement de la relation de cointégration entre le nombre d'hospitalisations et les deux autres variables.

4.1. Le modèle final

En nous référant à l'équation (3), initialement, nous avons choisi $p = k_1 = k_2 = 7$, compte tenu que les données utilisées étaient journalières. À l'issue des opérations de réduction séquentielle, les paramètres estimés par moindres carrés ordinaires se présentent comme suit :

Tableau 2. Paramètres estimés

Variables	Coefficients	Statistique t	p-valeur
LHOSP (-1)	- 0,252	- 11,11	< 0,01
LVAC (-1)	- 0,061	- 10,40	< 0,01
LCONT (-1)	0,265	11,69	< 0,01
DLHOSP (-1)	- 0,497	- 9,72	< 0,01
DLHOSP (-2)	- 0,317	- 5,66	< 0,01
DLHOSP (-3)	- 0,139	- 2,07	0,03
DLCONT (-2)	- 0,135	- 5,61	< 0,01
DLCONT (-3)	- 0,127	- 5,74	< 0,01
DLCONT (-4)	- 0,178	- 5,63	< 0,01
DLCONT (-5)	- 0,108	- 4,71	< 0,01
DLCONT (-6)	- 0,118	- 5,97	< 0,01
BIN1	- 0,976	- 28,92	< 0,01
BIN2	0,563	9,97	< 0,01
BIN3	- 0,173	- 26,06	< 0,01
R^2 corrigé = 0,75		$F(14, 211) = 2,84 e^{+16}$, p-valeur = 0,000000	

Remarquons d'abord que tous les coefficients sont significativement différents de zéro, leur p-valeur étant inférieure à 1%, à l'exception de celui de DLHOSP (-3), dont la p-valeur est égale à 3%. Remarquons aussi que le coefficient de détermination est égal à 75%, ce qui est

satisfaisant, tandis que le test F conduit au rejet de l'hypothèse de nullité de l'ensemble des coefficients du modèle.

Pour le reste, voici quelques remarques qui résultent de la lecture du tableau :

- (i) la constante est absente des paramètres estimés, eu égard à ce qu'elle n'est pas significativement différente de zéro.
- (ii) la variable LVAC n'intervient pas dans le mécanisme à correction d'erreur ; par contre, les variables DLCONT, moyennant des décalages compris entre 2 et 6 jours, et DLHOSP, retardée de 1 à 3 jours, jouent un rôle majeur dans l'ajustement à court terme du modèle à la relation d'équilibre décrite par les trois premières variables du tableau. À noter d'ailleurs que les paramètres estimés correspondant à ces variables décalées sont tous négatifs, de sorte que l'accroissement du nombre d'hospitalisations est corrigé négativement par rapport à ce que cet accroissement serait en fonction de la seule relation de cointégration.
- (iii) enfin, les trois variables binaires (*dummy variables*), notées BIN1, BIN2 et BIN3, correspondent à une rupture intervenant respectivement le 1^{er} juillet, le 2 juillet et le 8 juillet de l'année 2021 ; elles sont toutes trois pleinement significatives.

Avant de passer à l'analyse de la relation d'équilibre, il faut encore préciser que le modèle a été soumis à plusieurs tests qui, pour la plupart, confortent l'équation estimée :

- un test de log-linéarité, basé sur des régressions auxiliaires, rend la valeur $TR^2 = 3,74$, la p-valeur associée étant égale à 0,29, ce qui conduit à la conclusion que la forme fonctionnelle log-linéaire est appropriée.
- le test de Harvey-Collier, dont l'hypothèse nulle est la stabilité structurelle des paramètres estimés, donne la statistique $t(210) = 0,279$, avec une p-valeur égale à 0,29 ; on est donc conduit à accepter cette hypothèse nulle ; l'application du test CUSUM aboutit au même résultat de stabilité structurelle des paramètres.
- l'absence de multicollinéarité entre les 13 variables du modèle final est avérée au vu des facteurs d'augmentation de la variance (*Variance Inflation Factors*), qui sont tous inférieurs à 10.
- Le test de Bera-Jarque avec comme hypothèse nulle la normalité des résidus d'estimation, rend les résultats suivants :

$$\chi^2(2) = 2,643 \quad \text{p-valeur} = 0,27.$$

On en déduit que ces résidus sont normalement distribués.

Par contre, il existe un effet ARCH dans les résidus ($LM = TR^2 = 29,39$ p-valeur = 0,009). En d'autres termes, leur variance est non constante dans le temps, hétéroscédastique.

4.2. La relation de cointégration

En faisant fond sur le tableau 2, la relation d'équilibre statique – nous supprimerons donc toute référence temporelle dans la notation des variables – qui résulte de l'estimation du modèle ARDL, est la suivante :

$$LHOSP = -0,241 LVAC + 1,052 LCONT. \quad (4)$$

Eu égard à ce que les trois variables sont exprimées en logarithmes, les coefficients de LVAC et LCONT représentent des élasticités, c'est-à-dire des rapports de taux de variation relative. Dès lors, deux remarques s'imposent :

1. lorsque le nombre de personnes vaccinées s'accroît de 1 pourcent, alors le nombre de personnes hospitalisées diminue, lui, d'environ un quart de pourcent.
2. par contre, une augmentation de 1 pourcent des cas de contamination se traduit par un accroissement des hospitalisations d'un peu plus d'un pourcent ; à l'inverse, si les contaminations se réduisent d'un pourcent, alors les hospitalisations diminueront d'un pourcentage légèrement supérieur.

Nous allons à présent comparer ces résultats avec ceux obtenus en appliquant la procédure multivariée préconisée par Johansen (1988, 1995). Cette dernière commence par calculer deux statistiques de tests destinés à déterminer le nombre de relations de cointégration – le *rang de cointégration* du vecteur considéré dans les termes de Johansen – entre un ensemble de n variables non stationnaires et intégrées du même ordre : il s'agit en l'occurrence du test de la trace et de celui de la valeur propre maximale.

Le tableau 3 fournit les deux statistiques et les p-valeurs correspondantes, sachant que dans le cas qui nous occupe, trois variables sont prises en considération, de sorte qu'il n'existe au maximum que deux relations de cointégration.

Tableau 3. Tests du rang de cointégration

Rang	Stat. Trace	p-valeur	Stat. VP max	p-valeur
0	120,52	0,0000	98,88	0,0000
1	21,64	0,0009	19,23	0,0012
2	2,41	0,1416	2,41	0,1429

Il ressort des deux tests qu'il existe au plus une relation de cointégration entre les trois variables LHOSP, LVAC et LCONT.

La méthode de Johansen permet ensuite, dans un second temps, d'estimer la relation de cointégration qui a été isolée à l'étape précédente. Voici les estimateurs qui en résultent tout en précisant que pour faciliter la comparaison, nous avons ajouté dans le tableau qui suit nos propres estimateurs, tels qu'ils figurent dans l'équation 4.

Tableau 4. Comparaison des estimations

	LHOSP	LVAC	LCONT
ARDL	1	- 0,241	1,052
Johansen	1	- 0,263	1,112

À l'examen du tableau, il apparaît immédiatement que les estimations réalisées en suivant la procédure multivariée de Johansen sont fort proches de celles obtenues à l'aide du modèle ARDL. (Les deux élasticités sont tout au plus légèrement supérieures dans le premier cas tout en conservant le même signe.)

Ce fait à lui seul indique bien la robustesse de l'estimation de la relation de cointégration entre les trois variables, qui est le sous-produit de notre modèle économétrique dynamique à correction d'erreur.

5. Conclusions

Deux implications empiriques découlent des valeurs prises par les élasticités des hospitalisations par rapport à la vaccination et aux contaminations. D'une part, accroître le nombre de personnes complètement vaccinées diminuera les hospitalisations pour cause de Covid-19. D'autre part, puisque les gestes barrières et le port du masque chirurgical, en tout cas en intérieur, réduisent, tout comme le vaccin, les contaminations, il s'ensuit que le respect de ces mesures de précaution sanitaire est de nature à diminuer le nombre d'hospitalisations.

Sur un plan économétrique plus général, un débouché naturel de ce travail consisterait à prolonger l'analyse en envisageant des séries de données de panel, où l'on considérerait un ensemble de pays et une période temporelle déterminée.

Références bibliographiques

- Bagchi, B., Chatterjee, S., Ghosh, R. and Dandapat, D. (2020), *Coronavirus Outbreak and the Great Lockdown. Impact on Oil Prices and Major Stock Markets Across the Globe*, Springer, Singapore.
- Biesmans, F., Dubois, D. et Renda, J.C. (2021), *Comprendre le coronavirus*, Éditions du Coq, Flémalle.
- Bismans, F. et Majetti, R. (2021), « Marchés financiers, dette et récession en période de Covid-19 : l'exemple de la zone euro », in Eboué, C. (dir.), *Finance, banque, microfinance. Où va la richesse créée ?*, Larcier, Bruxelles.
- Engle, R.F. and Granger, C.W.J. (1987). « Co-integration and Error Correction: Representation, Estimation, and Testing », *Econometrica*, 55, pp. 251-276.
- Ferguson, N. M., Laydon, D. *et al.* [2020], « Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand », Imperial College, London, 16-03-2020, 20 pages. Doi: <https://doi.org/10.25561/77482>.
- Hendry, D.F. (2000). *Econometrics. Alchemy or Science ?*, Oxford University Press, Oxford.
- Hendry, D.F. and Doornik, J.A. (2014). *Empirical Model Discovery and Theory Evaluation. Automatic Selection methods in Econometrics*, Cambridge-London, Massachusetts Institute of Technology Press.
- Johansen, S. (1988). « Statistical Analysis of Cointegration Vectors », *Journal of Economic Dynamics and Control*, 12, pp. 231-254.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- Patterson, K. (2010). *A Primer for Unit Root Testing*, Palgrave Macmillan, Basingstoke-New York.
- Perron, P. (1989), « The Great Crash, the Oil Price Shock and the Unit Root Hypothesis », *Econometrica*, 57, pp. 1361-1401.
- Perron, P. (1997), « Further Evidence of Breaking Trend Functions in Macroeconomic Variables », *Journal of Econometrics*, 80, pp. 355-385.
- Perron, P. (2007). Dealing with Structural Breaks, in Mills, T.E. and Patterson, K. (eds), *Palgrave Handbook of Econometrics*, Palgrave-Macmillan, Basingstoke-New York, volume 1, pp. 278-352.
- Pesaran, H., Shin, Y. and Smith, R.J. (2001). « Bounds Testing Approaches to the Analysis of Level relationships », *Journal of Applied Econometrics*, 16, pp. 289-326.
- Sargan, D. (1964), « Wages and Prices in the United Kingdom: A Study in Econometric Methodology », in Hart, P.E., Mills, G. and Whitaker, J.K. (eds), *Econometric Analysis for National Economic Planning*, Butterworth, London, pp. 25-63.

- Stock, J. and Watson, M.W. (1993). « A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems », *Econometrica*, 61, pp. 783-820.
- Zivot, E., & Andrews, D.W.K. (1992). « Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis », *Journal of Business and Economic Statistics*, 10, pp. 251-270.