

THESIS / THÈSE

MASTER EN INGÉNIEUR DE GESTION À FINALITÉ SPÉCIALISÉE EN ANALYTICS & DIGITAL BUSINESS

Modélisation des hospitalisations associées au Covid-19 à l'aide de la Théorie des Valeurs Extrêmes

WALGRAFFE, Nicolas

Award date:
2023

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Modélisation des hospitalisations associées au Covid-19 à l'aide de la Théorie des Valeurs Extrêmes

Nicolas WALGRAFFE

Directeur : Prof. A.KIRILIOUK

Mémoire présenté
en vue de l'obtention du titre de
Master 120 en ingénieur de gestion, à finalité spécialisée
en Analytics & Digital Business

ANNEE ACADEMIQUE 2022-2023

Remerciements

Je voudrais commencer par exprimer ma profonde gratitude à mon directeur de mémoire, le Professeure Kiriliouk, pour son soutien, ses conseils avisés et son expertise inestimable tout au long de ce processus. Le chemin jusqu'à ce point n'aurait pas été possible sans sa précieuse assistance et sa grande disponibilité.

Je tiens à remercier sincèrement mes camarades de classe pour les discussions stimulantes, les échanges d'idées et les précieux conseils qu'ils m'ont prodigués tout au long de mes études.

Je souhaite également exprimer ma gratitude à ma famille et à mes amis pour leur soutien indéfectible, les encouragements fréquents et la motivation transmise.

Enfin, je suis reconnaissant envers l'Université de Namur d'avoir mis à ma disposition les ressources nécessaires à la réalisation de ce mémoire et d'avoir créé un environnement propice à la réalisation de ce travail.

Finalement, mes remerciements s'adressent à tous ceux qui ont contribué, de près ou de loin, à l'élaboration de ce mémoire. Votre soutien a été inestimable et a joué un rôle crucial dans sa réalisation, finalisant mon parcours académique.

Merci du fond du cœur,
Nicolas Walgraffe

Résumé

Le présent mémoire explore le comportement extrême d'un échantillon d'observations portant sur les hospitalisations quotidiennes associées au Covid-19 en Belgique et au Portugal. En recourant à la Théorie des Valeurs Extrêmes (TVE), le but est d'estimer si le risque d'une recrudescence pandémique de Coronavirus est à craindre au cours des dix prochaines années. Une distribution dite « Pareto Généralisée » (GPD) est ajustée aux données historiques extrêmes dépassant un seuil sensible en vue d'estimer les paramètres du modèle créé. A partir de ces paramètres, des niveaux de retour de la pandémie à moyen terme sont calculés pour répondre à notre question de recherche. Une seconde distinction est apportée afin d'évaluer l'influence du type de mesures gouvernementales, définies comme « fortes » ou « faibles », sur la probabilité d'être confronté une nouvelle fois à de fortes augmentations du nombre d'hospitalisations. Nous serons en mesure de conclure que la Belgique est plus susceptible de faire face à des augmentations spectaculaires du niveau d'hospitalisations mais que dans le même temps, notre pays peut considérablement diminuer ce risque étant donné sa meilleure gestion politique par rapport au Portugal.

Summary

This thesis explores the extreme behaviour of a sample of observations of daily hospitalisations associated with Covid-19 in Belgium and Portugal. Using Extreme Value Theory (EVT), the aim is to estimate whether there is a risk of a pandemic outbreak of Coronavirus over the next ten years. A so-called "Generalized Pareto Distribution" (GPD) is fitted to extreme historical data exceeding a sensitive threshold in order to estimate the parameters of the model created. Based on these parameters, medium-term pandemic return levels are calculated to answer our research question. A second distinction is made in order to assess the influence of the type of government measures, defined as "strong" or "weak", on the probability of being exposed once again to sharp increases in the number of hospital admissions. We will be able to conclude that Belgium is more likely to face spectacular increases in the level of hospitalisations, but at the same time, our country can drastically reduce this risk given its better political management than Portugal.

Table des matières

1	Introduction	7
1.1	Contexte.....	7
1.2	Importance et justification de la thématique	7
1.3	Problématiques centrales et questions de recherche	8
1.4	Synthèse de la méthodologie suivie.....	8
1.5	Plan du mémoire	9
2	Analyse descriptive des données	10
2.1	Données épidémiologiques : nombre d'hospitalisations liées au Covid-19	10
2.1.1	Source de la base de données	10
2.1.2	Variable étudiée.....	10
2.1.3	Description des paramètres centraux et de dispersion	10
2.1.4	Séries temporelles	12
2.1.5	Distributions des données	13
2.1.6	Cas des différences logarithmiques d'hospitalisations quotidiennes.....	14
2.2	Données gouvernementales : indice de sévérité des mesures non-pharmaceutiques ... 15	
2.2.1	Source de la base de données	15
2.2.2	Variable étudiée.....	15
2.2.3	Description des paramètres centraux et de dispersion	15
2.2.4	Séries temporelles	16
3	Introduction à la Théorie des Valeurs Extrêmes	17
3.1	Objectifs et spécificités de la TVE.....	17
3.2	Types d'approche.....	17
3.3	Méthode du dépassement de seuil	18
3.3.1	Modélisation de la fonction de distribution des valeurs extrêmes	18
3.3.2	Choix du seuil.....	19
3.3.3	Estimation des paramètres.....	19
3.3.4	Estimation des niveaux de retour	20
3.3.5	Vérification de la validité du modèle GPD	21
4	Revue de la Littérature	22
4.1	Applications de la TVE dans le domaine des soins de santé	22
4.1.1	Modélisation de la mortalité liée à la pneumonie et à la grippe	22
4.1.2	Modélisation des visites aux urgences dans un réseau hospitalier	22
4.1.3	Modélisation des pics de mortalité et de morbidité pour les maladies cardiovasculaires.....	22
4.1.4	Modélisation de données Covid-19 : les évènements de super-propagation	23
4.2	Analyse univariée : Indice composite de sévérité des mesures	25
4.2.1	Construction de l'Oxford Covid-19 Government Response Tracker	25
4.2.2	Constats relatifs à l'Ox-CGRT	26
4.2.3	Analyse géographique de l'Oxford COVID-19 Government Response Tracker	27
4.3	Analyse multivariée : Lien entre la sévérité des mesures gouvernementales et le nombre de décès Covid-19	29
4.3.1	Analyse de la corrélation par région du monde	29
4.3.2	Analyse de la corrélation par vague épidémiologique	30
5	Résultats de l'analyse des données extrêmes	33
5.1	Phase de préparation des données	33

5.1.1	Transformation de la variable « Hospitalisations quotidiennes » en différences logarithmiques	33
5.1.2	Application d'un décalage temporel.....	33
5.1.3	Sélection des différences logarithmiques stables.....	34
5.1.4	Séparation des « Hospitalisations quotidiennes » en fonction de l'intensité des mesures gouvernementales	34
5.2	Résultats issus de la Théorie des Valeurs Extrêmes	34
5.2.1	Analyse des hospitalisations	34
5.2.1.1	Détermination du seuil.....	34
5.2.1.2	Estimation des paramètres du modèle	36
5.2.1.3	Estimation de probabilités d'occurrences extrêmes et de niveaux de retour	37
5.2.2	Analyse des hospitalisations en fonction des mesures gouvernementales.....	39
5.2.2.1	Détermination du seuil.....	39
5.2.2.2	Estimation des paramètres du modèle	41
5.2.2.3	Estimation de probabilités d'occurrences extrêmes et de niveaux de retour	41
6	Conclusion	44
6.1	Synthèse des découvertes et des résultats.....	44
6.2	Réponses aux questions de recherche	44
6.3	Limitations de l'étude	45
7	Bibliographie	47
8	Annexes.....	50

Liste des abréviations utilisées dans ce mémoire

- BM : Block Maxima, méthode du maxima de bloc
- ECDC : Centre Européen de Prévention et de Contrôle des Maladies
- EEE : Espace Européen Economique
- Et Al. : abréviation latine qui signifie « et autres »
- GEV : Generalized Extreme Value distribution, distribution généralisée des valeurs extrêmes
- GPD : Generalized Pareto Distribution, distribution dite « Pareto Généralisée »
- Ox-CGRT : Oxford Covid-19 Government Response Tracker, le traceur de la réponse gouvernementale face au Covid-19, conçu par une équipe de l'Université d'Oxford
- POT : Peaks Over Threshold, la méthode du dépassement de seuil
- SARS-CoV-2 : nom officiel du nouveau coronavirus identifié le 9 janvier 2020 dans la ville de Wuhan. Dans ce mémoire, Les termes suivant désignent ce virus « Covid-19 », « Virus respiratoire », « Coronavirus ».
- SI : Stringency Index, la métrique « indice de sévérité » ,issue de l'Ox-CGRT, mesurant l'intensité des mesures gouvernementales
- TVE : Théorie des Valeurs Extrêmes
- UE : Union Européenne

1 Introduction

1.1 Contexte

Bien que nous ayons tendance à le renvoyer au second plan depuis quelques mois, le Covid-19 marquera à jamais l'époque que nous traversons. Jamais auparavant, l'économie mondiale n'avait été forcée de s'estomper au profit de la santé des êtres humains. Pourtant, c'est bien ce qu'il s'est passé au travers de plusieurs confinements en 2020 et 2021. En Belgique, nous avons été confrontés à trois confinements spectaculaires afin de diminuer les interactions physiques et sociales responsables de la propagation du virus. Notre quotidien a été complètement chamboulé du jour au lendemain sans que nous ayons pu le prévoir. Ces souvenirs d'une vie totalement différente resteront à jamais gravés dans les mémoires de chacun.

En 2023, il faut bien avouer que l'actualité touchant à ce virus SARS-CoV-2 est plutôt calme. En Belgique, le dernier fait saillant du 20 avril 2023 faisait état d'un passage au niveau 1 dans la gestion du virus¹. Il a été décidé par les experts du Risk Management Group et les autorités politiques de ne plus rendre obligatoire le port du masque buccal dans le secteur des soins de santé. Néanmoins, il a été rappelé que la surveillance et la préoccupation à l'égard du virus restaient de mise. L'année passée, les discussions à son sujet évoquaient plutôt les moyens de s'en protéger tel que la quatrième dose de vaccination et les inévitables craintes de variants du virus originel.

Or, le Covid-19 est toujours actif et circule toujours dans la population². Nous sommes encore loin d'affirmer que la pandémie est un sujet clos tant il convient de rester un minimum vigilant, à la lumière de la gravité des faits passés. De plus, de nombreux changements d'habitudes ont été précipités depuis son apparition : certains gestes barrières se maintiennent, la participation à des rassemblements sociaux et des sorties culturelles a diminué, les livraisons à domicile ont connu une croissance et l'équilibre professionnel souhaité, actuellement, exige une part non-négligeable de télétravail.

1.2 Importance et justification de la thématique

Les approches statistiques basées sur la Théorie des Valeurs Extrêmes (TVE) ont apporté des résultats prometteurs quant à la prévisibilité d'événements inhabituels dans des domaines aussi variés que l'hydrologie [Davison et Smith, 1990 ; Katz, 2002], la finance [Longin, 1998, 2000 ; Embrechts, 1999 ; Gençay et Selçuk, 2004], l'assurance [McNeil *et al.*, 1997 ; Rootzen et Tajvidi, 1997], etc. En extrapolant à partir de données historiques, cela a permis, par exemple, d'anticiper de potentielles inondations extrêmes et des crashes financiers. Cependant, les applications de cette théorie extrême dans le monde de la santé ont tendance à être rares. D'un point de vue de la planification des ressources, cela peut paraître surprenant tant la prévision de la probabilité d'événements exceptionnels dans un futur proche semble cruciale. Le passé nous l'a montré : des phénomènes aussi puissants qu'une épidémie non-anticipée ou des vagues significatives de chaleur engendrent des répercussions immédiates sur les services hospitaliers qui peuvent très vite être soumis à une pression importante, voir être complètement saturés.

De ce fait, il m'a paru intéressant de relier cette théorie TVE statistique avec une thématique qui n'a pas encore fait l'objet d'une étude précise. Il existe certaines publications concernant

¹<https://www.info-coronavirus.be/fr/news/le-niveau-de-gestion-%C3%A9pid%C3%A9miologique-passe-au-niveau-1-des-changements-pour-le-port-du-masque/>

²<https://www.rtbf.be/article/covid-le-nouveau-variant-eris-qui-gagne-du-terrain-en-france-a-t-il-un-impact-en-belgique-11239846>

la santé publique mais après plusieurs recherches, aucune n'aborde concrètement la pandémie mondiale du Covid-19. Cela peut s'expliquer par un intérêt moindre du monde des statistiques extrêmes pour le secteur sanitaire et par le caractère récent de la pandémie du Covid-19. En effet, bien que nous semblions de moins en moins affectés dans notre vie quotidienne par ce virus, celui-ci nous était encore totalement inconnu avant l'année 2020.

1.3 Problématiques centrales et questions de recherche

Dans un contexte où une mauvaise anticipation d'événements sanitaires extrêmes peut être d'une ampleur considérable, à la fois sur le monde hospitalier à cause d'une demande trop nombreuse, de matériel inadapté et un manque de personnel suffisamment formé, sur le plan économique via une baisse de l'activité et un ralentissement du progrès et sur le contingent humain à cause de la surmortalité, il est crucial d'apporter une estimation du risque de recrudescence d'une telle pandémie. A partir de la Théorie des Valeurs Extrêmes et des données d'hospitalisations associées au Coronavirus, nous nous demanderons si un retour pandémique est envisageable à moyen terme, en Belgique et au Portugal. Une dernière variable sera introduite afin d'établir une distinction entre des hospitalisations ayant eu lieu en période de faibles mesures gouvernementales et de fortes mesures. Dès lors, nous pourrons savoir si la réponse gouvernementale en période de Covid-19 est un facteur déterminant dans la lutte contre l'engorgement des hôpitaux et surtout sur les hospitalisations et décès causés par la maladie.

La première partie de l'analyse inclura uniquement la dimension hospitalière. Au cours de cette section, la question de recherche centrale de ce mémoire sera au centre de l'attention : « A l'aide de la Théorie des Valeurs Extrêmes, quel est le risque que la pandémie de Covid-19 resurgisse de manière fulgurante, dans les dix prochaines années, en Belgique et au Portugal ». Des sous questions seront également répondues : « Quelle est la probabilité quotidienne d'être confrontés à un niveau d'augmentation hospitalière similaire au pic maximal enregistré durant les années 2020 et 2021 ? » et « Peut-on affirmer qu'un de nos deux pays d'intérêt est plus susceptible de connaître une crise sanitaire aussi extrême dans le futur ? ».

Dans la deuxième partie de l'analyse, la question de recherche centrale sera discutée mais l'accent sera mis sur la nouvelle dimension étudiée, à savoir le type de mesures gouvernementales en vigueur. Les sous-questions suivantes seront abordées : « En Belgique et au Portugal, quelles sont les probabilités quotidiennes d'être confrontés à un niveau d'augmentation semblable aux pics passés, respectivement en présence de mesures faibles et fortes ? », « Peut-on affirmer que le degré de sévérité de la réponse gouvernementale influence la croissance du nombre d'hospitalisations associées au Covid-19 ? ».

1.4 Synthèse de la méthodologie suivie

La variable d'intérêt retenue pour l'analyse statistique est le nombre d'hospitalisations associées au Covid-19, bien qu'elle soit légèrement décalée dans le temps par rapport aux contaminations. Il a été choisi d'utiliser cette variable au lieu du nombre de décès ou encore du nombre de cas parce qu'il s'agit d'un indicateur dont le comptage fut homogène au cours du temps et entre les différents pays européens. La Belgique avait, par exemple, une façon inclusive et pas très objective de répertorier les décès³. En 2020, toutes les personnes suspectées

³<https://www.rtf.be/article/trop-transparente-la-belgique-va-revoir-sa-facon-de-compter-les-morts-du-covid-10483513>

d'être décédées du coronavirus en maison de repos, ont été comptabilisées et cela sans qu'il n'y ait eu de tests réalisés. Le nombre de cas n'est pas un indicateur fiable étant donné qu'il dépend du nombre total de tests réalisés sur la période considérée et du profil des personnes testées. Pour garantir l'hypothèse de stationnarité inhérente à la TVE, cette variable sera transformée en différence logarithmique. Les interprétations des résultats seront, par conséquent, exprimés en termes d'augmentation et de diminution du nombre d'hospitalisations. Pour élargir la perspective, notre pays, la Belgique, sera comparé avec le Portugal dans chacune des situations. Le Portugal était retenu parce qu'il offrait également des données hospitalières quotidiennes, à la différence de nos pays frontaliers. Ce pays ibérique s'approche également de la Belgique en nombre d'habitants : à la mi-2021, 10,3 millions pour la première nation citée contre 11,5 millions pour la seconde⁴. Cette comparaison pourra mettre en évidence des tendances ou des différences significatives. Le type de mesures gouvernementales, faible ou fort, ajouté dans la deuxième partie de l'analyse basée sur la TVE, a été traduit de l'indice de sévérité des mesures SI (Stringency Index). Des jours dont l'intensité de sévérité de la réponse est inférieure à la moyenne du SI sur la période étudiée, seront associées à des faibles mesures (vice-versa pour les fortes mesures).

Concernant l'application de la TVE à nos données, il a fallu faire un choix entre les approches de modélisation : celle des maxima de blocs ou celle des dépassements de seuil. Nous nous tournerons vers la méthode du dépassement de seuil parce qu'elle utilise plus d'information en considération de nos données quotidiennes. C'est pourquoi nous ajusterons une distribution GPD dite « Pareto Généralisée » à nos données extrêmes. Tout d'abord, nous choisirons un seuil adéquat, ayant pour objectif de garder uniquement un échantillon de valeurs extrêmes, supérieur au seuil. Ensuite, nous estimerons les paramètres d'échelle et de forme de la distribution GPD ajustée. Ce sont ces paramètres qui nous permettront d'extrapoler nos données historiques et de réaliser des prévisions grâce aux probabilités extrêmes et aux niveaux de retour.

1.5 Plan du mémoire

La structure suivie au cours de ce mémoire et l'articulation des différentes grandes parties peuvent être résumées comme suit. Dans un premier temps, une analyse descriptive des données hospitalières ainsi que de l'indice de sévérité des mesures gouvernementales sera présentée via les paramètres centraux et de dispersion. Les bases de l'exploration statistique seront ainsi déployées. Ensuite, une introduction aux fondements de la Théorie des Valeurs Extrêmes sera offerte, afin d'instaurer une base conceptuelle solide pour les analyses ultérieures. Par la suite, la littérature existante relative à notre sujet, sera parcourue. Cette revue bibliographique permettra d'appréhender l'état actuel des connaissances et les éventuelles lacunes dans la recherche existante. En outre, la TVE sera mise en œuvre, dans le but de faire part des résultats de l'analyse des données extrêmes. Ils seront accompagnés d'une explication méthodologique, à savoir la phase préparatoire des données, le choix du seuil de dépassement et l'estimation des paramètres du modèle. Enfin, la conclusion finale du mémoire constituera une synthèse des éléments clés. Elle récapitulera brièvement les points les plus importants des différentes étapes exposées et, surtout, les réponses à la question et aux sous-questions de recherche de cette introduction.

⁴ <https://fr.statista.com/statistiques/564181/population-totale-des-membres-de-l-union-europeenne-en-et-2050/>

2 Analyse descriptive des données

2.1 Données épidémiologiques : nombre d'hospitalisations liées au Covid-19

2.1.1 Source de la base de données

Concernant la source des données relative aux hospitalisations Covid-19, nous avons choisi de nous référer au Centre Européen de Prévention et de Contrôle des Maladies ECDC (European Center for Disease Prevention and Control). Cette source est fiable par le biais de son statut officiel de coordinateur de la surveillance des maladies infectieuses de l'Union Européenne et de L'Espace Européen Economique (EEE). Les Etats membres sont tenus de communiquer leurs données nationales conformément aux calendriers communs, aux définitions de cas et aux procédures convenues. La directive adoptée par l'UE portant sur l'open data exige que les entités publiques mettent à disposition ces informations en vue de leur réutilisation par des tierces personnes. Après les étapes de collecte, l'uniformisation et la validation, les données peuvent être consultées et téléchargées par voie électronique. Ce procédé de surveillance épidémiologique systématique, partagé de manière transparente et universelle, remplit différents objectifs. Il sert, entre autres, à permettre l'analyse de données par des particuliers afin d'améliorer et d'adapter la prévention et le contrôle des maladies. La recherche en matière de santé publique est ainsi facilitée. L'accès libre réduit la charge de travail de l'ECDC qui ne doit dès lors plus autoriser le partage des data à chaque utilisateur externe.

Les bases de données disponibles sur le site de l'ECDC⁵ sont continuellement mises à jour. Bien que le virus respiratoire soit beaucoup moins contaminant à l'heure actuelle, la collecte de données a toujours lieu. Les premières données enregistrées remontent à la fin du mois de février 2020 pour l'Italie et au cours des mois de mars et avril 2020 pour les autres pays européens. La fréquence des données fournies à l'ECDC varie d'un pays à l'autre. Celle-ci est tantôt quotidienne, tantôt hebdomadaire. Par exemple, la Belgique, le Portugal et le Danemark envoient des données épidémiologiques quotidiennes alors que la France, l'Allemagne et les Pays-Bas renseignent des moyennes hebdomadaires. Cela conduit à des tailles différentes d'échantillon par nation et l'agrégation des données offrent moins de variation dans le second ensemble de pays.

2.1.2 Variable étudiée

De la base de données rendue disponible par l'ECDC, nous avons choisi de nous concentrer sur les admissions hospitalières quotidiennes, une des classes présente sous l'en-tête « Indicateur ». La colonne « Valeur », variable quantitative discrète, relève le nombre d'individus comptés pour un indicateur, un pays et une date précise. L'intervalle de temps retenu commençait en mars 2020 lors de l'apparition du virus en Europe occidentale, et se terminait en octobre 2021. Nous estimions qu'au-delà de cette date⁶, la campagne de vaccination influençait grandement le niveau d'hospitalisations et qu'il était difficile de discerner réellement l'effet des mesures non-pharmaceutiques.

2.1.3 Description des paramètres centraux et de dispersion

Dans cette section, des statistiques descriptives seront discutées et comparées entre la Belgique et le Portugal, nos deux pays d'intérêt tout au long de ce mémoire. Ces statistiques permettent

⁵ <https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-covid-19>

⁶ <https://www.lalibre.be/belgique/societe/2021/11/05/coronavirus-les-personnes-non-vaccinees-sont-surrepresentees-aux-soins-intensifs-NO4SZN4TFNGY5GSEONHZLSBHCA/>

d'avoir une idée de la distribution des données et de l'étendue des valeurs observées. Une attention particulière sera consacrée à des résultats remarquables.

Paramètres	Pays	Belgique	Portugal
Taille de l'échantillon (n)		595	605
Moyenne		1776,6	1232
Médiane		1325	603
Ecart-type		1682,3	1464,2
Etendue		7331	6856
Ecart interquartile (IQR)		2055,5	789
Coefficient de variation		94,7	118,9

Tableau 1 : Principaux paramètres centraux et de dispersion des hospitalisations quotidiennes en Belgique et au Portugal

Tout d'abord, le niveau d'hospitalisation en raison d'une infection au SARS-CoV-2, peut être résumé globalement en examinant les paramètres centraux tels que la moyenne et la médiane ([Tableau 1](#)). En Belgique, la moyenne des hospitalisations s'élève à environ 1777 hospitalisations quotidiennes, tandis que la médiane est d'environ 1325. Cette nette prédominance de la moyenne sur la médiane indique la probable présence de valeurs extrêmement grandes, étirant la moyenne vers le haut. Ces valeurs aberrantes ont tendance à influencer de manière considérable la moyenne arithmétique étant donné qu'elle est moins robuste que la médiane. D'autre part, au Portugal, la moyenne des hospitalisations est d'environ 1232 hospitalisations quotidiennes, tandis que la médiane est plus basse, à environ 603. En comparant les deux pays, on observe que la Belgique obtient des scores plus élevés que le Portugal pour les paramètres centraux. Cela peut indiquer une charge hospitalière globalement plus élevée en Belgique. Cependant, d'autres facteurs et variables doivent être pris en compte pour une évaluation complète de la situation épidémiologique dans chaque pays.

Les paramètres de dispersion, tels que l'écart-type, le coefficient de variation, l'étendue et l'écart interquartile permettent de comparer la variabilité des données en Belgique et au Portugal ([Tableau 1](#)). En Belgique, l'écart-type des hospitalisations est de 1682 hospitalisations, signifiant qu'en moyenne, les écarts par rapport à ce paramètre central sont de 1682. Cela indique une dispersion relativement importante des données autour de la moyenne. Au Portugal, l'écart-type des hospitalisations est de 1464 admissions en centre hospitalier. Le coefficient de variation est de 118,9%, ce qui indique une variabilité relative plus élevée par rapport à la moyenne que celle observée en Belgique (94,7%).

Les valeurs prises par le paramètre d'écart interquartile dans ces deux pays sont nettement différentes. Ce paramètre représente l'étendue entre le premier quartile (Q1) et le troisième quartile (Q3) d'une distribution. L'IQR est particulièrement utile pour identifier la dispersion au sein de la moitié centrale d'une distribution étant donné qu'il ne tient pas compte des valeurs extrêmes, par définition. En Belgique, l'intervalle entre le Q1 et le Q3 est de 2055 hospitalisations. Au Portugal, l'IQR est de 789 hospitalisations. Ce résultat intéressant suggère que la distribution des hospitalisations au Portugal est potentiellement moins étendue et plus concentrée autour de la médiane. Par conséquent et en mettant en relation l'IQR avec l'étendue, nous pouvons nous attendre à un quatrième quartile de plus grande taille au Portugal. Cela peut expliquer la variabilité relative plus importante dans ce pays. Ces résultats sont visibles dans la boîte à moustache propre à chaque pays ([Annexe Figures A.1 et A.2](#)).

2.1.4 Séries temporelles

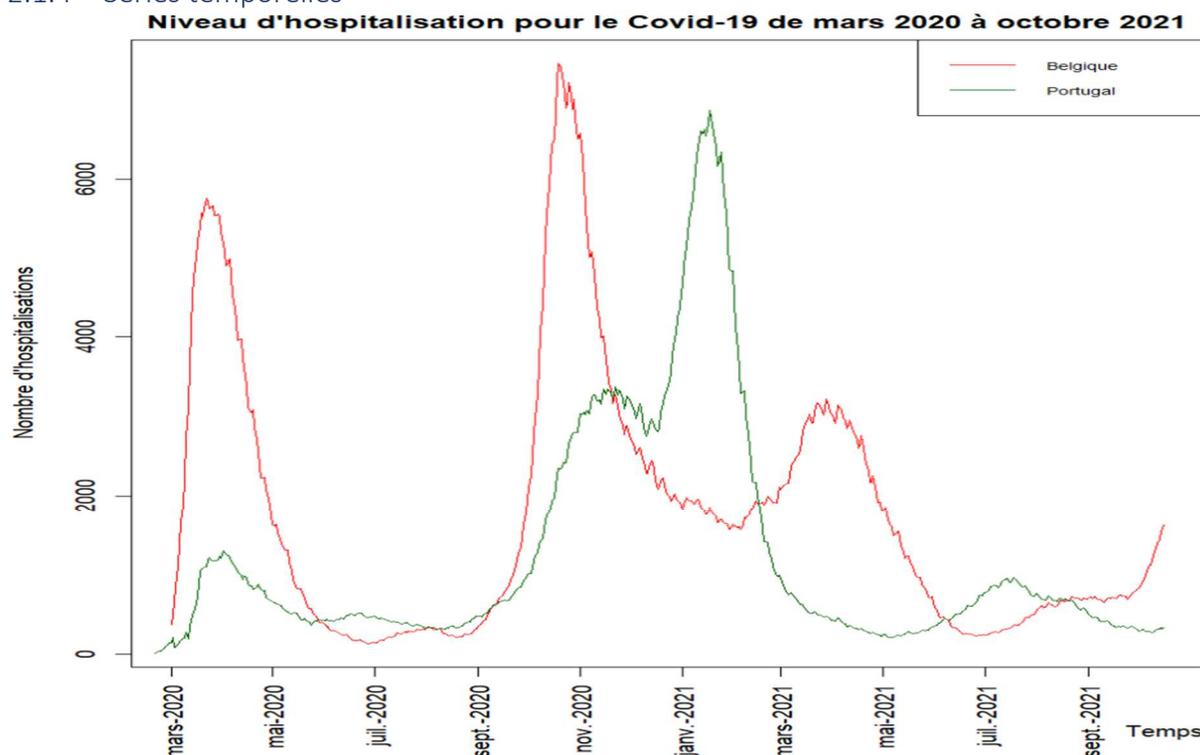


Figure 1 : Séries temporelles des hospitalisations quotidiennes en Belgique et au Portugal

La **Figure 1** nous fait part de la relation entre le niveau d'hospitalisation et le temps. Cette série temporelle pour la Belgique (rouge) met en évidence trois vagues de contaminations significatives : la première en mars et avril 2020, la seconde en octobre et novembre 2020 et la troisième en mars et avril 2021. Le pic le plus important d'hospitalisations liées au Covid-19 a été atteint au cours de la deuxième vague.

Au Portugal (vert), la tendance des trois vagues de contamination est moins marquée. La première vague a moins mis à mal le système hospitalier par rapport à la Belgique. Le pic d'hospitalisations a été enregistré un peu plus tard, en février 2021, au cours de la troisième vague. Nous pouvons également remarquer que le Portugal a dû faire face à un niveau relativement haut d'hospitalisations Covid-19 entre novembre 2020 et mars 2021. Pendant ces quatre mois, le niveau n'est jamais descendu en-dessous de 3000 hospitalisations.

2.1.5 Distributions des données

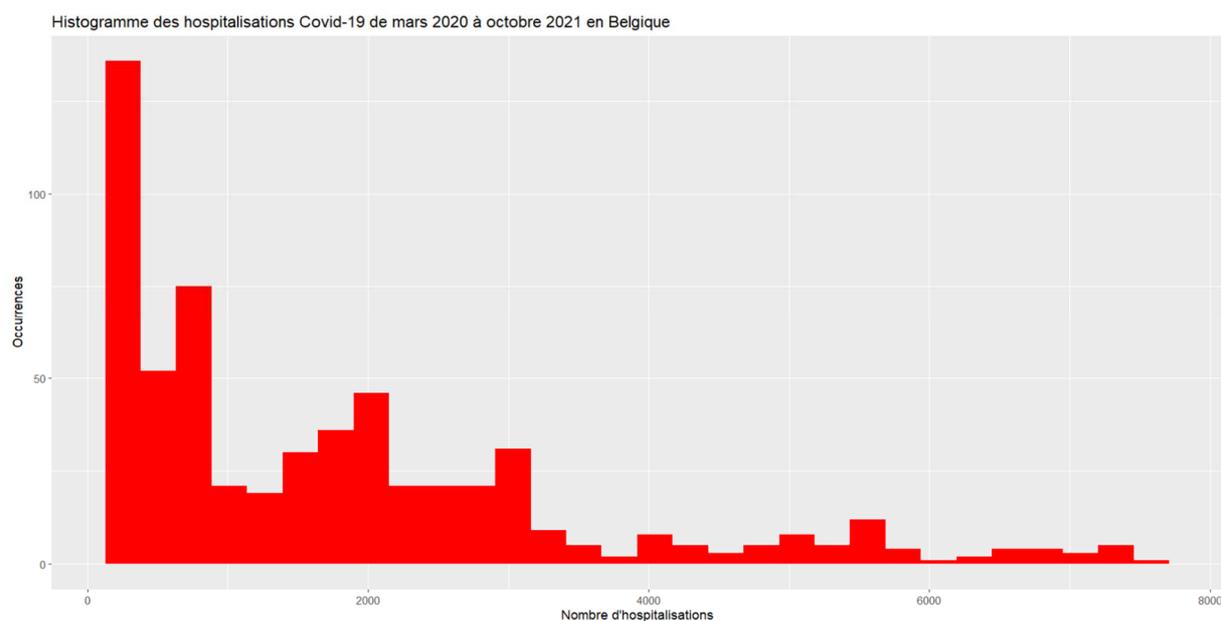


Figure 2 : Histogramme du nombre d'hospitalisations belges

L'histogramme permet de visualiser la répartition des valeurs d'une variable numérique en intervalles réguliers appelés « classes », en l'occurrence le nombre d'hospitalisations associées au virus respiratoire. Le but principal d'un histogramme est de fournir une représentation visuelle de la distribution des données. Les tendances et les caractéristiques importantes de la variable étudiée sont semblables quel que soit le pays étudié.

Dans les deux cas (**Figure 2** et **Annexe Figure A.3**), nous pouvons voir très nettement que la forme de la distribution est asymétrique à droite (skewness positif). En effet, la majorité des données se retrouvent à gauche du graphe et une minorité conséquente d'observations apparaît dans la queue de distribution de droite. Les valeurs très élevées sont donc plus fréquentes qu'une distribution normale qui ne peut être appliquée à cette distribution. Ces valeurs étirant la distribution à droite s'expliquent par le niveau relativement haut d'hospitalisations lors des pics de contamination.

De manière plus locale, la Belgique (**Figure 2**) semble compter trois modes à 200, 2000 et 3000 hospitalisations alors que le Portugal (**Annexe Figure A.3**) en recense plutôt deux, à 400 et 3400 hospitalisations. Les données tendent à être plus rassemblées autour de la médiane pour le Portugal comme l'indiquait l'écart interquartile. C'est moins le cas pour la Belgique, où l'étalement des données est plus visible. Dans ce pays, il faut aller au-delà de 3000 hospitalisations pour rencontrer des classes comptant moins de 40 hospitalisations.

En conclusion, l'analyse descriptive met en évidence la présence de valeurs extrêmes et d'une queue lourde de distribution à droite dans les hospitalisations liées au Covid-19. La présence de valeurs extrêmes souligne l'intérêt d'approfondir l'analyse en utilisant des méthodes de la théorie des valeurs extrêmes. Cette théorie se concentre spécifiquement sur l'étude des événements rares et extrêmes, permettant ainsi de mieux comprendre les caractéristiques et les impacts de ces valeurs extrêmes.

2.1.6 Cas des différences logarithmiques d'hospitalisations quotidiennes

Dans la section 5 du mémoire, nous n'analyserons pas les hospitalisations quotidiennes extrêmes telles quelles mais plutôt des transformations logarithmiques de cette variable. La TVE requiert une hypothèse de stationnarité des données et lorsque nous nous penchons sur la **Figure 1**, nous pouvons très vite conclure que les données fluctuent considérablement afin de garantir cette hypothèse. Dès lors, il est également intéressant d'avoir un aperçu de notre variable d'intérêt transformée : les différences logarithmiques népériennes du nombre d'hospitalisations liées au Covid-19. Il n'est pas nécessaire de proposer des valeurs de paramètres centraux et de dispersion étant donné que l'analyse basée sur la TVE, se focalise uniquement sur le quantile extrême.

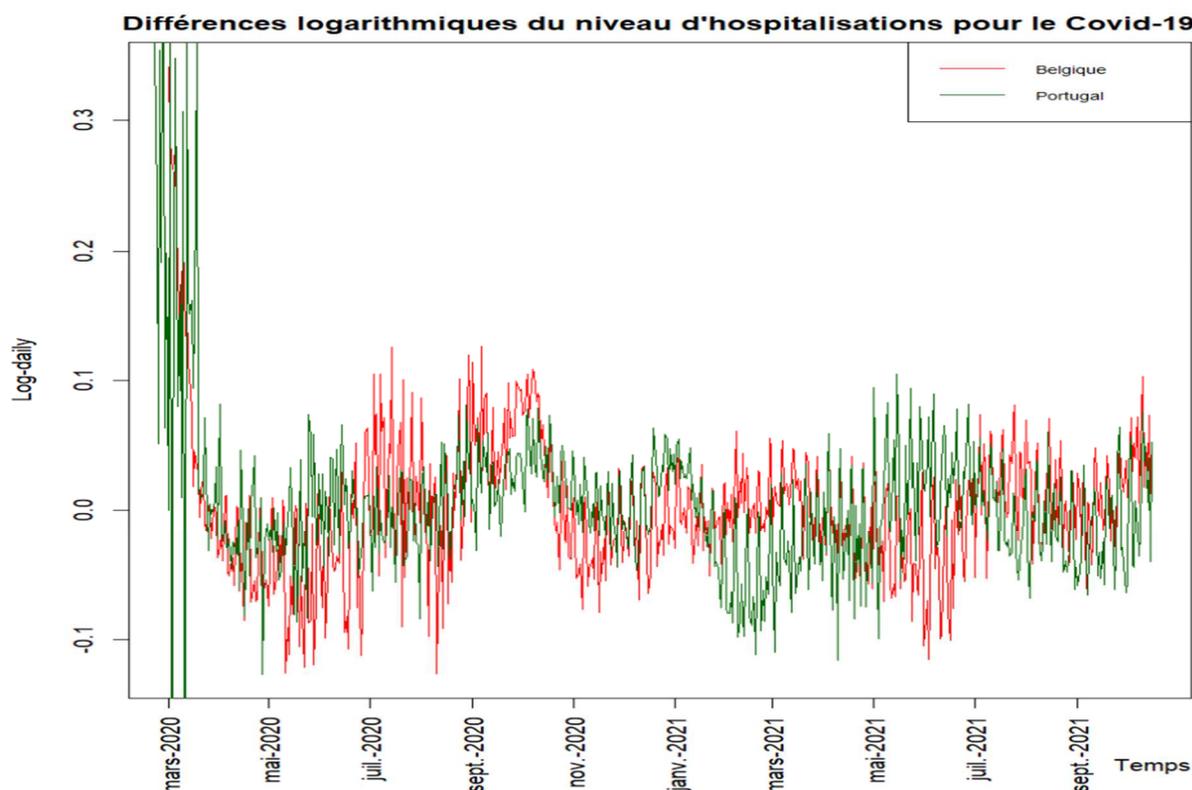


Figure 3 : Séries temporelles montrant les différences logarithmiques du nombre d'hospitalisations Covid-19, de mars 2020 à octobre 2021.

La **Figure 3** semble mieux respecter l'hypothèse de stationnarité et une distribution moins dépendante du temps que la Figure 1. Cependant, nous veillerons à exclure les observations de mars 2020, suite au comportement non-stationnaire constaté lors de ce premier mois. Après cette modification, la population finale conservée pour l'analyse extrême, oscillera la grande majorité du temps entre des valeurs de $-0,1$ à $0,1$. L'absence de pics supérieurs à $0,12$ dès le mois d'avril 2020, s'explique par le fait que le nombre d'hospitalisations n'est jamais retombé à 0 après l'arrivée du virus.

Les **Annexes Figures A.4 et A.5** montrent que les histogrammes des log-daily présentent approximativement des formes de cloche, contrairement à celles des **Figure 2 et Annexe Figure A.3**. Les distributions semblent centrées autour d'un logarithme journalier égal à 0. Cela démontre que les données ont bien été stabilisées par les logarithmes. Cependant, nous pouvons toujours apercevoir qu'il existe une queue de distribution à droite, visiblement plus marquée pour la Belgique (**Annexe Figure A.4**).

2.2 Données gouvernementales : indice de sévérité des mesures non-pharmaceutiques

2.2.1 Source de la base de données

La seconde base de données utilisée dans ce mémoire provient du tracker Ox-CGRT (Oxford Coronavirus Government Response Tracker) développé par l'équipe de recherche de l'Université d'Oxford dirigé par le Professeur Thomas Hale. Il calcule un indice de sévérité SI, une mesure composite de neuf indicateurs de réponse. Les neuf paramètres utilisés pour calculer l'indice de rigueur sont les suivants : fermetures d'écoles, fermetures de lieux de travail, annulations d'événements publics, restrictions sur les rassemblements publics, fermetures des transports publics, obligations de rester chez soi, campagnes d'information du public, restrictions sur les mouvements internes et contrôles des voyages internationaux. Ces mesures sont dites « non-pharmaceutiques » parce qu'elles ne requièrent pas d'interventions médicales. Les données sont collectées à partir de diverses sources publiques, notamment les annonces officielles des gouvernements, les sites web des ministères de la santé, les médias et d'autres sources fiables. L'équipe de recherche d'Oxford vérifie et compile ensuite ces informations pour créer une base de données complète et actualisée des réponses gouvernementales. Cet ensemble de données est disponible en libre accès et les données y étaient actualisées jusqu'en décembre 2022.

2.2.2 Variable étudiée

Au moment du téléchargement de cette base de données⁷ en décembre 2022, la base de données enregistrait 242 207 observations distinctes, réparties entre 248 pays et régions du monde. A elle seule, la Belgique avait fourni 1043 entrées différentes. Il s'agissait de 1017 observations pour le Portugal, les données étant compilées quotidiennement de mars 2020 à décembre 2022. Au total, cette base de données extrêmement complète établissait des données pour 67 variables. La variable la plus importante nous concernant portait le nom « Stringency Index », c'est-à-dire l'indice de sévérité des mesures non-pharmaceutiques. Cette variable quantitative continue est calculée comme la moyenne de neuf mesures précises, chacune prenant une valeur comprise entre 0 et 100. Un score élevé indique une réponse plus stricte (une valeur de 100 étant la réponse la plus stricte).

2.2.3 Description des paramètres centraux et de dispersion

A titre descriptif, la moyenne et la médiane de l'indice de sévérité des mesures pour la Belgique valaient respectivement 55,81 et 54,63. Les degrés de sévérité mesurés comme les plus extrêmes, pour la période avant l'effet de la vaccination, étaient de 81,48 pour le maximum et 11,11 pour le minimum. Au Portugal, les paramètres centraux (moyenne et médiane) étaient de 65,29 et 63,89. Nous pouvons affirmer que les mesures prises par le gouvernement portugais étaient en moyenne 10,5 points plus sévères que celles prises par son homologue belge. Les valeurs prises par l'indice portugais oscillaient entre 87,96 (valeur maximale) et 5,56 (valeur minimale). Le paramètre de dispersion écart-type était un peu plus élevé pour le pays ibérique : 15,54 pour la Belgique et 12,35 pour le Portugal.

Plus tard, lorsque nous devrons scinder les hospitalisations selon le type de mesures en vigueur, cela sera réalisé en fonction des moyennes de l'indice de sévérité. A titre d'illustration, les observations belges correspondant à des mesures fortes seront celles dont le SI était supérieur à la moyenne de 55,81, au moment de la contamination.

⁷ <https://ourworldindata.org/covid-stringency-index>

2.2.4 Séries temporelles

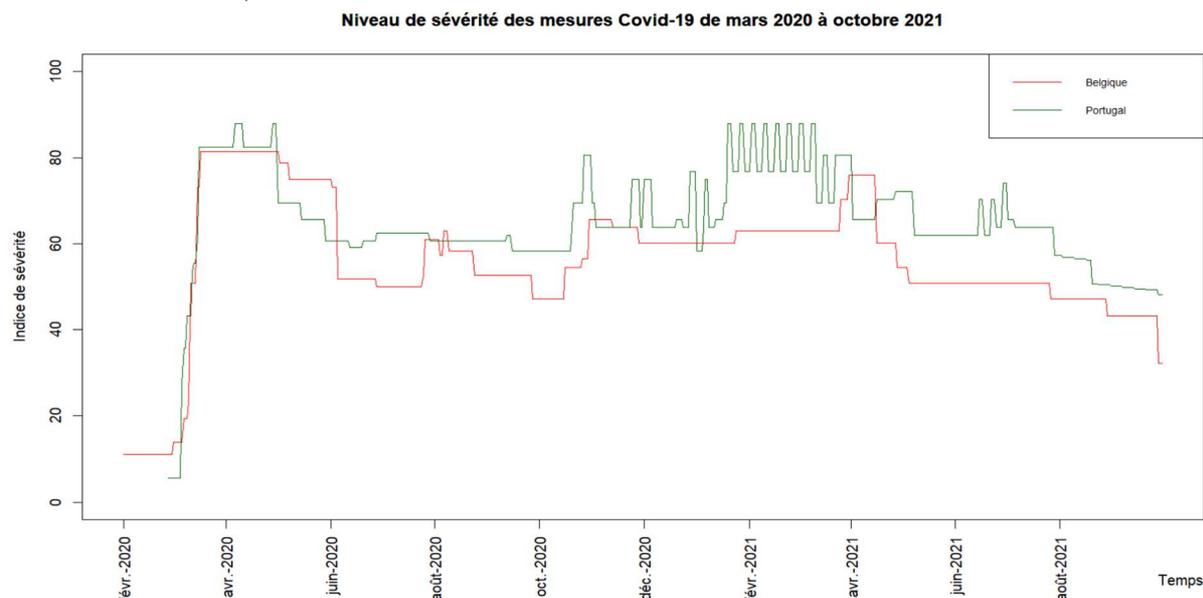


Figure 4 : Indice de sévérité des mesures Covid-19 de mars 2020 à octobre 2021

Le graphique (**Figure 4**) met en lumière deux séries temporelles. Il montre l'évolution de la rigueur des mesures non-pharmaceutiques au cours du temps en Belgique et au Portugal. A l'image des trois vagues de contaminations déterminées à partir des hospitalisations quotidiennes, en Belgique, le graphique renseigne également la présence de 3 sommets locaux. En fait, cette intensification des mesures s'est déclenchée en réaction à la hausse des hospitalisations. Par exemple, la première vague d'hospitalisation a eu lieu en mars et en avril 2020 mais le pic atteint par les mesures est visible en avril et mai 2020. La réponse gouvernementale était, par conséquent, directement dictée par la hausse inquiétante de la métrique « nombre d'hospitalisations liées au Covid-19 ». Lors des périodes de creux, nous pouvons voir que le SI s'est stabilisé à un niveau proche de 50 en Belgique.

La tendance prise par la courbe portugaise se rapproche à celle de la Belgique, à l'exception qu'elle prend des valeurs plus élevées. A plusieurs moments dans le temps, le gouvernement portugais a choisi de faire respecter des mesures ayant un degré supérieur à 80. Hormis en mai et juin 2020, la réponse gouvernementale était plus importante dans ce pays ibérique. Pendant un an et demi, l'indice n'est jamais descendu sous la barre de 60.

3 Introduction à la Théorie des Valeurs Extrêmes

3.1 Objectifs et spécificités de la TVE

La théorie des valeurs extrêmes a pour but d'évaluer les événements rares ayant une magnitude plus importante que d'habitude afin d'en estimer les répercussions et les pertes ou gains qui en découlent. Cette théorie est unique en son genre du fait qu'elle se concentre directement sur la queue de la distribution. En réalité, seules les données extrêmes sont retenues pour approximer les paramètres des modèles de la TVE. Un cadre mathématique précis est proposé [Coles, 2001] visant à étudier le comportement stochastique des extrêmes d'une distribution.

L'un des objectifs majeurs pour les personnes recourant à la TVE, est d'estimer la probabilité de certains événements rares, à partir d'une étude des valeurs extrêmes d'un échantillon. Des extrapolations au-delà des données historiques observées d'une distribution extrême sont rendues possibles, via le calcul d'une probabilité et le niveau de retour, jouant le rôle de quantile, qui lui est associé. Après avoir isolé les occurrences extrêmes, les paramètres de la fonction de distribution sont estimés. Ensuite, nous pouvons évaluer la probabilité d'événements aussi extrêmes, voir encore plus sévères que ceux précédemment enregistrés. La seconde métrique vise à estimer des quantiles extrêmes, appelés niveaux de retour. Cela fait référence au calcul de la limite d'une zone critique.

Le grand intérêt de la TVE réside au niveau de l'estimation des paramètres des modèles de la théorie des valeurs extrêmes : ceux-ci sont déterminés uniquement à partir des observations extrêmes faisant partie des queues de distribution. Cela contraste avec l'univers Gaussien avec lequel nous sommes familiers dont les paramètres d'espérance et d'écart-type sont estimés selon la totalité des observations et dont le Théorème Centrale Limite se base purement sur la moyenne de l'échantillon [Raggad, 2009 ; Charras-Garrido et Lezard, 2013]. Une approche gaussienne aurait tendance également à sous-estimer un futur événement extrême lorsque le passé récent ne contient pas de données particulièrement alarmantes. Or, il est crucial d'apporter une évaluation précise quant à l'impact d'un risque extrême tant ce sont ces aléas qui sont les plus impactants sur les plans humains, financiers et environnementaux. Par conséquent, mesurer un risque extrême selon la loi Normale n'est généralement pas l'idéal. En se limitant aux occurrences les plus exceptionnels, la TVE garantit un meilleur ajustement à la queue de la loi [Chiu et al., 2016].

Cette théorie a vu le jour entre 1920 et 1940 grâce aux travaux de Fréchet, Fisher, Tippet, Gumbel et Gnedenko. D'un point de vue théorique, il a été rapidement défini que la distribution des valeurs extrêmes ne peut appartenir qu'à trois classes distinctes lorsque l'on modélise le maximum d'un ensemble de variables aléatoires. Il s'agit des familles suivantes : Weibull, Gumbel et Fréchet. Ces familles se distinguent par différentes formes de comportement de la queue de la fonction de distribution.

3.2 Types d'approche

La modélisation des événements extrêmes selon la TVE se présente sous la forme de deux approches asymptotiques : l'approche classique des maxima de blocs BM (Block Maximas)

et l'approche du dépassement de seuil POT (Peaks Over Threshold). Dans le premier cas, on parle d'une distribution généralisée des valeurs extrêmes (GEV) qui est ajustée aux données et dans le second cas, il s'agit d'une distribution généralisée de Pareto (GPD). L'approche POT est généralement préférée car elle exploite plus d'informations sans avoir la contrainte de devoir relever les valeurs expérimentales à intervalles réguliers. En effet, il est fort probable que des événements extrêmes proches dans le temps soient exclus de l'analyse TVE par maximas de blocs parce qu'ils se trouvaient dans le même bloc, c'est-à-dire le même intervalle de temps. Si une série temporelle complète d'observations journalières est disponible, alors une meilleure utilisation de l'information est rendue possible par la méthode du dépassement de seuil et plus d'observations historiques sont mobilisés. Au cours de cette revue théorique, l'accent sera mis sur cette deuxième méthode étant donné que nous l'appliquerons pour nos hospitalisations Covid-19 journalières.

3.3 Méthode du dépassement de seuil

L'approche appelée « dépassement de seuil » est une approche de la TVE qui est utilisée pour modéliser les valeurs extrêmes d'une variable donnée. Cette approche suppose que les valeurs extrêmes de cette variable se produisent lorsque la variable dépasse un certain seuil. Cela consiste à diviser l'ensemble des données en deux parties : les valeurs inférieures au seuil et les valeurs supérieures au seuil [Coles, 2001 ; Davison et Smith, 1990].

3.3.1 Modélisation de la fonction de distribution des valeurs extrêmes

Si la distribution de la population F était connue, la distribution des dépassements de seuil serait également connue. Étant donné que, dans les applications pratiques, ce n'est pas le cas, on recherche des approximations largement applicables pour des valeurs élevées du seuil. Par conséquent, cette distribution inconnue des extrêmes est approximée par la distribution généralisée de Pareto (GPD). Le modèle GPD est un modèle paramétrique qui peut être ajusté aux données extrêmes pour estimer les paramètres de la distribution de Pareto [Coles, 2001].

Théorème 3.4.1 : Soit X_1, X_2, \dots une séquence de variables aléatoires indépendantes dont la fonction de distribution commune est F . Définissons M_n (la valeur maximale d'un processus au cours de n unités de temps d'observation) comme :

$$M_n = \max\{X_1, \dots, X_n\}$$

Dénotons un terme arbitraire de la séquence X_i par X , et supposons que F satisfasse le théorème 3.4.2 (**Annexe Théorème A.6**), c'est-à-dire que la fonction de distribution non-dégénérée G est membre de la famille GEV dans le cadre des maximas de blocs, de sorte que pour n suffisamment grand,

$$Pr\{M_n \leq z\} \approx G(z)$$

Où,

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}$$

Pour des estimateurs $\mu \in \mathbb{R}$ et $\sigma > 0$ et $\xi \neq 0$. Alors, pour une valeur suffisamment grande de u , la fonction de distribution de $(X - u)$, conditionnellement à $X > u$, est approximativement :

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} \quad (3.1)$$

Défini sur $\{y : y > 0 \text{ et } (1 + \xi/\tilde{\sigma}) > 0\}$, où $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

□

La famille de distributions H définie par l'équation (3.1) est appelée famille de Pareto généralisée. Les paramètres de la GPD sont déterminés de manière unique par ceux de la distribution GEV issue de la méthode des maxima de blocs. En particulier, le paramètre ξ dans (3.1) est égal à celui de la distribution GEV correspondante.

3.3.2 Choix du seuil

Le seuil est la valeur de la variable au-delà de laquelle les observations sont considérées comme des valeurs extrême. Cependant, la question du choix du seuil est analogue au choix de la taille des blocs dans l'approche des maxima par bloc, ce qui implique un compromis entre le biais et la variance. Le choix arbitraire du seuil par le statisticien, influence considérablement les résultats de l'analyse, parce qu'il détermine les observations qui seront incluses dans l'analyse des valeurs extrêmes. Un seuil trop bas est susceptible de violer la base asymptotique du modèle, entraînant un biais. A l'inverse, un seuil trop élevé génère peu d'excès avec lesquels le modèle peut être estimé, ce qui conduit à une variance élevée [Davison et Smith, 1990].

Le théorème 3.4.1 propose le cadre suivant pour la modélisation des valeurs extrêmes. Les données brutes consistent en une séquence de mesures indépendantes et identiquement distribuées x_1, \dots, x_n . Les événements extrêmes sont identifiés en définissant un seuil élevé u , pour lequel les dépassements sont $\{x_i : x_i > u\}$. Marquons ces dépassements par $x_{(1)}, \dots, x_{(k)}$, et définissons ces valeurs excédentaires par $y_j = x_{(j)} - u$, pour $j = 1, \dots, k$. D'après le théorème 3.4.1, les y_j peuvent être considérés comme des réalisations indépendantes d'une variable aléatoire dont la distribution peut être approximée par un membre de la famille de Pareto généralisée. L'inférence consiste à ajuster la famille de Pareto généralisée aux dépassements de seuils observés, puis à vérifier le modèle et à l'extrapoler.

Le chercheur peut s'aider de deux techniques graphiques pour approximer et choisir le seuil critique. Soit via le « mean residual life plot », devenant plus ou moins linéaire autour de u , soit via la technique visant à ajuster une GPD en fonction d'une gamme de seuils u . Pour cette dernière, il faut également sélectionner la valeur du seuil où une certaine stabilité apparaît sur le graphe.

3.3.3 Estimation des paramètres

La distribution paramétrique généralisée de Pareto (GPD) comprend deux paramètres estimés par la méthode de maximum de vraisemblance : le paramètre d'échelle σ et de forme ξ . Ils sont estimés une fois que le seuil est choisi.

Supposons que les valeurs y_1, \dots, y_k sont les k dépassements du seuil u . Pour $\xi \neq 0$, la log-vraisemblance est dérivé de (3.1) de la manière suivante :

$$\ell(\sigma, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^k \log(1 + \xi y_i / \sigma) \quad (3.2)$$

Dans le cas où $\xi = 0$, la log-vraisemblance est obtenue sous la forme suivante :

$$\ell(\sigma) = -k \log \sigma - \sigma^{-1} \sum_{i=1}^k y_i \quad (3.3)$$

La maximisation analytique de la log-vraisemblance n'est pas possible. Des techniques numériques sont nécessaires, en prenant soin d'éviter les instabilités numériques lorsque ξ s'approche de 0 dans (3.2).

Le paramètre d'échelle σ décrit la dispersion des valeurs extrêmes par rapport au seuil choisi. Plus précisément, il correspond à l'inverse de la pente de la distribution des valeurs extrêmes au-delà du seuil. Une interprétation courante du paramètre d'échelle est qu'il mesure l'importance relative des extrêmes dans la distribution. En effet, si le paramètre d'échelle est élevé, cela indique que les valeurs extrêmes sont relativement fréquentes et ont donc un impact plus important sur la distribution globale de la variable étudiée. Le paramètre de forme ξ décrit la queue de la distribution des valeurs extrêmes.

Si $\xi < 0$ (Weibull), la distribution des observations dépassant le seuil u , a une limite supérieure et finie égale à $(u - \sigma) / \xi$. Si $\xi > 0$ (Fréchet), la distribution n'a pas de limite supérieure. Si $\xi = 0$, la distribution est également non-bornée et cela correspond à une distribution exponentielle de paramètre $1/\sigma$. Cela revient à dire que la densité de H décroît de manière polynomiale pour la famille de Fréchet et de manière exponentielle pour la distribution de Gumbel. En quelque sorte, les taux de décroissance de la distribution d'origine F varient selon le type de famille, soulignant l'importance prise par les données extrêmes ou non [Coles, 2001].

3.3.4 Estimation des niveaux de retour

Il est généralement plus commode d'interpréter les modèles de valeurs extrêmes en termes de quantiles ou de niveaux de retour, plutôt qu'en termes de valeurs brutes prises par les paramètres estimés. Supposons donc qu'une distribution de Pareto généralisée avec les paramètres σ et ξ est un modèle approprié pour les observations dépassant le seuil u pour une variable X . C'est-à-dire, étant donné que $x > u$, la probabilité suivante présente le risque que la valeur extrême (x) soit au moins atteinte [Coles ; 2001] :

$$\Pr\{X > x \mid X > u\} = \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi}$$

Par conséquent, le niveau x_m qui est dépassé en moyenne une fois toutes les m observations, est la solution de :

$$\Pr\{X > x\} = \zeta_u \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi} \quad (3.4)$$

Où le terme $\zeta_u = \Pr\{X > u\}$, désigne la proportion de l'échantillon qui est supérieur au seuil u . En réarrangeant, nous obtenons le niveau de retour x_m :

$$x_m = u + \frac{\sigma}{\xi} [(m\zeta_u)^\xi - 1] \quad (3.5)$$

Dans l'équation (3.5), les paramètres σ et ξ peuvent être facilement substitués par les estimateurs du maximum de vraisemblance correspondants. Une estimation de ζ_u , est requise. Pour ce faire, nous pouvons compter sur l'estimateur naturelle ;

$$\hat{\zeta}_u = \frac{k}{n}. \quad (3.6)$$

Une fois l'équation complétée (3.5) par l'estimation de ζ_u (3.6), il est possible d'estimer le $m^{\text{ème}}$ niveau de retour, à condition que m soit suffisamment grand pour assurer que $x_m > u$ et à condition que ξ soit différent de zéro. Par construction, nous pouvons dire que x_m correspond au niveau de retour pour la $m^{\text{ème}}$ observation.

De manière à déterminer des niveaux de retour annuels z_N (3.7), il est souvent plus cohérent de remplacer la variable m par $N \times n_y$. Si le niveau de retour de la $N^{\text{ème}}$ année est estimé, alors N sera multiplié par n_y , à savoir le nombre d'observations quotidiennes en un an (365).

$$z_N = u + \frac{\sigma}{\xi} [(Nn_y\zeta_u)^\xi - 1] \quad (3.7)$$

3.3.5 Vérification de la validité du modèle GPD

Des graphiques dits de « diagnostic » sont pertinents de manière à évaluer la **qualité du modèle** GPD ajusté à l'échantillon extrême de données. On retrouve des plots utilisés couramment dans le monde statistique tels que le probability plot, le quantile plot et le density plot. Pour que le modèle de Pareto généralisée soit acceptable, le probability plot et le Q-Q plot doivent être constitués de points qui forment approximativement une ligne droite. Quant à la fonction de densité du modèle GPD ajusté, elle peut être comparée à un histogramme des dépassements de seuil.

4 Revue de la Littérature

4.1 Applications de la TVE dans le domaine des soins de santé

4.1.1 Modélisation de la mortalité liée à la pneumonie et à la grippe

Au travers d'applications peu communes de la Théorie des Valeurs Extrêmes, le but est de montrer que la TVE peut s'appliquer à la santé publique et en épidémiologie, plus précisément, pour prédire de futurs événements extrêmes. Afin d'illustrer cela, l'accent a été mis sur la modélisation des pics de mortalité due à la pneumonie et à la grippe [Thomas et al. ; 2016].

Pour le cas de maladie respiratoire et grippale, la GEV [Coles ; 2001] a été ajustée aux valeurs extrêmes issues de la méthode des maximas de blocs. Thomas et al. (2016) ont utilisé des données de mortalité hebdomadaire pour les États-Unis entre 1962 et 2011. Après la modélisation des pics, ils ont constaté que les pics de mortalité étaient plus fréquents pour les personnes de plus de 65 ans. Les résultats ont montré que les modèles basés sur la théorie des valeurs extrêmes étaient plus précis que les modèles traditionnels pour prédire les pics de mortalité dus à la pneumonie et à la grippe.

4.1.2 Modélisation des visites aux urgences dans un réseau hospitalier

Une modélisation des visites aux urgences dans un réseau hospitalier parisien, au début des années 2000 [Thomas et al. ; 2016], fut également une intéressante application du champ des possibles de la TVE.

La même méthodologie a été appliquée que dans la première application (4.1.1). Les auteurs ont utilisé des données de visites aux urgences pour un réseau de 12 hôpitaux parisiens entre 2003 et 2010. Après l'application de la TVE, ils ont constaté que les pics de visites étaient plus fréquents lors des périodes de fortes chaleurs. Les résultats ont montré que les modèles basés sur la théorie des valeurs extrêmes étaient utiles pour prédire les pics de visites aux urgences dans un réseau hospitalier et pour aider à planifier les ressources médicales nécessaires. Les modèles basés sur la théorie des valeurs extrêmes peuvent aider les décideurs en santé publique à mieux planifier les ressources et à minimiser les risques pour la santé publique [Thomas et al. ; 2016].

4.1.3 Modélisation des pics de mortalité et de morbidité pour les maladies cardiovasculaires

Cette troisième application concrète se focalise sur les pics d'hospitalisations et de décès pour les maladies cardiovasculaires (CVD), dans les métropoles de Montréal et de Québec, au Canada, pour la période 1981-2011 [Chiu et al., 2016]. Les niveaux de retour des pics ont été calculés à partir de la modélisation extrême.

Dans le cas des hospitalisations liées aux CVD à Montréal (**Figure 5**), Les pics d'hospitalisations sont extraits à l'aide de la méthode des dépassements de seuil [Coles ; 2001] avec un seuil fixé à 95 pourcents. Cette fois-ci, la distribution généralisée de Pareto (GPD) est ajustée aux données extrêmes. La distribution des hospitalisations pour CVD suit une distribution de valeurs extrêmes de type Weibull, avec un paramètre estimé de forme ξ de -0,37 et un paramètre estimé d'échelle σ de 13,14. Les auteurs ont utilisé leur modèle pour prédire le nombre d'hospitalisations extrêmes qui devraient avoir lieu à l'avenir. En guise d'extrapolation, le niveau de retour d'un an est de 200 hospitalisations par jour, signifiant que

l'on s'attend à dépasser 200 admissions quotidiennes, une fois durant les 12 prochains mois. Le niveau de retour de 10 ans était de 214 hospitalisations par jour [Chiu et al., 2016].

Pour les décès liés aux CVD au Québec (**Figure 6**), les pics de décès sont extraits des données brutes par la méthode des blocs auquel une distribution généralisée des valeurs extrêmes est appliquée. Après avoir réalisé des tests de WW et WX, la taille de bloc qui semblait convenir est 180 jours, c'est-à-dire des pics biannuels. Cela équivaut à 63 pics. Les paramètres estimés associés à la GEV sont $\mu = -0,14$, $\sigma = 8,57$ et $\xi = 1,23$. Le niveau de retour nous renseigne, qu'en moyenne une fois par an, huit décès seront atteints. Le niveau de retour associé à une période de 10 ans est de 11 décès [Chiu et al., 2016].

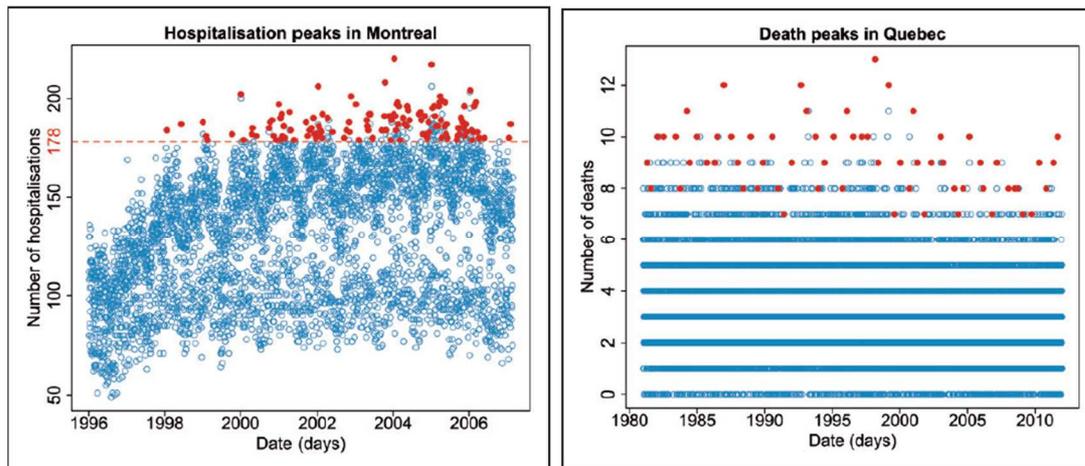


Figure 5 : Identification des pics d'hospitalisations à Montréal avec la méthode de dépassement de seuil

Figure 6 : Identification des pics de décès au Québec selon la méthode des maxima de blocs

La compréhension et la modélisation des pics de santé représentent un défi important en santé publique, parce que les techniques "classiques" basées sur les événements moyens, ne sont pas adéquates pour de tels événements [Raggad, 2009 ; Charras-Garrido et Lezaud, 2013]. La modélisation des pics conduit à des niveaux de retour d'hospitalisations et de décès, qui peuvent être utiles pour la prise de décisions en matière d'organisation des soins de santé. Cette étude peut également être utile pour rechercher des facteurs externes possiblement corrélés aux pics, tels que les conditions météorologiques, le stress social ou la pollution de l'air. Les résultats de cette étude ne signifient pas que les études "classiques" sont dépassées mais plutôt que les deux types d'études doivent être menés en complémentarité.

4.1.4 Modélisation de données Covid-19 : les événements de super-propagation

La modélisation des cas extrêmes de transmission communautaire du Covid-19 est une question cruciale pour évaluer les risques de transmission et prendre des décisions de santé publique appropriées. Cependant, cette modélisation est souvent difficile en raison de la nature complexe de la pandémie. Le cas pandémique décrit propose une méthode pour modéliser les cas extrêmes de transmission communautaire en utilisant des distributions discrètes de Pareto généralisées (GPD) [Daouia et al., 2023]. Les événements de super-propagation (SSE) peuvent être définies comme des foyers, des éclosions soudaines de contaminations dans lesquels un individu donné (le cas index) infecte un nombre de personnes (cas secondaires) nettement supérieur à un indicateur, tel que la moyenne des cas secondaires. Pour le SARS-CoV-1 et le SARS-CoV-2 [Cevik et al., 2020], les événements ayant déclenché plus de six cas secondaires ont été considérés comme des SSE.

Autrefois, il a été affirmé, en se basant sur la TVE, que les évènements dits de super-propagation sont à queue grasse, exprimé par un paramètre de forme compris entre 0 et 1 [Wong et Collins ; 2020]. Or, ce constat majeur s’est basé sur un regroupement restreint ($n = 60$) incluant à la fois les évènements du virus SARS-CoV-1 et du SARS-CoV-2. Les chercheurs [Daouia et al., 2023] ont remis en question ce résultat présent dans la littérature pour deux raisons. La qualité de l’analyse statistique de Wong et Collins est mauvaise car elle se base sur une taille d’échantillons de 60 observations historiques, ce qui est jugé trop faible. Ajouté à cela, le rassemblement des deux virus de la même famille de maladies respiratoires, n’a pas été jugé comme étant un bon choix. Il est possible, très facilement, de se rendre compte que les comportements statistiques des SSE pour le SARS-CoV-1 et le SARS-CoV-2 diffèrent.

Pour apporter une solution au premier problème, les chercheurs ont fait le choix d’analyser un plus grand échantillon [Daouia et al., 2023]. Ils ont sélectionné une base de données de cas secondaires en Inde pour l’année 2020 ($n = 88\,527$) et deux autres ($n = 5165$ et $n = 33\,903$) se focalisant sur la Corée du Sud à deux périodes différentes pour capturer deux variants différents du virus. Pour ces trois bases de données, les évènements de super-propagation faisaient référence au SARS-CoV-2 et aucun regroupement n’a été effectué. La distribution GPD discrète a, ensuite, été appliquée aux pics de propagation. Les valeurs excédentaires (les évènements $Z - u$), supérieurs au seuil $u = 6$, ont été conservées de manière à estimer les paramètres de la méthode des dépassements de seuil.

Ignorer la nature discrète de la variable aléatoire Z (le nombre de personnes infectées par le cas index) est une chose à éviter. Dans toutes les datasets analysées, les valeurs prises par Z oscillaient entre les nombres entiers 0 et 39. Vu le nombre de données, il était courant d’observer des égalités entre les valeurs prises par les SSE. Grâce à l’utilisation de la distribution GPD discrète, cela mène à un ajustement plus lisse et stable du paramètre de forme ξ à estimer par rapport aux estimateurs de Hill et celui issu de la GDP continue (**Figure 7**). Par exemple, l’estimation du paramètre ξ est d’environ 0,24 pour les cas secondaires en Inde. La GDP discrète confirme bien la distribution à queues grasses mentionnée par Wang et Collins mais de manière bien moins flagrante.

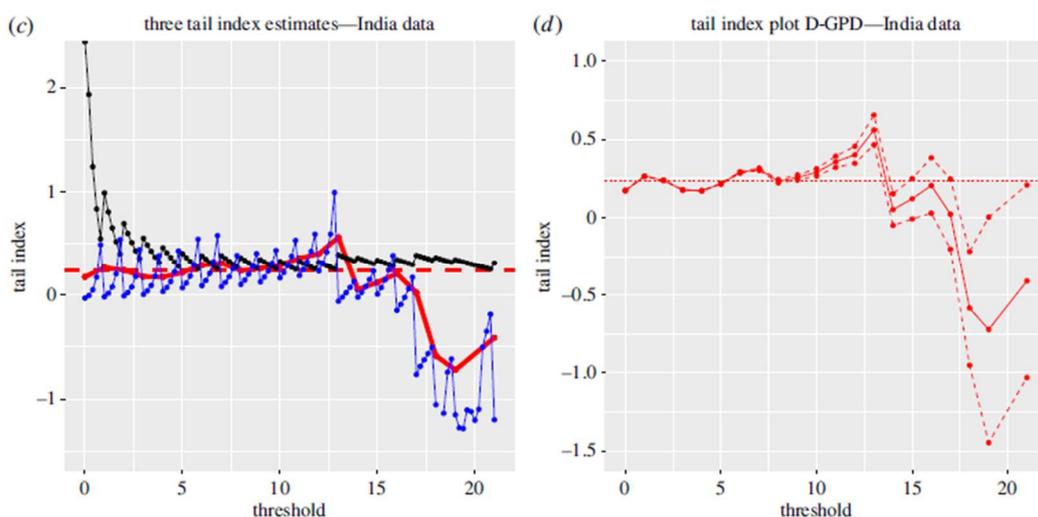


Figure 7 : Estimation du paramètre de forme ξ selon la GPD discrète (rouge), la GDP continue (bleu) et l’estimateur d’Hill (noir).

En conclusion, cette application aux événements de super-propagation, présente une méthode prometteuse pour modéliser les cas extrêmes de transmission communautaire du SARS-CoV-2 en utilisant des distributions GPD. Les résultats montrent que la distribution de Pareto discrète a fourni une meilleure estimation du paramètre de forme ξ . L'analyse de ces ensembles de données a permis de conclure que la distribution du nombre de cas secondaires est soit à queue grasse, mais avec un indice de valeur extrême ξ faible, soit même à queue de type exponentiel. Cette modélisation à l'aide de la GDP discrète peut être utilisée pour évaluer les risques de transmission future. Cela pourrait être utile pour guider les décisions de santé publique en identifiant les zones à risque élevé de transmission communautaire.

4.2 Analyse univariée : Indice composite de sévérité des mesures

4.2.1 Construction de l'Oxford Covid-19 Government Response Tracker

La base de données, appelée « Ox-CGRT » présente une vue d'ensemble d'indicateurs des politiques gouvernementales en réponse à la pandémie de COVID-19 [Hale et al. ; 2021]. L'Oxford COVID-19 Government Response Tracker (Ox-CGRT), est conçu pour être utilisé par les chercheurs, les décideurs politiques et le grand public afin de mieux comprendre la manière dont les gouvernements ont réagi à la pandémie. Hale et al. (2021) précisent la manière dont l'Oxford Covid Government Response Tracker a été développé. Les interventions politiques sont encodées en tant que séries temporelles, enregistrant l'intensité et la portée d'un indicateur dans un lieu et temps donné. Le choix du score enregistré selon l'indicateur est propre aux 400 collecteurs de données. Par conséquent, la précision et la cohérence des données sont dépendantes du jugement humain et de l'expertise contextuelle

Il semble important de souligner que le dataset Ox-CGRT enregistre uniquement le nombre et le degré d'intensité des réponses gouvernementales face au Covid-19. En aucun cas, ces données permettent de mesurer la qualité de leur mise en œuvre, ni le degré d'application par les citoyens [London School of Economics and Political Science, 2020]. Les données mises à notre disposition doivent être considérées comme un élément clé parmi d'autres pour comprendre l'adoption de politiques et les liens avec la propagation du virus, les décès attribués au Covid-19 et le comportement humain.

Le but principal des chercheurs était de construire un indice global, appelé indice composite, pour évaluer la sévérité de l'ensemble des mesures politiques adopté par un pays à un temps donné. Cet indice composite se base sur une série de sous-indices, chacun mesurant un aspect différent de la réponse gouvernementale à la pandémie (Tableau 2). Les sous-indices comprennent des indicateurs tels que les restrictions de mouvement, les fermetures d'entreprises, les politiques de test et de traçage, les restrictions de rassemblement, les politiques de voyage et les campagnes de vaccination. Chaque sous-indice est pondéré en fonction de son importance relative dans la réponse générale du gouvernement à la pandémie. Au niveau de la collecte, les données sont enregistrées à partir de sources gouvernementales officielles, telles que les sites web des ministères de la santé, les annonces officielles des gouvernements et les rapports des médias locaux [Hale et al., 2020]. Les données sont ensuite codées selon une échelle numérique standardisée, qui va de 0 à 100, où 0 représente l'absence totale de politique et 100 représente la politique la plus stricte. Les scores de chaque sous-indice sont ensuite agrégés pour produire un score composite pour chaque pays et chaque jour de la pandémie.

Table 1 OxCGRT Indicators			
ID	Name	Type	Targeted/ general?
Containment and closure			
C1	School closing	Ordinal	Geographic
C2	Workplace closing	Ordinal	Geographic
C3	Cancel public events	Ordinal	Geographic
C4	Restrictions on gathering size	Ordinal	Geographic
C5	Close public transport	Ordinal	Geographic
C6	Stay-at-home requirements	Ordinal	Geographic
C7	Restrictions on internal movement	Ordinal	Geographic
C8	Restrictions on international travel	Ordinal	No
Economic response			
E1	Income support	Ordinal	Sectoral
E2	Debt/contract relief for households	Ordinal	No
E3	Fiscal measures	Numerical	No
E4	Giving international support	Numerical	No
Health systems			
H1	Public information campaign	Ordinal	Geographic
H2	Testing policy	Ordinal	No
H3	Contact tracing	Ordinal	No
H4	Emergency investment in health care	Numerical	No
H5	Investment in COVID-19 vaccines	Numerical	No
H6	Facial coverings	Ordinal	Geographic
H7	Vaccination policy	Ordinal	Funding
Miscellaneous			
M1	Other responses	Text	No

For detailed descriptions and coding information, see the project website: <https://www.bsg.ox.ac.uk/covidtracker>.

Table 2 OxCGRT index composition and values							
	GRI	CHI	Stringency index	ESI	Legacy stringency index	Maximum value (N_i)	Flag (F_i)
k	13	11	9	2	7		
C1	x	x	x		x	3 (0, 1, 2, 3)	Yes=1
C2	x	x	x		x	3 (0, 1, 2, 3)	Yes=1
C3	x	x	x		x	2 (0, 1, 2)	Yes=1
C4	x	x	x			4 (0, 1, 2, 3, 4)	Yes=1
C5	x	x	x		x	2 (0, 1, 2)	Yes=1
C6	x	x	x			3 (0, 1, 2, 3)	Yes=1
C7	x	x	x		x	2 (0, 1, 2)	Yes=1
C8	x	x	x		x	4 (0, 1, 2, 3, 4)	No=0
E1	x			x		2 (0, 1, 2)	Yes=1
E2	x			x		2 (0, 1, 2)	No=0
E3							
E4							
H1	x	x	x		x	2 (0, 1, 2)	Yes=1
H2	x	x				3 (0, 1, 2, 3)	No=0
H3	x	x				2 (0, 1, 2)	No=0
H4							
H5							
H6	x	x				4 (0, 1, 2, 3, 4)	Yes=1
H7	x	x				5 (0, 1, 2, 3, 4, 5)	Yes=1
M1							

k is the number of component indicators in an index (see Methods); x indicates that an indicator contributes to that index.

Tableau 2 : Tableau récapitulatif des indicateurs mesurés par l'Ox-CGRT et classification en sous-indices

4.2.2 Constats relatifs à l'Ox-CGRT

Les créateurs de l'Ox-CGRT ne se sont pas seulement limités à nous offrir une base de données précise, transparente et facilement transposable à d'autres analyses. Ils ont également exploité leur base de données pour examiner les politiques gouvernementales mises en place dans 186 pays et tirer d'importantes conclusions. Le constat principal de cette analyse est le suivant : les politiques gouvernementales ont varié considérablement d'un pays à l'autre plus le temps passait [Hale et al. ; 2021].

En effet, les résultats montrent que, au début de la pandémie, de nombreux gouvernements ont mis en place des mesures similaires sans tenir compte des données épidémiologiques nationales. La plupart des gouvernements ont adopté en même temps un haut de niveau de réponse à la mi-mars 2020. A ce stade de la pandémie mondiale, des mesures fortes étaient appliquées aussi bien par des gouvernements faisant face à des dizaines de morts mais également par des états n'ayant pas encore enregistré leur dixième cas.

Les mois suivants ont été caractérisés par plus de variations au niveau de la réponse gouvernementale des pays. Le schéma commun n'a plus été observé d'une manière aussi unanime : la phase d'adoption étant suivie de politiques spécifiques propres aux gouvernements, qu'elles soient maintenues ou suspendues. Au fil du temps, ils ont commencé à adapter leurs propres politiques en fonction de la situation épidémiologique et de l'impact économique sur les entreprises et les travailleurs. Les pays asiatiques ont généralement obtenu des scores plus élevés en raison de leur réponse rapide et coordonnée à la pandémie, tandis que

les pays européens et américains ont enregistré des scores plus faibles en raison de la complexité de leurs systèmes politiques et de santé.

A côté de cela, une autre constatation frappante concerne l'ordre d'adoption des mesures politiques à travers les pays [Hale et al. ; 2021]. Plus spécifiquement, en 2020, il y avait plus de 50 pourcents de chance qu'un pays tiré au sort ait mis en place des campagnes publiques d'information, des contrôles sur les voyages internationaux et des mesures de test dans les 20 jours suivant la première réponse du gouvernement. Des politiques de soutien économique ont eu tendance à être déployées globalement dans un second temps. Au cours des deux premiers mois d'assouplissement des politiques, les politiques de fermeture et de confinement ont été réduites tandis que les politiques de soutien économique et les politiques de santé ont été maintenues à leur niveau maximal peu importe le pays observé. Cependant, des divergences ont été recensées au niveau de la force des réponses gouvernementales, de leur couverture géographique, de l'intensité et la fréquence dont elles ont été réimposées ultérieurement.

4.2.3 Analyse géographique de l'Oxford COVID-19 Government Response Tracker

A l'image du point précédent (4.2.2), il a été tenté de savoir quelles mesures ont été privilégiées par les gouvernements sur une période de deux ans. Cette fois-ci, des macro-tendances ont pu être dégagées avec des enseignements différents selon l'endroit où l'on se trouve sur le globe [Hale ; 2022]. Le but ici n'est pas de chercher un lien de causalité entre la sévérité des mesures et des effets sanitaires. Il est bon de souligner que ce nouvel article de Hale propose un recul plus important sur la situation étant donné la plus grande période considérée.

En ce qui concerne le premier mois d'adoption des mesures Covid-19, nous pouvons à nouveau parler d'effet d'imitation entre les pays du monde entier [Alemanno, 2020]. Cela peut être aperçu graphiquement à la **Figure 8** : en un laps de temps excessivement court, quasiment 100 pourcents des pays ont adopté les mêmes restrictions. A partir de ce pic atteint au printemps 2020, le nombre de gouvernements employant des mesures a connu une diminution constante, hormis des recrudescences associées aux variants Delta et Omicron. En revanche, une restriction a suivi une courbe totalement différente : il s'agit de l'obligation du port du masque qui s'est généralisée à la quasi-totalité des pays.

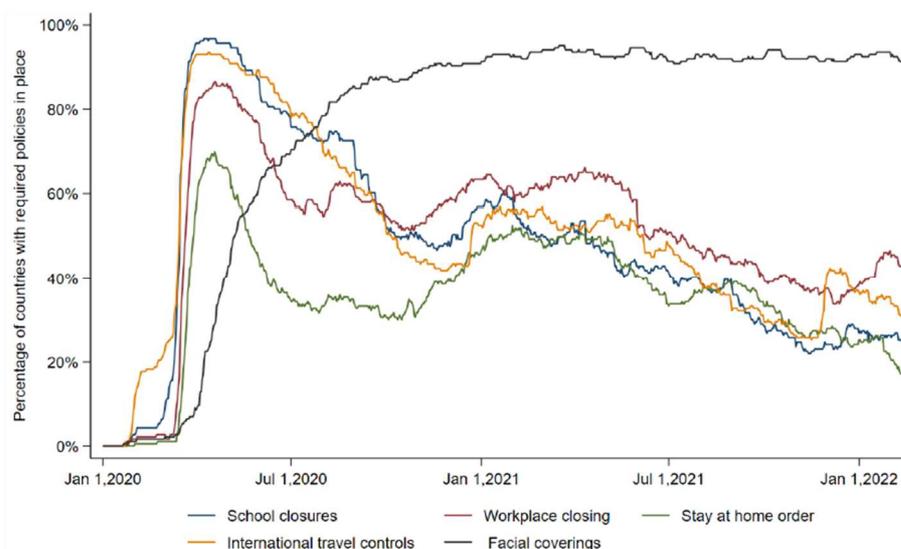


Figure 8 : Evolution de la couverture géographique des mesures Covid-19 de 2020 à 2022

Les auteurs ont dégagé deux régimes de gestion de la pandémie. Il y avait tout d'abord les pays « zero-covid », c'est-à-dire ceux qui cherchaient à éliminer le Covid-19 à travers des confinements radicaux et le maintenir à distance grâce à des mesures strictes de contrôle des déplacements [Hale ; 2022]. Cette approche était particulièrement courante en Asie orientale et dans la région du Pacifique. D'autre part, la majorité des pays d'Europe, d'Afrique et des Amériques visaient à réduire la propagation du virus sans pour autant l'éliminer. Les restrictions étaient renforcées ou atténuées au fur et à mesure des vagues d'infection.

Afin d'avoir une évaluation approximative des pays qui se sont le mieux débrouillés, des macro-patterns ont été dégagés. Ceux-ci regroupaient quatre dimensions telles que la santé économique et des métriques en matière de santé pour comprendre si elles étaient corrélées avec les réponses gouvernementales (Figure 9). La combinaison idéale de ces facteurs est un pays qui a connu un nombre faible de jours sous l'index SI70, un taux de croissance du PIB positif, un nombre raisonnable de morts tout en vaccinant une grande proportion de sa population.

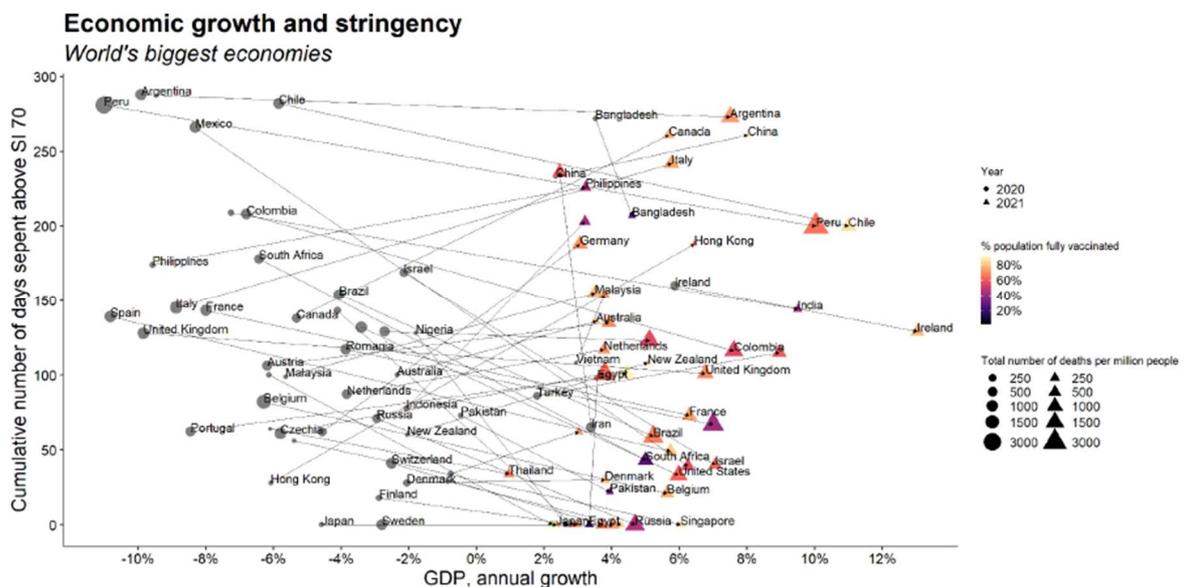


Figure 9 : Graphique mettant en relation la croissance économique, l'indice de sévérité des mesures, le nombre de morts et la proportion de la population vaccinée par pays pour les années 2020 et 2021.

Pour la période 2020-2022, des tendances ont pu être assignées à chaque région du monde, là où des clusters se sont créés. De manière générale, l'Asie de l'Est et les pays du Pacifique s'en tirent mieux que le reste du monde. Ils ont connu une croissance économique positive et un nombre de morts par habitant contrôlé. Quant à l'Europe, la croissance économique est comparable mais la divergence est notable au niveau des longues périodes de restrictions strictes et vis-à-vis du niveau de mortalité qui est nettement supérieur aux deux autres régions du globe. Pour finir, l'Amérique du Sud, les Caraïbes et l'Afrique subsaharienne ne montrent pas de tels regroupements. La croissance économique et l'indice de sévérité y étant peu homogènes entre états. Concernant la mortalité, on retrouve néanmoins des niveaux comparables à ceux rencontrés en Europe pour la plupart des pays d'Amérique Latine.

4.3 Analyse multivariée : Lien entre la sévérité des mesures gouvernementales et le nombre de décès Covid-19

4.3.1 Analyse de la corrélation par région du monde

Cette section évalue la relation entre les mesures de réponse gouvernementale à la pandémie de COVID-19 et le nombre de décès dus à la maladie. L'analyse relatée a mobilisé des données de 170 pays et les statisticiens ont utilisé un modèle de régression pour évaluer l'impact des différentes mesures de réponse sur le nombre de décès [Hale et al. ; 2020]. L'objectif de cette analyse était d'avoir une première évaluation rapide de l'impact de la sévérité des mesures gouvernementales sur le taux de croissance de la mortalité associée au Covid-19. Les résultats de l'étude ont montré que les pays qui ont adopté ces mesures plus tôt dans la pandémie, ont eu une meilleure performance dans la réduction du nombre de décès.

A partir du tracker Ox-CGRT, neuf interventions non-pharmaceutiques parmi les 19 ont été choisies par les auteurs pour étudier la relation. Cela inclut les fermetures des écoles et des lieux de travail, les restrictions sur les voyages internationaux et nationaux, l'interdiction des rassemblements publics, etc. Ensemble, ces neuf métriques forment l'indice de sévérité SI. Cet indice est également la variable indépendante de la régression linéaire. De la sorte, une valeur ordinale globale, comprise entre 0 et 100, est obtenue quotidiennement pour 170 pays.

Concernant la variable dépendante de la régression, les données sur les cas et décès du Covid-19 proviennent du Centre Européen de Prévention et de Contrôle des Maladies (ECDC) et de la Johns Hopkins University pour les régions autonomes chinoises, Hong Kong et Macau. L'indicateur de mortalité a été traduit selon deux aspects. Premièrement, la croissance de la mortalité due au Covid-19 était abordée en termes de différence logarithmique quotidienne. Dans un second temps, le nombre maximal quotidien de nouveaux morts était répertorié de manière logarithmique également. De la sorte, deux types d'information étaient capturés : les premières faisant référence à la pente de la courbe de mortalité et les secondes, aux pics de mortalité.

Avant de réaliser la régression, les auteurs ont réalisé quelques ajustements. Dans la première partie de leur méthode, ils ont cherché à inclure une nouvelle variable indépendante mesurant la vitesse des réponses gouvernementales. Cette dernière a été définie comme le nombre de jours entre le premier cas Covid-19 enregistré par un pays et le jour auquel l'index SI a dépassé le score de 40 sur 100. Ensuite, il a fallu contrôler l'effet du temps. Par exemple, une relation positive entre le SI actuel et les nouveaux morts ne signifient pas nécessairement que des mesures sévères augmentent la mortalité mais plutôt que la hausse des morts déclenche une politique de réponse gouvernementale. Pour résoudre ce problème, la décision prise était de retarder la variable explicative SI de six semaines dans les modèles de séries temporelles [Donovan et al. ; 2021]. Cela permettait d'accorder les interventions gouvernementales passées avec les changements actuels du niveau de mortalité.

Les régressions, adoptant la méthode des moindres carrés ordinaires, pouvaient être construites afin d'estimer la corrélation entre la vitesse et la force de la sévérité des mesures et la croissance de la mortalité et le nombre quotidien maximale de nouveaux morts. Trois grands enseignements étaient tirés de ces régressions [Hale et al. ; 2020] : les deux premiers à propos de la vitesse d'implémentation des mesures et le troisième concernant le degré de sévérité des mesures. Le premier résultat a montré une relation positive entre le nombre maximale de décès quotidiens et une réponse gouvernementale retardée. Chaque jour de retard supplémentaire pour atteindre la valeur de 40 sur l'indice de sévérité SI depuis le premier cas dans un pays

donné est associé à neuf décès quotidiens supplémentaires au pic de la courbe épidémique. A propos du taux de croissance, il existait une même corrélation positive. Chaque jour de retard supplémentaire pour atteindre 40 à l'indice SI correspond à une hausse du taux de croissance de 0,083% du nombre de décès quotidiens. Enfin, au plus l'indice de sévérité d'un pays était élevé, au moins le taux de croissance de mortalité était important. Chaque point supplémentaire obtenu à l'indice SI est associé à une baisse moyenne du taux de croissance des décès de 0,072%, six semaines plus tard.

Deux derniers outputs (**Figure 10 et 11**) se focalisent sur les effets à l'échelle des pays. Le Botswana et la Corée du Sud confirment les constats précédents. Ayant agi vite et de manière sévère, ils ont connu très peu de décès au sein de leur population. A l'inverse, le gouvernement espagnol est seulement intervenu après le premier mort attribué au Covid-19. Cela a donc pris plus de temps pour que son taux de croissance converge vers des valeurs acceptables. Les Etats-Unis se sont fait remarquer par la plus lente et la plus faible réponse [Piquero et Kurland ; 2022]. Par conséquent, c'est le pays qui a mis le plus de temps à reprendre le contrôle sur l'épidémie.

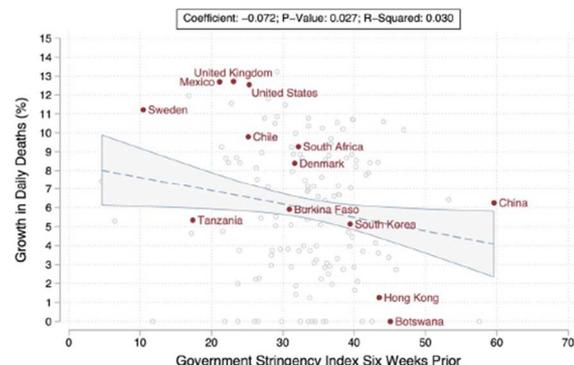
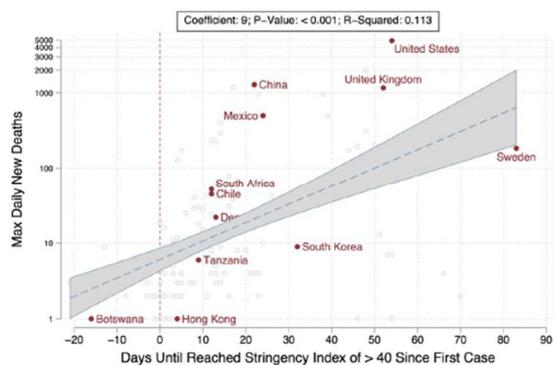


Figure 10 : Régression linéaire entre la vitesse d'implémentation des mesures gouvernementales et le nombre quotidien maximale de nouveaux décès

Figure 11 : Régression linéaire entre la force des mesures gouvernementales et la croissance quotidienne de la mortalité

En conclusion, les résultats de la régression linéaire ont clairement montré qu'au plus sévère et au plus tôt était la réponse nationale, au plus bas était le nombre de morts, six semaines plus tard. Au fil du temps, cela se traduit par des différences importantes au niveau du nombre total de morts. Ces découvertes rajoutent de la crédibilité aux mesures non-pharmaceutiques en l'absence de la mise sur le marché d'un vaccin à cette période-là de la pandémie.

4.3.2 Analyse de la corrélation par vague épidémiologique

Cette dernière insertion de connaissances scientifiques vise à explorer la relation entre les mesures de réponse gouvernementales et le nombre de décès liés à la COVID-19 à l'échelle mondiale au cours de plusieurs vagues pandémiques [Hale, Angrist, Hale, et al., 2021]. La base de données Ox-CGRT a été utilisée pour mesurer les différents types de mesures de réponse gouvernementales mises en place dans les pays du monde entier. La période analysée présente une plus grande étendue que celle du point précédent (4.3.1), où l'évaluation se limitait à des observations jusqu'en mai 2020.

De manière similaire à l'article précédent, neuf interventions non-pharmaceutiques de l'OxCGRT ont été retenues et ensemble, elles forment l'indice de sévérité SI. Cette variable explicative a été ensuite mise en relation avec le registre des décès Covid-19 provenant de la Johns Hopkins University. Il s'agit de données quotidiennes, du 1^{er} janvier 2020 au 11 mars

2021, dans 113 pays. Cette démarche vise à garder uniquement les pays ayant connu au moins, en moyenne, un mort journalier sur cet horizon de temps. Afin de connaître si les mesures gouvernementales sont efficaces à travers les différentes phases pandémiques, il était crucial de séparer les pays selon leur nombre de vagues. Cette méthode présente comme avantage une meilleure calibration du modèle à la pente spécifique de la courbe au sein de chaque vague, par rapport à un modèle universel. Même à l'intérieur d'un pays, les vagues ont souvent des pics, des creux et des pentes très variés.

Quant à la définition d'une vague pandémique, il n'y a pas de consensus précis dans la littérature actuelle. Cependant, plusieurs études ont relevé des caractéristiques communes aux vagues telles qu'une plus grande vulnérabilité des personnes âgées et un changement considérable de la structure démographique des cas et décès, d'une vague à l'autre [Hale, Angrist, Hale, et al., 2021]. On peut résumer une vague comme étant une phase de la maladie qui est plus conséquente qu'une flambée sporadique et qui comprend une phase ascendante et une phase descendante subséquente.

Différentes étapes ont été suivies pour identifier les vagues dans les 113 pays via une approche fondée sur les données. Premièrement, il a fallu localiser les pics et les creux de tous les pays afin de séparer les potentielles vagues. Dans un second temps, une régression LOWESS (locally weighted scatterplot smoothing) a été ajustée au nuage de points. Grâce à la courbe lissée en découlant (**Figure 12**), cela a permis de ne pas identifier du bruit et des fluctuations soudaines. Enfin, les auteurs se sont accordés pour considérer une vague distinctivement si la période creuse entre deux vagues est plus grande qu'un mois.

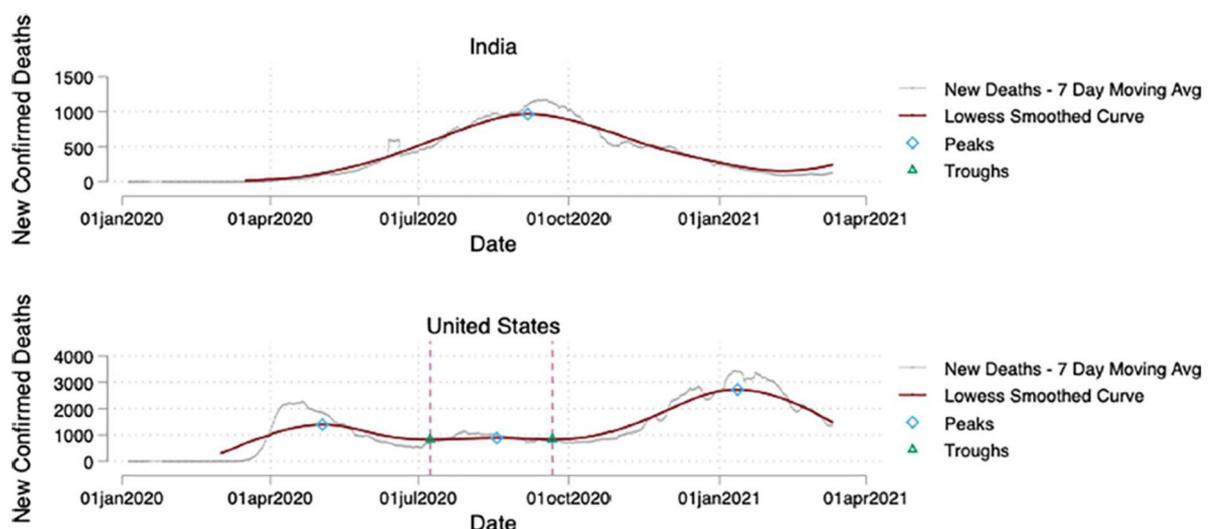


Figure 12 : Identification des vagues pandémiques par la régression LOWESS en Inde et aux États-Unis

Au niveau de la méthodologie, les données relatives aux décès sont analysées sous la forme de logarithme naturel quotidien dans un pays. Cette transformation log tient compte de la trajectoire exponentielle de la courbe épidémiologique et facilite l'interprétation de la régression OLS parce que les coefficients représentent des variations en pourcentage. Lors de cette étude, les chercheurs ont appliqué un retard de quatre semaines à la variable explicative pour refléter au mieux le délai entre le changement de comportement des individus, la transmission, la contamination et finalement le décès. Cette estimation a été basée sur la littérature existante [Li et al. ; 2021].

De manière générale, il a une nouvelle fois été montré qu'au plus la réponse gouvernementale était forte dans un pays, au moins il y avait de morts liés au SARS-CoV-2. Les résultats globaux ont montré une relation persistante entre les politiques et les décès, quels que soient les pays et les vagues. Des différences ont été observées dans le nombre de vagues Covid-19 que les pays ont rencontrés. En mars 2021, la plupart des pays (63) avaient expérimenté deux vagues, 40 seulement une vague et dix pays étaient confrontés à trois vagues ou plus [Hale, Angrist, Hale, et al., 2021]. Dans les 40 pays ayant été confrontés à une vague, une réduction quotidienne de 0,005% des décès était associée à chaque point supplémentaire de sévérité de l'index SI. Dans les 63 pays ayant connu trois vagues, des réductions de la mortalité de 0,021%, 0,031% et 0,028% étaient associées respectivement à la première, deuxième et troisième vague. Tous les coefficients étant significativement différents de zéro, les p-values ne dépassant pas le seuil 0,001. Etant donné que différents groupes de pays ont fait face à une, deux ou trois vagues de pandémie, il n'est pas possible de comparer l'ampleur des coefficients entre les différents groupes. De plus, il est logique de s'attendre à une plus grande corrélation négative pour les pays touchés par plusieurs vagues parce que les données tendent à être plus extrêmes et moins étalées dans le temps par rapport à une vague unique.

En guise de discussion, les articles remettaient en question la pertinence des mesures non-pharmaceutiques dans le temps. Certes, il a été prouvé qu'elles avaient un effet réel sur le niveau de mortalité mais pouvait-on s'attendre à ce que cela dure ? D'un côté, l'on pouvait supposer une plus grande efficacité des gouvernements tant ils ont appris les enseignements de leurs politiques Covid. En effet, ils pourraient mieux cibler leurs mesures grâce à leurs différentes expériences passées. D'un autre côté, la lassitude de la population face à des interventions politiques qui n'en finissent plus est un facteur à prendre en compte. Si la population ne suit plus les recommandations des politiques, même les mesures les plus efficaces ne porteraient pas leurs fruits.

5 Résultats de l'analyse des données extrêmes

5.1 Phase de préparation des données

5.1.1 Transformation de la variable « Hospitalisations quotidiennes » en différences logarithmiques

Dans le cadre de cette analyse basée sur la Théorie des Valeurs Extrêmes (TVE), l'hypothèse de stationnarité et l'hypothèse d'indépendance temporelle des hospitalisations quotidiennes associées au Covid-19 qui en découle, nous oblige à transformer nos données de départ. Comme montré dans la section 2, il est clair que ces dernières, ne sont ni indépendantes, ni identiquement distribuées tant elles dépendent du temps (**Figure 1**). Le recours au log quotidien réduit l'échelle des données et a tendance à homogénéiser le champ de données. Cela nous assure de sélectionner des observations tout au long de la période étudiée et pas que durant le pic le plus important.

Par définition des logarithmes, cette approche se concentre sur des variations relatives en pourcentage. Par conséquent, les interprétations des résultats seront exprimées en termes d'augmentation et de diminution du nombre d'hospitalisations quotidiennes et pas selon un nombre absolu d'hospitalisations quotidiennes.

5.1.2 Application d'un décalage temporel

De manière à associer l'effet d'un package de mesures prise à un certain temps avec le niveau d'hospitalisations qui y est associé, il était important d'estimer le temps moyen entre une infection au Covid-19 et l'admission potentielle en soins hospitaliers. Selon plusieurs sources scientifiques, cette période oscillerait, en moyenne, entre neuf et douze jours. Cela comprend la phase d'incubation, c'est-à-dire la période entre la contamination et l'apparition des premiers symptômes. Sa durée moyenne a été estimée entre cinq et six jours. Ensuite, il faut ajouter le temps entre ces symptômes visibles et la prise en charge en milieu hospitalier pour cette contamination grave au Covid-19. Une étude anglaise⁸ a démontré que cette seconde phase durait en moyenne quatre jours. Une étude belge⁹ présentait un intervalle plus large de trois à dix jours. A partir de ces recherches, nous en avons déduit qu'il se passait en moyenne dix jours entre une infection Covid-19 et son hospitalisation. Bien sûr, il s'agit d'une moyenne qui ne reflète pas exactement la réalité pour chaque individu et le type de variant. En effet, une personne âgée de plus de 65 ans se rendrait en moyenne deux jours plus tard à l'hôpital. Les variants tels qu'Omicron ont été caractérisés par un très haut degré de contamination, cela se manifestait via une phase d'incubation plus courte, jusqu'à quatre jours en moyenne.

Dès lors, nous avons appliqué un lag de dix jours à la date d'implémentation des mesures non-pharmaceutiques. Cela correspondrait au temps moyen pour être admis en centre hospitalier. Par conséquent, la réponse gouvernementale coïncidait avec le niveau d'hospitalisations. Nous ne pouvions pas garder notre base de données d'origine parce que le nombre d'hospitalisations pour un certain jour ne reflète pas immédiatement la politique sanitaire adoptée en temps réel.

⁸<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticle/wavesandlagsocovid19inenglandjune2021>

⁹<https://pubmed.ncbi.nlm.nih.gov/33080869/>

5.1.3 Sélection des différences logarithmiques stables

Bien que les log-différences ont rendu les valeurs prises par les hospitalisations quotidiennes plus stables, nous remarquons une zone de turbulence pour les relevés les plus anciens. Effectivement, les logarithmes variaient considérablement dans le courant du mois de mars et au début du mois d'avril 2020. Ces données avaient tendance à se stabiliser à partir du 4 avril 2020. En retirant les différences logarithmiques antérieures, la non-stationnarité est ainsi diminuée. Nous nous limitons au 4 avril 2020 pour veiller à toujours englober la première vague de contamination, se terminant à la fin du mois de mai 2020. En résumé, notre analyse s'étendra du 4 avril 2020 au 31 octobre 2021, le moment où il a été jugé que la vaccination commençait à exercer une certaine influence sur le niveau d'hospitalisations.

5.1.4 Séparation des « Hospitalisations quotidiennes » en fonction de l'intensité des mesures gouvernementales

Dans un second temps, une approche multivariée a été prévue afin de mettre en relation le nombre d'hospitalisations associées au Covid-19 avec le niveau de sévérité des mesures politiques. Les hospitalisations ont donc été divisées en deux échantillons, à la fois en Belgique et Portugal, selon l'indice de sévérité SI dont nous disposons. Dès lors, nous avons décidé de scinder nos deux populations à partir de la moyenne du SI et en tenant compte du lag estimé de dix jours. Les valeurs prises par les moyennes de l'index SI sont respectivement de 57,76 et 65,37 pour la Belgique et le Portugal pour l'intervalle de temps du 4 avril 2020 au 31 octobre 2021. Les sous-ensembles de données relatifs à des mesures faibles correspondaient aux jours d'hospitalisations dont l'indice SI était inférieur à sa moyenne dix jours plus tôt. A l'inverse, les échantillons restants d'hospitalisations belges et portugaises regroupaient les enregistrements pour lesquels le niveau de rigueur des mesures associé était supérieur au niveau moyen des mesures sur la période étudiée. Cette distinction, entre des hospitalisations ayant eu lieu en présence d'une politique sanitaire faible ou forte, permettra d'évaluer s'il existe réellement des différences au niveau du comportement extrême entre ces deux sous-ensembles.

5.2 Résultats issus de la Théorie des Valeurs Extrêmes

Dans le cas des données d'hospitalisations quotidiennes, l'attention est portée sur les jours où les différences logarithmiques du nombre d'hospitalisations dépassent un certain seuil prédéfini. Au cours de cette phase de modélisation, il y a deux principaux éléments à décoder : le seuil choisi (u) et les paramètres estimés de la distribution généralisée de Pareto. Nous commencerons par sélectionner le seuil idéal au travers de différents quantiles et enfin pour ce seuil choisi, nous serons en mesure d'estimer les paramètres d'échelle, de forme et de position. Cette sélection du seuil est cruciale parce qu'il déterminera quelles observations seront retenues pour l'analyse des queues de distribution.

5.2.1 Analyse des hospitalisations

Dans un premier temps, les modèles de la GPD pour la Belgique et le Portugal seront comparés sans tenir compte de l'influence de la réponse gouvernementale sur le nombre de personnes hospitalisés à cause du virus.

5.2.1.1 Détermination du seuil

Dans la pratique, le seuil de dépassement fait référence à la valeur prise par un quantile de distribution. Les quantiles les plus fréquemment utilisés comme seuils dans la GPD sont généralement compris entre 90 et 99 pourcents. Un seuil de 90 pourcents correspond à une probabilité de dépassement de 10 pourcents alors qu'un seuil de 99 pourcents correspond à une probabilité de dépassement d'un pourcent. Au plus le seuil est élevé, au plus l'accent est

mis sur les événements extrêmes et critiques mais au moins l'échantillon contient d'observations. Il fallait, dès lors, tenir compte de ce compromis entre précision et manque de données.

Nous avons utilisé la méthode du maximum de vraisemblance à une gamme de seuils, afin d'estimer un modèle GDP. En guise d'output, nous obtenions deux graphiques montrant les estimations du maximum de vraisemblance et les intervalles de confiance des paramètres de forme et d'échelle pour une gamme de seuils (Figure 13 et 14). L'intervalle de seuil u acceptable apparaît lorsque les estimateurs des paramètres sont stabilisés. Pour la Belgique, les paramètres d'échelle et de forme semblaient fixes, visuellement, pour des seuils allant de 0,07 à 0,08. Le barème garantissant une certaine forme de constance pour le Portugal, allait de 0,06 jusque 0,07. De plus, les intervalles de confiance présentaient le plus haut niveau de précision pour ces valeurs de seuils.

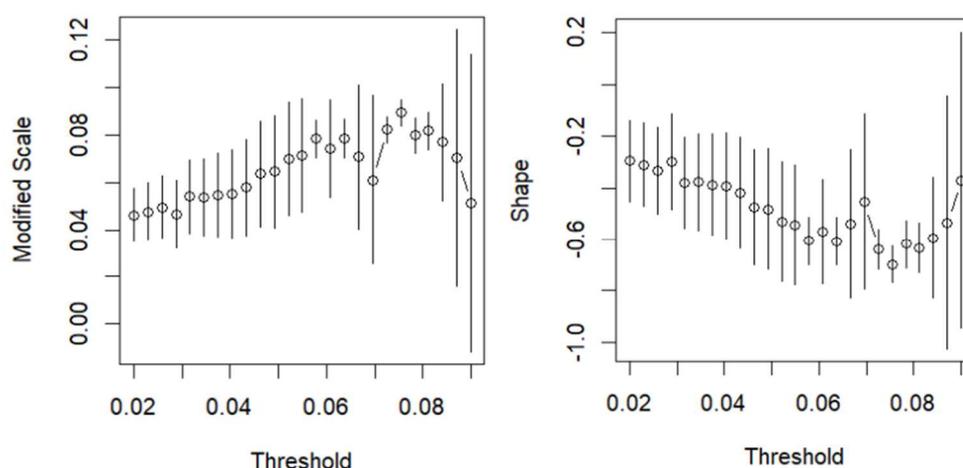


Figure 13 : Estimation du seuil optimal à partir des paramètres d'échelle et de forme en Belgique

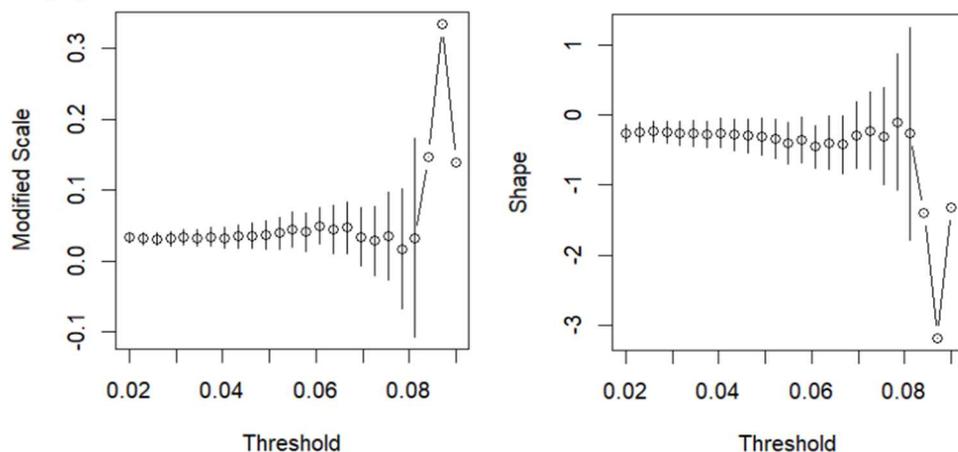


Figure 14 : Estimation du seuil optimal à partir des paramètres d'échelle et de forme au Portugal

Maintenant que nous avons une idée de la valeur prise par le seuil, il fallait trouver un quantile de distribution qui correspondait aux intervalles plus ou moins stables. Pour la Belgique (Figure 13), Des log-différence de 0,02 ; 0,04 ; 0,06 ; 0,08 ; 0,10 correspondaient respectivement aux quantiles 69,44 ; 82,64 ; 91,15 ; 95,14 ; 97,92. Pour le Portugal (Figure 14), il s'agit des quantiles 68,06 ; 85,59 ; 94,97 ; 98,44 ; 99,83 ; pour ce même intervalle logarithmique de 0,02 à 0,10. Après avoir remplacé ces valeurs critiques par les quantiles, nous

en déduisant que le quantile 95% faisait partie de l'intervalle de seuils stabilisés dans les deux populations étudiées. Il pouvait donc être choisi comme seuil de référence pour la modélisation GPD. Les quantiles 95% de la Belgique et du Portugal correspondant respectivement aux valeurs 0,079 et 0,060. Nous confirmons la pertinence de ces seuils en appliquant un Quantile-Quantile plot aux observations dépassant u (**Figure 15 et 16**). Dans les deux territoires analysés, les observations extrêmes s'alignaient relativement près de la ligne de référence diagonale. Cela suggérait une bonne concordance entre les cinq pourcents des observations jugées extrêmes et la distribution théorique suivie par la GPD.

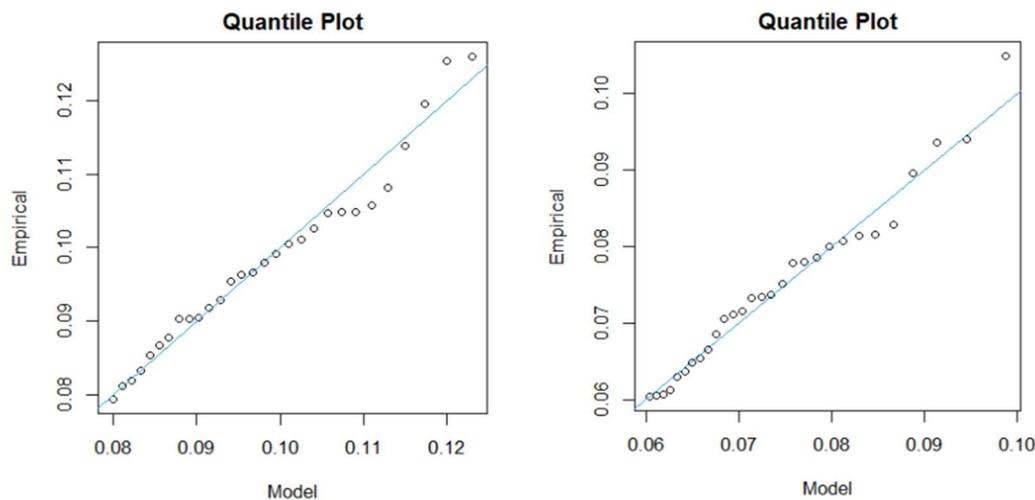


Figure 15 et 16 : Quantile-Quantile plots aux observations logarithmiques dépassant le seuil u (à gauche pour la Belgique et à droite pour le Portugal).

L'application du quantile 95 pourcents au sein de nos deux populations résultait en la création d'un échantillon de 29 données excédents le seuil. Ces 29 différences logarithmiques élevées sont issues des 576 données quotidiennes, à savoir le nombre de jours entre le 4 avril et le 31 octobre 2021. Notre modèle GPD pourra donc être construit selon un nombre de log-daily équivalent à, environ, un mois. Graphiquement (**Annexe Figures A.7 et A.8**), ces données extrêmes se caractérisent par les pics dépassant le seuil en rouge pour la Belgique et en vert pour le Portugal. A titre comparatif, les différences logarithmiques des hospitalisations associées au Covid-19 dépassant le seuil, sont plus éparpillées au Portugal qu'en Belgique. Dans ce dernier pays, la majorité des valeurs extrêmes sont issues des mois de juillet, septembre et octobre 2020. Cela s'explique par le début de deuxième vague qui fut terrible en Belgique, provoquant un taux de croissance conséquent d'hospitalisations.

5.2.1.2 Estimation des paramètres du modèle

Dans cette partie, nous nous intéressons à l'estimation des paramètres du modèle portant sur la loi généralisée de Pareto ou GPD. Ce modèle a été ajusté aux données dites « excédentaires », c'est-à-dire celles surpassant le seuil de 95 pourcents pour les transformations logarithmiques en Belgique et au Portugal. Grâce aux paramètres d'échelle σ et de forme ξ , nous serons en mesure de caractériser les queues de la distribution.

Les paramètres sont obtenus (**Tableau 3**) à l'aide de la méthode du maximum de vraisemblance. Pour chaque paramètre, nous fournissons également un intervalle de confiance de 95 pourcents. Cet intervalle nous donne une mesure de l'incertitude entourant notre estimation des paramètres. Il indique la plage de valeurs plausibles pour chaque paramètre avec un niveau de confiance de 95%.

Paramètres \ Pays	Belgique	Portugal
Seuil 95%	0,079	0,060
Echelle σ	0,032 [0,026 ; 0,038]	0,021 [0,012 ; 0,030]
Forme ξ	-0,640 [-0,730 ; -0,550]	-0,401 [-0,740 ; -0,062]

Tableau 3 : Paramètres estimées de la GPD pour l'analyse univariée

Les paramètres d'échelle (également appelé paramètre de dispersion) pour la Belgique et le Portugal sont relativement proches. Ce paramètre nous renseigne une information quant au degré d'aplatissement de la distribution : quelle est l'amplitude des valeurs extrêmes dans la distribution ? La variabilité est légèrement plus importante pour la Belgique étant donné le paramètre d'échelle plus élevé de 0,009 unité. Nous pouvons comparer ce constat avec les conclusions de la partie 2.1. Dans cette dernière, nous avons vu que la variabilité portant sur l'ensemble des données était plus importante au Portugal. Or, lorsque nous nous limitons aux cinq pourcents les plus importants, cette précédente trouvaille n'est pas d'application au regard des valeurs prises par les paramètres d'échelle estimés.

En ce qui concerne les paramètres de forme pour ces deux pays, ils tendent à prendre des valeurs distinctes (l'intervalle de confiance étant plus restreint en Belgique). Cet estimateur se concentre sur la queue de la distribution des valeurs extrêmes. Un paramètre de forme de -0,640 pour les hospitalisations Covid-19 belges indique que la distribution des excédents présente une queue légère étant donné que cet estimateur est négatif. Le paramètre de forme est moins négatif au Portugal. Bien que ce paramètre reste négatif autour d'un intervalle de confiance de 95% pour les données portugaises, nous pouvons en déduire que la queue de distribution de droite est plus épaisse. La valeur de ξ est environ 0.24 unité plus élevée dans ce pays. Grâce à l'estimation précédente des paramètres, nous pouvons conclure que la GPD associée à nos données est bornée à droite étant donné qu'il s'agit d'une distribution de Weibull ($\xi < 0$).

5.2.1.3 Estimation de probabilités d'occurrences extrêmes et de niveaux de retour

Généralement, il est plus intéressant d'interpréter les modèles de valeurs extrêmes en termes de risque de rencontrer un phénomène aussi extrême dans le futur ou grâce au niveau de retour qui joue le rôle de quantile de la fonction de distribution estimée. Une explication des valeurs prises par les paramètres de la GPD n'est pas suffisante pour effectuer des prévisions tant il s'agit d'une vue statique du comportement historique de la queue de distribution. En calculant différentes probabilités et niveaux de retour pour la Belgique et le Portugal, nous serons en mesure de déterminer quel pays est le plus propice à expérimenter, de nouveau, une croissance considérable de son nombre d'hospitalisations liées au Covid-19.

Tout d'abord, nous pouvons faire une estimation de la probabilité de rencontrer dans le futur, une différence logarithmique quotidienne égale ou supérieure à $x_{\max} = 0,126$. Ce coefficient précis de x , correspond à la valeur maximale recensée pour notre intervalle de temps étudié en Belgique. Cette valeur étant sous la forme d'un logarithmique en base e , nous devons l'interpréter en termes d'augmentation relative du niveau quotidien d'hospitalisations. En résolvant une simple équation logarithmique, nous pouvons remarquer que ce maxima de 0,126 équivalait à une augmentation de 13,43 pourcents car $\text{Ln}(x_{\max}) = 0,126 \Leftrightarrow x_{\max} = e^{0,126} = 1,1343$. Si l'on reparaît la base de données finale, le 21 septembre 2020, il y a bien eu la

plus forte augmentation d'hospitalisations en Belgique. Nous sommes passés de 432, le 20 septembre 2020, à 490 admissions, le lendemain, soit une différence de 13,43 pourcents.

Cette lecture des résultats exprimée en pourcents plutôt qu'en niveau maximal absolu d'hospitalisation, est même une bonne chose. Cela rendra la comparaison possible entre nos deux pays d'analyse et ce même si la Belgique et le Portugal n'ont pas le même nombre d'habitants, ni même une distribution semblable du nombre d'hospitalisations.

En appliquant la formule (3.4) de la section 4, la probabilité quotidienne de constater une augmentation de la même ampleur, en Belgique, est de 0,06 pourcent (5.1).

$$\Pr\{X > 0,126\} = 0,05 \times \left[1 - 0,640 \times \left(\frac{0,126-0,079}{0,032}\right)\right]^{-1/-0,640} = 0,0006 \quad (5.1)$$

Cela peut paraître faible mais il faut se rappeler le caractère extrêmement rare de tels événements. La probabilité associée à une croissance belge de 10 pourcents, c'est-à-dire le quantile logarithmique de 0,0953, est de 2,70 pourcents.

Au Portugal, le maxima atteint par le log-daily était de 0,105. De la même façon à ce qui a été fait pour la Belgique, nous devons traduire cette valeur en termes d'augmentation relative. Cela signifie que l'augmentation maximale quotidienne du nombre d'hospitalisations portugaises était de 11,07 pourcents : $\text{Ln}(x_{\max}) = 0,105 \Leftrightarrow x_{\max} = e^{0,105} = 1,1107$. La probabilité quotidienne d'au moins égal ou de dépasser ce niveau dans le futur, est de 0,04 pourcent (5.2).

$$\Pr\{X > 0,105\} = 0,05 \times \left[1 - 0,401 \times \left(\frac{0,105-0,060}{0,021}\right)\right]^{-1/-0,401} = 0,0004 \quad (5.2)$$

Le risque pour les citoyens portugais de faire face à une croissance de 10 pourcents des admissions Covid-19, est quant à lui de 0,32 pourcent. Nous en déduisons que les Portugais sont moins susceptibles que les Belges d'expérimenter d'importantes augmentations soudaines du nombre d'hospitalisations liées au Coronavirus. En effet, les citoyens du plat pays ont 1,5 fois (0,6/0,4) plus de chance (ou plutôt de malchance), quotidiennement, de connaître une augmentation au moins aussi conséquente que celle observée en 2020. Cela ne veut pas pour autant dire que la Belgique doit s'attendre à souffrir, une fois de plus, de cette pandémie. Il y a tellement d'autres paramètres à tenir compte, comme la capacité hospitalière nationale, le niveau des mesures non-pharmaceutiques, le degré futur de dangerosité du virus, etc.

Dans un second temps, le niveau de retour x_m nous permettra d'apprécier le niveau d'une croissance hospitalière qui est excédé, en moyenne, une fois toutes les m observations. Il est souvent plus pratique d'exprimer ce niveau selon une échelle annuelle. En appliquant la formule (3.7), nous pouvons estimer que le niveau de retour belge correspondant à une période de 10 ans, est de 0,1272 (5.3).

$$z_{10} = 0,079 + \frac{0,032}{-0,640} \times [(100 \times 3650 \times 0,05)^{-0,640} - 1] = 0,1272 \quad (5.3)$$

Ce quantile logarithmique équivaut à une augmentation de 13,56 pourcents ($e^{0,1272} = 1,1356$). Cela signifie, qu'en moyenne, une augmentation quotidienne de 13,56 pourcents du nombre d'hospitalisations Covid-19, sera excédée une fois tous les 10 ans.

Pour le Portugal, ce niveau de retour de 10 ans est de 0,1059 (5.4), soit 11,17 pourcents.

$$z_{10} = 0,060 + \frac{0,021}{-0,401} \times [(100 \times 3650 \times 0,05)^{-0,401} - 1] = 0,1059 \quad (5.4)$$

Comme on pouvait s'y attendre après nos premiers résultats, des niveaux extrêmes de croissance hospitalière auront tendance à être plus fréquents en Belgique. Le niveau maximal

de croissance atteint en Belgique pour la prochaine décennie, serait environ 1,21 fois supérieur à celui subi par le Portugal.

5.2.2 Analyse des hospitalisations en fonction des mesures gouvernementales

Dans cette seconde sous-section, des modèles de la GPD pour la Belgique et le Portugal seront créés en tenant compte du niveau de sévérité des mesures qui était en vigueur. Les observations pour lesquelles une faible réponse gouvernementale était déployée, sont rassemblées en un sous-échantillon au sein duquel les paramètres de la distribution extrême sont estimés comme dans le cas de l'analyse univariée. Il en va de même pour les jours d'hospitalisations qui étaient associés à des mesures gouvernementales élevées. Par conséquent, quatre modèles ont été construits.

5.2.2.1 Détermination du seuil

La recherche menant au choix du quantile ayant le rôle de seuil fut plus complexe pour cette seconde analyse. La base de données étant divisée en deux sous-ensembles de plus petite taille, garder le quantile de 95 pourcents comme limitation semblait impossible. En effet, cela aurait eu pour conséquence d'appliquer une distribution de la TVE à un faible nombre d'observations extrêmes. Il est également important de se rappeler que les 576 journées retenues ont été scindées dans nos deux pays d'intérêt, selon la moyenne du SI. Dès lors, les nouveaux sous-ensembles pouvaient être déséquilibrés. Pour ne pas perdre de l'information et donc garder un échantillon raisonnable de données extrêmes, il a été décidé, premièrement, d'essayer un quantile de 90 pourcents au type de réponse gouvernementale, dit « mesures faibles ». Nous confirmions la pertinence de ces seuils en appliquant un Quantile-Quantile plot aux observations dépassant u (**Annexe Figure A.9**) et en jetant un œil aux graphes montrant les différentes valeurs prises par les paramètres selon une gamme de seuils (**Annexe Figures A.10 et A.11**).

Ce déséquilibre dans la répartition des données engendre logiquement, un autre déséquilibre dans les échantillons retenus pour l'ajustement au modèle GPD. Abaisser le seuil à un quantile de 90 pourcents s'est avéré insuffisant pour la réponse gouvernementale, dite « mesures fortes ». Le manque de données extrêmes y était conséquent : 29 valeurs extrêmes en Belgique et 23 au Portugal. Un second souci était perceptible avec le quantile 90 : les intervalles de confiance du paramètre de forme étaient trop larges et manquaient de précision. Nous avons donc essayé d'abaisser le seuil de dépassement à 85 pourcents pour les observations liées à des mesures fortes. Le nombre d'excédents augmentait alors de 29 à 43 en Belgique et de 23 à 35 au Portugal. La fourchette proposée par l'intervalle de confiance de 95 pourcents pour ξ , diminuait d'une amplitude d'environ 0,9 à 0,5 pour la Belgique et de 0,8 à 0,7 pour le Portugal.

Pour valider ce choix de seuil au quantile 85, nous avons sollicité le graphe mettant en relation l'estimation des paramètres selon le maximum de vraisemblance, en fonction d'une gamme de seuils (**Figure 17 et 18**). Il était bien visible que les estimateurs gagnaient en stabilité et en précision jusque respectivement 0,02 et 0,03, pour la Belgique et le Portugal. Nous pouvions voir que considérer des seuils plus extrêmes, n'avaient pas de sens. Ces seuils approximatifs de 0,02 et 0,03 correspondait quasi-exactement au quantile 85 pourcents. Les QQ-plots nous rassuraient quant à la crédibilité de ces seuils (**Annexe Figure A.12**).

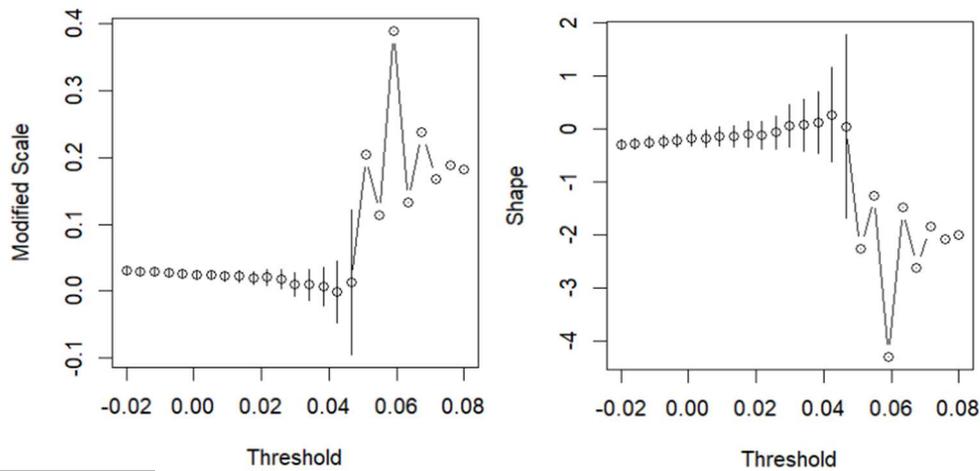


Figure 17 : Estimation du seuil optimal à partir des paramètres d'échelle et de forme en Belgique, selon des mesures fortes adoptées

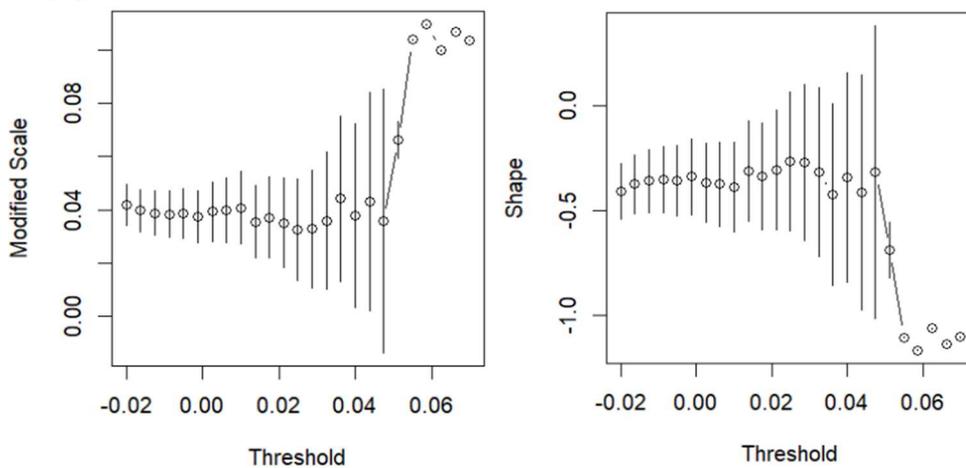


Figure 18 : Estimation du seuil optimal à partir des paramètres d'échelle et de forme au Portugal, selon des mesures fortes adoptées

Lorsque des mesures fortes étaient en place, le seuil équivalait à 0,021 et 0,030 pour respectivement la Belgique et le Portugal. En présence de mesures faibles, ce même seuil était de 0,074 et de 0,51. Au regard des valeurs logarithmiques prises par les seuils, il était évident que l'on relevait moins d'enregistrements extrêmes lorsque la réponse gouvernementale était sévère. Au Portugal, pour obtenir autant d'observations (35 points) dépassant le seuil que dans la situation de mesures faibles, il a fallu diminuer le quantile de 5 pourcents. Le seuil était alors 1,7 fois supérieur pour les mesures faibles, soulignant que les 15 pourcents d'observations incluses, étaient plus extrêmes. En Belgique, le seuil est environ 3,5 fois supérieur, en présence de mesures faibles. Cela mène à penser que les politiques non-pharmaceutiques décidés par ces deux nations ont eu un grand effet sur les pics de contamination : des mesures fortes limitent drastiquement l'apparition d'évènements extrêmes.

5.2.2.2 Estimation des paramètres du modèle

	Paramètres	Belgique	Portugal
Mesures faibles	Seuil 90%	0,074	0,051
	Echelle σ	0,040 [0,034 ; 0,046]	0,021 [0,012 ; 0,030]
	Forme ξ	-0,747 [-0,811 ; -0,683]	-0,283 [-0,604 ; 0,038]
Mesures fortes	Seuil 85%	0,021	0,030
	Echelle σ	0,018 [0,011 ; 0,025]	0,025 [0,013 ; 0,037]
	Forme ξ	-0,097 [-0,359 ; 0,165]	-0,267 [-0,656 ; 0,122]

Tableau 4 : Paramètres estimés de la TVE pour l'analyse des hospitalisations quotidiennes en fonction des mesures gouvernementales

En présence d'une politique sanitaire permissive (mesures faibles) en Belgique et au Portugal, les valeurs prises par les paramètres de la GPD ont tendance à se rapprocher aux paramètres estimés lors de l'analyse 5.2.1. Excepté pour le paramètre de forme qui diffère de quelques dixièmes mais qui reste pratiquement toujours négatif pour l'intervalle de confiance de 95 pourcents, les deux autres paramètres prennent des valeurs semblables. Ces points de similitude peuvent s'expliquer par le fait que les données extrêmes dépassant le seuil de 95 pourcents de l'analyse univariée, constitue majoritairement le sous-ensemble de données extrêmes en présence de mesures faibles. En fait, les 29 observations excédentaires de l'analyse 5.2.1 ont, pour la plupart, été enregistrées lors d'une réponse gouvernementale faible.

Dès lors, il est intéressant de se focaliser sur la partie inférieure du **Tableau 4**, où le modèle GDP a surtout été construit sur des données mises de côté jusqu'à présent. La principale différence dans ce nouveau modèle basé sur les log-différences en présence de mesures fortes, réside dans le paramètre de forme belge. En effet, celui-ci est pour la première fois proche de 0. Cela signifie que la queue de distribution à droite est légèrement plus lourde. Pour cet échantillon, les données extrêmes seraient plutôt fréquentes par rapport à une GDP standard. Au Portugal, nous ne remarquons pas de changements considérables en jetant un œil à l'estimation des paramètres. Bien qu'il ait été réduit par l'introduction du quantile 85 pourcents, il est nécessaire d'aborder l'intervalle de confiance du paramètre de forme avec précaution.

5.2.2.3 Estimation de probabilités d'occurrences extrêmes et de niveaux de retour

Comme mentionné antérieurement, l'analyse extrême des hospitalisations Covid-19 menée à partir de la TVE, reste très simplifiée tant une multitude d'autres variables ont été négligées. Dans cette dernière section, la force de la réponse gouvernementale nationale consistera en un nouvel aspect joint aux hospitalisations historiques enregistrées. Les nouvelles estimations de la probabilité d'être confronté au Coronavirus et des niveaux de retour remarquables seront proposés, en tenant compte de l'influence politique dans la prolifération d'un tel virus. Après avoir ajusté et retardé les données d'hospitalisations selon l'intensité de la réponse gouvernementale décidée dix jours plus tôt, nous étions capables statistiquement parlant, d'extrapoler l'effet ou non, des mesures politiques sur les risques d'augmentation hospitalière. Bien qu'une distinction entre la Belgique et le Portugal sera toujours établie, l'accent sera surtout mis sur les résultats de la TVE en présence de mesures faibles ou fortes.

Estimons, en premier, les probabilités de faire face à nouveau dans le futur, à la différence logarithmique de la même ampleur que celles enregistrées comme étant maximales en période de mesures faibles et fortes. En Belgique, les log-return les plus importants étaient de 0,126 et de 0,091, respectivement en présence de mesures faibles et fortes. Sans grande surprise, le maxima de 0,126 de la sous-section 5.2.1.3 a été atteint lorsque le gouvernement belge faisait preuve de plus de souplesse vis-à-vis de sa population. Lorsque les hospitalisations belges liées au Covid-19 correspondaient à des journées où les mesures adoptées étaient faibles, une augmentation maximale quotidienne de 13,43 pourcents était calculée. Nous utiliserons également ce x_{\max} de 0,126 pour le calcul de la probabilité extrême avec des mesures fortes. Cela permettra de comparer, rigoureusement l'effet de la réponse gouvernementale pour un niveau d'augmentation hospitalière similaire au sein des deux populations.

En appliquant les nouveaux paramètres de la section 5.2.2.2 à la formule (3.4), les probabilités quotidiennes de connaître dans le futur une augmentations hospitalière Covid-19 de 13,43 points sont de 0,090 pourcent (5.5) et 0,003 pourcent (5.6), pour respectivement un régime faible de mesure et un régime fort.

$$\Pr\{X > 0,126\} = 0,10 \times \left[1 - 0,747 \times \left(\frac{0,126-0,074}{0,040}\right)\right]^{-1/-0,747} = 0,0009 \quad (5.5)$$

$$\Pr\{X > 0,126\} = 0,15 \times \left[1 - 0,097 \times \left(\frac{0,126-0,021}{0,018}\right)\right]^{-1/-0,097} = 0,00003 \quad (5.6)$$

Bien que ces risques restent extrêmement faibles au vu de leurs valeurs proches de 0, nous pouvons constater que la probabilité est 30 fois plus élevé lorsque le gouvernement belge met en application des mesures dont la force est inférieure à la moyenne de l'indice de sévérité SI.

Pour le Portugal, la probabilité de faire face une fois de plus à une augmentation de 11,07 points ($e^{0,105} = 1,1107$), est de 0,1 pourcent en présence de mesures faibles (5.7) et de 0,04 pourcent en présence de mesures fortes (5.8).

$$\Pr\{X > 0,105\} = 0,10 \times \left[1 - 0,283 \times \left(\frac{0,105-0,051}{0,021}\right)\right]^{-1/-0,283} = 0,001 \quad (5.7)$$

$$\Pr\{X > 0,105\} = 0,15 \times \left[1 - 0,267 \times \left(\frac{0,105-0,030}{0,025}\right)\right]^{-1/-0,267} = 0,0004 \quad (5.8)$$

Le risque d'atteindre des degrés extrêmes d'augmentation des hospitalisations serait donc 2,5 fois supérieur lorsque le gouvernement portugais octroie plus de libertés à sa population. Ce constat encore plus implacable en Belgique, montre à quel point l'intervention gouvernementale est cruciale pour éviter que le niveau de croissance pandémique croisse exponentiellement.

En utilisant la formule (3.7) et en remplaçant les paramètres du modèle GDP par les valeurs adéquates, le dernier indicateur de niveau de retour pour une période de 10 ans, a été estimé pour les 4 situations : Belgique/mesures faibles, Belgique/mesures fortes, Portugal/mesures faibles et Portugal/mesures fortes. En Belgique, le quantile estimé pour ce niveau de retour de 10 ans était de 0,1269 pour des mesures faibles (5.9) et de 0,1059 pour des mesures fortes (5.10) selon l'indice SI.

$$z_{10} = 0,074 + \frac{0,040}{-0,747} \times [(100 \times 3650 \times 0,10)^{-0,747} - 1] = 0,1269 \quad (5.9)$$

$$z_{10} = 0,021 + \frac{0,018}{-0,097} \times [(100 \times 3650 \times 0,15)^{-0,097} - 1] = 0,1059 \quad (5.10)$$

Les résultats suggèrent une augmentation quotidienne de 11,17 pourcents ($e^{0,1059} = 1,1117$), une fois tous les 10 ans pour des mesures fortes, alors qu'elle serait de 13,53 pourcents ($e^{0,1269} = 1,1353$) en présence de mesures faibles. Cela va de sens avec notre première trouvaille, soulignant une plus grande probabilité de retour lorsque les mesures belges étaient faibles. Ici,

une augmentation extrême, excédée une fois tous les 10 ans, aurait 2,36 points de plus lorsque les mesures gouvernementales sont faibles.

Au Portugal, pour lequel les paramètres de la GDP avaient le moins varié entre les deux types de mesures, les niveaux de retour de 10 ans confirmaient, également, les résultats précédemment renseignés. L'augmentation quotidienne de 11,76 pourcents en présence de mesures faibles (5.11) et de 11,20 pourcents en présence de mesures fortes (5.12), étaient surpassées, en moyenne, une fois tous les 10 ans.

$$z_{10} = 0,051 + \frac{0,021}{-0,283} \times [(100 \times 3650 \times 0.10)^{-0,283} - 1] = 0,1112 \quad (5.11)$$

$$z_{10} = 0,030 + \frac{0,025}{-0,267} \times [(100 \times 3650 \times 0.15)^{-0,267} - 1] = 0,1062 \quad (5.12)$$

Le caractère plus extrême rencontré pour les mesures faibles se confirmait également pour le niveau de retour de 10 ans. En effet, la croissance du nombre d'hospitalisations serait 0,56 point plus élevé si le gouvernement adopte des mesures faibles.

Que ce soit pour le calcul d'une probabilité quotidienne extrême ou du niveau de retour, nous remarquons que les divergences sont moins importantes pour le Portugal. Nous pouvons déjà nous en douter au regard des valeurs prises par les paramètres estimés, qui tendent à être semblables. Nous pouvons penser que les autorités belges ont été particulièrement efficace pour ralentir la prolifération des hospitalisations. L'analyse descriptive a mis en lumière le fait que l'indice SI était pratiquement toujours plus élevé et moins variable au Portugal lors des 18 mois analysés. Cette rigueur constante en termes d'interventions non-pharmaceutiques, explique très certainement l'impact plus faible des mesures fortes, tant les Portugais s'étaient accoutumés à ce type de mesures. D'autres facteurs qualitatifs, comme un meilleur calibrage des mesures par rapport à la réalité du terrain en milieu hospitalier ou encore une plus grande discipline et mise en pratique des mesures de la population belge, sont des raisons potentielles à englober.

6 Conclusion

6.1 Synthèse des découvertes et des résultats

Tout d'abord, l'analyse descriptive des hospitalisations Covid-19 a mis en évidence une charge hospitalière moyenne plus élevée en Belgique, à savoir 1777 personnes par rapport à 1232 au Portugal. La **Figure 2** et l'**Annexe Figure A.3** nous garantissaient une forme de distribution asymétrique à droite. Cela était une bonne nouvelle : des valeurs extrêmes étaient bien présentes en plus grand nombre qu'une distribution normale. Etant donné l'impossibilité d'utiliser les données hospitalières brutes pour la TVE, nous nous penchions sur les transformations logarithmiques en base e (**Figure 3**). Concernant l'indice de sévérité SI issu du Ox-CGRT, la trouvaille majeure porte sur la moyenne du score des pays entre nos deux pays d'intérêt. Pendant un an et demi, le score moyen pris par le SI au Portugal, était supérieur de 10 points à celui de la Belgique.

La revue de la littérature existante a confirmé qu'il était pleinement pertinent de mener une analyse basée sur la TVE aux logarithmes des hospitalisations. La modélisation des événements de super-propagation s'apparentait le plus à ce que nous souhaitions dans la partie 5. Une distribution GPD discrète était ajustée aux foyers hyper-contaminateurs et le paramètre de forme était notamment estimé. Il était également intéressant de comprendre comment la base de données « Ox-CGRT » avait été construite, son interprétabilité et les différentes variantes d'index qu'elle contenait. Enfin, plusieurs études montraient un lien positif significatif entre l'indice SI, peu importe la localisation mondiale et la vague de contaminations en question.

A l'image de ce qu'il avait pu être réalisé dans la littérature, les hospitalisations étaient transformées sous la forme logarithmique et un décalage temporel de dix jours était appliqué au SI belge et portugais. Les résultats étaient obtenus en suivant, méthodiquement, les étapes décrites dans la partie 3 du mémoire. Le choix du seuil correspondant à un quantile de 95 pourcents était validé par les **Figures 13 et 14** pour l'analyse univariée. Les paramètres estimés à partir de la GPD ajusté à nos cinq pourcents de données extrêmes étaient relativement proches. La mise en pratique de ces estimateurs grâce aux formules (3.4 et 3.7) faisait état d'un risque d'augmentation future des hospitalisations un peu plus important pour la Belgique. Dans le cadre de l'analyse multivariée, les seuils choisis étaient associés respectivement à des quantiles de 90% et 85% pour les mesures faibles et fortes. Concernant les probabilités extrêmes quotidiennes d'être confrontés une nouvelle fois au pic maximal d'augmentation hospitalière, cette valeur était nettement plus élevée lorsque la réponse gouvernementale était faible, quel que soit le pays. Aux dépens des libertés individuelles, cela prouvait que les gouvernements ont un rôle considérable à jouer pour ralentir la progression d'une épidémie tel que celle du Coronavirus. Nous pouvons résumer merveilleusement bien cette dernière trouvaille avec la citation du philosophe latin Sénèque (de 4 ACN à 65 PCN) « On doit punir, non pour punir, mais pour prévenir »¹⁰.

6.2 Réponses aux questions de recherche

Après s'être rappelé tous les éléments clés de ce mémoire, c'est le bon moment pour répondre à notre question de recherche centrale qui est la suivante : « A l'aide de la Théorie des Valeurs Extrêmes, quel est le risque que la pandémie de Covid-19 resurgisse de manière fulgurante, dans les dix prochaines années, en Belgique et au Portugal ». Nous pouvons très facilement répondre, de manière quantifiée, à cette question à l'aide des probabilités calculées dans le

¹⁰ <http://evene.lefigaro.fr/citation/doi-punir-non-punir-prevenir-6532.php>

point 5.2.1.3. En Belgique, la probabilité quotidienne de connaître une croissance des augmentations hospitalières d'au moins 13,43 pourcents (le pic de la période étudiée), est de 0,06 pourcent. Pour le Portugal, la probabilité quotidienne de revoir une augmentation de 11,07 pourcents, est de 0,04 pourcent. Le risque pour les citoyens belges est 1,5 fois supérieur au risque portugais. A la lumière de ces probabilités, le risque quotidien d'être confronté, une nouvelle fois, à une augmentation conséquente des hospitalisations est très faible. Il est nécessaire d'évaluer ce risque à plus long terme à l'aide du niveau de retour. La métrique de niveau de retour de dix ans est pertinente pour connaître le niveau d'augmentation qui pourrait être excédé une fois dans les dix prochaines années. Pour la Belgique, une augmentation des hospitalisations d'au moins 13,56 pourcents est attendue, en moyenne, une fois pour cette période. Pour le Portugal, il s'agit d'une augmentation de 11,17 pourcents, correspondant au niveau de retour de dix ans.

Dès lors, nous sommes en mesure de répondre à la question suivante : « Peut-on affirmer qu'un de nos deux pays d'intérêt est plus susceptible de connaître une crise sanitaire aussi extrême dans le futur ? ». La Belgique semblerait plus susceptible d'être confrontée, à nouveau, à des niveaux d'augmentations hospitalières spectaculaires. Les différences entre les métriques calculées sont faibles pour les deux pays mais lorsque l'on parle de plusieurs milliers de personnes hospitalisées, une augmentation supplémentaire de deux pourcents peut vite représenter des centaines de personnes hospitalisées en plus. Nous pouvons tenter d'expliquer cette nouvelle positive pour le Portugal par sa population qui est inférieur de dix pourcents et par le niveau de sévérité des mesures qui était, en moyenne, dix points plus élevés. Ce risque portugais moins extrême à l'air logique étant donné que la modélisation s'est basée sur des observations historiques dont la réponse gouvernementale était sévère.

Concernant les questions relatives au risque d'augmentation hospitalière en fonction du type de mesures, fortes ou faibles, il est évident que les probabilités de revoir des pics aussi élevés qu'en 2020 et 2021, sont nettement supérieures si les décideurs politiques optent pour des mesures faibles. Oui, nous pouvons affirmer que le degré de sévérité de la réponse gouvernementale influence la croissance du nombre d'hospitalisations associées au Covid-19. Respectivement en Belgique et au Portugal, le risque de faire face, une nouvelle fois, au pic atteint, est 30 fois et 2,5 fois supérieur si des mesures faibles sont en vigueur. En Belgique, une augmentation extrême, excédé une fois tous les 10 ans, aurait 2,36 points de plus lorsque les mesures gouvernementales sont faibles. Lorsque l'on compare les deux pays, nous remarquons que les divergences entre les types de mesure sont moins importantes au Portugal. Nous pouvons en conclure que bien que la Belgique ait une plus grande probabilité de connaître une croissance des hospitalisations, ses mesures politiques ont mieux fonctionné, en comparaison avec le gouvernement portugais. Nous pouvons tenter d'expliquer cela par une meilleure adéquation des mesures avec la situation épidémiologique et une plus grande discipline de la population.

6.3 Limitations de l'étude

Bien qu'une dimension supplémentaire ait été ajoutée dans la section 5.2.2, ce travail reste perfectible dans bon nombre d'aspects. Premièrement, nous avons choisi d'appliquer des seuils correspondant aux quantiles 85, 90 et 95 pourcents. A partir de cette méthode graphique, nous avons retenu des quantiles étant des multiples de cinq pour faciliter l'analyse. Il est certain qu'il existe des méthodes non-mentionnées permettant de choisir des seuils encore plus précis que les approximations que nous avons utilisées.

La Belgique et le Portugal sont deux pays d'Europe similaires du point de vue du nombre d'habitants. En revanche, les climats, les cultures, les mentalités, les propensions à obéir aux ordres politiques, l'application des gestes barrières sont particulièrement différentes. Par cela, nous voulons dire qu'il existe d'autres variables que nous n'avons pas inclus dans le modèle alors qu'elles nous auraient permis de réaliser des extrapolations plus précises. Ces facteurs cités, difficiles à quantifier, peuvent expliquer une partie des fluctuations du niveau d'hospitalisations à un temps et dans un pays donné.

Malgré ces différences intrinsèques entre la Belgique et le Portugal, nous avons constamment comparé ces deux pays centraux dans notre analyse. Il aurait été plus judicieux de comparer la Belgique avec les Pays-Bas et le Portugal avec l'Espagne étant donné que ce sont des pays plus proches culturellement ainsi qu'au niveau des conditions météorologiques [Roumagnac et al., 2021]. Malheureusement, cela n'était possible puisque le Centre Européen de Prévention et de Contrôle des Maladies n'offraient pas des données quotidiennes pour ces pays. Dès lors, employer des données hebdomadaires auraient rendu l'analyse extrême très peu précise au vu du manque de données.

Concernant la division des données en deux sous-populations pour la section 5.2.2, nous nous sommes limités à créer deux classes, à savoir les jours d'hospitalisations où des mesures fortes avaient été adoptées dix jours plus tôt et les journées liées à des mesures faibles. Cette division peut paraître simplifiée étant donné que des journées avec un score de 10 et 50 à l'index SI, seront toutes rassemblées dans le même échantillon visant à prédire les niveaux de retour en présence de mesures gouvernementales faibles. Pour obtenir des résultats encore plus remarquables, il aurait été plus judicieux de diviser une nouvelle fois ces deux échantillons afin d'obtenir quatre classes de mesures et, par conséquent, plus d'homogénéité dans les ensembles de données. Une autre possibilité pour remédier à cette classification en deux groupes est une analyse continue. Dans ce cas-là, les paramètres d'échelle et de forme seraient modélisés comme des fonctions linéaires de l'indice de sévérité SI.

7 Bibliographie

Alemanno, A. (2020). The European response to COVID-19 : from regulatory emulation to regulatory coordination ? *European journal of risk regulation*, 11(2), 307-316. <https://doi.org/10.1017/err.2020.44>

Cevik, M., Tate, M., Lloyd, O., Maraolo, A. E., Schafers, J., & Ho, A. (2020). SARS-COV-2, SARS-COV-1 and MERS-COV Viral Load Dynamics, Duration of viral shedding and Infectiousness : A Living Systematic Review and Meta-Analysis. *Social Science Research Network*. <https://doi.org/10.2139/ssrn.3677918>

Charras-Garrido, Myriam; Lezaud, Pascal. Extreme Value Analysis: an Introduction. *Journal de la société française de statistique*, Tome 154 no. 2, pp. 66-97, 2013. http://www.numdam.org/item/JSFS_2013_154_2_66_0

Chisadza, C., Clance, M., & Gupta, R. (2021). Government effectiveness and the COVID-19 pandemic. *Sustainability*, 13(6), 3042. <https://doi.org/10.3390/su13063042>

Chiu, Y., Chebana, F., Abdous, B., Bélanger, D., & Gosselin, P. (2016). Mortality and Morbidity Peaks Modeling : An Extreme Value Theory approach. *Statistical Methods in Medical Research*, 27(5), 1498-1512. <https://doi.org/10.1177/0962280216662494>

Coles, Stuart. An Introduction to Statistical Modeling of Extreme Values. *Springer London*, 2001. DOI.org (Crossref), <https://doi.org/10.1007/978-1-4471-3675-0>

Daouia, A., Stupfler, G., & Usseglio-Carleve, A. (2023). Extreme value modelling of SARS-COV-2 community transmission using discrete generalized Pareto distributions. *Royal Society Open Science*, 10(3). <https://doi.org/10.1098/rsos.220977>

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the royal statistical society series b-methodological*, 52(3), 393-425. <https://doi.org/10.1111/j.2517-6161.1990.tb01796.x>

Donovan, S., Brett-Major, D. M., & Lawler, J. V. (2021). Lag between COVID-19 cases and subsequent deaths. *Research Square (Research Square)*. <https://doi.org/10.21203/rs.3.rs-434947/v2>

Fisher, M., & Marshall, A. P. (2009). Understanding descriptive statistics. *Australian Critical Care*, 22(2), 93-97. <https://doi.org/10.1016/j.aucc.2008.11.003>

Ghasemi, A., Boroumand, Y., & Shirazi, M. (2020). How do governments perform in facing COVID-19 ? *MPRA Paper*. <https://ideas.repec.org/p/pramprapa/99791.html>

Hale, T., Hale, A. J., Kira, B., Petherick, A., Phillips, T., Sridhar, D., Thompson, R. N., Webster, S., & Angrist, N. (2020). Global assessment of the relationship between government response measures and COVID-19 deaths. *medRxiv (Cold Spring Harbor Laboratory)*. <https://doi.org/10.1101/2020.07.04.20145334>

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5(4), 529-538. <https://doi.org/10.1038/s41562-021-01079-8>

Hale, T., Angrist, N., Hale, A. J., Kira, B., Majumdar, S., Petherick, A., Phillips, T., Sridhar, D., Thompson, R. N., Webster, S., & Zhang, Y. (2021). Government responses and COVID-19 deaths : Global evidence across multiple pandemic waves. *PLOS ONE*, 16(7), e0253116. <https://doi.org/10.1371/journal.pone.0253116>

Hale, T., Cameron-Blake, E., do Foclo, M., Furst, R., Green, K., Phillips, T., Sudarmawan, A., Tat-low, H., & Zha, H. (2022, March). What have we learned from tracking every government policy on COVID-19 for the past two years? *University of Oxford*. <https://www.bsg.ox.ac.uk/sites/default/files/2022-03/What-have-we-learned-from%20tracking-two-years-of-COVID-responses-BSG-research-note-March-2022.pdf>

Herby, Jonas and Jonung, Lars and Hanke, Steve (2022). A Literature Review and Meta-Analysis of the Effects of Lockdowns on COVID-19 Mortality – II. *Munich Personal REPEC Archive*. (s. d.). <https://mpra.ub.uni-muenchen.de/113732/>

Li Y, Campbell H, Kulkarni D, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect Dis*. 2021; 21(2):193–202. [https://doi.org/10.1016/S1473-3099\(20\)30785-4](https://doi.org/10.1016/S1473-3099(20)30785-4) PMID: 33729915

London School of Economics and Political Science. (2020). How effectively have governments responded to COVID-19 so far ? - *LSE Research online*. <https://eprints.lse.ac.uk/105490/>

Papastathopoulos, Ioannis, et Jonathan A. Tawn. Extended Generalised Pareto Models for Tail Estimation. *Journal of Statistical Planning and Inference*, vol. 143, n° 1, janvier 2013, p. 131-43. DOI.org (Crossref), <https://doi.org/10.1016/j.jspi.2012.07.001>

Pierron, D., Pereda-Loth, V., Mantel, M., Moranges, M., Bignon, E., Alva, O., Kabous, J., Heiske, M., Pacalon, J., David, R., Dinnella, C., Spinelli, S., Monteleone, E., Farruggia, M. C., Cooper, K. W., Sell, E., Thomas-Danguin, T., Bakke, A. J., Parma, V., . . . Bensafi, M. (2020). Smell and taste changes are early indicators of the COVID-19 pandemic and political decision effectiveness. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-18963-y>

Piquero, A. R., & Kurland, J. (2022). More stringent measures against COVID-19 are associated with lower cases and deaths in Florida and Miami-Dade. *American Journal of Emergency Medicine*, 53, 262-263. <https://doi.org/10.1016/j.ajem.2021.04.066>

Raggad, Bechir. Fondements de la théorie des valeurs extrêmes, ses principales applications et son apport à la gestion des risques du marché pétrolier. *Mathématiques et sciences humaines*, no. 186, été 2009, DOI : <https://doi.org/10.4000/msh.11069>.

Roumagnac, A., De Carvalho Filho, E., Bertrand, R., Banchereau, A., & Lahache, G. (2021). Étude de l'influence potentielle de l'humidité et de la température dans la propagation de la

pandémie COVID-19. *Médecine de Catastrophe - urgences collectives*, 5(1), 87-102.
<https://doi.org/10.1016/j.pxur.2021.01.002>

Thomas, M., Lemaitre, M., Wilson, M. L., Viboud, C., Yordanov, Y., Wackernagel, H., & Carrat, F. (2016). Applications of extreme value theory in Public Health. *PLOS ONE*, 11(7), e0159312. <https://doi.org/10.1371/journal.pone.0159312>

Wong, F., & Collins, J. J. (2020). Evidence that coronavirus superspreading is fat-tailed. *Proceedings of the National Academy of Sciences of the United States of America*, 117(47), 29416-29418. <https://doi.org/10.1073/pnas.2018490117>

8 Annexes

Figure A.1 : Boite à moustache pour les hospitalisations quotidiennes en Belgique
Boite à moustache des hospitalisations Covid-19 en Belgique

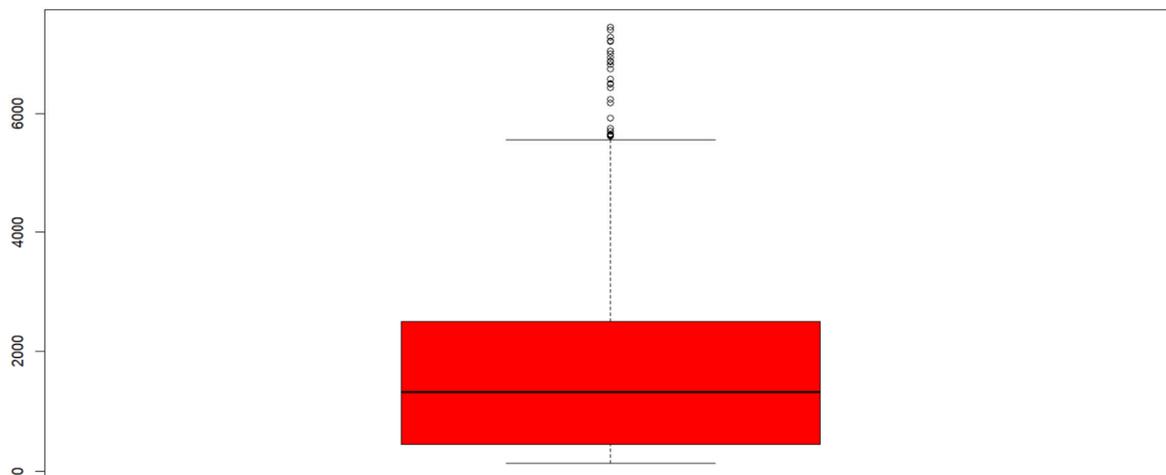


Figure A.2 : Boite à moustache pour les hospitalisations quotidiennes au Portugal
Boite à moustache des hospitalisations Covid-19 au Portugal

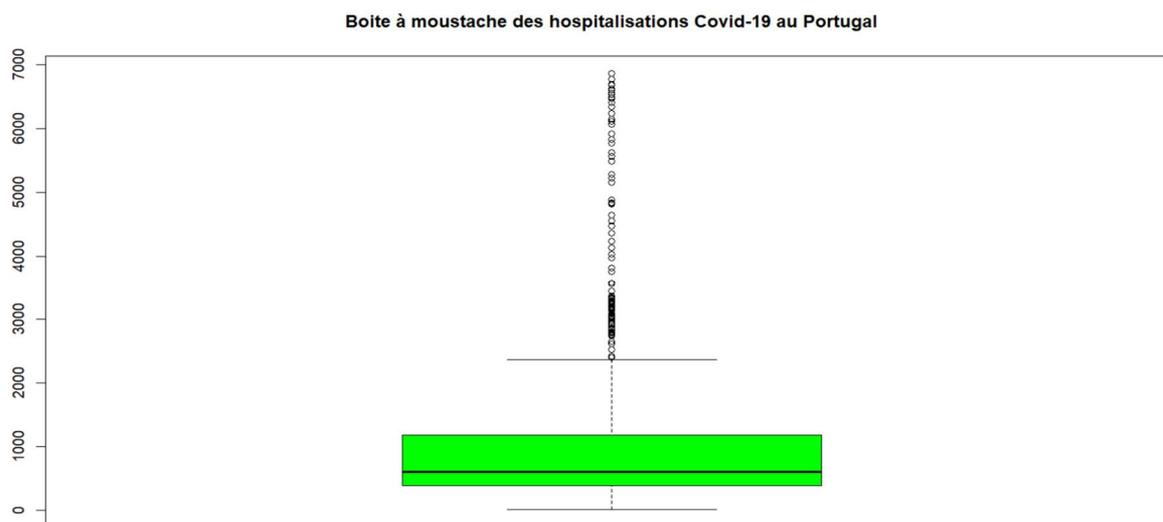


Figure A.3 : Histogramme du nombre d'hospitalisations portugaises

Histogramme des hospitalisations Covid-19 de mars 2020 à octobre 2021 au Portugal

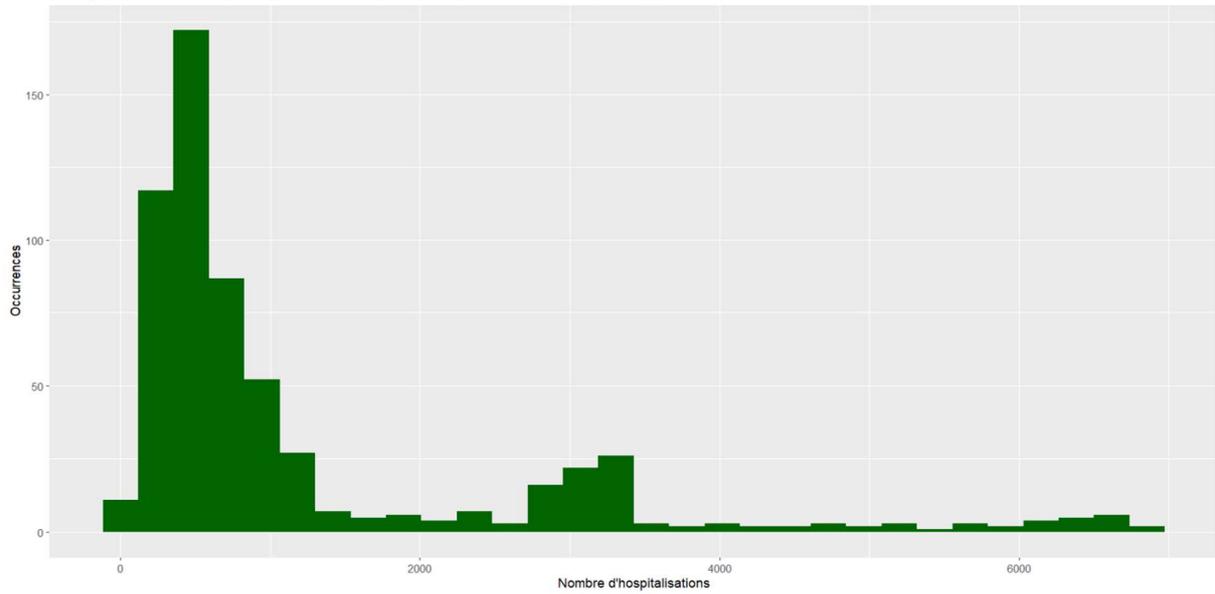


Figure A.4 : Histogramme des différences logarithmiques du nombre d'hospitalisations Covid-19 en Belgique

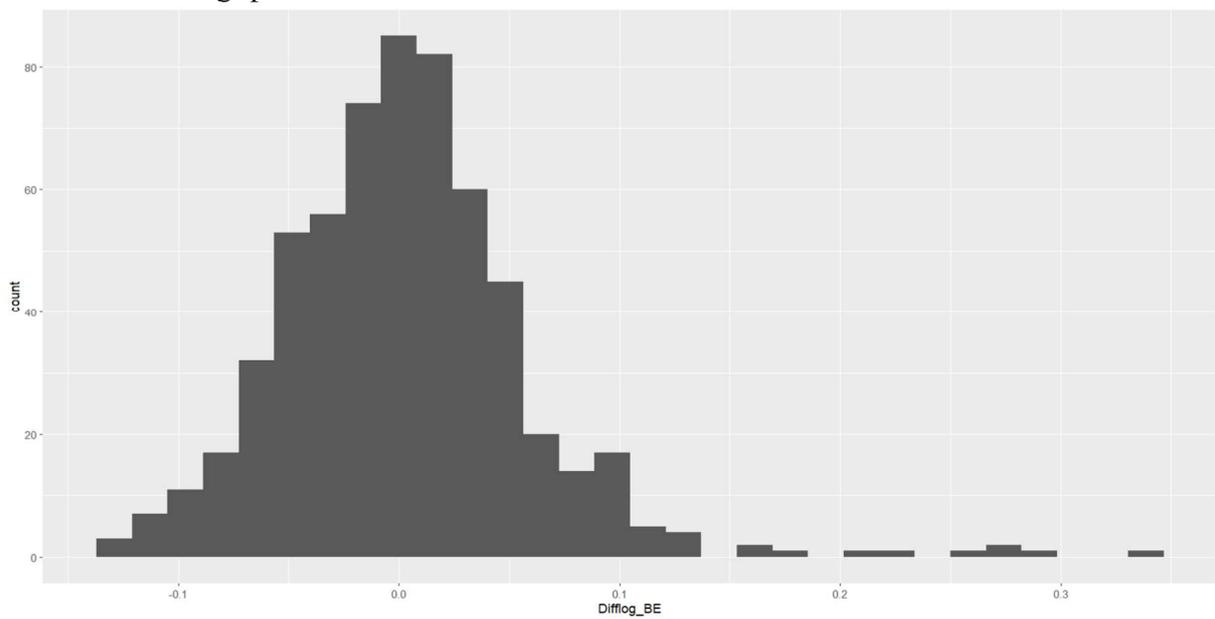
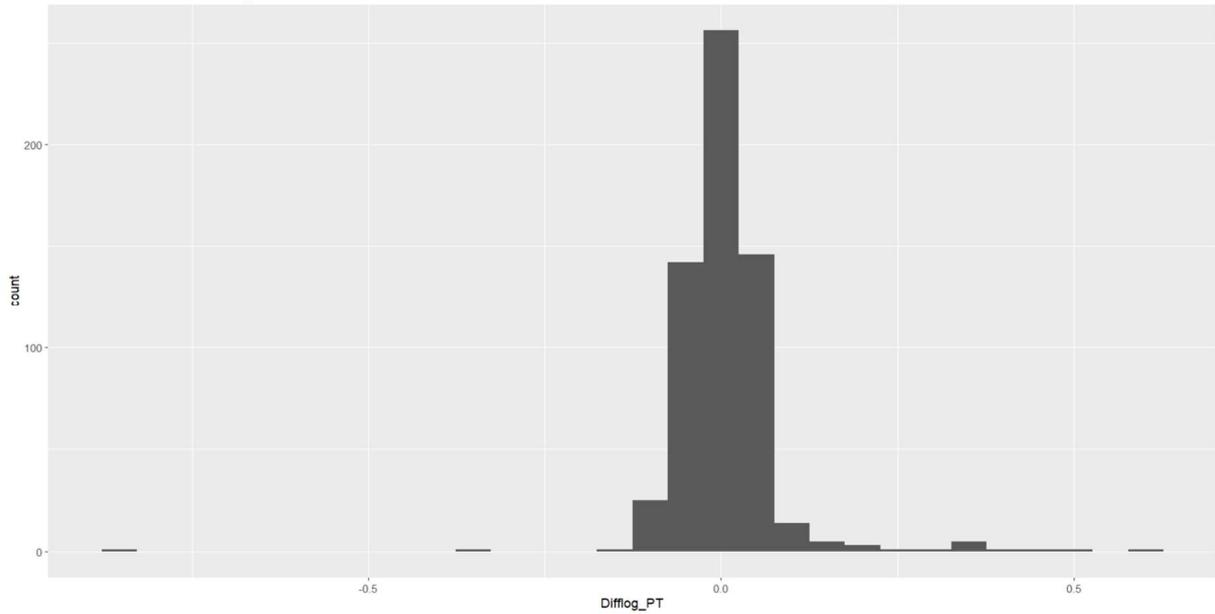


Figure A.5 : Histogramme des différences logarithmiques du nombre d'hospitalisations Covid-19 au Portugal



Théorème A.6 :

Théorème 3.4.2 : S'il existe une séquence de constantes $\{a_n > 0\}$ et $\{b_n\}$ tel que

$$\Pr\{(M_n - b_n)/a_n \leq z\} \rightarrow G(z) \text{ as } n \rightarrow \infty$$

Pour une fonction de distribution non-dégénérée G , alors G est un membre de la famille GEV

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}$$

Défini sur $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, où $-\infty < \mu < \infty$, $\sigma > 0$ et $-\infty < \xi < \infty$. □

L'interprétation de la limite du théorème 3.4.2 comme une approximation pour les grandes valeurs de n suggère l'utilisation de la famille GEV pour modéliser la distribution.

Figure A.7 : Série temporelle des transformations logarithmiques entre le 4 avril et le 31 octobre pour la Belgique

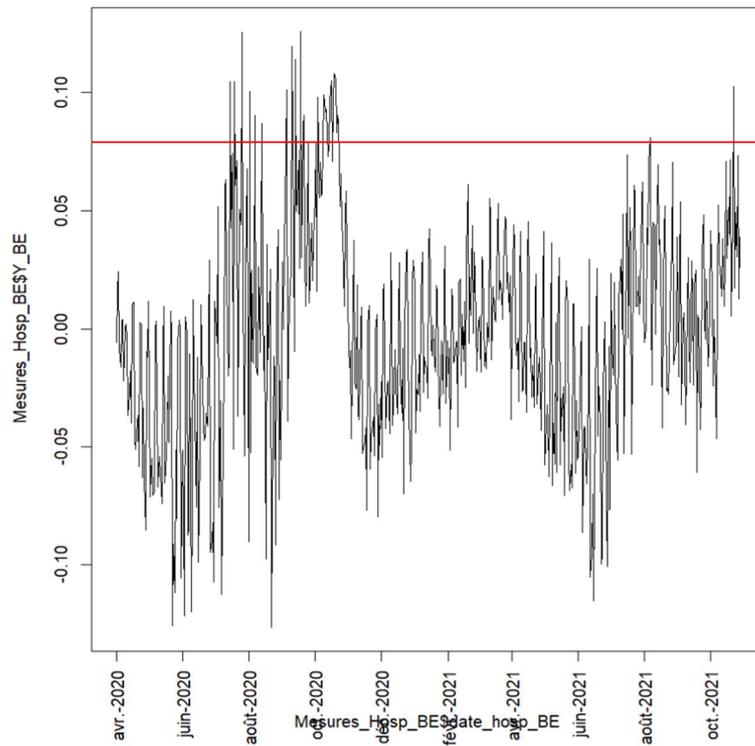


Figure A.8 : Série temporelle des transformations logarithmiques entre le 4 avril et le 31 octobre pour le Portugal

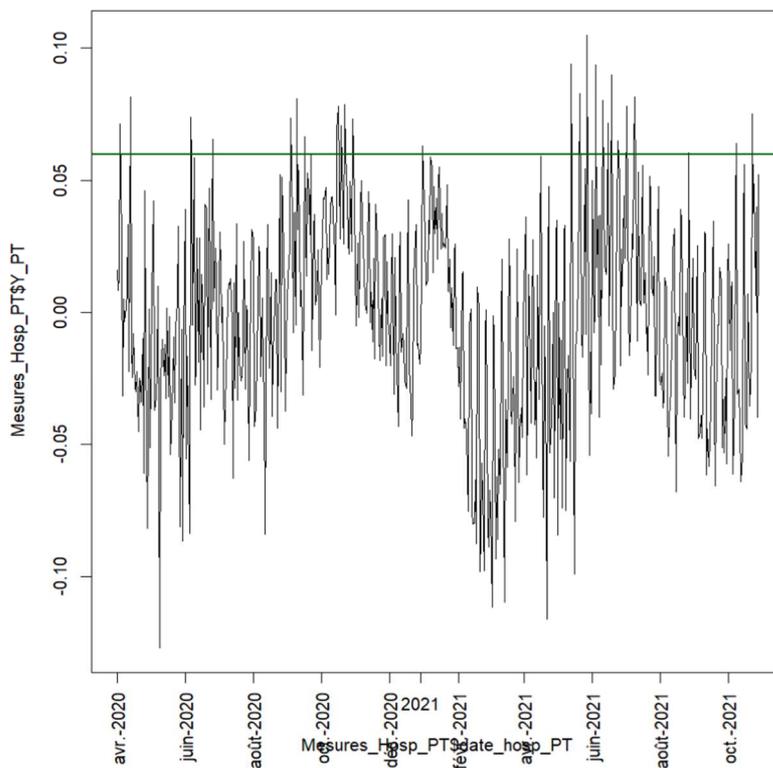


Figure A.9 : Quantile-Quantile plots des observations logarithmiques dépassant le seuil u , en présence de mesures faibles (à gauche pour la Belgique et à droite pour le Portugal).

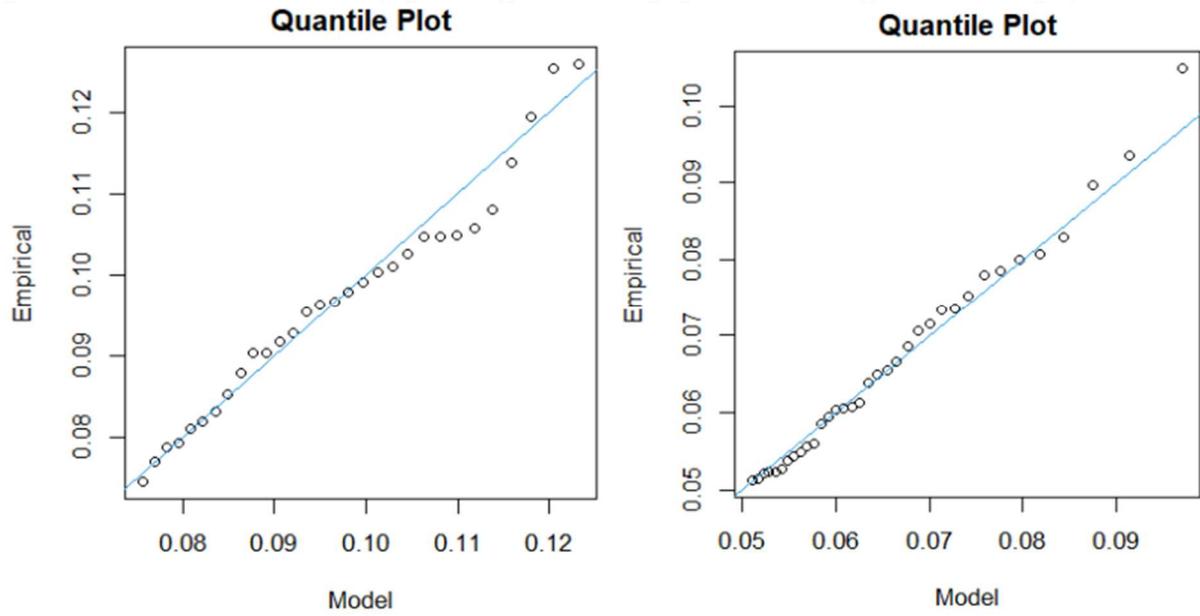


Figure A.10 : Estimation du seuil optimal à partir des paramètres d'échelle et de forme en Belgique, selon des mesures faibles adoptées

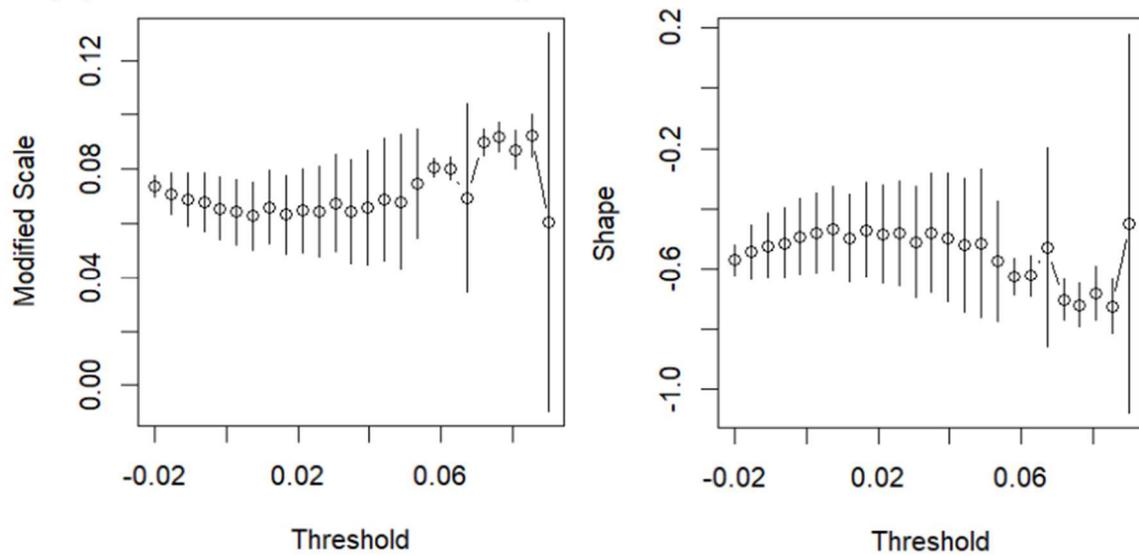


Figure A.11 : Estimation du seuil optimal à partir des paramètres d'échelle et de forme au Portugal, selon des mesures faibles adoptées

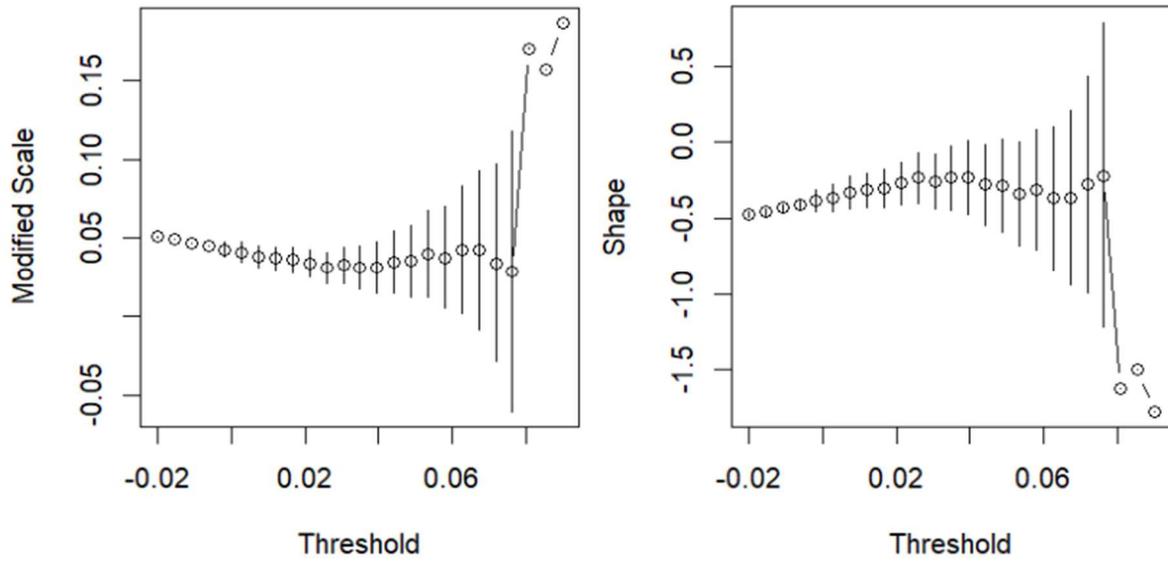


Figure A.12 : Quantile-Quantile plots des observations logarithmiques dépassant le seuil u , en présence de mesures fortes (à gauche pour la Belgique et à droite pour le Portugal).

