

**Mémoire présenté devant le jury de l'ISUP pour l'obtention du diplôme de
Statisticien Mention Actuariat et l'admission à l'Institut des Actuaires**

Le .../.../2012

Par : Hervé FRAYSSE

**Titre : Construction d'un générateur de scénarios économiques à sauts permettant la prise en compte de
scénarios de crises**

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité ci-dessus

*Membres présents du jury de l'Institut signature
des Actuaires*

.....

Membres présents du jury de la filière

Entreprise :

Nom :

Signature :

Directeur de mémoire en entreprise :

Nom :

Signature :

Invité :

Nom :

Signature :

***Autorisation de publication et de mise
en ligne sur un site de diffusion de
documents actuariels (après
expiration de l'éventuel délai de
confidentialité)***

Signature du responsable entreprise

Signature du candidat

Secrétariat

Bibliothèque :

Note de synthèse

Mots clés : générateur de scénarios économiques (GSE), risques croissants, crise économique, inflation, modélisation ARIMA, algorithme de Box et Tiao, loi Pareto hybride, optimisation de portefeuille.

Le futur n'est plus la projection du passé, mais la remise en cause de ce dernier. Ainsi s'exprime Jean-Paul Delevoye dans l'*Actuariel*, n°6. Cette pensée introduit parfaitement la problématique que nous avons tenté d'élucider au cours de notre étude consistant à créer un générateur de scénarios économiques ; l'objectif final de nos travaux étant d'optimiser l'allocation d'actifs d'une SICAV créée pour s'adapter au contexte de risques croissants du début du XXI^{ème} siècle. Comment créer un générateur de scénarios économiques adapté à l'environnement économique actuel ? Dans la mesure où il n'est plus envisageable d'utiliser les données passées comme seul moyen de calibration des équations régissant un générateur de scénarios économiques, quelle alternative peut-on proposer ? Notre solution consiste à intégrer des sauts dont les paramètres sont déterminés à dire d'experts, en adaptant à notre problématique la méthode d'analyse d'intervention proposée par Box et Tiao. L'outil proposé doit s'inscrire dans la politique de gestion de la SICAV pour laquelle il a été créé car il vise à optimiser l'allocation d'actifs de ce fonds ; une difficulté consistait à proposer un équilibre entre les approches qualitative et quantitative utilisées pour gérer cette SICAV, tout en garantissant une rentabilité performante.

Tout d'abord, nous rappellerons l'ensemble des éléments démontrant la nécessité de la construction d'un tel générateur de scénarios économiques.

Ensuite, nous décrirons les principales caractéristiques de l'outil créé en insistant sur les choix réalisés et la démarche adoptée. Nous parlerons notamment des différentes modifications apportées à la structure initiale du modèle.

Enfin, nous nous intéresserons à la méthode choisie pour intégrer l'outil créé dans la stratégie de gestion de la SICAV et pour interpréter les projections, ainsi que les résultats obtenus.

Les premières lignes du mémoire sont consacrées à la présentation de l'intérêt du sujet centré autour de deux axes principaux :

- l'évolution du contexte économique avec fortes incertitudes et chocs majeurs,
- la modification du contexte prudentiel.

Nous illustrons le premier point en étudiant brièvement la situation politico-économique mondiale qui montre qu'il existe de nombreux conflits de plusieurs ordres et dans de nombreuses parties du monde : des conflits militaires, par exemple en Syrie ou plus globalement au Moyen Orient ; des conflits économiques provoqués notamment par l'émergence de pays tels que la Chine ou le Brésil ; mais aussi, des conflits politiques. De plus, nous citons également une étude réalisée par François-Xavier Chevalier, montrant que d'un point de vue économique, nous constatons un bouleversement dans le cycle long de Kondratieff, ce qui traduit un changement économique important. Pour résumer cette nouvelle situation mondiale, Denis Kessler affirme que depuis le début du XXI^{ème} siècle, nous sommes entrés *dans un contexte de chocs majeurs et de fortes incertitudes*. Dans ce contexte instable, la rationalité des investisseurs est remise en cause et cela rend indispensable l'utilisation de générateurs de scénarios économiques.

Le second point est illustré par le rappel de la nouvelle directive européenne pour les compagnies d'assurance, *Solvabilité 2*, qui impose la génération de scénarios économiques et qui prendra effet au premier janvier 2014.

Ainsi, les modèles traditionnels d'allocation d'actifs basés sur la reproduction du passé ne nous semblent plus adaptés à ce nouveau contexte économique et à cette nouvelle exigence prudentielle. En effet, un modèle autorégressif simple ne permet plus une allocation d'actifs efficace dans ce contexte très changeant. L'ensemble de ces éléments nous a convaincus de la nécessité de considérer un modèle autorégressif, couplé à l'étude de l'intégration de sauts dans les équations du modèle, afin de prendre en compte la nouvelle situation économique dans notre générateur de scénarios économiques. Nous décrivons les principales caractéristiques de ce dernier ci-après.

La majeure partie de ce mémoire est consacrée à la description du modèle créé et à la justification des choix réalisés. Nous avons choisi de construire un générateur de scénarios économiques en utilisant la structure en cascade proposée par David Wilkie en 1985. Le choix de cette structure se justifie à plusieurs égards : il s'agit d'une structure simple, complète et modifiable facilement. Par ailleurs, une étude réalisée sur ce modèle a prouvé son efficacité. Dans la suite, nous présentons les ajustements successifs apportés au modèle initial pour l'adapter à notre problématique.

En premier lieu, nous proposons d'utiliser des données mensuelles sur la période [2000 ; 2011]. Ceci résulte du fait que l'on souhaite prendre en compte l'augmentation, perçue depuis le début du XXI^{ème} siècle, de la probabilité d'observer des crises majeures. L'utilisation de données mensuelles permet de s'assurer que l'on possède assez de données et s'inscrit dans notre volonté d'optimiser l'allocation d'actifs d'un fonds. Par ailleurs, nous

présentons les résultats obtenus pour quatre pays particuliers : les États-Unis, la France, l'Allemagne et le Japon. Le GSE est toutefois applicable à d'autres pays, sous réserve de la qualité des données.

En deuxième lieu, nous avons étudié la pertinence d'utiliser le processus autorégressif d'ordre un, proposé par D. Wilkie, pour modéliser l'inflation, grandeur centrale du générateur construit. Nous proposons une étude de séries temporelles en choisissant le critère de choix d'Akaike et en considérant l'ensemble des processus SARIMA. Nous concluons que le processus le plus adapté pour modéliser l'inflation est un SARIMA(1,1,1)(0,0,1)₁₂, quel que soit le pays considéré. Nous retenons donc cette modélisation. Notons en outre que nous avons choisi de considérer l'inflation en excluant l'évolution des prix du tabac et des produits alimentaires, car ces deux grandeurs provoquent une instabilité trop importante dans l'évolution de l'inflation.

En troisième lieu, nous nous sommes intéressés à un point central du mémoire relatif aux méthodes d'intégration de sauts dans les équations du modèle. Nous avons testé successivement trois méthodes. Tout d'abord, nous avons essayé d'utiliser une loi uniforme couplée à la modification des paramètres des résidus des équations : nous n'avons pas retenu cette méthode car la modification des paramètres des résidus introduisait trop de subjectivité. Ensuite, nous nous sommes intéressés aux lois de Poisson, adaptées à la modélisation d'événements rares : cette étude ne nous a pas pleinement convaincus et nous a persuadés qu'il fallait prévoir un processus de retour à l'équilibre après chaque saut. En effet, ordinairement, les crises durent le temps nécessaire à la correction des excès spéculatifs et on observe un retour à l'équilibre après effacement de ces excès. Enfin, nous avons étudié l'analyse d'intervention proposée par Box et Tiao : cette théorie permet de prendre en compte, lors de la modélisation SARIMA d'une série chronologique, des événements extérieurs au modèle. On apporte ainsi au modèle statistique une information supplémentaire de type qualitatif qui est intégrée au modèle, de manière additive, à l'aide de variables binaires exogènes. Les sauts ne figurant pas dans la structure intrinsèque des données, nous avons testé différents modèles en probabilisant la fonction de sauts et en déterminant certains paramètres à dire d'experts au lieu de les estimer. Nous avons finalement choisi de retenir le modèle d'intervention ponctuelle à effet rémanent dit *Temporary Change*. Ce dernier possède trois paramètres : la probabilité d'occurrence d'un saut, son amplitude et la durée de retour à l'équilibre. C'est ici que nous intégrons la partie qualitative indispensable pour s'adapter à la politique de gestion de la SICAV : les trois paramètres de la fonction de sauts sont déterminés à dire d'experts.

En quatrième lieu, nous avons souhaité également modifier la méthode de simulation des résidus des équations du modèle. En effet, traditionnellement, on observe que les résidus sont modélisés par une loi normale, alors qu'en réalité les résidus ne vérifient pas les hypothèses de normalité. Nous avons fait ce constat sur nos données en combinant l'analyse

statistique basique et les tests. On a observé que les résidus étaient leptokurtiques et asymétriques. Ainsi, nous avons choisi d'utiliser une loi Pareto hybride pour modéliser les résidus : il s'agit d'une loi dont la partie centrale est une loi normale et dont les queues de distribution sont modélisées par des lois de Pareto généralisées. Cette méthode permet de prendre en compte les caractères leptokurtique et asymétrique des résidus. Nous les modélisons donc en estimant quatre paramètres : la moyenne, l'écart-type et les deux paramètres de queues.

Nous avons décrit ci-dessus les principaux changements réalisés. Nous allons maintenant nous intéresser à l'utilisation du générateur de scénarios pour répondre à l'objectif de notre recherche, cité en introduction : optimiser l'allocation d'actifs d'une SICAV créée pour s'adapter au contexte de risques croissants du début du XXI^{ème} siècle. A ce titre, ce fonds est composé exclusivement d'actions internationales spécifiques résistant bien aux crises, des *actions résilientes*. Nous avons donc établi une démarche précise permettant une allocation efficace.

La première étape consiste à sélectionner des secteurs particuliers et à définir l'ensemble des actions dans lesquelles le gérant peut investir. Cette étape a été réalisée par le gérant du fonds au moment de la création de la SICAV.

La seconde étape consiste à établir des projections sur les croissances futures du produit intérieur brut (PIB) et des dettes de différents pays. L'idée retenue initialement était de trouver un modèle, relativement simple, permettant de projeter ces grandeurs macroéconomiques : le choix de la projection de l'évolution des dettes s'explique par le fait qu'on observe un impact de plus en plus important de celles-ci sur la croissance des pays. Cependant, nous n'avons pas trouvé de modèles simples et satisfaisants, nous avons donc choisi une autre voie : l'utilisation des indicateurs économiques avancés. Nous avons réalisé une étude afin de tester la fiabilité des indicateurs avancés de l'OCDE (plus précisément le *composite leading indicator* - CLI), et dans une moindre mesure, la qualité de l'indicateur de l'*Institut für Wirtschaftsforschung* (IFO). Nous avons conclu qu'ils étaient très performants, notamment pour les États-Unis, comme en témoigne le tableau ci-dessous : la première colonne indique le nombre de périodes de ralentissements observés, les seconde et troisième, le pourcentage de prédictions manquées (PM) et de prédictions non réalisées (PNR), la quatrième le pourcentage de bonnes prédictions (PB) et la cinquième la corrélation entre le CLI et le PIB pour chaque pays.

	Période [2000 ; 2005]					Période [2007 ; 2010]				
	Obs.	P.M.	P.N.R.	P.B.	Corr.	Obs.	P.M.	P.N.R.	P.B.	Corr.
Allemagne	8	37	7	57	0,55	2	2	20	80	0,85
France	6	15	13	72	0,79	3	9	24	67	0,65
États-Unis	6	11	21	68	0,70	3	0	19	81	0,95
Japon	3	10	17	72	0,87	3	20	22	58	0,85

TABLE 1 – Tableau récapitulatif du pouvoir prédictif du CLI sur le PIB par pays

Un défaut qu'il faut considérer est le fait que ces indicateurs prévoient un nombre trop élevé de retournements. De plus, les résultats obtenus pour l'Allemagne sont moins concluants, c'est pourquoi nous avons également considéré l'IFO. Malgré cela, nous retenons ces indicateurs avancés pour la sélection des pays dans lesquels nous souhaitons investir plus particulièrement.

La troisième étape a pour objectif de sélectionner un nombre d'actions, environ quarante, par le biais d'une analyse macroéconomique. Cette partie est indispensable pour être en accord avec le principe de gestion de la SICAV. Elle représente une partie qualitative importante et permet de ne pas considérer, par exemple, les actions peu liquides.

La quatrième étape consiste à utiliser le GSE créé : nous calibrons l'ensemble des équations pour les quarante valeurs sélectionnées ainsi que pour quatre indices particuliers servant à illustrer nos travaux : le CAC40 pour la France, le DAX30 pour l'Allemagne, le SP500 pour les États-Unis et le NIKKEI pour le Japon. Nous obtenons plusieurs projections. Les graphiques ci-dessous représentent l'évolution attendue de l'inflation à un an pour les quatre pays considérés :

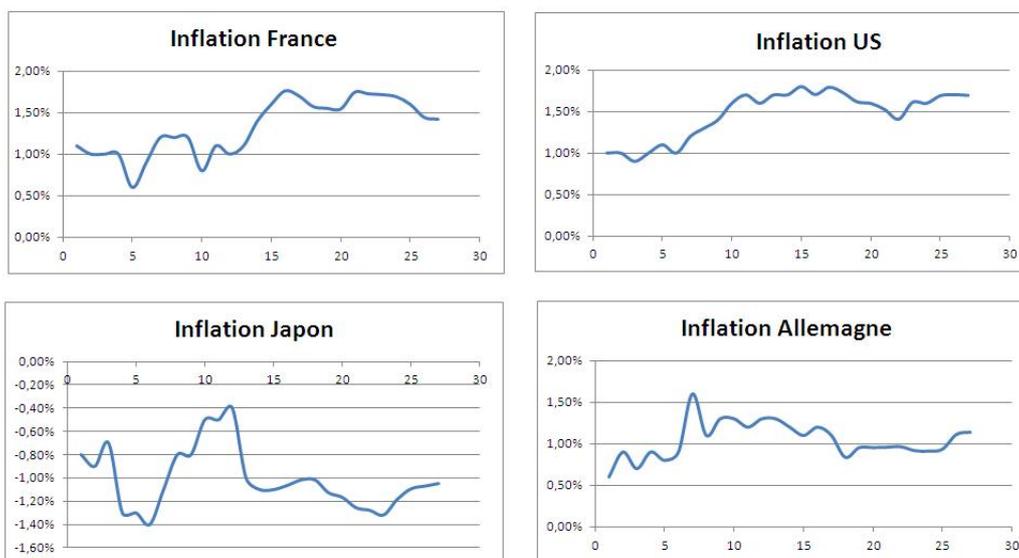


FIGURE 1 – Projections de l'inflation à 1 an

A l'issue de cette étape, nous obtenons les projections de l'ensemble des actifs que nous avons sélectionné. Notons que pour modéliser le prix des actions, nous projetons d'une part le rendement des dividendes, d'autre part, le montant des dividendes. L'évolution du prix des actions se déduit aisément en divisant le second par le premier.

La cinquième étape de ce processus est le calcul d'une utilité pertinente à partir des projections obtenues. En effet, il convient de choisir une formule d'utilité adaptée afin d'optimiser de manière efficace l'allocation d'actifs. Nous avons retenu une combinaison de la *skewness* et de la *VaR* à un an ou cinq ans avec un seuil de 5%. Ce choix traduit notre volonté de ne pas utiliser le critère moyenne-variance qui n'est pas efficace en période de crise. Notons que le calcul de l'utilité des titres ne constitue pas le coeur du mémoire mais n'est qu'une application du modèle créé.

L'indice le plus performant avec la formule d'utilité choisie est l'indice français, ce qui nous semble cohérent avec nos critères de choix : en effet, les pays qui résistent le mieux aux crises sont ceux possédant des stabilisateurs automatiques performants. Cette donnée est traduite dans les courbes par le coefficient d'asymétrie : dans le cas de la France, ce coefficient est positif, ce qui augmente l'utilité du titre.

Nous pouvons déduire de cette étude que l'utilisation d'un processus à sauts pour projeter la valeur d'actifs est indispensable dans le contexte à risques croissants de ce début de XXI^{ème} siècle. Par ailleurs, l'application de processus de gestion par scénarios nous paraît un bon compromis entre le *tout quantitatif* et le *tout qualitatif*. Enfin, précisons que ce mémoire de recherche s'inscrit dans une volonté de prise en compte de chocs dans les modèles de gestion d'actifs. Plusieurs travaux supplémentaires pourraient venir compléter l'étude initiée : l'implémentation précise d'un modèle de taux en incluant un processus de sauts à l'aide de martingales ; ou bien, l'étude de concepts mathématiques différents, par exemple les copules, pour simuler une corrélation plus précise entre les résidus ; ou encore, le rajout d'un processus pour l'immobilier performant. Ces sujets pourraient faire l'objet de prochains mémoires.

Executive summary

Keywords : economic scenario generator (ESG), economic crisis, inflation, ARIMA, Box and Tiao's algorithm, hybrid Pareto distribution, portfolio optimization.

*Future is not the projection of past anymore, but the calling into question of the latter*¹. That is what Jean-Paul Delevoye says in the French scientific revue, *Actuariel*, n°6. This idea perfectly introduces the question that we tried to solve in this paper. This study consists in creating an economic scenario generator (ESG) in order to optimize the strategic asset allocation of an investment fund created to take into account the current increasing economic risks of the beginning of the 21st century. How can an ESG appropriate to the current economic context be created? As it is no longer conceivable to use the past data to calibrate the equations of an economic scénario generator, which other option can be suggested? Our solution consists in integrating jumps which parameters are determined « according to experts », by adapting the Box and Tiao algorithm to our problem. The tool created has to respect the management policy of the unit trust, which it was created for because its target is to optimize the asset allocation of this fund. One of the difficulties was to suggest a balance between the qualitative and quantitative approaches used to manage this unit trust, while assuring the fund's performance.

Firstly, we will present the elements showing the necessity of the creation of such an economic scenario generator.

Secondly, we will describe the main characteristics of the tool created. We will insist on the choices we made and on the approach we adopted. We will particularly talk about the different modifications brought to the initial model structure.

Finally, we will bend on the chosen method in order to integrate the tool in the strategic management approach of the unit trust and to interpret the projections as well as the results of the studies.

The first lines of this paper focus on the interest of the subject dealt with. There are two main axes :

1. *Le futur n'est plus la projection du passé, mais la remise en cause de ce dernier.*

- the evolution of the economic context with strong uncertainty and major shocks,
- the modification of the prudential context.

We illustrated the first point by briefly studying the world politico-economic situation which shows that several different conflicts take place in many parts of the world such as : military ones, for example in Syria or more generally in the Middle East countries ; economic ones, particularly provoked by developing countries, such as China or Brazil ; and political ones too. Besides, we also quoted a study realized by François-Xavier Chevalier showing that from an economic point of view, we observe a modification in the Kondratieff's long cycle, which reflects an important economic shift. To sum up this new world situation, Denis Kessler says that since the beginning of the 21st century, we have entered *a context of major shocks and strong uncertainty*². In this unstable context, the use of an ESG is essential because it is during such a period that the investors' rationality is called into question.

The second point is illustrated by the reminder of the new European directive for insurance company, *Solvency 2*, which imposes the achievements of economic scenarios and which will take effect on the first of January 2014. In this way, traditional models, based on the use of past, does not seem adapted to this new economic and prudential context anymore. Indeed, a simple autoregressive model does not allow an efficient asset allocation in this shifting context. All this elements convinced us of the necessity to integrate jumps in the ESG's autoregressive equations. We are going to describe the main characteristics of this ESG bellow.

The main part of this paper is dedicated to the description of the model created and to the choices we made. We decided to build an ESG using the cascade structure proposed by David Wilkie in 1985. We can justify this choice at several levels : it is a simple and complete structure which can be easily modified. Besides, a study realized on this model showed its efficiency. Now, we present the successive adjustments we made to the initial model to adapt it to our problematics.

Firstly, we proposed to use monthly data, considering the period [2000 ; 2011]. We made this choice because we wanted to take into account the increase of the probability of major crisis seen since the beginning of the 21st century. The use of monthly data ensures that we have enough data and reflects our will to optimize the asset allocation of a fund. Besides, we studied four particular countries : the United States, France, Germany, Japan. Nevertheless, the ESG is applicable to other countries, subject to data qualities.

Secondly, we studied the adequacy of the use of a first order autoregressive model, proposed by D. Wilkie, to model the inflation, the central variable of the created ESG. We proposed a study of temporal series, choosing the Akaike criterion and considering all

2. *Dans un contexte de chocs majeurs et de fortes incertitudes*

the SARIMA processus. Our conclusion was that the most adapted processus to model inflation is a SARIMA(1,1,1)(0,0,1)₁₂, whatever the country. Furthermore, it should be noticed that we chose to considerate inflation, excluding the evolution of tobacco and food prices because these two quantities introduce an instability too important in the inflation's evolution.

Thirdly, we bent on one of the central points of this paper : the methods to introduce jumps in the equations of the model. We successively tested three methods. First of all, we tried to use the uniform distribution with a modification of residual parameters of the equations ; we did not keep this method because the modification of residuals introduced too much subjectivity. Then, we studied the use of Poisson distribution, appropriated to the modelling of rare events ; this study did not convince us and persuaded us to introduce a process of return to overall balance after each jump. Indeed, usually, crisis should remain in place for as limited a period as necessary to erase the speculative excess and we observed a return to overall balance after this period. Finally, we studied the Box et Tiao's algorithm : this theory allows to take into account exterior events, during the SARIMA modelling of a chronological series. We thus added to the model a qualitative information which is integrated in an additive way by the use of binary exogen variable. Since the jumps are not in the intrinsic structure of data, we tested different models, probalizing the jumps function and by estimating some parameters, according to experts, instead of the traditionnal estimation. We chose the tempory change model which has three parameters : the probability of a jump, its amplitude et its duration return to overall balance. Here, we integrated the essential qualitative part in order to be in accordance with the management policy of the unit trust : the three parameters are determinated according to experts.

Fourthly, we tried to modify the traditional modelling of residuals. Indeed, usually, we observe that residuals are modelled by using the standard normal distribution wherease the residuals do not verify the required hypothesis. We made this observation on our data by combining the basic statistical analysis and the tests. We noticed that residuals were leptokurtic and asymmetric, that is why we chose to use a hybrid Pareto distribution to model residuals : it consists in using normal distribution and generalized Pareto distribution to respectively model the central part and the tails of the distribution. This method allowed us to take into account the leptokurtic and asymmetric characteristics of the residuals. Therefore, we modelled residuals by estimating four parameters : average, standard deviation and the two parameters of tails.

We described above the main changes we made on the model. Now, we will focus our attention on the use of ESG. As we said in the introduction, the aim of our study was to optimize the asset allocation of a unit trust created to adapt itself to the context of increasing risks observed since the beginning of the 21st century. As such, this fund

is exclusively composed of specific international actions resisting well to the crises. We established a precise approach allowing an efficient allocation.

The first step consists in selecting particular sectors and defining all the shares in which the manager can invest. This step was realized by the manager of the unit trust.

The second step consists in establishing projections on the future growths of the gross domestic product (GDP) and the debts of various countries. The initial idea was to find a model, relatively simple, allowing to simulate these macroeconomic variables. The choice of the projection of the evolution of the debts is explained by the fact that we observed an increasing impact of these on the GDP of countries. However, we did not find a simple and satisfactory model; we thus chose another way : the use of the advanced economic indicators. We realized a study to test the reliability of the leading indicators of the OECD (more exactly *leading indicator composite* - CLI), and to a lesser extent, the quality of the indicator of the *Institut für Wirtschaftsforschung*(IFO). We concluded that they were very efficient, particularly for the United States, as showed in the table below. The first column indicates the number of periods of observed slowing down, the second and third, the percentage of missed predictions and unimplemented predictions (MP and UP), the fourth the percentage of good predictions (GP) and the fifth the correlation between the CLI and the GDP for each country.

	Period [2000 ; 2005]					Période [2007 ; 2010]				
	Obs.	M.P.	U.P.	G.P.	Corr.	Obs	M.P.	U.P.	G.P.	Corr.
Germany	8	37	7	57	0,55	2	2	20	80	0,85
France	6	15	13	72	0,79	3	9	24	67	0,65
The United States	6	11	21	68	0,70	3	0	19	81	0,95
Japan	3	10	17	72	0,87	3	20	22	58	0,85

TABLE 2 – Summary of the predictive power of the CLI on the GDP by country

A drawback that we need to take into account is the fact that these indicators plan a number of reversals too high. Furthermore, the results obtained for Germany are less convincing, that is why we also considered the IFO. Nevertheless, we retained these leading indicators for the selection of the countries in which we wish to invest.

The third step aims at selecting a number of shares, approximately forty, by using a macroeconomic analysis. This part is essential to be in agreement with the principle of management of the unit trust. It represents an important qualitative step and allows not to consider, for instance, the non liquid shares.

The fourth step consists in using the created ESG. We calibrated all the equations for the forty values selected as well as for the four particular indexes to illustrate our work : CAC40 for France, DAX30 for Germany, SP500 for the United States and NIKKEI for

Japan. We obtained several projections. The graphs below represent the evolution expected from the inflation in one year for the four countries :

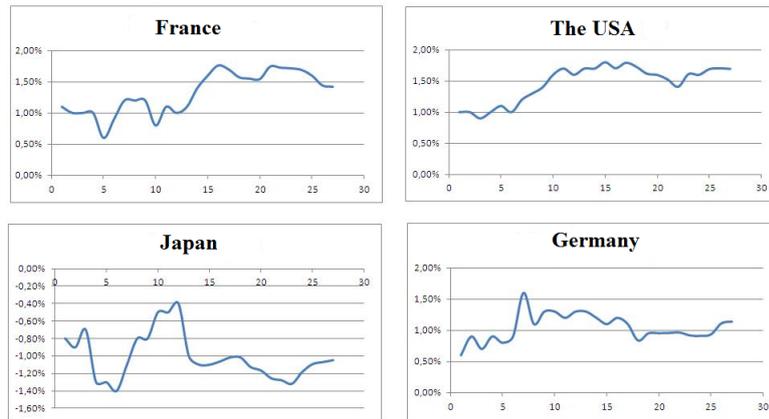


FIGURE 2 – One year inflation projection

At the end of this step, we obtained the projections of all the selected assets. Let us note that to model the equity price evolution of stock's price, we model, on the one hand the dividend yield and, on the other hand, the amount of dividends. We can easily deduce the evolution of the equity price by dividing the second by the first.

The fifth step of this process is the calculation of a relevant utility from the obtained projections. Indeed, it is necessary to choose a formula of utility adapted to optimize in an effective way the asset allocation. We retained a combination of *skewness* and *VAR* in one year or five years with a threshold of 5%. We made this choice because we did not want to use the criterion average-variance which is not effective in times of crisis. Let us note that the calculation of the utility does not constitute the heart of the study but is only an application of the created model.

The most performing index with the chosen utility formula is the French index. This conclusion seems to coincide with our criterion of choice : indeed, the countries which resist best the crises are the ones possessing successful automatic stabilizers. This data is translated in curves by the coefficient of asymmetry : in the case of France, this coefficient is positive, which increases the utility of the index.

We can deduce from this study that the use of a process with jumps to model the value of assets is essential in the context of the increasing risks of the beginning of the 21st century. Besides, the application of process of management by scenarios appears to us as a good compromise between *all quantitative* and *all qualitative*. Finally, let us precise that this paper lies within a will to take into account shocks in the asset management model. Several other studies could complete this study, such as : a precise implementation of rates model including a jump process with the use of martingales; or the study of

different mathematical concepts, such as the use of copula to simulate a more accurate correlation between residuals ; or whether, the add of an efficient property process. All this ideas could be thoroughly investigated in future papers.

Résumé

Mots clés : générateur de scénarios économiques, risques croissants, crise économique, inflation, modélisation ARIMA, algorithme de Box et Tiao, loi Pareto hybride, optimisation de portefeuille.

L'objectif de ce mémoire est de présenter un générateur de scénarios économiques construit à l'aide de modèles mathématiques à sauts et permettant la prise en compte de scénarios centraux lors des projections.

Cette étude possède deux visées principales : d'un point de vue actuariel, nous introduisons la notion de sauts dans les générateurs de scénarios économiques, notion peu utilisée en actuariat mais pourtant indispensable du fait de l'augmentation sensible du risque de crises majeures ; d'un point de vue financier, nous présentons une méthode d'optimisation de portefeuille en adoptant une gestion dite par les risques.

Ce mémoire a été réalisé dans le cadre d'un stage de fin d'études pour une société de gestion indépendante, **Diamant Bleu Gestion**. L'étude proposée vise à améliorer l'optimisation d'un fonds spécialisé dans la recherche de *valeurs résilientes* et géré selon une approche à la fois macroéconomique et de sélection de valeurs. Nous proposons de modéliser la croissance économique et la croissance des dettes en étudiant notamment les indicateurs avancés de l'OCDE pour faciliter l'approche *top down* ; le calcul de l'utilité des titres, suite aux projections réalisées par l'intermédiaire du générateur de scénarios, s'intègre dans l'approche *bottom up*.

Le générateur de scénarios est construit à partir d'une structure de base en cascade, celle du modèle de Wilkie (1985), à laquelle plusieurs modifications sont apportées : intégration de sauts avec l'**algorithme de Box et Tiao** ; modélisation de la loi des résidus par une **loi Pareto Hybride** ; considération de données mensuelles et modification de la modélisation de l'inflation.

Dans les cinq premières parties de ce mémoire, nous présentons les outils théoriques permettant de construire le modèle. Les parties suivantes portent sur la calibration et l'application du modèle créé.

Abstract

Keywords : economic scenario generator, economic crisis, inflation, ARIMA, Box and Tiao's algorithm, hybrid Pareto distribution, portfolio optimization.

The purpose of this document is to demonstrate how to create an economic scenario generator which allows the user to include probability of crisis thanks to process with jumps.

This paper has two main objectives : from an actuarial point of view, we will try to introduce the notion of the inclusion of jumps with the use of **Box and Tiao's algorithm** in temporal series. This is not frequently applied in actuarial science but this is, from our point of view, essential in the current economic and politic context. From a financial point of view, the aim is to help the optimization of an equity portfolio with the inclusion of probability of crisis.

The economic scenario generator is created for an investment fund of an independent management company, **Diamant Bleu Gestion**. The investment fund is an international equity one with a non-common management method. Indeed, the fund manager uses a double approach. The first one is a top down approach to focus its investments on few sectors : to quantitatively justify this step, we model the evolution of growth and debt's growth by studying the leading indicators proposed by OECD. The second one is a bottom up approach to do the stock picking : the calculus of the use of each equity of the portfolio with the economic scenarios generator represent a great help to monitor the portfolio.

To create our model, we will make use of the Wilkie model and his cascade structure. Then, we will do two significant changes : we will modify the equation of inflation and we will change the residuals modeling by using a new probability distribution : the hybrid Pareto distribution.

In the firts five parts, we will introduce the useful theoretical tools needed to understand the model. We will attach a particular importance to two notions : the Box and Tiao algorithm and the hybrid Pareto distribution.

In the other parts, we will show the calibration of the equations and the results of our study and we will discuss about potential improvements.

Remerciements

Avant toute chose, je tiens à adresser des remerciements à plusieurs personnes ayant rendu possible la réalisation de ce mémoire de fin d'études.

Je souhaite remercier, en premier lieu, Arnaud Clément-Grandcourt, gérant de la SICAV pour laquelle le modèle présenté est construit et président du jury de l'institut des actuaires, pour m'avoir permis de travailler avec lui sur un sujet au coeur de l'actualité, pour avoir répondu à mes questions durant tout mon stage en me donnant des conseils avisés et pour m'avoir suivi de près durant les six mois passés à Diamant Bleu Gestion. Mes remerciements s'adressent également à toute l'équipe de Diamant Bleu Gestion, notamment Christian Jimenez, Hugues Le Maire et Jonathan Jimenez, respectivement président, directeur général et gérant, pour m'avoir accueilli au sein de leur entreprise et m'avoir aidé à surmonter certaines difficultés.

J'exprime également toute ma gratitude à mes enseignants de l'Institut de Statistiques de l'Université de Paris qui m'ont permis d'acquérir les bases nécessaires à l'élaboration de ce mémoire : en particulier, Michel Delecroix, directeur de l'ISUP ; Olivier Lopez, mon tuteur de stage qui m'a permis d'effectuer de sensibles améliorations ; et Jacques Chevalier, professeur de mathématiques financières, qui m'a transmis son goût pour la finance.

Enfin, j'adresse mes remerciements mais également mes excuses aux personnes qui m'ont soutenu durant mon cursus scolaire et qui ne figurent pas dans les personnes sus-citées.

Sommaire

1	Introduction	11
1.1	Pourquoi s'agit-il d'un sujet central?	11
1.2	Les critiques	11
1.3	La gestion en période de crises	12
1.3.1	Quelques mots sur la gestion prudente	12
1.3.2	Les indicateurs avancés	13
1.4	Objectifs du mémoire	14
2	Le générateur de scénarios de Wilkie	16
2.1	La structure du modèle de Wilkie	16
2.1.1	Généralités	16
2.1.2	Les équations régissant le modèle	17
2.2	Avantages, inconvénients et autres modèles	19
2.2.1	Avantages	20
2.2.2	Inconvénients	20
2.2.3	Autres modèles	21
3	Les séries temporelles : modèles ARMA, ARIMA et SARIMA	25
3.1	Processus autorégressif et à moyenne mobile	25
3.1.1	Définitions	25
3.1.2	Estimation	26
3.2	Processus non stationnaires	27
3.2.1	Définitions	27
3.2.2	Modèle ARIMA	27
3.2.3	Modèle SARIMA	28
3.3	Les tests	29
3.4	La méthodologie de Box et Jenkins	30
4	Intégration de sauts dans une série temporelle	34
4.1	Le contexte politico-économique mondial	34
4.2	Intégration de sauts en utilisant des lois uniforme ou de poisson	36

4.2.1	Modification des paramètres des résidus	36
4.2.2	Utilisation de la loi de Poisson	36
4.3	L'algorithme de Box et Tiao (1975)	37
4.3.1	Le modèle	38
4.3.2	Les différents types de sauts	39
5	Modélisation des résidus des équations	42
5.1	Loi de Pareto généralisée	42
5.1.1	Généralités	42
5.1.2	Fonction quantile	43
5.2	Loi de Pareto hybride	43
5.2.1	Généralités et densité de probabilité	43
5.2.2	Contraintes sur les paramètres	44
5.2.3	Fonction de répartition et fonction quantile	44
6	Modélisation de l'inflation	46
6.1	Présentation des données	46
6.2	Application de l'algorithme de Box et Jenkins	46
6.2.1	Stationnarisation et différenciations	46
6.2.2	Désaisonnalisation	48
6.2.3	Paramètres optimaux	49
6.2.4	Validation de la modélisation	51
6.3	Inclusion de sauts dans l'équation	52
6.4	Lissage	54
7	Modélisation de la croissance et de la croissance des dettes	55
7.1	Modélisation de la croissance	56
7.1.1	Définitions	56
7.1.2	Choix du modèle et calibrage	56
7.1.3	Indicateurs de l'OCDE	60
7.2	Indicateur IFO	62
7.3	Modélisation de la croissance des dettes	62
7.3.1	Notion de solde primaire stabilisant	62
7.3.2	Modèle retenu	63
8	Calibrage du modèle et simulations	66
8.1	Données et calibrage des équations	66
8.1.1	Données	66
8.1.2	Calibrage	68
8.2	Méthodes de simulations	69

8.2.1	Générateurs de nombres aléatoires	69
8.2.2	Générations de lois de probabilité	70
8.3	Projections et comparaisons	71
8.3.1	Projections obtenues	71
8.3.2	Comparaisons	75
9	Optimisation d'un portefeuille	77
9.1	Les indicateurs de risque	77
9.1.1	Variance, <i>skewness</i> et <i>value at risk</i>	77
9.1.2	Calcul de la <i>value at risk</i>	77
9.2	Actions résilientes	78
9.3	Fonction d'utilité	78
9.4	Méthode d'optimisation	79
9.4.1	Processus global d'optimisation	79
9.4.2	Calcul de l'utilité	80
9.4.3	Résultats	81
A	Description des tests utilisés	91
B	Critère du maximum de vraisemblance	96
C	Modèle de taux : Vasicek en termes HJM	97
D	Résultats sur les indicateurs avancés	103
E	Coefficients retenus pour calibrer les résidus	105
F	Le bootstrap non paramétrique par blocs	106

Chapitre 1

Introduction

1.1 Pourquoi s'agit-il d'un sujet central ?

Les générateurs de scénarios économiques (GSE) sont utilisés à la fois dans les institutions financières et dans les compagnies d'assurance.

Pour les compagnies d'assurance, la projection des actifs permet d'assurer un pilotage technique cohérent : une des caractéristiques spécifiques aux compagnies d'assurance étant d'avoir des passifs à long terme, la projection des actifs est indispensable, notamment pour la gestion dite actif-passif. Par ailleurs, les nouvelles normes réglementaires en assurance, notamment « Solvabilité 2 », imposent la génération de scénarios économiques. En effet, pour être en conformité avec cette directive, une compagnie d'assurance doit être en mesure de générer des scénarios lors du calcul des provisions techniques relatives aux engagements d'assurance contenant des options et garanties financières ainsi que lors de l'estimation du capital de solvabilité par le biais d'un modèle interne.

Pour les institutions financières, les générateurs de scénarios économiques sont également utilisés, notamment pour réaliser des processus d'optimisation de portefeuille. En effet, ils permettent aux investisseurs et aux gérants d'une part, de prendre des décisions basées sur le passé et sur l'avenir ; d'autre part, de prendre des décisions plus rationnelles. Ainsi, plus l'environnement est instable, plus les générateurs de scénarios économiques sont utiles car c'est en période de crises que la rationalité des investisseurs est remise en cause.

Dans la période de crises à répétition que nous traversons, il paraît donc clair que la réalisation d'un générateur de scénarios économiques incluant des probabilités de crises est un sujet d'actualité à la fois pour les compagnies d'assurance et pour les institutions financières.

1.2 Les critiques

Par définition, un générateur de scénarios économiques, en tant que modèle, permet de produire des représentations simplifiées d'une réalité dont la complexité et l'imprévisibilité

sont inépuisables. A ce sujet, de nombreux auteurs et statisticiens ont exprimé des critiques et des réserves, comme en témoignent les quelques citations suivantes :

- Tous les modèles sont faux, certains sont utiles. (Georges Box)
- Un modèle simple est toujours faux ; un modèle complexe est inutilisable. (Paul Valéry)
- La statistique est l’art de préciser ce qu’on ignore. (Adolphe Thiers)

Ces citations traduisent le fait que les projections obtenues à l’issue de ce mémoire ne seront pas la prévision de la réalité : on pourra conclure que les projections obtenues seront proches des résultats réels, sous une certaine probabilité. La réticence de plusieurs personnes à l’égard de l’utilisation de modèles est souvent dû au fait qu’ils sont assimilés à des *boîtes noires*. Pour éviter cette critique, nous nous efforcerons d’être le plus transparent possible tout au long de ce mémoire.

Par ailleurs, nous devons admettre que nous sommes contraints de considérer un cadre d’étude qui n’est en fait pas la réalité et que pour modéliser le futur, on se repose en grande partie sur le passé car il s’agit des seules informations que l’on possède. Actuellement, dans l’environnement à risques croissants de ce début de XXI^{ème} siècle, du fait notamment de l’émergence de la Chine, il semble légitime de penser que le passé ne présage pas de l’avenir ou du moins, qu’il est nécessaire d’utiliser des informations exogènes au cadre initial pour que les projections obtenues soient fiables.

Afin de répondre à ces différentes critiques, nous proposons dans ce mémoire de créer un générateur de scénarios économiques relativement simple et permettant à l’utilisateur d’intégrer des probabilités de crises, déterminées à partir de données exogènes et conjoncturelles

1.3 La gestion en période de crises

En période de crises, différents types de gestion sont à proscrire. Nous souhaitons tout d’abord rappeler quelques notions fondamentales de la gestion en période de crises. Le deuxième paragraphe de cette partie sera consacré à l’introduction des indicateurs avancés qui s’avèrent très utiles dans ce type d’environnement.

1.3.1 Quelques mots sur la gestion prudente

Ce mémoire étant consacré à la prise en compte de crises, il nous semble nécessaire d’évoquer quelques principes fondamentaux de la gestion à adopter sans lesquels l’interprétation des projections réalisées serait trompeuse. En effet, comme le dit Arnaud Clément-Grandcourt dans [2], anticiper une crise est vraiment très difficile et il est dangereux de s’illusionner sur le sujet. L’examen critique des crises, a posteriori, permet de comprendre certains comportements irrationnels de gérants et d’éviter certains écueils. En temps de crise, trois méthodes de gestion sont envisageables :

- Adopter une philosophie prudente dans les marchés baissiers basée sur la gestion des risques et la diversification en *valeurs résilientes* (nous définirons cette notion dans

- le dernier chapitre),
- Privilégier les investissements dans les fonds gérés par des gérants qui ont été habiles dans les cas précédents,
 - S’efforcer d’être dans les trois premiers déciles du troupeau des concurrents.

Un gérant prudent est un gérant qui limite le nombre et l’importance de ses erreurs en prenant des risques mesurés après avoir bien pris en compte tous les paramètres et qui ne surestime pas sa capacité de prévision et de compréhension des évènements. La *gestion benchmarkée*, trop proche de la gestion indicielle, est donc à éviter par manque de valeur ajoutée. C’est pour mieux évaluer les risques et pour faire mieux que le benchmark que nous construisons le générateur de scénarios économiques.

Par ailleurs, il faut noter que la méthodologie prudentielle progresse du fait des crises : augmentation du nombre de scénarios, contrôle des risques ordinaires (*due diligence, best practise*), contrôle des risques qui sortent de l’ordinaire (*stress scénarios*), contrôle des risques extrêmes (*worst case scénarios*). Cette méthodologie va continuer de se développer du fait de la difficulté à maîtriser la spéculation et les crises. Notre générateur de scénarios économiques s’inscrit dans cet environnement et vise à augmenter le nombre de prises de décisions prudentes et rationnelles.

Dans ce contexte de crises, il est néanmoins nécessaire de garder à l’esprit qu’être prudent nécessite parfois d’être offensif comme le résume parfaitement cette citation : *a good offense is sometimes prudent*. Les gérants contrariants de Boston et d’Edimbourg sont hardis mais néanmoins prudents.

Pour faciliter une prise de décision cohérente et rationnelle, le recours aux indicateurs avancés peut être d’une grande efficacité.

1.3.2 Les indicateurs avancés

En introduction, nous souhaitons également évoquer l’utilisation d’indicateurs économiques dont l’évolution reflète avec un peu d’avance celle de l’économie réelle. En période de crises, les indicateurs avancés permettent d’anticiper des retournements de conjoncture et de modifier, par exemple, la composition de son portefeuille. Cependant, le nombre de mois d’avance dépend de l’indicateur choisi et il est nécessaire de tester sa fiabilité en période de crises. Nous présenterons les indicateurs que nous proposons d’utiliser au chapitre sept de ce mémoire, leur place dans la stratégie d’optimisation de portefeuille ainsi que leur fiabilité en période de crise.

Actuellement, les indicateurs les plus couramment utilisés sont l’indice des prix à la consommation ou à la production, l’indice de production industrielle ainsi que la balance des paiements. Nous verrons que l’on retient les Composite Leading Indicator (CLI) construits par l’OCDE, qui n’ont pas manqué de récession mais qui prévoient un nombre de ralentissements trop important. Nous nous intéresserons également à l’IFO qui est un très bon prédicteur de l’évolution de l’économie allemande.

1.4 Objectifs du mémoire

L'objectif de ce mémoire est de réaliser un générateur de scénarios économiques basé sur le modèle de Wilkie, dans sa première version de 1985. Nous proposons une méthode permettant la prise en compte de crises majeures telles qu'une baisse brutale de l'inflation ou un choc pétrolier ou encore un choc alimentaire impliquant de l'inflation... Nous allons remplacer un unique GSE par plusieurs GSE pondérés. Le but de cette démarche est de diminuer le nombre de scénarios totalement impossibles en considérant d'une part que des GSE spécialisés produisent des scénarios plus réalistes que des GSE généraux et d'autre part qu'un seul GSE avec équipondération des scénarios conduit à une trop forte influence du passé. Les GSE créés seront pondérés à l'aide d'indicateurs avancés.

Par ailleurs, notons qu'un modèle autorégressif à sauts permet d'élargir les possibilités d'anomalies et donc la gamme des scénarios. L'inclusion de sauts dans les équations, permettra la prise en considération de crises extraordinaires : c'est ici que se trouve la part de l'originalité de ce mémoire. Cette notion a été introduite par Taleb dans [21] et est devenue incontournable, dans le contexte actuel.

Les cinq types de crises que l'on souhaite modéliser sont les suivants :

- Générer des crises extraordinaires et les crises induites ultérieurement parmi un catalogue de crises envisageables c'est-à-dire des cygnes noirs suivis de papillons noirs : ce genre de crise ne peut être prévue ; par contre les crises induites sont envisageables. Il peut s'agir de crises économique, financière, bancaire, sociale, de crise inflationniste en plusieurs épisodes (stop and go) ou encore de crise d'endettement.
- Générer des accidents inflationnistes ou plus généralement des chocs envisageables c'est-à-dire des papillons noirs : choc alimentaire, monétaire, pétrolier avec des conséquences sociales...
- Générer des récessions ordinaires
- Générer une période de réduction de dettes sans restructuration, sans défaut ni crise sociale c'est-à-dire une déflation à la japonaise.
- Générer une période de réduction des dettes avec restructuration aléatoire, récessions, crises sociales et bancaires.

Afin de pondérer les GSE créés, nous avons étudié la possibilité d'utiliser différentes méthodes :

- Attribution subjective d'une probabilité pour pouvoir utiliser les GSE, agrégation bayésienne.
- Attribution de pondération en se référant à des indicateurs avancés de conjoncture, des indicateurs globaux de l'OCDE et de l'IFO, des indicateurs d'inflation et d'aversion au risque.

L'attribution de probabilités à chaque GSE est délicate c'est pourquoi elle ne constituera pas l'objet central du mémoire. En effet, cette étape possède intrinsèquement une part de subjectivité. C'est la raison pour laquelle nous nous contenterons de présenter la construction du générateur de scénarios à sauts et que nous ne nous attarderons pas sur

l'attribution des probabilités.

Par ailleurs, l'optimisation d'un portefeuille est délicate et est soumise à l'appréciation de l'utilisateur. En effet, le choix de la fonction d'utilité utilisée peut être discuté. Nous verrons, dans le dernier chapitre, la fonction d'utilité retenue et la démarche mise en place. Nous justifierons les choix réalisés. Toutefois, cela ne représente pas le coeur du mémoire.

Les quatre chapitres qui suivent visent à introduire les notions théoriques nécessaires à la construction du GSE. Dans le chapitre suivant, nous présentons les équations régissant le générateur de Wilkie. Dans le chapitre trois, nous étudions les outils relatifs aux séries temporelles. Les chapitres quatre et cinq s'intéressent respectivement à l'inclusion de sauts dans une série temporelle et à la présentation d'une distribution originale, la loi de Pareto hybride.

Les chapitres six, sept et huit seront consacrés à la construction de notre GSE : nous présenterons le modèle retenu pour l'inflation, pour la croissance ainsi que le calibrage des données et les simulations obtenues.

Enfin, le dernier chapitre est un exemple d'application du modèle créé : nous optimisons un portefeuille d'actions en mettant en place une démarche de recherche d'*actions résilientes*.

Chapitre 2

Le générateur de scénarios de Wilkie

Les premières tentatives actuarielles de modélisation de scénarios économiques proviennent des travaux de David Wilkie, actuaire écossais. Il s'agit d'un modèle en cascade dans lequel l'inflation joue un rôle central : toutes les valeurs des autres variables dépendent de l'inflation. Dans la première version du modèle, datant de 1985, l'inflation est modélisée par un processus autorégressif d'ordre un ; dans la seconde, celle de 1995, Wilkie la modélise par un processus plus complexe, un modèle autorégressif et conditionnellement hétéroscédastique (ARCH). Nous allons ici présenter le modèle dans sa version la plus ancienne.

Dans ce chapitre, nous allons dans un premier temps, rappeler les équations utilisées par Wilkie. La littérature étant abondante à ce sujet, nous invitons le lecteur à se reporter à la bibliographie pour trouver des informations supplémentaires sur ce générateur notamment [1], [20] et [3].

Dans un second temps, nous allons présenter les avantages du modèle de Wilkie par rapport aux autres ainsi que les différentes critiques à l'égard de ce modèle.

2.1 La structure du modèle de Wilkie

2.1.1 Généralités

En 1985, Wilkie publie [25] dans lequel il décrit un modèle intégré. Ce premier modèle se singularise par une modélisation des prix des actions à partir des taux et des montants de dividendes. Ce modèle inclut par ailleurs le taux d'inflation et les taux d'intérêt à long terme.

Le modèle de Wilkie repose sur cinq variables fondamentales :

- Q_t : L'indice des prix à la consommation
- D_t : L'indice du montant des dividendes
- Y_t : L'indice du rendement des dividendes
- C_t : Le taux d'intérêt à long terme
- B_t : Le taux d'intérêt à court terme

Comme nous l'avons indiqué en introduction, ce modèle possède une structure en cascade dont la variable centrale est l'inflation. En effet, ce modèle part du constat que le taux d'inflation permet de retrouver l'ensemble de l'état du marché. Ainsi, les équations qui accompagnent ce modèle permettent de retrouver les prix des actions, de l'immobilier... C'est la raison pour laquelle nous avons souhaité apporter une attention particulière à l'équation de l'inflation afin de l'adapter à la période actuelle. En effet, le processus autorégressif sur le taux d'inflation continu proposé par David Wilkie nous paraît peu adapté à nos données. Nous développerons plus précisément ce point dans le chapitre six.

La figure 2.1 illustre la structure du modèle en cascade de Wilkie et le rôle fondamental de l'inflation :

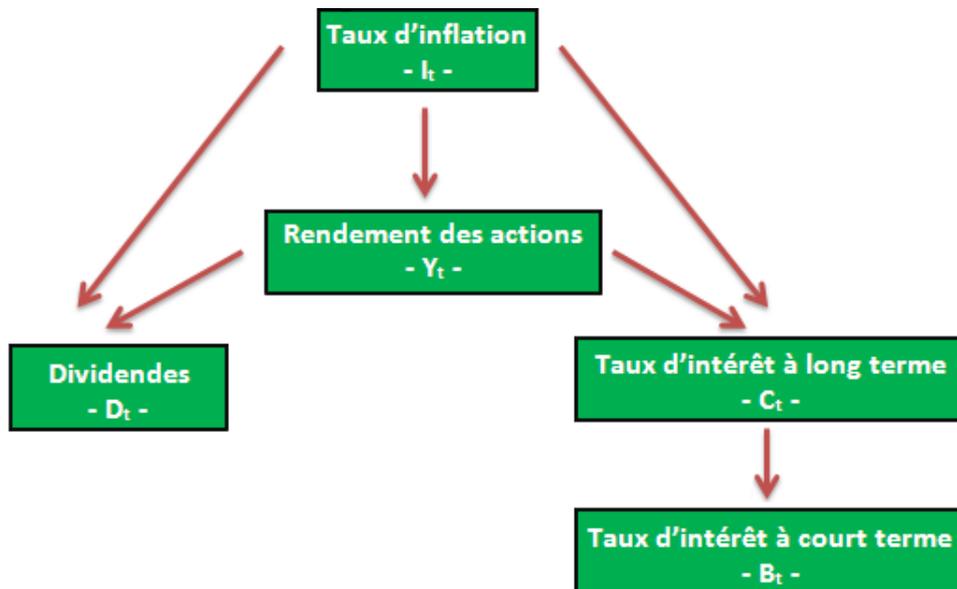


FIGURE 2.1 – Modèle de Wilkie (1985)

2.1.2 Les équations régissant le modèle

Détaillons, variable par variable, les équations régissant le modèle de Wilkie dans sa version de 1985.

Inflation

Le processus d'inflation est modélisé par un modèle autorégressif d'ordre un :

$$I_t - I_m = \alpha_I \times (I_{t-1} - I_m) + \sigma_I \times \epsilon_{I,t}$$

Avec :

- $I_t = \ln(Q_t/Q_{t-1})$: taux d'inflation continu
- I_m : inflation à long terme
- α_I : coefficient de retour à la moyenne
- σ_I : écart type du bruit
- $\epsilon_{I,t}$: bruit suivant une loi gaussienne centrée et réduite

Les paramètres de cette équation sont estimés par maximum de vraisemblance conditionnelle. Cet estimateur correspond en réalité à l'estimateur des moindres carrés ordinaires. Le calibrage de cette équation ne pose donc pas de problème majeur.

Actions

Le processus des actions est modélisé en utilisant deux processus :

- le processus de rendement des dividendes
- le processus de montant des dividendes

On en déduit aisément le prix de l'action en divisant le second par le premier.

Processus de rendement des dividendes

Wilkie suppose que le rendement des dividendes suit le processus suivant :

$$y_t - y_m = \alpha_Y \times I_t + \beta_Y \times (y_{t-1} - y_m) + \sigma_Y \times \epsilon_{Y,t}$$

Avec :

- $y_t = \ln(Y_t)$ où Y_t est l'indice du rendement des dividendes
- y_m : logarithme du rendement des dividendes à long terme
- I_t : taux d'inflation continu
- α_Y : taux de corrélation entre inflation et rendement des dividendes
- β_Y : taux de retour à la moyenne du rendement des dividendes
- σ_Y : écart type du bruit $\epsilon_{Y,t}$ suivant une loi gaussienne centrée réduite.

Nous pourrions estimer les paramètres de cette équation en réalisant une régression sur l'inflation continue et sur les valeurs retardées du processus.

Processus de montant des dividendes

Le modèle du montant des dividendes est le suivant :

$$d_t - d_m = \alpha_D \times I_t + \beta_D \times M_t + \mu_D \times \sigma_Y \times \epsilon_{Y,t-1} + \xi_D \times \sigma_D \times \epsilon_{D,t-1} + \sigma_D \times \epsilon_{D,t}$$

Avec :

- $d_t = \ln(\frac{D_t}{D_{t-1}})$ où D_t est l'indice du montant des dividendes
- d_m : taux de croissance de long terme du montant des dividendes
- I_t : taux d'inflation continu
- M_t : lissage exponentiel du taux d'inflation continu
- α_D : taux de corrélation entre inflation et montant des dividendes
- β_D, μ_D et ξ_D : coefficients à estimer
- σ_Y et σ_D : écart type des bruits respectivement $\epsilon_{Y,t}$ et $\epsilon_{D,t}$.

Obligations

Wilkie propose une modélisation pour les taux d'intérêt long terme et court terme. Des taux long terme, on déduit aisément les taux court terme.

Taux longs

$$C_t = \alpha_C \times M_t + RC_t$$

Avec :

- M_t : lissage exponentiel du taux d'inflation continu
- $RC_t = RC_m \times \exp(N_{RC,t})$
- $N_{RC,t} = \alpha_{RC} \times (RC_{t-1} - RC_m) + \beta_{RC} \times (RC_{t-2} - RC_m) + \gamma_{RC} \times (RC_{t-3} - RC_m) + \mu_{RC} \times \sigma_Y \times \epsilon_{Y,t} + \sigma_{RC} \times \epsilon_{RC,t}$

Taux courts

Dans le modèle Wilkie, le lien de causalité entre les taux court terme et les taux long terme est exprimé par l'équation suivante :

$$\ln(B_t) = \ln(C_t) + BN_t$$

$$BN_t - BN_m = \alpha_B \times (BD_{t-1} - BD_m) + \mu_B \times \sigma_{RC} \times \epsilon_{RC,t} + \sigma_B \times \epsilon_{B,t}$$

Avec les mêmes notations que précédemment.

Dans la suite de notre mémoire, nous ne nous intéresserons pas aux deux processus de taux. En effet, d'une part la modélisation proposée par Wilkie nous semble inadéquate avec ce que nous souhaitons modéliser; d'autre part, l'objectif final de ces travaux est d'optimiser un portefeuille d'actions internationales. Si la modélisation des taux avait été nécessaire, nous aurions privilégié l'utilisation du modèle Health-Jarrow-Morton (HJM), ce modèle paraissant plus adapté que celui de Wilkie dans la mesure où il permet de simuler l'ensemble de la courbe des taux, à toutes les dates. Par ailleurs, nous aurions privilégié ce modèle car il reste relativement simple. Nous décrivons ce modèle en annexe, en proposant très brièvement une adaptation à la structure de Wilkie.

De la même façon, si la modélisation des actifs immobiliers avait été nécessaire, nous aurions étudié la possibilité d'utiliser deux modèles :

- un premier modèle autorégressif pour modéliser l'évolution des loyers; en effet, on remarque que les indices de loyer sont peu volatiles.
- un modèle plus complexe pour modéliser l'évolution de la valeur du capital : en effet, on observe que les modèles simples et sans sauts donnent des résultats peu probants.

2.2 Avantages, inconvénients et autres modèles

Le modèle de Wilkie est un modèle économétrique, calibré à partir de données historiques, selon des méthodes de séries temporelles. Le choix des équations du modèle doit répondre au double objectif de réalisme, qui nécessite une spécification en adéquation avec les données historiques, et de simplicité, qui exige une spécification relativement simple pour le temps de calcul et la lisibilité des résultats des simulations.

2.2.1 Avantages

Deux des avantages majeurs du modèle de Wilkie sont les suivants : d'une part, il s'agit d'un modèle simple, car comme nous l'avons vu ci-dessus, les équations régissant le modèle ne présentent aucune complexité mathématique ; et d'autre part, ce modèle est complet, car il est possible de modéliser toute sorte d'actifs : actions, obligations ou immobiliers. Par ailleurs, comme nous l'avons rappelé, il s'agit d'un modèle en cascade : selon nous, il est clair que pendant les crises (ce qu'on souhaite prévenir), les modèles par corrélation sont beaucoup plus instables (par exemple le modèle d'Ahlgrim).

De plus, ce modèle est caractérisé par une autre qualité : la possibilité de l'améliorer facilement tout en conservant la même structure. C'est d'ailleurs ce qu'on propose de faire dans ce mémoire : nous gardons la structure de base du modèle mais nous proposons certaines améliorations.

En outre, une étude rétrospective a été réalisée sur ce modèle [26] afin de tester sa robustesse. Elle conclut que le modèle de Wilkie est plus performant dans sa première version (1985) que dans sa deuxième (1995). Pour cette raison, nous nous intéresserons au modèle dans sa forme initiale.

Enfin, la littérature abondante sur ce modèle permet d'éviter certains écueils et permet à ce modèle d'évoluer.

2.2.2 Inconvénients

Les critiques relatives au modèle de Wilkie portent essentiellement sur la spécification et la qualité de l'adéquation du modèle. Nous nous limiterons à rappeler les principaux problèmes liés à l'utilisation du modèle de Wilkie, le lecteur pouvant trouver un détail précis des inconvénients dans [3].

Plusieurs actuaires anglais ont fortement critiqué le modèle : citons notamment Kitts (1990), Huber (1995) et Geoghegan et al. (1992). Les critiques portent principalement sur deux éléments :

La spécification du modèle et la dépendance aux paramètres initiaux

Huber (1995) souligne la surparamétrisation du modèle de Wilkie qui implique une forte dépendance des résultats aux paramètres de calibrage initiaux. Les processus comportant un retour à la moyenne et le choix de la moyenne conditionnent les résultats futurs et les projections. Ainsi, les choix de spécifications doivent être faits avec prudence.

Les biais liés à l'estimation

Kitts (1990) montre que la série d'inflation, initialement utilisée par Wilkie, n'est pas stationnaire. De ce fait, il critique le choix de modéliser la série d'inflation par un processus autorégressif d'ordre un.

Geoghegan et al. (1992) montrent que les résidus du modèle ne respectent pas les hypothèses de normalité, d'indépendance et de variance constante. La distribution des

résidus est asymétrique et leptokurtique et les résidus présentent des périodes de forte volatilité, alternant avec des périodes de faible volatilité.

Enfin, Huber(1995) montre que le modèle de Wilkie est biaisé par quelques données aberrantes correspondant à des années de forts chocs inflationnistes en Grande Bretagne (1920, 1940 et 1974). Si ces années sont exclues de la régression, le coefficient de corrélation entre rendement des dividendes et inflation n'est plus significatif.

Des critiques théoriques du modèle sont également prononcées. En effet, le modèle de Wilkie est avant tout empirique et deux hypothèses centrales, au XX^{ème} siècle, de la théorie économique ne sont pas respectées : l'efficacité des marchés financiers et l'absence d'opportunité d'arbitrage. Cependant, il s'avère que ces deux critiques sont remises en cause en période de crises et que la plupart des modèles de gestion d'actifs ne respectent pas ces deux principes.

Les améliorations que l'on va apporter au GSE consiste à répondre aux différentes critiques pour améliorer le modèle initial. Nous sommes conscients que d'autres modèles étaient envisageables mais nous avons choisi le modèle de Wilkie qui constitue selon nous, un bon arbitrage entre complexité statistique et rapidité de simulation. Rappelons en outre que la complexité des extensions statistiques améliore l'adéquation au passé sans nécessairement améliorer la qualité de la prévision.

Par ailleurs, dans la mesure où nous souhaitons réaliser des projections sur un an ou cinq ans, l'utilisation de ce modèles est tout à fait justifiable. Avant de présenter les outils théoriques utilisés dans ce mémoire, nous présentons, dans le paragraphe suivant, les autres modèles que l'on aurait pu choisir.

2.2.3 Autres modèles

Certains actuaires proposent des modèles alternatifs intégrant les dernières avancées des méthodes de séries temporelles. Pour cette partie, nous vous invitons à consulter le mémoire de Charles Keita [12] qui a réalisé une comparaison des modèles qui existent actuellement. Avec son autorisation, nous nous sommes inspirés du tableau de synthèse présent dans son mémoire. Les tableaux suivants récapitulent les modèles existant, leurs avantages et leurs inconvénients :

Modèle	Type de modèle	Modèle - inflation	Modèle - Taux	Modèle - action	Avantages	Inconvénients
Modèle de Wilkie	ALM, en cascade	AR(1)	Déduit de l'inflation	Déduit de l'inflation	Liens entre les variables, améliorations possibles	Paramètres figés, modèle inflation, beaucoup de paramètres
Modèle de Kaufman, Klett, Gadamer adapté	ALM, en cascade	Déduit des taux court-terme	HJM	Black et Scholes	Simplicité, liens entre les variables, modèle taux à 2 facteurs	Paramètres figés, modélisation action
Modèle de Jarrow-Yildirim	ALM, en cascade	Taux de change entre taux réels et taux nominaux	HJM		Robuste, modèle moins proche du déterministe, littérature abondante	Paramètres figés, modélisation des taux, calibrage de trois browniens corrélés, pas de modélisation des autres actifs
Modèle de l'AAA	ALM et lien par corrélation	Pas de modélisation	Modèle d'équilibre à 3 facteurs et volatilité stochastique	Modèle Black et Scholes à volatilité stochastique	Modélisation action et processus adaptés pour les taux	Pas d'inflation, modèle à volatilité stochastique très complexe, aucune cohérence

TABLE 2.1 – Comparaison des modèles (1/3)

Modèle	Type de modèle	Modèle - inflation	Modèle - Taux	Modèle - action	Avantages	Inconvénients
Modèle de la SOA-CAS	Modèle spécifique inflation et modèle de marché	Ornstein-Ulhenbeck	Hull et White à deux facteurs	Log normale à 2 régimes	Processus spécifiques adaptés à chaque actif, cohérence assurée pour les taux et l'inflation, modèle de taux, modélisation action non-gaussienne	Pas de cohérence avec les actions, Ornstein-Ulhenbeck équivalent à du déterministe sur le long terme, paramètres figés
Modèle de Mercuro, Benhamou et Koelher	Modèle de marché	Analogie avec le modèle LMM	Non pris en compte	Non pris en compte	Autorise toutes les déformations de la courbe inflation, robuste, éloigné du déterministe	Calibrage sur données gré à gré, étude sous la probabilité risque neutre, pas de modélisation action et immobilier
Modèle VAR	ALM et Modèle multi-varié	ARMA	ARMA	ARMA	Simplicité, cohérence, neutralité, outils propres aux séries temporelles	Beaucoup de paramètres, paramètres figés, modélisation gaussienne

TABLE 2.2 – Comparaison des modèles (2/3)

Modèle	Type de modèle	Modèle - inflation	Modèle - Taux	Modèle - action	Avantages	Inconvénients
Modèle VECM	ALM et Modèle multi-varié	ARMA	ARMA	ARMA	Evolution des cor-réactions dans le temps, littérature abondante, outils propres aux séries temporelles	Beaucoup de paramètres, difficulté d'estimation des probabilités de transition
Modèle TAR	Modèle multi-varié à changement de régime	ARMA	ARMA	ARMA	Traduit des sauts de marché, distribution non normale pour les actifs, outils propres aux séries temporelles, économiquement réaliste	Beaucoup de paramètres, paramètres figés, pas de changement de régime
Modèle SETAR	Modèle multi-varié à changement de régime	ARMA	ARMA	ARMA	Distribution normale pour les actifs, outils propres aux séries temporelles, économiquement réaliste	Beaucoup de paramètres, difficulté d'estimation des probabilités de transition
Modèle STAR	Modèle multi-varié à changement de régime	ARMA	ARMA	ARMA	Transitions lisses, outils propres aux séries temporelles	Beaucoup de paramètres, difficulté d'estimation des probabilités de transition

TABLE 2.3 – Comparaison des modèles (3/3)

Chapitre 3

Les séries temporelles : modèles ARMA, ARIMA et SARIMA

Ce chapitre a pour objectif d'introduire les notions qui vont être ultérieurement utilisées. Pour construire un générateur de scénarios, le recours aux séries temporelles est indispensable. En effet, nous disposons de séries de nombres qu'il convient d'analyser afin de définir la modélisation la plus adaptée à notre utilisation. Ainsi, la modélisation retenue pour mettre au point des prévisions sera différente de celle utilisée si l'on souhaite modéliser seulement. Par ailleurs, le principe de parcimonie, visant à éviter un surparamétrage du modèle retenu, doit être adopté.

Dans une première partie, nous rappellerons la définition d'un processus autorégressif et à moyenne mobile ainsi que les méthodes existantes pour estimer les paramètres d'un tel modèle.

Dans une seconde partie, nous introduirons la notion de processus sur les séries non stationnaires. Nous invitons le lecteur à consulter attentivement cette partie car nous avons utilisé ces méthodes pour modéliser l'inflation, comme nous le verrons dans le chapitre six. Dans une troisième partie, nous présenterons les tests utilisés lors de nos travaux pour vérifier la pertinence de notre modélisation. Lors de la modélisation d'une série temporelle, de nombreux tests sont indispensables pour vérifier la qualité de la modélisation : ainsi nous présenterons successivement les tests de stationnarité, d'indépendance et de normalité.

Dans une quatrième partie, nous évoquerons la méthodologie de Box et Jenkins, très efficace pour choisir, en fonction d'un critère de choix, le modèle le plus adapté pour modéliser une série temporelle.

3.1 Processus autorégressif et à moyenne mobile

3.1.1 Définitions

Nous rappelons ici brièvement les définitions de différents processus temporels : un processus autorégressif, un processus à moyenne mobile et enfin un processus autorégressif et à moyenne mobile, que l'on abrège avec le sigle ARMA. Nous introduisons des notations que l'on utilisera dans la suite du mémoire. Ces processus jouent un très grand rôle dans la modélisation statistique des séries temporelles. Ils ont été popularisés par Box and Jenkins (1976) qui ont développé une démarche claire permettant de déterminer les paramètres

d'un processus. Ils permettent une représentation relativement parcimonieuse des séries temporelles stationnaires. Avant de présenter globalement un processus ARMA, arrêtons-nous sur ses deux composantes en définissant les processus autorégressifs (AR) puis les processus à moyenne mobile (MA).

Processus autorégressif Ces processus jouent un rôle important dans les applications car ils peuvent être extrapolés linéairement à partir de p valeurs précédentes, à un bruit blanc près. Mathématiquement, cela se traduit par une expression de la forme suivante :

$$X_t = \sum_{i=1}^p a_i X_{t-i} + \epsilon_t$$

En utilisant l'opérateur retard B , ce processus s'écrit de la manière suivante :

$$A(B)X_t = \epsilon_t$$

où A est le polynôme de degré p possédant les coefficients a_i définis ci-dessus.

Processus moyenne mobile Il s'agit des processus vérifiant l'équation suivante :

$$X_t = \sum_{i=1}^q b_i \epsilon_{t-i} = \Theta(B)\epsilon_t$$

où Θ est un polynôme de degré q dont les coefficients sont les b_i . Dans un tel modèle, on suppose que l'influence des chocs passés se manifeste au travers d'une fonction linéaire. L'appellation de ce modèle, moyenne mobile d'ordre q , est due à X_t , moyenne mobile appliquée aux variables aléatoires ϵ_t . Le terme « moyenne » doit être pris dans un sens très large dans la mesure où la somme des coefficients n'est pas nécessairement égale à un.

Processus ARMA Supposons que X est un processus stationnaire au second ordre et centré. Le lecteur consultera le paragraphe sur les séries stationnaires pour trouver une définition précise de ces notions. Un processus ARMA(p,q) s'écrit de la manière suivante :

$$\sum_{i=1}^p a_i X_{t-i} = \sum_{i=1}^q b_i \epsilon_{t-i} \text{ i.e}$$

$$A(B)X_t = \Theta(B)\epsilon_t$$

Afin de déterminer les degrés optimaux des polynômes A et B , Box et Jenkins ont proposé une méthodologie. Nous la décrivons dans la dernière partie de ce chapitre.

3.1.2 Estimation

En général, pour estimer les paramètres d'un modèle, il est naturel de chercher à calculer l'estimateur du maximum de vraisemblance. Cette dernière est toutefois difficile à calculer. Si on suppose que le bruit blanc est gaussien, il s'agit en effet de calculer :

$$L_{\xi, \sigma^2} = \frac{1}{(2\pi)^{\frac{n}{2}} (\det R_n)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} x' R_n^{-1} x\right)$$

où $\xi = (a_1, a_2, \dots, a_p, b_1, \dots, b_q)$ est le vecteur des paramètres, R_n est la matrice de Toeplitz à l'ordre n et $x = (x_1, \dots, x_n)$ sont les valeurs observées de la série ; n est donc le nombre d'observations, qui peut être très grand. Il s'agit de calculer les valeurs de ξ et σ^2 qui rendent la quantité ci-dessus maximale. Cette opération est délicate eu égard au calcul de l'inverse de la matrice de Toeplitz.

Plusieurs méthodes numériques, notamment de type récursif, ont été développées pour mener à bien le calcul de l'estimateur et ont donné lieu à une abondante littérature en lien avec les séries chronologiques. Par ailleurs, la plupart des logiciels statistiques possèdent, déjà programmé, ce genre de calcul, c'est pourquoi nous ne les décrivons pas ici.

Pour un processus autorégressif, on montre que la méthode de Yule-Walker est la plus performante. Notons qu'il est également possible d'utiliser la méthode des moindres carrés conditionnels.

3.2 Processus non stationnaires

Comme nous l'avons indiqué dans le paragraphe précédent, la modélisation par un ARMA nécessite que la série à modéliser soit stationnaire. Nous présentons ci-dessous deux modèles pour les séries non stationnaires.

3.2.1 Définitions

Tout d'abord, rappelons les définitions de la stationnarité au sens fort et au sens faible.

Stationnarité au sens fort

Une série temporelle, à valeurs réelles et en temps discret $(X_t)_{t \in \mathbb{N}}$ est dite stationnaire au sens fort si pour toute fonction f mesurable, on a :

$$\forall (t, k, n) \in \mathbb{N}^3, f(X_t, \dots, X_{t+n}) \text{ et } f(X_{t+k}, \dots, X_{t+k+n}) \text{ ont même loi.}$$

Stationnarité au sens faible (ou au second ordre)

Soit un processus temporel à valeurs réelles et en temps discret $(X_t)_{t \in \mathbb{N}}$. Il est dit stationnaire au sens faible (ou de second ordre) s'il vérifie les trois propriétés suivantes :

$$\left\{ \begin{array}{ll} \mathbb{E}[X_t] = \mu < \infty & \forall t \in \mathbb{N} \\ \mathbb{E}[X_t^2] < \infty & \forall t \in \mathbb{N} \\ Cov[X_k, X_{k+t}] = Cov[X_0, X_t] & \forall (t, k) \in \mathbb{N}^2 \end{array} \right.$$

Nous allons présenter à présent deux modélisations pour les processus non stationnaires.

3.2.2 Modèle ARIMA

Le terme ARIMA est un terme anglais signifiant *AutoRegressive Integrated Moving Average*. Un processus X est un processus ARIMA(p, d, q), s'il existe un entier naturel d

tel que le processus $Y_t = (1 - B)^d X_t$ est un ARMA(p,q). Autrement dit, il existe deux polynômes $A(z)$ et $C(z)$ de degré respectif p et q et un bruit tels que :

$$(1 - B)^d A(B)X_t = C(B)\epsilon_t$$

En réalité l'opérateur $\Delta = (1 - B)$ permet de stationnariser la série en éliminant les tendances. Le paramètre d est défini en réalisant des tests de stationnarité. En général, il est égal à un ou deux.

Cependant, certaines séries ne sont pas modélisables de façon parcimonieuse avec un tel processus : il s'agit des séries possédant une saisonnalité. En effet, il est possible de capter l'autocorrélation saisonnière, en considérant un ARIMA avec des ordres élevés (de l'ordre de douze par exemple pour une saisonnalité annuelle si l'on considère des données mensuelles), mais cela nécessite l'estimation de toutes les autocorrélations intermédiaires qui ne sont parfois pas significatives. Ainsi, la modélisation obtenue est très biaisée et souffrira vraisemblablement d'un problème de surparamétrisation.

De ce fait, généralement, le recours à une structure plus générale, incluant stationnarité, tendance et saisonnalité, est requis : il s'agit des modèles SARIMA. Le paragraphe suivant présente ces modèles permettant de capter la saisonnalité tout en restant fidèle au principe de parcimonie.

3.2.3 Modèle SARIMA

Le terme SARIMA est un terme signifiant *Seasonnal AutoRegressive Integrated Moving Average*. On suppose que la série satisfait l'équation :

$$(1 - B)^d(1 - B^s)^D A(B)\phi(B^s)X_t = C(B)\Theta(B^s)\epsilon_t$$

Où p, P, d, s, D, q, Q sont des entiers naturels inconnus. Les polynômes $A(z)$, $\phi(z)$, $C(z)$ et $\Theta(z)$ sont à coefficients inconnus.

Chaque terme à une signification précise :

- d : différence. Ce coefficient est le même que celui de l'ARIMA, il correspond au nombre de différenciations nécessaires pour que la série à modéliser devienne stationnaire.
- s : saison. Permet de capter la saisonnalité du processus.
- D : différence saisonnière.
- $A(z)$ et $\phi(z)$ correspondent à la partie autorégressive du processus respectivement pour la partie non saisonnière et saisonnière.
- $C(z)$ et $\Theta(z)$ correspondent à la partie moyenne mobile du processus respectivement pour la partie non saisonnière et saisonnière.

La modélisation d'une série temporelle par un tel processus s'avère parfois judicieux. Nous verrons dans le chapitre six que pour modéliser l'inflation, nous avons considéré un processus de cette forme.

À présent, nous présentons les tests dont nous nous sommes servis pour calibrer les équations.

3.3 Les tests

Tests de stationnarité

Comme nous l'avons évoqué ci-dessus, un des points fondamentaux concernant les séries temporelles est de savoir si celles-ci sont stationnaires ou non. Si elles ne le sont pas, les séries ne sont pas modélisables par un processus ARMA, sauf si elles sont stationnarisées en amont de la modélisation.

Ainsi, avant toute modélisation, il convient de tester la stationnarité de la série. Différents tests sont à notre disposition, nous allons évoquer brièvement les tests utilisés. Une description plus précise est réalisée en annexe.

Tout d'abord, il existe une catégorie de tests appelée les tests de racine unitaire. Ils permettent de déceler la présence d'une racine supérieure ou égale à l'unité dans le polynôme caractéristique du processus. Il s'agit des tests de Dickey-Fuller et de Dickey-Fuller augmenté.

Par ailleurs, le test de Philipps Perrond peut être également utilisé : en effet, Phillips et Perrond proposent une correction au test de Dickey-Fuller simple afin de régler le problème de l'autocorrélation et de l'hétéroscédasticité des erreurs.

Enfin, nous avons également utilisé le test de Kwiatkowski, Phillips, Schmidt et Shin (KPSS). Contrairement aux tests précédents, l'hypothèse nulle de ce test est celle de la stationnarité.

Tests d'indépendance

Après avoir choisi un type de modélisation, il convient de s'assurer que les résidus obtenus sont indépendants. A cette fin, il existe plusieurs tests : Ceux généralement utilisés sont les tests du type *portemanteau*, mais il existe également le test de Breusch-Godfrey qui s'avère efficace pour notre problématique. Ce test étant peu connu, nous en proposons une description plus précise.

- Les tests de type *portemanteau*

Afin de tester l'indépendance des résidus issus d'une modélisation, les *tests du portemanteau* sont couramment utilisés : il s'agit des tests de Ljung-Box et de Box-Pierce. Ces tests visent à évaluer la corrélation entre les résidus et se fondent sur les estimateurs classiques de la corrélation et de la variance.

Une remarque nous semble cependant nécessaire. En effet, nous lisons dans la littérature que les tests de type *portemanteau* testent l'indépendance des résidus. Ceci n'est pas tout à fait exact : en effet, ils ne testent pas l'indépendance au sens strict du terme de la série mais celle de non corrélation jusqu'à un certain décalage dans le temps. Un autre test peu connu mais très puissant pour tester la blancheur des résidus est le test de Breusch-Godfrey que nous présentons au paragraphe suivant.

- Le test de Breusch-Godfrey

Le test de Breusch-Godfrey, connu aussi sous le nom de test ML (pour multiplicateur

de Lagrange) permet d'éviter certains écueils des tests de type *portemanteau* ou du test de Durbin Watson sur l'autocorrélation. Les statisticiens Breusch et Godfrey ont proposé un test général en ce sens qu'il prend en compte :

- des régresseurs non stochastiques tels que les valeurs décalées de la variable indépendante
- des systèmes autorégressifs d'ordre plus élevé tels que AR(1), AR(2), etc.
- des moyennes simples, des moyennes mobiles ou encore des termes d'erreur sous forme de bruit blanc.

Nous ne rentrerons pas dans les détails mathématiques très précis mais nous tenons à expliquer les principes généraux de ce test.

Supposons que le terme d'erreur suive un processus autorégressif du $p^{\text{ème}}$ ordre, défini comme suit :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t$$

où ϵ_t est un terme d'erreur de type bruit blanc. L'hypothèse qu'il faut tester est la suivante :

$$\mathcal{H} : \rho_0 = \rho_1 = \dots = \rho_p = 0$$

Cette expression indique qu'il n'y a pas d'autocorrélation de quelque ordre que ce soit. Le test de Breusch-Godfrey comprend les étapes suivantes :

- Obtenir les valeurs estimées de \hat{u}_t
- Régresser \hat{u}_t sur les X_t d'origine.
- Calculer le R^2 à partir de la dernière régression.

Si la taille de l'échantillon est grande, Breusch et Godfrey ont montré que $(n-p)R^2$ suit une loi du khi deux à p degrés de liberté. Si $(n-p) \times R^2$ est supérieur à la valeur critique du Khi deux pour le niveau de signification choisi, nous rejetons l'hypothèse nulle, ce qui signifie qu'au moins un paramètre est statistiquement et significativement différent de zéro.

Ce test étant plus adapté à notre problème, nous le privilégierons dans la suite de notre exposé.

Tests de normalité

La normalité est un point important de ce mémoire. En effet, nous avons besoin de tester la normalité des résidus des équations du modèle. Ces tests peuvent également être utiles pour étudier la loi que suivent nos observations simulées.

Afin de tester la normalité d'une série de nombres, il existe plusieurs tests que nous présentons en annexe. Notons que l'hypothèse \mathbb{H}_0 sera systématiquement l'adéquation de notre échantillon à une loi normale.

3.4 La méthodologie de Box et Jenkins

La méthodologie de Box et Jenkins donne une méthode robuste permettant de déterminer la modélisation la plus adaptée à la série que l'on souhaite modéliser. Dans cette section, nous nous contenterons de présenter l'aspect théorique de la méthodologie. Nous avons appliqué cette méthode pour déterminer la modélisation la plus adaptée de l'infla-

tion car la modélisation proposée par Wilkie était inadaptée à nos données. Pour prendre connaissance des résultats obtenus et de la démarche suivie afin de proposer une modélisation de l'inflation par un processus SARIMA, le lecteur se reportera au chapitre six qui concerne la calibration du modèle.

Avant de présenter le principe de la méthodologie de Box et Jenkins, il nous semble nécessaire de rappeler six principes fondamentaux à retenir lors d'une modélisation :

- Si le nombre de retards considérés est trop faible, le modèle obtenu aura un biais de spécification.
- Si le nombre de retards considérés est trop élevé, le modèle n'aura pas été estimé de façon aussi efficace que possible.
- Un modèle simple est toujours préférable à un modèle plus compliqué : on appliquera donc le principe dit de parcimonie ou principe du *rasoir d'Ockham*.
- Dans les modèles AR(p) ou MA(q), il est préférable d'avoir un polynôme continu. Par exemple, si $p=4$, on inclut tous les retards de un à quatre sans exception.
- Dans la plupart des cas, il est nécessaire de fixer un retard maximal raisonnable. En effet, si par le calcul, on trouve des paramètres trop élevés, par exemple $p = 12$, $q = 14$, le modèle obtenu sera surparamétré.
- Dans tous les cas, le modèle final retenu devra avoir des résidus bruits blancs. Toutefois, il se pourrait que les coefficients estimés ne soient pas toujours significatifs aux seuils habituels. L'idée fondamentale ici est de trouver les paramètres optimaux et non pas d'estimer des coefficients significatifs.

Stationnarisation et désaisonnalisation de la série

Une première étape consiste à appliquer les tests de stationnarité évoqués à la partie précédente. Dans le cas où la série n'est pas stationnaire, il est possible de la différencier. La plupart du temps, pour obtenir une série stationnaire il suffit de différencier à l'ordre un ou à l'ordre deux.

Ensuite, il convient de désaisonnaliser la série différenciée. Pour cela, Box et Jenkins proposent de tracer la fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF) de la série stationnarisée. La présence éventuelle d'un ou plusieurs pics significatifs nous permet de déterminer s'il faut considérer un paramètre saisonnier.

Détermination des retards maximaux

Par ailleurs, l'ACF et la PACF nous permettent de déterminer le nombre de retard maximaux respectivement de la partie moyenne mobile et de la partie autorégressive. Pour cela, il suffit d'appliquer la règle suivante sur la série stationnaire et désaisonnalisée :

- Le nombre de retard maximal de la partie moyenne mobile est le dernier pic significatif de l'ACF.
- Le nombre de retard maximal de la partie autorégressive est le dernier pic significatif de la PACF.

Détermination des paramètres optimaux

L'étape suivante consiste à sélectionner un critère de choix afin de déterminer les paramètres optimaux. Pour cela, nous avons le choix entre différents critères : le critère du

maximum de vraisemblance, le critère d'Akaiké, le critère de Schwarz, ou encore les critères de Rissanen ou Hannan et Quinn.

- Le critère du maximum de vraisemblance

La méthode du rapport de vraisemblance est bien connue. Il s'agit de formuler une série d'hypothèses à tester. Le lecteur trouvera une description plus détaillée de ce critère en annexe.

- Les critères bayésiens de Schwarz et d'Akaiké

Nous avons retenu ces critères dans notre mémoire, c'est la raison pour laquelle nous souhaitons les expliquer davantage.

L'approche adoptée est différente de celle du critère précédent. Le principe de base de ces critères est que si l'on compare simplement le maximum de vraisemblance atteint par chaque modèle, alors il est logique que les modèles ayant le plus de paramètres soient les mieux adaptés aux données. En effet, un modèle qui contient beaucoup de paramètres aura nécessairement un très bon ajustement à l'intérieur de l'échantillon (faible variance des résidus notamment) mais pourrait être moins performant au niveau prévisionnel. D'ailleurs, il a souvent été constaté qu'un modèle contenant beaucoup de paramètres était par construction plus enclin à expliquer tous les phénomènes observés durant la période étudiée même si ces mêmes phénomènes étaient le plus souvent très spécifiques et peu susceptibles de se reproduire dans le futur. On appelle parfois ce phénomène l'« illusion d'un R^2 élevé » ou de surajustement.

Une remarque qui découle de ce qu'on a dit ci-dessus est que si un modèle est un cas particulier d'un autre modèle comme c'est le cas pour les modèles AR par exemple, alors le modèle avec le plus de paramètres aura indéniablement un maximum de vraisemblance plus élevé ; et ce, même si le « vrai modèle » est celui qui a le moins de paramètres.

Les critères d'Akaike et de Schwarz permettent de pénaliser les modèles en fonction du nombre de paramètres. Ils permettent de choisir les paramètres optimaux en réalisant un compromis entre le *fit* (qui mesure l'adéquation aux données et qui s'améliore en ajoutant des paramètres) et la parcimonie (qui donne plus de poids aux modèles avec moins de facteurs).

Akaike et Schwarz ont montré que lorsque la taille de l'échantillon tend vers l'infini, minimiser les critères suivants permet de trouver les paramètres appropriés :

$$Akaike(h) = \ln(\sigma_h^2) + \frac{2h}{T}$$
$$Schwarz(h) = \ln(\sigma_h^2) + \frac{h \times \ln(T)}{T}$$

Où T est le nombre d'observations, σ_h est l'écart type des résidus et h est le nombre de paramètres testés.

Le premier terme correspond à l'ajustement du modèle tandis que le deuxième est une pénalité qui dépend du nombre de paramètres inclus. Ainsi quand on ajoute un paramètre, deux effets interviennent :

- La variance estimée diminue (premier terme),

- La pénalité augmente au rythme $2/T$ pour Akaike et $\ln T/T > 2/T$ pour Schwarz qui est plus pénalisant.

L'impact total dépendra de l'importance relative des deux termes.

Nous ne présentons pas les autres critères envisageables dans la mesure où nous ne les avons pas utilisés dans notre mémoire.

Tests sur les résidus

Après avoir choisi un des critères proposés ci-dessus (nous rappelons que la liste n'est pas exhaustive) et avoir déterminé les paramètres optimaux en fonction du critère de choix, il convient enfin de vérifier que les résidus sont des bruits blancs. Pour cela, nous avons proposé des tests dans la partie précédente.

Dans le cadre de notre modélisation, il est intéressant de tester si les résidus sont normaux ou pas. A priori, il n'y a pas de raison qu'ils le soient.

Ainsi, nous avons présenté dans ce chapitre les fondements des séries temporelles en introduisant des notations que nous utiliserons par la suite. Nous avons décrit une méthode de détermination des paramètres que nous appliquerons afin de déterminer une modélisation de l'équation de l'inflation, base du générateur de scénarios que l'on souhaite réaliser.

Chapitre 4

Intégration de sauts dans une série temporelle

L'originalité de ce mémoire est le fait de considérer le contexte économique actuel et ainsi de prévoir la possibilité de prendre en compte des crises liées à des événements de nature exogène. Le début du XXI^{ème} siècle ayant vu croître considérablement les risques de crises ainsi que leur probabilité d'occurrence, il nous semble inenvisageable de construire des modèles sans pouvoir les prendre en compte.

Ce chapitre vise à présenter les différentes méthodes testées pour intégrer des crises dans le générateur de scénarios économiques et à justifier le choix de l'utilisation de l'une d'entre elles.

Après avoir fait un bref retour sur la situation politico-économique mondiale depuis le début du XXI^{ème} siècle afin de justifier l'intégration d'un processus de prise en compte des crises dans le générateur, nous présenterons les différentes méthodes d'intégration de sauts dans un processus de séries temporelles et expliquerons leurs avantages et leurs inconvénients.

4.1 Le contexte politico-économique mondial

Il est incontestable d'affirmer que le contexte politico-économique mondial est actuellement bouleversé. En effet, depuis les années 2000, de nombreuses crises, de plusieurs types (financières, politiques...), ont perturbé les systèmes financiers et politiques. Notons que les crises se sont succédé, depuis le début du XXI^{ème} siècle, à une fréquence élevée et inédite.

Plusieurs crises financières ont perturbé les marchés financiers. En premier lieu, trois bulles spéculatives ont éclaté : celle des valeurs technologiques en 2000 à la suite au développement d'internet et, en 2007, les bulles de l'immobilier et du crédit venues des États-Unis. En second lieu, la crise des dettes souveraines et la crise européenne ont perturbé les marchés financiers à partir de 2010.

La stabilité mondiale a été affectée également par plusieurs conflits militaires persistants ou débutants : le conflit israélo-palestinien, les guerres en Afghanistan et en Irak, les émeutes du *printemps arabe* en 2010, les révoltes au Sénégal ainsi qu'en Syrie.

Participent également à cette instabilité, les mouvements de révolte en Russie et la situation économique de la Chine. En effet, la part importante et croissante du budget chinois consacrée à la défense est un facteur à prendre en compte : la Chine prévoit une augmentation annuelle de son budget de 19% jusqu'en 2015 afin d'atteindre un budget total de 240 milliards de dollars soit quatre fois plus que le budget de la défense du Japon qui est le deuxième en Asie. Cet élément tend à prouver que la situation asiatique n'est pas pacifique.

Enfin, les crises multiples peuvent éclater en raison du déséquilibre économique existant entre les pays émergents, notamment la Chine, et les pays développés. Il n'y a aucune raison pour que le PIB par habitant de la Chine soit inférieur à celui du Japon, ce qui implique une multiplication par quatre du PIB chinois dans les prochaines années. En effet, le PIB par habitant en Chine est de 20% alors qu'il est de l'ordre de 85% au Japon. Sous la pression de la concurrence asiatique et de quelques autres pays, l'Europe va nécessairement perdre des parts de marché au profit de la Chine et continuer son processus de désindustrialisation. Des crises d'adaptation seront obligatoires et c'est pourquoi des scénarios multiples et variés sont à prévoir. Olivier Blanchard, conseiller de Christine Lagarde au fonds monétaire international, a même récemment évoqué la possibilité de grande dépression.

Pour conclure, citons un dernier argument économique corroborant le fait que des crises sont à prévoir et qu'il est donc nécessaire d'inclure des sauts dans le générateur. L'émergence de la Chine a créé une modification de l'évolution de l'économie mondiale. En effet, la puissance démographique de la Chine crée un déséquilibre qui perturbe les théories économiques. Ainsi, par exemple, en temps normal, l'économie suit un cycle long d'environ cinquante six ans, dit *cycle de Kondratieff*. Or, ce cycle long est actuellement perturbé. Pour appuyer ce propos, rappelons à l'appui de l'ouvrage de François-Xavier Chevalier [6], les cinq saisons d'un cycle long :

- **Le printemps** : boom inflationniste, le cycle des quantités est haussier.
- **L'été** : il s'agit d'un régime de croisière dans lequel l'économie se porte bien.
- **La stagflation** : croissance en baisse et hyperinflation.
- **L'été indien** : période de prospérité factice idéale pour la création d'une bulle spéculative. L'inflation est en baisse, de même que les taux ; la croissance est morose. De ce fait, les marchés financiers sont puissants. A la fin de cet été indien, un krach est provoqué par un endettement devenu trop important.
- **La dépression** : suite à l'explosion d'une bulle, les prix des actifs diminuent et les banques ont des problèmes de solvabilité.

D'après l'étude de François-Xavier Chevalier, nous devrions être au début d'un cycle long, dans la période *Été* mais ce n'est pas le cas. En effet, la croissance diminue alors qu'on devrait observer une hausse...

Ainsi, comment ne pas introduire dans notre GSE la possibilité d'intégrer des scénarios de crises dans un tel contexte instable ? Nous développons dans les paragraphes suivants,

les méthodes étudiées pour introduire des sauts dans les équations.

4.2 Intégration de sauts en utilisant des lois uniforme ou de poisson

La littérature concernant l'introduction de sauts dans une série temporelle n'est pas très abondante. En effet, ces techniques n'ont pas encore été trop étudiées, notamment pour les séries financières. Ainsi, dans un premier temps, nous avons testé des lois simples comme les lois de Poisson.

4.2.1 Modification des paramètres des résidus

Une méthode permettant d'introduire des sauts dans une série temporelle consiste à modifier les paramètres des résidus après avoir simulé une loi uniforme. La méthode consiste à simuler une loi uniforme $[0;100]$ et à modifier les paramètres des résidus en fonction de la valeur simulée. Si la valeur simulée est inférieure à la probabilité que l'on souhaite donner au scénario (par exemple si la valeur simulée est inférieure à 5), on modifie les paramètres des résidus.

Cette méthode a été proposée dans [19] et appliquée à un processus autorégressif d'ordre un. Dans ce cas, les résidus ne sont constitués que d'une variable aléatoire, gaussienne par hypothèse. Nous avons appliqué cette méthode à notre modèle mais elle nous semble peu précise pour deux raisons :

- la partie *moyenne mobile* des processus auxquels nous souhaitons intégrer des crises ne sont pas aussi simples que celle d'un processus autorégressif d'ordre un. En effet, par exemple, nous verrons ultérieurement que nous ne retiendrons pas un processus AR(1) pour modéliser l'inflation. Ainsi, la méthode consistant à modifier les paramètres de la partie *moyenne mobile* semble inappropriée.
- cette méthode, même dans le cas d'une modélisation par un processus autorégressif d'ordre un, nous semble inadéquate. En effet, un éventuel choc se traduit, dans la méthode proposée, par une modification de l'espérance et de la variance de la variable aléatoire gaussienne. Ainsi, les résidus obtenus peuvent perdre leur caractère « bruits blancs ». Par ailleurs, comment modifier la variance de manière efficace ?

Pour ces deux raisons, nous n'avons pas retenu cette méthode.

4.2.2 Utilisation de la loi de Poisson

Naturellement, l'utilisation d'un processus de Poisson, appelé également processus des *événements rares*, paraît adapté à la modélisation de crises, événement rare par hypothèse.

Les applications relatives au processus de Poisson ont longtemps été limitées aux événements tels que les files d'attente ou les arrivées de bateaux dans un port. L'étude de Ladislaus Bortkiewicz concernant les accidents dus aux coups de pied de cheval dans les armées est un exemple célèbre d'application de ce processus.

Cependant, depuis une trentaine d'années, son champ d'application s'est considérablement élargi, ce qui nous amène à penser qu'il est possible de l'adapter dans notre cas. En effet,

cette loi est maintenant utilisée dans beaucoup de domaines : les télécommunications, le contrôle de qualité statistique, la description de certains phénomènes liés à la désintégration radioactive, la biologie, la météorologie... Par ailleurs, elle est parfois utilisée en finances, pour modéliser la probabilité de défaut d'un crédit.

De ce fait, nous avons étudié la possibilité d'ajouter un processus de Poisson simple dans l'équation de l'inflation afin de modéliser une crise.

Un souci majeur que l'on a rapidement rencontré est que la prise en compte de ce saut translate l'ensemble de la courbe et ne prévoit pas de retour à l'équilibre. Ainsi, une baisse de l'inflation, par exemple, n'est jamais véritablement compensée, ce qui ne correspond pas à ce qu'il se passe réellement. En effet, ordinairement, les crises durent, le temps nécessaire à la correction des excès spéculatifs et on observe un retour à l'équilibre après effacement de ces excès. Le temps d'effacement est plus ou moins long en fonction de la crise subie. Par ailleurs, si l'on insère des sauts à l'intérieur du processus, cela implique que les valeurs prédites par le processus temporel choisi prennent en compte les éventuels sauts. La prise en compte de ces sauts, issus de scénarios non présents dans les données, rend la modélisation biaisée.

En conclusion, si l'intégration d'un processus de Poisson simple dans notre cas est envisageable, nous avons cherché à l'affiner pour pouvoir prendre en compte le retour à l'équilibre. Ceci fait l'objet de la prochaine partie.

4.3 L'algorithme de Box et Tiao (1975)

L'algorithme de Box et Tiao permet de prendre en compte des perturbations survenant sur des séries chronologiques. Lors de la modélisation d'une série, l'analyse d'intervention de Box et Tiao prend en compte les interventions extérieures et fournit une mesure de l'impact de celles-ci sur la série, comme Ferrara et Guegan le décrivent dans [15] à l'appui de données de la RATP.

Dans notre cas, la date des perturbations, l'ampleur de la modification et le temps de retour à l'équilibre ne sont pas connus. Cependant, cette méthode peut être utilisée en attribuant une probabilité à ces deux quantités. C'est ce que nous proposons dans la suite.

Box et Tiao (1975) sont à l'origine de la théorie dite de l'*analyse d'intervention* qui permet de prendre en compte différentes interventions extérieures au modèle lors de la modélisation d'une série chronologique. L'apport de l'analyse d'intervention à une modélisation de série de type SARIMA se situe principalement au niveau de l'information disponible à l'utilisateur pour modéliser cette série. En effet, l'approche de Box et Jenkins présentée au chapitre précédent utilise simplement l'information quantitative contenue dans les données alors que l'analyse d'intervention permet d'ajouter de manière additive une information de type qualitatif.

Le but de cette partie est de mettre en évidence l'intérêt de la théorie de l'analyse d'intervention pour la simulation de sauts dans une série temporelle. Nous présentons une

utilisation de l'analyse d'intervention différente de celle proposée initialement par Box et Tiao. Nous précisons les différents types de processus que l'on peut envisager, nous les illustrons à partir de simulations, puis nous appliquons cette méthodologie aux données réelles.

Pour avoir plus de détails sur l'analyse d'intervention, le lecteur pourra consulter différents travaux réalisés pour compléter les travaux de Box et Tiao, notamment [22].

4.3.1 Le modèle

La théorie de l'analyse d'intervention développée par Box et Tiao permet de prendre en compte, lors de la modélisation SARIMA d'une série chronologique, des événements extérieurs au modèle. On apporte ainsi au modèle statistique une information supplémentaire de type qualitatif qui est intégrée, de manière additive, au modèle à l'aide de variables binaires exogènes. On pourra ainsi intégrer des scénarios de crises dans notre générateur de scénarios économiques.

Dans un premier temps, nous faisons une présentation théorique du modèle d'intervention de Box et Tiao. Nous présenterons au paragraphe suivant des cas particuliers que l'on a testé afin de choisir le modèle à retenir pour nos simulations. Le modèle d'intervention proposé par Box et Tiao se présente sous la forme suivante :

$$X_t = C + \frac{s(B) \times B^b}{\delta(B)} \xi_t + SARIMA_t$$

Avec :

- X_t : série à modéliser,
- s et δ sont des polynômes dont le degré est à déterminer,
- B : opérateur retard présenté au chapitre précédent,
- C est une constante à déterminer.

La fonction déterministe $\delta^{-1}(B)s(B)B^b\xi_t$ représente l'effet de l'intervention qui vient s'ajouter de manière additive au modèle SARIMA. Elle est appelée fonction d'intervention. Dans l'équation ci-dessus, la suite de variables aléatoires $(\xi_t)_{t>0}$ représente l'effet d'une intervention extérieure à la date t' , mise sous la forme d'une variable déterministe qui prend pour valeur un ou zéro suivant la présence ou l'absence de l'intervention. Cette variable est, en général, modélisée par deux classes de fonctions :

- une fonction en forme de saut, c'est-à-dire :

$$\begin{aligned} \xi_t &= S_t^{(t')} = 0 \text{ si } t < t' \\ \xi_t &= S_t^{(t')} = 1 \text{ si } t \leq t' \end{aligned}$$

- une fonction en forme d'impulsion, c'est-à-dire :

$$\begin{aligned} \xi_t &= P_t^{(t')} = 0 \text{ si } t \neq t' \\ \xi_t &= P_t^{(t')} = 1 \text{ si } t = t' \end{aligned}$$

On remarque toutefois que grâce à l'égalité suivante : $(1-B)S_t^{(t')} = P_t^{(t')}$, on peut toujours passer d'un saut à une impulsion.

Plus généralement, la série chronologique peut être perturbée par k interventions de nature différente. Avec les notations précédentes, le modèle d'intervention s'écrit alors :

$$X_t = C + \sum_{j=1}^k \frac{s_j(B) \times B^{b_j}}{\delta_j(B)} \xi_t^{(T_j)} + SARIMA_t$$

Une hypothèse fondamentale lors de l'utilisation de l'analyse d'intervention est que la structure du modèle soit la même avant et après l'intervention. Ainsi, afin d'estimer les paramètres du modèle, il faut fixer deux sous-ensembles de données correspondant à l'évolution du processus avant et après la date d'intervention qui est fixée de manière déterministe.

Dans notre cadre, ceci n'est pas nécessaire car le saut sera simulé aléatoirement et n'est pas présent dans la structure intrinsèque des données servant à calibrer l'équation. Ce problème d'estimation ne se pose donc pas. Cependant, cette remarque indique que le processus à sauts sera traité indépendamment du SARIMA.

En ce qui concerne la forme de la fonction d'intervention, il n'existe pas de méthode précise permettant de la déterminer. Néanmoins, il existe différents types de fonctions permettant de s'adapter à la forme graphique que l'on souhaite modéliser, suite à l'effet de l'intervention extérieure. Dans le paragraphe suivant, nous allons nous intéresser à quelques types de fonctions d'intervention classiques et justifier le choix de l'une d'entre elles.

4.3.2 Les différents types de sauts

Comme nous l'avons évoqué dans le paragraphe précédent, le choix d'une fonction d'intervention est assez subjectif car il dépend de la forme graphique et des propriétés géométriques que l'on souhaite modéliser. Dans ce paragraphe, nous allons présenter trois types de modèles d'intervention différents en utilisant une fonction d'impulsion comme variable exogène binaire.

Modèle avec intervention ponctuelle

Supposons, d'une part, que l'on observe une trajectoire qui est la réalisation d'un processus SARIMA. Supposons, d'autre part, qu'à une date $t = t'$ l'évolution de cette trajectoire soit perturbée par un choc extérieur que l'on a schématisé par le graphique ci-dessous.

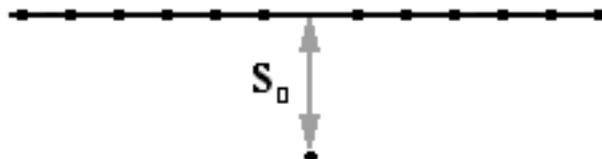


FIGURE 4.1 – Représentation locale de l'intervention ponctuelle

Le modèle d'intervention s'écrit alors sous la forme suivante :

$$X_t = C + s_0 \times P_t^{(t')} + SARIMA_t$$

Le paramètre s_0 représente l'impact de l'intervention extérieure sur la série. Son estimation permet de donner une mesure de cet impact. Dans notre cas, la valeur de ω_0 sera déterminée, de manière exogène, dans nos scénarios.

Dans la littérature anglo-saxonne, ce modèle est dit de type *Additive Outlier*.

Modèle avec intervention à effet rémanent

Nous pouvons également supposer que l'intervention au temps $t = t'$ vient perturber l'évolution de la série avec un effet qui décroît progressivement dans le temps, de façon exponentielle. Cette intervention est schématisée par le graphique ci-dessous :

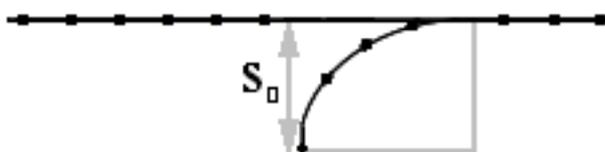


FIGURE 4.2 – Représentation locale de l'intervention à effet rémanent

Le modèle d'intervention s'écrit alors sous la forme suivante :

$$X_t = C + \frac{s_0}{1 - \lambda B} \times P_t^{(t')} + SARIMA_t$$

Le paramètre λ est un paramètre strictement compris entre zéro et un qui mesure la vitesse de décroissance de l'effet de l'intervention : s'il est proche de zéro, cela signifie que l'impact est quasiment ponctuel (on se ramène alors au cas précédent) et s'il est proche de un, cela signifie que l'impact se prolonge dans le temps.

Dans la littérature anglo-saxonne, ce modèle est dit de type *Temporary Change* ou *Transient Change*.

Modèle avec intervention à effet rémanent et changement de niveau

Enfin, il est possible de supposer que l'intervention vienne perturber l'évolution de la série avec un effet rémanent suivi d'un changement de niveau. Cette intervention est schématisée par le graphique ci-dessous.

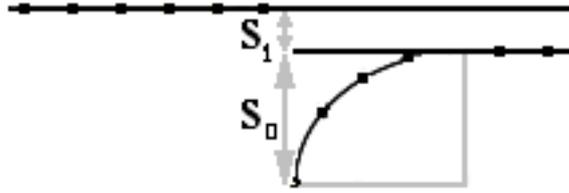


FIGURE 4.3 – Représentation locale de l'intervention à effet rémanent et changement de niveau

Le modèle d'intervention s'écrit alors sous la forme suivante.

$$X_t = C + \left(\frac{s_0}{1 - \lambda B} + \frac{s_1}{1 - B} \right) \times P_t^{(t')} + SARIMA_t$$

Le paramètre s_1 représente la différence de niveau consécutive au choc : il sera donc positif si le niveau observé après le choc est supérieur au niveau observé avant le choc et, négatif sinon. La somme de s_0 et s_1 représente l'impact de l'intervention extérieure sur la série. De même que précédemment, λ mesure la vitesse de décroissance de l'effet de l'intervention. Dans la littérature anglo-saxonne, ce modèle est dit de type *Level Shift*.

Toutes les méthodes présentées ci-dessus ont été testées pour intégrer des sauts dans l'équation de l'inflation. Nous verrons, dans le chapitre consacré à la modélisation de l'inflation, la méthode retenue et les raisons pour lesquelles ladite méthode a été retenue. Par ailleurs, nous avons appliqué la méthode choisie à toutes les équations du modèle.

Chapitre 5

Modélisation des résidus des équations

Nous avons signalé, dans le chapitre consacré à la présentation du modèle de Wilkie, qu'un inconvénient majeur de ce modèle, est le caractère non normal des résidus des équations. En effet, comme nous le montrerons dans les prochains chapitres, les tests de normalité sont tous rejetés. Une approximation réalisée couramment est de considérer tout de même les résidus comme normaux afin de pouvoir obtenir des projections facilement. Dans ce mémoire, nous avons souhaité présenter une approche différente afin de ne pas réaliser cette approximation courante. Nous avons alors cherché des lois, modélisables relativement facilement et permettant de mieux s'adapter aux résidus de nos équations. Après plusieurs tentatives vaines, nous avons décidé de modéliser les résidus par une loi de Pareto hybride, introduite récemment dans [8]. Ce chapitre vise à présenter brièvement cette loi.

5.1 Loi de Pareto généralisée

5.1.1 Généralités

Tout d'abord, il est nécessaire de présenter la loi de Pareto généralisée qui est nécessaire à notre modélisation. Cette loi a été introduite notamment pour l'étude des valeurs extrêmes par Pickands (1975). Sa fonction de répartition et sa densité sont données, dans l'ordre, par les expressions suivantes :

$$F(x) = \begin{cases} 1 - (1 + \xi \times x/\beta)^{-1/\xi} & \text{si } \xi \neq 0, \\ 1 - \exp(-x/\beta) & \text{si } \xi = 0, \end{cases}$$

$$f(x) = \begin{cases} \beta^{-1}(1 + \xi \times x/\beta)^{-1/\xi-1} & \text{si } \xi \neq 0, \\ \beta^{-1}\exp(x/\beta) & \text{si } \xi = 0, \end{cases}$$

où $\beta \geq 0$, $x \geq 0$ quand $\xi \geq 0$ et $0 \leq x \leq -\beta/\xi$ quand $\xi \leq 0$.

Le paramètre ξ est le paramètre de forme. Il s'agit du paramètre définissant l'épaisseur de la queue. Le paramètre β est un paramètre d'échelle.

Il est possible d'ajuster une loi de Pareto à un ensemble de données en appliquant des méthodes classiques telles que la méthode des moments ou celle du maximum de vraisemblance. D'autres méthodes plus spécifiques existent comme le recours à l'estimateur de Zhang. Dans notre mémoire, nous avons appliqué la méthode du maximum de vraisemblance. Le lecteur trouvera des informations plus précises sur différentes méthodes d'ajustement en consultant le mémoire sus-cité.

5.1.2 Fonction quantile

Nous rappelons ci-dessous l'expression de la fonction quantile de la loi de Pareto généralisée nécessaire notamment pour simuler cette loi par la méthode d'inversion.

$$F^{-1}(x) = \begin{cases} \frac{\beta}{\xi}((1-x)^{-\xi} - 1) & \text{si } \xi \neq 0, \\ -\beta \ln(1-x) & \text{si } \xi = 0, \end{cases}$$

5.2 Loi de Pareto hybride

5.2.1 Généralités et densité de probabilité

Carreau et Bengio ont proposé d'utiliser une loi gaussienne en remplaçant une des queues par une loi de Pareto généralisée. Dans son mémoire, William Gehin propose de modéliser les deux queues de la distribution par deux lois de Pareto et le centre par une loi gaussienne. Schématiquement, la densité est représentée comme suit :

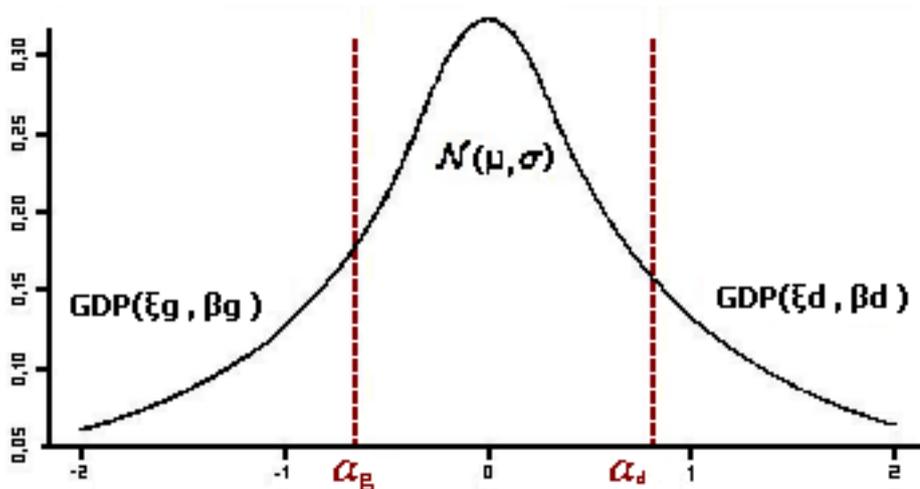


FIGURE 5.1 – Densité de probabilité d'une loi de Pareto Hybride

Ainsi, en notant :

- n la fonction densité de la loi normale de moyenne μ et d'écart-type σ .
- g_g et g_d les fonctions de densité des lois de Pareto généralisées respectivement à gauche et à droite.

La fonction de densité de la loi Pareto hybride est donc définie par :

$$f(x) = \begin{cases} g_g(x)/\gamma & \text{si } x \leq \alpha_g, \\ n(x)/\gamma & \text{si } \alpha_g \leq x \leq \alpha_d, \\ g_d(x)/\gamma & \text{si } x \geq \alpha_d. \end{cases}$$

où γ est un coefficient de normalisation assurant que f est bien une densité. Il est défini par l'équation suivante :

$$\gamma = 2 + N\left(\frac{\alpha_d - \mu}{\sigma}\right) - N\left(\frac{\alpha_g - \mu}{\sigma}\right)$$

où N est la fonction de répartition de la loi normale centrée réduite.

5.2.2 Contraintes sur les paramètres

Des restrictions sont nécessaires afin d'assurer la continuité et la dérivabilité de la densité. Nous nous contentons ici de rappeler les résultats :

$$\begin{cases} \alpha_g &= \mu - \sigma \times \sqrt{W(z_g)} \\ \beta_g &= \frac{\sigma(1+\xi_g)}{\sqrt{W(z_g)}} \end{cases}$$

avec $z_g = \frac{(1+\xi_g)^2}{2\pi}$, et où W est la fonction de Lambert.

Symétriquement, on a :

$$\begin{cases} \alpha_d &= \mu + \sigma \times \sqrt{W(z_d)} \\ \beta_d &= \frac{\sigma(1+\xi_d)}{\sqrt{W(z_d)}} \end{cases}$$

avec $z_d = \frac{(1+\xi_d)^2}{2\pi}$.

Ainsi, pour ajuster une loi de Pareto hybride, il faut estimer quatre paramètres : la moyenne μ , l'écart-type σ , et les deux paramètres de forme ξ_g et ξ_d des lois de Pareto généralisées. Nous noterons ainsi la loi de Pareto hybride par $\mathcal{H}(\mu, \sigma, \xi_g, \xi_d)$.

5.2.3 Fonction de répartition et fonction quantile

La fonction de répartition et la fonction quantile de la loi de Pareto hybride sont également utiles. Pour les simulations, nous avons utilisé la fonction inverse de la fonction de répartition donnée par l'expression suivante :

$$F^{-1}(x) = \begin{cases} G_g^{-1}(\gamma x) & \text{si } x \leq \alpha_g, \\ \mu + \sigma N^{-1}(\gamma x + N(\frac{\alpha_d - \mu}{\sigma}) - 1) & \text{si } \frac{1}{\gamma} \leq x \leq \frac{1 + N(\frac{\alpha_g - \mu}{\sigma}) - N(\frac{\alpha_d - \mu}{\sigma})}{\gamma}, \\ G_d^{-1}(\gamma x + N(\frac{\alpha_g - \mu}{\sigma}) - N(\frac{\alpha_d - \mu}{\sigma})) & \text{si } x \geq \frac{1 + N(\frac{\alpha_g - \mu}{\sigma}) - N(\frac{\alpha_d - \mu}{\sigma})}{\gamma} \end{cases}$$

avec

$$F_g^{-1}(x) = \begin{cases} \alpha_g + \frac{\beta_g}{\xi}((1 - x^{-\xi_g})) & \text{si } \xi_g \neq 0, \\ \alpha_g + \beta_g \ln(x) & \text{si } \xi_g = 0, \end{cases}$$

et

$$F_d^{-1}(x) = \begin{cases} \alpha_d + \frac{\beta_d}{\xi_d}((1 - x)^{-\xi_d} - 1) & \text{si } \xi_d \neq 0, \\ \alpha_d - \beta_d \ln(1 - x) & \text{si } \xi_d = 0, \end{cases}$$

Dans toutes les équations de notre générateur de scénarios, nous modéliserons les résidus par une loi de Pareto généralisée en estimant les paramètres par la méthode du maximum de vraisemblance.

Notons que la modélisation par une loi de Pareto hybride est efficace principalement en période de crises. En effet, dans ces périodes, la loi normale conduit à une sous-estimation des valeurs extrêmes. Notre générateur de scénarios étant construit pour s'adapter au mieux aux crises, la modélisation des résidus par cette loi nous a semblé importante même si elle augmente le nombre de paramètres de notre générateur.

Chapitre 6

Modélisation de l'inflation

Ce chapitre est consacré à la modélisation de l'inflation. Après avoir présenté les données utilisées, nous allons appliquer l'algorithme de Box et Jenkins afin de définir la modélisation la plus adaptée. Puis, nous déterminerons une méthode d'intégration de sauts à l'aide de l'algorithme de Box et Tiao.

6.1 Présentation des données

Dans une première démarche, nous avons considéré l'inflation sur une grande période, incluant les chocs pétroliers, les guerres mondiales... Nous nous sommes rapidement rendu compte que modéliser une telle série par un AR(1) n'était pas adapté à ce qu'on souhaitait modéliser et qu'aucune interprétation ne serait alors possible. Nous avons réalisé une étude permettant de stationnariser les séries et d'identifier le nombre de retards optimal pour pouvoir modéliser la série. Cependant, à l'issue de ce travail, nous avons décidé de considérer d'une part une autre période et, d'autre part, une autre série de données.

En effet, nous avons choisi de considérer la *core inflation* qui est l'inflation sans les prix du pétrole et des produits alimentaires. Nous avons fait ce choix car la période choisie, de dix ans, pour calibrer les équations, implique qu'il faut considérer des séries relativement stables si l'on souhaite les calibrer de manière fiable. Or, durant les dix dernières années, les prix des produits alimentaires et le prix du pétrole fluctuent énormément, ce qui, si on les avait considérés, rendrait la série difficilement modélisable, qui plus est par un processus autorégressif d'ordre un.

6.2 Application de l'algorithme de Box et Jenkins

Afin de prendre en compte le plus de types de séries chronologiques possible, nous avons considéré les processus SARIMA présentés au chapitre deux. Nous avons appliqué l'algorithme de Box et Jenkins afin de déterminer les paramètres optimaux et la valeur des coefficients pour chaque pays.

6.2.1 Stationnarisation et différenciations

Dans un premier temps, nous réalisons des tests de stationnarité sur la série des taux d'inflation bruts. Le tableau ci-dessous donne les *p-values* des tests pour quatre pays :

Tests	Dickey-Fuller	Phillips-Perron	KPSS
France	0,2924	0,6641	< 0,01
États-Unis	0,4673	0,4179	< 0,01
Japon	0,4955	0,3596	< 0,01
Allemagne	0,4595	0,07311	0,01607

TABLE 6.1 – Série brute - test de stationnarité

Ainsi, il est clair que les séries des taux d'inflation non différenciées ne sont pas stationnaires.

Rappelons ici que les deux premiers tests sont des tests de racine unitaire servant à tester la non stationnarité de la série tandis que le troisième teste la stationnarité de la série autour d'une tendance. Ainsi, on pourra conclure que la série étudiée est stationnaire si les *p-values* des deux premiers tests sont inférieures à 0.05 et si celle du troisième est supérieure à 0.05.

Pour nous convaincre définitivement de la non-stationnarité des séries, nous avons tracé la fonction d'autocorrélation (ACF) des séries. En effet, rappelons qu'une série stationnaire fluctue autour d'une valeur moyenne et que sa fonction d'autocorrélation décline rapidement vers 0. En pratique, si une série présente des autocorrélations positives pour un grand nombre de décalages, alors elle nécessite d'être différenciée. Nous constatons sur les graphiques suivants que c'est le cas ici :

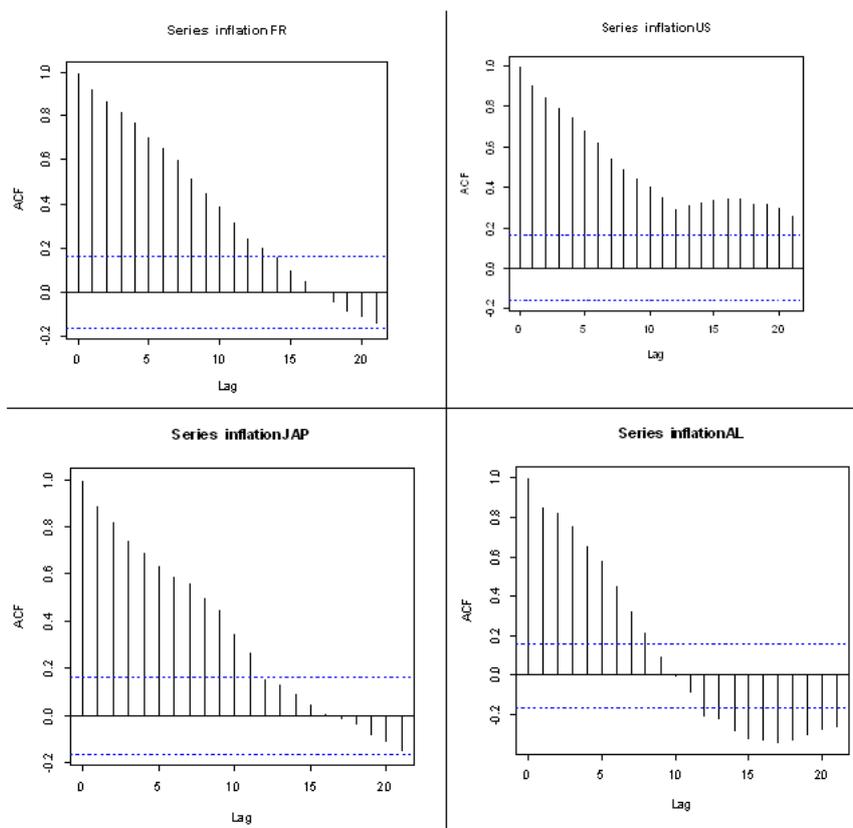


FIGURE 6.1 – ACF des séries brutes

Nous avons donc cherché à rendre les séries stationnaires en les différenciant à l'ordre un. Après différenciation, nous obtenons les résultats ci-dessous :

Tests	Dickey-Fuller	Phillips-Perron	KPSS
France	< 0,01	< 0,01	> 0,1
États-Unis	< 0,01	< 0,01	> 0,1
Japon	< 0,01	< 0,01	> 0,1
Allemagne	< 0,01	< 0,01	> 0,1

TABLE 6.2 – Série différenciée - test de stationnarité

Nous constatons que les tests valident l'hypothèse de la stationnarité des séries différenciées.

La série différenciée à l'ordre un peut donc être considérée comme stationnaire.

Remarque : *Pourquoi ne pas choisir de différencier une nouvelle fois ? Cela permettrait-il d'avoir une série plus stationnaire que la précédente ?*

Afin de répondre à ces interrogations, il convient de signaler que la différenciation d'une série tend à introduire des autocorrélations négatives. Si l'autocorrélation de décalage un est égale à zéro ou est négative, la série n'a pas besoin d'être différenciée. Par ailleurs, si l'autocorrélation de décalage un est inférieure à -0.5, cela signifie que la série est surdifférenciée.

L'ordre optimal de différenciation est souvent celui pour lequel l'écart-type est minimal. Un accroissement de l'écart-type doit donc être considéré comme un symptôme de surdifférenciation.

Comme nous le verrons dans la suite, les séries ne sont pas surdifférenciées dans notre cas et il n'est pas nécessaire de les différencier une deuxième fois car cela augmente l'écart type de la série.

6.2.2 Désaisonnalisation

Afin de déterminer les paramètres optimaux du SARIMA, nous avons tracé la fonction d'autocorrélation partielle (PACF) et l'ACF des séries. Ci-dessous, nous avons représenté le corrélogramme ainsi que le corrélogramme partiel de la France et des États-Unis. Ceux de l'Allemagne et du Japon sont relativement similaires à ces derniers.

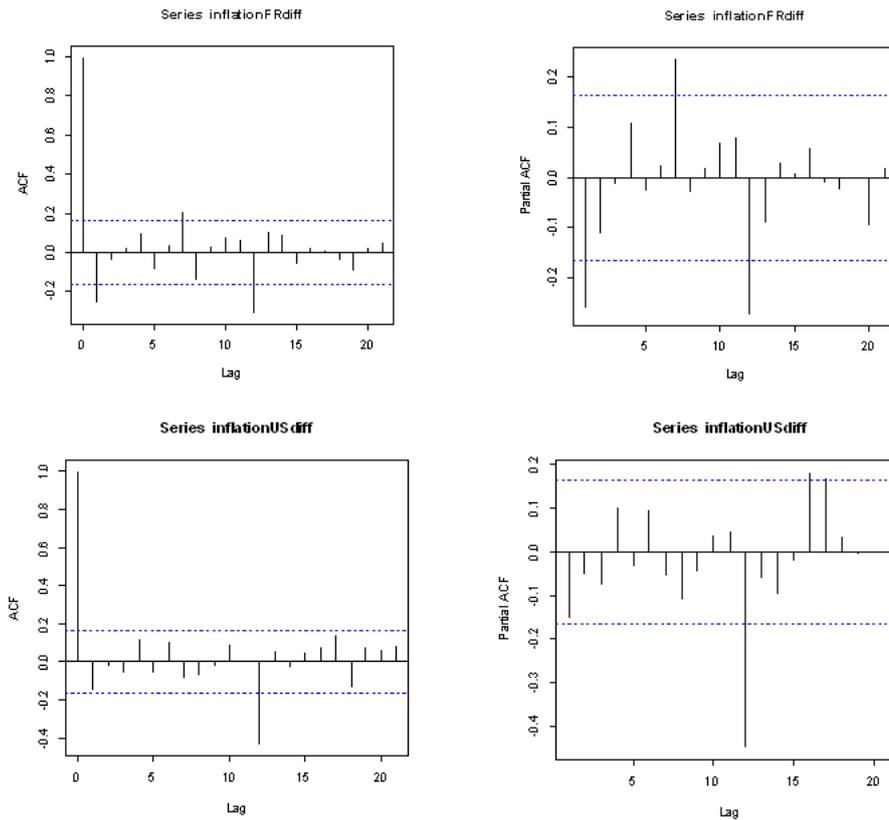


FIGURE 6.2 – ACF et PACF des séries différenciées à l'ordre un

Les ACF et PACF laissent apparaître une saisonnalité de période douze (pic significatif pour les ordres douze).

Pour confirmer cela, il est intéressant de tracer le spectrogramme de Fourier afin de vérifier que la fréquence correspondant à un évènement annuel (ou biennuel) se démarque des autres. Par ailleurs, nous pouvons également tracer l'ACF et la PACF des séries saisonnièrement différenciées pour constater qu'il existe bien une saisonnalité de période douze. Nous ne représentons pas ces courbes ici mais elles confirment la présence d'une saisonnalité annuelle.

Notons que d'un point de vue macroéconomique, il est clair que l'indice d'inflation est sensible à la saisonnalité. En effet, les prix subissent des pressions différentes suivant les mois. En décembre par exemple, les fêtes de Noël ont un effet cyclique sur les prix qui est différent de celui d'octobre.

Nous considérons donc une saisonnalité de douze mois dans la suite de notre étude.

6.2.3 Paramètres optimaux

Afin que le modèle ne soit pas surparamétré, nous avons étudié la possibilité d'utiliser des modèles du type SARIMA(1,1,1)(1,0,1)₁₂, SARIMA(1,1,1)(0,0,1)₁₂...

Pour choisir le modèle le plus adapté parmi tous ces modèles, nous avons décidé d'utiliser le critère bayésien d'Akaike (AIC).

Le tableau ci-dessous donne les valeurs du critère AIC calculées pour le jeu de données de la France :

Critères	AIC
SARIMA(1,1,1)(1,0,0)12	-1437
SARIMA(1,1,1)(1,0,1)12	-1440
SARIMA(1,1,1)(1,1,0)12	-1239
SARIMA(1,1,1)(1,1,1)12.	-1272
SARIMA(1,1,1)(0,0,0)12	-1421
SARIMA(1,1,1)(0,0,1)12	-1441
SARIMA(1,1,1)(0,1,0)12	-1186
SARIMA(1,1,1)(0,1,1)12	-1262

TABLE 6.3 – Valeurs du critère AIC - France

Finalement, le tableau suivant donne l'ordre des SARIMA les plus adaptés pour notre modélisation au sens du critère AIC pour chaque pays :

Critères	Paramètres du SARIMA
France	SARIMA(1,1,1)(0,0,1)12
États-Unis	SARIMA(1,1,1)(0,0,1)12
Japon	SARIMA(1,1,1)(1,0,0)12
Allemagne	SARIMA(1,1,1)(1,0,1)12

TABLE 6.4 – Modélisation optimale selon le critère AIC

Nous constatons que les paramètres optimaux sont identiques pour la France et les États-Unis mais qu'ils diffèrent pour le Japon et l'Allemagne. Cependant, la comparaison des critères AIC nous a conduit à utiliser le SARIMA(1,1,1)(0,0,1)12 pour tous les pays. En effet, concernant le Japon et l'Allemagne, la valeur du critère est quasiment identique pour le SARIMA optimal et pour le SARIMA similaire à celui des Etats Unis et de la France, comme le montre les tableaux suivants :

Allemagne	AIC
SARIMA(1,1,1)(1,0,1)12	-1333
SARIMA(1,1,1)(0,0,1)12	-1331

TABLE 6.5 – Valeurs du critère AIC - Allemagne

Japon	AIC
SARIMA(1,1,1)(1,0,0)12	-1413
SARIMA(1,1,1)(0,0,1)12	-1411

TABLE 6.6 – Valeurs du critère AIC - Japon

Ainsi, nous modéliserons l'inflation par un modèle SARIMA(1,1,1)(0,0,1)12 pour tous les pays.

6.2.4 Validation de la modélisation

Afin de valider l'utilisation de cette modélisation, il convient de s'assurer que les résidus obtenus sont des bruits blancs. Pour cela, nous pouvons avoir recours à des tests classiques tels que les tests de Ljung-Box et de Box-Pierce. Comme nous l'avons précisé dans le chapitre deux, nous avons choisi d'utiliser le test de Breusch-Godfrey. Nous signalons tout de même que les tests de Ljung-Box et de Box Pierce sont également vérifiés. Voici les résultats trouvés :

Test	Breusch-Godfrey
France	0,6082
États-Unis	0,5128
Japon	0,5542
Allemagne	0,4789

TABLE 6.7 – Valeurs de la statistique de Breusch-Godfrey

Par ailleurs, il est intéressant de confirmer ces tests en traçant l'ACF et la PACF des résidus qui ne doivent pas avoir de pic significatif. Nous présentons ci-dessous les résultats obtenus pour la France, ceux des autres pays étant similaires.

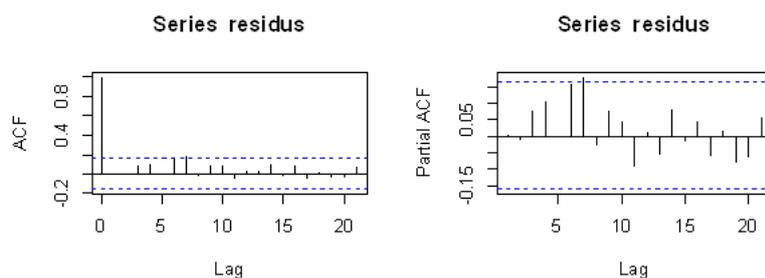


FIGURE 6.3 – ACF et PACF des résidus

Les *p-values* des tests étant supérieures à 0.05 et l'ACF et la PACF ne présentant pas de pic très significatif, on peut en conclure que les résidus sont bien des bruits blancs.

L'hypothèse d'indépendance des bruits blancs étant validée, nous confirmons la modélisation de l'inflation par un processus SARIMA(1,1,1)(0,0,1)₁₂.

Par ailleurs, nous tenons à signaler que cette modélisation qui offre une estimation plus juste que la modélisation proposée par Wilkie (1985) comporte seulement 3 paramètres i.e. un paramètre de plus que pour l'AR(1). Nous n'avons donc pas de problème de sur-paramétrisation.

Enfin, nous avons également testé la normalité des résidus afin de savoir si la modélisation de ces derniers par une loi de Pareto hybride était justifiée. Nous avons appliqué différentes méthodes : la méthode « statistique » et les tests. Nous donnons ci-dessous les valeurs des *skewness*, *kurtosis* et *p-values* relatives au test de Jarque-Bera.

Tests	Skewness	Kurtosis	Jarque-Bera
France	-0,4406	0,7383	0,0142
États-Unis	-1,356	8,144	$< 2,2 \cdot 10^{-16}$
Japon	-0,4953	4,182	$< 2,2 \cdot 10^{-16}$
Allemagne	0,5658	2,966	$1,499 \cdot 10^{-14}$

TABLE 6.8 – Résultats des tests sur les résidus

Les *p-values* concernant le test de Jarque-Bera sont inférieures à 0.05. Par ailleurs, les *skewness* et les *kurtosis* sont respectivement significativement différents de 0 et 3. Nous en déduisons que l’hypothèse de normalité des résidus est rejetée.

Notons qu’il existe différentes méthodes pour remédier à ce problème. Une des méthodes consisterait à modéliser la volatilité et à ne plus la considérer constante : il s’agit d’un processus GARCH. C’est ce que propose Wilkie dans sa version de 1995. Des études ont prouvé que l’utilisation de cette méthode a été moins efficace que la modélisation initiale.

Comme nous l’avons précisé plus haut, nous avons choisi de modéliser les résidus par une loi de Pareto hybride dont les paramètres ont été estimés par maximum de vraisemblance. Ce choix se justifie tout à fait en analysant les résultats du tableau ci-dessous. Les kurtosis des distributions des résidus sont tous supérieurs à zéro. Précisons que nous indiquons dans le tableau la valeur du kurtosis normalisé, c’est-à-dire que le kurtosis d’une distribution normale serait égale à zéro. Cela traduit le caractère leptokurtique de la distribution des résidus, en particulier pour les résidus relatifs aux données américaines. Par ailleurs, les coefficients d’asymétrie des distributions des résidus sont différents de zéro : cela montre que les distributions ne sont pas symétriques. Ces deux caractéristiques peuvent être pris en compte en utilisant une modélisation par une loi Pareto hybride : en modifiant les queues de distribution, il est possible d’adapter plus précisément un modèle adéquat. Nous présentons en annexe le calibrage retenu pour les quatre pays présentés ci-dessus.

Notons que si d’un point de vue théorique cette modélisation est plus précise, d’un point de vue pratique, les résultats obtenus ne sont pas significativement différents dans la mesure où les résidus ont des variances faibles et des moyennes nulles. Les figures 8.3 et 8.4 illustrent cette remarque.

6.3 Inclusion de sauts dans l’équation

Afin d’inclure des sauts dans l’équation de l’inflation, nous avons testé les différentes méthodes évoquées au chapitre trois. Nous avons rapidement rejeté l’hypothèse de modéliser des sauts en utilisant une loi de Poisson car cela n’était pas adapté à notre souhait. En effet, à l’instar de la méthode *additive outliers*, ce type d’intervention n’est pas adapté à notre étude car si, comme on l’a indiqué précédemment, on observe après les crises un retour à l’équilibre, ce retour n’est pas immédiat. La durée de retour à l’équilibre doit être

adaptée au type de crises que l'on rencontre ou que l'on simule. Le graphique ci-dessous représente une trajectoire obtenue en utilisant ce type d'intervention. Les hypothèses utilisées sont les suivantes : saut de -2% avec une probabilité de 5%. Il illustre le fait que l'impact d'une baisse de l'inflation est immédiatement effacé si on utilise ce type d'intervention.

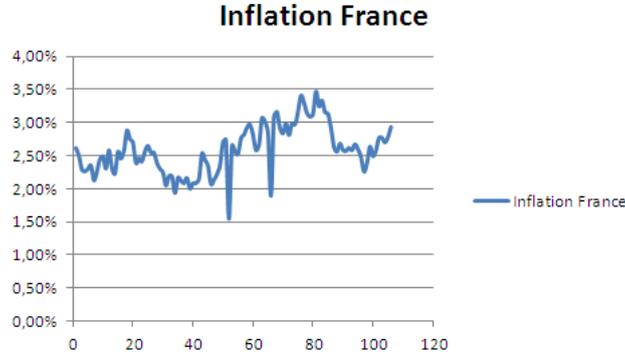


FIGURE 6.4 – Représentation locale de l'intervention ponctuelle

Nous avons ensuite testé la méthode d'intervention ponctuelle avec effet rémanent proposé par Box et Tiao : ce modèle semble parfaitement en accord avec ce qu'on souhaite modéliser i.e. des chocs dont l'effet n'est pas immédiatement résorbé. Par ailleurs, il nous semble intéressant de pouvoir, par l'intermédiaire du paramètre λ , modéliser un impact plus ou moins long du choc. Cependant, un problème majeur de cette méthode peut résider dans l'estimation des trois paramètres de sauts : amplitude, probabilité d'occurrence et paramètre de résorption. Dans notre cas, ce problème ne se pose pas : en effet, c'est par l'intermédiaire de ces trois paramètres intégrés dans le modèle par le biais d'une interface que le gérant du fonds peut intégrer son analyse macroéconomique. Les valeurs des coefficients sont donc déterminés « à dire d'experts ».

Comme nous l'avons vu, il est possible d'affiner encore le modèle d'intervention en considérant une intervention ponctuelle avec effet rémanent et changement de niveau. Nous n'avons pas retenu ce modèle pour ne pas surparamétrer le modèle et nous avons considéré qu'estimer une différence de niveau, « à dire d'experts » introduisait trop de subjectivité.

En conclusion, après avoir testé les différentes méthodes évoquées ci-dessus, nous avons choisi d'utiliser l'algorithme de Box et Tiao avec la considération d'une fonction d'intervention avec effet rémanent sans changement de niveau et dont l'estimation des paramètres est réalisée « à dire d'experts ».

Nous modéliserons de ce fait l'inflation par le processus suivant :

$$I_t = \sum_{j=1}^k \frac{s_0}{1 - \lambda B} P_t^{(T_j)} + SARIMA_t$$

6.4 Lissage

Certaines équations du générateur de Wilkie nécessitent le lissage de la courbe de l'inflation. Wilkie propose un lissage exponentiel. Nous avons étudié l'éventualité d'utiliser le lissage de Winters qui a l'avantage de pouvoir prendre en compte la saisonnalité de la série lissée mais n'avons pas retenu cette méthode car son apport était limité.

Nous avons décidé de conserver le lissage exponentiel proposé par Wilkie.

Chapitre 7

Modélisation de la croissance et de la croissance des dettes

La construction de scénarios de croissance mondiale a été remise en cause par la mauvaise qualité des prévisions effectuées, notamment dans les années 1960 et 1970. En effet, dans la plupart des cas, un rattrapage général des niveaux de vie des pays industrialisés les plus riches était prévu, y compris par les pays en développement, en trois ou quatre décennies. Ces prévisions ont été particulièrement inexactes, une grande partie du monde restant jusqu'à présent exclue de ce phénomène de rattrapage. En dépit de ces échecs passés, des hypothèses de croissance par pays restent tout de même indispensables.

La construction de scénarios de croissance mondiale à plus ou moins long terme reste donc nécessaire, malgré l'incertitude inhérente à cet exercice. En outre, les notions de croissance et particulièrement de croissance des dettes sont devenues particulièrement indispensables depuis le début du XXI^{ème} siècle. Le lecteur pourra constater le rôle fondamental que jouent les prévisions de conjoncture dans la sélection des actions servant à optimiser le portefeuille.

Dans ce mémoire, nous ne proposons pas de modélisation très précise de la croissance : en effet, des modèles complexes existent, par exemple le modèle de Solow, mais leur utilisation nécessite un temps d'étude important. Le lecteur intéressé pourra consulter [14], consacrée à la modélisation de la croissance du PIB par un modèle de Mankiw, Romer et Weil (1992) ou encore [5].

Nous réaliserons ici une étude moins complexe et adopterons une démarche en deux étapes pour prendre des décisions relatives à la valeur de la croissance : d'une part, nous réaliserons des projections à l'aide d'un modèle relativement simple que nous allons présenter ; d'autre part, entrera dans notre processus, la prise en compte des indicateurs avancés de l'OCDE.

Dans la deuxième partie de ce chapitre, nous aborderons le problème de la modélisation de l'endettement d'un pays, qui est devenue un facteur fondamental ces dernières années : nous présenterons notamment la notion de solde primaire stabilisant la dette.

7.1 Modélisation de la croissance

7.1.1 Définitions

Tout d'abord, il est important de préciser que nous souhaitons modéliser la croissance du produit intérieur brut (PIB) de chaque pays et non pas la croissance de la production de chaque pays. Cette dernière mesure a été abandonnée par l'OCDE au mois d'Avril 2012. En effet, en France par exemple, le secteur industriel ne représente plus qu'environ 9% de la croissance du PIB alors qu'auparavant, elle représentait environ 30% de la croissance, ce qui explique qu'on la modélisait. Cette chute résulte des diminutions des marges des produits industriels : un grand nombre d'entreprises françaises produit à perte en France et réalise ses bénéfices dans des pays émergents tels que la Chine, l'Inde, le Brésil ou la Roumanie.

Par ailleurs, pour modéliser la croissance des entreprises, on considérera de manière approximative qu'il s'agit de la croissance du pays dans lequel elle a son activité. Pour les entreprises mondiales, on affinera le processus en essayant d'étudier chaque entreprise individuellement. Ceci a pour objectif de ne pas surparamétrer le modèle. On étudiera la corrélation sur le passé entre la croissance de chaque entreprise et la croissance du pays pour définir si la modélisation est cohérente ou non.

Enfin, avant d'aborder la modélisation de la croissance à proprement parler, nous souhaitons faire une remarque concernant le lien entre la croissance économique et la croissance des dettes d'un Etat d'une part et, d'une entreprise d'autre part :

- quand une société est en croissance, elle s'endette.
- quand un pays est en croissance, il se désendette, excepté en période de guerre où la situation est inversée.

Après avoir présenté quelques généralités sur la croissance, nous présentons le modèle que nous avons retenu.

7.1.2 Choix du modèle et calibrage

A long terme, une approximation classique consiste à considérer qu'il existe un lien entre le taux d'intérêt nominal, le taux d'inflation et le taux d'intérêt réel. Plus précisément, on a la relation suivante :

$$\text{Taux d'intérêt Nominal} = \text{Taux d'Inflation} + \text{Taux d'intérêt Réel}$$

Par ailleurs, il existe également un lien, à long terme, entre le taux d'intérêt réel et le taux de croissance réel de l'économie : en effet, ils sont généralement considérés comme identiques. Ainsi,

$$\text{Taux d'intérêt réel à long terme} = \text{Taux de croissance réel à long terme}$$

Ces deux remarques préliminaires montrent que le taux de croissance est fonction de la valeur des taux longs et du taux d'inflation. Toutefois, à l'appui de l'article [16], qui présente une étude sur le lien entre l'évolution de la croissance et celle des taux réels à long terme, aux États-Unis, nous avons constaté que ce lien n'était pas toujours observé. En effet, le graphique ci-dessous présente l'évolution des taux réels et de la croissance observée depuis 1960. Les séries utilisées sont les suivantes :

- taux longs : taux des obligations d'Etat, 10 ans, émis par le Trésor américain,
- inflation : indice des prix à la consommation (IPC) en glissement annuel et en fréquence trimestrielle,
- croissance : évolution du produit intérieur brut (PIB) en glissement annuel et en fréquence trimestrielle.

La série des taux réels est obtenue en retranchant l'inflation des taux longs.

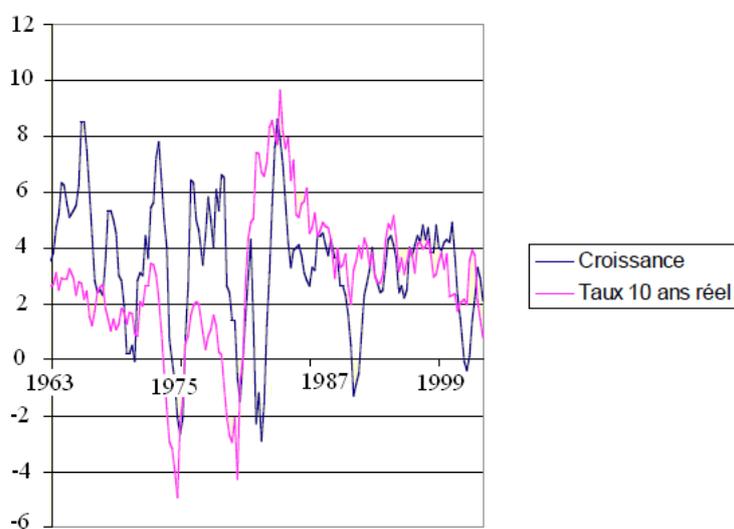


FIGURE 7.1 – Evolution de la croissance américaine

De plus, sur la période [2000 ; 2011] que l'on considère, voici l'évolution constatée :

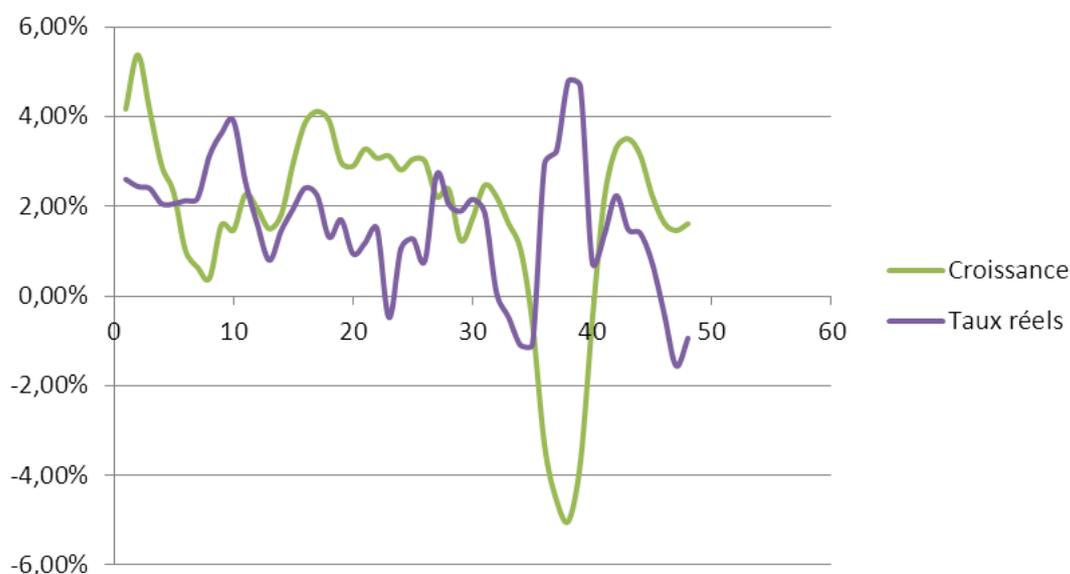


FIGURE 7.2 – Evolution de la croissance américaine

Par ailleurs, nous avons effectué quelques calculs de corrélation sur la période [2000 ; 2011] afin d'étudier le lien entre les taux longs, l'inflation et la croissance du PIB.

Voici les résultats obtenus :

Corrélation avec Taux Long	0,4237
Corrélation avec inflation	0,5750
Corrélation avec Taux Réels	-0,2914
Croissance moyenne	1,81%
Taux réels moyens	1,62%

TABLE 7.1 – Statistiques Taux réels et croissance

Globalement, nous observons que la baisse des taux réels anticipe la baisse de la croissance dans plusieurs cas. L'étude citée en début de paragraphe précise que la corrélation entre la croissance et les taux réels 10 ans depuis 1963 est de 0.24 et conclut qu'il n'existe pas de relation stable sur longue période entre les taux réels et la croissance de l'économie américaine.

Du reste, à l'appui des calculs réalisés sur la période à laquelle on s'intéresse, il nous semble délicat d'utiliser l'approximation de l'égalité des taux réels et de la croissance, les statistiques montrant que cette modélisation n'est pas appropriée.

Par ailleurs, rappelons que dans le GSE que l'on construit, nous simulons l'évolution de la *core inflation* et non de l'inflation totale. Ainsi, la série obtenue en retranchant la *core inflation* des taux long terme n'est pas celle des taux réels. Le coefficient de corrélation entre cette série et la série de la croissance est toutefois de 0.40.

Une méthode étudiée pour anticiper l'évolution future de la croissance consiste à considérer les taux longs comme un indicateur avancé de la croissance. Cependant, même si l'on constate qu'effectivement, globalement, les taux longs anticipent l'évolution de la croissance, le délai entre les deux évolutions n'est pas constant ce qui rend nos prévisions peu fiables. Dans la prochaine partie, nous présenterons un indicateur avancé de l'OCDE (le Composite Leading Indicator (CLI)) qui est performant.

Malgré ces remarques, nous souhaitons tout de même utiliser le fait que les évolutions des taux d'inflation et des taux nominaux sont corrélées avec l'évolution de la croissance. Ainsi, nous avons étudié différentes méthodes pour modéliser une relation du type :

$$\text{Croissance} = f(\text{inflation, taux longs})$$

Une première idée était d'utiliser des modèles à correction d'erreurs. Or, pour utiliser ce type de modèle et ainsi pouvoir appliquer le théorème d'Engel et Granger, il faut s'assurer que les séries étudiées sont cointégrées. Or, l'étude de la série de la croissance nous a montré qu'elle était stationnaire alors que les séries de l'inflation et des taux longs sont intégrées à l'ordre un. En effet, voici les *p-values* obtenues pour les séries brutes des États-Unis :

	Dickey-Fuller	Philipps-Perrond	KPSS
Croissance	0,0142	0,0401	> 0,1
Inflation	0,4694	0,6141	< 0,01
Taux réels	0,2336	0,3709	0,0150

TABLE 7.2 – *P-values* des tests de stationnarité

Lorsque l'on différencie une fois les deux séries non stationnaires, on obtient des séries stationnaires. Ainsi, après cette constatation, nous avons conclu que ce type de modélisation n'était pas adapté à notre étude.

Dans ces conditions, il reste une méthode simple envisageable pour modéliser la croissance : il s'agit de la régression linéaire multiple. Nous choisissons donc de réaliser une régression pour modéliser la croissance sur ces deux variables que l'on peut modéliser à l'aide du générateur de Wilkie. L'équation régissant l'évolution de la croissance est donc du type :

$$G_t = k + \alpha_G * I_t + \beta_G * C_t + \epsilon_t$$

Où, $(G_t)_{t \in \mathbb{N}}$ représente la série de la croissance, en glissement annuel.

Voici la valeur des coefficients calculés pour les États-Unis ainsi que les résultats des tests de significativité :

	Valeur	Standard error	P-values test student
Constante	0,0004	0,0002927	0,1033
Inflation	-0,12	0,05885	0,0366
Taux longs	0,21	0,04037	$5,78 \cdot 10^{-7}$

TABLE 7.3 – Coefficients et tests

Notons que les coefficients obtenus sont significatifs (à 5%) sauf la constante. En effet, les *p-values* obtenues en réalisant le test de Student sont inférieures à 0.05. Néanmoins, si nous choisissons de retirer la constante de la modélisation, cela fait diminuer l'*adjusted R²*. Pour cette raison, nous avons décidé de la conserver.

Nous nous contentons de présenter les résultats obtenus pour les États-Unis. La même démarche a été appliquée pour les autres pays.

Le R^2 global obtenu est de 19%, ce qui n'est pas très élevé. Nous retenons tout de même cette modélisation bien qu'ayant conscience qu'il s'agit d'une approche relativement élémentaire.

Pour remédier au problème du R^2 , nous proposons d'étudier les indicateurs avancés de l'OCDE pour affiner ces prévisions. Le paragraphe suivant est consacré à l'étude des indicateurs de l'OCDE pouvant nous permettre de prendre des décisions rationnelles et fiables.

7.1.3 Indicateurs de l'OCDE

Les indicateurs conjoncturels avancés sont des séries chronologiques descriptives donnant des informations anticipées qui servent d'outils d'analyse et de prévision macroéconomique. Ils peuvent être de plusieurs types : des indicateurs avancés de conjoncture économique, de retournement de la conjoncture, de crise de change, de crise monétaire, etc. Ces indicateurs avancés peuvent donc nous être utiles pour prévoir l'évolution future de la croissance d'un pays et anticiper un éventuel retournement.

Le principe d'un indicateur avancé est simple : étant donné le retard enregistré dans la mise à disposition des principales statistiques économiques (série de référence), on essaie d'identifier un ensemble de variables (appelées indicateurs avancés) disponibles plus rapidement et qui évoluent préalablement à la série de références. Dans notre cas, nous avons considéré successivement deux séries de références : le produit intérieur brut (PIB) et l'indice des prix à la consommation (IPI). Notons que la considération de l'une ou l'autre de ces séries est équivalente pour prévoir les crises ou les récessions.

Cependant, l'utilisation des indicateurs avancés est confrontée à deux types de risques : le risque de faux signaux et le risque de signaux manquants. Une analyse précise des indicateurs avancés est donc indispensable afin de tester leur fiabilité et leur prévision.

Le National Bureau of Economic Research (NBER) des États-Unis a défini cinq critères pour évaluer le comportement des indicateurs :

- Rôle de mesure : il doit exister un lien économique clair entre l'indicateur et la série de référence.
- Qualité de la mesure.

- Homogénéité des avances : l’horizon de la prévision doit être suffisamment stable pour donner des signaux crédibles.
- Possibilité d’identification rapide des retournements cycliques : cette propriété est relative au risque de signaux manquants qui découleraient d’un délai de détection trop long.
- Rapidité et fréquence de la communication des statistiques : la rapidité dans la fourniture de l’information est essentielle pour la pertinence de l’indicateur.

Une étude précise sur notre période de référence [2000 ; 2011] montre que pour prévoir les retournements de la croissance des pays étudiés de façon fiable, le *Composite Leading Indicator* (CLI) est le plus performant, notamment pour les États-Unis. En effet, avec l’aide d’étudiants de l’ENSAE, nous avons comparé le pouvoir prédictif de quatre indicateurs sur cette période et voici les résultats obtenus (nous présentons ci-dessous les résultats obtenus pour les États-Unis) :

Indicateurs et activité économique	Nombre de mois optimal de prévisions	Ralentissements non prédits	Ralentissements prédits mais non réalisés	Prédictions réalisées
CLI	5	0	17	82
SP500	6	23	27	50
ECRI	6	21	24	55
Conference board	7	16	17	67

TABLE 7.4 – Indicateurs avancés - pouvoir prédictif sur le PIB

Cette étude nous a permis de conclure que le CLI était l’indicateur le plus performant actuellement pour prévoir les retournements de conjoncture. Cependant, il convient de rester vigilant car cet indicateur a tendance à prévoir trop de retournements. L’étude de la mise en place de *stop loss* en cas de signe de retournement peut être envisagée afin de se protéger contre une éventuelle baisse.

Par ailleurs, une étude plus précise du CLI, sur une période plus large, nous a permis de définir le nombre de mois d’anticipation du CLI, pour différents pays. Elle nous a permis de conclure que le CLI n’était pas performant pour les pays émergents et qu’il possède une certaine robustesse pour d’autres pays, notamment la France et l’Allemagne. Le tableau ci-dessous récapitule les résultats pour les quatre pays que l’on étudie plus précisément, en considérant l’IPI comme série de référence :

Pays/Période	Sans crise	Période 74-81	Période 02-05	Période 07-08
France	7 mois	6 mois	5 mois	7 mois
États-Unis	6 mois	6 mois	5 mois	4,5 mois
Allemagne	5 mois	4 mois	5 mois	3,5 mois
Japon	4 mois	5 mois	2 mois	< 1 mois

TABLE 7.5 – Période d’anticipation du CLI sur l’IPI

Nous proposons en annexe quelques résultats plus détaillés relatifs à cette étude. Nous n'avons pas souhaité les indiquer dans le corps du mémoire car cette étude possède un caractère actuariel limité.

7.2 Indicateur IFO

En plus de l'étude précédente, nous avons étudié l'indice calculé mensuellement par l'*Institut für Wirtschaftsforschung* (IFO) qui permet de suivre la confiance de 7000 entrepreneurs allemands vis à vis de leur économie nationale. De part l'importance de l'économie allemande dans la zone euro, la valeur de l'IFO, publiée le dernier mercredi de chaque mois par l'institut de conjoncture, peut influencer grandement le niveau de l'euro sur le marché des changes et donc l'évolution du marché. Le nombre de mois d'anticipation de cet indice est de six mois ; par ailleurs, l'IFO possède la particularité de n'avoir manqué aucune récession de l'économie allemande depuis 1993. Ainsi, nous nous intéresserons à son évolution pour réaliser nos scénarios.

Pour conclure, nous proposons la table de décision suivante pour prendre des décisions relatives à la croissance :

Indicateurs/Scénarios	Projections GSE	Interprétation Indicateurs	Conclusion
Scénario 1	Croissance	Positif	Croissance
Scénario 2	Croissance	Retournement	Zone de perturbation
Scénario 3	Décroissance	Positif	Zone de perturbation
Scénario 4	Décroissance	Retournement	Décroissance

TABLE 7.6 – Table de décisions

7.3 Modélisation de la croissance des dettes

7.3.1 Notion de solde primaire stabilisant

Afin d'étudier la croissance des dettes des pays, introduisons la notion de solde primaire stabilisant la dette. Cette notion est primordiale dans l'évolution de la dette. N'ayant pas trouvé de démonstration claire de l'expression du solde primaire stabilisant la dette d'un pays, nous avons souhaité proposer notre propre démonstration. Définissons dans un premier temps différentes variables :

- B_t : dette à la date t en euros
- D_t : dette à la date t en pourcentage du PIB
- S_t : solde sur la période $[t-1 ; t]$ en euros
- P_t : solde sur la période $[t-1 ; t]$ en euros hors charge de la dette appelé *Solde primaire*
- p_t : solde primaire sur la période $[t-1 ; t]$ en pourcentage du PIB
- p_t^* : solde primaire stabilisant sur la période $[t-1 ; t]$ en pourcentage du PIB
- i_t : taux d'intérêt apparent moyen de la dette
- c_t : taux de croissance du PIB

La dette vérifie l'équation d'accumulation suivante :

$$B_{t+1} = B_t - S_{t+1}$$

Cette équation peut s'écrire en fonction du solde primaire de la façon suivante :

$$B_{t+1} = B_t \times (1 + i_t) - P_{t+1}$$

En écrivant cette équation en pourcentage du PIB, cela donne :

$$\frac{B_{t+1}}{PIB_{t+1}} = \frac{B_t \times (1 + i_t) - P_{t+1}}{PIB_{t+1}}$$

$$D_{t+1} = \frac{B_t}{PIB_t} \times \frac{1 + i_t}{1 + c_t} - p_{t+1}$$

$$D_{t+1} = D_t \times \left(1 + \frac{i_t - c_t}{1 + c_t}\right) - p_{t+1}$$

Ainsi, on obtient la variation de la dette sur une année, en fonction de la dette en t-1, du solde primaire, du taux de croissance du PIB et du taux d'intérêt apparent moyen de la dette :

$$D_{t+1} - D_t = \left(\frac{i_t - c_t}{1 + c_t}\right) \times D_t - p_{t+1}$$

Le solde primaire stabilisant est le solde à partir duquel la dette en t est supérieure à la dette en t+1 i.e. il vérifie $D_{t+1} - D_t = 0$. Soit :

$$p_{t+1}^* = \left(\frac{i_t - c_t}{1 + c_t}\right) \times D_t$$

Si le solde primaire de l'année (Recettes-dépenses hors service de la dette) est supérieur au solde primaire stabilisant, la dette diminue. Par ailleurs, plus l'intérêt payé sur la dette est faible et plus la croissance est forte, plus le solde primaire stabilisant est petit et donc plus il est possible que le solde primaire effectif de l'année soit supérieur au solde primaire stabilisant.

7.3.2 Modèle retenu

A partir de cette notion, nous allons maintenant proposer une méthode pour modéliser l'évolution de la dette d'un pays. Celle-ci dépend, comme nous l'avons vu dans la partie précédente, des déficits successifs constatés chaque année. Ces déficits dépendent de la politique mise en place, ce qui rend les prévisions relativement complexes. Ainsi, afin de modéliser l'évolution de la dette, nous allons devoir faire plusieurs scénarios. Nous souhaitons réaliser des projections sur un an et cinq ans. Rappelons le déficit annuel, la dette ainsi que la croissance des quatre pays étudiés durant les dernières années, en % du PIB :

Pays	France	États-Unis	Allemagne	Japon
2003	-4,1%	-5,0%	-4,0%	-7,9%
2004	-3,6%	-4,4%	-3,8%	-6,2%
2005	-3,0%	-3,3%	-3,3%	-6,7%
2006	-2,3%	-2,2%	-1,6%	-1,6%
2007	-2,7%	-2,9%	-0,3%	-2,4%
2008	-3,3%	-6,3%	-0,1%	-2,2%
2009	-7,5%	-11,3%	-3,0%	-8,7%
2010	-7,0%	-10,6%	-3,3%	-8,1%

TABLE 7.7 – Déficit annuel

Pays	France	États-Unis	Allemagne	Japon
2003	63,3%	60,2%	65,4%	158,0%
2004	65,2%	61,2%	68,8%	165,5%
2005	66,8%	61,4%	71,2%	175,3%
2006	64,1%	60,8%	69,3%	172,1%
2007	64,2%	62,0%	65,3%	167,0%
2008	68,2%	71,0%	69,3%	174,1%
2009	79%	84,3%	76,4%	194,1%
2010	82,3%	93,6%	87,0%	199,7%

TABLE 7.8 – Dette annuelle

Pays	France	États-Unis	Allemagne	Japon
2003	0,9%	2,5%	-0,2%	1,4%
2004	2,5%	3,6%	1,2%	2,7%
2005	1,8%	3,1%	0,8%	1,9%
2006	2,5%	2,7%	3,4%	2,0%
2007	2,3%	1,9%	2,7%	2,4%
2008	-0,1%	0,0%	1,0%	-1,2%
2009	-2,7%	-2,7%	-4,7%	-6,3%
2010	1,5%	2,9%	3,6%	3,9%

TABLE 7.9 – Croissance annuelle

Afin de modéliser l'évolution de la dette d'un pays, nous sommes contraints de faire des scénarios sur l'évolution du déficit (primaire ou non). Voici quelques exemples de scénarios que l'on peut envisager :

- stabilisation du déficit
- déficit réduit en 5 ans à 3% du PIB
- déficit réduit en 5 ans à 0
- augmentation du déficit de 10%

Notons que ces scénarios dépendent de la politique mise en place et qu'ils doivent être actualisés afin que la modélisation obtenue soit cohérente.

Une fois que les scénarios ont été choisis, nous proposons de les pondérer afin d'en déduire l'évolution de la dette : pour cela, nous comparons l'évolution ce déficit par rapport à la valeur du solde primaire stabilisant la dette en prenant comme hypothèse que le taux d'intérêt apparent de la dette est stable. Cette dernière hypothèse est acceptable dans la mesure où les projections sont réalisées sur un horizon moyen terme.

La méthode décrite ci-dessus nous permet d'obtenir une modélisation de la croissance des dettes pour chaque pays étudié. Néanmoins, nous sommes conscients que la modélisation présentée n'est pas idéale et qu'il conviendrait de l'affiner pour obtenir des modélisations plus pertinentes et moins déterministes : en effet, le recours a des scénarios sur l'évolution du déficit introduit nécessairement une part de déterminisme dans le processus. Cette modélisation ayant un impact limité dans l'application que nous proposons au dernier chapitre, nous avons préféré conserver un modèle simple.

En conclusion, nous avons décidé d'accorder plus d'importance aux indicateurs de l'OCDE et de l'IFO que ce que nous avons prévu initialement. La performance de ces indicateurs étant prouvée, nous leur accorderons une place privilégiée dans le processus d'optimisation.

Chapitre 8

Calibrage du modèle et simulations

Ce chapitre est consacré aux calibrages des équations relatives au générateur de Wilkie présentées au chapitre trois. Nous ne revenons pas ici sur le calibrage de l'équation de l'inflation car nous avons consacré un chapitre entier à cette grandeur centrale.

Dans une première partie, nous allons présenter les données utilisées ainsi que la valeur de certains paramètres. Par souci de concision, nous ne donnerons pas la valeur de l'estimation de tous les paramètres.

La seconde partie sera consacrée aux simulations. Nous rappelons différentes méthodes permettant de simuler une loi de probabilité, en particulier une loi gaussienne. Par ailleurs, nous revenons rapidement sur les générateurs de nombres aléatoires.

La troisième partie sera consacrée aux résultats des simulations. Nous commenterons et analyserons les projections obtenues.

8.1 Données et calibrage des équations

8.1.1 Données

Cette partie est très importante car il est impossible d'obtenir des projections fiables et interprétables si les données initiales ne sont pas pertinentes. De ce fait, nous avons attaché une importance particulière au choix des données d'une part, en déterminant une période précise, un pas ainsi qu'un nombre de pays à considérer et, d'autre part, en retraitant les données obtenues.

Choix de la période

Tout d'abord, il est important de définir une période commune à toutes les données ainsi qu'un pas. Nous avons choisi de considérer les données mensuelles sur la période [2000 ; 2011]. Cette période nous semblait appropriée pour prendre en compte les changements de contexte politico-économique que l'on a évoqués au chapitre quatre. Par ailleurs, ayant défini cette période, il nous semblait opportun de considérer des données mensuelles : en effet, des données annuelles ou trimestrielles nous donneraient une profondeur d'historique trop faible pour que notre modélisation soit fiable.

Afin d'obtenir l'ensemble des données nécessaires à la modélisation, nous avons utilisé Bloomberg.

Remarque 1 : *Pourquoi ne considère-t-on pas les données des années 1930 tant le rapprochement avec les années 2000 a été fait ?*

Nous avons décidé de ne pas prendre en compte les données des années trente, eu égard au traitement très politique de cette crise. En effet, le président de la réserve fédérale des États-Unis a pris des décisions discutables a posteriori qui ont engendré trois crises bancaires. Par ailleurs, il est difficile d'obtenir l'ensemble des données nécessaires de cette époque.

Remarque 2 : *Pourquoi n'a-t-on pas non plus considéré les données des années 1990 ?*

Nous avons choisi de ne pas le faire car la période actuelle nous semble réellement différente de celle d'il y a vingt ans, période durant laquelle la moitié des caisses d'épargne a disparu.

Choix des pays

Nous avons choisi de calibrer des équations pour six pays, avec des caractéristiques différentes :

- Les États-Unis : ce pays est à l'origine du départ de la majorité des cycles économiques. Il est donc intéressant de modéliser l'évolution des variables de ce pays. Par ailleurs, dans l'application que nous avons réalisée, nous considérons quarante valeurs américaines afin d'optimiser un portefeuille.
- La France et l'Allemagne : il s'agit des deux pays moteurs de la zone euro, à l'instant où ces lignes sont écrites, ils traduisent bien l'évolution future de la zone euro.
- Le Japon : pays atypique caractérisé par une récession chronique depuis la crise de 1989. Réalisant un générateur de scénarios économiques basé en partie sur l'étude macroéconomique de la croissance des dettes, il est indispensable de modéliser l'évolution de ce pays.
- L'Italie et l'Espagne : Ces deux pays souffrent d'une dette publique colossale. Ils ont choisi deux manières différentes pour endiguer la crise de la dette. Il nous semble important de les étudier.

Notons toutefois que le générateur économique créé peut s'adapter à tous les pays. Il convient seulement de s'assurer de la bonne qualité des données afin d'obtenir des estimations fiables.

Retraitement des données

Comme nous l'avons indiqué dans la partie concernant la description du générateur de scénarios de Wilkie, il est parfois nécessaire de retraiter les données afin de rendre les séries modélisables par des séries temporelles ; plus précisément, il faut utiliser des transformations rendant les séries stationnaires.

Voici les séries que l'on a considérées pour l'ensemble des données :

Série brute	Séries modifiées
Core Inflation : CPI	I(t)
Rendement des dividendes Y(t)	ln(Y(t))
Montant des dividendes D(t)	ln(D(t)/D(t-1))
Taux long : B(t)	ln(B(t))
Taux court : C(t)	ln(C(t)/C(t-1))

TABLE 8.1 – Données considérées

Nous avons réalisé des tests de stationnarité à l'aide des tests de Dickey-Fuller, de Philipps-Perrond et de KPSS. Nous avons constaté que les séries modifiées vérifiaient bien les tests de stationnarité.

8.1.2 Calibrage

Les séries étant modifiées, nous pouvons calibrer les équations. Pour réaliser cette étape, nous avons utilisé le logiciel R qui s'avère un outil puissant et rapide pour ce genre de calculs. Nous avons procédé au calibrage de chacune des équations, pour chacun des six pays. Par ailleurs, dans le chapitre suivant, nous avons calibré les équations de l'inflation, du rendement des actions et du montant des dividendes pour plusieurs actions. Voici, par exemple, les coefficients que l'on a déterminés pour ces trois premières équations correspondant au jeu de données des États-Unis. L'ensemble des coefficients utilisés pour calibrer les lois Pareto hybrides est présenté en annexe. Pour plus de lisibilité, nous rappelons les équations utilisées :

$$I_t = SARIMA_t$$

$$y_t - y_m = \alpha_Y \times I_t + \beta_Y \times (y_{t-1} - y_m) + \sigma_Y \times \epsilon_{Y,t}$$

$$d_t - d_m = \alpha_D \times I_t + \beta_D \times M(t) + \mu_D \times \epsilon_{Y,t-1} + \xi_D \times \sigma_D \times \epsilon_{D,t-1} + \sigma_D \times \epsilon_{D,t}$$

Nous notons les paramètres du SARIMA relatif à l'inflation de la façon suivante :

- AR : coefficient relatif à la partie autorégressive,
- MA : coefficient relatif à la partie *moyenne mobile* non saisonnière,
- SMA : coefficient relatif à la partie *moyenne mobile* saisonnière.

Coefficients	AR	MA	SMA	Ecart type
Valeurs	-0,070	-0,13	-0,85	0.0013

TABLE 8.2 – Paramètres - Inflation

Coefficients	y_m	α_Y	β_Y	σ_Y
Valeurs	-0,63	-0,28	0,81	0,052

TABLE 8.3 – Paramètres - Rendement des dividendes

Coefficients	d_m	α_D	β_D	μ_D	ξ_D	σ_D
Valeurs	0,0034	-0.023	-0,024	-0,010	0,015	0,0039

TABLE 8.4 – Paramètres - Montant des dividendes

Nous tenons à préciser deux points importants concernant les estimations présentées dans les tableaux ci-dessus :

- Les coefficients autres que ceux estimés ci-dessus, comme les coefficients de la « partie sauts » des équations, sont déterminés de manière déterministe, à dire d’experts, en fonction de la probabilité de crises que l’on souhaite modéliser.
- L’ensemble des coefficients présentés ci-dessus ne sont pas significatifs. En effet, nous avons réalisé des tests de student pour tester la significativité de chaque coefficient et il ressort que certains ne le sont pas. Par exemple, pour l’équation du montant des dividendes, la constante d_m n’est pas significative dans le cas des États-Unis. Cependant, dans la mesure où l’*adjusted R²* ne diminuait pas, nous avons préféré conserver l’ensemble des paramètres afin de garder la structure du modèle de Wilkie. De plus, rappelons le sixième principe évoqué au chapitre trois : l’idée fondamentale ici est de trouver les paramètres optimaux et non pas d’estimer des coefficients significatifs.

Par ailleurs, comme nous l’avons signalé dans la partie relative à la modélisation des résidus, nous avons utilisé la méthode du maximum de vraisemblance pour estimer les paramètres de queue des lois de Pareto hybrides.

8.2 Méthodes de simulations

Les équations étant calibrées, il convient à présent d’effectuer les simulations. Pour cela, il est nécessaire de générer des nombres aléatoires, quelle que soit la méthode de simulation choisie.

Dans un premier paragraphe, nous rappelons les points importants concernant la génération de nombres aléatoires. Pour cela, nous conseillons au lecteur soucieux d’avoir plus de détails à ce sujet de consulter [13] . Nous nous contentons de souligner les points qui nous semblent importants.

Dans un second paragraphe, nous présentons la méthode choisie pour simuler les résidus. Enfin, nous présentons quelques trajectoires obtenues lors de la simulation du jeu de données des États-Unis.

8.2.1 Générateurs de nombres aléatoires

Tout d’abord, rappelons que la génération de nombres aléatoires uniformes est très importante pour construire un générateur de scénarios économiques car c’est elle qui apporte l’aléa dans l’algorithme. Il faut donc attacher une importance particulière à cette source d’aléa. Le modèle perdra tout son intérêt si la source d’aléa devient cyclique par exemple. Par ailleurs, rappelons qu’en informatique, l’aléatoire n’existe pas et que nous avons juste la possibilité de générer des séries qui ressemblent à l’aléa.

Il existe plusieurs types de générateurs regroupés en deux classes :

- les générateurs congruentiels linéaires ou multiplicatifs,
- les générateurs quasi aléatoires dont la construction est basée sur des suites à discrédance faible telles que les suites de Van Der Corput, Halton, Fauré et Sobol ou l’algorithme de Tore.

La discrédance d’une suite traduit le fait que celle-ci est bien répartie sur son ensemble de définition : plus la discrédance est faible, plus la répartition de la suite sur son ensemble de définition est bonne. Des études ont montré que la discrédance des suites à discrédance faible est meilleure que celle des suites pseudo-aléatoires.

Le générateur *Random* d’Excel est un générateur congruentiel. Il a été prouvé que celui-ci est moins efficace que le générateur obtenu avec les suites de Fauré ou de TORE.

Pour tester le caractère aléatoire des séries simulées, il existe plusieurs tests notamment les tests de Diehard ou les tests du National Bureau of Standards. Rotz et al ont démontré dans [24] que le générateur d’Excel réussit les tests cités ci-dessus. Ils ont noté que pour voir apparaître une répétition, il faut une série de 10^{13} tirages. Par ailleurs, ils donnent une méthode pour améliorer le processus *Random* d’Excel consistant à utiliser la fonction *Randomize* avant chaque appel de la fonction *Random*.

Cette étude réalisée par Rotz et al nous incite à utiliser la fonction *Random* d’Excel, malgré les constatations présentées dans le mémoire de J. Kouassi.

8.2.2 Générations de lois de probabilité

Disposant de nombres aléatoires, il convient à présent de modéliser des réalisations de lois de probabilité. Pour notre mémoire, nous avons besoin de modéliser des lois normales et des lois de Pareto généralisées. Pour simuler une loi normale, il existe de nombreuses méthodes. Citons les méthodes les plus connues :

- méthode d’inversion de la fonction de répartition avec l’algorithme de De Moro ou la méthode d’inversion d’Acklam.
- méthode du rejet.
- méthode de simulation par transformation de variables uniformes, notamment la méthode de Box-Muller.

Le lecteur trouvera une description précise de ces méthodes dans le mémoire de J. Kouassi [13].

Pour simplifier nos programmes, nous avons choisi d’utiliser la méthode d’inversion de la fonction de répartition. Dans le chapitre consacré à la loi de Pareto hybride, nous avons présenté une expression de la fonction de répartition inverse de la loi de Pareto hybride ainsi que de la loi de Pareto généralisée. Par ailleurs, nous utiliserons directement la fonction d’Excel *Loi normale inverse* pour simuler des réalisations de la loi normale. Nous avons tout de même cherché à coder cette fonction directement sur VBA ce qui accélère l’algorithme. Le lecteur intéressé trouvera le code en annexe.

8.3 Projections et comparaisons

8.3.1 Projections obtenues

Par souci de concision, nous ne présentons pas l'ensemble des courbes obtenues. Cela aurait un intérêt limité. Toutefois, pour illustrer nos travaux, nous souhaitons tout de même présenter quelques trajectoires. Ainsi, nous représentons une trajectoire possible concernant l'inflation et l'indice de référence pour la France, les États-Unis, le Japon et l'Allemagne dans le cas du générateur sans saut et dans le cas du générateur avec un saut sous certaines probabilités. Nous représentons ici des projections à cinq ans pour le générateur sans saut et à un an pour le générateur avec saut.

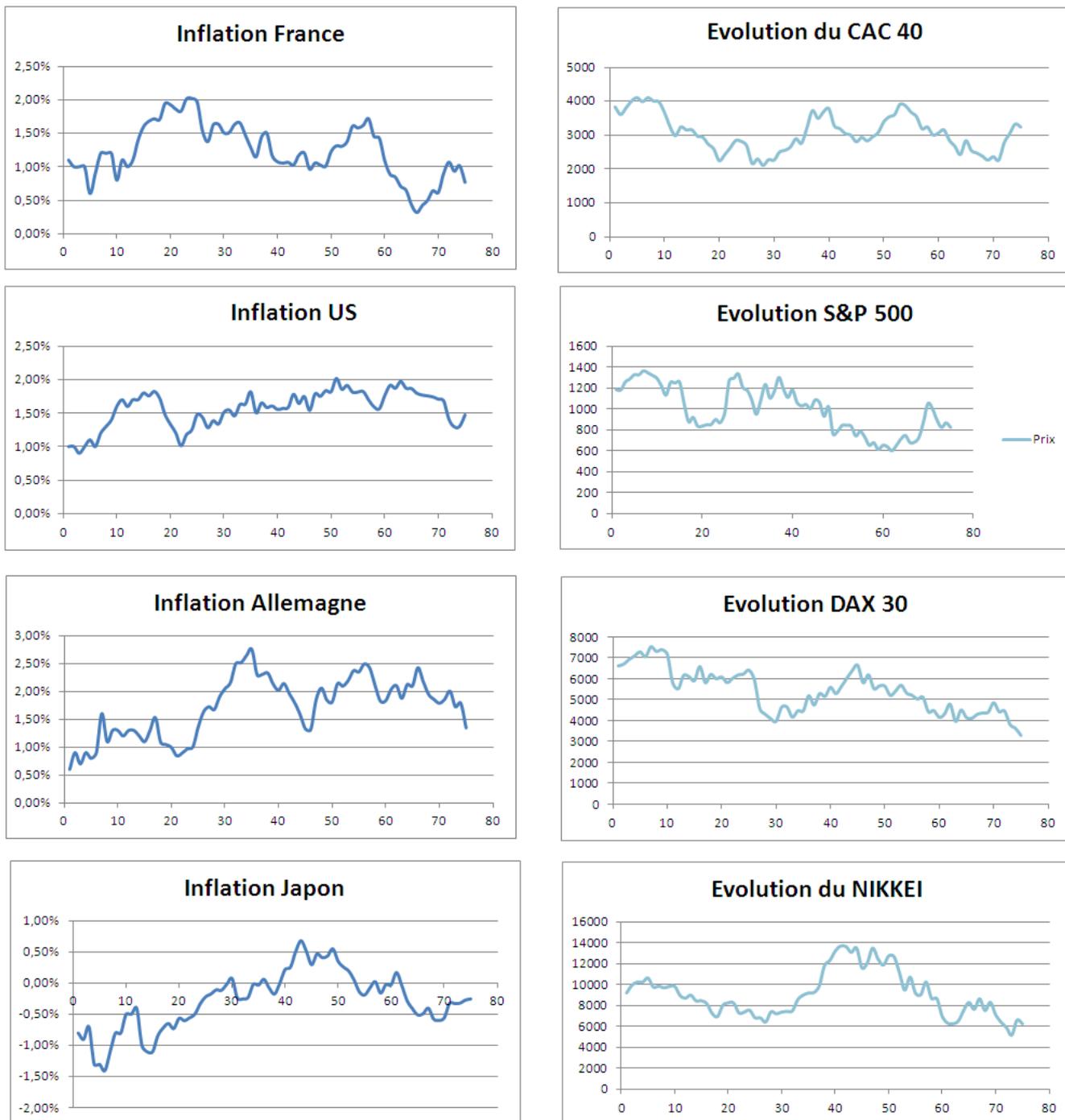


FIGURE 8.1 – Exemple de projections obtenues avec le GSE sans saut

Les trajectoires présentées ci-dessus ont été obtenues à l'aide du GSE sans saut. En réalisant la moyenne de 3000 trajectoires, nous obtenons une projection relativement stable. Les trajectoires obtenues ci-dessus sont envisageables. Cependant, nous ne pouvons pas faire plus de commentaires car il s'agit de quelques trajectoires parmi plusieurs milliers.

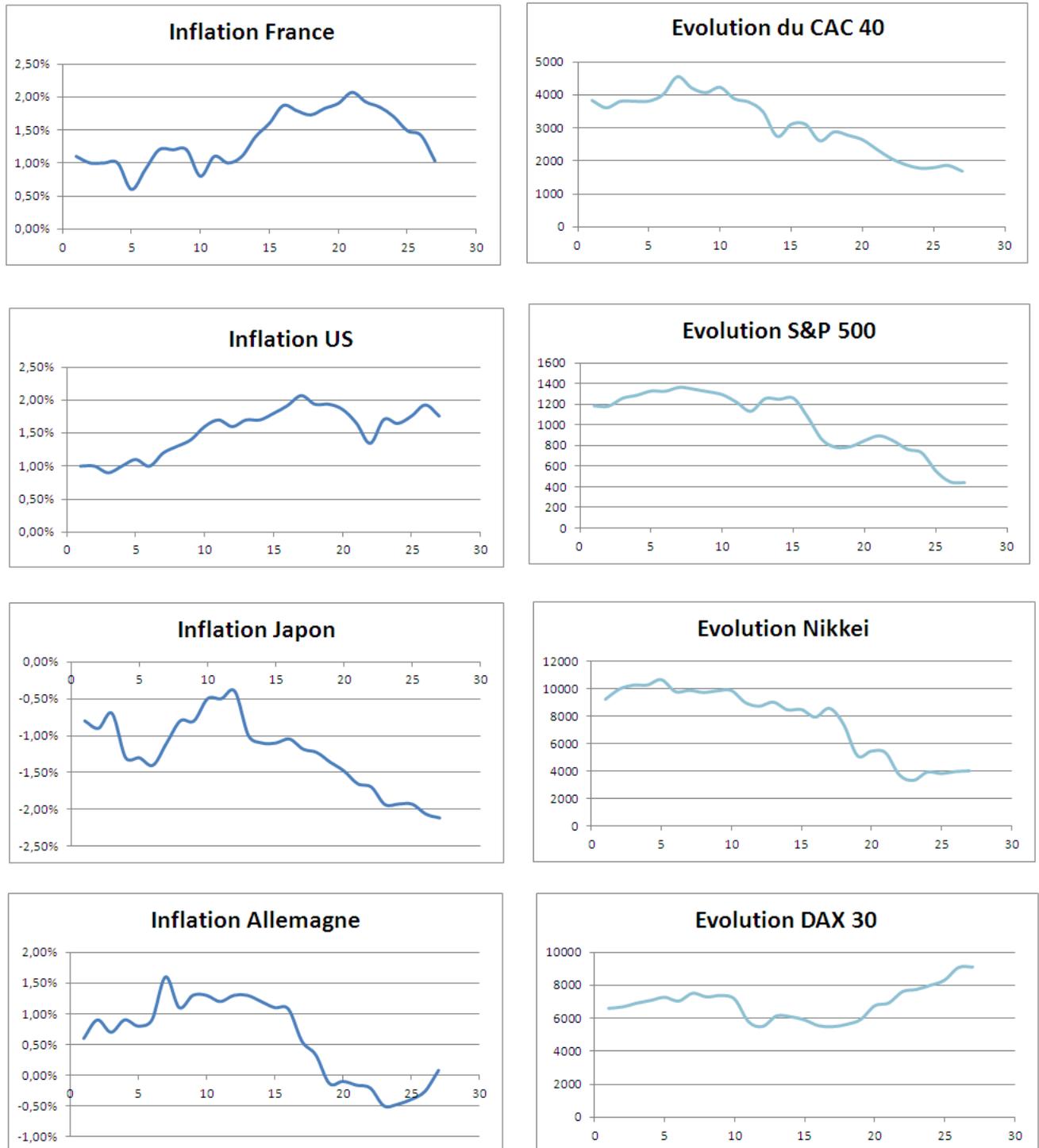


FIGURE 8.2 – Exemple de projections obtenues avec le GSE avec sauts

Les trajectoires présentées ci-dessus ont été obtenues à l'aide du GSE avec saut. Nous avons utilisé les hypothèses suivantes :

- pour l'inflation : une chute de 1% avec une probabilité de 5% et un retour en 3 ans ;
- pour les indices : une chute de 20% avec une probabilité de 5% et retour en 1 an.

Nous observons que l'inflation allemande, par exemple, devient négative dans la projection présentée et que le DAX reste à un niveau stable. Cependant, de même que pour les

projections précédentes, nous ne pouvons pas faire plus de commentaires car il s'agit de quelques trajectoires parmi plusieurs milliers.

Par ailleurs, il est intéressant de présenter les résultats de l'évolution de l'inflation obtenus en faisant la moyenne de 1000 simulations. Nous représentons ici l'évolution attendue à un an.

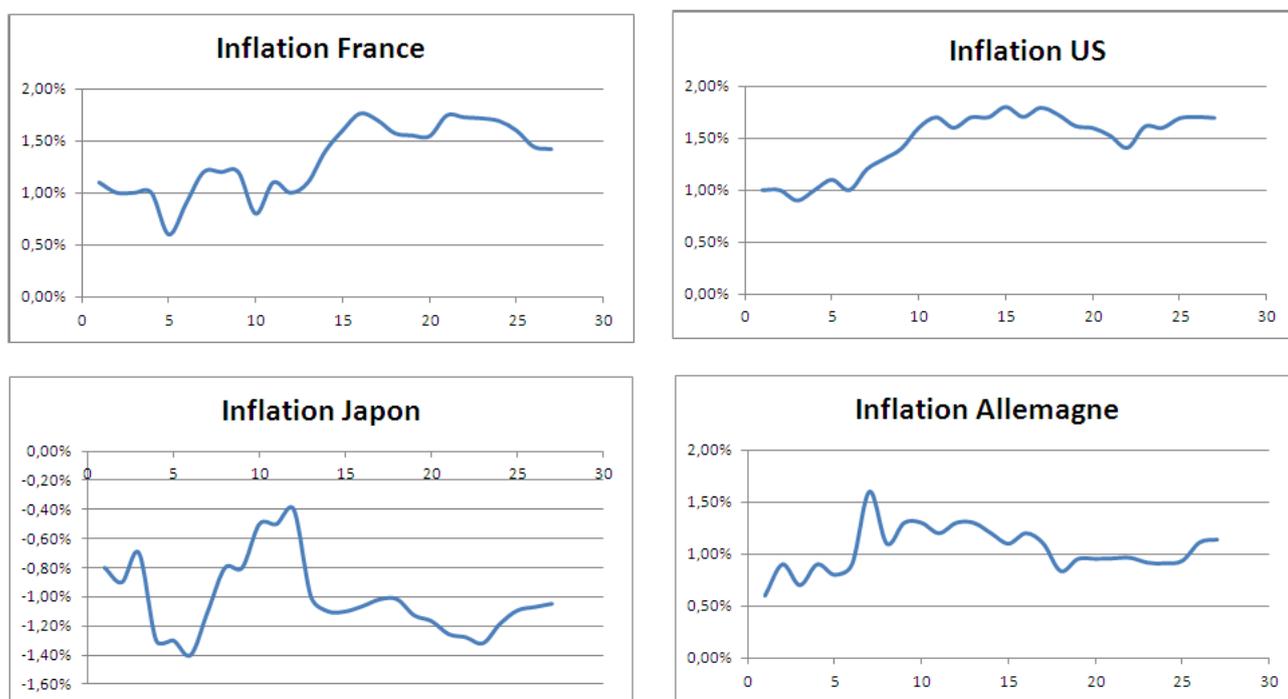


FIGURE 8.3 – Projections à 1 an

A partir de 1000 simulations, nous obtenons un résultat stable.

Les résultats présentés ci-dessus suggèrent que l'inflation dans la zone euro, après avoir remonté légèrement, va diminuer de nouveau en fin d'année. En effet, on observe qu'en France, l'indice des prix à la consommation progresse de +1,5% sur une année glissante (décembre 2011 - décembre 2012); en Allemagne, il progresse de +1,0%. Dans la crise déflationniste dans laquelle l'Europe est entrée, résultant des politiques d'austérité mises en place, ces projections semblent réalistes. Le fait que l'inflation allemande augmente moins que l'inflation française est un résultat cohérent dans la mesure où les politiques allemandes prônent une politique d'austérité beaucoup plus exigeante que celle souhaitée par la France.

Aux États-Unis, l'inflation continuera à progresser d'après nos projections. La politique de relance des États-Unis, avec notamment les « quantitative easings » (QE) avait pour but de maintenir l'activité du pays et a permis de maintenir l'inflation à un niveau assez élevé. La projection nous semble réaliste.

Enfin, d'après nos projections, le Japon restera, à horizon un an, dans la crise déflationniste initiée depuis les années 1990.

8.3.2 Comparaisons

Il est également intéressant de faire différentes comparaisons suite au choix que l'on a fait tout au long de notre étude. Dans un premier temps, nous allons comparer le pouvoir prédictif de l'équation de l'inflation. Ensuite, nous évaluerons les différences constatées en modélisant les résidus par une loi de Pareto hybride ou pas une loi normale.

Pouvoir prédictif

Différents critères peuvent être utilisés pour estimer le pouvoir prédictif d'un modèle. Nous présentons ici les résultats obtenus pour trois critères classiques :

- la somme des carrés des résidus (SCR)
- l'erreur quadratique moyenne (MSE)
- la racine du carré moyen des erreurs (RMSE).

Nous avons comparé les résultats obtenus en utilisant le processus autorégressif d'ordre un proposé par Wilkie (1985) et notre modélisation SARIMA. Pour cette comparaison, nous avons conservé la modélisation des résidus par une loi normale afin de pouvoir évaluer uniquement l'impact du changement du processus.

Voici les résultats obtenus pour la France et les États-Unis :

Indicateurs	Modélisation	AR(1)	SARIMA
	SCR		2,8045
MSE		0,0206	$2,26 \cdot 10^{-4}$
RMSE		0,1436	0,015

TABLE 8.5 – Pouvoir prédictif - Inflation France

Indicateurs	Modélisation	AR(1)	SARIMA
	SCR		1,33
MSE		0,0094	$1,67 \cdot 10^{-6}$
RMSE		0,097	0,0013

TABLE 8.6 – Pouvoir prédictif - Inflation US

Les deux tableaux précédents montrent que le pouvoir prédictif du modèle retenu est supérieur à celui du modèle proposé par Wilkie, ce qui confirme que notre modélisation est plus appropriée. En effet, tous les indicateurs calculés pour notre modèle sont inférieurs à ceux du modèle initial ce qui traduit un pouvoir prédictif supérieur. Par exemple, la comparaison des *Roots Mean Square Error* (RMSE) obtenues pour le jeu de données de la France est cent fois supérieur avec la modélisation que l'on a retenue. Nous pouvons faire la même remarque pour les États-Unis.

De plus, nous pouvons noter que les projections obtenues ne sont pas très différentes de celles obtenues avec le générateur de Wilkie et qu'elles suivent la même tendance (Nous ne les représentons pas ici). Cela nous permet d'affirmer que notre objectif initial est respecté : en effet, en améliorant certains aspects du modèle, nous sommes tout de même parvenus à conserver sa structure.

Pareto hybride vs Loi normale

Il est également intéressant de comparer les résultats obtenus en simulant les résidus par des lois Pareto hybrides et par des lois gaussiennes. Nous constatons que les résultats ne sont pas très différents, ce qui confirme que notre modélisation des résidus en utilisant la loi Pareto hybride n'est pas aberrante. De surcroît, ce choix possède l'avantage d'être modifiable facilement en ajustant les paramètres de queue. Pour illustrer notre propos, voici l'évolution de l'inflation lorsque l'on suppose que les résidus sont simulés avec une loi gaussienne :

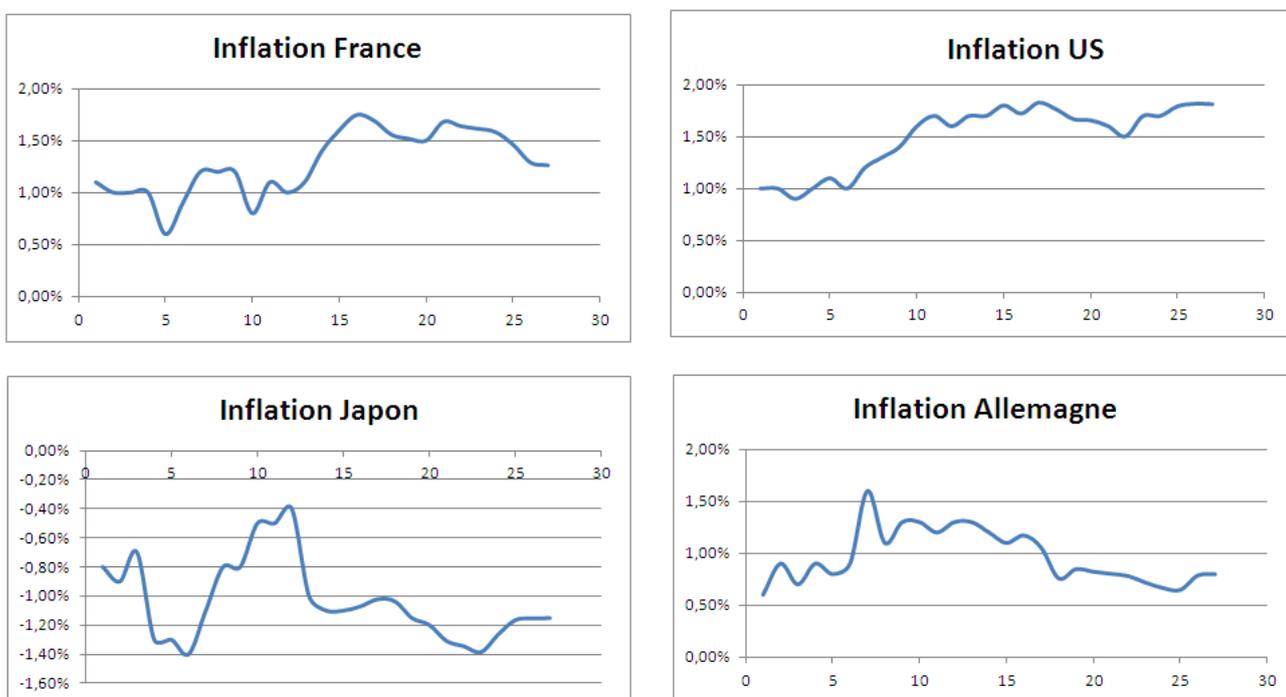


FIGURE 8.4 – Projections à 1 an avec la loi normale

En comparant ces courbes avec celles de la figure 8.3, nous observons que les évolutions sont semblables. Notons que nous faisons la même remarque concernant les projections relatives aux indices, c'est la raison pour laquelle nous ne les avons pas présentées ici.

En conclusion de cette partie, nous pouvons considérer que les changements apportés au GSE initial sont pertinents dans la mesure où les projections obtenues sont réalistes, comme en témoignent les projections de l'inflation à un an ou les résultats de l'application proposée au chapitre suivant. Par ailleurs, notons que le GSE obtenu est robuste : en effet, en faisant la moyenne de 1000 simulations, nous obtenons des projections stables.

Chapitre 9

Optimisation d'un portefeuille

Ce chapitre vise à exploiter le générateur de scénarios économiques créé et à l'appliquer pour optimiser un portefeuille d'actions internationales. Nous présentons ici une méthode pour calculer l'utilité d'un portefeuille afin d'améliorer son rendement. Notons que la méthode appliquée n'est pas classique. En effet, elle consiste à choisir les actions résistant le mieux aux crises : on désignera par l'expression *actions résilientes* ce type d'actions. Pour cela, nous projetons tous les titres en intégrant des probabilités de crises majeures. Après avoir rappelé la définition ainsi que les méthodes de calcul de la *value at risk* (VaR), nous présenterons précisément la démarche de gestion flexible par les risques mise en place pour optimiser le portefeuille du fonds.

9.1 Les indicateurs de risque

9.1.1 Variance, *skewness* et *value at risk*

Traditionnellement, on distingue trois indicateurs de risque : la variance, le *skewness* et la *value at risk*. Nous avons choisi de ne pas retenir la variance car nous considérons que cet indicateur n'est pas pertinent, qui plus est en période de crise. En effet, la variance ne fait pas de distinction entre les variations négatives et positives. Par ailleurs, cet indicateur possède beaucoup de bruit dans les périodes de crises, ce qui rend son estimation délicate. Nous privilégierons donc une combinaison de la VaR et du *skewness* ; nous critiquerons toutefois cette approche qui a tendance à favoriser les pays à forts stabilisateurs automatiques.

Un coefficient positif du *skewness* traduira le fait que les rendements sont étalés vers la droite ; un coefficient négatif, qu'ils sont étalés vers la gauche.

Par ailleurs, nous avons choisi de considérer la VaR plutôt que ces dérivés par souci de simplicité. En effet, il est plus facile d'expliquer à des investisseurs potentiels, non statisticiens, le principe de la VaR plutôt que celui des dérivés de la VaR comme la TVaR. Notons toutefois que l'utilisation de la VaR introduit déjà le concept de perte maximale, ce qui nous permet d'introduire, au paragraphe 9.2, le principe des *actions résilientes*.

9.1.2 Calcul de la *value at risk*

Le calcul du *skewness* ne pose pas de problème. Celui de la VaR est plus délicat. En effet, il existe plusieurs méthodes pour la calculer : la méthode delta normale, l'approxi-

mation delta gamma, la méthode du bootstrap paramétrique ou encore celle du bootstrap non paramétrique. Toutes les méthodes sont décrites de manière précise dans [11] qui traite de l'utilité de la VaR et de son calcul.

Après avoir testé l'ensemble de ces méthodes qui donnent des résultats similaires, nous avons choisi la méthode du bootstrap non paramétrique. Ce choix résulte notamment du fait que l'on simule les résidus par une loi Pareto hybride et non plus une loi normale. Un point est fait en annexe sur cette méthode.

Par ailleurs, nous souhaitons rappeler la définition de la *value at risk* qui est parfois méconnue. Une définition littéraire est la suivante : la *value at risk* représente la perte potentielle maximale d'un investisseur sur la valeur d'un actif ou d'un portefeuille d'actifs financiers compte tenu d'un horizon de détention et d'un intervalle de confiance.

Mathématiquement, cela se traduit de la façon suivante :

$$\mathbb{P}(X \leq VaR) = p$$

Où, dans notre cas, X représente la distribution des rendements et p une probabilité à attribuer de manière déterministe. Classiquement, p est souvent égal à 5% ou 10%.

9.2 Actions résilientes

Rappelons que le fonds pour lequel l'outil a été créé est un fonds d'actions internationales ayant pour caractéristique principale de bien résister aux crises. Pour cela, il doit être constitué d'actions baissant peu en période de crise : c'est ce qu'on appelle les *actions résilientes*. De fait, le processus d'optimisation mis en place visera à tester la résistance des actions à différents types de chocs.

Actuellement, le gérant du fonds tente d'identifier les *actions résilientes* en faisant une analyse macroéconomique du marché et en calculant certains indicateurs. Le modèle proposé dans ce mémoire a pour objectif de compléter l'analyse macroéconomique déjà réalisée et n'a pas vocation à se substituer à cette dernière. En effet, la recherche d'*actions résilientes* nécessite une double approche, comme nous l'avons indiqué en introduction : une approche *top down* suivie d'une approche *bottom up*.

9.3 Fonction d'utilité

A partir des deux indicateurs présentés, nous souhaitons calculer l'utilité de chaque titre afin de maximiser l'utilité du portefeuille. L'expression de la fonction d'utilité retenue s'écrit de la manière suivante :

$$U = \text{Espérance} + a \times \text{Skewness} + b \times \text{Value-at-risk}$$

Les valeurs de a et b sont déterminées en considérant deux éléments : le backtesting et la valeur des indicateurs avancés de l'OCDE. Ces coefficients peuvent être modifiés en fonction de la période dans laquelle nous sommes. De plus, nous avons considéré les règles suivantes :

- $a > 0$: en effet, si le *skewness* est négatif, cela traduit le fait que les rendements sont étalés à gauche de la moyenne. Ainsi, il nous semble légitime que cela fasse diminuer l'utilité. A l'inverse, un *skewness* positif doit faire augmenter l'utilité du titre.
- $b < 0$: la VaR correspond à la perte maximale sous une certaine probabilité. Ainsi, plus elle est faible, plus cela traduit le caractère résilient de l'action et donc plus cela augmente l'utilité. Les VaR calculées étant négatives, il faut donc que le coefficient b soit également négatif.

Remarque : *Comment déterminer de manière efficace les valeurs de a et b ? Quelles sont les valeurs retenues ?*

Pour les calculs réalisés dans ce mémoire, nous avons retenu les valeurs suivantes : $a = 1/2$ et $b = -1/2$. Nous sommes conscients que ces valeurs peuvent être remises en cause. Cependant, il s'agit de celles utilisées par le gérant pour gérer la SICAV. Elles ont été estimées par backtesting et sont cohérentes avec les contraintes que l'on a exposées. Une méthode de détermination de ces valeurs est délicate à mettre en place. Néanmoins, nous pouvons modifier les paramètres en suivant les principes suivants : si l'économie s'améliore, plus précisément, si les indicateurs avancés indiquent un retournement de l'économie, nous diminuerons la valeur du coefficient de la VaR et augmenterons la valeur du coefficient d'asymétrie. Nous ferons l'inverse si l'économie se détériore. Cependant, nous n'avons pas fait de recherche plus précise à ce sujet, qui n'est pas le coeur de notre mémoire. Nous avons cependant effectué quelques tests pour tester la robustesse des paramètres.

Notons également qu'il est possible d'utiliser une fonction d'utilité totalement différente de celle que l'on a présentée ci-dessus. Cela ne remet pas en cause la méthode d'optimisation que l'on va exposer.

9.4 Méthode d'optimisation

Les parties suivantes visent à présenter la méthode d'optimisation mise en place. Dans un premier temps, nous présentons la méthode générale. Ensuite, nous abordons plus particulièrement le calcul de l'utilité du portefeuille. Enfin, nous présentons certains résultats obtenus.

9.4.1 Processus global d'optimisation

Le générateur de scénarios présenté est intégré dans une étape particulière du processus d'optimisation. En effet, d'autres étapes sont nécessaires pour mener à bien la gestion du portefeuille. Nous présentons ci-dessous les étapes que l'on a suivies. Notons que les étapes une et deux sont réalisées en grande partie par le gérant du fonds.

Etape 1 : Choisir un univers d'investissement

Dans un premier temps, il convient de limiter de façon claire l'univers d'investissement. Le fonds pour lequel l'outil est construit est focalisé sur les secteurs liés aux domaines suivants : l'énergie, les matières premières rares et/ou précieuses, les matières premières alimentaires, la santé, l'eau, la lutte contre le réchauffement climatique et la construction d'infrastructures.

Etape 2 : Sélection de valeurs

Dans un second temps, il faut sélectionner une certaine quantité de valeurs en fonction de différents critères qualitatifs : connaissance des entreprises, expérience ou secteur connu. A ce stade, nous sélectionnons 40 à 50 valeurs. Afin d'apporter une partie quantitative à cette étape, nous proposons le recours aux indicateurs avancés et aux modélisations de la croissance du PIB et de la croissance des dettes présentés au chapitre sept. En fonction des projections obtenues, nous privilégions les domaines cycliques, défensifs ou offensifs.

Etape 3 : Projections des valeurs sélectionnées

A l'aide du générateur de scénarios, nous pouvons projeter les titres sélectionnés dans l'étape précédente.

Etape 4 : Calcul de l'utilité

Nous proposons le détail de cette étape dans le paragraphe suivant. Nous calculons l'utilité de chaque titre, puis, en fonction des valeurs trouvées, nous optimisons le portefeuille en maximisant son utilité.

9.4.2 Calcul de l'utilité

Comme nous l'avons indiqué dans l'introduction de ce mémoire, à partir du principe du générateur intégrant des possibilités de sauts, nous avons construit une démarche de recherche d'*actions résilientes*. Ainsi, nous soumettons chaque action du portefeuille à cinq différents types de crises. Les probabilités à affecter à chaque scénario dépendent de la conjoncture et sont données par le gérant du fond. Les cinq différents types de crises sont les suivants :

- crises extraordinaires
- accidents inflationnistes
- récessions ordinaires
- période de réduction des dettes sans restructuration
- période de réduction des dettes avec restructuration

Notons que l'utilisateur de l'outil créé a la possibilité de rajouter ou de supprimer certains types de crises, en fonction de la conjoncture.

Pour plus de clarté, le schéma ci-dessous illustre la méthode générale mise en place concernant le calcul de l'utilité :

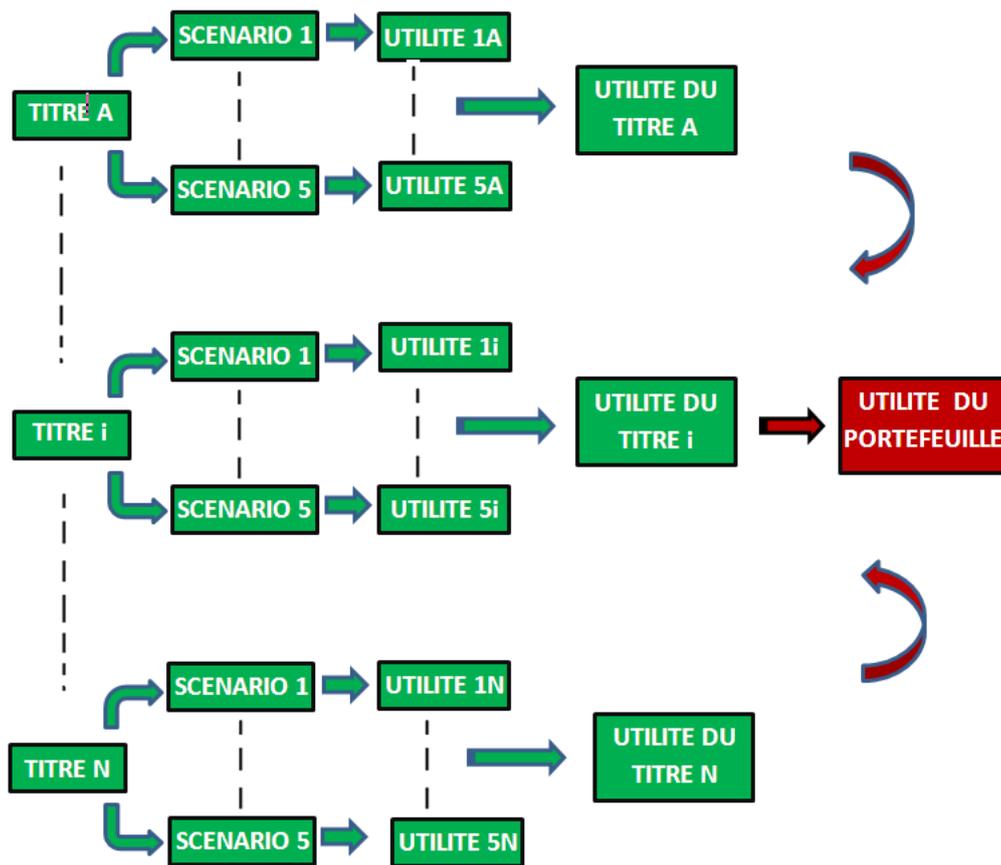


FIGURE 9.1 – Processus d’optimisation

La partie suivante est consacrée aux résultats obtenus.

9.4.3 Résultats

Dans la continuité de ce qu’on a fait tout au long de ce mémoire, nous avons choisi de présenter les résultats obtenus pour les quatre indices : États-Unis, France, Allemagne et Japon.

Les cinq générateurs de scénarios sont pondérés avec des valeurs qui varient en fonction de la conjoncture. Ainsi, les valeurs utilisées pour les calculs présentés dans ce mémoire seront vraisemblablement différentes de celles utilisées dans trois ou six mois. Elles proviennent des hypothèses réalisées dans la société pour gérer le fonds pour lequel l’outil a été construit. Précisons que l’objectif de ce mémoire n’est pas d’étudier la pertinence des pondérations présentées.

Les deux tableaux ci-dessous donnent les hypothèses que l’on a utilisées pour les calculs réalisés :

	Pondération	Choc inflation	Choc actions
Crises extraordinaires	10%	-1%	-40%
Accidents inflationnistes	15%	2%	-20%
Récessions ordinaires	45%	0%	-20%
Période de réduction des dettes sans restructuration	20%	0%	-10%
Période de réduction des dettes avec restructuration	10%	0%	-10%

TABLE 9.1 – Hypothèses sur les scénarios - Zone euro et États-Unis

	Pondération	Choc inflation	Choc actions
Crises extraordinaires	10%	0%	-40%
Accidents inflationnistes	5%	3%	-20%
Récessions ordinaires	45%	0%	-20%
Période de réduction des dettes sans restructuration	20%	0%	-10%
Période de réduction des dettes avec restructuration	20%	0%	-10%

TABLE 9.2 – Hypothèses sur les scénarios - Japon

A partir de ces hypothèses, nous avons calculé l'utilité de chaque indice. Pour illustrer ces calculs, nous donnons ci-dessous le détail des résultats obtenus pour la France et l'Allemagne.

	Value at risk	Skewness	Moyenne	Utilité
Scénario 1	-0,13	0,61	-0,068	0,30
Scénario 2	-0,060	0,72	-0,042	0,35
Scénario 3	-0,15	0,44	-0,11	0,19
Scénario 4	-0,049	0,40	-0,029	0,19
Scénario 5	-0,059	-0,041	-0,018	-0,0092
TOTAL	-0,11	0,44	-0,069	0,20

TABLE 9.3 – Résultats France

	Value at risk	Skewness	Moyenne	Utilité
Scénario 1	-0,14	0,06	-0,07	0,03
Scénario 2	-0,07	0,40	-0,05	0,19
Scénario 3	-0,18	0,07	-0,11	0,02
Scénario 4	-0,05	-0,01	-0,03	-0,01
Scénario 5	-0,04	-0,67	-0,02	-0,33
TOTAL	-0,12	0,03	-0,07	0,0029

TABLE 9.4 – Résultats Allemagne

Nous observons que les résultats obtenus sont cohérents.

En effet, les valeurs des VaR évoluent suivant les mêmes ordres de grandeurs pour la France et l'Allemagne, de même que les moyennes.

Par contre, nous observons que le *skewness* relatif à l'indice de la France est supérieur quel que soit le scénario considéré. Cela traduit le fait que, selon le modèle, la distribution du titre français est plus étalée vers la droite que celle du titre allemand. Ainsi, avec la fonction d'utilité choisie, il est légitime d'obtenir une utilité supérieure pour l'indice français. Cela traduit le fait que selon le modèle, l'indice français est plus résilient que l'indice allemand.

Le tableau suivant propose le résultat de l'utilité obtenue pour les quatre pays étudiés :

Pays	Utilité
France	20 %
États-Unis	7,9 %
Allemagne	0,29 %
Japon	-14 %

TABLE 9.5 – Utilité des indices

Ces résultats suggèrent le fait que l'indice de la France est le plus résilient.

De prime abord, ce résultat peut paraître étonnant : en effet, actuellement, l'Allemagne semble plus à même d'affronter les crises que la France, en témoigne le niveau très bas de l'obligation d'État allemand qui donne même parfois un rendement réel négatif ! Cependant, il est possible d'expliquer ce résultat qui tend à remettre en cause l'utilisation du *skewness* dans le calcul de l'utilité. En effet, la France, à l'inverse des trois pays présentés ci-dessus, possède des piliers stabilisateurs : le système de sécurité sociale français est très protecteur et cela peut s'avérer utile en période de crise, notamment pour éviter les crises sociales. Ainsi, l'optimisation d'un portefeuille à l'aide du *skewness* favorise les pays à forts stabilisateurs automatiques. Afin de tester la cohérence de ce résultat, nous avons réalisé le calcul d'utilité des indices sur des périodes antérieures. Ce calcul a confirmé que l'indice français avait un *skewness* plus élevé que les autres indices étudiés.

Par ailleurs, afin de tester l'influence des paramètres de la fonction d'utilité, nous avons testé la sensibilité des résultats à des modifications de ces paramètres. Voici les résultats trouvés pour la France et l'Allemagne :

Coefficients (a,b)	Utilité France	Utilité Allemagne
(0,3 ; -0,7)	14 %	2,0 %
(0,4 ; -0,6)	17 %	1,1 %
(0,5 ; -0,5)	20 %	0,29 %
(0,6 ; -0,4)	24 %	-0,71 %
(0,7 ; -0,3)	27 %	-1,62 %

TABLE 9.6 – Tests - fonction d'utilité

Nous observons que quels que soient les paramètres choisis, l'utilité obtenue pour l'indice français est supérieure à celle obtenue pour l'indice allemand. Or, ce qui nous intéresse ici n'est pas l'écart entre la valeur de l'utilité des indices mais la relation d'ordre. La fonction d'utilité semble donc robuste.

Toutefois, nous avons conscience que la conclusion de cette étude dépend de la fonction d'utilité choisie et qu'elle peut être remise en cause. Par exemple, en utilisant le kurtosis, autre indicateur caractéristique d'une distribution, nous privilégierons les États-Unis, puis l'Allemagne et enfin la France. Par ailleurs, en considérant la variance, nous obtenons des résultats différents : en effet, nous privilégierons dans ce cas, l'Allemagne, puis la France et enfin les États-Unis.

Avant de conclure, nous tenons donc à rappeler que le but de ce mémoire n'était pas l'interprétation des résultats ou encore la détermination d'une fonction d'utilité complexe, mais simplement la réalisation de l'outil permettant d'obtenir les projections.

Conclusion et perspectives

Le but de ce mémoire était de présenter un modèle de générateur de scénarios économiques adapté à l'environnement à risques croissants de ce début du XXI^{ème} siècle. Ce modèle était destiné à l'optimisation d'un portefeuille d'actions et à la détection d'*actions résilientes* ; la nécessité de la recherche de telles actions se justifiant par le contexte politico-économique mondial.

Afin de mener à bien ce projet, nous avons suivi les étapes mentionnées ci-dessous :

- Application de l'algorithme de Box et Jenkins pour la modélisation de l'inflation ainsi que pour la réalisation de tests sur les autres équations du modèle.
- Application et comparaison de différentes méthodes d'inclusion de sauts dans une série temporelle : adaptation de l'analyse d'intervention de Box et Tiao en probalissant certains paramètres et en déterminant « à dire d'experts » les autres.
- Modélisation des résidus des équations par la loi Pareto hybride.
- Etude précise des indicateurs avancés de l'OCDE et de l'IFO afin d'anticiper l'évolution de la conjoncture.
- Projections en utilisant le GSE de Wilkie modifié.
- Calcul de l'utilité de plusieurs titres.

Les résultats obtenus sont satisfaisants. En effet, l'application du modèle donne des résultats cohérents et plus pertinents que les résultats obtenus avec le générateur de Wilkie non modifié. Cela s'explique notamment par le fait que le générateur de Wilkie était adapté pour un horizon long terme.

Une nouveauté que l'on a présentée est la modification de la loi des résidus des équations. Cette dernière nous semblait indispensable dans la mesure où les tests de gaussianité étaient tous rejetés. En outre, pour le générateur créé, il est important de pouvoir modifier l'épaisseur des queues de distribution, même si cela augmente le nombre de paramètres à prendre en compte.

Par ailleurs, nous avons rencontré différents types de difficultés qui nous ont contraints de modifier notre projet initial. Par exemple, les modélisations de la croissance du PIB et de la croissance des dettes n'ont pas été aussi pertinentes qu'escompté et le recours aux indicateurs avancés a finalement été privilégié. Le fait que l'horizon de projection soit de un an (ou cinq ans) est une des causes des problèmes rencontrés ; le fait que l'aspect politique soit très présent, dans l'évolution des dettes notamment, en est une autre. L'étude réalisée

sur les indicateurs a prouvé leur efficacité en période de crise. Selon nous, leur utilisation est plus appropriée que toute tentative de modélisation, qui plus est sur un horizon court ou moyen terme.

Enfin, dans le but de poursuivre cette étude, visant à optimiser un portefeuille, il pourrait être envisageable d'étudier d'autres concepts mathématiques, notamment les copules. En effet, l'approche présentée dans ce mémoire consiste à optimiser un portefeuille d'actions et n'est pas adaptée à un portefeuille diversifié contenant de l'immobilier, des obligations et d'autres actifs. Pour de tels portefeuilles, il pourrait être envisageable de garder la structure du modèle de Wilkie ainsi que les modifications que l'on a apportées et de modéliser une interdépendance entre les résidus en utilisant la copule de Clayton. L'utilisation de cette copule se justifie par le fait qu'elle permette la modélisation d'une dépendance dyssymétrique. Par ailleurs, comme nous l'avons évoqué, il serait intéressant d'étudier un modèle de taux différent de celui proposé par Wilkie, en essayant par exemple d'adapter le modèle HJM. Ces tentatives de modélisation pourront faire l'objet d'une future recherche ou d'un prochain mémoire.

Bibliographie

- [1] A.D. WILKIE, M.A., F. F. *More on a stochastic asset model for actuarial use*. 1995.
- [2] ARNAUD CLEMENT-GRANDCOURT, J. J. *Gestion des risques financiers et papillons noirs - méthodes qualitatives*. Hermes, 2009.
- [3] BERNAY, A. Rentabilité des actifs à long terme et risque inflation : enjeux de modélisation pour l'assurance, 2008.
- [4] BLATTBERG, R. C. *Evaluation of the Power of the Durbin-Watson Statistic for Non-First Order Serial Correlation Alternatives*, vol. 55. MIT Press, 1973.
- [5] BRUNO TISSOT, L. S. Revisiting recent productivity developments across oecd countries, 2005.
- [6] CHEVALIER, F.-X. *Le bonheur économique*, vol. 55. Albin Michel, 1998.
- [7] G. FULOP, G. G. *Transition of the OECD CLI system to a GDP-based business cycle target*. OECD Publishing, Mars 2012.
- [8] GEHIN, W. Modélisation des queues de distribution des rendements des actifs financiers. application à la mesure et à la gestion du risque de marché , 2012.
- [9] HERVÉ, P. *Allocation d'actifs - Théories et pratiques*. Economica, 2007.
- [10] JACQUES ANAS, L. F. *Detecting cyclical turning points : the abcd approach and two probabilistic indicators*, vol. 2. 2004.
- [11] JORION, P. *Value at risk*. Third édition, 2007.
- [12] KEITA, C. Génération de scénarii stochastiques économiques en vue d'une étude d'allocation d'actifs en assurance-vie, 2011.
- [13] KOUASSI, J. Utilisation de générateurs quasi-aléatoires dans la valorisation des stock options avec conditions de performance, 2011.
- [14] KOUSNETZOFF, N. Croissance économique mondiale : un scénario de référence à l'horizon 2030, 2001.
- [15] LAURENT FERRARA, D. G. Analyse d'intervention et prévisions. problématique et application des données de la ratp, 1999.
- [16] MANAGEMENT, S. A. Croissance et taux réels aux états-unis, 2003.
- [17] MARCELLE CHAUVET, S. P. *Coincident and leading indicators of the stock market*, vol. 1. 2000.
- [18] MICHEL AGLIETTA, LAURENT BERREBI, A. C. *Banques centrales et globalisation*. Groupama asset management, 2009.

- [19] NICOLAS HUCHET, A. M. E. J. R. Générateur de scenarii économiques et applications, 2011.
- [20] PLANCHET, THÉRON, K. *Scénarios économiques en assurance*. Economica, 2009.
- [21] TALEB, N. N. *The Black Swan - The impact of the highly improbable*. 2007.
- [22] TIAO, G. *Autoregressive moving average models, intervention problems and outlier detection in time series*, vol. 5. Handbook of statistics, 1985.
- [23] VINCENT ALHENC-GELAS, JEAN-BATISTE BERNARD, B. R. W. Z. Recherche autour des indicateurs avancés de récessions aux etats-unis, 2010.
- [24] WENDY ROTZ, ERIC FALK, A. J. A comparison of random number generators used in business, 2001.
- [25] WILKIE, D. *A stochastic investment model for actuarial use*. 1985.
- [26] WILKIE, CAIRNS, K. S. Revisiting the wilkie investment model, 2008.

Lexique

Définitions

Approche *bottom up* : méthode consistant à sélectionner des titres sans attacher trop d'importance à l'examen d'un secteur industriel ou à l'étude de la conjoncture économique d'ensemble d'une société.

Approche *top down* : méthode utilisée par plusieurs gérants consistant à penser que la conjoncture économique générale ainsi que la vigueur d'un secteur industriel donné ont un impact considérable sur le rendement des titres et qu'il est nécessaire de réaliser une recherche à ces deux niveaux d'analyse afin de choisir des valeurs pour son portefeuille.

Backtesting : procédure qui consiste à tester un modèle ou une méthode en l'appliquant aux données historiques.

Core *inflation* : inflation dont plusieurs éléments extrêmement fluctuants comme les matières premières agricoles ou énergétiques ont été retirés.

Discrépance d'une suite : critère caractérisant la répartition des valeurs d'une suite sur son ensemble de définition.

Indicateurs de risque : mesure qui tente de résumer, par un nombre unique, le risque lié à l'incertitude sur la valeur future d'une variable aléatoire.

Loi de Pareto généralisée : loi de probabilité introduite par la théorie des valeurs extrêmes pour modéliser l'excédent d'une variable aléatoire par rapport à un seuil suffisamment élevé.

Loi de Pareto hybride : loi de probabilité consistant en une loi gaussienne dont les deux queues ont été tronquées et remplacées par des lois de Pareto généralisées.

Skewness : coefficient de dissymétrie, correspond à une mesure de l'asymétrie de la distribution d'une variable aléatoire réelle.

Kurtosis : coefficient d'aplatissement, correspond à une mesure de l'aplatissement de la distribution d'une variable aléatoire réelle.

Stabilisateur automatique : mécanisme structurel qui fait de la dépense publique un soutien de la conjoncture durant les périodes de ralentissement de l'activité économique et un frein en période d'accélération : les impôts tendent spontanément à diminuer en même temps que l'activité, tandis que les prestations sociales tendent spontanément à augmenter (indemnisation du chômage notamment), l'évolution étant inverse en cas d'emballement.

Valeur résiliente : titre financier ayant comme caractéristique principale d'être performant en période de crises.

Acronymes

ACF et PACF : Fonction d'autocorrélation et Fonction d'autocorrélation partielle.

AIC : Akaike Information Criterion.

CLI : Composite Leading Indicator.

ECRI : Economic Cycle Research Institute.

IFO : Institut für Wirtschaftsforschung.

IPI : Indice des Prix à la Consommation.

NBER : National Bureau of Economic Research.

OCDE : Organisation de coopération et de développement économique.

PIB : Produit Intérieur Brut.

SARIMA : Autoregressive integrated moving average.

SICAV : Société d'Investissement à Capital Variable.

VaR : Value at risk.

Annexe A

Description des tests utilisés

Tests de stationnarité

1. Les tests de racine unitaire

- Le test de Dickey-Fuller

Nous présentons ci-dessous le principe du test de Dickey-Fuller à l'ordre un. Le test permet de déceler la présence d'une racine supérieure ou égale à l'unité dans le polynôme caractéristique du processus lorsque celui-ci s'écrit sous une des formes suivantes :

- $X_t = a_1 \times X_{t-1} + \epsilon_t$
- $X_t = c + a_1 \times X_{t-1} + \epsilon_t$
- $X_t = c + b \times ta_1 \times X_{t-1} + \epsilon_t$

Nous testons :

- $\mathcal{H}_0 : a_1 = 1$
- $\mathcal{H}_1 : a_1 < 1$

En pratique, il suffit de soustraire X_{t-1} aux deux membres des équations précédentes puis d'estimer les coefficients par les moindres carrés et enfin de tester la nullité du coefficient $a_1 - 1$ par le test de Student classique. Nous ne détaillons pas plus, le lecteur intéressé trouvera une littérature abondante sur internet et dans les ouvrages spécialisés.

- Le test de Dickey-Fuller augmenté

C'est le même type de test sauf que la condition laisse place au cas plus général où le processus peut être un AR(p-1). Le test vérifie la présence d'une racine unitaire dans le polynôme caractéristique. Nous considérons donc que si la p-value est inférieure à 5%, nous rejetons l'hypothèse nulle au profit de l'hypothèse alternative qui énonce que la racine est strictement inférieure à l'unité et donc que la série est stationnaire.

- Le test de Philipps Perrond

Phillips et Perrond (1988) proposent une correction non paramétrique au test de Dickey-Fuller simple afin de régler le problème de l'autocorrélation et de l'hétéroscédasticité des erreurs. Afin d'éliminer les paramètres de nuisance, associés à l'existence de corrélations dans la composante stochastique du processus générateur de données, qui perturbent les résultats des tests de racine unitaire de Dickey-Fuller, Phillips et Perron suggèrent de joindre à la statistique de Student du coefficient autorégressif, un facteur de correction,

fondé sur des estimateurs convergents des paramètres de nuisance, qui élimine cette dépendance asymptotique. Il se déroule en deux étapes :

- estimation par les moindres carrés ordinaires des trois modèles de base des tests de Dickey-Fuller et calcul des statistiques associées
- estimation d'un facteur correctif établi à partir de la structure des covariances des erreurs des modèles précédemment estimée de telle sorte que les transformations réalisées conduisent à des distributions identiques à celles de Dickey-Fuller standard.

2. Le test de Kwiatkowski, Phillips, Schmidt et Shin (KPSS)

Contrairement aux trois tests ci-dessus, l'hypothèse nulle de ce test est celle de la stationnarité (autour d'une constante ou d'une tendance déterministe linéaire). La statistique du test utilisée correspond à la statistique du test du score lorsque les bruits sont indépendants, identiquement distribués et suivent une loi normale. Cependant, cette statistique est corrigée de façon à tenir compte de l'autocorrélation des résidus dans le cas général.

Tests d'indépendance

- Les tests de type *portemanteau*

Afin de tester l'indépendance des résidus issus d'une modélisation, les *tests du portemanteau* sont couramment utilisés : il s'agit des tests de Ljung-Box et de Box-Pierce. Ces tests visent à évaluer la corrélation entre les résidus et se fondent sur les estimateurs classiques de la corrélation et de la variance.

Nous rappelons ci-dessous l'expression de l'estimateur de la corrélation :

$$\beta_\epsilon(h) = \frac{\frac{1}{n-h} \sum_{i=1}^{n-h} \epsilon_i \epsilon_{i+h}}{\frac{1}{n} \sum_{i=1}^n \epsilon_i^2}$$

Si les résidus sont indépendants, cette valeur est égale à 0. Les expressions des statistiques de Ljung-Box et de Box-Pierce sont les suivantes, avec $k < E[n/4]$:

$$Q_{BP} = n \times \sum_{i=1}^k \beta_\epsilon^2(i)$$

$$Q_{LB} = n \times (n + 2) \times \sum_{i=1}^k \frac{\beta_\epsilon^2(i)}{n - i}$$

Ces deux statistiques suivent une loi du Khi deux à k degrés de liberté. Si les statistiques sont « trop grandes », le test sera rejeté car cela traduit une certaine corrélation entre les résidus.

- Le test de Breusch-Godfrey

Le test de Breusch-Godfrey, connu aussi sous le nom de test ML (pour multiplicateur de Lagrange) permet d'éviter certains écueils des tests de type portemanteau ou du test de Durbin Watson sur l'autocorrélation. Nous l'avons présenté dans le corps du mémoire, p30 et p52.

Tests de normalité

Nous exposons ci-dessous le principe de différents tests de normalité. L'hypothèse \mathbb{H}_0 est systématiquement l'adéquation de notre échantillon à une loi normale.

- L'étude statistique

Avant de faire des tests statistiques plus ou moins complexes, une étude statistique des données est intéressante. En effet, elle permet de discerner des points extrêmes à éventuellement retirer ou permet d'anticiper les résultats : ainsi, il est intéressant de tracer l'histogramme des fréquences ainsi que les box-plot de chaque jeu de données.

- Le test de Shapiro-Wilk

Ce test est très puissant pour les petits effectifs ($n \leq 50$). La statistique du test est définie par :

$$W = \frac{\sum_{i=1}^{E[\frac{n}{2}]} a_i \times (x_{(n-i+1)} - x_{(i)})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

où $x_{(i)}$ est la série des données triées, $E[\frac{n}{2}]$ la partie entière de $\frac{n}{2}$, et les a_i sont les coefficients du test de SHAPIRO-WILK, qui sont fournis dans des tables spécifiques.

Nos séries comportant généralement plus de 50 données, nous n'avons pas utilisé ce test.

- Le test de Jarque-Bera

Ce test est fondé sur les coefficients du skewness et du kurtosis. Il permet d'évaluer les écarts simultanés de ces coefficients avec ceux de la loi normale. Ce test se révèle très intéressant lorsque les effectifs testés sont élevés, ce qui est notre cas. La statistique de ce test est définie par l'expression ci-dessous, qui suit une loi du Khi deux à deux degrés de liberté :

$$JB = \frac{n}{6} \left(S^2 + \frac{K^2}{4} \right)$$

avec S et K, les skewness et Kurtosis de PEARSON, donnés par :

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}} \text{ et } K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

- Le test de d'Anderson-Darling

Le test d'ANDERSON DARLING est une variante du test d'adéquation entre les échantillons de KOLMOGOROV SMIRNOV. Il se focalise néanmoins plus sur les queues de distribution que ce dernier.

La statistique de ce test est définie par la relation suivante :

$$A = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\ln(F_i) + \ln(1 - F_{n-i+1})]$$

Où $F_{(i)}$ est la fréquence théorique de la loi de répartition normale centrée et réduite associée à la valeur standardisée de $z_{(i)} = \frac{x_{(i)} - \bar{x}}{s}$ et n est la taille de l'échantillon.

Les valeurs critiques définissant la région de rejet ne dépendent pas de la taille de l'échantillon et sont répertoriées dans le tableau suivant :

α	1%	5%	10%
$A_{critique}$	1.035	0.752	0.631

TABLE A.1 – Tableau de certaines valeurs critiques du test d'ANDERSON-DARLING

- Le test de Lilliefors

Le test de LILLIEFORS est lui aussi une variante du test de KOLMOGOROV SMIRNOV. Ce test est dit sensible à l'inadéquation de la distribution empirique avec la loi théorique au niveau du centre de la distribution (là où les écarts n'ont que peu d'effets), mais il est moins efficace lorsque l'inadéquation se situe au niveau des queues de distribution. Sa statistique est :

$$D = \max_{i=1, \dots, n} \left(F_i + \frac{i-1}{n}, \frac{i}{n} - F_i \right)$$

avec les mêmes notations que ci-dessus.

On peut approcher les valeurs critiques théoriques $D_{critique}$ pour les grands effectifs (plus grand que 30) par les formules suivantes :

α	1%	5%	10%
$D_{critique}$	$\frac{1.031}{\sqrt{n}}$	$\frac{0.886}{\sqrt{n}}$	$\frac{0.805}{\sqrt{n}}$

TABLE A.2 – Tableau d'approximation de certaines valeurs critiques du test de LILLIEFORS

Annexe B

Critère du maximum de vraisemblance

Voici un exemple d'utilisation du critère du maximum de vraisemblance pour un processus AR(p). Dans ce cas, on peut construire le test suivant :

$$\mathcal{H}_0 = 4$$

$$\mathcal{H}_1 = 6$$

Pour réaliser ce test, on estime alors le modèle AR contraint sous \mathcal{H}_0 et le modèle AR non contraint sous \mathcal{H}_1 $p = 6$. Ceci revient à tester si les coefficients de X_{t-5} et X_{t-6} sont conjointement égaux à zéro ou non.

La statistique LR définie ci-dessous suit une loi du Khi deux :

$$LR = T[\ln\sigma_c^2 - \ln\sigma_{NC}^2] \sim \chi^2(q)$$

où

- q est le nombre de restrictions imposées (dans l'exemple ci-dessus 2)
- T est le nombre d'observations
- σ_c et σ_{NC} sont les écarts types calculés respectivement sous \mathcal{H}_0 et \mathcal{H}_1 .

Si on ne peut pas rejeter \mathcal{H}_0 , les coefficients de X_{t-5} et X_{t-6} ne sont pas statistiquement différents de zéro et le modèle AR(4) est préférable au modèle AR(6) sur la base du principe de parcimonie. On poursuit l'analyse en examinant $\mathcal{H}_0 : p = 2$ $\mathcal{H}_1 : p = 4$ et ainsi de suite jusqu'à ce \mathcal{H}_0 soit rejetée.

Annexe C

Modèle de taux : Vasicek en termes HJM

Dans le chapitre deux de ce mémoire, nous avons présenté le modèle de taux utilisé par Wilkie qui consiste à modéliser les taux instantanés à l'aide d'une série chronologique en incorporant des corrélations avec l'inflation et le taux de dividendes. Bien que n'ayant pas utilisé de modèle de taux dans l'application de notre mémoire - rappelons que notre objectif final est d'optimiser l'allocation d'un portefeuille d'actions uniquement - il nous a semblé indispensable de s'intéresser au modèle de taux qui est un élément indispensable d'un GSE, en particulier pour les compagnies d'assurance pour lesquelles les obligations sont un composant majeur de l'actif. C'est pourquoi nous présentons cette annexe.

Le modèle de taux proposé par Wilkie ne nous semble pas adapté car il ne permet pas de construire une courbe des taux futurs. Ainsi, nous avons étudié différents modèles répondant à l'objectif de simplicité et permettant de construire une courbe des taux. Nous avons conclu que l'utilisation du modèle Vasicek en termes Health-Jarrow-Morton (HJM) était à privilégier. En effet, d'une part, ce modèle permet de prendre en compte la courbe des taux initiale comme paramètre, ce qui s'avère indispensable : c'est pour cette raison que les modèles factoriels usuels (Vasicek et Cox Ingersoll et Ross) ne peuvent être utilisés et qu'il est préférable de se tourner vers les modèles de déformation de la courbe des taux, comme le modèle HJM ; d'autre part, ce modèle est simple et a déjà été étudié : réalisé en 1990, ce modèle est très utilisé et fournit des formules fermées permettant d'obtenir la dynamique des prix des zéro-coupon.

Dans un premier paragraphe, nous allons présenter le modèle dans sa version initiale.

Dans un second paragraphe, nous présenterons le modèle de taux adapté à la structure du modèle de Wilkie.

Dans un troisième paragraphe, nous présenterons brièvement une méthode d'introduction de sauts dans le modèle.

Le modèle Vasicek en termes HJM

Ce paragraphe a pour objectif de présenter l'aspect théorique du modèle HJM.

Notations

Tout d'abord, introduisons les notations utilisées dans la suite pour décrire le modèle.

P(0,T) : prix zéro-coupon Il s'agit du prix actuel d'une obligation ne versant pas de coupon et ayant une maturité T.

F(i,t,T) : prix zéro-coupon forward Il correspond au prix, à horizon t, d'une obligation zéro-coupon de maturité t+T, calculé à la date i ($i < t < T$). On a la relation suivante :

$$F(i, t, T) = P^i(t, T)$$

Tous les calculs présentés dans la suite seront réalisés à la date $i=0$. Nous posons pour simplifier les écritures $P^0(t, T) = P(t, T)$.

R(t,T) : taux d'intérêt zéro-coupon Il correspond au taux relatif à une obligation zéro-coupon, nous avons la relation suivante :

$$R(t, T) = -\frac{\ln P(t, T)}{T - t}$$

Le taux d'intérêt instantané Il est applicable à un intervalle de temps infinitésimal et se définit par la relation suivante :

$$R(t, t + \Delta t) = -\frac{\ln P(t, t + \Delta t)}{\Delta t} \text{ avec } \Delta t \geq 0$$

f(t,T) : taux forward instantané Il s'agit du taux d'intérêt instantané applicable à une date future et donné par la relation suivante :

$$f(t, T) = \frac{\partial}{\partial T} \ln P(t, T)$$

Le taux Swap Un swap standard est l'échange de flux entre deux institutions bancaires. C'est l'échange d'un flux à taux variable contre un flux à taux fixe. A l'émission du Swap, nous évaluons l'évolution du taux variable, le taux fixe est ainsi fixé de sorte que la somme des flux entrant soit égale à celle des flux sortant. Le taux ainsi appliqué s'appelle le taux Swap.

Le cadre Health-Jarrow-Morton

Le modèle HJM est un cadre de travail permettant de modéliser à terme les taux d'intérêt. Il a pour avantage de chercher à s'ajuster à la structure des taux d'intérêt. De manière générale, il définit un modèle multifactoriel, c'est-à-dire utilisant plusieurs sources d'incertitudes (plusieurs mouvements browniens). L'équation différentielle suivie par la dynamique des taux à terme est la suivante :

$$df(t, T) = \mu(t, T)dt + \sum_{i=1}^n \sigma_i(t, T)dW_i(t)$$

avec $\sigma_i(t, T) = \sigma_i(t) \exp^{-\lambda_i(T-t)}$ où λ_i est une constante et $\sigma_i(t)$ est une fonction déterministe du temps.

Par la suite, nous allons uniquement nous intéresser au modèle à un facteur, ayant pour paramètre le processus de volatilité $\sigma(t, T)$.

Description du modèle de Vasicek à un facteur en termes HJM

Ce modèle permet d'évaluer la courbe des taux courts, en supposant qu'ils suivent un processus de retour à la moyenne. Comme indiqué dans le paragraphe précédent, l'équation différentielle stochastique suivie par les taux forwards instantanés sous HJM est la suivante :

$$df(t, T) = \mu(t, T)dt + \sigma(t, T)dW_t$$

Avec :

- $\sigma(t, T) = \sigma e^{-(\alpha(T-t))}$
- α est une constante positive
- W_t décrit un mouvement brownien géométrique

Dans le cadre HJM, en faisant l'hypothèse d'absence d'opportunité d'arbitrage et sous la probabilité risquée neutre, il existe une relation simple entre le processus de dérive et l'expression de $\sigma(t, T)$, donnée par l'équation ci-dessous :

$$\mu(t, T) = \sigma(t, T) \int_t^T \sigma(t, s) ds = \frac{\sigma^2}{\alpha} (e^{-\alpha(T-t)} - e^{-2\alpha(T-t)})$$

Par conséquent, l'équation stochastique du taux forward instantané est la suivante :

$$df(t, T) = \frac{\sigma^2}{\alpha} (e^{-\alpha(T-t)} - e^{-2\alpha(T-t)}) dt + \sigma e^{-\alpha(T-t)} dW_t$$

Par intégration, nous obtenons l'équation suivante :

$$f(t, T) = f(0, T) - \frac{\sigma^2}{2\alpha^2} (1 - e^{-\alpha(T-t)})^2 + \frac{\sigma^2}{2\alpha^2} (1 - e^{-\alpha T})^2 + \sigma \int_0^t e^{-\alpha(T-s)} dW_s$$

Nous en déduisons donc le processus des taux courts :

$$r(t) = f(t, t) = f(0, t) + \frac{\sigma^2}{2\alpha^2} (1 - e^{-\alpha t})^2 + \sigma \int_0^t e^{-\alpha(T-s)} dW_s$$

Ce processus de simulation admet comme paramètre d'entrée, la courbe des taux forward instantanés $f(0, T)$. Par exemple, dans le cadre de la directive solvabilité 2, la courbe des taux swap interbancaires est fixée. A partir de ces taux swap, il est nécessaire de déduire les taux zéro-coupon afin de calculer les taux forward instantanés. Le paragraphe suivant vise à décrire la méthodologie à appliquer.

Mise en place de l'algorithme

Passage des taux Swap aux taux zéro-coupon

La courbe des taux initiale devant être considérée est une courbe des taux swap. La

première étape consiste à calculer les taux zéro-coupon correspondants.

A l'émission, la valeur d'un swap est nulle. Par conséquent, la valeur des flux futurs actualisés au taux zéro-coupon est égale à celle actualisée au taux swap. Ces derniers peuvent donc être définis comme étant un ensemble de taux actuariels au pair.

Considérons qu'une obligation d'échéance n ans, payant un coupon de $C\%$ et ayant un taux actuariel à n ans de $C\%$, côte 100. Cela se traduit par :

$$\sum_{i=1}^{n-1} \frac{C}{(1+C)^i} + \frac{1+C}{(1+C)^n} = 1$$

L'actualisation des coupons faite par les taux zéro-coupon ne modifie pas le prix de l'obligation, nous avons donc également :

$$\sum_{i=1}^{n-1} \frac{C}{(1+\tau_{zc_i})^i} + \frac{1+C}{(1+\tau_{zc_n})^n} = 1$$

Où τ_{zc_i} : taux zéro-coupon de maturité i .

En faisant varier n et en procédant de proche en proche, nous obtenons une formule récursive sur les taux zéro-coupon :

$$\tau_{zc_n} = \frac{1+C_n}{1-C_n \sum_{i=1}^{n-1} (1+\tau_{zc_i})}$$

Où C_n est le taux Swap à n ans .

Calcul des taux forward instantanés

Le paramètre d'entrée du modèle est la courbe initiale des taux forward instantanés, nous devons donc les calculer à partir des taux zéro-coupons. Nous utilisons l'égalité suivante pour calculer les taux forward instantanés initiaux :

$$P(0, T) = \exp\left(-\int_0^T f(0, s) ds\right)$$

Où,

- $P(0, T)$: prix d'un zéro-coupon d'échéance T .
- $f(0, s)$: taux forward instantané débutant en date s pour une période infinitésimale.

On en déduit que :

$$\ln(P(0, T+h)) - \ln(P(0, T)) = -\int_T^{T+h} f(0, s) ds$$

En supposant que h est assez faible, nous obtenons l'expression suivante :

$$f(0, T) = \frac{1}{h}((T+h)R_{zc}(0, T+h) - TR_{zc}(0, T))$$

Discrétisation et simulation

Le modèle utilisé pour simuler la courbe des taux courts est décrit en temps continu. Dans la pratique, il convient de le discrétiser. Nous pouvons exprimer le processus des

taux courts sous la forme suivante :

$$r(t+\delta) - e^{-\alpha\delta}r(t) = f(0, t+\delta) - e^{-\alpha\delta}f(0, t) + \frac{\sigma^2}{2}(K^2(t+\delta) - e^{-\alpha\delta}K^2(t)) + \sigma e^{-\alpha\delta} \int_t^{t+\delta} e^{-\alpha(t-s)} dW_s$$

Le calcul de l'intégrale stochastique donne :

$$\int_t^{t+\delta} e^{-\alpha(t-s)} dW_s = \sqrt{\frac{e^{2\alpha\delta} - 1}{2}} \epsilon$$

Nous obtenons finalement :

$$r(t + \delta) - e^{-\alpha\delta}r(t) = f(0, t + \delta) - e^{-\alpha\delta}f(0, t) + \frac{\sigma^2}{2}(K^2(t + \delta) - e^{-\alpha\delta}K^2(t)) + \sqrt{L(\delta)}$$

Avec :

- $K(t) = \frac{1 - e^{-\alpha t}}{\alpha}$
- $L(t) = \frac{\sigma^2}{2}(1 - e^{-2\alpha t})$
- α et σ sont deux constantes positives
- ϵ suit une loi normale centrée réduite

Modèle HJM adapté au modèle de Wilkie

Comme nous l'avons signifié dans l'introduction de ce mémoire, l'ensemble des modifications apportées au modèle de Wilkie ne doit pas altérer la structure initiale du modèle. Ainsi, afin de respecter cela, il convient d'intégrer un paramètre relatif à l'inflation et un autre relatif à l'équation du rendement des dividendes. La méthode d'intégration de ces paramètres n'a pas été étudiée précisément et nous nous limitons à présenter le concept global. Il conviendrait d'affiner le lien entre l'inflation et le taux forward afin de déterminer une manière pertinente d'intégration de cette relation dans les équations. Ceci n'étant pas utile pour notre application, nous n'avons pas développé ce point.

Intégration de sauts dans le modèle de taux

Notre GSE étant créé pour s'adapter au crise, il convient de permettre à l'utilisateur d'intégrer des probabilités de crises pour l'équation des taux. Contrairement aux autres équations du modèle pour lesquelles nous avons adapté l'algorithme de Box et Tiao, nous conseillons d'utiliser pour les taux, un autre modèle. En effet, il nous semble opportun de créer une chaîne de Markov permettant de simuler le changement de rating du pays, secteur ou entreprise étudié.

Nous considérons donc une chaîne de Markov à 10 états qui correspondent aux rating attribués par les agences de notation : AAA, AA, A, BBB, BB, B, CCC, CC, C et D. En cas de différence entre les trois agences de notation les plus importantes, il est préférable de considérer une moyenne des ratings. Il est également envisageable de choisir une agence de notation de référence.

Les probabilités de transition sont déterminées à dire d'experts, indépendamment pour chaque pays, secteur ou entreprise que l'on souhaite étudier. Il convient d'associer à chaque transition, un saut dont l'amplitude est à déterminer. Pour cela, il serait intéressant de

réaliser une étude précise de l'historique des sauts constatés suite aux changements de notation. Notons que le principe suivant est à considérer : une diminution de rating se traduira par une augmentation de la valeur du taux ; une diminution sera observée dans le sens contraire.

Dans cette annexe, nous avons souhaité présenter un modèle de taux, adapté au modèle de Wilkie et incluant la prise en compte de scénarios de crises. Nous avons proposé un modèle relativement simple et permettant d'intégrer l'analyse macroéconomique du gérant.

Annexe D

Résultats sur les indicateurs avancés

Une étude des indicateurs avancés a été réalisée afin de comparer leur performance pour prévoir l'évolution de l'économie des pays. Nous en avons étudié quatre :

- le Composite Leading Indicator (CLI)
- le SP 500
- l'ECRI (Conseil économique Research Institute)
- le Conference Board

Comme nous l'avons montré dans le chapitre 7, l'indicateur avancé le plus performant est le CLI. Cependant, il est à utiliser avec précaution car il a tendance à prévoir trop de retournements. Nous présentons ici, quelques résultats plus détaillés.

L'étude de la performance de ces indicateurs pour prédire les ralentissements de l'économie américaine depuis les années 2000 donne les résultats suivants :

	Nombre de mois optimal de prévision	Ralentissements non prédits (%)	Ralentissements prédits non réalisés (%)	Prédictions réalisées (%)
CLI	5	0	17	82
SP500	6	23	27	50
ECRI	6	21	24	55
Conference Board	7	16	17	76

TABLE D.1 – Comparaison du pouvoir prédictif de quatre indicateurs depuis 2000

L'étude de la performance de ces indicateurs pour prédire les ralentissements de l'économie américaine durant la période de crise [2007 ; 2010] donne les résultats suivants :

	Nombre de mois optimal de prévision	Ralentissements non prédits (%)	Ralentissements prédits non réalisés (%)	Prédictions réalisées (%)
CLI	2	0	9	91
SP500	9	44	0	56
ECRI	9	44	0	59
Conference Board	9	15	0	85

TABLE D.2 – Comparaison du pouvoir prédictif de quatre indicateurs sur la période [2007;2010]

Annexe E

Coefficients retenus pour calibrer les résidus

Nous avons simulé tous les résidus de notre modèle par des lois Pareto hybride. Nous avons estimé les coefficients en utilisant le logiciel R. Nous donnons ci-dessous les paramètres retenus :

Résidus	Moyenne	Ecart-type	ξ_g	ξ_d
France - inflation	0.0003	0.00049	-0.101	-0.278
France - Dividend Yield	-0.013	0.025	-0.182	0.070
France - Montand Dividendes	-0.0077	0.0039	0.911	1.109
US - inflation	0.0001	0.0004	0.023	-0.1722
US - Dividend Yield	-0.0067	0.0169	-0.316	-0.062
US - Montand Dividendes	0.0026	0.0040	0.21	0.16
Allemagne - inflation	0.0002	0.0006	-0.143	-0.028
Allemagne - Dividend Yield	-0.014	0.028	-0.256	0.043
Allemagne - Montand Dividendes	0.001	0.0033	0.51	0.20
Japon - inflation	-1,325e-05	0.0004	0.131	0.0028
Japon - Dividend Yield	-0.005	0.00014	-0.294	-0.093
Japon - Montand Dividendes	-0.004	0.003	0.312	0.644

TABLE E.1 – Tableau des coefficients relatifs aux résidus des équations

Annexe F

Le bootstrap non paramétrique par blocs

Cette méthode est décrite de manière très claire dans [9] (pages 100-101). Nous présentons ici une synthèse de cette méthode.

Le principe de la méthode du bootstrap non paramétrique est de tirer et de concaténer des blocs de points consécutifs tirés à partir de positions choisies aléatoirement dans la série des rendements considérée. La méthode du bootstrap non paramétrique par blocs repose sur ce même principe, mais au lieu de tirer un point puis un autre, on considère des blocs de tailles aléatoires. Par exemple, on commence avec un bloc de taille un et puis on rajoute itérativement le point avec probabilité $1 - p$. La méthode par blocs possède l'avantage de pouvoir rendre compte d'éventuels phénomènes de mémoire présent dans la série étudiée. En effet, en choisissant un bloc de plusieurs valeurs successives, cette méthode ne détruit pas un éventuel lien entre deux valeurs successives.

Pour calculer les VaR de notre mémoire (voir p77 et suivantes), nous avons choisi des blocs de deux points : ceci correspond à l'existence d'une mémoire longue faible dans les rentabilités des séries considérées, ce qui est une propriété connue des rentabilités des séries financières.

Voici le code VBA que nous avons utilisé :

```
Sub BootstrapNonPara()  
Dim i, j, mean As Integer  
Dim alpha As Single  
For j = 1 To 1000  
For i = 1 To 50  
Randomize  
alpha=Int((27-1)*Rnd()+1)  
Cells(2*i-1,j) = Excel.WorksheetFunction.VLookup(alpha, Range("B1 :D100"),3)+mean  
Cells(2*i,j) = Excel.WorksheetFunction.VLookup(alpha+1, Range("B1 :D100"),3)+mean  
Next i  
Next j  
End Sub
```

TABLE F.1 – Code VBA - Bootstrap non paramétrique