

**Mémoire présenté devant l'ENSAE
pour l'obtention du diplôme d'Actuaire ENSAE
et l'admission à l'Institut des Actuaires**

le 2 Novembre 2015

Par : Pierrick PIETTE

Titre : Assurance Agricole et Images Satellites : Produit d'Assurance et Arbitrage
Financier

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

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

*Membre présent du jury de l'Institut des
Actuaires :*

M. Gérard CROSET

M. Michel FROMENTEAU

Membres présents du jury de l'ENSAE :

M. Nicolas BARADEL

Mme Caroline HILLAIRET

Signature : Entreprise :

Nom : AXA Corporate Solutions

Ville : Paris, France

Directeur de mémoire en entreprise :

Nom : M. Kevin DEDIEU

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)**

Secrétariat :

Bibliothèque :

Signature du responsable entreprise :

Signature du candidat :

ENSAE ParisTech, 3 Avenue Pierre Larousse, 92245 Malakoff cedex

RESUME

Mots-clés : assurance agricole, assurance paramétrique, CBOT, États-Unis, image satellite, maïs, marché des futures, NASA, NDVI, prévision des prix, réaction des marchés, rendement agricole, revenu agricole, soja, télédétection

Notre mémoire étudie l'apport potentiel des images satellites de la NASA, plus spécifiquement de l'indice NDVI (*Normalized Difference Vegetation Index*), à l'assurance agricole que ce soit pour le risque de rendement agricole ou le risque de prix. Nous commençons par rappeler les fondamentaux de l'assurance paramétrique ainsi que le contexte de l'assurance agricole aux États-Unis.

Dans un second temps nous analysons et tarifons un produit d'assurance surface-rendement classique (sans images satellites) dont l'indice est fondé sur les statistiques gouvernementales. Ce type de produit est soumis à des contraintes de données, dont notamment la fiabilité de ces dernières, qui ne sont pas toujours vérifiées. L'utilisation d'images satellites permet de se préserver de certaines de ces contraintes et donc facilite le développement de l'assurance agricole dans des régions du globe qui n'y avaient pas accès auparavant. Afin d'illustrer ce point, nous structurons et tarifons un produit d'assurance portant sur les mêmes risques que le produit d'assurance surface-rendement précédemment analysé.

Dans un troisième temps nous montrons l'importance du risque de prix dans l'assurance agricole américaine. En analysant l'impact des rapports gouvernementaux sur le marché des futures CBOT (*Chicago Board of Trade*) portant sur les matières premières agricoles, nous montrons que les estimations précoces des rendements agricoles publiées par le NASS (*National Agricultural Statistics Service*) sont considérées par les acteurs comme une information utile. Nous créons ensuite un modèle fondé sur l'indice NDVI permettant l'estimation précoce de cette information, et donc des changements dans les prix de marchés. Nos résultats peuvent servir de base à une stratégie d'arbitrage statistique et donc diminuer le risque de prix transféré aux assureurs. Nous proposons des pistes d'extension et d'amélioration de notre méthode, notamment par l'augmentation de la précision des images satellites mais qui demande une forte puissance de calcul.

ABSTRACT

Key words: agricultural insurance, agricultural revenue, agricultural yield, CBOT, corn, futures market, markets reaction, NASA, NDVI, parametric insurance, price forecasting, remote sensing, satellite image, soybean, United States

Our thesis studies the potential contribution of NASA satellite images, more precisely the NDVI (*Normalized Difference Vegetation Index*) images, to the agricultural insurance industry, either concerning the agricultural yield risk management or agricultural price risk. In a first part we remind some fundamentals of parametric insurance and agricultural insurance in the United States.

In a second part we analyze and price a classic yield-area insurance product (without satellite data) based on an index created from governmental statistics. Some data constraints have to be verified for this kind of insurance product, especially the reliability of the data which is not always the case around the world. The use of satellite images allows insurers to avoid some of these constraints and makes the agricultural insurance development easier in some parts of the world where the farmers used to not have access to it. To illustrate this point, we structure and price a NDVI based insurance product on the same risks as the are-yield insurance studied just before.

In a third part we show the importance of the price risk in the American agricultural insurance industry. By analyzing the impacts of governmental report on the CBOT (*Chicago Board of Trade*) futures market, we enlighten that the early estimations of agricultural yields published by the NASS (*National Agricultural Statistics Service*) are considered by the market as a useful information. We then create a prediction model for this information based on the NDVI time series which also allows us to estimate the new prices. Our results can be used as a base for a statistical arbitrage strategy and thus decrease the price risk transferred to the insurers. We propose some extensions and some ways of improvement for our method, especially by increasing the precision of the satellites images which, however, requires a significant increase of the computational power.

REMERCIEMENTS

Je remercie en premier lieu M. Tanguy Touffut qui m'a donné l'occasion d'effectuer ce mémoire au sein de l'équipe d'assurance paramétrique d'AXA Corporate Solutions dans les meilleures conditions qui soient.

Je remercie bien entendu chaleureusement mon maître de stage M. Kevin Dedieu pour sa grande disponibilité et ses conseils précieux durant la totalité de mon stage. Je tiens aussi à présenter mes plus profonds remerciements à M. Claudio Busarello pour son accompagnement, ses connaissances et sans qui cette étude n'aurait certainement pas eu lieu.

Je remercie également l'ensemble des personnes qui m'ont aidé au sein d'AXA CS, en particulier Mlle Madeleine Latapie, Mlle Leila Miri et M. Thomas Ozello qui ont rendu ce stage des plus agréables par leur bonne humeur constante et leur disponibilité.

Enfin j'adresse mes remerciements à M. Xavier Milhaud, mon encadrant académique à l'ENSAE ParisTech, pour ses conseils avisés pour ce mémoire.

SOMMAIRE

Introduction Générale	7
1. Assurance Paramétrique	8
1.1. Principe	8
1.1.1. Climat et économie.....	8
1.1.2. Transfert de risque climatique	10
1.2. Littérature actuarielle.....	11
1.3. Enjeux de l'assurance paramétrique	12
1.3.1. Avantages et défauts actuariels	12
1.3.2. Avantages économiques	13
1.3.3. Enjeux technologiques.....	14
2. Assurance Agricole aux États-Unis	16
2.1. Agriculture américaine.....	16
2.2. Les <i>Farm Bill</i>	17
2.3. Les limites de notre étude	19
Assurance Rendement	21
3. Description de l'assurance rendement	22
3.1. Structuration de l'assurance rendement	22
3.2. Enjeux actuariels de l'assurance rendement	23
3.2.1. Tarification.....	23
3.2.2. Gestion des sinistres	24
3.2.3. Provisionnement.....	24
4. Tarification de l'assurance rendement	25
4.1. Structure	25
4.2. Données	26
4.3. Analyse tendancielle des rendements	27
4.3.1. Régression linéaire.....	27
4.3.2. Régression kernel gaussien.....	28
4.3.3. Régression LOESS.....	30
4.4. Tarification	31
4.4.1. Modélisation des rendements	31
4.4.2. Modélisation des queues de distribution.....	32

4.4.3.	Résultats.....	33
5.	Assurance rendement par images satellites.....	35
5.1.	Littérature et enjeux actuariels.....	35
5.1.1.	Description de l'indice NDVI.....	35
5.1.2.	Revue littéraire de l'indice NDVI.....	37
5.1.3.	Enjeux Actuariels de l'indice NDVI.....	38
5.2.	Construction d'un indice NDVI.....	40
5.2.1.	Données NDVI utilisées.....	40
5.2.2.	Corrélation entre NDVI et rendement agricole.....	42
5.3.	Tarifification.....	45
5.3.1.	Analyse de l'indice.....	45
5.3.2.	Structuration du produit d'assurance.....	45
5.3.3.	Modélisation de l'indice NDVI.....	46
5.3.4.	Résultats.....	47
	Assurance Revenu.....	48
6.	Description de l'assurance revenu.....	49
6.1.	Structuration de l'assurance revenu.....	49
6.2.	Enjeux de l'assurance revenu.....	49
6.2.1.	Taille de l'assurance revenu aux Etats-Unis.....	49
6.2.2.	Conséquences sur la gestion financière.....	50
7.	Rapports Gouvernementaux.....	51
7.1.	Revue littéraire.....	51
7.1.1.	Fondamentaux des prix futures agricoles.....	51
7.1.2.	Impact des rapports gouvernementaux.....	52
7.1.3.	Données utilisées.....	54
7.2.	Rapport WASDE.....	57
7.2.1.	Méthodologie.....	57
7.2.2.	Statistiques descriptives.....	59
7.2.3.	Résultats.....	60
7.3.	Estimations précoces du NASS.....	63
7.3.1.	Méthodologie.....	63
7.3.2.	Résultats.....	65
8.	Recherche d'arbitrage.....	69
8.1.	Prévision des estimations précoces de rendement.....	69

8.1.1.	Méthodologie	69
8.1.2.	Statistiques Descriptives.....	71
8.1.3.	Résultats.....	72
8.2.	Limites et pistes d'améliorations	76
8.3.	Extensions.....	77
Conclusion Générale		78
Glossaire		80
Bibliographie		81
Bibliographie générale.....		81
Bibliographie sans auteur précis		84
Mémoires d'actuariat.....		85

PREMIERE PARTIE
INTRODUCTION GENERALE

Chapitre 1

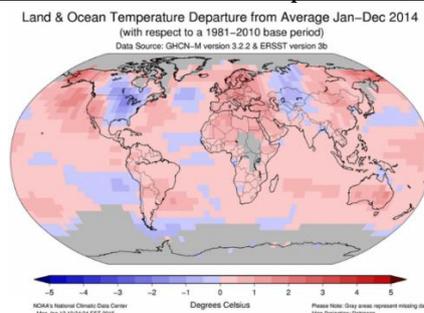
1. ASSURANCE PARAMETRIQUE

1.1. PRINCIPE

1.1.1. CLIMAT ET ECONOMIE

Le changement climatique et la multiplication probable des catastrophes naturelles associée auront un impact significatif sur l'économie mondiale, et donc sur l'assurance, dans les prochaines décennies. Selon la NASA la température moyenne de surface de la Terre a augmenté d'environ 0,8 °C depuis 1880. Cette tendance est largement due à la hausse du dioxyde de carbone et des autres émissions de l'activité humaine dans l'atmosphère. De plus la majorité de ce réchauffement s'est produit durant les 30 dernières années. Le réchauffement climatique n'implique pas forcément des hivers plus doux, mais plutôt une multiplication des anomalies extrêmes. Ainsi en 2014, la température moyenne de surface de la Terre a été la plus haute depuis le début des relevés, soit probablement l'un des plus chaud du millénaire. Dans le même temps le climat peut fortement différer d'une région à l'autre, par exemple l'hiver 2013-2014 fut l'un des plus froids pour le nord-est des États-Unis, le second plus chaud pour la France depuis 110 ans et l'un des plus pluvieux pour le Royaume-Uni depuis un siècle. Le mois de février 2015 fut le mois le plus sévère dans l'histoire des relevés : selon l'Administration Nationale Océanographique et Atmosphérique (NOAA, *National Oceanic and Atmospheric Administration*) plus de 1 000 records de températures minimales et 900 de hauteurs de neige ont été battus. La *Figure 1.1* montre les anomalies de température en 2014, la *Figure 1.2* montre le réchauffement climatique à la station météorologique de Paris-Montsouris (station de référence pour la France).

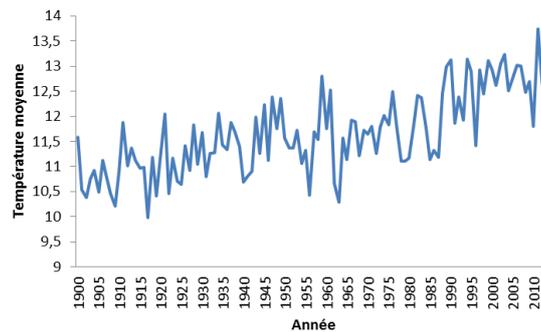
Figure 1.1 : Anomalies de température, 2014



Source : *National Oceanic and Atmospheric Administration*¹

¹ <https://www.ncdc.noaa.gov>

Figure 1.2 : Température moyenne annuelle, Paris Montsouris, 1900-2014



Source : *Speedwell Weather*²

Dans ce contexte la météo-sensibilité des entreprises devient une problématique de plus en plus importante pour la gestion du risque et les assurances. Certains secteurs sont plus impactés que d'autres, nous donnons quelques exemples chiffrés³ qui ne forment évidemment pas une liste exhaustive des secteurs météo-sensibles :

- (i) Les basses températures, la pluie et les chutes de neiges sont des risques importants pour le secteur de la construction. Le ciment ne peut pas être utilisé lorsqu'il fait trop froid. Le printemps froid de 2013 a fait baisser de 17% le chiffre d'affaires du secteur de construction allemand.
- (ii) Le transport est fortement touché par les vagues de froid et de chutes de neige. Les compagnies aériennes doivent faire face aux dégivrages, retards et annulations de vols. Durant le premier trimestre 2014, une compagnie aérienne a perdu 200 millions de dollars à cause de la vague de froid dans le Nord-Est des États-Unis.
- (iii) L'énergie est aussi impactée par les anomalies météorologiques. Les mois doux de janvier et février 2014 en France ont entraîné une chute de 20% de la distribution de gaz de chauffage.
- (iv) Les produits saisonniers de nourriture et de boisson sont sensibles au climat. Par exemple les ventes de bières ont baissé de 15% en 2012 dues aux faibles températures et aux précipitations importantes. La vague de froid de 2013 a provoqué une chute de 20% des ventes de glaces en Europe Occidentale.

Parmi ces secteurs, il en est un qui est plus météo-sensible que les autres : l'agriculture. Les risques climatiques sont responsables de la qualité, des rendements agricoles, des coûts et des revenus financiers. Ce risque est d'autant plus élevé pour les pays en voie de développement où l'agriculture représente une grande partie de l'économie et où des sécheresses ou des

² www.speedwellweather.com

³ Exemples tirés de **Parametric Insurance : A Fitting Solution for the Weather-Sensitive** (2015)

inondations peuvent détruire des récoltes entières. Il est donc important de structurer et de mettre à disposition au plus grand nombre des produits de transfert de risques climatiques.

1.1.2. TRANSFERT DE RISQUE CLIMATIQUE

Deux types principaux de produits de transfert de risques climatiques sont actuellement à disposition des différents acteurs : les dérivés climatiques et les produits d'assurance paramétrique. Nous nous intéressons dans notre mémoire à la deuxième catégorie de solutions. L'avantage principal de l'assurance paramétrique sur les dérivés climatiques est la flexibilité de la structuration, le désavantage est que les dérivés climatiques reviennent souvent moins chers car ils sont cotés en bourse. Nous ne rentrons pas plus en détails à propos des dérivés climatiques.

Le principe de l'assurance paramétrique est la construction d'un indice corrélé avec la variable d'intérêts (*e.g.* ventes, rendements agricoles, qualité des cultures, etc.). Cet indice doit cependant respecté certaines propriétés :

- (i) observable et facilement mesurable ;
- (ii) objectif ;
- (iii) vérifiable indépendamment ;
- (iv) communiqué dans un délai convenable ;
- (v) cohérent dans le temps.

Cet indice n'est pas forcément météorologique, par exemple un rendement agricole d'une région publié par une autorité indépendante, et sa structuration, qui peut être complexe (assurance multi-indicielle, flexibilité temporelle, flexibilité spatiale, etc.), est créée après une analyse de météo-sensibilité de la variable d'intérêt.

Finalement nous aboutissons a un produit d'assurance sur-mesure. Cette flexibilité peut être résumée comme suit: « *The beauty in parametric insurance is its customizability. In order to build the most appropriate cover, we bring together our client, brokers, and our team of engineers and weather experts to come up with the best-fit solution. We customize by location, dates, trigger, so each client receives a cover that accurately reflects his risk⁴* » (T. Touffut⁵). Un produit d'assurance paramétrique basique peut être résumé par le profil des indemnités associées définies par les trois variables suivantes : le point d'attache, le point de sortie et la valeur d'un point d'indice. Ces indemnités sont calculées à l'aide de la formule des montants compensatoires suivante,

⁴ « L'élégance de l'assurance paramétrique est sa personnalisation. Afin de construire la couverture la plus appropriée, nous rassemblons notre client, nos courtiers, notre équipe d'ingénieurs et d'experts climatiques pour arriver au meilleur ajustement de la solution. Nous personnalisons en fonction de la localisation, des dates, des points d'attache, afin que chaque client reçoive une couverture qui reflète au mieux son risque »

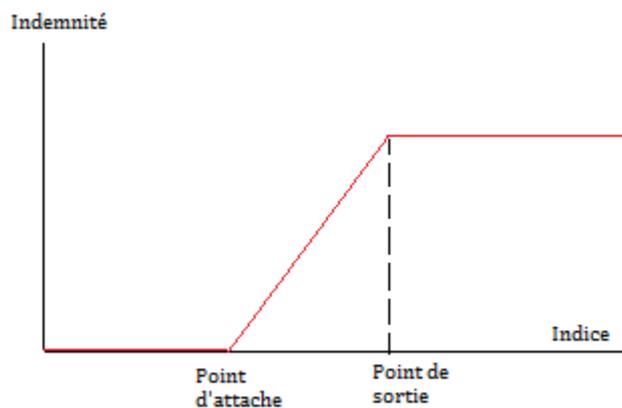
⁵ Head of parametric insurance à AXA Corporate Solutions

$$P = V \times \mathbb{1}_{I \geq A} \min\{I - A; S - A\} \quad (1.1)$$

où P est l'indemnité reçue, V est la valeur d'un point d'indice, I est la valeur de l'indice observée, A le point d'attache et S le point de sortie. La *Figure 1.3* représente graphiquement les indemnités de l'assurance paramétrique basique. Bien entendu il ne s'agit là que de la structure la plus répandue sur le marché (paiements linéaires) où l'indice est corrélé positivement avec les pertes. Une quasi-infinité d'autres structures peuvent être appliquées, dans notre mémoire nous nous restreignons cependant à ce type de produit. Néanmoins, dans la manière dont nous définissons nos indices dans notre étude, les risques correspondent à la partie basse de l'indice, *i.e.* plus l'indice est bas plus la perte est importante (corrélé négativement avec les pertes). Les indemnités sont alors calculées grâce à la formule suivante,

$$P = V \times \mathbb{1}_{I \leq A} \min\{A - I; A - S\} \quad (1.2)$$

Figure 1.3 : Exemple d'indemnités d'une assurance paramétrique



Dans le cas de l'assurance paramétrique le point d'attache peut être vu comme une franchise, cependant il ne joue pas le même rôle. Dans la plupart des contrats d'assurance (auto par exemple), la franchise est mise en place pour combattre l'aléa moral. L'assurance indiciaire est beaucoup moins sujette à ce problème comme nous le notons dans la partie 1.3.1. Le point d'attache et le point de sortie servent plutôt à déterminer la part du risque transférée à la compagnie d'assurance. En ce sens le produit d'assurance paramétrique doit être perçu comme un contrat de réassurance XS.

1.2. LITTÉRATURE ACTUARIELLE

Plusieurs mémoires d'actuariat discutent déjà de l'utilité des produits d'assurances paramétriques ou des dérivés climatiques. Vendé (2003) discute de la corrélation entre les différents indices climatiques dans la tarification d'une couverture indiciaire de réassurance Cat. Denizot et Plouhinec (2008) évaluent les implications des différents modèles de température dans la tarification des dérivés climatiques. Gilles et Finas (2011) s'intéressent quant à eux à la gestion du risque de vent pour l'énergie éolienne avec une construction de mathématiques

financières de dérivés climatiques pour répondre à ce problème. Nguyen et Ritleng (2014) étudient un produit d'assurance paramétrique sur le risque de pluie en Jamaïque.

Les mémoires que nous citons ci-dessus fondent leurs analyses sur des indices purement météorologiques dont l'application n'est pas l'agriculture. D'après nos recherches, le mémoire qui se rapproche le plus de notre étude est Koch (2011) où l'auteur analyse la création d'un produit d'assurance rendement agricole par indice climatique. Bien que nos indices ne soient pas climatiques (rendement publié par une agence publique ou indice de réflectivité), les Chapitre 4 et Chapitre 5 répondent au même type de problématique. Plusieurs autres mémoires discutent des produits d'assurance agricole. Nous citons à ce sujet Leprince (2001) qui décrit deux produits d'assurance aux États-Unis, rendement et revenus, qui correspondent respectivement à notre deuxième et troisième partie. Bien que les produits d'assurance soient les mêmes que dans nos travaux, la méthodologie est différente. Leprince (2001) décrit les produits et ses implications dans la gestion du risque, nous analysons pour notre part l'apport des images satellites pour ces produits.

1.3. ENJEUX DE L'ASSURANCE PARAMETRIQUE

1.3.1. AVANTAGES ET DEFAUTS ACTUARIELS

L'assurance paramétrique a de nombreux avantages d'un point de vue actuariel. L'utilisation d'indices suivant les propriétés énoncées à la partie 1.1.2. permet de proposer une réponse simple à certaines problématiques actuarielles.

La structuration du produit d'assurance sur des indices indépendants de l'assuré, par exemple le bénéficiaire ne peut *a priori* pas influencer la température moyenne à une station météorologique, permet de supprimer l'aléa moral. Le risque et les indemnités reçues étant fondés sur ces indices extérieurs à l'assuré, celui-ci n'a donc aucune possibilité d'augmenter ce risque. De même, dans la majorité des cas lorsque le contrat d'assurance est signé suffisamment en avance, l'assuré ne détient pas plus d'information sur le risque couvert par le contrat d'assurance que la compagnie à laquelle il a transféré le risque. L'anti-sélection disparaît alors. Pour illustrer ce point nous prenons l'exemple d'une entreprise de construction voulant se couvrir contre le risque de froid pendant l'hiver en choisissant une assurance paramétrique sur l'indice de température moyenne des mois de novembre à mars. Les données météorologiques (historiques, prévisions journalières ou saisonnières, cycles climatiques, etc.) sont accessibles avec le même degré de précision pour l'assureur et l'assuré. De plus, dans la majorité des cas, l'analyse des experts climatiques de l'assurance est plus avancée que celle de la gestion des risques de l'entreprise de construction. Ainsi sur le risque couvert (risque de froid) l'assureur détient plus d'information que l'assuré.

Cependant l'assurance paramétrique n'a pas que des avantages. Le premier inconvénient est l'augmentation du risque de base pour l'assuré. En effet, même si la structuration du produit de l'indice et du calcul des indemnités se fait après une étude poussée de la sensibilité de l'activité de l'assuré à cet indice, le risque couvert n'est jamais directement celui du bénéficiaire. Le risque que les montants compensatoires ne soient pas en totale adéquation avec la perte réelle est donc plus important. De plus dans certains cas les risques ne sont pas indépendants, notamment en assurance indicielle agricole. Plusieurs contrats peuvent être fondés sur le même indice ou différents indices mais fortement corrélés positivement, par exemple la température moyenne à Lyon et la température moyenne à Paris. L'assurance paramétrique fait ainsi face à de nombreux risques systémiques.

1.3.2. AVANTAGES ECONOMIQUES

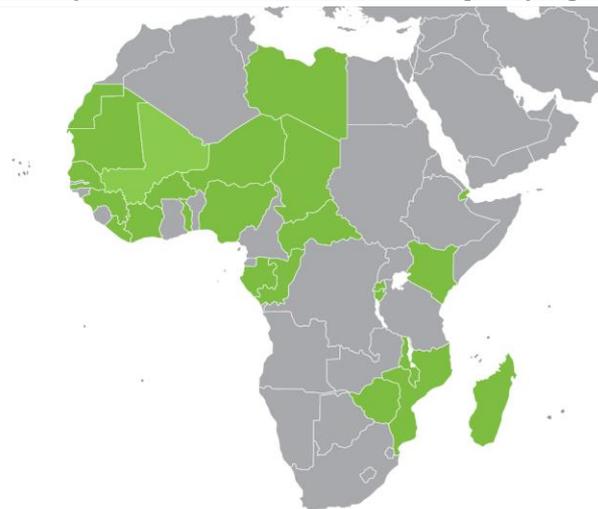
Pratiquement l'utilisation de produit d'assurance paramétrique permet d'effectuer certains gains économiques pour la compagnie d'assurance. L'intérêt économique majeur de l'assurance indicielle est le coût de la gestion des sinistres. En effet les « sinistres » sont calculés de manière objective à partir de l'indice choisi. La valeur de ce dernier est observable depuis n'importe quel endroit avec un accès internet, qu'importe la localisation réelle du risque, il n'y a donc aucune utilité à engager un expert qui se déplacerait. De par sa nature objective et vérifiable indépendamment, le nombre de litige, et donc de frais judiciaires, est restreint par rapport aux autres branches de l'assurance dommage. De plus le calcul des indemnités est fait dès la publication de l'indice, soit souvent moins d'un an après l'occurrence du risque : de quelques jours pour les indices météorologiques, à quelques mois pour les rendements agricoles. Cette rapidité se retrouve donc naturellement dans le paiement des sinistres, limitant les provisions pour l'assurance et augmentant la rapidité d'indemnisation pour l'assuré.

Ce faible coût de l'assurance paramétrique a permis son expansion dans des pays en voie de développement, notamment dans l'assurance agricole en Afrique. Il permet de proposer des produits assurantiels économiquement viables là où les produits plus classiques seraient trop chers pour la population locale. Pour exemple nous pouvons citer le programme d'assurance indicé sur les précipitations de l'*African Risk Capacity*⁶ (ARC). L'indice, développé avec le Programme Alimentaire Mondial (*World Food Programme*⁷), est fondé sur les précipitations observées à partir de données satellites pour une résolution de 10km par 10km. Rassemblant pour l'instant 4 pays (Kenya, Mauritanie, Niger, Sénégal), ce programme est à terme destiné aux 27 pays membres de l'ARC Agency. La *Figure 1.4* représente les pays membres de l'ARC. En 2014 les paiements de ce programme ont atteint environ 24 millions de dollars.

⁶ <http://www.africanriskcapacity.org/>

⁷ <http://fr.wfp.org/>

Figure 1.4 : Pays membre de l'African Risk Capacity Agency

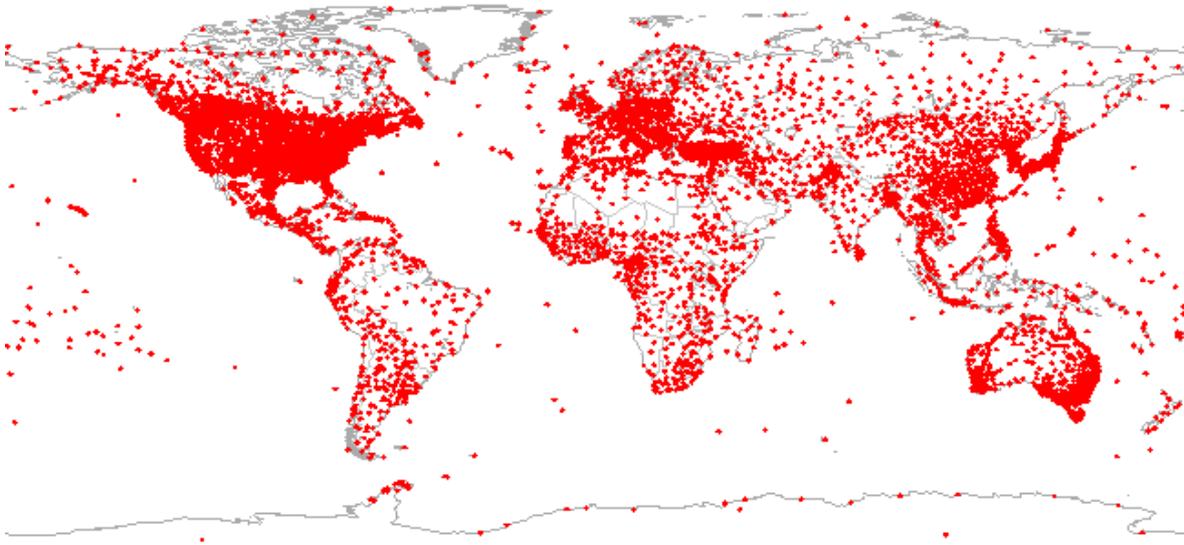


1.3.3. ENJEUX TECHNOLOGIQUES

La multiplication des données a un fort impact dans l'assurance paramétrique. La résolution spatiale des images satellites ne cesse de s'améliorer, atteignant aujourd'hui en moyenne 250 mètres, et dans quelques années nous serons sûrement capables d'obtenir une résolution de l'ordre du mètre. La résolution temporelle est aussi très importante, les données sont maintenant disponibles à un rythme horaire pour les variables météorologiques classiques (température, vent, précipitation, etc.). Le nombre de variables observables ne cessent d'augmenter avec les nouvelles technologies (indice de réflexion, évapotranspiration, etc.). Enfin à ces données satellites se rajoutent les données fournies par les stations météorologiques dont l'historique est disponible sur plus d'un siècle pour certaines. La *Figure 1.5* représente la localisation des stations météorologiques dans le monde.

La multiplication des données disponibles fait de l'assurance paramétrique un des secteurs de l'assurance les plus réceptifs au *Big Data*, permettant ainsi de proposer des produits d'assurance toujours plus précis et disponibles dans toutes les régions du monde. Le développement du *Big Data* dans l'assurance paramétrique est un enjeu majeur pour son expansion. Nous percevons d'ailleurs dans notre étude l'importance de la puissance de calcul pour nos analyses et c'est l'un des points qui a limité nos travaux (cf. Chapitre 8).

Figure 1.5 : Localisation des stations météorologiques dans le monde



Source : National Oceanic and Atmospheric Administration⁸ (NOAA)

⁸ <http://www.ncdc.noaa.gov/>

Chapitre 2

2. ASSURANCE AGRICOLE AUX ÉTATS-UNIS

Pour tout ce mémoire nous avons conservé les unités de mesure américaines, non métriques, car elles sont la norme dans la littérature économique agricole. Ainsi nous utilisons les boisseaux US (environ 25,4 kg pour le maïs, 27,2 kg pour le soja), l'acre (40% d'un hectare) et le dollar américain.

2.1. AGRICULTURE AMERICAINE

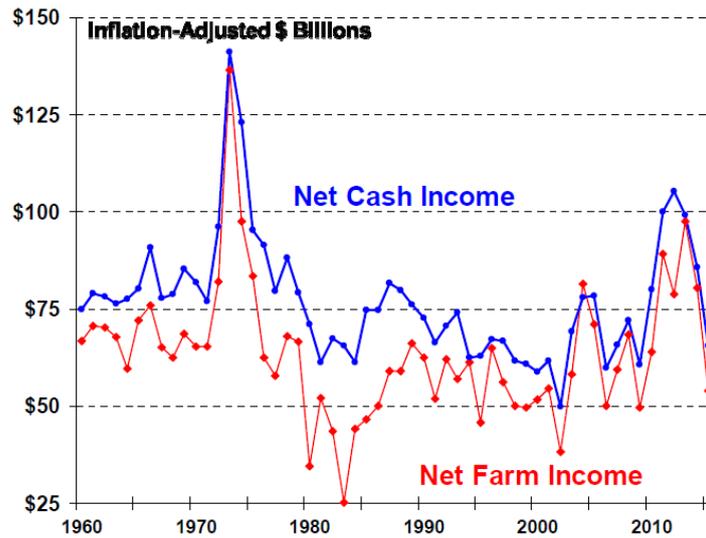
L'industrie agricole américaine est un secteur important de l'économie des États-Unis. L'agriculture américaine est la première puissance agricole au monde mais aussi l'une dont les bases de données sont les plus détaillées. Le début de la statistique agricole américaine date de George Washington (1791), ces collectes et analyses de données sont restées très importantes encore aujourd'hui ce qui en fait un sujet parfait pour les études économiques ou statistiques. D'après le recensement sur l'agriculture de 2012⁹, il y a 2,1 millions d'exploitations avec une superficie moyenne de 434 acres par exploitation. La valeur totale des produits agricoles vendus est d'environ 395 milliard de dollars. En 2013, l'agriculture américaine représentait 1,4% du PIB des États-Unis.

A cause de la nature aléatoire des récoltes et de la variabilité des prix des denrées agricoles, l'agriculture est une branche où les revenus sont soumis à une forte volatilité comparée à d'autres secteurs économiques. L'agriculture américaine n'échappe pas à cette règle. Schnepf (2015) souligne le fait que non seulement les revenus varient dans le temps, ainsi l'USDA estime par exemple une baisse de 32% des revenus nets agricoles pour 2015 à 73,6 milliard de dollars, mais aussi géographiquement à travers les États-Unis due notamment à des facteurs climatiques et des conditions de marché. La *Figure 2.1* montre l'évolution des revenus agricoles annuels de 1960 à 2015 aux États-Unis corrigés de l'inflation.

Cette variabilité de revenus des agriculteurs les incite à acheter des produits permettant de lisser leurs résultats sur plusieurs années. Les assurances agricoles répondent à cette demande et sont de ce fait très prisées : 88% des exploitations éligibles à l'assurance aux États-Unis ont souscrit un produit d'assurance en 2014, soit l'équivalent de plus de 10 milliard de dollars de prime. Historiquement la première assurance agricole aux États-Unis a été autorisée dans les années 1930 à cause de la Grande Dépression et du *Dust Bowl* (« Bassin de poussière »).

⁹ 2012 Census of Agriculture

Figure 2.1 : Évolution des revenus agricoles aux États-Unis, corrigés de l'inflation 1960-2015



Source : Schnepf (2015)

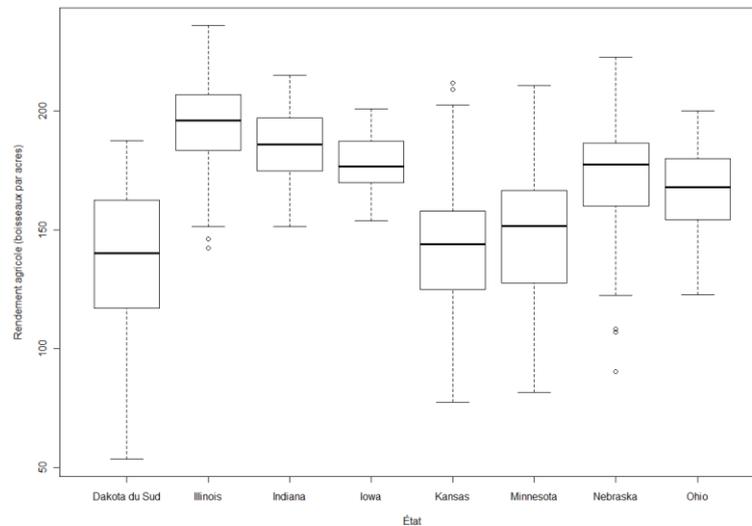
2.2. LES FARM BILL

Les deux facteurs principaux impactant les revenus des agriculteurs sont les rendements agricoles obtenus et les prix des denrées. Le risque d'un mauvais rendement est spécifique à chaque exploitation : les conditions climatiques, qualité du sol, produits utilisés, techniques agricoles mises en place sont d'autant de variables particulières propres à chaque agriculteur. Il en résulte que les rendements varient fortement d'un État à l'autre, voire même d'un comté à l'autre. La Figure 2.2 montre cette sensibilité par comté pour les rendements agricoles du maïs sur les États de la *Corn Belt* pour l'année 2014. Cependant, même si les rendements ont des valeurs fortement différentes d'une zone à l'autre, il existe une corrélation significativement positive entre eux. Par exemple dans le cas d'une sécheresse importante une grande partie des récoltes de la *Corn Belt* peut être touchée. Sur une région géographique restreinte, typiquement la *Corn Belt*, le risque de rendement peut donc être considéré comme un risque systémique. Le risque de prix est le même pour toutes les exploitations, les prix donnés par la Bourse de Chicago étant communs pour tous les acteurs du marché. Ainsi le risque de prix est aussi un prix systémique car il touche toutes les exploitations en même temps.

La nature systémique des risques dans l'assurance agricole, que ce soit pour le prix ou le rendement dans une moindre mesure, est un problème pour les assureurs. Si les sinistres (baisse des rendements agricoles ou des prix) surviennent pour tous les assurés en même temps, le risque de faillite de la compagnie d'assurance est beaucoup plus important. Pour faire face à ce risque, les primes des produits d'assurances sont fortement augmentées, afin d'obtenir des réserves suffisantes, et deviennent rapidement économiquement insoutenables pour les agriculteurs. En assurance agricole les catastrophes sont la norme plutôt que l'exception. Pour

éviter cet effet la politique agricole des États-Unis, les *Farm Bill*, participe financièrement à l'achat d'assurance pour les agriculteurs.

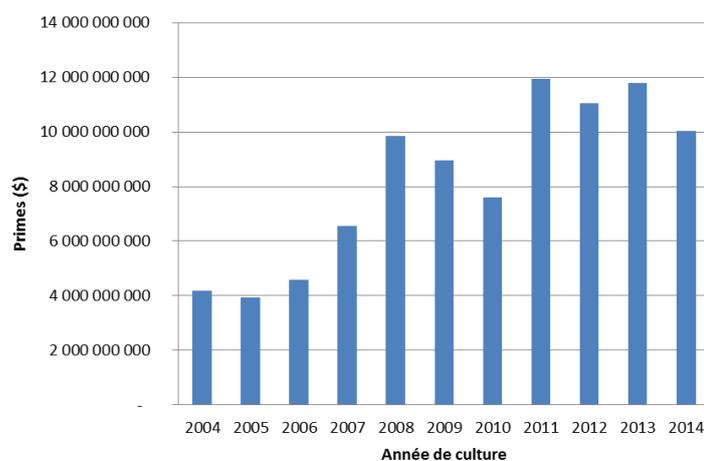
Figure 2.2 : Rendements agricoles des comtés de la *Corn Belt*, récolte de maïs 2014.



Les *Farm Bill* sont l'équivalent de la PAC (Politique Agricole Commune) aux États-Unis depuis 1938, faisant l'objet d'un vote tous les cinq ans au congrès américain. A la date de rédaction du mémoire, le *Farm Bill 2014* est en vigueur et est doté d'une enveloppe de 950 milliards de dollars jusqu'en 2018. De cette enveloppe environ 10% sont destinés pour les programmes d'assurances agricoles. Contrairement à la PAC dont le principal outil est l'aide directe, le *Farm Bill* fait recours à plusieurs systèmes assurantiels. Ainsi les agriculteurs américains ne reçoivent des financements seulement s'ils subissent des pertes dans leur production. Ce choix de structure de l'aide apportée aux agriculteurs a permis à l'assurance agricole de se développer nettement plus aux États-Unis qu'en Europe car les primes sont devenues abordables pour les agriculteurs. En 2014, sur 10 milliards de dollars de prime pour les assurances agricoles, environ 6,2 milliards ont été payés par le gouvernement via le *Farm Bill*. Cette croissance de l'assurance agricole est représentée par la *Figure 2.3*.

Les produits d'assurances agricoles peuvent être séparés en deux catégories : ceux prenant en compte le risque de prix et ceux qui ne le couvrent pas. Le *Farm Bill* propose dans ses programmes les deux types de produits. La deuxième partie de ce mémoire se concentre sur le risque de rendement et plus particulièrement sur les produits d'assurance surface-rendement (*area-yield insurance*). Notre troisième partie traite des produits permettant aux agriculteurs de couvrir leurs revenus, *i.e.* à la fois le rendement agricole et le prix de vente.

Figure 2.3 : Évolution du marché de l'assurance agricole aux États-Unis, 2004-2014



Source : 2015 Crop Insurance Update, Securing America's Farmers

2.3. LES LIMITES DE NOTRE ETUDE

Nous concentrons notre étude sur les cultures du maïs et du soja qui sont deux des trois cultures majeures des États-Unis (la troisième étant le blé). Spatialement nous nous restreignons à la zone de la *Corn Belt* (littéralement « la ceinture de maïs ») qui est une région de forte production agricole du mid-ouest des États-Unis. Plus précisément les 8 États d'intérêts de ce mémoire sont l'Illinois, l'Indiana, l'Iowa, le Kansas, le Minnesota, le Nebraska, l'Ohio et le Dakota du Sud. La *Figure 2.4* montre la localisation de ces États. Les *Figure 2.5* et *Figure 2.6* représentent la production de maïs et de soja par comté en 2014 aux États-Unis. Dans cette zone ainsi délimitée l'assurance agricole représentait en 2014 environ 2,56 milliards de dollars de prime pour le maïs et 1,36 milliards pour le soja, pour respectivement 34,2 et 19,5 milliards de dollars de production assurée.

Bien que les modèles que nous mettons en place dans la deuxième partie soient fondés sur les données des États-Unis, l'importance de notre travail réside plutôt dans la méthodologie appliquée. Ces résultats peuvent être étendus à d'autres zones géographiques relativement facilement. Ainsi, bien que l'assurance rendement par indice satellite soit une alternative à l'assurance rendement classique aux États-Unis, elle devient beaucoup plus innovatrice dans son application aux pays en voie de développement, comme indiqué dans la partie 1.3.2. En revanche la troisième partie est plus spécifique à l'agriculture nord-américaine. Certes nous proposons des pistes d'extension pour d'autres régions au Chapitre 8, mais elles nécessitent une adaptation nettement plus importante.

Figure 2.4 : Les États d'intérêts de notre étude aux États-Unis

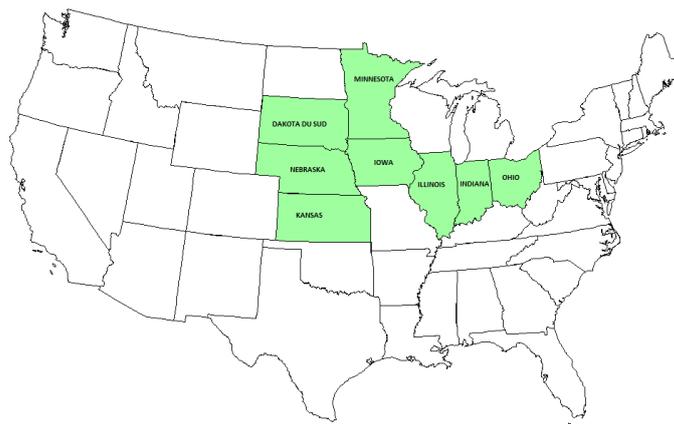
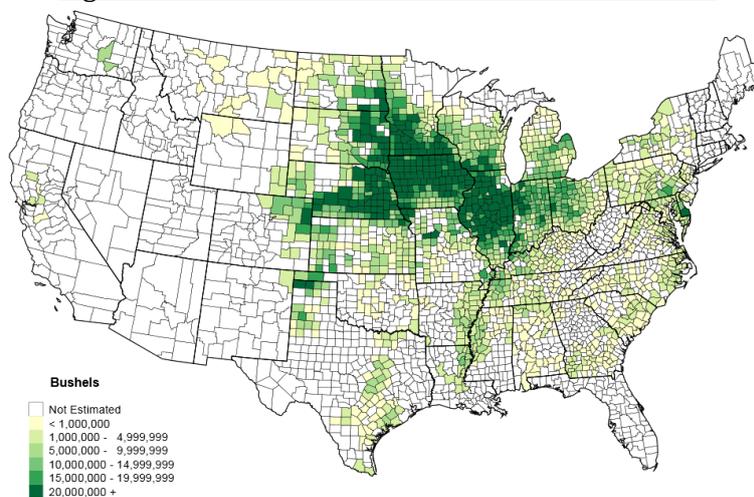
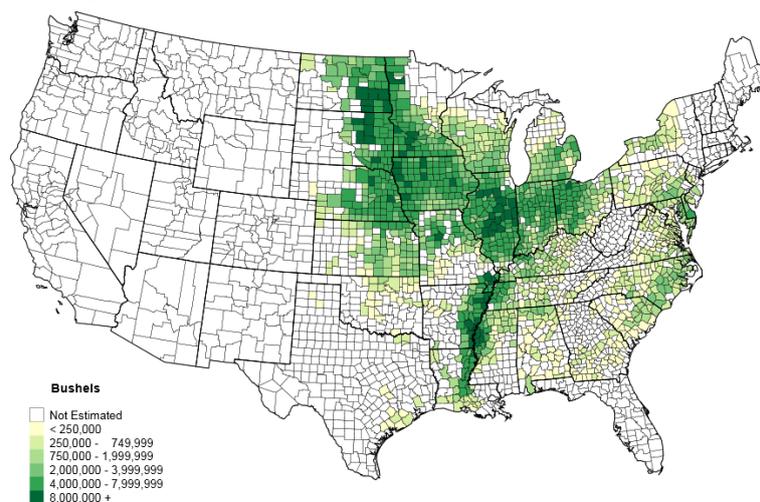


Figure 2.5 : Production de maïs aux États-Unis, 2014



Source : NASS

Figure 2.6 : Production de soja aux États-Unis, 2014



Source : NASS

DEUXIEME PARTIE
ASSURANCE RENDEMENT

Chapitre 3

3. DESCRIPTION DE L'ASSURANCE RENDEMENT

3.1. STRUCTURATION DE L'ASSURANCE RENDEMENT

L'assurance rendement, c'est-à-dire qui couvre seulement la perte due à une chute des rendements agricoles de l'exploitant, est l'un des produits d'assurance agricole les plus basiques. Dans certains cas les indemnités perçues sont calculées d'après l'avis d'experts agricoles, cependant ce mémoire se restreint aux assurances paramétriques. Plus précisément nous nous focalisons sur les produits d'assurance surface-rendement (*area-yield insurance*), *i.e.* les indemnités sont proportionnelles à la surface exploitée et calculées en fonction d'un rendement agricole de référence.

La fin de l'assurance paramétrique est de réduire les coûts de gestion et le risque d'aléa moral, l'utilisation du rendement agricole de l'exploitation elle-même serait donc mal venue. En effet l'assurance devrait faire appel à un expert (et donc le rémunérer) afin de déterminer le rendement de référence, ce qui augmenterait les coûts de gestion. De plus l'exploitant ayant la possibilité d'influer fortement sur ce rendement, il y a alors un risque d'aléa moral. Ces deux raisons amènent donc les assureurs à utiliser d'autres rendements agricoles de référence fortement corrélés avec ceux de l'exploitant, facilement accessibles et sur lesquels l'agriculteur assuré n'a qu'une influence très limitée. La réponse à cette demande est dans la très grande majorité des cas un rendement agricole fourni par une institution publique, typiquement le ministère de l'agriculture, pour la zone administrative dans laquelle se trouve l'exploitation.

Une fois ce rendement de référence choisi, la structure de l'assurance suit celle d'une assurance paramétrique classique décrite dans le Chapitre 1. Le point d'attache et le point de sortie sont souvent exprimés en pourcentages d'un rendement agricole dit attendu ou normal. Dans les produits d'assurance agricole classiques des États-Unis, ce rendement attendu est la moyenne olympique sur 5 ans du rendement de référence, *i.e.* la moyenne des rendements agricoles des 5 dernières années auxquels l'on soustrait le maximum et le minimum. Enfin il ne reste que la valeur de la production à évaluer. Cette dernière peut correspondre au prix des futures sur les marchés financiers pour la culture assurée ou au coût fixe de la culture pour le producteur. Aux États-Unis tous les produits d'assurance rendement surface se fondent sur le prix des futures de la Bourse de Chicago (CBOT, *Chicago Board of Trade*).

3.2. ENJEUX ACTUARIELS DE L'ASSURANCE RENDEMENT

Les produits d'assurance agricole de type surface-rendement nécessitent que certaines propriétés soient vérifiées sur les données utilisées pour que l'assureur accepte le transfert de risque.

3.2.1. TARIFICATION

Afin de pouvoir tarifier le produit d'assurance, l'assureur doit être en capacité de se procurer un historique des rendements agricoles de référence suffisamment long dont les données sont suffisamment fiables. Dans la pratique un historique de 10 ou 15 ans peut être considéré comme assez long. Le problème qui se pose le plus souvent dans l'assurance agricole mondiale est la fiabilité des données. Dans ce mémoire nous nous intéressons au cas des États-Unis, dont le Ministère de l'Agriculture (*USDA, United States Department of Agriculture*) dispose de fonds importants pour l'estimation des récoltes des différentes cultures à un niveau administratif détaillé. Ce n'est cependant pas le cas de tous les pays. Les assureurs peuvent considérer que les moyens mis en œuvre par les institutions publiques pour la détermination des rendements agricoles ne sont pas fiables, et refusent de fonder leurs calculs sur ces données. Cette situation apparaît souvent dans les pays d'Afrique instables politiquement ou des pays avec des conflits internes, *e.g.* l'Ukraine. Dans ces cas l'utilisation de produits d'assurance fondés sur des données satellitaires devient une solution de remplacement, les assureurs considérant la NASA comme fiable. Nous donnons un exemple de ce type de produit dans le Chapitre 5.

Une autre variable qui peut influencer sur la volonté de l'assureur à accepter le transfert de risque est le type de culture dont il est question. Dans ce mémoire nous analysons seulement des cultures dites annuelles, *i.e.* dont la plante ne vit qu'une année. Le maïs « naît » de la graine après avoir été planté et « meurt » une fois récolté. Ce n'est par exemple pas le cas des arbres fruitiers. La vigne, dont la demande en assurance agricole est assez importante en France, ne meurt pas après les vendanges. D'une année sur l'autre il s'agit de la même plante, ainsi les récoltes ne peuvent pas toujours être considérées comme indépendantes. Par exemple une bonne récolte, c'est-à-dire avec beaucoup de fruits, peut fatiguer la plante et celle-ci donnera alors moins de fruits l'année suivante même si les conditions (météorologiques, engrais, maladies, etc.) sont inchangées. Ce phénomène est difficile à prendre en compte pour les assureurs qui n'ont souvent pas les connaissances nécessaires pour l'analyser en détail. Les agriculteurs fruitiers quant à eux sont plus sensibles à ce problème, ils disposent donc de plus d'information. Nous retrouvons dans le cas de la sélection adverse que l'assurance paramétrique tente d'éviter.

3.2.2. *GESTION DES SINISTRES*

Tout comme la tarification, la gestion des sinistres nécessite que les données fournies par l'institution publique soient fiables. Mais la principale problématique sur les données pour la gestion des sinistres est le fait que l'exploitant ne puisse pas influencer significativement sur le rendement agricole de référence. Avec l'utilisation de statistiques agricoles gouvernementales ce problème ne se pose que rarement pour les exploitations de taille moyenne, cependant les produits d'assurance peuvent être achetés par des coopératives regroupant des dizaines d'agriculteurs d'une même région. Dans ce dernier cas la coopérative peut représenter une proportion non négligeable de la production de la région administrative sur laquelle est fondé l'indice. Potentiellement, en cas de mauvaise récolte, les exploitants agricoles pourraient alors décider de changer de comportement afin de toucher plus d'indemnités car leurs rendements influent fortement sur la statistique publique. Nous retrouvons alors un risque d'aléa moral pour l'assureur.

Plusieurs méthodes peuvent être employées pour contrer cet effet. La première est d'utiliser si possible un rendement de référence sur une zone administrative plus grande sur lequel la coopérative aura une influence plus limitée, quitte à augmenter le risque de base. La seconde possibilité est d'utiliser un prix d'indemnisation des denrées agricoles plus bas que le marché afin d'inciter les agriculteurs à faire de leur mieux dans tous les cas. Nous retournons ici à l'utilisation d'une franchise. Enfin une dernière possibilité est de changer le rendement agricole de référence en choisissant un indice purement météorologique sur lequel les assurés ne peuvent avoir aucune influence. C'est d'ailleurs le sujet du mémoire d'actuariat de Koch (2011).

3.2.3. *PROVISIONNEMENT*

Dans le cas de l'assurance surface-rendement, le rendement agricole de référence est publié par une institution publique. Certains pays fournissent des estimations précoces de ces rendements, c'est notamment le cas des États-Unis, mais pour la majorité des pays les statistiques définitives sur les cultures ne sont publiées qu'un ou deux mois après la récolte. Ce délai de publication peut obliger les assureurs à effectuer des provisions pour sinistres non déclarés. Afin de quantifier ces provisions, la compagnie d'assurance peut utiliser des indices, satellitaires ou météorologiques, fortement corrélés avec le rendement agricole de référence mais dont les données sont disponibles avant la récolte. Nous ne développons pas spécifiquement une telle méthode dans ce mémoire, mais l'analyse faite dans le Chapitre 5 montre qu'elle ne serait pas complexe à mettre en place.

Chapitre 4

4. TARIFICATION DE L'ASSURANCE RENDEMENT

Dans ce chapitre nous donnons une méthode simplifiée de la tarification de l'assurance rendement. Nous fondons ce travail sur des hypothèses qui peuvent être remises en question, de plus nous nous intéressons qu'au calcul de la prime pure, *i.e.* de la perte moyenne estimée. Nous pensons que la tarification précise d'assurance agricole puisse faire l'objet d'un mémoire d'actuariat à part entière. Plusieurs mémoires d'actuariat en discutent certainement déjà, notamment Batisse et Mercuzot (2009) ainsi que Fosso et Seretti (2010). Cependant, malheureusement pour notre étude, ces mémoires sont confidentiels à la date de rédaction.

4.1. STRUCTURE

Dans ce chapitre nous nous intéressons à deux produits d'assurance rendement classiques : sur le maïs en Iowa et sur le soja en Illinois. Les indices de référence sont les estimations des rendements agricoles des cultures étudiées publiés par l'USDA en janvier pour les zones géographiques analysées. Le rendement normal est défini comme la moyenne olympique sur les 5 dernières années (moyenne des 5 derniers rendements agricoles auxquels on soustrait le maximum et le minimum). Le point d'attache est fixé à 5% du rendement normal. Le point de sortie est fixé à 30%. Pour le prix des matières premières nous utilisons 4\$ par boisseaux pour le maïs et 9\$ par boisseaux pour le soja¹⁰. Enfin nous donnerons les différentes valeurs pour une superficie de 1000 acres (400 hectares), ce qui revient par exemple à étudier un produit d'assurance pour un agriculteur possédant un champ de 1000 acres en Iowa. Le *Tableau 4.1* résume la structure des produits que nous étudions.

Tableau 4.1 : Structure de l'assurance rendement étudiée

Culture	Iowa Maïs	Illinois Soja
Rendement normal	167	49,7
Point d'attache (%)	5%	5%
Point d'attache	158,7	47,2
Point de sortie (%)	30%	30%
Point de sortie	116,9	34,8
Superficie (acres)	1 000	1 000
Prix (\$/boisseaux)	4	9
Somme assurée (\$)	133 600	89 100

¹⁰ Il s'agit de l'ordre de grandeur des prix des futures à échéance décembre 2015 et novembre 2015 pour respectivement le maïs et le soja à la date de rédaction.

4.2. DONNEES

Pour une assurance rendement classique les données principales nécessaires à la tarification sont l'historique des rendements agricoles de la zone de référence pour la culture assurée. Ces données sont pour la majeure partie des cas fournies par le bureau de statistique national du pays dans lequel se trouve la culture à assurer. Dans notre cas il s'agit du *National Agricultural Statistics Service*¹¹ (NASS), le service national de statistiques agricoles des États-Unis. Ce service met en libre accès de multiples données concernant l'agriculture américaine, telles que la production, le rendement agricole, la superficie plantée, etc. disponibles par type de culture (maïs, soja, blé, orge, etc.), zones géographiques (national, État, comté) sur des historiques pouvant atteindre plus de cent années pour certaines variables. En tout ce sont environ 33 millions d'observations rendues disponibles grâce à l'outil Quick Stats¹². Pour information le service équivalent en France est l'Agreste¹³ qui, de même, met à disposition ses données par l'intermédiaire de l'outil Disar¹⁴.

Dans cette partie nous nous intéressons à la tarification de deux exemples d'assurance rendement : une sur le soja en Illinois et une autre sur le maïs en Iowa. Ces deux États font partie de la *Corn Belt* et sont des producteurs importants de maïs et de soja aux États-Unis ce qui implique un fort potentiel en terme de clients pour de l'assurance agricole, comme le montre le *Tableau 4.2*. Les primes en 2014 pour ces deux cultures, tous produits d'assurances confondus, représentaient environ 700 millions de dollars. La localisation géographique de ces deux États est montrée à l'aide de la *Figure 2.4*.

Tableau 4.2 : Assurance agricole, soja dans l'Illinois et maïs dans l'Iowa, 2014

État	Illinois	Iowa
Culture	Soja	Maïs
Surface plantée (millier d'acres)	9 800	13 700
Production (millions de boisseaux)	548	2 367
Nombre de polices	68 520	79 312
Somme assurée (millions \$)	3 771	8 737
Superficie assurée (milliers d'acres)	8 278	12 798
Primes (millions \$)	200	501
Taux de sinistres	22%	228%

Source : 2015 Crop Insurance Update, Securing America's Farmers; NASS

¹¹ <http://www.nass.usda.gov/>

¹² <http://quickstats.nass.usda.gov/>

¹³ <http://agreste.agriculture.gouv.fr/>

¹⁴ <https://stats.agriculture.gouv.fr/disar/>

Nous disposons des rendements agricoles annuels à partir de 1924 pour le soja dans l'Illinois et de 1866 pour le maïs de l'Iowa. Le maïs et le soja étant des cultures annuelles aux États-Unis, nous avons une observation par année. Cependant nous décidons de nous restreindre à la période 1950-2014, soit 65 ans d'historique, afin d'éviter d'éventuelles variations dues à la seconde guerre mondiale ou au *Dust Bowl*. Ces séries temporelles sont représentées dans les *Figure 4.1* et *Figure 4.2*. Nous constatons une évolution significative des rendements agricoles au cours du temps pour les deux cultures étudiées. Nous devons donc corriger les données brutes de cette tendance avant de modéliser les rendements.

Figure 4.1 : Rendement du soja, Illinois

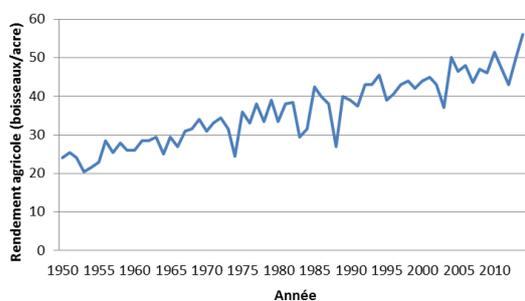
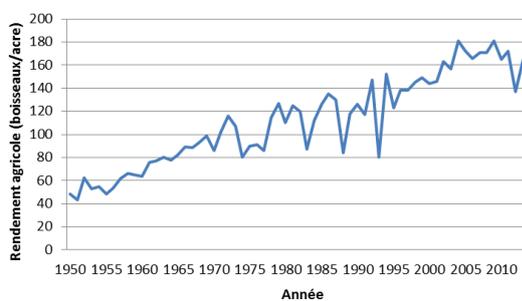


Figure 4.2 : Rendement du maïs, Iowa



4.3. ANALYSE TENDANCIELLE DES RENDEMENTS

Comme nous venons de le constater, les rendements agricoles que nous étudions pour cette étude de cas ont une tendance significative. La correction de cette tendance est une étape cruciale dans la tarification du produit d'assurance. Selon les techniques employées la prime peut être très différente. Pour montrer cette sensibilité, nous décidons d'appliquer trois des techniques les plus répandues dans le monde assurantiel agricole : la régression paramétrique linéaire, la régression non-paramétrique kernel gaussien et la régression non-paramétrique LOESS (*Local regrESSion*).

Pour simplifier la compréhension des parties suivantes les modèles de régressions seront notés sous la forme suivante,

$$E(Y|X) = m(X) + \epsilon \quad (4.1)$$

où Y est la variable aléatoire des rendements agricoles, X sont les années d'observation et ϵ est le terme d'erreur. Nous préciserons donc pour les différents modèles la fonction m .

4.3.1. REGRESSION LINEAIRE

Le modèle de régression linéaire est un des plus ancien modèle utilisé pour la modélisation des rendements agricoles. Elle est fondée sur l'estimation suivante de la fonction m ,

$$\hat{m}(x) = \beta_0 + \beta_1 \cdot x \quad (4.2)$$

Lorsque nous l'appliquons à nos données nous trouvons un rendement moyen de 175,2 boisseaux par acre pour le maïs dans l'Iowa et 49,4 boisseaux par acre pour le soja dans l'Illinois. Les écarts-types sont respectivement de 13,9 et 3,3 boisseaux par acre. Les *Figure 4.3* et *Figure 4.4* représentent les données corrigées de la tendance en utilisant la régression linéaire.

Figure 4.3 : Régression linéaire sur le rendement du soja, Illinois, 1950-2014

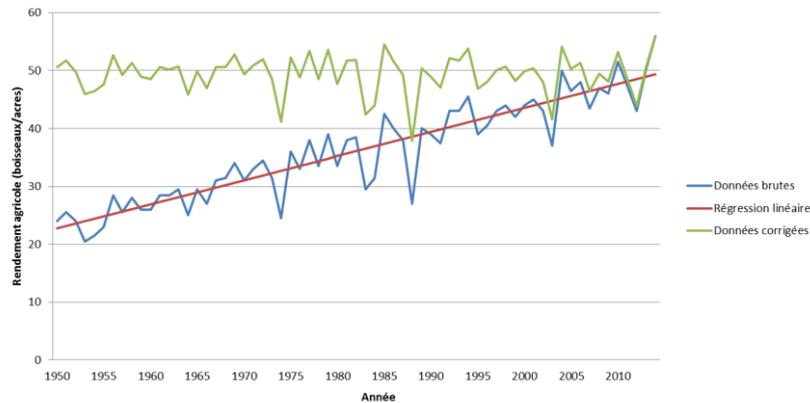
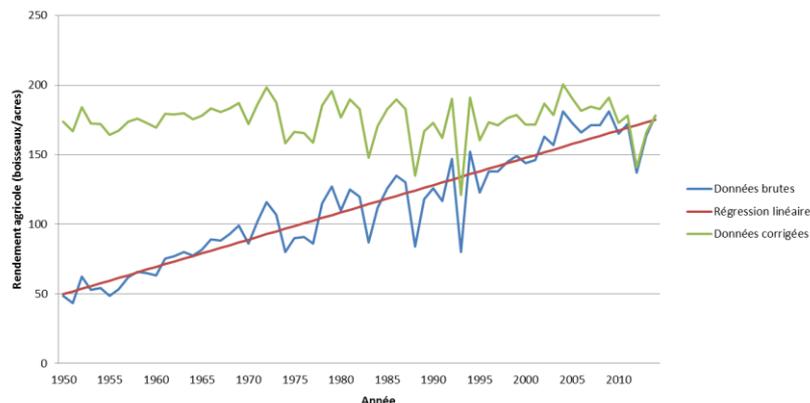


Figure 4.4 : Régression linéaire sur le rendement du maïs, Iowa, 1950-2014



4.3.2. REGRESSION KERNEL GAUSSIEN

L'utilisation des modèles de régression non-paramétrique est plus récente. Elle est motivée par le fait que l'évolution des rendements est due à des progrès techniques dans le monde agricole (engrais, création de nouvelles variétés hybrides, véhicules agricoles, etc.) dont les effets ne sont pas linéaires dans le temps mais apparaissent plutôt par sauts. Les modèles non-paramétriques permettent de prendre en compte ce dynamisme. Une représentation générale des estimations de la fonction m dans le cas de modèles non-paramétriques est notée, en suivant Pagan et Ullah (1999), de la façon suivante,

$$\hat{m}(x) = \sum_{i=1}^n w(x_i, x) \cdot y_i \quad (4.3)$$

où $w(x_i, x)$ représente le poids associé à la i -ème observation y_i , avec un poids dépendant de la distance entre le point x_i et le point x .

Dans le cas de la régression kernel la fonction de pondération est définie comme suit,

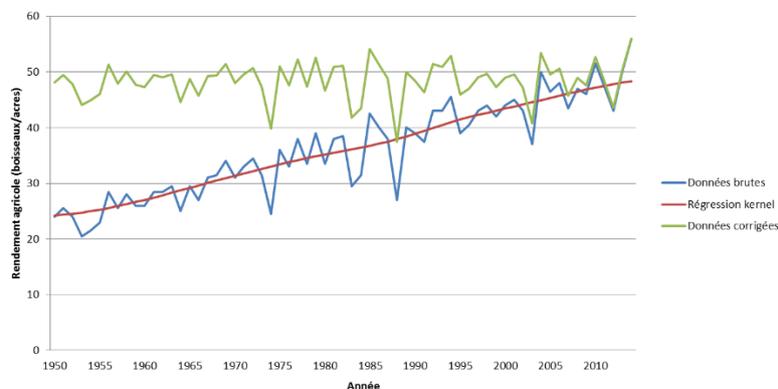
$$w_h(x_i, x) = \frac{K_h(x_i - x)}{\sum_{i=1}^n K_h(x_i - x)} \quad (4.4)$$

où $K(\cdot)$ est un noyau¹⁵ et h un paramètre de lissage, appelé aussi bande passante. La régression kernel utilisée dans les études de rendements agricoles, et que nous utilisons dans ce mémoire, est le kernel gaussien, dont le noyau est défini ci-dessous,

$$K_h(u) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{1}{2} u^2 h^{-2}\right) \quad (4.5)$$

Lorsque nous appliquons cette méthode à nos données de rendements agricoles nous obtenons des rendements moyens plus faibles qu'avec la régression linéaire de la partie précédente pour le maïs de l'Iowa et le soja de l'Illinois. Les rendements moyens ainsi calculés sont respectivement 166,1 et 48,4 boisseaux par acre. La *Figure 4.5* et la *Figure 4.6* montrent la correction appliquée aux données par la méthode de la régression kernel gaussien. Cette différence a bien entendu une influence sur la modélisation, et donc sur la valeur de la prime finale.

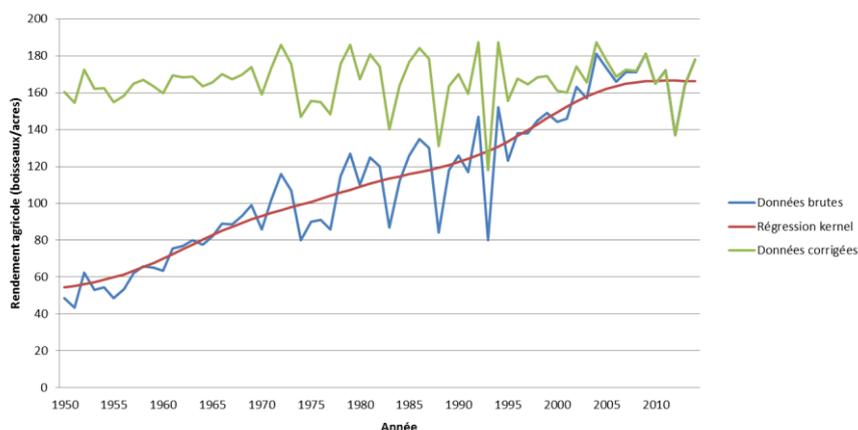
Figure 4.5 : Régression kernel gaussien sur le rendement du soja, Illinois, 1950-2014



¹⁵ Un noyau est une fonction non-négative, intégrable et à valeurs réelles, notée K vérifiant les conditions suivantes :

- (i) $\int_{-\infty}^{+\infty} K(u) du = 1$
- (ii) $K(-u) = K(u)$

Figure 4.6 : Régression kernel gaussien sur le rendement du maïs, Iowa, 1950-2014



4.3.3. REGRESSION LOESS

La régression LOESS est, comme la régression kernel gaussien, une méthode non-paramétrique. L'estimation de la fonction m se fait donc de la manière décrite par les formules (4.3) et (4.4). Dans notre analyse par régression LOESS nous utilisons la fonction de noyau suivante,

$$K_h(u) = \left(1 - \frac{|u|^3}{h^3}\right)^3 \cdot I\left(\frac{|u|}{h} < 1\right) \quad (4.6)$$

En appliquant cette régression LOESS à nos données les nouveaux rendements agricoles moyens sont alors 173,1 et 50,2 boisseaux par acre, respectivement pour le maïs en Iowa et le soja dans l'Illinois. Les *Figure 4.7* et *Figure 4.8* représentent la correction des données par méthode LOESS sur la période 1950-2014.

Figure 4.7 : Régression LOESS sur le rendement du soja, Illinois, 1950-2014

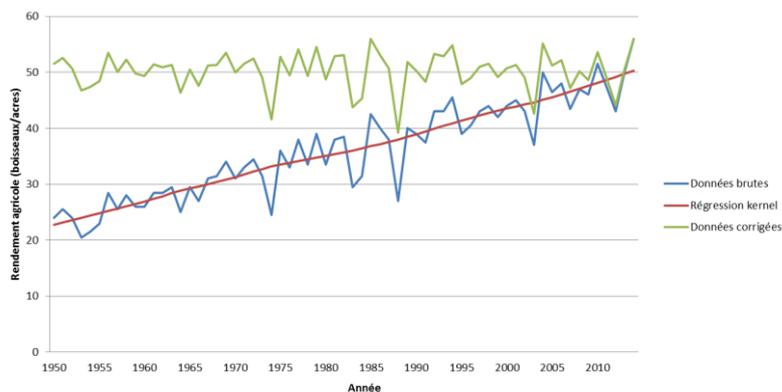
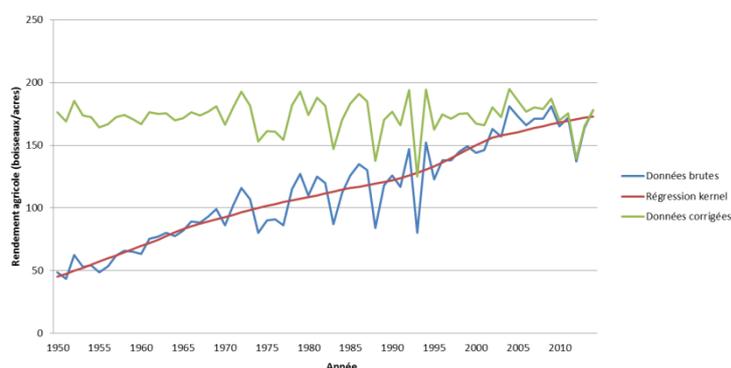


Figure 4.8 : Régression LOESS sur le rendement du maïs, Iowa, 1950-2014



4.4. TARIFICATION

4.4.1. MODELISATION DES RENDEMENTS

La modélisation des rendements agricoles par une loi normale est un sujet fortement discuté dans la littérature économique agricole depuis le milieu des années 1960, avec notamment Day (1965) qui affirme que la distribution normale des rendements agricoles « est plus une exception qu'une règle ». Aucun consensus ne se dégage pour l'instant de la littérature (Anderson, 1974 ; Just et Weninger, 1999 ; Ramirez *et al.*, 2001 ; Atwood *et al.*, 2002 ; Shaik 2010). De plus l'hypothèse d'homoscédasticité est souvent aussi remise en cause (Yang *et al.*, 1992 ; Kapiamba, 2005 ; Ker et Tolhurst, 2013). Cependant nous décidons de nous affranchir de ces questions, et nous considérons une approche simplifiée de la modélisation des rendements agricoles. En effet ce ne sont pas des points centraux dans nos travaux sur l'utilisation des images satellites en assurance agricole.

Ainsi nous choisissons une modélisation des rendements par une loi normale à l'aide des deux premiers moments des données corrigées de la tendance. Comme nous l'avons remarqué dans la partie précédente, le choix de la méthode de régression influe sur ces deux premiers moments. Le *Tableau 4.3* rassemble les différents résultats que nous obtenons sur la moyenne et la variance des rendements agricoles selon la méthode de régression appliquée.

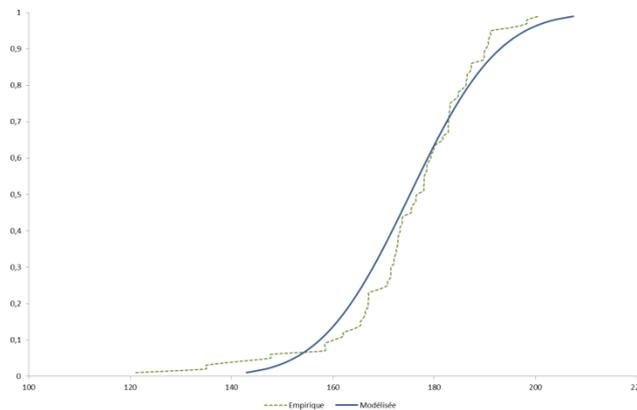
Tableau 4.3 : Moments des rendements agricoles selon trois méthodes de régressions

Culture	Moment	Méthode de régression		
		Linéaire	Kernel Gaussien	LOESS
Soja	Moyenne	49,4	48,4	50,2
Illinois	Variance	10,8	10,9	10,7
Maïs	Moyenne	175,2	166,1	173,1
Iowa	Variance	192,0	166,8	169,9

4.4.2. MODELISATION DES QUEUES DE DISTRIBUTION

Les « sinistres » de l'assurance rendements sont présents dans la queue de distribution, *i.e.* quand le rendement agricole est faible. Or le choix de la loi normale dans la modélisation implique dans certains cas un décalage entre les probabilités empiriques et modélisées, notamment dans les queues de distributions. La *Figure 4.9* représente ce décalage dans le cas de la modélisation des rendements agricoles du maïs en Iowa avec la méthode de régression linéaire.

Figure 4.9 : Fonctions de répartition des rendements agricoles du maïs en Iowa, 1950-2014



Afin de palier à ce décalage, nous modifions le poids de la queue de distribution nous choisissons un rendement de référence Y_{ref} à partir duquel la fonction de répartition est modifiée et un facteur de modification f . En fonction de la distance à ce point, nous diminuons le rendement agricole modélisé par un facteur exponentiel selon la formule suivante,

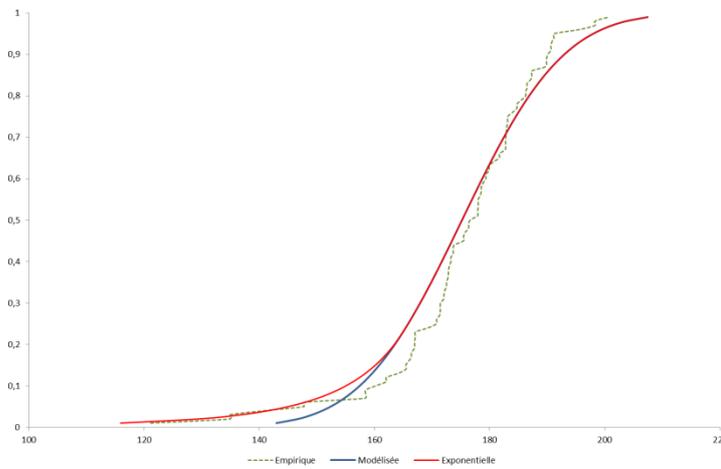
$$Y_{modifié} = Y_{ref} - (Y_{ref} - Y) \cdot e^{\left| \frac{Y_{ref} - Y}{Y_{ref}} \right| \cdot f} \quad \text{si } Y \leq Y_{ref} \quad (4.7)$$

Nous obtenons alors bien une diminution du rendement agricole modélisé,

$$Y - Y_{modifié} = (Y - Y_{ref}) \cdot \left(1 - e^{\left| \frac{Y_{ref} - Y}{Y_{ref}} \right| \cdot f} \right) \geq 0 \quad (4.8)$$

La *Figure 4.10* représente cette modification exponentielle de la queue de distribution lors de la modélisation des rendements agricoles du maïs en Iowa avec la méthode de régression linéaire. Cette méthode permet de modéliser les rendements agricoles de manière plus prudente.

Figure 4.10 : Fonctions de répartition des rendements agricoles du maïs en Iowa avec modification exponentielle, 1950-2014



4.4.3. RESULTATS

La modélisation ainsi faite , nous simulons 10 000 tirages indépendants des rendements auxquels nous appliquons la formule des montants compensatoires associée à la structure du produit d'assurance. Enfin, à l'aide d'une simple moyenne sur ces 10 000 montants, nous déduisons la perte moyenne estimée, donc la prime pure. Nous nommons cette méthode l'approche paramétrique car nous modélisons la loi d'un paramètre, ici par exemple le rendement agricole du maïs en Iowa, puis nous appliquons une méthode de Monte-Carlo. Une autre méthode, plus simple, peut être utilisée pour calculer la prime pure. Elle consiste à appliquer la formule des montants compensatoires aux données historiques corrigées de leur tendance, et de calculer la moyenne sur ces 65 montants. Cette méthode est nommée approche historique.

Le *Tableau 4.4* rassemble les valeurs ainsi calculées des primes pures pour les différentes méthodes de régression (linéaire, kernel gaussien, LOESS) et de calcul de la prime pure (paramétrique, historique). Nous remarquons de grandes différences, pouvant aller du simple au double selon la méthode utilisée. Dans notre cas nous remarquons que la régression kernel gaussien est plus conservatrice que les autres méthodes de régression. De même l'approche paramétrique implique des primes pures plus importantes.

Tableau 4.4 : Résultats des calculs de primes pures, exprimées en \$

	Iowa Maïs		Illinois Soja	
	Paramétrique	Historique	Paramétrique	Historique
Linéaire	7 284	5 577	6 801	5 161
Kernel Gaussien	14 774	9 242	9 538	7 068
LOESS	7 654	5 928	5 094	3 853

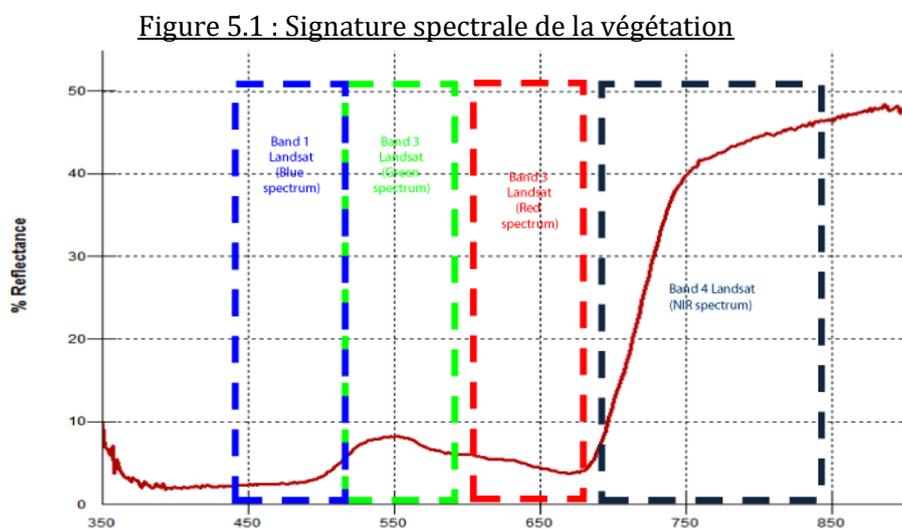
Chapitre 5

5. ASSURANCE RENDEMENT PAR IMAGES SATELLITES

5.1. LITTERATURE ET ENJEUX ACTUARIELS

5.1.1. DESCRIPTION DE L'INDICE NDVI

La réflectivité (*reflectance*) est la proportion d'énergie électromagnétique réfléchi à la surface d'un matériau ayant une épaisseur telle que ce facteur de réflexion ne change pas en augmentant son épaisseur¹⁶. Kumar et Silva (1972) ont été les premiers à étudier en détails l'activité chlorophyllienne d'une culture et de la réflectivité associée. La réflectivité spectrale d'une culture se distingue d'après leurs travaux dans la région proche de l'infrarouge ($\lambda = 700 - 1300 \text{ nm}$) et dans la région visible du rouge-orange ($\lambda = 550 - 700 \text{ nm}$) du spectre électromagnétique. A cause de l'absorption chlorophyllienne, les cultures ont de faibles réflectivités dans les régions du bleu et du rouge, avec une réflectivité un peu plus importante dans la région du vert : c'est pourquoi nous les voyons de couleur verte. La *Figure 5.1* montre la réflectivité moyenne d'une culture.



Source : Hadjimitsis et Papadavid (2013)

L'énergie proche de l'infrarouge est fortement reflétée par la surface de la plante et la quantité de cette réflexion est déterminée par les propriétés du tissu foliaire (Kumar et Silva, 1972). Ces caractéristiques anatomiques peuvent être influencées par des facteurs

¹⁶ Bureau de traduction, Travaux publics et services gouvernementaux du Canada

environnementaux tels que la salinité du sol (Gausman et Cardenas, 1968) le stade de développement (Gausman *et al.*, 1970), à l'humidité du sol ou le bilan nutritif (Gausman *et al.*, 1969). Le contraste entre le sol et la végétation est à son maximum dans la région du rouge et du proche infrarouge. Ainsi la réflectivité d'une culture peut être utilisée pour calculer différents indices de végétation corrélés avec les paramètres agronomiques et biologiques de la plante. Colwell (1956) fut le premier à utiliser des images infrarouges aériennes pour surveiller des maladies dans les cultures.

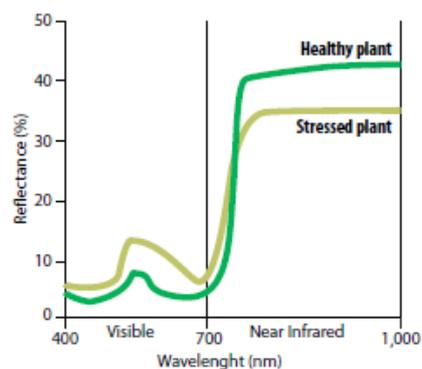
Dans notre mémoire nous utilisons l'un de ces nombreux indices : le NDVI (*Normalized Difference Vegetation Index*). Utilisé pour la première fois par Rouse *et al.* (1973), il est sûrement l'indice de végétation le plus étudié et le plus appliqué que ce soit dans le monde académique ou celui de l'assurance agricole. Il permet de détecter et de quantifier l'activité chlorophyllienne en incluant dans son calcul à la fois la réflectivité dans le proche infrarouge et dans le rouge. Il est calculé de la façon suivante,

$$NDVI = \frac{R_{NIR} - R_R}{R_{NIR} + R_R} \quad (5.1)$$

où R_{NIR} est la réflectivité dans le proche infrarouge (*Near InfraRed*) et R_R est la réflectivité dans la région du rouge visible (*Red*). Une canopée dense de végétation implique des valeurs positives, entre 0,3 et 0,8, de cet indice. Des arbustes ou des plaines d'herbe ont des valeurs modérées, de l'ordre de 0,2 à 0,3. Les forêts tropicales quant à elles ont des valeurs plus élevées, de 0,6 à 0,8. Les valeurs NDVI pour un sol nu est généralement faible, de l'ordre de 0,1. Enfin les valeurs négatives de NDVI correspondent aux sols stériles tels que les pierres, le sable ou la neige.

Cet indice est corrélé avec la santé de la plante. En effet si une culture est en bonne santé, elle absorbe la majorité de la lumière visible et reflète une grande proportion de la lumière dans le proche infrarouge. Au contraire une culture stressée reflète plus de lumière visible et moins de lumière du proche infrarouge. La *Figure 5.2* donne une représentation des signatures spectrales d'une plante en bonne santé et d'une plante stressée.

Figure 5.2 : Différence de signature spectrale d'une plante selon sa santé



Source : Govaerts et Verhulst (2010)

5.1.2. REVUE LITTÉRAIRE DE L'INDICE NDVI

L'indice NDVI a fait l'objet de nombreux papiers et d'utilisations différentes dans la littérature économique agricole, notamment depuis son calcul grâce à des images satellites mises en libre accès. Nous décrivons rapidement dans cette sous-partie deux utilisations possibles de l'indice NDVI. Elles ne forment évidemment pas la liste exhaustive des recherches menées jusqu'à maintenant, mais elles correspondent aux deux pans majeurs de la littérature sur les applications des données NDVI.

La première, et celle qui nous intéresse en premier lieu, est l'estimation précoce du rendement agricole. Plusieurs modèles de prédiction reposant pour la majorité sur une régression linéaire ont été étudiés pour notamment le maïs (Prasad *et al.*, 2006), le soja (Ma *et al.*, 2001) et le blé (Mkhabela *et al.*, 2011). D'autres techniques plus sophistiquées sont aussi mises en place, ainsi Li *et al.* (2007) utilisent des réseaux de neurones afin d'estimer le rendement agricole à partir des données NDVI. Plusieurs variables issues de la série temporelle NDVI peuvent être utilisées dans les modèles. Ainsi Mkhabela *et al.* (2011) estiment le rendement agricole en n'utilisant que la moyenne NDVI sur la période de croissance des cultures. Zhang *et al.* (2012) au contraire séparent la série en deux périodes : de la plantation à l'épiaison et de l'épiaison à la maturité. Le nombre de modèle est très important car pour chaque étude les auteurs rencontrent des spécificités géographiques allant du Zimbabwe (Svotwa *et al.*, 2014) à la Chine (Zhang *et al.*, 2012) en passant par la Hongrie (Ferencz *et al.*, 2004) ou les États-Unis (Prasad *et al.*, 2005) ; des bases de données avec notamment la différence de résolution du satellite (*Advanced Very High Resolution Radiometer* ou *Moderate-Resolution Imaging Spectroradiometer*) et donc de son historique ; de la culture étudiée (maïs, soja, blé, orge, riz, tabac, etc.). De plus certains auteurs ajoutent d'autres indices météorologiques dans leurs modèles d'estimation précoce des rendements agricoles afin d'améliorer la corrélation. Ainsi Prasad *et al.* (2006) utilisent dans leur modèle l'indice NDVI, mais aussi un indice de précipitation, un indice de température et un indice d'humidité du sol. Cependant toutes ces études s'accordent à conclure que l'indice NDVI est significativement corrélé avec le rendement agricole final (R^2 entre 0,4 et 0,9). Cette estimation peut se faire environ 2 mois avant le début des récoltes en Amérique du Nord.

La seconde utilisation des données NDVI s'est fortement développée depuis quelques années avec l'augmentation de la résolution des images satellites, le pixel est en effet passé d'une taille de plusieurs kilomètres de côté à seulement 250 mètres maintenant. Il est maintenant devenu possible de distinguer les champs d'après les images satellites NDVI. Pour certains États, des calques de données de géolocalisation sont disponibles afin de connaître, année par année, le type de culture sur chaque pixel. Cependant ces données restent marginales, et la grande

majorité des pays ne possède pas de telles cartes des cultures. L'analyse des estimations précoces de rendements montre que les séries NDVI n'ont pas les mêmes caractéristiques selon la culture. Ainsi Pervez et Brown (2010) arrivent à déterminer l'irrigation ou non d'un champ à partir du pic de la série NDVI à une résolution de 250 mètres avec une exactitude de 92% pour la Californie et 75% pour les Grandes Plaines. Peterson *et al.* (2011) se fondent aussi sur ce pic maximal de NDVI pour déterminer, pixel par pixel, l'irrigation pour plusieurs types de plante (blé, maïs, soja, etc.) avec une justesse de 88%. Kastens *et al.* (2005) utilisent quant à eux le premier et le second ordre de la série NDVI et les données historiques gouvernementales afin de déterminer la culture par pixel de 1km de côté. La création d'une telle carte est utile car elle permet d'améliorer significativement les modèles d'estimation précoce de rendements présentés précédemment (Maselli et Rembold, 2001).

5.1.3. ENJEUX ACTUARIELS DE L'INDICE NDVI

La littérature économique agricole s'est pour l'instant surtout attardée à l'étude des corrélations entre l'indice NDVI et les rendements agricoles ou le type de culture. L'apport de son application concrète en sciences actuarielles est un thème beaucoup moins abordé, même si certains auteurs tels que Turvey et Mclaurin (2012) ou Vrieling *et al.* (2014) s'y attardent pour respectivement les États-Unis et le Kenya. Turvey et Mclaurin (2012), bien qu'ils estiment que l'indice NDVI puisse être utile pour la structuration de produit d'assurance dans certaines circonstances, avertissent que son application ne doit pas devenir une généralité et qu'une réflexion doit être menée par les assureurs et les agriculteurs avant d'investir du temps et de l'argent dans de telles solutions. Néanmoins l'utilisation de NDVI, que ce soit pour la création de nouveaux produits assurantiels ou l'amélioration de la gestion de produits plus classiques, répond à certaines problématiques actuarielles.

L'innovation la plus importante qu'apporte l'utilisation du NDVI est, d'après nous, la possibilité qu'elle offre de structurer une assurance indicielle lorsque les données sur les rendements agricoles ne sont pas suffisamment fiables. En effet les valeurs des rendements agricoles ne sont utilisées que pour déterminer la structure du produit et non la tarification ou l'indemnisation. Cela permet à l'assureur de proposer des produits là où l'assurance rendement indicielle classique impliquerait un risque supplémentaire, *i.e.* dans le cas où les données de l'agence publique peu fiables (cf. Chapitre 3). C'est pourquoi l'assurance indicielle sur le NDVI, ou d'autres indices satellitaires, est fortement développée en Afrique et dans d'autres pays en voie de développement où les statistiques nationales ne sont pas considérées comme suffisamment sûres. Cependant, même si ce risque n'est plus porté par l'assureur, il est transféré sous forme de risque de base à l'assuré. En effet si les données historiques des rendements agricoles sont peu fiables, le modèle d'estimation du rendement par NDVI est donc moins fiable.

La structure et l'indemnisation pour l'assuré risquent de ne pas refléter complètement son risque, et donc augmentent le risque de base.

D'un autre côté ce risque de base supporté par l'assuré peut être diminué par l'utilisation adéquate de l'indice NDVI. Avec les nouvelles résolutions d'images NDVI (250 mètres de côté pour un pixel) les assureurs peuvent extraire la série temporelle NDVI d'un champ en particulier et ne sont plus réduits à une région de plusieurs kilomètres carrés. En se fondant sur les rendements agricoles historiques de ce champ, que l'agriculteur assuré connaît sûrement, l'assureur peut alors structurer un produit sur cet indice. Si la corrélation entre l'indice ainsi obtenu et le rendement agricole réel est plus importante que la corrélation entre ce dernier et l'estimation du rendement agricole par l'agence publique, alors le risque de base diminue par rapport à l'assurance rendement indiciaire classique.

Par contre cette diminution du risque de base est accompagnée d'une augmentation du risque d'aléa moral. Si l'indice NDVI est celui des champs de l'agriculteur, alors la modification du comportement de ce dernier (non-utilisation de pesticides, moins d'irrigation, etc.) se répercutera sur la série NDVI. Afin de contrer cet effet deux méthodes principales peuvent être proposées. Premièrement une franchise sur le prix de l'indemnité peut être appliquée, *e.g.* au lieu de prendre 4\$ par boisseau pour le maïs, l'assurance ne supporterait que des pertes à 2\$ par boisseau. Nous revenons dans ce cas à l'utilisation classique d'une franchise. Deuxièmement l'indice NDVI peut être couplé (voire remplacé entièrement) avec un indice météorologique dans un fonctionnement à double point d'attache : l'indemnisation ne commencera que si les deux indices (NDVI et climatique) dépassent les seuils définis. L'agriculteur n'ayant aucun moyen de modifier la météo, il a alors tout intérêt à ne pas modifier son comportement. Koch (2011) donne un exemple d'assurance agricole fondé sur des indices climatiques au Maroc.

Enfin la dernière amélioration que nous notons dans l'application du NDVI à l'assurance est le fait que l'indice est connu avant la récolte, et donc avant l'estimation finale des rendements agricoles. Dans le cas d'une assurance par NDVI, cela implique que l'assuré sera indemnisé plus rapidement comparé à une assurance rendement classique. Si le produit en question géré par l'assureur est une assurance rendement (ou revenus) indiciaire sur l'estimation du rendement agricole d'une agence publique, l'estimation précoce par NDVI permet d'obtenir en avance une estimation des indemnités à payer. En effet toutes les agences publiques n'investissent pas autant que l'USDA dans l'estimation précoces des productions agricoles, et la majorité se contente de fournir une estimation finale plusieurs mois après la récolte. Actuariellement parlant, l'analyse NDVI peut aboutir à une méthode de calcul des provisions de produits d'assurance agricole.

5.2. CONSTRUCTION D'UN INDICE NDVI

Dans ce chapitre nous nous attardons sur les séries temporelles NDVI de l'Iowa et de l'Illinois afin de conserver une suite logique avec le Chapitre 4 et la troisième partie. L'application d'une assurance rendement fondée sur un indice NDVI aux États-Unis n'est pas motivée par la non-fiabilité des données publiques, les estimations de l'USDA étant les plus crédibles dans la sphère agricole. L'utilisation d'un tel produit dans la pratique serait plutôt due à la volonté de réduire le risque de base par rapport à une assurance indicelle rendement fondé sur le rendement agricole de l'État en centrant la fenêtre d'étude sur la localisation des champs de l'assuré. Cependant l'analyse que nous menons serait exactement la même pour une assurance indicée sur le NDVI dans le cas où les données publiques ne serait pas fiables, *e.g.* en Afrique.

Le modèle que nous mettons en place est fondé sur le NDVI seul, si ce produit était mis en place pratiquement pour un agriculteur en particulier avec une extraction de données centrée sur ses champs, il serait bon de le coupler avec un indice météorologique (précipitation et/ou température) afin de réduire le risque d'aléa moral. Cependant, pour faire suite au Chapitre 4, nous étudions dans ce chapitre une assurance rendement NDVI sur le maïs sur la totalité de l'Iowa et le soja sur la totalité de l'Illinois. Le NDVI seul est donc suffisant.

5.2.1. DONNEES NDVI UTILISEES

Dans ce mémoire nous utilisons les données MODIS (*Moderate-Resolution Imaging Spectroradiometer*) qui nous sont fournies par le site pekko.geog.umd.edu¹⁷, résultant d'un projet de communication de la NASA (*National Aeronautics and Space Administration*), l'USDA (*United States Department of Agriculture*), de l'Université du Maryland et du GAM (*Global Agriculture Monitoring*). La résolution spatiale de ces données est de 250 mètres et une résolution temporelle de 16 jours (produit MOD44 16-day), *i.e.* tous les 16 jours nous obtenons une image représentant la moyenne des 15 derniers jours. Cette méthodologie consistant à analyser les données NDVI sur une période de 15 jours est due au fait que cet indice est fondé sur la réflectivité des plantes, or cette variable est très sensible aux éléments météorologiques tels que les nuages lorsqu'elle est analysée depuis un satellite. La publication de ces images est faite en jour de l'année constant (par exemple toujours le 17^{ème} jour de l'année). A ces images nous appliquons le calque standard disponible via MODIS (produit MOD12) afin de ne prendre en compte que les cultures, cependant pour des questions de disponibilité de données nous ne pouvons appliquer de calque pour différencier les différents types de cultures.

¹⁷ <http://pekko.geog.umd.edu/usda/beta/delivery>

Les données MODIS ne sont disponibles que seulement depuis 2000 (18 février 2000), année de la mise en service du satellite par la NASA. Les données NDVI qui étaient utilisées jusqu'alors provenaient des satellites équipés pour des images AVHRR (Advanced Very High Resolution Radiometer). Cette dernière série a été arrêtée en 2013. Nous choisissons d'utiliser le maximum de données disponibles sur les saisons 2000 à 2014 : du 18 février 2000 au 3 janvier 2015. Nous appliquons enfin une distinction géographique par État au sein de la *Corn Belt* (Illinois, Indiana, Iowa, Kansas, Minnesota, Nebraska, Ohio et Dakota du Sud). La *Figure 5.3* et la *Figure 5.4* sont des exemples d'images que nous utilisons. A partir de celles-ci une analyse pixel par pixel permet d'obtenir la valeur moyenne du NDVI pour la date et la région géographique données.

Figure 5.3 : Image NDVI (MOD44 16-days) de la *Corn Belt*, période du 10 Juin 2014 au 25 Juin 2014, 1 pixel = 2,5km

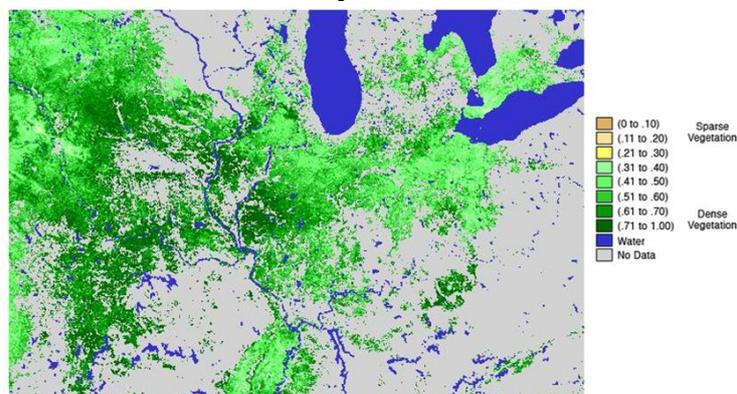
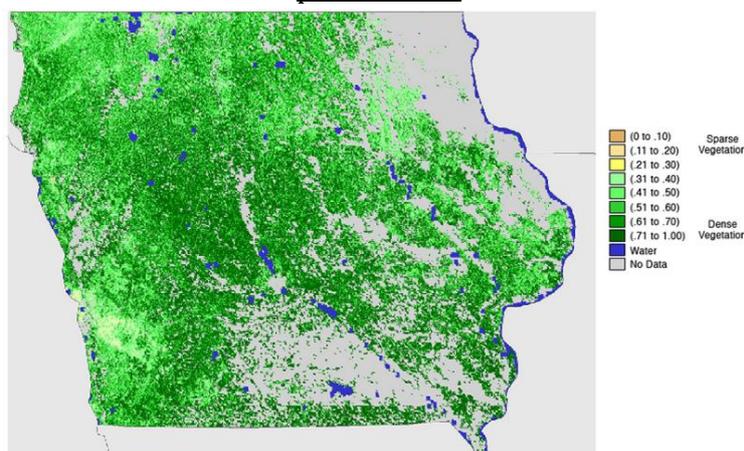


Figure 5.4 : Image NDVI (MOD44 16-days) de l'Iowa, période du 10 Juin 2014 au 25 Juin 2014, 1 pixel = 250m



Source : NASA, USDA, University of Maryland, Global Agriculture Monitoring¹⁸

Comme nous venons de l'indiquer dans la description des données, les images NDVI utilisées sont disponibles au jour de l'année constante. Pour plus de facilité dans les notations nous noterons les différentes périodes de l'année comme indiquées dans le *Tableau 5.1*. Nous tenons à

¹⁸ <http://pekko.geog.umd.edu/usda/beta/delivery>

faire remarquer à ce point que les données ne sont disponibles qu'à partir du 18 février 2000, ainsi pour l'année 2000 les périodes pour i allant de 1 à 3 sont manquantes. D'un point de vue pratique ce manque de données a un impact non-significatif pour notre étude. En effet pendant le mois de janvier la récolte du maïs ou du soja est terminée et la nouvelle saison n'est pas encore plantée. Cette notation de période sera conservée tout au long du mémoire, notamment au Chapitre 8.

Tableau 5.1 : Equivalence de notation des périodes et des jours de l'année

Période	Date de début	Date de fin	Jour de l'année (début)
1	01-janv	16-janv	1
2	17-janv	01-févr	17
3	02-févr	17-févr	32
4	18-févr	05-mars	48
5	06-mars	21-mars	66
6	22-mars	06-avr	82
7	07-avr	22-avr	97
8	23-avr	08-mai	113
9	09-mai	24-mai	129
10	25-mai	09-juin	145
11	10-juin	25-juin	160
12	26-juin	11-juil	176
13	12-juil	27-juil	192
14	28-juil	12-août	208
15	13-août	28-août	223
16	29-août	13-sept	239
17	14-sept	29-sept	254
18	30-sept	15-oct	270
19	16-oct	31-oct	286
20	01-nov	16-nov	301
21	17-nov	02-déc	317
22	03-déc	18-déc	333
23	19-déc	03-janv	349

5.2.2. CORRELATION ENTRE NDVI ET RENDEMENT AGRICOLE

En suivant une méthodologie semblable à Mkhabela *et al.* (2011) nous créons nos indices sur les pics de NDVI observés sur les États de l'Iowa et de l'Illinois durant la période 2000-2014. Pour la période i de l'année n , nous notons $V_{i,n}$ la valeur observée du NDVI. Nous définissons le pic NDVI comme la somme cumulée des NDVI entre les périodes 12 et 16 incluses, *i.e.* entre fin juin et mi-septembre. Pour l'année n nous notons le pic NDVI par P_n , il est calculé par la formule suivante,

$$P_n = \sum_{i=12}^{16} V_{i,n} \quad (5.1)$$

Pour les États de l'Iowa et de l'Illinois, la fenêtre temporelle sur laquelle nous définissons le pic correspond aux périodes durant lesquelles les moyennes des valeurs NDVI sont supérieures à 0,70. Afin d'étudier l'évolution de la corrélation nous définissons aussi $P_{i,n}$ le pic NDVI de l'année n vu à la période i . Cette variable est calculée de la manière suivante,

$$P_{i,n} = \begin{cases} 0, & \text{si } i < 12 \\ \sum_{t=12}^{\min\{i,16\}} V_{t,n} \end{cases} \quad (5.2)$$

Nous remarquons que pour $i \geq 16$, nous avons l'égalité $P_{i,n} = P_n$.

Nous étudions la corrélation entre le rendement agricole et la série temporelle des NDVI, nous procédons par application des régressions linéaires en suivant l'équation suivante pour i de 13 à 16,

$$Y = \beta_{i,0} + \beta_{i,1} \cdot X + \beta_{i,2} \cdot P_i + \epsilon_i \quad (5.3)$$

où Y est la variable aléatoire des rendements agricoles, X est la variable déterministe des années, P_i la variable des pic NDVI vu à la date i et ϵ_i un terme d'erreur. Les résultats des régressions sont rassemblés dans le *Tableau 5.2*. Nous observons que dans le cas du maïs d'Iowa la corrélation augmente plus la période à laquelle nous observons le pic est avancée, passant d'un R^2 ajusté de 0,17 pour la période 14 à 0,85 en période 16. Nous ne retrouvons cependant pas ce résultat dans l'analyse des rendements agricoles du soja en Illinois. Les corrélations augmentent jusqu'à la période 15 en atteignant un R^2 ajusté de 0,63. Cependant cette corrélation diminue à 0,50 lors de l'utilisation du pic vu à la période 16. Nous supposons que ce phénomène est dû à l'utilisation de la valeur NDVI sur l'ensemble des cultures et non pas seulement sur le soja. Des interférences de la série temporelle NDVI provenant de la culture du maïs peuvent diminuer l'information disponible sur le pic NDVI du soja, et donc la corrélation avec les rendements agricoles de cette dernière culture.

Tableau 5.2 : Régression des rendements agricoles par NDVI, 2000-2014

Période	Iowa			Illinois			
	14	15	16	13	14	15	16
$\widehat{\beta}_0$	-2863	-3890*	-3463***	-1443**	-1587**	-1680***	-1430**
$\widehat{\beta}_1$	1,4 .	1,8*	1,6***	0,7**	0,8**	0,8***	0,7**
$\widehat{\beta}_2$	93 .	111**	101***	27*	24**	22**	15*
R^2	0,17	0,47	0,85	0,52	0,59	0,63	0,50

Note : Un point correspond à une significativité à un niveau 10%, une étoile correspond à une significativité à 5%, deux étoiles à 1% et trois étoiles à 0,1%.

Ce phénomène d'interférence peut être très fortement diminué avec l'utilisation d'un calque des différentes cultures grâce à une analyse pixel par pixel de la série NDVI en suivant une

méthodologie similaire à celle développée par Kastens *et al.* (2005). Cependant cette méthode demande une puissance de calcul et un temps de développement auxquels nous n'avons malheureusement pas eu accès dans le cadre de ce mémoire. De plus l'utilisation pratique des assurances agricoles par NDVI se fait généralement sur des surfaces plus restreintes où les emplacements des cultures sont connus. Par exemple dans le cas de l'assurance d'une coopérative, cette dernière est capable de fournir les coordonnées de ses champs, et l'analyse de l'indice NDVI se serait faite sur ce calque.

L'indice que nous choisissons pour l'assurance par NDVI du maïs en Iowa est donc le pic NDVI entier, quant au soja en Illinois nous préférons prendre comme indice le pic NDVI vu à la période 15. Les *Figure 5.5* et *Figure 5.6* représentent graphiquement les corrélations que nous obtenons entre la série temporelle des NDVI, l'année et le rendement agricole pour les deux cultures étudiées dans cette partie.

Figure 5.5 : Modélisation des rendements agricoles du soja en Illinois, 2000-2014

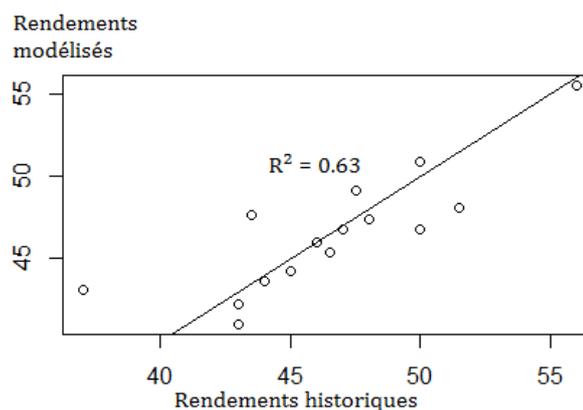
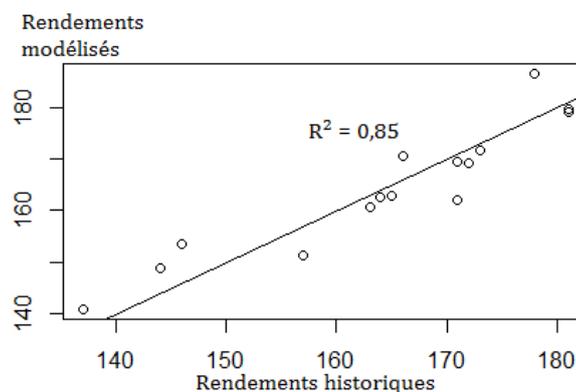


Figure 5.6 : Modélisation des rendements agricoles du maïs en Iowa, 2000-2014



5.3. TARIFICATION

5.3.1. ANALYSE DE L'INDICE

Les *Figure 5.7* et *Figure 5.8* représentent l'historique des indices de pics NDVI choisis pour la structuration de l'assurance rendement par NDVI. Nous remarquons, qu'au contraire des historiques des rendements agricoles, aucune tendance ne semble se dégager dans cette série temporelle. Nous n'appliquons donc aucune régression aux données et fondons notre modélisation sur les données historiques des indices de pics NDVI. La moyenne empirique de l'indice choisi pour le maïs d'Iowa est de 3,16 avec un écart-type de 0,13 ; la moyenne de celui pour le soja d'Illinois est de 3,94 et son écart-type est de 0,12.

Figure 5.7 : Pic NDVI, soja Illinois

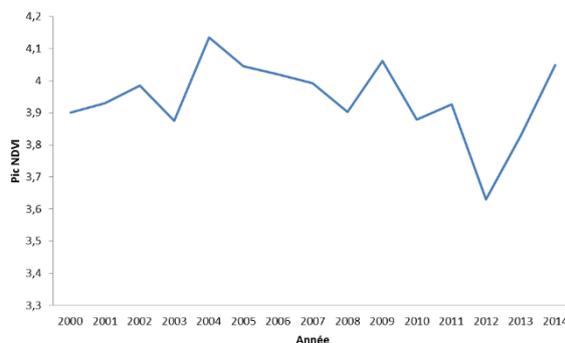
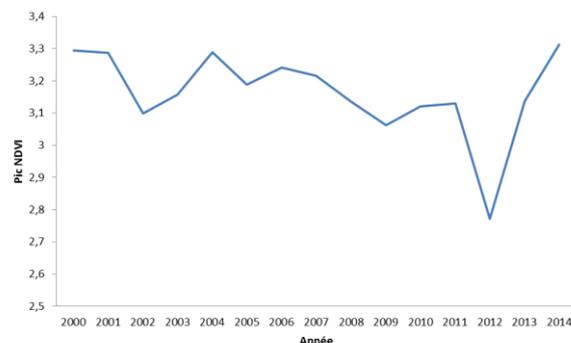


Figure 5.8 : Pic NDVI, maïs Iowa



5.3.2. STRUCTURATION DU PRODUIT D'ASSURANCE

Nous structurons maintenant le produit d'assurance paramétrique sur le rendement fondé sur l'indice que nous venons de définir. Nous gardons le même type d'appétit du risque qu'au chapitre précédent pour ce produit d'assurance NDVI. Les rendements agricoles moyens attendus pour l'année 2015 sont, d'après les modèles résumés par le *Tableau 5.2*, de 177,4 boisseaux de maïs par acre en Iowa et de 53,1 boisseaux par acre de soja en Illinois. Ces rendements agricoles attendus sont plus importants que ceux trouvés par la méthode de la moyenne olympique du Chapitre 4, résumé dans le *Tableau 4.1*. Nous appliquons les mêmes point d'attache (5%) et point de limite (30%) qu'au chapitre précédent, puis à l'aide de notre modélisation du rendement agricole en fonction du NDVI et de l'année, nous retrouvons les points d'attaches et de sorties correspondants en valeur NDVI. Ces résultats sont rassemblés dans le *Tableau 5.3*.

Les différentes variables de structure du produit d'assurance rendement par NDVI ont été choisies à partir des estimations des rendements agricoles par nos modèles, c'est pourquoi nous

avons intégré cette équivalence dans le *Tableau 5.3*. Cependant le produit d'assurance que nous étudions dans ce chapitre ne prend en compte que la structure en termes d'indice NDVI.

Tableau 5.3 : Structure de l'assurance rendement par indice NDVI

	Iowa Maïs		Illinois Soja	
	Rendement	NDVI	Rendement	NDVI
Attendu	177,4	3,94	53,1	3,16
Franchise (% rendement)	5%		5%	
Point d'attache	168,5	3,86	50,5	3,04
Limite (% rendement)	30%		30%	
Poinr de sortie	124,2	3,42	37,2	2,42
Superficie (acres)	1 000		1 000	
Prix (\$/ indice)	4	405	9	194
Somme assurée (\$)	178 200		120 280	

5.3.3. MODELISATION DE L'INDICE NDVI

Afin de modéliser l'indice NDVI nous suivons la même méthodologie que pour la modélisation des rendements agricoles du Chapitre 4. En utilisant les deux premiers moments empiriques nous déduisons les lois normales par lesquelles nous modélisons dans un premier temps les valeurs des indices de pics NDVI. Puis nous appliquons une augmentation exponentielle de la queue de distribution afin de palier au décalage entre les probabilités historiques et les probabilités modélisées. Les différentes fonctions de distribution ainsi obtenues (empiriques, modélisées, exponentielles) sont représentées dans les *Figure 5.9* et *Figure 5.10*.

Figure 5.9 : Modélisation du pic NDVI en Illinois, 2000-2014

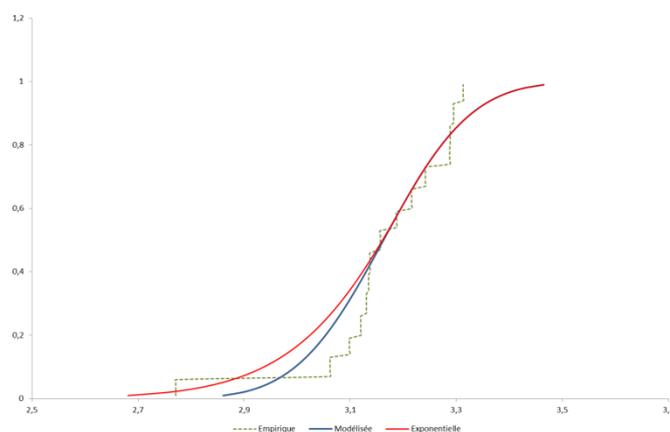
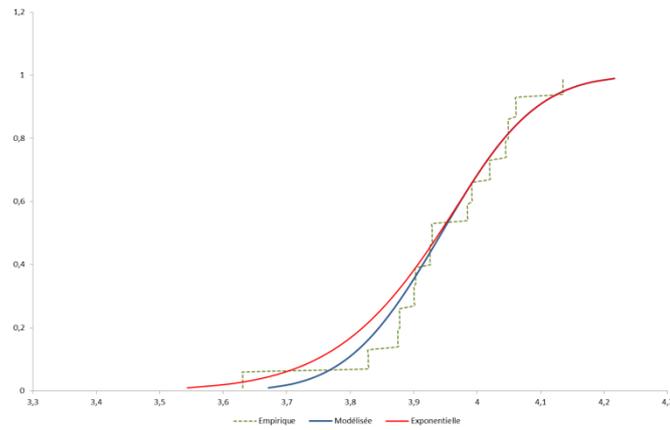


Figure 5.10 : Modélisation du pic NDVI en Iowa, 2000-2014



5.3.4. RESULTATS

Nous continuons de suivre la méthodologie de tarification du Chapitre 4 en simulant 10 000 tirages de pics NDVI, modélisés dans la partie précédente. Nous appliquons ensuite la formule des montants compensatoires sur ces données pour obtenir 10 000 versements simulés. Nous calculons enfin la moyenne de ces montants pour obtenir la prime pure. Nous comparons cette prime pure de l'approche paramétrique à la prime pure de l'approche historique, *i.e.* en appliquant la formule des montants compensatoires aux données historiques.

Les résultats de ces calculs sont rassemblés dans le *Tableau 5.4*. Nous remarquons que l'approche paramétrique est la plus conservatrice dans notre étude. Cela est en concordance avec ce que nous trouvons au Chapitre 4.

Tableau 5.4 : Résultats des calculs de primes pures, exprimées en \$

	Iowa Maïs	Illinois Soja
Paramétrique	11 815	5 273
Historique	7 039	3 477

TROISIEME PARTIE
ASSURANCE REVENU

Chapitre 6

6. DESCRIPTION DE L'ASSURANCE REVENU

6.1. STRUCTURATION DE L'ASSURANCE REVENU

L'assurance revenu couvre, comme l'indique son nom, les risques d'une chute de revenus de l'agriculteur. Comme nous l'avons vu précédemment il y a deux risques majeurs pour les revenus agricoles : le risque de rendement et le risque de prix. Nous avons déjà analysé le risque de rendement dans la deuxième partie de ce mémoire. Dans cette troisième partie nous nous attardons sur le risque de prix transféré à l'assureur.

A l'instar de l'assurance rendement, un produit type d'assurance revenu est fondé sur un indice de revenu de référence. Ce revenu de référence est la multiplication d'un rendement de référence par un prix de référence. Le rendement de référence est le même que dans les produits d'assurance de surface-rendement. Le prix de référence est le plus souvent le prix des futures à la bourse agricole locale, dans le cas des cultures étudiées, maïs et soja des États-Unis, il s'agit respectivement des prix des futures à échéance décembre et novembre cotés à la bourse de Chicago. Une fois ce revenu de référence défini, le produit d'assurance se structure de la même façon qu'un produit d'assurance surface-rendement avec un point d'attache et un point de sortie en pourcentage du revenu de référence.

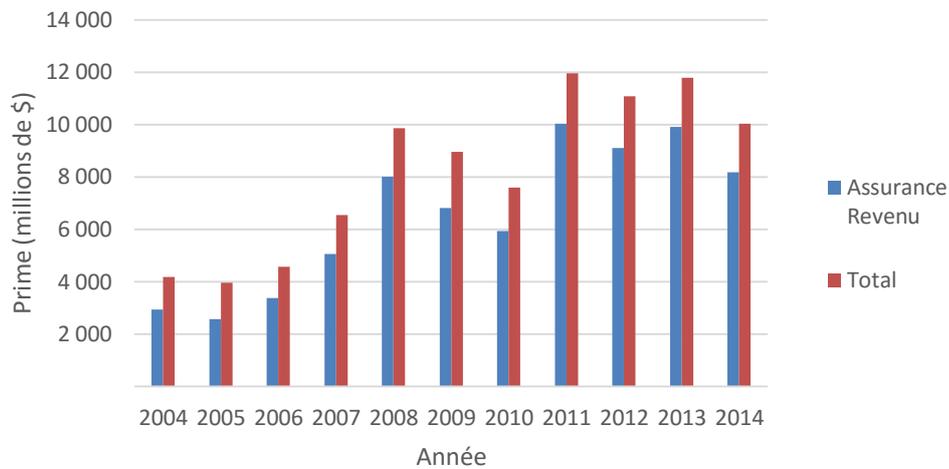
6.2. ENJEUX DE L'ASSURANCE REVENU

Les enjeux du risque de rendement agricole ayant déjà été analysés dans la deuxième partie, nous décidons de ne plus nous y attarder pour ce focaliser sur le risque de prix.

6.2.1. TAILLE DE L'ASSURANCE REVENU AUX ÉTATS-UNIS

Bien qu'en France les assurances revenus soient encore très marginales, notamment dû au fait que la PAC ne fournit que très peu d'aide pour la souscription d'assurance agricole, ce marché est fortement développé aux États-Unis. En 2014, sur les 10 milliards de dollars de prime d'assurance agricole, plus de 80% étaient destinés pour des produits d'assurance revenu pour une somme à assurer de plus de 83 milliards de dollars. La *Figure 6.1* montre que ce phénomène n'est pas récent et que les assurances revenus ont depuis 10 ans représenté la majorité des produits souscrits aux États-Unis. Les problématiques que posent les produits d'assurance revenus sont donc importantes pour les compagnies américaines.

Figure 6.1 : Marché de l'assurance revenu aux Etats-Unis, 2004-2014



Source : 2015 Crop Insurance Update, Securing America's Farmers

6.2.2. CONSEQUENCES SUR LA GESTION FINANCIERE

La conséquence directe d'un tel marché de l'assurance revenu aux Etats-Unis est que les compagnies se doivent de gérer un important portefeuille de futures sur les matières premières agricoles. Ainsi les assureurs sont très alertes aux questions de stratégies de couvertures sur les marchés dits de *commodities*. Des stratégies de couverture du risque de prix à l'aide de produits dérivés de matières premières ont déjà été développées, c'est notamment le sujet qu'étudient El Alaoui et Kreutzberger (2009). Malheureusement cette analyse est encore confidentielle à la date de rédaction du mémoire.

Dans cette troisième partie nous analysons le marché des futures sur le maïs et le soja de la bourse de Chicago afin de montrer l'existence d'une stratégie d'arbitrage statistique sur ce marché à l'aide des images NDVI. Une telle stratégie serait d'une forte utilité pour les assureurs qui, comme nous le remarquons, sont fortement exposés aux fluctuations des prix des matières premières agricoles. Elle permettrait de diminuer le risque de prix transféré aux compagnies.

Chapitre 7

7. RAPPORTS GOUVERNEMENTAUX

7.1. REVUE LITTÉRAIRE

7.1.1. FONDAMENTAUX DES PRIX DES FUTURES AGRICOLES

Au début des années 1850, afin de se protéger de prix défavorables sur les marchés, les agriculteurs commencèrent à vendre leurs récoltes aux marchands de Chicago en utilisant des contrats à terme pour lesquels la date et le prix étaient fixés à l'avance. Ce procédé est devenu une généralité, et des investisseurs ont commencé à chercher d'éventuels profits à l'aide de ces contrats, sans l'intention de détenir la récolte une fois la maturité arrivée à l'échéance. Ces contrats étaient alors échangés de main en main plusieurs fois avant maturité. Le manque de marché régulé engendra une situation chaotique (Schnepf, 2006). C'est pour cela qu'en 1865, le premier marché d'échange de futures, le *Chicago Board Of Trade* (CBOT) fut créé.

Aujourd'hui les contrats à terme sur les ressources naturelles sont échangés partout dans le monde. Les principales denrées agricoles en termes de volume d'échange dans ces marchés sont le maïs, le soja et le blé. Pour les récoltes stockables, dont font parties les trois ressources citées ci-dessus, les prix des futures jouent un rôle important pour la connaissance du prix au comptant. Baldi *et al.* (2011) analysent la relation entre le prix au comptant et les prix des futures pour les marchés du maïs et du soja. Ils montrent que les prix des futures sont utilisés comme référence pour les prix au comptant, et qu'un choc dans le prix des futures se répercute dans les prix au comptant.

Les doutes sur les quantités et les prix font partie intégrante des marchés de futures sur les ressources agricoles (Spratt, 2013). A la vue de ces caractéristiques et afin de permettre aux participants d'estimer correctement les prix il y a un fort besoin de régulation et de transparence concernant l'information sur l'offre et la demande des différentes ressources au niveau international. Fajarnes (2011) affirme que les données sur l'estimation de l'offre, des conditions climatiques, date de plantation des cultures et de récoltes jouent un rôle majeur dans le marché des futures agricoles à moyen et long terme. La distribution de ces informations est cruciale pour l'obtention d'un marché efficient.

Dans ce contexte nous allons nous intéresser plus particulièrement à une institution qui participe grandement à la circulation de ces informations : le Département d'Agriculture des États-Unis (USDA). L'USDA publie régulièrement différents rapports au public à propos du

marché des ressources agricoles. Ces rapports rassemblent des estimations sur la production des cultures et la demande au niveau national mais aussi pour d'autres pays. Isengildina-Massa *et al.* (2008) affirment que les informations produites par l'USDA sont fortement utilisées comme référence dans le marché de par leur nature exhaustive, objective et ponctuelle. Ainsi les différents acteurs du marché des futures agricoles sont très attentifs aux rapports de l'USDA.

Bien que tous les rapports de l'USDA contiennent des informations précieuses, il en est un qui attire plus l'attention des acteurs que les autres : le *World Agricultural Supply and Demand Estimates* (WASDE). Schnepf (2006) le décrit comme « la pierre angulaire des rapports de l'USDA » (« *cornerstone of USDA reports* »). Ce rapport fournit une estimation, denrée par denrée et pays par pays (pour les majeurs producteurs), de la demande et de l'offre. D'août à novembre et en janvier, il est accompagné du rapport sur l'estimation de la production des cultures aux États-Unis (*Crop Production*) du service national de statistique agricole (NASS, *National Agricultural Statistics Service*). Ces deux rapports sont préparés simultanément dans une zone sécurisée pour éviter la fuite de données et donc de délit d'initié. Il est publié à 8h30 de la côte Est, avant l'ouverture des marchés au CBOT entre le 9 et le 12 du mois (Vogel et Bange, 1999).

7.1.2. IMPACT DES RAPPORTS GOUVERNEMENTAUX

Une grande partie de la littérature économique agricole utilise les rapports de l'USDA, que ce soit le contenu ou simplement la date de publication, afin de tester si le marché des denrées agricoles est efficient. Pour ce faire ils supposent que des changements significatifs dans les prix de marché après une publication de l'USDA sont une preuve que les rapports ont un intérêt d'information. Ainsi les approches événementielles des publications USDA et leurs impacts sur les prix des futures ont été fortement appliqués. Selon Irwin *et al.* (2001) la notion fondamentale de l'approche événementielle est que dans un marché efficient, si les prix réagissent à la publication d'une information alors cette information est précieuse pour les acteurs du marché. L'événement étudié est alors la publication de l'information. La méthodologie de l'approche événementielle possède une longue histoire de travaux dans la littérature économique. L'une des premières études de ce genre revient à Dolley (1933).

Sumner et Mueller (1989) ont été les pionniers dans l'utilisation des tests statistiques (*t-test*, *F-test*, test du *Chi-2*) afin de prouver l'influence de la publication du WASDE sur les prix des futures. Ils analysent les changements de prix des futures au CBOT pour le maïs et le soja dans une fenêtre de 12 jours autour des dates de publications des rapports d'estimation des récoltes de l'USDA. Leur analyse est conduite sur la période 1961 à 1982. Les résultats de leur étude prouvent que la publication de ces rapports de l'USDA affecte le marché des futures à la bourse de Chicago. Pour ce faire ils étudient les prix de clôture, plus précisément leur changement en valeur absolue ainsi que la variance des changements. L'utilisation de *t-tests*, *F-tests* et de tests

du *Chi-2* permet de rejeter l'hypothèse nulle sous laquelle il n'y a pas de différence significative entre la valeur absolue des changements ou leur variance pour le jour de publication du rapport USDA et les autres jours de la fenêtre d'analyse. D'après leur étude, Sumner et Mueller (1989) indiquent que la réaction des marchés aux publications de l'USDA semble plus forte durant la période 1961-1979 qu'à partir de 1980. Fortenbery et Sumner (1990) ont effectué une étude très similaire sur la période 1969-1989 à propos de la réaction des marchés du maïs et du soja à la publication des rapports WASDE. Les résultats sont en concordance avec ceux de Sumner et Mueller (1989). Ils estiment notamment que la valeur de ces rapports a diminué depuis 1985, considérant même que l'USDA n'apportait plus d'information significativement précieuse. Ils considèrent que les acteurs sont devenus suffisamment qualifiés pour anticiper l'information, et ainsi, qu'aucune réaction des marchés n'est observée.

Garcia *et al.* (1997) ont étudié la valeur de l'information des rapports d'estimation des récoltes de l'USDA et des estimations du secteur privé (Conrad Leslie d'une part et Sparks Company, INC. de l'autre) durant la période 1971-1992 sur le maïs et le soja. Ils utilisent trois tests différents : la précision des estimations, la réaction des prix à l'annonce de l'information en approche événementielle, et le consentement à payer. Les résultats de leurs études ne montrent pas de diminution de la valeur des rapports de l'USDA depuis 1980, contrairement à Fortenbery et Sumner (1990). Ils justifient cette différence par le fait de ne pas avoir utilisé les mêmes tests, et estiment que les leurs (Garcia *et al.*, 1997) sont plus puissants. A propos des tests de précisions, Garcia *et al.* (1997) estiment que les sources privées ou publiques ont la même exactitude dans leurs estimations depuis le milieu des années 1980. Concernant le consentement à payer, les auteurs montrent que les traders des contrats à termes sur le maïs et le soja seraient prêts à payer afin de connaître en avance les estimations de l'USDA.

Irwin *et al.* (2001) donnent une autre limite de l'étude, et donc de ses résultats, de Fortenbery et Sumner (1990). En effet ces derniers étudient le changement de prix sur un rythme fermeture à fermeture (*close-to-close*). Or le rapport WASDE est publié en dehors des horaires d'ouverture de la Bourse de Chicago : après la fermeture avant mai 1994, ou avant l'ouverture après mai 1994. Ainsi l'influence du rapport WASDE peut être masquée par l'impact d'autres informations arrivant pendant la session. Irwin *et al.* (2001) prônent logiquement l'approche événementielle sur un changement des prix de la fermeture à l'ouverture (*close-to-open*) afin de seulement analyser l'impact de la publication du rapport WASDE. Ainsi les auteurs utilisent l'approche événementielle sur une fenêtre de 13 jours, durant la période 1985-1998, pour les futures du maïs et du soja. Leurs résultats indiquent que pendant toute la période de l'étude, les rapports WASDE qui comportent les estimations de récoltes (*NASS Crop Production*) ont un impact significatif sur les prix de marché.

La majorité des papiers dont nous venons d'évoquer forme les fondamentaux de l'étude de l'impact du rapport WASDE sur le marché des futures de denrées agricoles. Ils donnent une réponse à la question de l'existence de l'impact. Mais un autre pan de la littérature économique agricole s'attarde maintenant à quantifier cet impact, voire à l'évaluer. Ainsi nous avons vu apparaître récemment des études qui tentent d'estimer la réaction des prix, n'utilisant plus seulement la date de publication mais aussi l'information quantitative de ces rapports. Ainsi Milacek et Brorsen (2013) quantifient la valeur de marché du rapport WASDE pour le maïs et le soja aux États-Unis à l'échelle journalière. Pour ce faire, ils utilisent les stocks de fin (*ending stocks*) exprimés en boisseaux présents dans chaque rapport WASDE pendant la période 1975-2012. Les auteurs suggèrent que les mois les plus profitables en connaissance du rapport coïncident avec les mois avec les estimations de récoltes (*NASS*) ainsi que les mois importants pour le cycle des cultures elles-mêmes. Lehecka (2013) quantifie la réaction des prix futures en utilisant le pourcentage de cultures en excellente ou bonne condition donné dans les rapports de progression des cultures du Bureau des Statistiques Agricoles (*Agricultural Statistics Board*) de l'USDA. L'auteur applique une régression linéaire sur la période 1986-2012 afin de quantifier la relation entre le changement dans les prix et la variable d'intérêt. Les résultats de l'étude montrent alors une influence significativement négative de la bonne santé des cultures sur le prix, ce qui est intuitif dans le cadre de la théorie de l'offre et de la demande.

7.1.3. *DONNEES UTILISEES*

Pour cette partie de notre étude nous nous focalisons sur la période 1980-2014. Ce choix sera argumenté ci-dessous.

L'historique des prix des futures que nous utilisons sont les prix quotidiens de la Bourse de Chicago (*CBOT future prices*) pour le maïs et le soja. Contrairement à Irwin *et al.* (2001) et Isengildina-Massa (2008), nous avons choisi d'utiliser les contrats à termes dont la maturité est décembre de l'année en cours (et décembre de l'année suivante pour le mois de décembre) pour le maïs et novembre pour le soja. Dans les deux papiers cités ci-dessus, les auteurs utilisent les futures dont la maturité est la plus proche. Dans le marché standardisé il existe 5 différentes maturités pour les CBOT futures portant sur le maïs (mars, mai, juillet, septembre et décembre) et 7 pour le soja (janvier, mars, mai, juillet, août, septembre et novembre). Les auteurs justifient ce choix par le fait que ces contrats sont les plus échangés, et donc les plus liquides. De notre côté nous justifions le choix de la maturité de décembre pour le maïs et novembre pour le soja par notre approche assurantielle et non financière de l'événement. En effet, dans le cadre des assurances de revenus (*Revenue Protection*) qui couvrent les revenus des producteurs, les prix à assurer dans les États de la *Corn Belt* sont en fonction des prix futures à maturité de décembre

pour le maïs et à maturité de novembre pour le soja¹⁹. De plus ces contrats sont aussi fortement échangés avec par exemple pour le maïs une moyenne de 68 675 ordres effectués lors d'une session quotidienne dans l'année, valeur qui monte à 108 820 pour les mois de juin à novembre. Ces données ont été récupérées sur le site d'*Open Data Quandl*²⁰, plus précisément celle provenant de l'Open Data Financial Project²¹.

Concernant l'historique des dates de publications des rapports WASDE de l'USDA ou des rapports NASS, nous les avons obtenus directement depuis le site de l'USDA²².

Enfin nous utilisons aussi les estimations précoces des rendements des récoltes faites durant les mois d'août à novembre et en janvier du Service National de Statistiques Agricole (NASS), ainsi que les conditions et les progressions des cultures. Ces données ont été obtenues à l'aide de l'outil Quick Stats²³ mis à disposition par l'USDA. Nous avons téléchargé les valeurs pour 9 zones géographiques d'intérêt : la totalité du territoire des États-Unis, et 8 États de la *Corn Belt* (Illinois, Indiana, Iowa, Kansas, Minnesota, Nebraska, Ohio et Dakota du Sud). Ces États ont été choisis car ils sont les producteurs de maïs et de soja les plus importants aux États-Unis, comme le montrent les *Tableau 7.1* et *Tableau 7.2* ainsi que la *Figure 7.1* ci-après.

Tableau 7.1 : Production de maïs des États de la *Corn Belt*

Année	Production de maïs en million de boisseaux									
	ILLINOIS	INDIANA	IOWA	KANSAS	MINNESOTA	NEBRASKA	OHIO	SOUTH DAKOTA	CORN BELT	ÉTATS-UNIS
2014	2 350	1 085	2 367	566	1 178	1 602	611	787	10 546	14 216
2013	2 100	1 032	2 140	504	1 294	1 614	649	803	10 137	13 829
2012	1 286	597	1 877	375	1 374	1 292	438	535	7 775	10 755
2011	1 939	840	2 356	449	1 194	1 536	490	653	9 457	12 314
2010	1 947	898	2 153	577	1 292	1 469	523	570	9 429	12 425
2009	2 053	934	2 407	598	1 244	1 584	537	707	10 064	13 067
2008	2 130	874	2 189	483	1 166	1 394	409	585	9 229	12 043
2007	2 284	981	2 377	508	1 146	1 472	542	542	9 851	13 038
2006	1 817	845	2 050	345	1 103	1 178	471	312	8 121	10 531
2005	1 709	889	2 163	466	1 192	1 271	465	470	8 623	11 112
Moyenne	1 962	897	2 208	487	1 218	1 441	513	596	9 323	12 333
Pourcentage	16%	7%	18%	4%	10%	12%	4%	5%	76%	100%

Source : NASS

¹⁹ 2015 Crop Insurance Update, Securing America's Farmers

²⁰ www.quandl.com

²¹ www.ofdp.org

²² <http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1194>

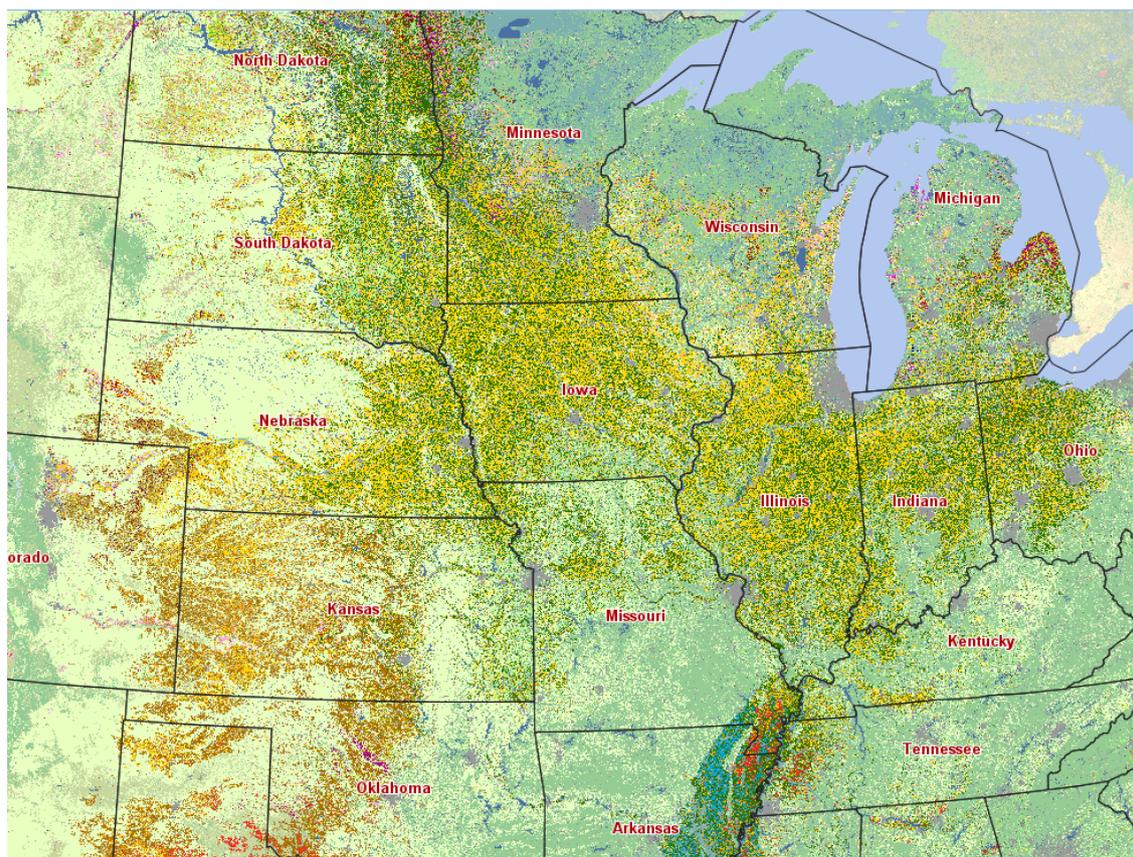
²³ <http://quickstats.nass.usda.gov/>

Tableau 7.2 : Production de soja des États de la *Corn Belt*

Année	Production de soja en million de boisseaux									ÉTATS-UNIS
	ILLINOIS	INDIANA	IOWA	KANSAS	MINNESOTA	NEBRASKA	OHIO	SOUTH DAKOTA	CORN BELT	
2014	548	307	506	143	305	289	254	230	2 582	3 969
2013	474	267	421	131	278	255	222	185	2 234	3 358
2012	384	225	419	88	305	207	207	144	1 978	3 042
2011	423	241	475	103	275	261	218	151	2 147	3 097
2010	466	259	496	140	329	268	220	157	2 335	3 331
2009	430	267	486	162	285	259	222	176	2 287	3 361
2008	429	244	450	120	265	226	161	138	2 033	2 967
2007	360	220	449	86	267	196	199	136	1 914	2 677
2006	482	284	510	99	323	251	217	131	2 296	3 197
2005	439	264	525	105	309	235	202	135	2 215	3 068
Moyenne	444	258	474	118	294	245	212	158	2 202	3 207
Pourcentage	14%	8%	15%	4%	9%	8%	7%	5%	69%	100%

Source : NASS

Figure 7.1 : Lieux des différentes cultures dans la *Corn Belt* pour la saison 2014.



Notes : les champs de maïs sont représentés en jaune, le soja en vert foncé et le blé en marron

Source : NASS

Pour 2013, l'estimation précoce d'octobre n'est pas présente, c'est pourquoi nous avons choisi de ne pas conserver cette année. Les estimations de production d'août à novembre ne sont

pas disponibles avant 1980 pour la plupart des zones géographiques sur lesquelles nous travaillons. Nous avons ainsi pris le maximum de données disponibles auxquelles nous avons accès. Pour le soja, nous notons aussi le manque d'estimations précoces de rendements agricoles pour les États du Kansas et du Dakota du Sud avant l'année 1990 non incluse.

7.2. RAPPORT WASDE

7.2.1. METHODOLOGIE

Dans la lignée d'Irwin *et al.* (2001) et Isengildina-Massa *et al.* (2008), nous procédons dans un premier temps à une approche événementielle afin de vérifier que l'information publiée à travers le WASDE affecte bien les prix futures CBOT. Pour cela nous nous attardons sur la valeur moyenne des changements des prix. Le but de cette première partie n'est pas d'estimer le sens et l'importance du changement mais seulement de prouver son existence.

Pour ce faire nous introduisons un indice temps t et un indice événement i . La fenêtre de temps que nous prendrons autour de de l'événement sera de 5 jours, ainsi t prendra ses valeurs dans l'intervalle $[-5 : 5]$ telque :

- $t = 0$ indique la session de Bourse de Chicago qui suit immédiatement la publication du rapport WASDE, il s'agit de l'événement ;
- $t = -5, \dots, -1$ représentent les sessions CBOT avant la publication du rapport ;
- $t = 1, \dots, 5$ représentent les sessions CBOT après la publication du WASDE.

L'analyse sur une fenêtre de 10 jours autour l'événement est un choix rendu classique par les travaux de Sumner et Mueller (1989), Fortenbery et Sumner (1990) et Isengildina *et al.* (2008).

L'USDA publie parfois des corrigés des rapports USDA quelques jours après la première publication, ainsi le nombre total de publications de WASDE durant la période 1980-2014 s'élève à 441. Nous décidons de ne pas prendre en compte ces corrigés en tant qu'événement. En effet certaines informations auront déjà été communiquées, l'influence ne sera pas comparable avec les premières publications. Finalement nous gardons 419 observations d'événements : 35 années d'historique, avec chaque année 12 observations (une par mois), auxquelles on soustrait l'événement d'octobre 2013 car il n'y a pas eu de publication de WASDE.

Sous l'hypothèse d'un marché efficient, dès qu'une nouvelle information (le rapport WASDE dans notre cas) devient disponible les prix doivent s'ajuster immédiatement à cet apport de nouvelles données. Le rapport WASDE est publié avant l'ouverture de la session CBOT, les prix doivent théoriquement réagir entre la fermeture de l'ancienne session et l'ouverture de la nouvelle. Nous décidons donc de suivre la méthodologie d'Irwin *et al.* (2001) qui consiste à étudier le changement de prix entre la fermeture et l'ouverture (*close-to-open*). Nous définissons alors le rendement brut de fermeture-ouverture comme suit,

$$r_{t,i} = \ln\left(\frac{p_{t,i}^o}{p_{t-1,i}^c}\right) \times 100 \quad (7.1)$$

où $p_{t,i}^o$ est le prix CBOT d'ouverture du future de la denrée agricole de la session t de l'événement i , et $p_{t-1,i}^c$ est le prix CBOT de fermeture du future de la denrée agricole de la session $t - 1$ de l'événement i . Le rendement absolu de fermeture-ouverture est quant à lui défini de la façon suivante,

$$|r_{t,i}| = \left| \ln\left(\frac{p_{t,i}^o}{p_{t-1,i}^c}\right) \times 100 \right| \quad (7.2)$$

Afin de déterminer si la publication du rapport WASDE a un impact significatif sur les prix nous utilisons un *F-test* sur un ratio de variances, proposé par Isengildina *et al.* (2008). L'hypothèse nulle est que les variances des rendements pour la session de publication du rapport et celles des sessions avant et après la publication sont identiques. Cela revient ainsi à supposer qu'aucune différence significative n'est observée dans les changements de prix entre le jour de publication et les autres jours de la fenêtre. Les deux variances que nous calculons sont alors la variance des rendements bruts lors de la session de publication d'une part, et la variance des rendements bruts lors des 10 sessions de la fenêtre d'encadrement de la publication d'autre part.

La variance des rendements bruts lors de la session de publication est définie comme suit,

$$\sigma_P^2 = \frac{1}{N-1} \sum_{i=1}^N (r_{0,i} - \bar{r}_P)^2 \quad (7.3)$$

où N est le nombre de publication de rapports WASDE utilisés dans l'estimation, $r_{0,i}$ est le rendement brut de fermeture-ouverture lors de la i -ème publication et \bar{r}_P est l'estimation de la moyenne des rendements bruts de fermeture-ouverture lors des sessions des N publications. Nous calculons cette dernière variable de la façon classique suivante,

$$\bar{r}_P = \frac{1}{N} \sum_{i=1}^N r_{0,i} \quad (7.4)$$

La variance des rendements bruts lors des sessions de la fenêtre d'encadrement de la publication est définie de la manière suivante,

$$\sigma_{WP}^2 = \frac{1}{N * 10 - 1 * N} \sum_{i=1}^N \sum_{t \in W} (r_{t,i} - \bar{r}_{WP})^2 \quad (7.5)$$

où W est l'ensemble des indices des sessions précédant et suivant la publication, pratiquement on définit W comme l'ensemble suivant $\{-5, -4, \dots, -1, 1, 2, \dots, 5\}$, et \bar{r}_{WP} est l'estimation de la moyenne des rendements bruts de fermeture-ouverture lors des sessions encadrant la session de publication. Nous calculons ce dernier paramètre comme suit,

$$\bar{r}_{WP} = \frac{1}{10 * N} \sum_{i=1}^N \sum_{t \in W} r_{t,i} \quad (7.6)$$

La F -statistique est alors calculée de la façon suivante,

$$F = \frac{\sigma_P^2}{\sigma_{WP}^2} \quad (7.7)$$

dont la distribution d'échantillonnage suit sous l'hypothèse nulle une F distribution d'après Isengildina-Massa *et al.* (2008).

Toujours en suivant les travaux d'Isengildina-Massa *et al.* (2008), nous effectuons aussi un test de Kruskal-Wallis (Kruskal et Wallis, 1952) sur les rendements absolus d'ouverture fermeture. Ce test étant non-paramétrique il permet de vérifier que les résultats obtenus avec le F -test ne sont pas dus qu'à l'hypothèse de normalité.

7.2.2. STATISTIQUES DESCRIPTIVES

Le *Tableau 7.3* rassemble quelques éléments de statistiques descriptives à propos du rendement fermeture-ouverture des futures CBOT à maturité décembre sur le maïs et à maturité novembre pour le soja sur la période de janvier 1980 à décembre 2014. La moyenne du rendement brut est faible et non significativement différente de zéro, contrairement à la moyenne du rendement absolu qui est significatif, ce qui montre la variabilité dans les changements de prix. Les skewness et kurtosis sont significatifs, ce qui semblent indiquer que les rendements bruts sur les futures ne sont pas normaux. Nous validons cette hypothèse avec un test de Jarque-Bera. Ce résultat est en concordance avec les travaux de Sumner et Mueller (1989), Isengildina-Massa (2008) et Lehecka (2013).

Tableau 7.3 : Statistiques descriptives sur les rendements fermeture-ouverture des futures CBOT sur le maïs (maturité décembre) et le soja (maturité novembre), 1980-2014

	Maïs		Soja	
	r	r	r	r
Moyenne	-0,02	0,52	-0,01	0,5
Écart type	1,26	1,26	0,94	0,94
Variance	1,59	1,59	0,88	0,88
Min	-16,93	0,00	-28,21	0,00
1er Quartile	-0,31	0,11	-0,28	0,11
Médiane	-0,08	0,28	-0,04	0,27
3ème Quartile	0,21	0,59	0,26	0,58
Max	23,95	23,95	13,41	28,21
Skewness	4,46	9,86	-3,28	9,18
Kurtosis	112	159	119	200
Jarque-Bera	4,37.e6***	9,12.e6***	4,98.e6***	14,37.e6***
Observation	8824	8824	8824	8824

7.2.3. RESULTATS

Dans un premier temps nous allons donner une vision graphique de l'influence du WASDE sur les prix futures CBOT. Bien que n'étant pas une preuve statistique en soit, cette première approche permet de mieux appréhender le phénomène que nous étudions. Nous présentons pour différentes périodes la moyenne des rendements absolus de fermeture-ouverture ($|r_{t,i}|$) pendant la session du CBOT qui suit la publication du rapport WASDE ($t = 0$), mais aussi les sessions suivantes ($t = 1, \dots, 5$) et précédentes ($t = -1, \dots, -5$). Ces résultats sont montrés de la *Figure 7.2* à la *Figure 7.7*. Cette moyenne est effectuée sur les 419 publications entre 1980 et 2014. Nous observons que, durant la session qui suit directement la publication du rapport WASDE le rendement absolu moyen est 1,6 fois supérieur aux sessions suivantes et précédentes pour le maïs et 1,7 fois supérieur pour le soja. Ce facteur augmente lors des mois où le WASDE est accompagné du rapport sur la production NASS, atteignant 3,7 pour le mois de janvier pour le maïs et 3,4 pour le soja durant ce même mois. Cependant certains mois, l'influence de l'événement étudié semble moins significative, par exemple en juillet où ce facteur est même inférieur à 1, atteignant 0,9 pour le maïs et le soja.

Figure 7.2 : Tous les mois, maïs

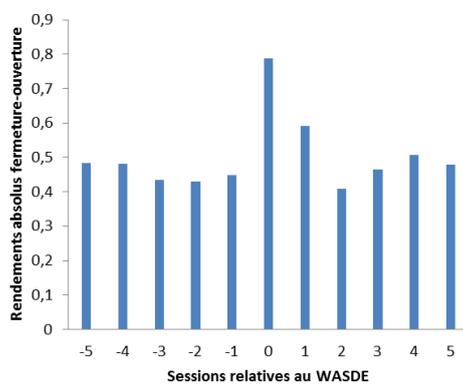


Figure 7.3 : Tous les mois, soja

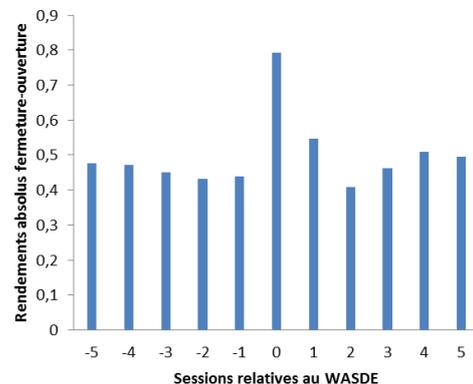


Figure 7.4 : Janvier, maïs

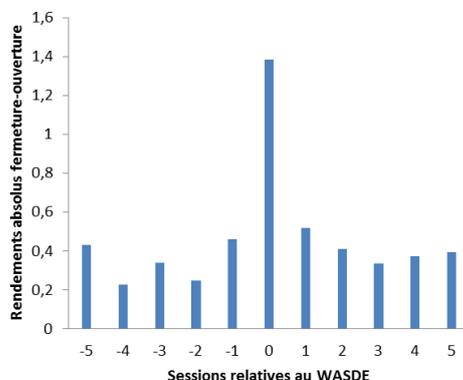


Figure 7.5 : Janvier, soja

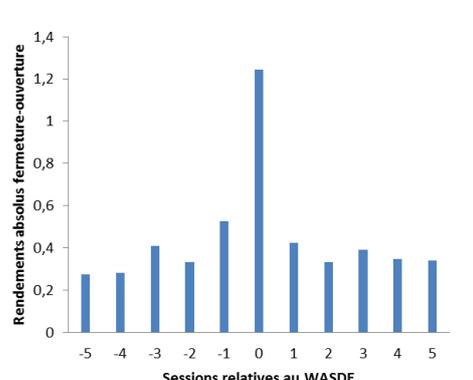


Figure 7.6 : Juillet, maïs

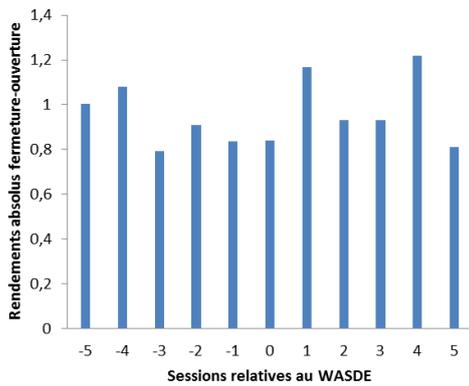
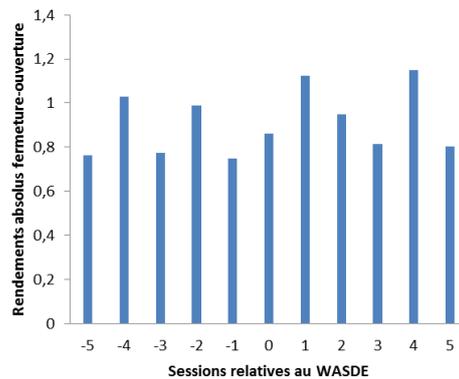


Figure 7.7 : Juillet, soja



Le *Tableau 7.4* et le *Tableau 7.5* présentent les résultats de l'analyse sur l'impact de la publication du rapport WASDE sur les prix des futures CBOT pour le maïs entre janvier 1980 et décembre 2014. La variance des rendements fermeture-ouverture durant les sessions suivant l'événement est 2,9 fois plus importante que pendant les autres sessions de la fenêtre de l'événement pour le maïs, et ce facteur atteint 3 pour le soja. Les deux tests que nous appliquons indiquent que la variance durant la session d'intérêt est significativement plus élevée au niveau 1% pour les deux denrées agricoles. L'information contenue dans le rapport est donc en général précieuse pour les acteurs du marché des futures CBOT en changeant les estimations des participants et en réduisant l'incertitude.

Si l'on se restreint aux seuls mois où le rapport WASDE est accompagné du rapport NASS sur la progression des cultures, la variance des rendements lors de l'événement est alors 5,7 fois plus importante que durant les autres jours considérés pour le maïs et 5,3 pour le soja. Cette augmentation de la variabilité des prix est confirmée par les deux tests à un niveau de 1%. La publication simultanée des rapports WASDE et NASS a donc plus d'impact sur les prix futures. Ces résultats sont en concordance avec les travaux effectués précédemment par Fortenbery et Sumner (1990), Irwin *et al.* (2001) et Isengildina-Massa *et al.* (2008). Au contraire, lors de la publication du rapport WASDE seul, la variance de l'événement est seulement 1,17 fois supérieure pour les deux cultures étudiées. Seul le test non paramétrique pour le soja indique que ce changement est significatif à un niveau de 5%. Cela montre que la publication du rapport NASS est très précieuse pour les acteurs du marché CBOT.

Nous effectuons une analyse mois par mois de l'impact de la publication des rapports gouvernementaux sur les prix futures. Les deux mois où l'augmentation de la variance est la plus importante sont janvier et août pour le maïs, dont les variances de la session de publication sont respectivement 12,8 et 5,6 fois plus importantes que les variances de la fenêtre d'analyse, respectivement. Cela peut s'expliquer par le fait qu'il s'agit des dates où la première estimation

(août) et l'estimation finale (janvier) de rendement agricole, donc d'offre, est publiée par le NASS. Cette augmentation est indiquée comme significative par les deux tests appliqués à un niveau de 1%. Pour le soja les deux mois où l'augmentation de la variance est la plus importante sont janvier et octobre, avec une variance respectivement 8,2 fois et 9,6 fois plus élevée lors de la publication du WASDE. Les autres mois dont la hausse de la variabilité des prix est significative pour les deux cultures sont notamment les mois de septembre à novembre, ce qui correspond aux mois durant lesquels le maïs et le soja sont récoltés aux États-Unis. Ainsi, si l'on omet la non significativité de l'impact en septembre d'après le test de Kruskal-Wallis pour le maïs et de même le mois de novembre pour le soja, les mois dont l'influence de la publication sont les mois de récoltes, *i.e.* d'août à novembre²⁴, et le mois d'estimation finale des récoltes, *i.e.* janvier. Seule la publication du WASDE de décembre n'a pas d'impact significatif sur les prix futures CBOT d'après les deux tests. Ce résultat semble indiquer que le rapport sur la production des cultures NASS fournit une information plus précieuse pour les acteurs du marché que le rapport WASDE.

Tableau 7.4 : Test de la volatilité des rendements des futures CBOT sur le maïs pour le rapport WASDE, rendements fermeture-ouverture, Janvier 1980-Décembre 2014

Rapports	N	Variance de la session WASDE	Variance avant et après la session WASDE	Différence entre les deux variances	F-statistique	Kruskal-Wallis statistique
Tous les mois	419	1,81	0,62	1,19	2,93***	23,99***
WASDE et NASS	174	3,30	0,58	2,72	5,65***	42,03***
Seulement WASDE	245	0,76	0,65	0,11	1,18*	1,05
Janvier	35	4,10	0,32	3,78	12,64***	25,08***
Février	35	0,50	0,15	0,35	3,36***	2,61
Mars	35	0,35	0,34	0,01	1,03	0
Avril	35	0,49	0,31	0,18	1,60**	0,43
Mai	35	0,88	0,56	0,32	1,58**	1,19
Juin	35	1,19	0,89	0,30	1,33	0,49
Juillet	35	1,41	1,91	-0,50	0,74	0,31
Août	35	5,47	0,97	4,50	5,66***	14,40***
Septembre	35	1,21	0,42	0,79	2,90***	0,05
Octobre	34	4,22	0,79	3,43	5,35***	7,05**
Novembre	35	1,49	0,41	1,08	3,61***	10,02***
Décembre	35	0,32	0,35	-0,03	0,91	0,93

Notes : N indique le nombre de rapport. Les rendements sont calculés comme la différence des logarithmes népériens des prix multipliée par 100. Deux étoiles indiquent une significativité à un niveau de 5%, et trois étoiles à niveau de 1%.

²⁴ Usual Planting and Harvesting Dates for U.S. Field Crops

Tableau 7.5 : Test de la volatilité des rendements des futures CBOT sur le soja pour le rapport WASDE, rendements fermeture-ouverture, Janvier 1980-Décembre 2014

Rapports	N	Variance de la session WASDE	Variance avant et après la session WASDE	Différence entre les deux variances	F-statistique	Kruskal-Wallis statistique
Tous les mois	419	1,88	0,62	1,19	3,02***	39,00***
WASDE et NASS	174	3,57	0,68	2,72	5,23***	48,67***
Seulement WASDE	245	0,68	0,58	0,11	1,17	7,01**
Janvier	35	2,95	0,36	3,78	8,19***	32,06***
Février	35	0,58	0,21	0,35	2,81***	1,45
Mars	35	0,31	0,37	0,01	0,86	1,1
Avril	35	0,52	0,17	0,18	3,07***	6,39**
Mai	35	0,77	0,36	0,32	2,13***	8,12**
Juin	35	0,48	0,83	0,30	0,58	0,87
Juillet	35	1,39	1,79	-0,50	0,78	5,81
Août	35	5,58	1,13	4,50	4,96***	8,96**
Septembre	35	1,80	0,54	0,79	3,32***	7,86**
Octobre	34	6,43	0,67	3,43	9,54***	7,77**
Novembre	35	1,13	0,71	1,08	1,60**	5,74
Décembre	35	0,60	0,30	-0,03	2,00***	0,07

Notes : N indique le nombre de rapport. Les rendements sont calculés comme la différence des logarithmes népériens des prix multipliée par 100. Deux étoiles indiquent une significativité à un niveau de 5%, et trois étoiles à niveau de 1%.

7.3. ESTIMATIONS PRECOCES DU NASS

7.3.1. METHODOLOGIE

Dans cette partie nous quantifions la relation entre les prix futures CBOT et l'estimation de la production des cultures faite par le Service National de Statistiques Agricoles (NASS) dont le rapport est publié en même temps que le WASDE pour les mois d'août à novembre et en janvier. Pour cela nous nous inspirons de la méthodologie de Lehecka (2013). Dans son travail l'auteur utilise le pourcentage de culture en excellente et bonne condition, nous utilisons de notre côté les estimations des rendements des cultures disponibles pour chaque État et pour l'ensemble des États-Unis.

Si l'on se place sous l'hypothèse que le dynamisme des prix à long terme est déterminé par un équilibre de l'offre et de la demande, un changement dans l'offre doit avoir un impact sur les prix. Plus spécifiquement si l'offre augmente, les prix baissent théoriquement. Depuis 2012 le rapport publié par NASS donne une estimation précoce de la production, notamment grâce à une estimation précoce des surfaces récoltées. Cependant cette estimation n'a pas beaucoup

d'années historiques. Néanmoins depuis 1980, les rapports NASS d'août à novembre et celui de janvier, comportent une estimation précoce²⁵, État par État, du rendement agricole des cultures.

Nous gardons les mêmes notations que nous avons introduites dans la partie de l'analyse de l'existence de l'impact de la publication du rapport WASDE sur les prix futures CBOT. Nous avons cependant besoin d'introduire de nouvelles notations pour cette partie de l'analyse.

Les nouveaux événements que nous étudions sont les sessions de bourse qui suivent la publication de l'estimation des rendements agricoles par le NASS. Nous notons le nouvel indice d'événements j . Dans l'historique que nous étudions nous avons 5 événements par an, et 34 années d'historique, soit 170 au total. Lors de chaque événement nous observons la réalisation de 9 différentes variables : les 8 États de la *Corn Belt* ainsi que l'ensemble des États-Unis, notées Y_k pour k de 1 à 9. Les réalisations sont les estimations des rendements agricoles fournies par le NASS, nous les notons $Y_{k,j}$. Nous ajoutons une variable dont l'observation n'est pas directement donnée dans le rapport NASS, il s'agit de l'estimation du rendement agricole de la *Corn Belt* $Y_{10,j}$. Pour cela nous définissons cette variable comme la moyenne des rendements agricoles des États de la *Corn Belt* pondérée par l'estimation précoce des surfaces plantées dans ces mêmes États.

Cependant nous cherchons à quantifier l'influence de l'apport d'une nouvelle information précieuse pour les acteurs du marché sur les prix futures CBOT. Les réalisations $Y_{k,i}$ prises séparément ne permettent pas de quantifier la nouvelle information qu'en tirent les acteurs, nous définissons donc la variable d'information $X_{k,j}$ pour k de 1 à 10 de la manière suivante,

$$X_{k,j} = \left(\frac{Y_{k,j}}{Y_{k,j-1}} - 1 \right) \times 100 \quad (7.8)$$

pour les événements ayant lieu durant les mois de septembre, octobre, novembre et janvier, et de la façon suivante pour les événements se déroulant en août de l'année N ,

$$X_{k,j} = \left(\frac{Y_{k,j}}{\overline{Y_{k,N}}} - 1 \right) \times 100 \quad (7.9)$$

où $\overline{Y_{k,N}}$ représente la moyenne olympique des estimations finales des rendements agricoles sur les 5 dernières années. En notant $Y_{k,n}^f$ l'estimation finale du rendement agricole pour la zone géographique k et l'année n , nous calculons alors $\overline{Y_{k,N}}$ comme suit,

$$\overline{Y_{k,N}} = \frac{1}{3} \left(\sum_{n=1}^5 Y_{k,N-n}^f - \max_{n \in [1:5]} \{Y_{k,N-n}^f\} - \min_{n \in [1:5]} \{Y_{k,N-n}^f\} \right) \quad (7.10)$$

En effet il serait absurde de comparer l'estimation du rendement agricole en août pour la récolte N avec l'estimation du rendement agricole en janvier pour la récolte $N - 1$. La variable $\overline{Y_{k,N}}$ représente le rendement agricole espéré par les acteurs du marché sans information autre que

²⁵ Les estimations d'août à novembre sont des estimations précoces, l'estimation de janvier est considérée comme l'estimation finale de la production pour les cultures de l'année qui vient de s'écouler.

l'historique des rendements agricoles. Nous supposons ici qu'il n'y a pas de nouvelles informations suffisamment importantes avant la publication du rapport d'août permettant aux acteurs du marché d'estimer le rendement agricole de cette année (hypothèse que nous savons fautive d'après Lehecka (2013), mais qui a le mérite d'être simple). Nous choisissons la moyenne olympique sur 5 ans car elle est une variable de référence dans le monde agricole, notamment en assurance avec le programme ARC (*Agricultural Risk Coverage Program*) de couverture des risques agricoles (Smith *et al.*, 2014).

Afin de vérifier et de quantifier l'impact de la publication des estimations précoces de rendement agricole, nous suivons la méthodologie décrite par Lehecka (2013). Ainsi nous testons l'hypothèse nulle sous laquelle l'estimation du rendement agricole publiée en même temps que le WASDE n'a pas d'impact sur les prix futures CBOT à l'aide des régressions suivantes pour k de 1 à 10,

$$r_{0,j} = \beta_{k,0} + \beta_{k,1} \cdot X_{k,j} + \epsilon_{k,j} \quad (7.11)$$

où $r_{0,j}$ est le rendement fermeture-ouverture à la session qui suit directement la publication du rapport j , il est calculé de la même manière que dans la formule (7.1). L'hypothèse nulle à tester est donc $H_0 : \beta_{k,1} = 0$ pour les différents k .

7.3.2. RESULTATS

Les *Tableau 7.6* et *Tableau 7.7* rassemblent les résultats des différentes régressions que nous avons effectuées afin de quantifier l'impact de l'estimation des rendements agricoles sur les prix futures CBOT, d'août 1980 à janvier 2015. Pour plus de lisibilité nous avons mis en gras les régressions pour lesquelles le coefficient $\widehat{\beta}_1$ estimé est significativement différent de zéro à un niveau 1%. Nous remarquons que pour l'ensemble des zones géographiques analysées, la publication de l'estimation du rendement agricole par le NASS a un impact significatif à un niveau 1%. De plus les estimations des $\widehat{\beta}_1$ de l'équation (7.11) sont négatives, ce qui correspond bien à l'intuition que nous avons. Une ré-estimation à la hausse des rendements agricoles et la publication de celle-ci par le NASS changent les attentes des acteurs du marché à propos de la production. L'estimation de la production est revue à la hausse et, par l'équilibre de l'offre et de la demande, les prix des futures baissent. L'estimation précoce des rendements agricoles est ainsi précieuse pour les acteurs du marché.

Les différentes zones géographiques ne semblent pas avoir toutes le même impact sur les prix futures CBOT à la vue de nos résultats. Ainsi quatre zones ont un impact plus significatif, à un niveau 0,1% pour l'ensemble des mois et sur la restriction de septembre à janvier, il s'agit de l'Iowa, le Nebraska, les États-Unis et la *Corn Belt*. D'après le *Tableau 7.1* nous observons que ces deux États sont parmi les 3 trois plus grands producteurs de maïs aux États-Unis. La production dans ces États représente respectivement environ 18% et 12% de la production totale du pays.

Un changement de production par rapport aux précédentes estimations dans ces États est donc jugé comme une information plus précieuse par les acteurs du marché qu'un changement dans d'autres États. Ce résultat est logique avec l'intuition et est confirmé par la significativité des zones nationale et *Corn Belt* au même niveau. D'après le *Tableau 7.1*, l'Illinois est un producteur important de maïs, environ 16%, cependant nos résultats ne montrent qu'une significativité à un niveau 1% pour le $\widehat{\beta}_1$ correspondant.

Nous restreignons notre étude à différentes périodes. Nous effectuons notamment une distinction entre août et les autres mois de publications. En effet la méthodologie n'est pas exactement la même entre ces deux périodes. Pour les quatre zones définies au-dessus, la valeur absolue des $\widehat{\beta}_1$ augmente lors de l'exclusion d'août de l'étude en gardant le même niveau de significativité à 0,1%. De plus la significativité des estimateurs de β_1 lors de la restriction à août, et donc de l'impact de la publication de l'estimation du rendement agricole sur les prix futures CBOT, diminue à un niveau à 5% pour les États-Unis, la *Corn Belt* et le Nebraska, et à un niveau 10% pour l'Iowa. Or si nous nous référons aux résultats présentés dans les *Tableau 7.4* et *Tableau 7.5* la publication des rapports WASDE et NASS en août augmente fortement la variance de la session qui la suit directement. Pour expliquer ce résultat, nous émettons l'hypothèse que notre choix de la moyenne olympique sur 5 ans n'est pas le meilleur indice possible. En effet nous considérons par ce fait que la publication du rapport NASS est la première information de l'année en ce qui concerne la progression des cultures. Or d'autres rapports à propos de la santé du maïs sont publiés avant le rapport NASS d'août, c'est le cas des rapports de Progression de Culture (*Crop Progress, CP*) publiés hebdomadairement à partir d'avril. Lehecka (2013) montre que la publication de ces derniers a un impact sur les prix futures CBOT. Ainsi, même sans l'existence de publication d'estimation chiffrée du rendement agricole avant août, l'estimation précoce du NASS de ce mois n'est pas la première information que le marché reçoit et prend en compte.

Nous effectuons une analyse mois par mois de l'impact de la publication de l'estimation du rendement agricole sur les prix futures CBOT. En concordance avec les résultats du *Tableau 7.5* et *Tableau 7.4*, l'influence de la publication est plus forte durant les mois d'octobre et novembre comparée avec le mois de septembre. Le deuxième point que nous notons est la relative faible significativité de l'impact de la publication finale des rendements agricoles (janvier) sur les prix à un niveau de 5% pour l'Indiana, les États-Unis et la *Corn Belt*. Nous avons cependant montré que la publication des rapports WASDE et NASS en janvier augmente significativement la variabilité des prix durant la session qui la suit directement. De plus la publication du WASDE de décembre n'implique pas de changement de prix significatif, donc si une information supplémentaire a affecté le marché entre les publications de novembre et de janvier, elle ne provient *a priori* pas du rapport qui les sépare.

**Tableau 7.6 : Impact de l'estimation précoce des rendements agricoles du rapport NASS sur les rendements des futures CBOT, Application au maïs,
Août 1980-Janvier 2015**

	ILLINOIS		INDIANA		IOWA		KANSAS		MINNESOTA		NEBRASKA		OHIO		SOUTH DAKOTA		ÉTATS-UNIS		CORN BELT	
	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$
Tous les mois	0,12 (0,14)	-0,06** (0,02)	0,13 (0,13)	-0,07*** (0,02)	0,14 (0,13)	-0,10*** (0,03)	0,15 (0,13)	-0,09*** (0,03)	0,13 (0,14)	-0,07** (0,03)	0,17 (0,13)	-0,14*** (0,04)	0,12 (0,14)	-0,08** (0,02)	0,17 (0,14)	-0,07** (0,02)	0,16 (0,13)	-0,13*** (0,03)	0,14 (0,13)	-0,12*** (0,03)
Réestimation (Sep-Jan)	0,00 (0,14)	-0,09* (0,04)	0,03 (0,14)	-0,13** (0,04)	-0,02 (0,14)	-0,13*** (0,04)	0,04 (0,14)	-0,11** (0,04)	0,01 (0,14)	-0,05 (0,04)	0,01 (0,14)	-0,17*** (0,05)	-0,01 (0,14)	-0,09* (0,04)	0,04 (0,14)	-0,06 (0,03)	0,02 (0,14)	-0,22*** (0,06)	0,00 (0,13)	-0,21*** (0,05)
Août	0,61 (0,40)	-0,05 (0,03)	0,57 (0,39)	-0,06 (0,03)	0,72 (0,39)	-0,09 (0,04)	0,62 (0,39)	-0,08 (0,04)	0,64 (0,39)	-0,08* (0,04)	0,83* (0,40)	-0,15* (0,07)	0,63 (0,39)	-0,07* (0,03)	0,67 (0,39)	-0,07* (0,03)	0,69 (0,39)	-0,10* (0,05)	0,67 (0,38)	-0,10* (0,04)
Septembre	-0,31 (0,20)	-0,02 (0,04)	-0,31 (0,19)	-0,04 (0,04)	-0,32 (0,20)	-0,03 (0,05)	-0,29 (0,19)	0,03 (0,06)	-0,33 (0,19)	-0,08 (0,06)	-0,33 (0,20)	-0,06 (0,06)	-0,34 (0,19)	-0,07 (0,04)	-0,29 (0,19)	-0,04 (0,04)	-0,33 (0,19)	-0,06 (0,06)	-0,33 (0,19)	-0,05 (0,06)
Octobre	0,21 (0,32)	-0,21** (0,07)	0,22 (0,33)	-0,33* (0,13)	0,36 (0,33)	-0,26** (0,09)	0,35 (0,33)	-0,37** (0,13)	0,16 (0,37)	-0,01 (0,10)	0,33 (0,32)	-0,31** (0,10)	0,14 (0,36)	-0,05 (0,12)	0,11 (0,36)	0,08 (0,13)	0,38 (0,33)	-0,45** (0,15)	0,35 (0,33)	-0,39** (0,13)
Novembre	0,12 (0,21)	-0,11 (0,09)	0,15 (0,21)	-0,16 (0,08)	0,06 (0,18)	-0,17*** (0,05)	0,11 (0,20)	-0,13 (0,06)	0,10 (0,20)	-0,13 (0,07)	-0,01 (0,20)	-0,22* (0,09)	0,23 (0,20)	-0,21** (0,07)	0,21 (0,21)	-0,11* (0,05)	0,15 (0,19)	-0,32** (0,10)	0,11 (0,18)	-0,28** (0,08)
Janvier	0,05 (0,36)	-0,10 (0,20)	0,18 (0,32)	-0,31* (0,15)	-0,08 (0,36)	-0,28 (0,21)	0,20 (0,33)	-0,12 (0,07)	0,09 (0,35)	-0,02 (0,11)	0,17 (0,34)	-0,19 (0,15)	0,03 (0,35)	-0,10 (0,10)	0,12 (0,33)	-0,08 (0,06)	0,09 (0,32)	-0,76* (0,32)	-0,06 (0,32)	-0,63* (0,25)

Note : Un point correspond à une significativité à un niveau 10%, une étoile correspond à une significativité à 5%, deux étoiles à 1% et trois étoiles à 0,1%.

Tableau 7.7 : Impact de l'estimation précoce des rendements agricoles du rapport NASS sur les rendements des futures CBOT, Application au soja.

Août 1980-Janvier 2015

	ILLINOIS		INDIANA		IOWA		KANSAS		MINNESOTA		NEBRASKA		OHIO		SOUTH DAKOTA		ÉTATS-UNIS		CORN BELT	
	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$
Tous les mois	0,18 (0,14)	-0,11*** (0,03)	0,19 (0,14)	-0,09*** (0,03)	0,16 (0,14)	-0,09** (0,03)	0,12 (0,19)	-0,05* (0,02)	0,15 (0,14)	-0,06* (0,03)	0,20 (0,14)	-0,09*** (0,03)	0,17 (0,14)	-0,11*** (0,02)	0,13 (0,20)	-0,04 (0,03)	0,21 (0,14)	-0,14*** (0,03)	0,2 (0,14)	-0,17*** (0,03)
Réestimation (Sep-Jan)	0,28 (0,16)	-0,11*** (0,03)	0,26 (0,16)	-0,09*** (0,03)	0,24 (0,16)	-0,09** (0,03)	0,23 (0,22)	-0,05* (0,02)	0,24 (0,16)	-0,08** (0,03)	0,29 (0,16)	-0,09** (0,03)	0,27 (0,16)	-0,11*** (0,03)	0,25 (0,22)	-0,05 (0,03)	0,30 (0,16)	-0,14*** (0,04)	0,08 (0,14)	-0,21*** (0,05)
Août	0,66 (0,40)	-0,07 (0,04)	0,69 (0,39)	-0,08* (0,05)	0,69 (0,41)	-0,07 (0,06)	0,69 (0,51)	-0,05 (0,03)	0,71 (0,41)	-0,07 (0,05)	0,80 (0,41)	-0,07 (0,05)	0,73 (0,36)	-0,13** (0,04)	0,68 (0,53)	-0,04 (0,05)	0,86* (0,40)	-0,12 (0,06)	0,74 (0,38)	-0,14* (0,06)
Septembre	-0,06 (0,23)	-0,06 (0,05)	-0,11 (0,23)	-0,07 (0,06)	-0,12 (0,21)	-0,13** (0,04)	-0,20 (0,32)	-0,03 (0,03)	-0,10 (0,23)	-0,06 (0,04)	-0,08 (0,23)	-0,06 (0,04)	-0,08 (0,23)	-0,05 (0,04)	-0,12 (0,31)	-0,07 (0,04)	-0,12 (0,23)	-0,12 (0,06)	-0,10 (0,022)	-0,15* (0,06)
Octobre	0,53 (0,39)	-0,25** (0,07)	0,51 (0,43)	-0,18* (0,08)	0,38 (0,43)	-0,12 (0,07)	0,30 (0,62)	-0,06 (0,11)	0,49 (0,44)	-0,16 (0,09)	0,37 (0,41)	-0,16* (0,07)	0,39 (0,44)	-0,12 (0,09)	0,29 (0,63)	0,01 (0,14)	0,48 (0,41)	-0,32* (0,12)	0,52 (0,41)	-0,29* (0,11)
Novembre	0,10 (0,20)	-0,10 (0,07)	0,05 (0,19)	-0,06 (0,06)	0,03 (0,18)	-0,09 (0,06)	0,24 (0,25)	-0,11* (0,05)	-0,03 (0,18)	-0,10 (0,05)	0,09 (0,19)	-0,10 (0,06)	0,07 (0,18)	-0,11* (0,05)	0,14 (0,26)	-0,07 (0,05)	0,14 (0,18)	-0,25* (0,10)	0,15 (0,17)	-0,27** (0,09)
Janvier	-0,22 (0,30)	-0,16 (0,14)	-0,14 (0,31)	-0,07 (0,10)	-0,17 (0,30)	-0,03 (0,12)	-0,31 (0,41)	-0,05 (0,08)	-0,13 (0,30)	0,13 (0,10)	-0,14 (0,30)	-0,10 (0,10)	-0,23 (0,30)	-0,11 (0,11)	-0,41 (0,41)	0,06 (0,11)	-0,15 (0,30)	-0,17 (0,21)	-0,17 (0,30)	-0,15 (0,22)

Note : Un point correspond à une significativité à un niveau 10%, une étoile correspond à une significativité à 5%, deux étoiles à 1% et trois étoiles à 0,1%.

Chapitre 8

8. RECHERCHE D'ARBITRAGE

8.1. PREVISION DES ESTIMATIONS PRECOCES DE RENDEMENT

8.1.1. METHODOLOGIE

Dans cette dernière partie nous cherchons l'existence d'une possibilité d'arbitrage en anticipant le contenu du rapport NASS à l'aide de l'analyse de la série temporelle NDVI. Nous n'avons trouvé aucun papier traitant de ce sujet, cependant nous sommes intimement persuadé de l'existence de telles analyses, non publiées, par des entreprises fortement impliquées dans les marchés des ressources agricoles (fonds spéculatifs, *commodities desks* des banques anglosaxonnes, etc.). Nous étudions ce phénomène sur la période 2000-2014, *i.e.* dans la limite des données satellitaires dont nous disposons. Les rapports NASS sont publiés durant les mois d'août à novembre et en janvier. Comme nous le remarquons dans le Chapitre 5 les valeurs NDVI après le mois d'octobre n'ont plus d'influence sur l'estimation du rendement agricole. Nous décidons donc de concentrer notre analyse sur les mois d'août à octobre entre 2000 et 2014 inclus. A quelques changements près, que nous signifions ci-dessous, nous gardons les mêmes notations que dans le Chapitre 5 et le Chapitre 7. Nous rappelons que le *Tableau 5.1* d'équivalence entre les périodes NDVI, notées i et les dates conventionnelles se trouve au Chapitre 5.

Dans un premier temps nous améliorons les estimations des rendements agricoles par rapport aux régressions de la formule (5.3). Nous modifions la définition du pic de NDVI $P_{k,i,n}$, *i.e.* la valeur cumulée durant le pic de l'année n pour la zone k vu à la période i , qui sera pris maintenant entre les périodes 11 et 17 incluses. Le calcul de $P_{k,i,n}$ se fait donc pour cette partie de la façon suivante,

$$P_{k,i,n} = \begin{cases} 0, & \text{si } i < 11 \\ \sum_{t=11}^{\min\{i,17\}} V_{k,t,n} & \end{cases} \quad (8.1)$$

Nous définissons aussi la variable aléatoire du maximum de NDVI de l'année n pour la zone k vu à la date i , que nous notons $M_{k,i,n}$, de la manière suivante,

$$M_{k,i,n} = \max_{t \leq i} V_{k,t,n} \quad (8.2)$$

Pour les 8 zones géographiques k (les 8 États d'intérêts) et les périodes i de 13 à 17, nous effectuons les régressions linéaires suivantes,

$$Y_{k,n} = \beta_{k,i,0} + \beta_{k,i,1} \cdot n + \beta_{k,i,2} \cdot M_{k,i,n} + \beta_{k,i,3} \cdot (P_{k,i,n} - M_{k,i,n}) + \epsilon_{k,i,n} \quad (8.3)$$

Une fois les estimateurs $\widehat{\beta}_{k,l,0}$, $\widehat{\beta}_{k,l,1}$, $\widehat{\beta}_{k,l,2}$ et $\widehat{\beta}_{k,l,3}$ obtenus, nous appliquons alors ces formules aux différentes années d'historiques disponibles. Nous calculons ainsi les estimations précoces des rendements agricoles par NDVI, que nous notons $\widehat{Y}_{k,l,n}^{NDVI}$ et données par la formule ci-dessous,

$$\widehat{Y}_{k,l,n}^{NDVI} = \widehat{\beta}_{k,l,0} + \widehat{\beta}_{k,l,1} \cdot n + \widehat{\beta}_{k,l,2} \cdot M_{k,i,n} + \widehat{\beta}_{k,l,3} \cdot (P_{k,i,n} - M_{k,i,n}) \quad (8.4)$$

En suivant la même méthodologie indiquée dans le Chapitre 7 nous analysons les évolutions de ces estimations précoces en fonction des périodes, et quantifions ces changements. Nous redéfinissons la variable aléatoire d'information supplémentaire $X_{k,j}$ des formules (7.8) et (7.9). Pour la publication d'un rapport NASS et donc un événement sur les marchés à analyser, nous avons maintenant non plus une estimation précoce du rendement agricole, mais deux estimations dues au rythme bimensuel des valeurs NDVI que nous utilisons. Nous décidons donc d'étudier trois types d'estimations précoces d'information supplémentaire $\widehat{X}_{k,l,n}^s$ pour $s \in \{1,2,3\}$, que nous calculons de la façon suivante pour les périodes 14, 15 (périodes avant la publication du rapport de septembre) et 16, 17 (périodes précédant la publication du rapport d'octobre) de l'année n ,

$$\widehat{X}_{k,l,n}^s = \left(\frac{\widehat{Y}_{k,l,n}^{NDVI}}{\widehat{Y}_{k,l-s,n}^{NDVI}} - 1 \right) \times 100 \quad (8.5)$$

De la même manière qu'à la formule (7.9) pour les périodes 12, 13, *i.e.* celles qui précèdent la publication du rapport d'août, nous préférons définir les variables d'estimations d'informations supplémentaires par la formule suivante,

$$\widehat{X}_{k,13,n} = \left(\frac{\widehat{Y}_{k,13,n}^{NDVI}}{\widehat{Y}_{k,n}} - 1 \right) \times 100 \quad (8.6)$$

où $\widehat{Y}_{k,n}$ est la moyenne olympique des estimations finales des rendements agricoles sur les 5 dernières années donnée par la formule (7.10). Suivant la méthodologie du Chapitre 7 nous ajoutons la neuvième zone géographique qu'est la *Corn Belt* entière. Les $\widehat{X}_{9,l,n}$ pour cette dernière sont définis comme une moyenne des $\widehat{X}_{k,l,n}$ pondérée par la surface de maïs ou de soja plantée dans chaque zone géographique k .

Enfin, nous relient cette estimation d'information supplémentaire avec les changements des prix futures CBOT sur le maïs en suivant la même méthodologie décrite par la formule (7.11). Pour cela nous effectuons les régressions suivantes,

$$r_{aug,n} = \beta_{k,aug,0} + \beta_{k,aug,1} \cdot \widehat{X}_{k,l,n} + \epsilon_{k,aug,n} \quad \text{pour } i \in \{12; 13\} \quad (8.7)$$

$$r_{sep,n} = \beta_{k,oct,0} + \beta_{k,sep,1} \cdot \widehat{X}_{k,l,n}^s + \epsilon_{k,sep,n} \quad \text{pour } (i, s) \in \{14; 15\} \times \{1,2,3\} \quad (8.8)$$

$$r_{oct,n} = \beta_{k,oct,0} + \beta_{k,oct,1} \cdot \overline{X_{k,l,n}^s} + \epsilon_{k,oct,n} \quad \text{pour } (i,s) \in \{16; 17\} \times \{1,2,3\} \quad (8.9)$$

où $r_{aug,n}$ (respectivement $r_{sep,n}$, $r_{oct,n}$) représente le rendement fermeture-ouverture de la session qui suit directement la publication des rapports WASDE et NASS d'août (respectivement de septembre et d'octobre) de l'année n , calculé de la même manière que dans la formule (7.1).

Comme nous l'avons déjà fait remarquer dans les parties précédentes, lors du mois d'octobre 2013 aucun rapport NASS ou WASDE n'a été publié, nous excluons donc l'année 2013 pour l'analyse de la régression (8.9).

8.1.2. STATISTIQUES DESCRIPTIVES

Avant toute analyse, nous tenons à rappeler au lecteur que les séries temporelles NDVI que nous étudions représentent le NDVI calculé sur l'ensemble des cultures des zones géographiques, en aucun cas il ne s'agit des séries temporelles associées à la seule culture du maïs. La *Figure 8.1* et la *Figure 8.2* représentent les courbes moyennes NDVI extraites correspondant respectivement à l'État de l'Iowa et à celui du Kansas, sur la période 2000-2014. Comme nous le constatons l'indice NDVI de l'Iowa et celui du Kansas ne suivent *a priori* pas le même type de série temporelle. Ceci est dû au fait que les cultures de maïs et de blé n'ont pas le même profil NDVI (le blé d'hiver est souvent mis en culture avec du soja, ce qui explique les deux pics de NDVI), or le Kansas est un grand producteur de blé au contraire de l'Iowa dont la culture principale est le maïs. Le *Tableau 8.1* montre l'allocation en termes de surface plantée des trois principales cultures agricoles pour les 8 États étudiés. Nous observons qu'en effet au Kansas la culture dont la surface plantée est la plus importante est celle du blé avec 57%, le maïs ne représentant seulement 23%.

Figure 8.1 : NDVI moyen en Iowa, 2000-2014

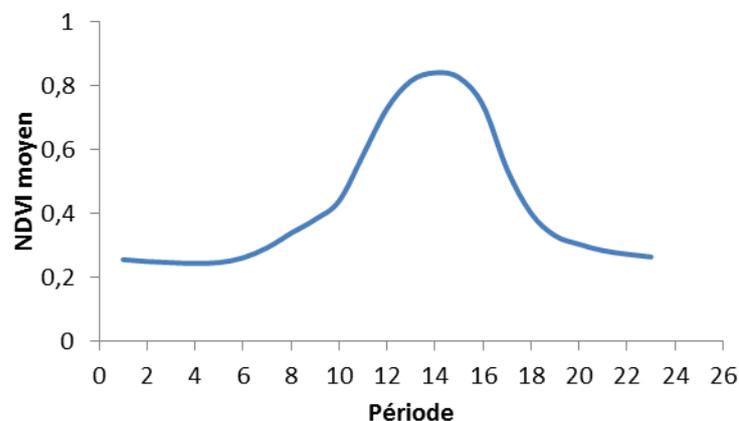


Figure 8.2 : NDVI moyen au Kansas, 2000-2014

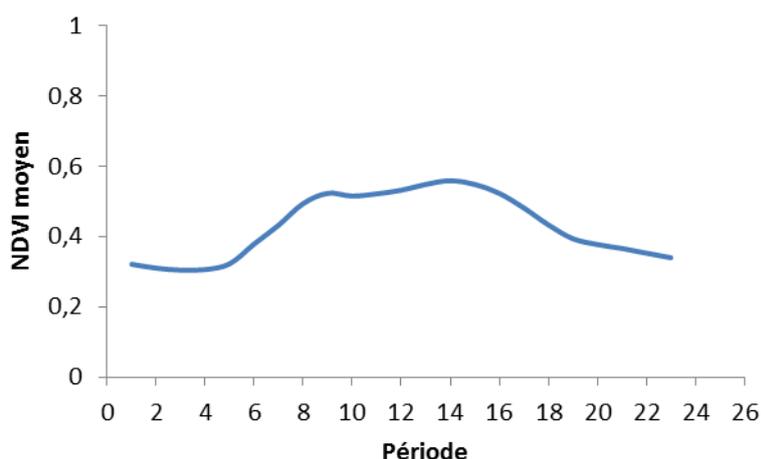


Tableau 8.1 : Moyennes des allocations des terres agricoles en fonction de la culture, 2000-2014

Culture	Illinois	Indiana	Iowa	Kansas	Minnesota	Nebraska	Ohio	South Dakota	
Acres (1)	Maïs	11,92	5,82	13,11	3,85	7,70	8,90	3,49	4,88
	Soja	9,66	5,41	9,92	3,30	7,14	4,82	4,54	4,25
	Blé	0,15	0,00	0,00	9,64	1,70	1,71	0,22	3,04
Pourcentage (2)	Maïs	55%	52%	57%	23%	47%	58%	42%	40%
	Soja	44%	48%	43%	20%	43%	31%	55%	35%
	Blé	1%	0%	0%	57%	10%	11%	3%	25%

Notes : (1) surface exprimée en million d'acres (2) le pourcentage est calculé sous l'hypothèse que les seules cultures sont le maïs, le soja et le blé.

Nous avons insisté sur le cas spécifique du Kansas car la courbe NDVI d'une culture de blé est visiblement très différente de celle d'une culture de maïs, mais le même problème d'interférence peut aussi exister entre la culture du maïs et du soja. Pour une meilleure compréhension des résultats que nous obtenons il est donc nécessaire de garder en mémoire que les indices NDVI calculés sont ceux de l'ensemble des cultures.

8.1.3. RESULTATS

Les résultats des équations données par les formules (8.7), (8.8) et (8.9) sur l'estimation précoce des changements de prix des futures CBOT par NDVI pour les 8 États et la zone *Corn Belt* entière durant la période 2000-2014 sont regroupés dans les *Tableau 8.2* et *Tableau 8.3*. Nous effectuons une analyse selon le mois de publication du rapport NASS. Nous observons que lorsque nous trouvons une corrélation significative (au minimum à un niveau 10%) entre l'estimation l'information supplémentaire par NDVI et le rendement fermeture-ouverture des futures CBOT, l'estimateur $\widehat{\beta}_1$ est négatif. Ainsi si, grâce à une estimation par NDVI, nous

prévoyons une hausse du rendement agricole, *i.e.* une potentielle hausse de la production et donc de l'offre, le prix des futures CBOT baissent. Ce résultat est en concordance avec la théorie de l'offre et la demande.

Pour la publication d'août seule l'estimation précoce d'information supplémentaire de l'Ohio par NDVI nous apparaît comme une variable significative (à un niveau de 5%) pour l'estimation précoce des changements de prix des futures CBOT sur le maïs. Ce résultat est en conformité avec ceux trouvés dans le Chapitre 7 (cf. *Tableau 7.6*) où l'estimation précoce du rendement agricole de l'Ohio publié par le NASS est considérée comme une information supplémentaire significative (à un niveau de 5%) par les acteurs du marché. Nous aurions pu attendre aussi à une relation significative pour les États du Nebraska ou du Minnesota, mais il semble que les estimations précoces des rendements agricoles par NDVI que nous effectuons pour ces deux États ne soient pas suffisamment proches de l'estimation précoce publiée par le NASS. A propos du soja nous trouvons de même une corrélation significative en août à un niveau de 5% pour les États de l'Illinois, l'Indiana, l'Ohio, le South Dakota et la *Corn Belt*. Ces zones géographiques correspondent bien aux zones où l'estimation précoce du NASS a un effet significatif sur les prix.

Concernant la publication de septembre, nous ne parvenons pas à estimer significativement les changements dans les prix des futures grâce à la série des NDVI. Seule l'Ohio (respectivement le Minnesota) montre une estimation significative à un niveau 10% si l'on compare les estimations des rendements agricoles par NDVI entre la période 15 et 14 (respectivement 12) pour le maïs (respectivement le soja). Ce résultat nous semble logique, en effet nous montrons dans les parties 7.2.3. (cf. *Tableau 7.4* et *Tableau 7.5*) et 7.3.2. (cf. *Tableau 7.6* et *Tableau 7.7*) que la publication de l'estimation des rendements agricoles par le NASS et du rapport en général ne semblent pas avoir d'impact significatif sur les prix de marché.

En octobre nos estimations des changements des prix futures CBOT par indice NDVI nous apparaissent comme significatives. Ainsi en comparant nos estimations précoces de rendements agricoles du maïs par NDVI entre les périodes 17 et 15, nous trouvons une corrélation significative à un niveau de 5% avec les rendements fermeture-ouverture des futures maïs pour l'Illinois, l'Indiana, le Kansas ainsi que la *Corn Belt*. Cette corrélation s'améliore en comparant la période 17 à la période 16, nous trouvons alors un niveau de significativité à 5% pour l'Iowa et le Minnesota, mais surtout une corrélation significative à un niveau de 1% pour la *Corn Belt*. De la même façon nous trouvons aussi des corrélations significatives à un niveau de 5% pour le soja des États de l'Indiana et l'Ohio en comparant les périodes 16 et 15, ainsi que pour la *Corn Belt*, l'Iowa et le Minnesota entre les périodes 17 et 15.

Tableau 8.2 : Estimations précoces des changements des prix futures CBOT sur le maïs par NDVI, 2000-2014

Période NDVI (1)	Publication NASS (2)	ILLINOIS		INDIANA		IOWA		KANSAS		MINNESOTA		NEBRASKA		OHIO		SOUTH DAKOTA		CORN BELT	
		$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$
13 - Moyenne (3)	Août	0,74 (0,66)	0,02 (0,06)	0,89 (0,65)	-0,06 (0,06)	0,61 (0,67)	0,09 (0,10)	0,68 (0,61)	-0,09 (0,06)	0,54 (0,63)	0,27 (0,19)	0,94 (0,72)	-0,07 (0,13)	0,75 (0,54)	-0,15* (0,06)	0,91 (0,34)	-0,10 (0,08)	0,82 (0,68)	-0,03 (0,11)
15 - 13	Septembre	-0,5 (0,29)	0,07 (0,06)	-0,05 (0,30)	0,01 (0,05)	-0,05 (0,30)	-0,01 (0,06)	-0,05 (0,30)	0,02 (0,06)	-0,05 (0,29)	0,11 (0,09)	-0,05 (0,29)	0,14 (0,12)	-0,06 (0,29)	-0,04 (0,04)	-0,05 (0,28)	0,010 (0,07)	-0,05 (0,30)	0,08 (0,10)
15 - 14	Septembre	-0,06 (0,29)	-0,16 (0,16)	-0,05 (0,28)	-0,12 (0,08)	-0,05 (0,30)	-0,14 (0,21)	-0,05 (0,30)	0,11 (0,13)	-0,05 (0,30)	-0,20 (0,29)	-0,05 (0,30)	0,11 (0,23)	-0,04 (0,27)	-0,17 (0,09)	-0,05 (0,30)	-0,05 (0,12)	-0,05 (0,29)	-0,24 (0,23)
17 - 15	Octobre	0,06 (0,64)	-0,56* (0,24)	0,00 (0,66)	-0,37* (0,17)	0,11 (0,75)	-0,38 (0,33)	0,04 (0,65)	-0,47* (0,21)	-0,03 (0,73)	-0,24 (0,18)	0,01 (0,70)	-0,37 (0,21)	0,03 (0,74)	-0,26 (0,21)	0,02 (0,78)	0,00 (0,14)	0,06 (0,64)	-0,69* (0,29)
17-16	Octobre	0,01 (0,67)	-0,68 (0,33)	-0,02 (0,71)	-0,46 (0,30)	-0,01 (0,61)	-1,26* (0,45)	0,03 (0,69)	-0,88 (0,46)	-0,18 (0,64)	-0,65* (0,26)	-0,06 (0,68)	-0,86 (0,43)	0,03 (0,75)	-0,28 (0,27)	0,03 (0,77)	-0,16 (0,24)	-0,07 (0,55)	-1,47** (0,42)

Notes : (1) période NDVI correspondant aux i et $i - s$ de la formule (8.5) (2) mois de la publication du rapport NASS pour le calcul du rendement fermeture-ouverture des futures CBOT (3) Moyenne olympique des rendements agricoles sur 5 ans, conformément à la formule (8.6) (4) Un point correspond à une significativité à un niveau 10%, une étoile correspond à une significativité à 5%, deux étoiles à 1% et trois étoiles à 0,1%.

Tableau 8.3 : Estimations précoces des changements des prix futures CBOT sur le soja par NDVI, 2000-2014

Période NDVI (1)	Publication NASS (2)	ILLINOIS		INDIANA		IOWA		KANSAS		MINNESOTA		NEBRASKA		OHIO		SOUTH DAKOTA		CORN BELT	
		$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$
12 - Moyenne (3)	Août	2,03** (0,59)	-0,27* (0,10)	1,91** (0,57)	-0,16* (0,06)	1,21 . (0,63)	-0,05 (0,13)	1,33* (0,60)	-0,05 (0,04)	1,26* (0,58)	-0,12 (0,08)	1,69* (0,61)	-0,15 . (0,08)	1,72** (0,43)	-0,25** (0,06)	1,58** (0,49)	-0,18** (0,06)	1,83** (0,48)	-0,34** (0,10)
13 - Moyenne (3)	Août	1,83** (0,60)	-0,21* (0,09)	1,88** (0,57)	-0,15* (0,06)	1,21 . (0,63)	-0,04 (0,11)	1,34* (0,58)	-0,06 (0,03)	1,25 . (0,59)	-0,12 (0,09)	1,72* (0,60)	-0,16 . (0,08)	1,61** (0,47)	-0,19** (0,05)	1,57** (0,51)	-0,18* (0,06)	1,70** (0,52)	-0,27* (0,09)
15 - 12	Septembre	0,27 (0,46)	0,03 (0,13)	0,27 (0,46)	0,05 (0,14)	0,26 (0,45)	-0,08 (0,11)	0,26 (0,45)	-0,03 (0,05)	0,27 (0,40)	-0,31 . (0,15)	0,27 (0,46)	-0,01 (0,09)	0,27 (0,46)	-0,01 (0,08)	(0,27 (0,45)	-0,06 (0,09)	0,26 (0,45)	-0,08 (0,15)
15 - 14	Septembre	0,25 (0,42)	-0,67 (0,42)	0,26 (0,46)	-0,11 (0,35)	0,26 (0,45)	-0,35 (0,39)	0,24 (0,44)	-0,11 (0,11)	0,25 (0,43)	-0,77 (0,57)	0,26 (0,46)	-0,07 (0,28)	0,26 (0,43)	-0,32 (0,23)	0,27 (0,45)	-0,09 (0,14)	0,24 (0,42)	-0,71 (0,45)
16 - 15	Octobre	0,36 (0,75)	-0,71 . (0,38)	0,28 (0,64)	-0,73* (0,24)	0,36 (0,74)	-0,58 . (0,30)	0,17 (0,79)	-0,29 (0,21)	0,30 (0,80)	-0,52 (0,41)	0,27 (0,83)	-0,21 (0,34)	0,23 (0,68)	-3,09* (1,20)	0,24 (0,84)	0,08 (0,21)	0,32 (0,72)	-0,95 . (0,45)
17 - 15	Octobre	0,29 (0,79)	-0,22 (0,17)	0,27 (0,80)	-0,21 (0,17)	0,27 (0,65)	-0,46* (0,16)	0,22 (0,81)	-0,13 (0,13)	0,18 (0,71)	-0,44* (0,20)	0,25 (0,81)	-0,22 (0,21)	0,27 (0,75)	-0,92 . (0,50)	0,26 (0,85)	-0,00 (0,16)	0,26 (0,71)	-0,56* (0,25)

Notes : (1) période NDVI correspondant aux i et $i - s$ de la formule (8.5) (2) mois de la publication du rapport NASS pour le calcul du rendement fermeture-ouverture des futures CBOT (3) Moyenne olympique des rendements agricoles sur 5 ans, conformément à la formule (8.6) (4) Un point correspond à une significativité à un niveau 10%, une étoile correspond à une significativité à 5%, deux étoiles à 1% et trois étoiles à 0,1%.

8.2. LIMITES ET PISTES D'AMÉLIORATIONS

Bien que nous obtenions des corrélations significatives dans ce chapitre entre nos estimations d'informations supplémentaires et les changements de prix futures CBOT, ces résultats seuls ne suffisent pas à mettre en place des stratégies financières qui permettraient aux gérants de portefeuilles de futures CBOT sur le maïs et le soja d'arbitrer statistiquement les marchés en estimant en avance le contenu des rapports gouvernementaux. Cependant ils montrent l'existence de tels arbitrages. Notre modèle peut être amélioré afin d'obtenir de meilleures corrélations et ainsi être dans la capacité de construire des stratégies financières. Nous proposons ci-après trois pistes d'amélioration, qui sont d'après nous les plus prometteuses. Nous n'avons pas eu l'opportunité de les mettre en place par manque de temps de développement et de ressources informatiques.

La première amélioration à apporter à notre modèle concerne les données NDVI. Comme nous le signifiions, nous utilisons une base de données regroupant des images de NDVI moyen disponible tous les 16 jours. Une base NDVI améliorée serait une moyenne mobile sur 16 jours disponible quotidiennement. Cette base n'existe pas directement, mais la NASA met à disposition quotidiennement les images de réflectivités pour différentes longueurs d'ondes de la plupart du globe²⁶. Ces données permettent de recréer la série temporelle journalière des NDVI pixel par pixel, puis en appliquant une simple moyenne mobile nous retrouvons la base de données souhaitée. Cette amélioration dans la précision temporelle permettrait d'optimiser le choix des périodes de NDVI, en effet entre chaque rapport nous n'aurions plus 2 mais environ 30 choix de valeur NDVI et donc d'estimations de rendements agricoles.

La seconde amélioration est à propos du calque appliqué aux données NDVI. Le calque que nous utilisons permet de différencier les terres arables des autres zones géographiques. Un calque amélioré permettrait de différencier les différents types de cultures. Un tel calque peut être obtenu par l'USDA²⁷ mais seulement en fin de saison, donc trop tard pour l'utilisation pratique des prévisions des changements de prix, ou par l'utilisation d'un algorithme d'analyse pixel par pixel pour la détermination du type de culture en suivant une méthodologie similaire à Kastens *et al.* (2005).

Ces deux pistes d'améliorations demandent un temps développement et de calcul important. En effet pour ce faire une analyse détaillée pixel par pixel sur l'historique est nécessaire pour l'ensemble de la zone géographique. Pour donner un ordre de grandeur, la superficie totale des 8 États (Illinois, Indiana, Iowa, Kansas, Minnesota, Nebraska, Ohio et Dakota du Sud) est de 1,345 millions de km². Chaque pixel est d'une superficie de 0,0625km², ce qui nous donne environ 21,5

²⁶ http://www.gscloud.cn/userfiles/file/MOD09_UserGuide.pdf

²⁷ <http://nassgeodata.gmu.edu/CropScape>

millions de pixels par image sur un historique d'environ 5350 jours. Les ressources temporelles et informatiques dont nous disposons dans le cadre de ce mémoire ne suffisent pas au développement d'une telle analyse.

Enfin la troisième piste d'amélioration que nous proposons pour notre modèle est la prise en compte d'indices climatiques en plus de l'indice agronomique qu'est le NDVI à la manière de Prasad *et al.* (2006) afin d'améliorer la précision d'estimation des rendements agricoles. Les données climatiques, comme la température, les précipitations ou l'ensoleillement, peuvent être obtenues soit par le biais de stations météorologiques ou par données satellites fournies gratuitement par la NASA²⁸.

8.3. EXTENSIONS

Dans ce mémoire nous appliquons notre méthodologie d'estimation précoce d'information supplémentaire des rapports gouvernementaux seulement au maïs et au soja pour les rapports WASDE. Quitte à faire quelques changements, cette approche peut être étendue. Dans cette partie nous donnons trois pistes d'extensions possibles de nos travaux que nous n'avons pas eu le temps d'appliquer.

La première extension, et sûrement la plus facile à implémenter, est de se focaliser sur d'autres cultures, quitte à changer les zones géographiques d'analyse. En premier lieu nous pensons au blé, qui est une culture importante des États-Unis. Mais le colza ou l'avoine sont aussi des candidats potentiels pour une telle analyse.

La seconde piste d'extension est de modifier les rapports gouvernementaux sur lesquels nous nous fondons. Lehecka (2103) montre en effet que les marchés des futures CBOT sur le maïs et le soja réagissent à l'annonce de l'information contenue dans les rapports hebdomadaires des progressions des cultures (*Crop Progress report*) du NASS. En estimant en avance cette information supplémentaire avec une méthodologie similaire à celle que nous appliquons dans ce mémoire, nous pourrions estimer la réaction des marchés en avance.

Enfin la dernière extension que nous proposons est d'analyser si sur d'autres marchés que le CBOT cette approche pourrait être mise en place. Notamment pour le blé, maïs et colza à la bourse NYSE Euronext de Paris. Pour ce faire il faut déterminer les rapports gouvernementaux ayant un impact significatif sur les prix des futures, puis analyser les zones géographiques de références, enfin construire un modèle semblable au nôtre pour estimer en avance l'information supplémentaire fournie par le rapport gouvernemental.

²⁸ <https://search.earthdata.nasa.gov/>

CONCLUSION GENERALE

Au terme de ce mémoire nous avons montré différentes améliorations que peuvent apporter les images satellites à l'assurance agricole en nous focalisons sur l'indice de réflectivité NDVI. Cet indice permet de quantifier l'activité chlorophyllienne et ainsi la santé de la culture étudiée, il est donc naturellement corrélé avec le risque de rendement agricole qui constitue le premier risque des exploitants. Mais en analysant cet indice sur une zone géographique étendue, nous avons montré que les images satellites permettent aussi une meilleure approche du risque de prix des agriculteurs.

Dans un premier temps nous avons proposé un exemple de produit d'assurance agricole fondé sur l'indice NDVI et l'avons comparé à un produit d'assurance surface-rendement plus classique utilisant les statistiques gouvernementales. Ce changement d'indice permet de proposer des assurances à des exploitants localisés dans des pays où les assureurs refusaient régulièrement jusqu'à lors toute souscription due à la non-fiabilité des données. De plus ces agriculteurs pour lesquels les compagnies ne souhaitaient pas accepter le risque sont souvent parmi les plus touchés par le changement climatique et les événements associés (sécheresses, températures extrêmes, etc.). Concrètement le développement de produits fondés sur des images satellitaires permet d'ouvrir de nouveaux marchés pour l'assurance agricole. Ce développement a déjà commencé et des programmes comme l'African Risk Capacity sont désormais bien implantés et rassemblent plusieurs dizaines de pays et d'assureurs. L'utilisation de tels indices peut dans d'autres cas être aussi un moyen de diminuer le risque de base en centrant les images satellitaires sur un champ en particulier, quitte à rajouter une contrainte sur un indice météorologique ou une franchise afin d'éviter tout aléa moral.

Dans un second temps nous avons analysé les marchés financiers qui régissent les prix des matières premières aux États-Unis. A travers les produits d'assurance revenu, les compagnies choisissent de supporter une partie du risque de prix des agriculteurs, et gèrent donc des portefeuilles relativement importants de produits financiers indexés sur les matières premières agricoles. Nous avons remarqué que les rapports gouvernementaux ont une forte influence sur des sessions précises connues d'avance. L'information contenue dans ces rapports est considérée comme précieuse par les différents acteurs. Nous montrons que cette information peut être estimée en avance à l'aide des images satellites NDVI. Les changements de prix sur les marchés pour certaines sessions peuvent alors aussi être estimés en avance, permettant ainsi la création d'une stratégie d'arbitrage statistique, fortement utile pour les assureurs. Le modèle que nous mettons en place ne montre que l'existence d'un tel phénomène. Avant de réfléchir à la création d'une stratégie financière il conviendrait de l'améliorer notamment en augmentant la précision des données mais qui passe par une augmentation de la puissance de calcul disponible. Les résultats obtenus et le modèle créé restent tout de même une première base utile pour de futurs travaux.

GLOSSAIRE

CBOT : *Chicago Board Of Trade*. Bourse de commerce de Chicago.

MODIS : *Moderate-Resolution Imaging Spectroradiometer*. Instruments d'observation scientifique couplés à un système embarqué satellitaire.

NASA : *National Aeronautics and Space Administration*. Agence spatiale gouvernementale des États-Unis.

NASS : *National Agricultural Statistics Service*. Département de statistiques agricoles des États-Unis.

NDVI : *Normalized Difference Vegetation Index*. Indice de végétation utilisant la réflectivité dans les régions du rouge et du proche infrarouge fournie par la NASA.

USDA : *United States Department of Agriculture*. Département de l'Agriculture des États-Unis, équivalent au Ministère de l'Agriculture.

WASDE : *World Agricultural Supply and Demand Estimates*. Rapport mensuel de l'USDA rassemblant les données sur la demande et l'offre mondiale en denrées agricoles.

BIBLIOGRAPHIE

BIBLIOGRAPHIE GÉNÉRALE

Anderson, J. R., **Simulation Methodology and Application in Agricultural Economics**, *Rev. Market. And Agr. Econ.*, No.42, 3-55, 1974.

Atwood, J., Shaik, S., Watts, M., **Can Normality of Yields be assumed for Crop Insurance?**, *Canadian Journal of Agricultural Economics*, 50:171-884, 2002.

Baldi, L., Peri, M., Vandone, D., **Spot and Futures Prices of Agricultural Commodities: Fundamentals and Speculation**, *Dipartimento di Scienze Economiche Aziendali e Statistiche*, Milan, Working Paper n.2011-03, 2011.

Cleveland, W. S., **Robust Locally Weighted Regression and Smoothing Scatterplots**, *Journal of the American Statistical Association*, Vol. 74, (368):829-836, 1979.

Cleveland, W. S., Devlin, S. J., **Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting**, *Journal of the American Statistical Association*, Vol. 83, No. 403, pp. 596-610, 1988.

Colwell, R. N., **Determining the Prevalence of Certain Cereal Crop Diseases by Means of Aerial Photography**, *Hilgardia*, 26:223-286, 1956.

Day, R. H., **Probability Distributions of Field Crop Yields**, *Journal of Farm Economics*, Vol. 47, No. 3, pp 713-741, 1965.

Dolley, J., **Characteristics and Procedure of Common Stock Split-Ups**, *Harvard Business Review*, 316-326, 1933.

Fajarnes, P., **An Overview of Major Sources of Data and Analyses Relating to Physical Fundamentals in International Commodity Markets**, *United Nations Conference on Trade and Development*, Discussion Paper No. 202, 2011.

Falk, B., Orazem, P., **A Theory of Future's Market Responses to Government Crop Forecasts**, *Economic Staff Paper Series*, Paper 34, 1984.

Ferencz, C., Bognar, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timar, G., Molnar, G., Pasztor, S., Steinbach, P., Szekely, B., Ferencz, O. E., Ferencz-Arkos, I., **Crop Yield Estimation by Satellite Remote Sensing**, *International Journal of Remote Sensing*, Vol. 25, No. 20, pp 4413-4149, 2004.

Fortenbery, T. R., Sumner, D. A., **The Effects of USDA Reports in Futures and Options Markets**, *NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, Chicago, IL, 1990.

Garcia, P., Irwin, S. H., Leuthold, R. M., Yang, L., **The Value of Public Information in Commodity Futures Markets**, *Journal of Economic Behavior & Organization*, Vol. 32, 559-570, 1997.

- Gausman, H. W., Cardenas, R., **Effect of Soil Salinity on External Morphology of Cotton Leaves**, *Agronomy Journal*, 60:566-576, 1968.
- Gausman, H. W., Allen, W. A., Myers, V. I., Cardenas, R., **Reflectance and Internal Structure of Cotton Leaves, *Gossypium Hirsutum***, *Agronomy Journal*, 61:374-376, 1969.
- Gausman, H. W., Allen, W. A., Cardenas, R., Richardson, A. J., **Relation of Light Reflectance to Histological and Physical Evaluations of Cotton Leaf Maturity**, *Applied Optics*, 9:545-552, 1970.
- Govaerts, B., Verhulst N., **The Normalized Difference Vegetation Index (NDVI) GreenSeeker Handheld Sensor: Toward the Integrated Evaluation of Crop Management**, *CIMMYT International Maize and Wheat Improvement Center*, Working paper, 2010.
- Hadjimitsis, D. G., Papadavid, G., **Remote Sensing of Environment – Integrated Approaches** ISBN 978-953-51-1152-8, InTech, 2013.
- Irwin, S. H., Good, D. L., Gomez, J. K., **The Value of USDA Outlook Information: An Investigation Using Event Study Analysis**, *NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, St. Louis, Missouri, April 23-24, 2001.
- Isengildina-Massa, O., Irwin, S. H., Good, D. L., Gomez, J. K., **The Impact of Situation and Outlook Information in Corn and Soybean Futures Markets: Evidence from WASDE Reports**, *Journal of Agricultural and Applied Economics*, 40, 1(April 2008):89-103, 2008.
- Just, R. E., Weninger, Q., **Are Crop Yields Normally Distributed?**, *American Journal of Agricultural Economics*, Vol. 81, No. 2, pp287-304, 1999.
- Kapiamba, L. F., **Modeling Heteroskedasticity of Crop Yield Distributions: Implications for Normality**, *American Agricultural Economics Association Annual Meeting*, Selected paper, 2005.
- Kastens, J. H., Kastens, T. L., Kastens, D. L. A., Price, K. P., Martinko, E. A., Lee, R.Y., **Image Masking for Crop Yield Forecasting Using AVHRR NDVI Time Series Imagery**, *Remote Sensing of Environment*, 99:341-356, 2005.
- Ker, A., Tolhurst, T., **On the Treatment of Heteroscedasticity in Area-Yield/Group-Risk Crop Insurance**, *Structure and Performance of Agriculture and Agri-products industry Network*, Working Paper #2013-05, 2013.
- Kruskal, W. H., Wallis, W. A., **Use of Ranks in One-Criterion Variance Analysis**, *Journal of the American Statistical Association*, Vol. 47, No. 260, pp. 583-621, 1952.
- Kumar, R., Silva, L., **Light Ray Tracing through a Leaf Cross Section**, *LARS Technical Reports*, Paper 16, 1972.
- Livanis, G. T., Salois, M. J., Moss, C. B., **A Nonparametric Kernel Representation of the Agricultural Production Function: Implications for Economic Measures of Technology**, *The 83rd Annual Conference of the Agricultural Economics Society*, Dublin, 2009.

Lehecka, G. V., **The Value of USDA Crop Progress and Condition Information: Reactions of Corn and Soybean Futures Markets**, *Journal of Agricultural and Resource Economics*, 39(1):88-105, 2013.

Li, A., Liang, S., Wang, A., Qin, J., **Estimating Crop Yield from Multi-temporal Satellite Data Using Multivariate Regression and Neural Network Techniques**, *Photogrammetric Engineering & Remote Sensing*, Vol. 73, No. 10, pp 1149-1157, 2007.

Ma, B. L., Dwyer, L. M., Costa, C., Cober, E. R., Morrison, M. J., **Early Prediction of Soybean Yield from Canopy Reflectance Measurements**, *Agronomy Journal*, 93:1227-1234, 2001.

Maselli, F., Rembold F., **Analysis of GAC NDVI Data for Cropland Identification and Yield Forecasting in Mediterranean African Countries**, *Photogrammetric Engineering and Remote Sensing*, 67(5), 593-602, 2001.

Milacek, T. T., Brorsen, B. W., **How much Would It Be Worthy to Know the WASDE Report In Advance?**, *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, St. Louis, MO., 2013.

Mkhabela, M. S., Bullock, P., Raj, S., Wang, S., Yang, Y., **Crop Yield Forecasting on the Canadian Prairies using MODIS NDVI Data**, *Agricultural and Forest Meteorology*, 151:385-393, 2011.

Pagan, A., Ullah, A., **Nonparametric Econometrics**, NY: Cambridge University Press, 1999.

Pervez, S., Brown, J. F., **Mapping Irrigated Lands at 250-m Scale by Merging MODIS Data and National Agricultural Statistics**, *Remote Sensing*, 2:2388-2412, 2010.

Peterson, D., Whistler, J., Egbert, S., Brown, J. C., **Mapping Irrigated Lands by Crop Type in Kansas**, *Pecora 18-Forty years of Earth Observation...Understanding a Changing World*, 2011.

Prasad, A. K., Chai, L., Singh, R. P., Kafatos, M., **Crop Yield Estimation Model for Iowa using Remote Sensing and Surface Parameters**, *International Journal of Applied Earth Observation and Geoinformation*, 8:26-33, 2006.

Ramirez, O. A., Misra, S. K., Field, J., **Are Crop Yields Normally Distributed?**, *American Agricultural Economics Association*, Selected paper, 2001.

Rouse, J. W., Haas R. H., Schell, J. A., Deering, D. W., **Monitoring Vegetation Systems in the Great Plains with ERTS**, *Third ERTS Symposium*, NASA SP-351 I, 309-317, 1973.

Schnepf, R., **Price Determination in Agricultural Commodity Markets: A Primer**, *Congressional Research Services – The Library of Congress*, Order Code RL33204, 2006.

Schnepf, R., **U.S. Farm Income Outlook for 2015**, *Congressional Research Service*, OrderCode R40152, 2015.

Shaik, S., **Normality of Indian Crop Yields: Application of Panel Analysis**, *Journal of Quantitative Economics*, Vol. 8, No. 1, 2010.

Smith, V. H., Johnson, J. B., Hewlett, J. P., **New Programs in the 2014 Farm Bill: Price Loss Coverage, Agricultural Risk Coverage and the Supplemental Coverage Agricultural**

Insurance Option for Wyoming Farms and Ranches, *Agricultural Marketing Policy Paper*, No. 44, 2014.

Spratt, S., **Food Price Volatility and Financial Speculation**, *Institute of Development Studies, Future Agricultures*, Working Paper 047, 2013.

Sumner, D. A., Mueller, R. A. E., **Are Harvest Forecasts News? USDA Announcements and Futures Market Reactions**, *American Journal of Agricultural Economics*, 71(1):1-8, 1989.

Svotwa, E., Masuka, A. J., Maasdorp, B., Murwira, A., Masocha, M., **Estimating Tobacco Crop Area and Yield in Zimbabwe Using Operational Remote Sensing and Statistical Techniques**, *International Journal of Agricultural Research and Review*, Vol. 2(5), pp 084-091, 2014.

Turvey, C. G., Mclaurin, M. K., **Applicability of the Normalized Difference Vegetation Index (NDVI) in Index-Based Crop Insurance Design**, *Weather, Climate & Society*, Vol. 4, 2012.

Vogel, F. A., Bange, G. A., **Understanding USDA Crop Forecasts**, *United States Department of Agriculture*, Miscellaneous Publication No. 1554, 1999.

Vrieling, A., Meroni, M., Shee, A., Mude, A. G., Woodard, J., Kees de Bie, C. A. J. M., Rembold, F., **Historical Extension of Operational NDVI Products for Livestock Insurance in Kenya**, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 28, pp 238-251, 2014.

Yang, S., Koo, W. W., Wilson, W. W., **Heteroskedasticity in Crop Yield Models**, *Journal of Agricultural and Resource Economics*, 17:103-9, 1992.

Zhang, H., Chen, H., Zhou, G., **The Model of Wheat Yield Forecast based on MODIS-NDVI – A Case Study of Xinxiang**, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume I-7, 2012.

BIBLIOGRAPHIE SANS AUTEUR PRÉCIS

Introduction to Risk Management, Understanding Agricultural Risk: Production, Marketing, Financial, Legal, Human Resources, *United States Department of Agriculture*, Risk Management Agency, December 1997.

Parametric Insurance: A Fitting Solution for the Weather-Sensitive, *AXA Group Communications Department*, 2015.

Usual Planting and Harvesting Dates for U.S. Field Crops, *Agricultural Handbook Number 628*, National Agricultural Statistics Service, United States Department of Agriculture, 1997.

2015 Crop Insurance Update, Securing America's Farmers, *Rain and Hail Insurance Society*, Johnston, IA, 2015.

2012 Census of Agriculture, *United States Department of Agriculture*, 2014.

MEMOIRES D'ACTUARIAT

Batisse, L., Mercuzot, Y., **Étude de l'assurance des rendements agricoles face aux risques climatiques**, *Mémoire d'actuariat*, CEA & PACIFICA, 2009.

Denizot, O., Plouhinec, C., **Pricing des Dérivés Climatiques: Quel Modèle de Température?**, *Mémoire d'actuariat*, ENSAE & Paris X Nanterre, 2008.

El Alaoui, I., Kreutzberger, E., **Assurances Agricoles : Couverture du Risque Prix en Utilisant des Produits Dérivés de Matières Premières**, *Mémoire d'actuariat*, ENSAE & PACIFICA, 2009.

Fosso, P., Seretti, L., **Assurance Revenu Agricole**, *Mémoire d'actuariat*, ENSAE & PREDICA, 2010.

Gilles, S., Finas, B., **La Gestion Indicielle du Risque Climatique: Application à l'Energie Eolienne**, *Mémoire d'actuariat*, ENSAE & AON Global Risk Consulting, 2011.

Koch, E., **Etude de Faisabilité d'une Assurance Rendement Basée sur Indice Climatique**, *Mémoire d'actuariat*, Université Paris-Dauphine & Laboratoire de Météorologie Dynamique de Jussieu, 2011.

Leprince, F., **Yield and Revenue Crop Insurances in the United States**, *Mémoire d'actuariat*, EURIA & AIM, 2001.

Nguyen, C., Ritleng, F., **Étude d'un Produit d'Assurance Paramétrique contre le Risque de Pluie Torrentielle en Jamaïque**, *Mémoire d'actuariat*, ENSAE, 2014.

Vendé, P., **Les Couvertures Indicielles en Réassurance Cat: Prise en Compte de la Dépendance Spatiale dans la Tarification**, *Mémoire d'actuariat*, ESSEC & ISUP & Benfield Paris, 2003.

Assurance Agricole et Images Satellites : Produit d'Assurance et Arbitrage Financier

Note de Synthèse

Mémoire présenté devant l'ENSAE pour l'obtention du diplôme
d'Actuaire ENSAE et l'admission à l'Institut des Actuaires

Pierrick Piette

Note de Synthèse

Mots-clefs : assurance agricole, assurance paramétrique, CBOT, États-Unis, image satellite, maïs, marché des futures, NASA, NDVI, prévision des prix, réaction des marchés, rendement agricole, revenu agricole, soja, télédétection

1. Fondamentaux de l'assurance agricole

Les agriculteurs, de par leur activité professionnelle, sont fortement exposés aux risques. Les deux risques majeurs rencontrés sont le risque de rendement et le risque de prix. Le premier consiste en une baisse des rendements agricoles de l'exploitation qui peut être causée par le manque d'engrais, les maladies, les conditions météorologiques, etc. Si l'agriculteur produit moins en quantité pour une surface équivalente ses revenus baissent. Le second risque vient de la variabilité des prix des matières premières. Si le prix de vente diminue pour une production équivalente les revenus de l'exploitant baissent. Ces deux risques impliquent une forte instabilité dans les revenus à travers les années pour les agriculteurs. De plus cette variation semble avoir tendance à augmenter à cause du changement climatique et des événements extrêmes associés à ce dernier. C'est donc tout naturellement que les exploitants agricoles cherchent des outils pour lisser leurs revenus d'une année sur l'autre.

Plusieurs de ces outils sont disponibles sur le marché : produits dérivés sur les matières premières, produits dérivés climatiques, assurance agricole, etc. Dans notre étude nous nous restreignons à l'étude de l'assurance agricole qui est l'outil le plus souvent utilisé aux États-Unis. C'est d'ailleurs sur ce pays, dont l'agriculture est la première mondiale en termes de production, que nous fondons nos exemples. En 2014 les primes d'assurance agricole y avaient dépassé les 10 milliards de dollars. Les méthodologies que nous développons peuvent néanmoins être étendues, avec plus ou moins de changements, à d'autres régions. Deux catégories de produits assurantiels existent : l'assurance rendement, ne couvrant que le risque de rendement, et l'assurance revenu, couvrant le risque de rendement et de prix. Notre étude s'attarde à explorer différents apports de l'utilisation d'images satellites de la NASA et plus spécifiquement de l'indice NDVI (*Normalized Difference Vegetation Index*) à ces deux produits d'assurance agricole.

2. Indice NDVI

L'indice NDVI est un indice de réflectivité permettant de mesurer l'activité chlorophyllienne d'une plante et donc sa santé. Il se fonde sur la spécificité du spectre de réflexion de la végétation dans le domaine du rouge et du proche infrarouge. Il est calculé de la manière suivante,

$$NDVI = \frac{R_{NIR} - R_R}{R_{NIR} + R_R} \quad (1)$$

où R_{NIR} est la réflectivité dans le proche infrarouge (*Near InfraRed*) et R_R est la réflectivité dans la région du rouge visible (*Red*). De nombreux papiers académiques et publications montrent sa forte corrélation (R^2 entre 0,4 et 0,9) avec les rendements agricoles pour de multiples types de cultures dans différents régions géographiques.

Les données que nous utilisons sont fournies gratuitement par la NASA à une résolution de 250 mètres. Nous disposons d'une image tous les seize jours sur la période du 18 février 2000 au 3 janvier 2015. Chaque valeur d'une image correspond à la moyenne des 16 derniers jours, en effet le NDVI étant un indice de réflectivité il est très sensible aux conditions météorologiques (nuages, pluie, etc.) quand il est observé depuis les satellites. Nous appliquons un calque permettant de ne prendre en compte que les cultures, cependant nous ne disposons pas de calques afin d'obtenir une série temporelle NDVI par type de culture. Les valeurs obtenues sont donc celles de l'ensemble des cultures (blé, maïs, soja, etc.) confondues, certaines problématiques d'interférence entre les séries peuvent apparaître et diminuer les corrélations.

Dans un premier temps nous analysons comment cet indice permet de se préserver de certaines contraintes de l'assurance rendement classique et ainsi favoriser le développement de l'assurance agricole dans des régions du globe où les assureurs étaient réticents à souscrire. Dans un second temps nous cherchons des opportunités d'arbitrages grâce au NDVI sur les marchés futures de la Bourse de Chicago (CBOT, *Chicago Board of Trade*) afin de permettre aux assureurs de diminuer le risque de prix qu'ils détiennent.

3. Assurance rendement

Il existe deux structures principales pour les produits d'assurance rendement : l'une pour laquelle les indemnités sont calculées d'après un expert qui s'est rendu sur place, l'autre où les montants compensatoires sont fondés sur un indice (assurance paramétrique). Nous nous intéressons à la seconde structure car c'est celle qui s'est développée aux Etats-Unis et qui est la plus prometteuse par ses coûts de gestion très faibles. Le produit le plus classique est l'assurance surface-rendement où l'indice est un rendement dit de référence. Un point d'attache et un point de sortie sont définis. Les indemnités sont alors calculées de la manière suivante,

$$I = V \times \mathbb{1}_{R \leq A} \min\{A - R; A - S\} \quad (2)$$

où I est l'indemnité reçue, V est la valeur d'un point d'indice, R est la valeur du rendement de référence, A le point d'attache et S le point de sortie. La valeur d'un point d'indice est tout simplement le produit de la surface assurée par le prix d'une unité de la culture assurée.

Le rendement de référence doit cependant vérifier quelques propriétés afin de pouvoir être utilisé dans la structure du produit. L'assureur doit être capable de se procurer un historique de

cette variable assez long (10-15 ans au minimum). Les données doivent être considérées comme suffisamment fiables. L'assuré doit avoir une influence limitée sur cet indice. Aux Etats-Unis comme en France le rendement de référence le plus souvent utilisé est l'estimation par le ministère de l'agriculture du rendement agricole de la zone administrative dans laquelle se trouve l'exploitation assurée. Cependant ces données ne sont pas toujours considérées comme fiables dans certains pays instables politiquement comme en Afrique ou en Ukraine par exemple. Dans ces cas-là, et sans autres indices disponibles, l'assureur va généralement préférer ne pas souscrire. Pour montrer la sensibilité de la tarification aux données, nous prenons deux exemples de produits (maïs et soja en Iowa et Illinois respectivement) pour lesquels nous estimons la prime pure suivant trois méthodologies de régressions. Nous observons alors des variations pouvant aller du simple au double selon la régression appliquée.

Cette problématique de fiabilité des données peut être résolue en utilisant un indice fondé sur le NDVI. En effet les données de la NASA sont considérées comme suffisamment fiables par les assureurs et l'obtention des valeurs de l'indice a un coût très limité (connexion à internet, téléchargement des images, calcul des valeurs via R ou un autre logiciel). Nous proposons donc une structure d'assurance rendement fondé sur le NDVI couvrant les mêmes risques que l'exemple étudié précédemment. L'indice utilisé correspond au pic de NDVI, c'est-à-dire durant la phase de croissance de la plante. Les valeurs prises en compte sont alors comprises entre le 26 juin et le 13 septembre. Nous montrons que cet indice est bien corrélé avec les rendements agricoles (R^2 de 0,85 pour le maïs et de 0,63 pour le soja). Nous proposons aussi une méthodologie de tarification de la prime pure pour les produits ainsi structurés.

Le choix d'un produit d'assurance NDVI peut aussi être motivé par la volonté de diminuer le risque de base porté par l'assuré. Avec les nouvelles résolutions des images (250 mètres) l'indice NDVI peut être centré sur une exploitation en particulier afin d'obtenir une meilleure corrélation qu'avec le rendement de référence gouvernemental. Dans ces cas-là l'utilisation du NDVI a un apport significatif même dans les pays avec des données publiques fiables, par exemple les Etats-Unis.

L'utilisation de l'indice NDVI peut cependant poser une nouvelle problématique : si les valeurs sont seulement celles de l'exploitation assurée il y a un risque d'aléa moral. En effet l'agriculteur peut influencer sur la santé de ses cultures et donc de l'indice NDVI. Il convient alors soit d'appliquer une franchise sur le prix d'indemnisation par unité de production, soit de coupler l'indice avec un indice météorologique sur lequel l'exploitant n'aura aucune influence possible.

4. Assurance revenu

L'assurance revenu est une catégorie de produits d'assurance agricole qui couvrent à la fois le risque de rendement et le risque de prix. Ces produits sont largement répandus aux Etats-Unis où

ils représentent depuis plusieurs années environ 80% des souscriptions (en 2014 les primes pour les assurances revenus s'élevaient à 8,18 milliards de dollars). Il existe de multiples structures possibles que nous ne détaillons pas. La différence majeure de l'assurance revenu face à l'assurance rendement est que le prix d'indemnisation par unité de production n'est pas connu à l'avance et est calculé à partir du marché des futures au moment de la récolte (novembre pour le soja et décembre pour le maïs). Ainsi les compagnies d'assurance sont contraintes de gérer un important portefeuille de produits dérivés sur les matières premières agricoles.

4.1. Publications gouvernementales et réactions des marchés

Dans un premier temps nous analysons les réactions des marchés face aux publications gouvernementales. En effet le Département de l'Agriculture des Etats-Unis (USDA, *United States Department of Agriculture*) publie chaque mois un rapport sur l'offre et la demande en denrées agricoles à travers le monde : le WASDE (*World Agricultural Supply and Demand Estimates*). Certains mois (janvier, août, septembre, octobre et novembre) il est accompagné du rapport sur les productions des cultures américaines de maïs et de soja publié par le NASS (*National Agricultural Statistics Service*). En suivant les méthodologies déjà développées à travers de nombreux papiers, nous montrons à l'aide de tests statistiques (*F-test* et *Kruskal-Wallis test*) que les rendements fermeture-ouverture des futures à maturité décembre et novembre ont une volatilité plus forte lors des sessions suivant la publication de ces rapports. En d'autres mots, le marché considère que la publication contient des informations précieuses et réagit à celles-ci.

Nous poussons notre analyse en quantifiant cette réaction et notamment la direction de celle-ci (à la hausse ou à la baisse). Pour ce faire nous considérons que l'information précieuse est contenue dans les estimations précoces des rendements agricoles. Ces données sont disponibles sur le site de l'USDA grâce à l'outil Quick Stats. Ainsi pour les mois de septembre, octobre, novembre et janvier nous estimons l'information publiée par la formule suivante,

$$X_j = \left(\frac{Y_j}{Y_{j-1}} - 1 \right) \times 100 \quad (3)$$

où Y_j représente la j -ème estimation précoce publiée par le NASS, X_j représente l'information précieuse pour le marché. Pour le mois d'août l'apport d'information est estimé à l'aide de la formule suivante,

$$X_j = \left(\frac{Y_j}{\bar{Y}_N} - 1 \right) \times 100 \quad (4)$$

où \bar{Y}_N représente la moyenne olympique des estimations finales des rendements agricoles sur les 5 dernières années (moyenne des cinq années auxquelles on soustrait le minimum et le maximum).

Enfin nous quantifions la réaction des marchés à cette estimation d'information à l'aide d'une régression linéaire. Cette étude est menée sur différentes zones géographiques : 8 Etats de la *Corn*

Belt séparément (Illinois, Indiana, Iowa, Kansas, Minnesota, Nebraska, Ohio et Dakota du Sud), l'ensemble des Etats-Unis, et l'ensemble des Etats ci-dessus. Les résultats montrent des corrélations significatives entre certaines estimations d'information et les réactions des marchés. Ces corrélations sont, comme nous l'attendions, négatives : une ré-estimation à la hausse des rendements agricoles correspond à une augmentation de l'offre, à demande constante les prix doivent donc baisser. Nous remarquons cependant que tous les mois et toutes les zones géographiques n'ont pas la même influence sur les prix.

4.2. Arbitrage financier fondé sur le NDVI

Nous essayons à présent d'estimer en avance cette information précieuse que le marché reçoit par la publication des rapports gouvernementaux. Pour ce faire nous utilisons la série temporelle des NDVI, plus précisément nos indices sont fondés sur les valeurs du pic comprises entre le 10 juin et le 29 septembre. Les deux indices nécessaires dans notre méthodologie sont la valeur maximale atteinte à l'image i notée M_i , et la valeur de l'ensemble du pic vu à la date i notée P_i . A chacune des images fournies par la NASA (tous les 16 jours) nous obtenons alors une estimation de rendement agricole grâce à la régression suivante,

$$Y_n = \beta_{i,0} + \beta_{i,1} \cdot n + \beta_{i,2} \cdot M_i + \beta_{i,3} \cdot (P_i - M_i) + \epsilon_{i,n} \quad (5)$$

où Y_n est l'estimation final du rendement agricole pour l'année n , et $\epsilon_{i,n}$ est un terme d'erreur. Cette analyse est effectuée sur les 8 Etats séparément et sur l'ensemble de ceux-ci. Comme nous ne disposons pas d'un calque différenciant les types de cultures, les résultats sur l'ensemble des Etats-Unis ne seraient pas bons à cause des interférences entre ces dernières.

Une fois ces régressions appliquées nous obtenons donc une estimation NDVI du rendement agricole pour l'image i de l'année n , notée $\widehat{Y}_{i,n}^{NDVI}$. Grâce à ces données nous estimons en avance l'information publiée par le NASS en suivant la même méthodologie que dans la partie précédente. Par exemple l'information précieuse contenue dans la j -ème publication de l'année de culture n sera estimée par $\widehat{X}_{j,n}^{NDVI}$ calculée à l'aide de la formule suivante,

$$\widehat{X}_{j,n}^{NDVI} = \left(\frac{\widehat{Y}_{i_1,n}^{NDVI}}{\widehat{Y}_{i_2,n}^{NDVI}} - 1 \right) \times 100 \quad (6)$$

où i_1 et i_2 sont deux images disponibles avant la publication telles que $i_1 \geq i_2$.

Enfin nous analysons la significativité de la corrélation entre ces estimations précoces d'information et les réactions des marchés à l'aide d'une régression linéaire. Nos résultats montrent que pour certaines publications et pour des images bien choisies, une corrélation significative apparaît. De plus, comme attendu, cette corrélation est négative. Cependant nous doutons que ces résultats puissent être suffisants pour la création d'une stratégie d'arbitrage financier afin de réduire le risque de prix détenu par les assureurs. Ce modèle doit être perçu

comme une base montrant l'existence d'une possibilité d'arbitrage, et doit être amélioré avant son utilisation pratique.

Nous donnons plusieurs pistes d'amélioration, dont notamment la possibilité d'augmenter la précision des images satellites : en appliquant un calque différenciant les cultures mais aussi en créant une série temporelle NDVI journalière. Nous n'avons pu mettre en place ces outils car ils demandent un temps de développement et surtout de calcul important dont nous ne disposions pas. Même si cette étude se focalise sur le maïs et le soja des Etats-Unis, elle peut être étendue à d'autres cultures comme le blé par exemple, ou à d'autres régions comme la bourse EURONEXT de Paris. Une dernière piste d'extension est de se focaliser sur d'autres rapports gouvernementaux, notamment l'hebdomadaire *Crop Progress* du NASS, qui ont aussi une influence sur le marché des futures CBOT.

5. Conclusion

Au terme de ce mémoire nous avons montré deux apports des images satellites à l'assurance agricole par le biais de l'indice de réflectivité NDVI. Cet indice propose une alternative à l'assurance surface-rendement classique. Il permet ainsi le développement de l'assurance agricole dans des régions du monde où les produits classiques n'étaient que très peu souscrits. Dans une moindre mesure il peut aussi diminuer le risque de base supporté par l'assuré. Une analyse plus poussée de la série temporelle des NDVI et des marchés financiers nous a permis de montrer l'existence de stratégie d'arbitrage sur les futures des matières premières agricoles. Cela permettrait de diminuer le risque de prix transféré aux assureurs par les produits d'assurance revenu. Cependant une amélioration, notamment par l'augmentation de la précision des images satellites et de la puissance de calcul, de notre modèle est nécessaire pour la mise en pratique d'une telle stratégie.

Executive Summary

Key words: agricultural insurance, agricultural revenue, agricultural yield, CBOT, corn, futures market, markets reaction, NASA, NDVI, parametric insurance, price forecasting, remote sensing, satellite image, soybean, United States

1. Basics of agricultural insurance

Farmers, through their professional activity, are highly exposed to risks. The two main risks faced are the yield risk and the price risk. The first one consist in a decrease of the agricultural yield of the farm which may be due to lack of fertilizer, diseases, climatic conditions, etc. If the farm produces less in term of quantity at constant prices, the revenue decreases. The second risk comes from the variation of the commodities prices. If the prices decrease at constant production, the revenue decreases. Both of these risks imply a high volatility of the revenue over the years for the farmers. Moreover this variation tends to increase due to the climate change and the extreme events associated to it. As a consequence famers are looking forward for tools to smooth earnings over the years.

Many of these tools are available on the market: commodity derivatives, weather derivatives, agricultural insurance, etc. In our study we restrict ourselves to the study of agricultural insurance which is the most frequently used tool in the USA. The USA, where agriculture is the most powerful worldwide, is also the country we based our examples on. In 2014 the agricultural insurance premiums had exceeded \$ 10 billion. Methodologies we develop however can be extended, with more or less of changes, to other regions. Two kinds of insurance-linked products exist: yield assurance, covering only the risk of yield, and revenue insurance, covering the risk of yield and price. Our study focuses on exploring different contributions of using satellite images from NASA and specifically NDVI (Normalized Difference Vegetation Index) to these agricultural insurance products.

2. Normalized Difference Vegetation Index

The NDVI is a reflectance index that measures the chlorophyll activity of a plant and therefore its health. It is based on the specificity of the vegetation reflectance spectrum in the red and near infrared. It is calculated as follows,

$$NDVI = \frac{R_{NIR} - R_R}{R_{NIR} + R_R} \quad (1)$$

where R_{NIR} is the reflectance in the near infrared (Near InfraRed) and R_R is the reflectance in the visible red region. Many academic papers and publications show the strong correlation (R^2 between 0.4 and 0.9) with crop yields for many kinds of crops in different geographical regions.

The data we use are provided free of charge by NASA at a resolution of 250 meters. We have a picture every sixteen days over the period from February 18th, 2000 to January 3rd, 2015. Each value of an image is the average of the last 16 days. In fact the NDVI is an index of reflectance which is very sensitive to weather conditions (clouds, rain, etc.) when it is observed from satellites. We apply a layer to only take into account the cultures, yet we do not have layers to obtain a NDVI time series by kind of crop. The values obtained are therefore those of all crops (wheat, corn, soy, etc.) combined, some interference issues between sets may occur and reduce correlations.

First we analyze how this index allows to avoid some of the constraints of classical area-yield assurance and thus promote the development of agricultural insurance in areas of the world where insurers did not want to subscribe. Secondly we seek arbitrage opportunities through NDVI on future markets of the Chicago Stock Exchange (CBOT, Chicago Board of Trade) to allow insurers to reduce the price risk they hold.

3. Yield insurance

There are two main structures for yield insurance products: one for which the compensation is calculated according to an expert who visited the site, the other where the indemnities are based on an index (parametric insurance). We are interested in the second structure because it is the one that has been developed in the US and is the most promising by its very low management costs. The classical product is area-yield insurance for which the index is called reference yield. A trigger and an exit point are defined. The allowances are then calculated as follows ,

$$I = V \times \mathbb{1}_{R \leq T} \min\{T - R; T - E\} \quad (2)$$

where I is the indemnity, V is the value of an index point, R is the observed reference yield, T the trigger point and E the exit point. The value of an index point is simply the multiplication between the insured area and the price of a unit of the insured crop.

The reference yield should however checks some properties in order to be used in the product structure. The insurer must be able to get a fairly long history of this variable (10-15 years minimum). The data should be considered sufficiently reliable. The insured farmer must have a limited influence on this index. In the US, as in France, the most often used reference yield is the estimate by the Department of Agriculture of agricultural yield of the administrative area in which the insured area is located. However these data are not always considered reliable in some politically unstable countries like Africa or Ukraine for instance. In such cases, without other index available, the insurer will usually preferred not to subscribe. To show the sensitivity of the pricing

to data, we take two examples of products (corn and soybeans, in Iowa and Illinois respectively) for which we estimate the pure premium through three different regressions methodologies. We then observe variations ranging from single to double depending on the regression applied.

These data reliability issues can be resolved using an index based on the NDVI. Indeed NASA data is considered reliable enough by insurers and obtaining index values has a very limited cost (internet connection, downloading images, calculate values using R or other software). We therefore propose a NDVI based yield insurance structure covering the same risks as the examples discussed above. The index used matches the peak NDVI, that is to say during the growth phase of the plant. The values are taken into account between June 26th and September 13th. We show that this index is correlated with actual agricultural yields (R^2 0.85 for corn and 0.63 for soybeans). We also propose a pricing methodology for the pure premium of the structured products.

Choosing a NDVI based insurance product can also be motivated by the desire to reduce basis risk held by the insured. With the new resolution images (250 meters) NDVI may be centered on a particular farm in order to obtain a better correlation than with the reference yield from governmental data. In these cases the use of NDVI has a significant contribution even in states with public reliable data such as the United States.

The use of NDVI may, however, poses a new problem: if the values are only those of the insured farm there is a risk of moral hazard. Indeed the farmer can influence the health of crops and therefore NDVI. It is then advisable to apply a deductible on the compensation price per unit of production, or to couple the NDVI index with a weather index on which the operator will have no possible influence.

4. Revenue insurance

The revenue insurance is a kind of agricultural insurance products that covers both the risk of yield and price risk. These products are widely used in the United States where they account for several years about 80% of premiums (in 2014 for instance the amount of revenue insurance premiums was 8.18 billion US dollars). There are many possible structures that we do not detail. The major difference between revenue insurance and yield insurance is that the compensation price per unit of production is not known in advance and is calculated from the futures market at the time of harvest (November for soybeans and December for corn). Thus insurance companies are forced to manage a large portfolio of derivatives on agricultural commodities.

4.1. Governmental reports and markets reactions

First we analyze the markets reactions to government reports. Indeed the Department of Agriculture (USDA, United States Department of Agriculture) publishes a monthly report on supply and demand for agricultural commodities worldwide: the WASDE (World Agricultural

Supply and Demand Estimates). Some months (January, August, September, October and November) it is published simultaneously with the report on the production of US crops of corn and soybeans published by the NASS (National Agricultural Statistics Service). By following the methods already developed through many papers, we show, using statistical tests (F-test and Kruskal-Wallis test), that closed-open returns of futures maturing in December and November were more volatile during session following the release of the reports. In other words, the market considers the report contains valuable information and then responds to it.

We push our analysis by quantifying this reaction, including the direction of it (upward or downward). To do this we suppose that valuable information is contained in the early estimates of crop yields. These data are available on the USDA website through Quick Stats tool. Thus for the months of September, October, November and January we suppose that the information published is giving by the following formula,

$$X_j = \left(\frac{Y_j}{Y_{j-1}} - 1 \right) \times 100 \quad (3)$$

where Y_j = represents the j -th early estimates published by NASS, X_j represents the valuable information for the market. For the month of August the new information is estimated using the following formula,

$$X_j = \left(\frac{Y_j}{\bar{Y}_N} - 1 \right) \times 100 \quad (4)$$

where \bar{Y}_N represents the Olympic average of the estimates of crop yields over the past five years (average of five years to which we subtract the minimum and maximum).

Finally we quantify the markets reactions to this information using linear regression. This study is conducted on different geographical regions: 8 states in the Corn Belt separately (Illinois, Indiana, Iowa , Kansas , Minnesota , Nebraska , Ohio and South Dakota), the whole United States, and all the States above together. The results show significant correlations between some information estimates and markets reactions. These correlations are, as expected, negative: a re-estimation to higher agricultural yields corresponds to an increase in supply at constant demand, therefore prices should fall. We note, however, that every month and all geographic areas do not have the same influence on prices.

4.2. NDVI based financial arbitrage

We try now to estimate in advance this valuable information the market receives from the publication of government reports. To do this we use the NDVI time series, specifically our indices are based on the values of the peak between June 10th and September 29th. The two indices needed in our methodology are the maximum value reached seen at the image i denoted M_i , and the value

of the peak seen at the image i denoted P_i . For each of the images provided by NASA (every 16 days) we thus obtain an estimate agricultural yield through the following regression,

$$Y_{i,n} = \beta_{i,0} + \beta_{i,1} \cdot n + \beta_{i,2} \cdot M_i + \beta_{i,3} \cdot (P_i - M_i) + \epsilon_{i,n} \quad (5)$$

where Y_n is the final estimation of agricultural yield for the year n , and $\epsilon_{i,n}$ is an error term. This analysis is performed on the 8 states separately and over the whole Corn Belt. As we do not have a layer to difference crop types, the results on all the USA would not be good because of interference between crops.

Once these regressions applied we thus get an estimate of NDVI agricultural yield for the image i of year n , denoted $\widehat{Y}_{i,n}^{NDVI}$. With this data we estimate in advance the information published by NASS following the same methodology as in the previous section. For example, the valuable information in the j -th publication of the crop year n is estimated by using $\widehat{X}_{j,n}^{NDVI}$ calculated with the following formula,

$$\widehat{X}_{j,n}^{NDVI} = \left(\frac{\widehat{Y}_{i_1,n}^{NDVI}}{\widehat{Y}_{i_2,n}^{NDVI}} - 1 \right) \times 100 \quad (6)$$

where i_1 and i_2 are two images available before the report's release such as $i_1 \geq i_2$.

Finally we analyze the significance of the correlation between these early estimates of information and markets reactions using linear regression. Our results show that for some publications and appropriate images, a significant correlation appears. Furthermore, as expected, the correlation is negative. But we doubt that these results can be sufficient to create a financial arbitrage strategy to reduce the price risk held by insurers. This model should be seen as a base, showing the existence of an arbitrage opportunity, and must be improved before its practical use.

We give several possible improvements, including the ability to increase the accuracy of satellite images: by applying a layer which allows to difference the crops but also by creating a daily NDVI time series. We could not develop such tools because they require an important development time and an even more important calculation power which we did not have. Although this study focuses on corn and soybeans from the United States, it can be extended to other crops such as wheat, for example, or other areas such as the Euronext Paris stock exchange. One last extension would be to focus on other reports, such as the weekly Crop Progress from the NASS, which also have an influence on the market.

5. Conclusion

To put it in a nutshell we showed two contributions of satellite images to agricultural insurance through the NDVI reflectance index. This index provides an alternative to conventional area-yield insurance. It allows the development of agricultural insurance in regions of the world

where traditional products used to not be subscribed. To a lesser extent it can also reduce the basic risk held by the insured farmer. Further analysis of the NDVI time series and financial markets has allowed us to show the existence of arbitrage strategies on commodities futures. This would lower the price risk transferred to insurers by revenue insurance products. However an improvement, in particular by increasing the accuracy of satellite images and the computing power, of our model is needed before the setting up of such a strategy.