

ESSEC

ISUP

Promotion 2010

Mémoire présenté devant

**L'Institut de statistique
de l'Université Pierre et Marie Curie**

et

L'Ecole Supérieure des Sciences Economiques et Commerciales

pour l'obtention du

Diplôme de l'Institut des Actuares

Assurance Finance

Par **Mlle Juliette Bugnon**

Sujet : Etude des opportunités d'investissement en assurance non-vie sur les marchés émergents

Lieu du stage : EMINCAP, Paris

Responsable du stage : M. Olivier Lopez
Responsable entreprise : M. Yann Drévilion

Invité(s) :

CONFIDENTIEL

Remerciements

Je tiens à remercier Messieurs Yann Dréville et Olivier Jaillon pour m'avoir proposé ce sujet avec lequel j'ai pu approfondir mes connaissances en macroéconomie des pays émergents, et pour m'avoir permis de réaliser mon mémoire au sein d'EMINCAP. Je remercie en particulier Monsieur Dréville pour son soutien dans la phase de recherche d'informations et sa disponibilité pour échanger sur le modèle.

J'adresse toute ma reconnaissance à Monsieur Jean-Marie Nessi. Son écoute et ses conseils m'ont apporté une aide décisive, tout au long du stage.

Je tiens également à remercier Monsieur Olivier Lopez, pour sa disponibilité et son aide à la rédaction de ce mémoire.

Je remercie Messieurs les professeurs de l'ISUP et du département Systèmes d'Informations et de Décision de l'ESSEC, dont les cours m'ont été très utiles dans la rédaction de ce mémoire.

Enfin, je souhaite exprimer toute ma gratitude à Monsieur Jean Pierre Indjehagopian, qui m'a fait découvrir l'économétrie et a su me communiquer son enthousiasme pour les études d'actuariat.

Résumé

Le marché de l'assurance automobile français, comme de nombreux marchés occidentaux, a aujourd'hui atteint un niveau de maturité qui ne laisse que peu de possibilités pour de nouveaux développements. Par ailleurs, la forte croissance économique de nouveaux pays, les « pays émergents », ouvre d'autres horizons à l'activité d'assurance. En particulier, l'exploitation de l'expérience française pour la détermination du meilleur fonctionnement d'une jeune compagnie dans un contexte de naissance ou de croissance rapide de l'assurance, peut lui permettre de devenir leader sur son marché.

C'est à partir de cette idée que s'est construit le projet EMINCAP ou Emerging Insurance Capital, créé par les fondateurs du groupe Protegys, Yann Drévilion et Olivier Jaillon, en 2008.

Dans ce mémoire, nous nous proposons d'apporter un soutien au processus décisionnel d'EMINCAP en définissant des critères de choix de pays représentant des opportunités d'investissement.

Mots-clés : pays émergents, assurance automobile, économétrie, statistiques multidimensionnelles, données manquantes, notation

Abstract

The French automobile insurance market, as all occidental developed markets, has nowadays reached a maturation point which does not leave a large area for new opportunities of growing business. Parrallely, the fast development of other countries, the “merging countries”, offers new horizons to the automobile insurance activity. Indeed, basing on the acquired experience on the French market, it seems that contributing to the development and optimization of a fresh insurance company, in a country where the activity is quite new and fast growing, shall help this company become a market leader.

The Project EMINCAP -for Emerging Insurance Capital- was created under this idea by Yann Drévilion and Olivier Jaillon (CEOs of the company PROTEGYS) in 2008.

This memoire is aimed to support EMINCAP in the decisional process, by proposing selection criteria for countries representing investment opportunities in the insurance sector.

Key-words: emerging markets, automobile insurance, econometrics, multidimensional statistics, missing data, notation

Sommaire

Remerciements	2
Résumé.....	3
Abstract.....	4
Introduction	7
Présentation d'EMINCAP	7
Une structure de fond d'investissement	7
Des caractéristiques spécifiques.....	7
Un projet.....	7
Définition des objectifs du mémoire	8
Problématique : la notion de pays émergent et de pays émergent en assurance	9
Qu'est ce qu'un « pays émergent » ?	9
Comment définir un pays émergent en assurance ?	10
Les apports de « la Société du Risque », d'Ulrich Beck.....	10
Critères recherchés	11
1. Recherche d'informations et constitution de la base de données	12
1.1. Sélection des critères	13
1.2. Présentation des critères retenus.....	14
1.2.1. Insurance Overview	14
1.2.2. Economic Overview.....	15
1.2.3. Outlook.....	16
1.2.4. Political & Legal Overview.....	17
1.2.5 Variables qualitatives	21
2. Préparation des données	24
2.1 Analyse de la table de données manquantes	24
2.1.1. Analyse de la capacité de reporting par pays	25
2.1.2 Construction des paliers de modélisation des données manquantes	26
2.2 Modélisation de l'appartenance à une Organisation Economique	31
2.2.1 Méthodologie employée.....	31
2.2.2 Construction de l'arbre de classification.....	32
2.2.3 Résultats et Interprétations	33
3. Consolidation des données manquantes.....	35
3.1. Choix de la Modélisation par ANCOVA	35
3.1.1. Définition et Principe de l'ANCOVA	35

3.1.2. Variables explicatives initiales	37
3.2 Application de l'ANCOVA	37
3.2.1. Données à consolider	37
3.2.2. Statistiques descriptives	37
3.2.3 Matrice de corrélation	38
3.3 Consolidation des données pour le palier 1.....	41
3.3.1. Variable « Corruption Level »	41
4. Sélection.....	67
4.1. « Profiling » des pays : Application de l'ACP.....	67
4.1.1. Méthodologie de l'Analyse en Composantes Principales.....	67
4.1.2. ACP sur base complète	68
4.1.3. ACP sur base réduite.....	76
4.1.4. Visualisation des regroupements par pays.....	85
4.1.5. Caractéristiques des classes.....	87
4.1.6. Résultats.....	88
4.2. Construction du modèle de notation.....	89
4.2.1. Construction des Clusters	89
4.2.2. Représentation graphique par pays.....	90
4.2.3. Règles de décision.....	96
4.2.4. Résultats.....	96
4.3. Synthèse des pays sélectionnés	97
Conclusion.....	98
Bibliographie	99
Annexes.....	101

Introduction

Présentation d'EMINCAP

Une structure de fond d'investissement

EMINCAP (Emerging Insurance Capital) est une société d'investissement créée par Olivier Jaillon et Yann Drévilion en 2008.

Comme tout fond d'investissement, son objectif principal est de lever des fonds, collecter des capitaux, objectif pour lequel la société doit se rendre attractive sur le marché.

L'investissement de type « Private Equity » implique un objectif de rentabilité élevée via des leviers ou stratégies d'investissement innovantes.

L'existence d'un risque relativement élevé, à hauteur de la rentabilité espérée, est prise en compte. On souhaite pouvoir se donner des indicateurs de mesure du risque encouru.

Des caractéristiques spécifiques

EMINCAP comporte deux spécificités principales qui le démarquent des autres fonds d'investissement :

Son portefeuille d'investissement est uniquement basé sur le marché de l'assurance,

La sélection se fait sur les marchés émergents en assurance, ce qui nécessite un repérage des opportunités d'investissement.

Un projet

L'idée part d'un constat fait sur l'état actuel du marché de l'assurance dans le monde : d'une part, on ne trouve aujourd'hui de forts taux et potentiels de croissance que sur des marchés dits émergents ; d'autre part, il existe une corrélation forte entre le développement du secteur de l'assurance et la croissance économique dans un pays donné.

Sur cette base, la société décide de construire un modèle d'investissement assez novateur, puisque spécialisé sur le domaine de l'assurance, qui constitue sur les marchés émergents une niche d'investissement. Des critères économiques et financiers plus classiques sont également à prendre en compte pour repérer les marchés émergents actuels, mais on souhaite limiter le scope aux seuls marchés émergents en assurance.

Une fois le processus de sélection mis en place, on se focalise sur un groupe de pays définis comme « marchés émergents en assurance » et représentant une opportunité d'investissement. L'étape suivante consiste à récolter des fonds pour investir dans une société sur un des marchés sélectionnés, avec un objectif de restructuration et d'optimisation. La société une fois réorganisée sera cédée à moyen terme (3 à 5 ans).

La croissance de la société s'accompagnera localement d'un développement économique, notamment via le financement de formations en assurance pour les manageurs. Il s'agit de développer une structure opérationnelle et autonome.

Le secteur privilégié au départ est celui de l'assurance automobile, pour lequel EMINCAP a le plus d'expérience, et qui comporte l'avantage d'une mise en place simplifiée et d'un besoin adapté à tous les marchés potentiels, ainsi qu'une croissance assez rapide.

Définition des objectifs du mémoire

L'objectif principal de ce mémoire est de venir en support à la prise de décision d'investissement d'EMINCAP. La mise en place d'un modèle décisionnel d'investissement comporte plusieurs étapes distinctes :

Tout d'abord, on doit constituer une base d'informations à caractère économique, social et assurantiel, qui nous permettront de mesurer le niveau de développement (économique et assurantiel) des pays. Il s'agit ici d'un travail de recherche de sources, mais aussi de traitement des données manquantes afin d'obtenir une base consolidée et complète.

On souhaite ensuite construire un modèle de notation un pays émergent en assurance. A partir de ces critères, on choisit un modèle de notation qui nous permettra d'effectuer un classement des pays par niveau d'opportunité d'investissement.

Pour les pays sélectionnés, on fournit une description plus approfondie de l'état économique et assurantiel.

Problématique : la notion de pays émergent et de pays émergent en assurance

Qu'est ce qu'un « pays émergent » ?

La notion de pays émergent n'est pas évidente a priori, et il existe de nombreuses définitions différentes, selon des critères n'incluant pas toujours les mêmes groupes de pays.

Pour le Centre d'études prospectives et d'informations internationales (CEPII), un pays est dit émergent si trois critères sont remplis :

- un niveau de richesse (revenu par tête moyen inférieur à 70% du niveau moyen des pays de l'OCDE),
- une participation croissante aux échanges internationaux de produits manufacturés (croissance des exportations de produits manufacturés, y compris les industries agroalimentaires supérieures de 2% en moyenne par an à la croissance des échanges mondiaux)
- une forte attraction de ce pays sur les flux internationaux de capitaux.

L'encyclopédie Larousse, dans l'article « émergent » issu du dossier sur le développement, nous donne une définition plus approfondie du concept d'« émergence » : un pays émergent se distingue des pays en développement par des résultats macroéconomiques (production industrielle, emploi) supérieurs et un taux de croissance élevé.

« Caractérisés par un brusque décollage économique, fondé sur l'industrialisation, et par un fort potentiel de croissance, les pays émergents ont quitté les rangs du tiers-monde dans les années 1970. Leurs performances, influant notamment sur l'indice de développement humain (I.D.H.) et sur le taux d'équipement des ménages, sont liées à la conquête des marchés internationaux par leurs produits d'exportation. »

Historiquement, les pays émergents sont d'anciens pays en développement, qui ont accéléré leur croissance. Les premiers pays correspondant à la définition d'émergence ont été les nouveaux pays industrialisés (N.P.I.) aussi appelés les « quatre dragons » d'Asie : la Corée du Sud, Taïwan, Singapour et Hong Kong, à qui se s'est ajoutée ensuite une seconde génération de nouveaux pays industriels en Asie du Sud-Est, les « tigres » : la Thaïlande, les Philippines et la Malaisie.

À l'expression « nouveaux pays industriels », très utilisée dans les années 1980, s'est substituée celle de « pays émergents ». Aujourd'hui, cette catégorie correspond à un groupe plus large de pays, tels que la Chine, l'Inde et l'Indonésie, ainsi que des pays latino-américains comme le Brésil et l'Argentine. Néanmoins, les pays émergents ne constituent pas une entité globale : leur histoire et leur développement de même que leurs structures économiques et sociales sont différentes, et il sera nécessaire de prendre en compte cette hétérogénéité dans l'optique d'une décision d'investissement.

Comment définir un pays émergent en assurance ?

Les apports de « la Société du Risque », d'Ulrich Beck.

L'étude de la théorie d'Ulrich Beck a pour objectif de comprendre ce qu'est la société du risque, quels sont les indicateurs économiques et les autres facteurs qui la caractérisent.

On souhaite ainsi déterminer comment on passe d'une société en développement, dont la première préoccupation est de résorber une situation de pénurie, à une société « moderne », où la production de richesses et de risques sont directement corrélées, et où la première préoccupation devient la recherche de protection contre ces risques : quelles sont les étapes de cette transition, quels indicateurs permettent de représenter cette évolution ?

L'ouvrage d'Ulrich Beck date de 1986. Par conséquent, il ne nous indique pas directement ce qui caractérise aujourd'hui les pays où on observe un fort développement de l'assurance, mais nous explique plutôt comment s'est faite l'évolution vers la « société du risque » pour les pays occidentaux.

Après avoir analysé les aspects historiques de cette transition, nous chercherons à créer un parallèle avec la situation actuelle et à déterminer quelles caractéristiques des pays en développement semblent déterminantes dans la définition d'un pays émergent en assurance.

Dans la première partie, Beck nous explique que la société du risque se construit parallèlement, avec un décalage temporel, à la société industrialisée.

En effet, dans la société préindustrielle, la principale difficulté est celle de la production de biens de consommation en quantité suffisante pour répondre à la demande. A partir du moment où un pays entre en révolution industrielle, le progrès technique va apporter la solution à la problématique de la pénurie, et un développement rapide des classes moyennes. Ces classes moyennes vont être la cible première d'une activité d'assurance : elles possèdent déjà un certain nombre de biens à assurer, mais ne sont pas suffisamment riches pour s'auto-assurer.

Beck s'intéresse tout particulièrement aux risques industriels et économiques, qui sont une conséquence directe du développement industriel et économique : il étudie par exemple la pollution, le risque nucléaire et le chômage, qui sont des risques de groupe. On peut néanmoins adapter cette théorie à des risques individuels : ainsi, la forte croissance du marché automobile dans un pays va impliquer un développement du trafic routier, une augmentation du risque d'accident et une volonté croissante de se protéger face à ce risque.

Tout d'abord, Beck indique les états de passage à la société du risque : « La nécessité de se prémunir du risque vient remplacer la nécessité de se prémunir de la pénurie ». En considérant le risque de pollution engendré par la production industrielle de masse, la prise de conscience du danger et la volonté de s'en prémunir se fait alors par échelle de zone menacée, et non plus par Etats-nations ou par blocs économiques. La mondialisation actuelle plus encore que la situation en 1986 brouille les pistes en matière de pollution, notamment par le biais des entreprises multinationales. On fait face à une internationalisation des risques industriels : un critère d'indication du niveau de pollution nous semble ainsi trop difficile à utiliser, et sa significativité serait a priori discutable à l'échelle d'un pays : un pays émergent en assurance serait-il alors un pays qui commence à développer des industries polluantes par lui-même ? A exporter des industries polluantes ? Par ailleurs, la prise de conscience vis-à-vis des dangers de la pollution est beaucoup plus importante aujourd'hui, avec par exemple 183 pays ayant ratifié le Protocole de Kyoto dans le monde.

A la fin du chapitre 1, Beck résume bien la transition observée en deux sensations : dans la société préindustrielle, la motivation première est « j'ai faim », et l'idéal, celui de l'égalité de la

répartition entre les classes ; dans la société du risque, l'idéal est celui de la sécurité pour tous, motivé par un « j'ai peur » collectif.

Dans un deuxième temps, on étudie une autre caractéristique essentielle de la société du risque : l'importance fondamentale des médias, moyens de communication et d'information. Beck affirme même qu' « écrire une sociologie politique et une théorie de la société du risque, c'est par définition écrire une sociologie du savoir ». Ces aspects sont plus facilement quantifiables par pays, par exemple en étudiant les taux relatifs de pénétration des téléphones mobiles, en forte croissance dans les pays en développement, ainsi que d'internet, véhicule puissant et mondialisé de l'information. Ces indicateurs auront aussi un intérêt supplémentaire dans l'optique d'une opportunité d'investissement en assurance : évaluer les possibilités sur un marché donné d'employer des techniques de vente innovantes sur le modèle du « mobile banking » ou de la banque à distance, en proposant des contrats d'assurance pouvant être souscrits et gérés par internet ou téléphone mobile.

Critères recherchés

Au final, comment cibler les pays recherchés ?

Nous avons tout d'abord vu que pour sélectionner un pays émergent, on doit se baser sur des critères de forte croissance, par exemple du PIB par habitant. On a aussi vu que les pays émergents voient leur situation économique s'améliorer significativement, et leur niveau de vie se rapprocher des standards occidentaux : il sera donc intéressant de prendre en compte l'indicateur de développement humain, ainsi que le niveau d'équipement en biens de consommation, par exemple en automobile ; ce dernier indicateur nous apportera aussi la taille du marché de l'assurance automobile pour un pays donné.

Pour prendre en compte l'aspect assurantiel de notre recherche, on intégrera aussi à nos critères les volumes de primes par pays, le taux de pénétration de l'assurance et les taux de croissance des primes d'assurance sur les dernières années. Nous avons aussi vu que le développement des moyens de communication semble être un bon indicateur du niveau d'émergence d'un pays : on intégrera les taux de pénétration par pays en téléphone mobile et internet.

Enfin, notre objectif n'est pas seulement de définir un ensemble de pays émergents en assurance, mais bien de cibler des opportunités d'investissement sur ce secteur.

Cela implique de prendre en compte plus spécifiquement les risques associés à ce type d'investissement. On sélectionnera donc des indicateurs de la qualité de l'environnement légal, du niveau de corruption, des risques économiques, financiers et politiques dans chaque pays. On ajoutera également un indicateur du positionnement des pays concernant les investissements étrangers.

1. Recherche d'informations et constitution de la base de données

Afin de construire une base au modèle décisionnel, on constitue une base de données sur un ensemble de critères. Parmi ceux-ci :

- des critères quantitatifs macroéconomiques
- des critères quantitatifs spécifiques au domaine de l'assurance
- des critères qualitatifs économiques et assurantiels

Dans cette partie nous définirons et étudierons l'ensemble des critères pris en compte pour la mise en place du modèle décisionnel. Pour chacun d'entre eux, nous analyserons sa méthode de construction, ses apports et ses limites dans le cadre du modèle décisionnel que nous souhaitons développer. Les sources de données sont indiquées pour chaque variable, les liens vers les bases de données d'origine sont indiqués en annexe.

1.1. Sélection des critères

Une fois la base de données constituée, on a à disposition 25 critères, parmi lesquels on choisit d'en retenir 23 que l'on va classer en quatre familles d'analyse :

- Insurance Overview
- Economic Overview
- Outlook
- Political and Legal Overview

Les variables ont été nommées en anglais à la demande de l'entreprise EMINCAP. Elles seront définies dans les pages suivantes.

L'Insurance Overview est composé de :

P&C spending per capita
Motor TPL
Penetration rate
Distribution Network

L'Economic Overview est composé de :

Vehicles per capita
Mobile phone penetration
Internet penetration
Population
GDP/capita
IDH

L'Outlook est composé de :

Population Growth Rate
Real annual GDP Growth
Vehicle ownership projection
P&C spending trend
Economic Freedom Index

Le Political & Legal Overview est composé de :

Economic Risk
Financial Risk
Political Risk
Corruption level
Legal framework
Control of Corruption
Political stability

On inclue également dans la base deux indices qualitatifs :

La zone géographique : le continent dans lequel se situe le pays,
L'organisation économique dont le pays fait partie.

1.2. Présentation des critères retenus

1.2.1. Insurance Overview

Les sociétés de réassurance publient annuellement des rapports sur l'état de l'assurance dans le monde. Les publications « Analyse Mondiale » 2007 et 2008 de Swiss-Re fournissent en particulier une base intéressante pour mieux appréhender les différentes zones de marchés émergents. Les données techniques en annexe viennent alimenter la base de données sur des critères assurantiels (volume et évolution des primes Vie et Non Vie) et économiques (PIB, population, taux d'inflation).

P&C spending per capita

Cette variable, issue de l'analyse Sigma 2008 de Swiss-Re, représente le montant en millions USD de primes Non Vie émises en 2008 par pays.

Elle donne une vue actualisée de l'état du marché de l'assurance Non Vie, avec une échelle permettant de comparer l'ensemble des pays.

Les données ne sont pas disponibles pour 75 pays sur 164, soit presque la moitié de notre cible, et ne sont pas accessibles via d'autres études.

Motor TPL

Cette donnée, fournie par l'AIDA Motor Working Party, représente de façon binaire le caractère obligatoire ou non de l'assurance automobile responsabilité civile dans chacun des pays étudiés.

La base ne comporte pas de données manquantes pour cet indice.

La limite de cette variable est qu'elle ne donne pas de précision sur la sévérité de la régulation en matière d'assurance automobile obligatoire, il n'est pas facile de mesurer le champ d'application réel de la mesure.

Penetration rate

Ce taux provient de l'étude Sigma 2008 de Swiss-Re. Il s'agit d'un ratio construit à partir des montants de primes d'assurance Non Vie émises par pays en fonction du PIB 2008 par pays. Les données sont exprimées sous forme de pourcentages.

Elle permet de mettre en évidence de manière assez simple les pays où le marché de l'assurance est déjà développé, et ceux qui possèdent encore un fort potentiel de croissance.

Cette variable comporte 75 données manquantes sur 164.

Distribution Network

Cette donnée, construite sur la base de recherches par pays, constitue un indicateur de l'étendue du réseau de distribution d'assurance Non Vie dans chaque pays. On s'intéresse à la présence de quatre différents canaux de distribution : Assurances directes, Brokers (Courtiers), Agents d'assurance, Banques.

Cette variable qualitative donne deux informations : la diversité du réseau de distribution déjà en place ainsi que sa taille.

Cette variable ne comporte pas de données manquantes par pays, mais la possibilité de données incomplètes, du fait de la difficulté à obtenir les informations (recherche au cas par cas). Les situations de monopole et d'assurance Islamique dite « takaful » sont néanmoins mises en évidence.

1.2.2. Economic Overview

Vehicles per capita

Le classement des pays par nombre d'immatriculation est obtenu par Wikipedia (article « List of countries by vehicles per capita »). Les données datent, à quelques exceptions près, de 2008. Cette donnée nous indique le volume du parc automobile immatriculé par pays. Les valeurs sont exprimées en ‰.

Cette variable comporte 41 données manquantes sur 164.

Mobile phone Penetration

Cet indice est fourni par l'International Telecommunication Union, l'agence des Nations Unies pour les technologies de l'information et de la communication. Il représente le taux d'équipement en téléphonie mobile en 2008 par pays.

La base de données est complète pour cet indice.

Internet Penetration

Cet indice est fourni par le site Internet World Stats qui propose des données en termes d'utilisation internet dans 233 pays. Il représente le taux de pénétration Internet en 2008 par pays.

La base de données est complète pour cet indice.

Population

Cette donnée représente la taille de la population par pays. Elle est issue de l'étude Sigma 2008 de Swiss-Re, complétée par les informations fournies par la Banque Mondiale pour les pays n'étant pas pris en compte par l'étude Sigma.

La base de données est complète pour cet indice.

GDP/capita

Cette donnée est issue de l'étude Sigma 2008 de Swiss-Re, complétée par les informations fournies par la Banque Mondiale pour les pays n'étant pas pris en compte par l'étude Sigma. Elle représente le PIB par habitant, qui nous donne une indication du niveau de développement économique du pays.

La base de données est complète pour cet indice.

IDH

L'Indicateur de Développement Humain est une donnée fournie par le PNUD (Programme des Nations Unies pour le Développement), une entité de l'ONU fondée en 1990, qui publie le Human Development Report.

L'IDH est un indice composite du niveau de longévité (mesuré par l'espérance de vie à la naissance), du niveau d'éducation (mesuré par le taux d'alphabétisation des adultes et le taux brut de scolarisation) et du niveau de vie (mesuré par le logarithme du PIB/habitant en parité de pouvoir d'achat). L'IDH est calculé comme la moyenne de ces trois éléments.

Nous avons choisi de prendre en compte le classement IDH de chaque pays plutôt que la valeur de l'IDH lui-même. Le classement est construit par ordre décroissant : les premiers pays sont ceux dont l'IDH est le plus élevé.

Notre base est complète pour cet indice.

1.2.3. Outlook

Population Growth Rate

Cet indice donne le taux de croissance de la population par pays. Les données sont fournies par le CIA World Factbook, Edition de Janvier 2009.

La base de données est complète pour cet indice.

Real annual GDP Growth

Cet indice donne le taux de croissance annuelle réelle du PIB par pays. Il provient de l'étude ICRG (International Country Risk Guide), où il est utilisé dans la construction de l'indicateur composite ICRG Economic Risk Rating.

La base de données est complète pour cet indice.

Vehicle ownership projection

Cet indice donne une estimation du taux de croissance annuel du parc automobile sur les 20 prochaines années. Les données sont obtenues à partir du travail de recherche « Vehicle Ownership and Income Growth, Worldwide: 1960-2030 » par Joyce Dargay (Janvier 2007).

Cet indice comporte 64 données manquantes sur 164.

P&C spending trend

Cette valeur est calculée à partir des données issues des études Sigma 2004 à 2008 de Swiss-Re. Elle représente le taux de croissance annuel moyen des dépenses d'assurance non-vie par pays sur les 5 dernières années disponibles. Celui-ci a été modélisé comme la moyenne des taux de croissance d'assurance non-vie de 2004 à 2008.

Cet indice comporte 75 données manquantes sur 164.

1.2.4. Political & Legal Overview

Pour étudier la liquidité et la stabilité des marchés, on choisit de travailler sur les indices composites ICGR (International Country Risk Guide) développés par le groupe PRS. Celui comprend trois indices agrégés : Risque Economique, Risque Financier et Risque Politique, que l'on peut analyser séparément. L'interprétation des scores est la même pour les deux indices Risque Economique et Risque Financier. Le score attribué est compris entre 0 et 50% pour chacun, avec des catégories (« clusters ») réparties comme suit :

- 0 à 24,9% : Risque très élevé
- 25 à 29,9% : Risque élevé
- 30 à 34,9% : Risque moyen/modéré
- 35 à 39,9% : Risque faible
- 40 à 50% : Risque très faible

Economic Risk

L'indice Risque Economique est construit à partir des composantes suivantes :

PIB per capita

Croissance réelle du PIB depuis 1990

Taux d'inflation annuel

Equilibre budgétaire / PIB

Equilibre de la balance courante / PIB (solde des flux monétaires d'un pays résultant des échanges internationaux de biens et services (balance commerciale), revenus et transferts courants)

Financial Risk

L'indice Risque Financier est construit à partir des composantes suivantes :

Dettes extérieures / PIB

Remboursement de la dette extérieure / Exportations de biens et services

Balance courante / Exportations de biens et services

Liquidité internationale nette / Réserves en nombre de mois d'importations

Stabilité des taux de change

Political Risk

L'indice Risque Politique est construit à partir des composantes suivantes :

- Stabilité gouvernementale
- Conditions socio-économiques
- Profil d'investissement
- Conflit interne
- Conflit externe
- Corruption
- Présence de l'armée en politique
- Tensions religieuses
- Loi et respect de l'ordre
- Tensions ethniques
- Niveau de responsabilité démocratique
- Qualité de l'administration

Corruption Level

L'indice Corruption Level (ou Indice de Perception de la Corruption) est construit par Transparency International. Il s'agit d'un organisme indépendant dont l'objectif est la lutte globale contre la corruption. L'indice est construit à partir de dix agences indépendantes rapportant sur le niveau de corruption perçu dans les secteurs public et politique de chaque pays.

La base est complète pour cette variable.

Worldwide Governance Indicators

Les Worldwide Governance Indicators (WGI) sont des indicateurs construits par la Banque Mondiale qui regroupent un ensemble de critères sur le thème de la gouvernance individuelle des Etats. Les données concernent 212 pays et territoires, sur la période 1996-2008.

Six axes sont considérés dans le rapport :

- Voice and Accountability
- Political Stability and Absence of Violence
- Government Effectiveness
- Regulatory Quality
- Rule of Law
- Control of Corruption

Ces six indicateurs sont constitués d'une agrégation de données de nombreuses sources différentes (35 bases de données, construites par 33 organisations). Les scores attribués constituent un classement en base 100. Ainsi, un score est de 70 signifie que le pays est classé « meilleur » que 70% des pays de l'étude, et « moins bon » que 30% d'entre eux, sur le critère mesuré.

On a choisi de prendre en compte les indicateurs de stabilité politique («Political Stability and Absence of Violence»), de qualité de du cadre légal («Legal Framework», Indice composite de 3 indicateurs en base 100 : « Government Effectiveness », « Regulatory Quality » et « Rule of Law ») et de contrôle de la corruption pour l'ensemble des pays concernés.

L'étude « Worldwide Governance Indicators » de la Banque Mondiale a deux objectifs majeurs :

- proposer aux entrepreneurs et investisseurs un « état 2009 » de la situation dans le monde en termes de corruption, gouvernance et sécurité
- se baser sur une méthodologie innovante qui met pays développés et pays émergents sur une ligne d'égalité a priori, et prendre en compte non seulement les statistiques officielles, mais aussi une étude basée sur des expériences de terrain : L'Executive Opinion Survey construite par le World Economic Forum.

Les conclusions de l'étude nous montrent que :

- la séparation pays développés / pays émergents est trop manichéenne en termes de gouvernance et de corruption : en effet, il existe parfois une différence plus flagrante entre des pays considérés comme développés, qu'avec d'autres pays considérés comme émergents. L'exemple le plus fréquent est celui des pays d'Europe du Sud (notamment l'Italie, mais aussi, l'Espagne, le Portugal et la France), par rapport aux Nouveaux Pays Industrialisés d'Asie (les 4 « Dragons » : Corée du Sud, Hong Kong, Singapour et Taiwan) ainsi que ceux en développement (les 5 « Tigres » : Thaïlande, Malaisie, Indonésie, Philippines et Brunei)
- les entreprises multinationales ne subissent pas seulement les lois et réglementations, elles jouent aussi un rôle dans la corruption mondiale. On observe notamment que, selon les pays d'implantations, les comportements de ces entreprises varient entre la maison-mère et les succursales situées dans d'autres zones géographiques.
- Les problématiques de sécurité et de coût du terrorisme sont plus complexes : elles semblent plus liées à la zone géographique concernée, et sont très variables selon le type de crime étudié : par exemple, les coûts pour les pays du G7 sont globalement relativement faibles, alors que le coût de la lutte contre le terrorisme dans ces pays est élevé.

Legal framework

Il s'agit d'un indice composite de 3 indicateurs en base 100 :

Government Effectiveness
Regulatory Quality
Rule of Law

La base est complète pour cette variable.

Control of Corruption

Cette variable correspond à l'indicateur Control of Corruption, en base 100.

La base est complète pour cette variable.

Political stability

Cette variable correspond à l'indicateur Political Stability and Absence of Violence, en base 100.

La base est complète pour cette variable.

Economic Freedom Index

L'Economic Freedom Index est construit par The Heritage Foundation en association avec le Wall Street Journal. La "Liberté Economique", telle qu'elle est définie dans la construction de cet indice, représente le droit de travailler, produire, consommer, investir de chaque individu dans un pays donné. Elle est mesurée selon dix axes :

- Business Freedom
- Trade Freedom
- Fiscal Freedom
- Government Spending
- Monetary Freedom
- Investment Freedom
- Financial Freedom
- Property rights
- Freedom from Corruption
- Labor Freedom

La base est complète pour cette variable.

1.2.5 Variables qualitatives

Zone

Cette variable classe les pays selon leur situation géographique :

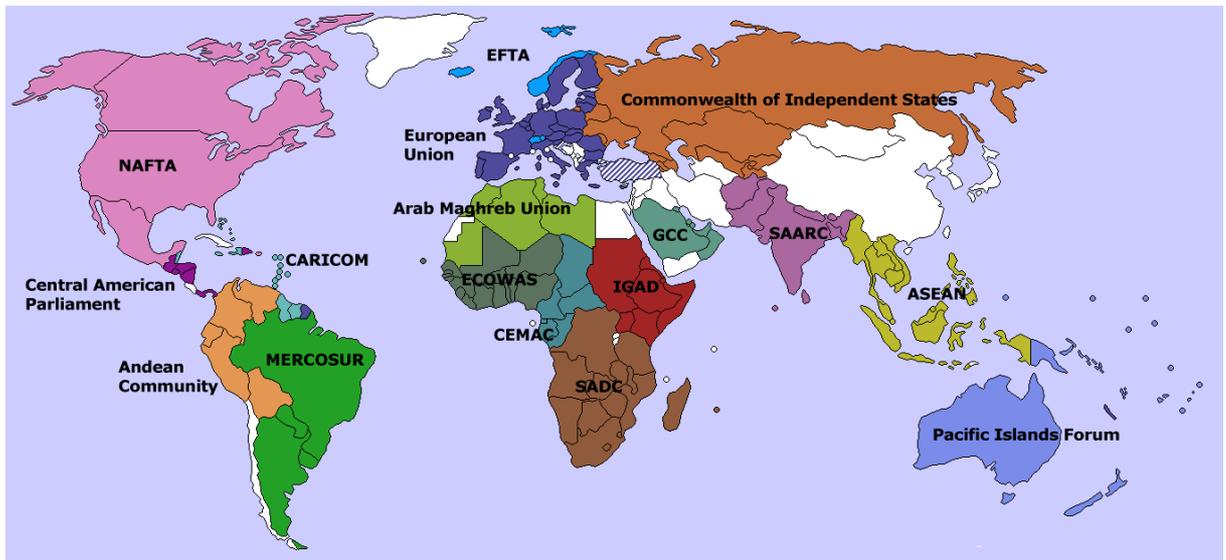
- Afrique
- Amérique Latine
- Asie
- Europe de l'Est
- Océanie

Auxquelles on a ajouté, pour les pays témoins de notre base :

- Amérique (du Nord)
- Europe

Organisation

Cette variable qualitative regroupe 16 organisations économiques internationales, présentées ci-après.



Andean Community

La communauté Andine est une association économique créée en 1996, dans la logique du Pacte Andin de 1969. Elle regroupe quatre pays : Bolivie, Colombie, Équateur et Pérou. Le Venezuela, auparavant membre de la communauté, s'est retiré en 2006 pour faire partie seulement du MERCOSUR.

Arab Maghreb Union

L'Arab Maghreb Union désigne l'organisation économique et politique créée en 1989 pour regrouper les pays du grand Maghreb : l'Algérie, Libye, le Maroc, la Mauritanie et la Tunisie.

ASEAN

L'Association des nations de l'Asie du Sud-Est (ASEAN) est une organisation politique, économique et culturelle regroupant dix pays d'Asie du Sud-Est. Elle a été fondée en 1967 dans le contexte de la guerre froide avec l'objectif de faire barrage aux mouvements communistes et développer la croissance dans la région. Aujourd'hui, l'association a pour but de renforcer la coopération et l'assistance mutuelle entre ses membres : Indonésie, Malaisie, Philippines, Singapour, Thaïlande, Brunei, Viêt Nam, Laos, Myanmar et Cambodge

CARICOM

CARICOM est l'acronyme anglais de la communauté caribéenne (Caribbean Community). Cette association, créée en 1973 par quatre pays de la zone caraïbe, compte aujourd'hui 20 membres : Antigua-et-Barbuda, Bahamas, Barbade, Belize, Dominique, Grenade, Guyana, Haïti, Jamaïque, Montserrat, Sainte-Lucie, Saint-Christophe-et-Niévès, Saint-Vincent-et-les Grenadines, Suriname, Trinité-et-Tobago, Îles Vierges britanniques, Îles Turques-et-Caïques, Anguilla, Îles Caïmans et Bermudes. La communauté comporte notamment un marché unique et un passeport commun.

CEMAC

La CEMAC (Communauté Économique et Monétaire de l'Afrique Centrale), créée en 1994, regroupe six États-membres partageant une monnaie commune, le franc CFA. Il s'agit du Cameroun, du Congo, du Gabon, de la Guinée Equatoriale, de la République Centrafricaine, et du Tchad.

Central American Parliament

Le Central American Parliament (Marché Commun Centraméricain) a été créé en 1960 et regroupe cinq pays d'Amérique Centrale : Costa Rica, Salvador, Guatemala, Honduras et Nicaragua. L'objectif principal de cette association est de créer une zone de libre-échange visant à contribuer au développement économique et commercial des pays.

Commonwealth of Independent States

Le Commonwealth of Independent States (en français Communauté des États indépendants) est une entité intergouvernementale composée de 11 anciennes républiques soviétiques : Biélorussie, Russie, Kazakhstan, Arménie, Ouzbékistan, Tadjikistan, Kirghizistan, Moldavie, Azerbaïdjan, Ukraine et Turkménistan.

ECOWAS

L'ECOWAS (Economic Community Of West African States) ou, en français, Communauté économique des États de l'Afrique de l'Ouest, est une organisation internationale régionale. Son but principal est de promouvoir la coopération et l'intégration avec pour objectif de créer une union économique et monétaire entre les pays de l'Afrique de l'Ouest. Créée en 1975 et compte aujourd'hui 15 membres : Bénin, Burkina Faso, Cap-Vert, Côte d'Ivoire, Gambie, Ghana, Guinée, Guinée-Bissau, Liberia, Mali, Niger, Nigeria, Sénégal, Sierra Leone et Togo.

GCC

Le "Gulf Cooperation Council" est une communauté politique et économique créée en 1981. Il est composé de six États membres : Arabie Saoudite, Bahreïn, Oman, Qatar, Emirats Arabes Unis et Koweït, auxquels s'ajoute un pays observateur mais non membre, le Yémen.

IGAD

L'IGAD (Intergovernmental Authority on Development) ou, en français, Autorité intergouvernementale pour le développement est une organisation économique régionale regroupant sept pays Est-Africains : Djibouti, Érythrée, Éthiopie, Kenya, Somalie, Soudan et Ouganda. Fondée en 1986 en réponse aux sécheresses et famines qui avaient dévasté la zone dans les années 1974 – 1984, la communauté a désormais pour objectif de contribuer au développement et à l'intégration économique de ses Etats membres.

MERCOSUR

Le MERCOSUR (Mercado Común del Sur) désigne la communauté économique des pays d'Amérique Latine. Le traité fondateur a été signé en 1991 par les quatre premiers membres : le Brésil, l'Argentine, le Paraguay et l'Uruguay, rejoints en 2006 par le Venezuela.

NAFTA

Le NAFTA (North American Free Trade Agreement) ou en français ALENA est un traité signé en 1994, créant une zone de libre échange entre les Etats Unis, le Canada et le Mexique.

Pacific Islands Forum

Le Forum des îles du Pacifique est une organisation politique internationale de coopération régionale créée en 1971. Il réunit les 16 pays indépendants et territoires associés de l'Océanie : Australie, Iles Cook, Iles Fidji, Kiribati, Iles Marshall, États fédérés de Micronésie, Nauru, Nouvelle-Zélande, Niue, Palaos, Papouasie-Nouvelle-Guinée, Salomon, Samoa, Tonga, Tuvalu et Vanuatu.

SAARC

La SAARC (South Asian Association for Regional Co-Operation), ou, en français, Association sud-asiatique pour la coopération régionale est une association fondée en 1983 et regroupant huit pays d'Asie du Sud : Afghanistan, Bangladesh, Bhoutan, Inde, Maldives, Népal, Pakistan et Sri Lanka. La SAARC encourage la coopération dans les domaines de l'agriculture, le développement rural, la science et les technologies, la culture, la santé, le contrôle de la natalité, ainsi que la lutte contre le narcotrafic et contre le terrorisme.

SADC

La SADC (Southern African Development Community) ou, en français, Communauté de Développement d'Afrique Australe est une organisation économique régionale visant au développement des pays d'Afrique Australe. Créée en 1992 pour remplacer la Conférence de coordination pour le développement de l'Afrique Australe, elle compte aujourd'hui 15 Etats membres : Angola, Botswana, République Démocratique du Congo, Lesotho, Madagascar, Malawi, Maurice, Mozambique, Namibie, Seychelles, Afrique du Sud, Swaziland, Tanzanie, Zambie et Zimbabwe.

Union Européenne

L'Union Européenne a été instituée par deux traités fondateurs : le traité de Rome (1957) et le traité de Maastricht (1992). Elle compte aujourd'hui vingt-sept Etats membres : Allemagne, Autriche, Belgique, Bulgarie, Chypre, Danemark, Espagne, Estonie, Finlande, France, Grèce, Hongrie, Irlande, Italie, Lettonie, Lituanie, Luxembourg, Malte, Pays-Bas, Pologne, Portugal, République tchèque, Roumanie, Royaume-Uni, Slovaquie, Slovénie et Suède. Trois pays sont actuellement candidats à l'intégration à l'Union Européenne : Ancienne République yougoslave de Macédoine, Croatie et Turquie.

2. Préparation des données

2.1 Analyse de la table de données manquantes

Lors de la construction de la base de données, la principale difficulté, outre la recherche elle-même, a été de se heurter à un nombre relativement important de données manquantes. Leur volume rend la construction d'un modèle de notation direct assez problématique.

Les deux solutions les plus simples proposées en analyse statistique pour traiter les données manquantes sont :

La suppression des observations comportant au moins une donnée non renseignée,

L'affectation d'une valeur estimée par la moyenne de la variable.

Aucune de ces solutions ne semble satisfaisante dans notre cas, la première nous faisant perdre trop d'opportunités potentielles, et la seconde donnant une interprétation faussée compte tenu du caractère très hétérogène de nos données.

On souhaite donc effectuer une analyse plus approfondie et tenter de remplacer ces données manquantes par des estimations que nous voulons les plus fines possibles.

Avant de construire le modèle de traitement des données manquantes, il convient de préparer les données et une première analyse d'information nous paraît intéressante : celle de la capacité (ou de la volonté) d'un pays à fournir une donnée économique ou assurantielle. L'objectif d'investissement implique en effet de sélectionner des pays qui semblent représenter une opportunité économique. D'une part, on ne pourra pas sélectionner un pays pour lequel on ne parvient à obtenir aucune information, d'autre part, le manque de transparence concernant les données économiques indique également que le pays est soit à un stade de développement économique trop peu avancé pour pouvoir fournir ce type de données, soit dans un système politique non favorable aux investissements étrangers.

On construit, à partir de notre base de données, une matrice binaire $[\delta_{ij}]$ où $\delta_{ij} = 1$ si la variable j est renseignée pour le pays i , et 0 sinon ($1 \leq i \leq 164$; $1 \leq j \leq 21$).

A partir de la table binaire construite, on souhaite obtenir deux types de résultats :

Une première analyse de la capacité de reporting économique des pays

Une base de travail par groupes de variables en fonction du nombre de données manquantes et des observations non renseignées, que nous allons utiliser pour déterminer l'ordre de modélisation des données manquantes par variable.

2.1.1. Analyse de la capacité de reporting par pays

On souhaite analyser la capacité de reporting économique des pays par Classification Ascendante Hiérarchique.

Pour cela, on construit tout d'abord une matrice de similarité. Comme nos données sont binaires, on va se baser sur la similarité calculée selon l'indice de Jaccard.

L'indice de Jaccard calcule un indice de similarité s_{ij} entre deux pays i et j selon la formule suivante :

$$s_{ij} = \frac{r_{ij}}{r_{i\bar{j}} + r_{\bar{i}j} + r_{ij}}$$

avec

$$\begin{aligned} r_{ij} &= \text{nombre de données renseignées pour } i \text{ et } j \\ r_{i\bar{j}} &= \text{nombre de données renseignées pour } i \text{ seulement} \\ r_{\bar{i}j} &= \text{nombre de données renseignées pour } j \text{ seulement} \\ r_{\bar{i}\bar{j}} &= \text{nombre de données renseignées ni pour } i \text{ ni pour } j \end{aligned}$$

L'indice de similarité s_{ij} prend des valeurs dans $[0 ; 1]$. On remarque que la somme des quatre éléments qui le composent vaut 1.

En d'autres termes, si l'on considère deux ensembles I et J tels que $I = \{\text{variables renseignées pour le pays } i\}$ et $J = \{\text{variables renseignées pour le pays } j\}$, l'indice de similarité de Jaccard correspond au ratio $\frac{I \cap J}{I \cup J}$.

On obtient la matrice de proximité, de laquelle on extrait la liste des pays considérés comme similaires, c'est-à-dire dont le seuil de similarité dans la matrice est supérieur à 0,95.

Pour des raisons de lisibilité, cette matrice sera présentée en annexe, sa taille correspondant au nombre de pays de la base : 164 x 164.

A partir de la matrice de similarité construite, on effectue un Classification Ascendante Hiérarchique (CAH).

La Classification Ascendante Hiérarchique consiste à regrouper entre eux des objets similaires selon tel ou tel critère, c'est-à-dire à répartir n individus, caractérisés par p variables X_1, X_2, \dots, X_p en un certain nombre de sous-groupes aussi homogènes que possible. Le nombre de sous-groupes choisi correspond à un niveau de précision donné : c'est le principe sous-jacent à une classification hiérarchique. Contrairement aux méthodes de classification non hiérarchiques (par partitionnement), le résultat d'une CAH n'est pas une partition de l'ensemble des individus.

C'est une hiérarchie de classes telles que :

- toute classe est non vide
- tout individu appartient à une (et même plusieurs) classes
- deux classes distinctes sont disjointes, ou vérifient une relation d'inclusion (l'une d'elles est incluse dans l'autre)
- toute classe est la réunion des classes qui sont incluses dans elle.

L'application de la Classification Ascendante Hiérarchique à la matrice de similarité des pays permet d'effectuer des regroupements de pays en fonction de leur nombre de données manquantes. A partir du découpage le plus fin (c'est-à-dire, chaque pays pris individuellement), on agrège progressivement les pays les plus ressemblants du point de vue du critère de similarité calculé précédemment avec l'indice de Jaccard, jusqu'à revenir à une unique classe contenant tous les éléments. Le regroupement des éléments en classes intermédiaires se fait selon la méthode de la liaison moyenne : la dissimilarité entre deux classes est la moyenne des dissimilarités entre les objets de chacune de ces deux classes.

Le dendrogramme (cf. Figure 1) permet de visualiser le regroupement progressif des pays en classes, en fonction du niveau de précision.

On sélectionne une troncature du dendrogramme avec un découpage en 8 classes de pays, ce qui correspond à une précision de 0,85. Ce découpage résulte d'un arbitrage entre la nécessité d'avoir un nombre suffisant de classes pour discriminer, et la décision d'éviter de construire une classe contenant un seul élément.

Au niveau de regroupement choisi, on calcule le barycentre de chaque classe, et la distance de chaque individu à son barycentre, afin de déterminer si le groupe est homogène ou relativement étalé. Ces données sont regroupées dans le Tableau 1.

Une interprétation économique de ce tableau nous laisse supposer que la classe 2 correspond à celle pour laquelle on a le plus de renseignements. Certaines associations, comme le Japon dans la classe 1 ou Taiwan dans la classe 3, semble assez étonnantes. Une fois la modélisation des données manquantes terminée, il sera intéressant de croiser la sélection des pays candidats à l'investissement avec ces classes.

2.1.2 Construction des paliers de modélisation des données manquantes

Afin de préparer les paliers de modélisation des données manquantes, on procède avec les variables de la même façon qu'avec les observations-pays précédemment. On construit dans un premier temps la matrice de proximité selon l'indice de similarité de Jaccard.

On effectue ensuite sur cette matrice une Classification Ascendante Hiérarchique qui va regrouper les variables par niveau de similarité en fonction de leurs données manquantes.

Les données manquantes par variable réparties selon le tableau présenté en annexe, où l'on retrouve également la matrice de proximité associée.

Le dendrogramme correspondant (Figure 2) permet des découpages en un nombre plus ou moins important de classes selon la précision souhaitée. Ici encore, la précision et le nombre de classes obtenus résultent du choix de limiter la construction de classes à une seule variable. On choisit de conserver une troncature du dendrogramme à une précision de 0,75, qui correspond à un découpage en 4 classes. Ce niveau de découpage nous semble suffisant compte tenu du nombre de variables à classer ensuite. Le choix d'une précision à 0,8 ou 0,85 aurait impliqué la construction d'une ou deux nouvelles classes constituées d'une seule variable (Economic Freedom Index et Vehicles per Capita). L'utilisation de la méthode de classification ascendante hiérarchique a été choisie afin de regrouper les variables par similitude, il ne nous semble pas opportun de choisir un niveau de précision qui nous contraint à les séparer par unité.

Tableau 1 : Regroupement des pays en classes

Classe	1	2	3	4	5	6	7	8	
Nombre d'éléments	21	81	34	13	7	4	2	2	
Variance intra-classe	0,233	0,13561976	0,263	0,307	0,129	0,000	0,000	0,000	
Distance minimale au barycentre	0,342	0,174506941	0,342	0,396	0,206	0,000	0,000	0,000	
Distance moyenne au barycentre	0,458	0,303704101	0,491	0,521	0,310	0,000	0,000	0,000	
Distance maximale au barycentre	0,668	0,918498392	0,789	0,704	0,488	0,000	0,000	0,000	
	Afghanistan	Afrique du Sud	Libye	Albanie	Antigua-et-Barbuda	Bénin	Bosnie-et-Herzégovine	Congo	Corée du Nord
	Barbade	Algérie	Macédoine	Arménie	Comores	Burundi	Éthiopie	Mali	Malte
	Belize	Allemagne	Madagascar	Azerbaïdjan	Dominique	Égypte	Kenya		
	Bhoutan	Angola	Malaisie	Belarus	Fidji	Jamaïque	Ouganda		
	Cambodge	Arabie saoudite	Malawi	Birmanie	Haïti	Liban			
	Cap-Vert	Argentine	Maroc	Bolivie	Maldives	Namibie			
	Djibouti	Australie	Maurice	Botswana	Mozambique	Turkménistan			
	Géorgie	Bahreïn	Mexique	Cuba	Ouzbékistan				
	Grenade	Bangladesh	Moldavie	Gambie	Philippines				
	Iran	Brésil	Népal	Ghana	Russie				
	Israël	Bulgarie	Nicaragua	Guinée	Rwanda				
	Japon	Burkina Faso	Niger	Îles Marshall	Somalie				
	Libéria	Cameroun	Nigeria	Inde	Tanzanie				
	Lituanie	Chili	Oman	Islande					
	Monténégro	Chine	Pakistan	Jordanie					
	Papouasie-Nouvelle-Guinée	Chypre	Palaos	Kazakhstan					
	République démocratique du Congo	Colombie	Pérou	Kirghizstan					
	Sri Lanka	Corée du Sud	Qatar	Mongolie					
	Togo	Costa Rica	Royaume-Uni	Panama					
	Vanuatu	Côte d'Ivoire	Saint-Christophe-et-Nevis	Paraguay					
	Yémen	Croatie	Sainte-Lucie	Pologne					
		Émirats arabes unis	Samoa	République centrafricaine					
		Équateur	São Tomé e Príncipe	République dominicaine					
		Espagne	Sénégal	République tchèque					
		Estonie	Serbie	Roumanie					
		Etats-Unis	Seychelles	Salvador					
		France	Sierra Leone	Syrie					
		Gabon	Slovaquie	Tadjikistan					
		Grèce	Slovénie	Taiwan					
		Guatémala	Soudan	Timor-Leste					
		Guinée-Bissau	Surinam	Ukraine					
		Guinée-Équatoriale	Swaziland	Uruguay					
		Guyana	Tchad	Vénézuéla					
		Honduras	Thaïlande	Zambie					
		Hongrie	Tonga						
		Îles Salomon	Trinité-et-Tobago						
		Indonésie	Tunisie						
		Irak	Turquie						
		Koweït	Vietnam						
		Laos	Zimbabwe						
		Lettonie							

Figure 2 : Dendrogramme – Regroupement progressif des variables classes

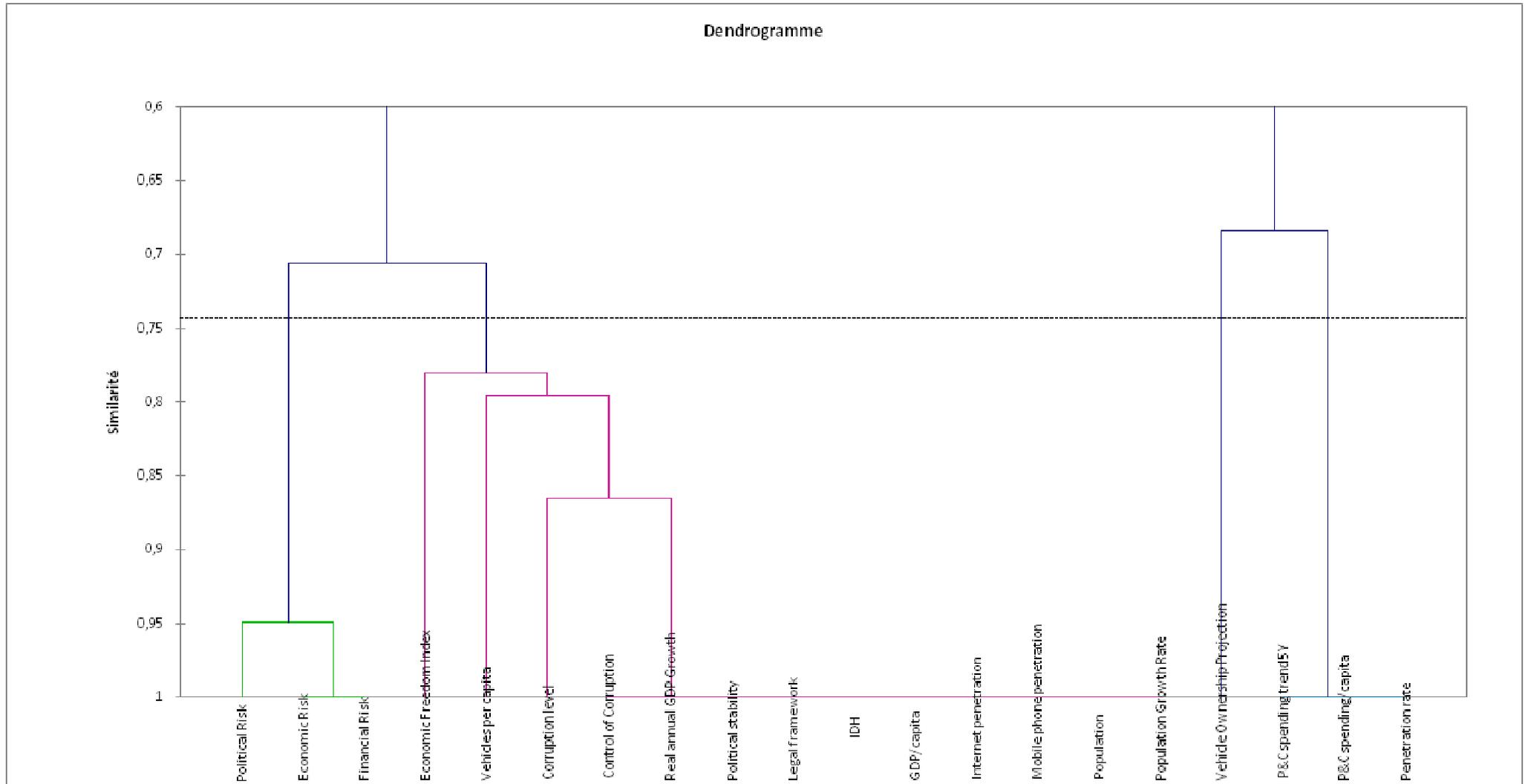


Tableau 2 : Regroupement des variables en classes

Classe	1	2	3	4
Objets	13	3	1	3
Variance intra-classe	0,083	0,008	0,000	0,000
Distance minimale au barycentre	0,133	0,051	0,000	0,000
Distance moyenne au barycentre	0,220	0,068	0,000	0,000
Distance maximale au barycentre	0,607	0,102	0,000	0,000
	Population	Economic Risk Financial Risk	Vehicle Ownership Projection	P&C spending/capita
	Population Growth Rate	Political Risk		Penetration rate
	Mobile phone penetration			P&C spending trend
	Internet penetration			
	GDP/capita			
	IDH			
	Legal framework			
	Political stability			
	Real annual GDP Growth			
	Control of Corruption			
	Corruption level			
	Economic Freedom Index			
	Vehicles per capita			

La classe 1 contient les dix variables pour lesquelles la base de données est complète. On visualise ces variables sur la matrice de proximité par une sous-matrice $\begin{pmatrix} 1 & \dots & 1 \\ \dots & 1 & \dots \\ 1 & \dots & 1 \end{pmatrix}$ construite avec les 10 premières lignes et 10 premières colonnes. Les trois autres variables contenues dans la classe 1 sont Corruption Level, Economic Freedom Index, et Vehicles per capita qui vont représenter notre premier palier de modélisation. Elles correspondent également aux variables pour lesquelles les nombres de données manquantes sont les moins élevés.

La classe 2 contient les trois variables de Risques économique, financier et politique issus de l'International Country Risk Guide. Les données manquantes sont les mêmes pour ces trois variables, il s'agit des pays non pris en compte par l'étude.

La classe 3 contient uniquement la variable Vehicle Ownership Projection, dont la répartition des données manquantes est sensiblement différente, puisqu'elle provient d'un travail de recherche à part.

Enfin, la classe 4 contient les données assurantielles de l'étude Sigma de Swiss Re. Elles comportent le plus de données manquantes et seront donc modélisées en dernier.

2.2 Modélisation de l'appartenance à une Organisation Economique

2.2.1 Méthodologie employée

La variable qualitative Organisation présentée plus haut comporte des données manquantes, simplement liées au fait que tous les pays de notre base n'appartiennent pas à l'une des 17 organisations économiques choisies. Il y a 28 données manquantes sur 164, et l'on souhaite réduire ce nombre afin d'optimiser les chances de modélisation des données manquantes.

La répartition des modalités de la variable Organisation est présentée en annexe.

On souhaite analyser les caractéristiques qui regroupent les éléments dans chacune de ces organisations en fonction des dix variables pour lesquelles la base de données est complète :

- Population
- GDP/capita
- Mobile phone Penetration
- Internet Penetration
- IDH
- Population Growth Rate
- Real Annual GDP Growth
- Legal Framework
- Political Stability
- Control of Corruption

On ajoute à cet ensemble de variables quantitatives la variable qualitative Zone, complète, qui donne une indication logique sur l'appartenance d'un pays à une organisation économique. En effet, comme l'indique la carte des organisations économiques présentée lors de la définition des variables, l'appartenance de deux pays à une même organisation économique implique l'appartenance à une même zone géographique.

On obtient les statistiques descriptives par variable, présentées en annexe.

Afin d'expliquer l'appartenance des pays à chaque modalité de la variable Organisation, sur la base des variables ci-dessus, on utilise un arbre de classification, qui nous permettra également de construire une forme de prédiction sur les pays pour lesquels la variable Organisation n'est pas renseignée.

On construit pour cela une matrice de corrélation entre les variables, présentée en annexe.

On voit que les variables les plus corrélées sont Legal Framework, Internet Penetration, Mobile phone penetration et Control of Corruption. D'un point de vue économique, cela signifie que la qualité de l'environnement politico-légal et du contrôle de la corruption sont fortement liés au niveau de développement des moyens de communication.

2.2.2 Construction de l'arbre de classification

On utilise la méthode CHAID (CHi-squared Automatic Interaction Detection) développée par Kass (1980) à partir de la méthode AID (Automatic Interaction Detection) de Morgan et Sonquist (1963).

La méthode CHAID nous permet de mettre en évidence des groupes dont les caractéristiques seront relativement homogènes, et à les comparer à la variable à expliquer Organisation. On cherche pour cela à déterminer les critères discriminants.

On procède à une première étape de séparation : à partir du nœud initial qui comprend la totalité des observations-pays, la meilleure variable de séparation est celle pour laquelle la p-value est la plus petite, en se basant sur le test du Khi-deux de Pearson, donné par la formule :

$$\chi^2 = \sum \frac{(f_t - f_0)^2}{f_t} = \sum \frac{(\text{Résidus})^2}{(\text{Effectif théorique})^2}$$

où $f_t = \text{effectif théorique}$
et $f_0 = \text{effectif observé}$

Pour chacune des variables explicatives, on teste son degré d'interaction avec la variable Organisation. On construit ainsi une ségrégation de plus en plus fine de la base, dont les critères de séparation successifs sont décrits dans l'arbre de classification.

L'algorithme proposé par Kass se décompose en 5 étapes.

Pour chaque variable explicative, on croise ses catégories (valeurs ou modalités) avec celles de la variable à expliquer, et on procède aux étapes 2 et 3.

On détermine les deux catégories qui sont les moins significativement différentes (selon le teste du Khi-deux) et si la p-value reste sous un certain seuil, on accepte de fusionner ces deux catégories.

Pour toutes les nouvelles catégories composées, constituées de 3 ou plus des catégories originales, on cherche la séparation binaire la plus significative. Si le niveau de significativité dépasse le seuil critique, on implémente cette segmentation et on retourne à l'étape 2.

Calculer le niveau de significativité de chaque composition de catégories de variables explicatives (« optimisée »), déterminer la plus significative et construire la segmentation selon cet ensemble de critères.

Pour les autres partitions qui n'ont pas encore été étudiées, reprendre à l'étape 1.

L'étape 1 nous a fait sélectionner la partition selon la variable Zone, puis, pour chaque catégorie (modalité) de Zone, on construit une segmentation.

On obtient l'algorithme de construction d'un arbre de classification donné sous la forme d'un ensemble de règles booléennes. La classification, ainsi que l'ensemble des règles, est présentée en annexe.

2.2.3 Résultats et Interprétations

L'arbre de classification affecte a posteriori une organisation par pays en fonction de la ségrégation construite. Le tableau correspondant, présenté en annexe, permet d'analyser pour quelles observations on observe une erreur de choix d'organisation, et indique pour les pays où l'organisation était non renseignée une éventuelle proposition de remplacement.

La matrice de confusion présentée en annexe nous indique le nombre de données mal classifiées par organisation économique.

Elle indique notamment que 5 organisations sur les 17 ne comportent plus aucune valeur après l'application de l'arbre. Ces organisations contiennent entre 2 et 7 éléments, dont les caractéristiques ne sont pas significativement différentes de ceux d'une autre organisation économique, dans laquelle elle va être « incluse » : l'utilisation du test du Khi-deux a en effet tendance à favoriser les classes contenant le plus d'éléments. On trouve une application de ce cas avec par exemple l'ensemble des pays de l'Andean Community et du Central American Parliament, du NAFTA et du MERCOSUR intégrés à CARICOM, celui-ci

contenant 11 éléments a priori et 32 a posteriori. L'IGAD est partagé entre Arab Maghreb Union, CEMAC et ECOWAS.

Au final, on décide de modifier les valeurs qui étaient a priori non renseignées par la valeur proposée a posteriori. Les pays supplémentaires dont l'organisation est alors renseignée sont :

Égypte	NC	→	Arab Maghreb Union
Seychelles	NC	→	Arab Maghreb Union
Iran	NC	→	ASEAN
Chili	NC	→	CARICOM
Costa Rica	NC	→	CARICOM
Cuba	NC	→	CARICOM
Corée du Nord	NC	→	Commonwealth of Independent States
Géorgie	NC	→	Commonwealth of Independent States
Burundi	NC	→	ECOWAS
Israël	NC	→	GCC
Mongolie	NC	→	SAARC

On remarque que la zone géographique de chaque pays a bien été respectée.

Même si elles ne représentent pas la situation véritable actuelle, notamment pour des raisons politiques, ces affectations basées sur l'analyse des composantes de ces pays montrent qu'ils présentent une cohérence économique avec leurs voisins.

On va utiliser ces données complétées dans la modélisation des données manquantes.

3. Consolidation des données manquantes

3.1. Choix de la Modélisation par ANCOVA

3.1.1. Définition et Principe de l'ANCOVA

La base d'indices économiques et assurantiels que nous avons constituée comporte des données manquantes, non disponibles. On choisit d'estimer les données manquantes à l'aide d'un modèle de régression linéaire, qui semble le plus applicable à des données économiques.

Comme nous souhaitons effectuer une régression à partir de plusieurs variables explicatives, nous allons utiliser un modèle de régression multiple, dont le modèle matriciel de base est :

$$Y = X \times a + \varepsilon$$

$$\text{où } Y = \begin{pmatrix} y_1 \\ \dots \\ y_t \\ \dots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \dots & x_{k1} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1t} & \dots & x_{kt} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{kn} \end{pmatrix}, a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_k \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_t \\ \dots \\ \varepsilon_n \end{pmatrix}$$

On veut estimer le vecteur a (le premier coefficient a_0 correspond au terme constant). On applique pour cela la méthode des Moindres Carrés Ordinaires, qui se base sur la minimisation de la Somme des Carrés des Erreurs.

Afin de valider le modèle obtenu, et d'accepter le remplacement des données manquantes par leur valeur estimée, on souhaite vérifier que le modèle construit a bien un sens, et mesurer la qualité de la représentation de la variable à expliquer par les variables explicatives. On effectue donc sur le modèle une analyse de variance.

Equation fondamentale d'analyse de la variance :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n \varepsilon_i^2$$

La variabilité totale (SCT) est égale à la variabilité expliquée (SCE) + la variabilité des résidus (SCR).

Quatre situations principales peuvent être rencontrées, pour lesquelles on utilisera une technique d'Analyse de Variance différente :

Avec une seule variable dépendante (à expliquer), et une ou plusieurs variables explicatives discrètes (qualitatives) on utilise l'analyse de variance simple ou ANOVA. L'analyse consiste à tester si les différences de variation dans chaque groupe (ou échantillon) défini par les modalités des variables explicatives s'écartent de manière significative de la valeur 0.

Avec toujours une seule variable à expliquer, mais une ou plusieurs variables explicatives, toutes continues, on fait appel au modèle de régression linéaire simple ou multiple : La régression consiste à évaluer les paramètres d'une fonction linéaire, à estimer à l'aide de cette fonction les valeurs de la variable à expliquer et à tester si les résidus (écarts entre valeurs mesurées et valeurs estimées) sont significativement proches de 0.

Lorsque certaines des variables explicatives sont discrètes (qualitatives), et d'autres continues, on réalise une analyse de la covariance ou ANCOVA. On cherche si les relations entre la variable à expliquer et les variables explicatives continues restent les mêmes pour toutes les modalités des variables explicatives discrètes.

Lorsqu'il y a plusieurs variables à expliquer à prendre en compte simultanément, au lieu de faire plusieurs analyses de variance (une par variable à expliquer), on peut réaliser une analyse de variance multiple. On teste si les différences de variation dans chaque groupe défini par les modalités des variables explicatives s'écartent de manière significative de la 0, et cela de manière simultanée dans l'ensemble des variables.

Dans le cas de notre base de données, nous avons bien des variables explicatives quantitatives et qualitatives. De plus, on souhaite modéliser les variables une par une afin de pouvoir utiliser, si la modélisation est suffisamment fine, les données manquantes estimées pour une variable lors de la modélisation de la variable suivante. La méthode la plus appropriée dans cette optique semble donc l'ANCOVA.

Modèle de régression de l'ANCOVA :

Soit p le nombre de variables quantitatives et q le nombre de variables qualitatives. On veut obtenir:

$$y_i = \beta_0 + \sum_{j=1}^{j=p} \beta_j x_{ij} + \sum_{j=1}^{j=q} \beta_{k(i,j),j} + \varepsilon_i$$

Où y_i représente la variable à expliquer, β_0 la constante, β_j le coefficient de la variable quantitative x_{ij} , $\beta_{k(i,j),j}$ le coefficient affecté à la modalité $k(i,j)$ de la variable qualitative j ($\beta_{k(i,j),j'} = 0$ si $j \neq j'$), et ε_i représente l'erreur de modélisation.

Les étapes de modélisation sont décrites au paragraphe suivant, en application de la modélisation de la première variable contenant des données manquantes : Corruption Level.

3.1.2. Variables explicatives initiales

On dispose d'une base d'information complète pour :

- Dix variables quantitatives :
 - Population
 - GDP/capita
 - Mobile phone Penetration
 - Internet Penetration
 - IDH
 - Population Growth Rate
 - Real Annual GDP Growth
 - Legal Framework
 - Political Stability
 - Control of Corruption
- Deux variables qualitatives : la zone géographique du pays, et l'organisation économique à laquelle il appartient, dont on souhaite également analyser l'impact.

3.2 Application de l'ANCOVA

3.2.1. Données à consolider

Ce premier palier a un nombre de données à compléter relativement faible, ce qui simplifie la validation a posteriori et permettra, si l'estimation semble suffisamment fine, d'inclure ces données complémentaires dans la table de variables explicatives de l'analyse ANCOVA pour la modélisation des paliers suivants.

Le tableau ci-dessous récapitule l'ensemble des données manquantes sur le premier palier:

	Corruption Level	Economic Freedom Index	Vehicles Per Capita
Données manquantes	7	7	41

3.2.2. Statistiques descriptives

3.2.2.1. Variables quantitatives

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Population	157	72514,000	1360445010,000	41979075,172	147529607,581
Population Growth Rate	157	-0,300	3,690	1,434	0,963
Mobile phone penetration	157	0,000	2,423	0,864	0,572
Internet penetration	157	0,001	0,893	0,224	0,226
GDP/capita	157	112,070	66016,000	8163,629	12567,989
IDH	157	2,000	1012,000	116,815	133,759
Real annual GDP Growth	157	-8,500	21,400	4,838	3,282
Legal framework	157	0,000	96,433	43,381	25,409
Political stability	157	0,000	95,200	41,261	25,548
Control of Corruption	157	0,000	98,500	42,835	26,493

On observe que les échelles de valeur des observations sont très hétérogènes. Néanmoins l'application de la modélisation par ANCOVA ne nécessite pas de centrer et réduire les variables, car les écarts d'échelles seront compensés par les valeurs des coefficients.

3.2.2.2. Variables qualitatives

Variable	Modalités	Effectifs	%	
Zone	Afrique	50	31,847	
	Amérique	1	0,637	
	Amérique Latine	28	17,834	
	Asie	29	18,471	
	Europe	6	3,822	
	Europe de l'Est	21	13,376	
	Moyen Orient	16	10,191	
	Océanie	6	3,822	
	Organisation	ASEAN	10	6,369
		Andean Community	5	3,185
Arab Maghreb Union		6	3,822	
CARICOM		10	6,369	
CEMAC		6	3,822	
Central American Parliament		7	4,459	
Commonwealth of Independent States		12	7,643	
ECOWAS		16	10,191	
GCC		7	4,459	
IGAD		6	3,822	
MERCOSUR		5	3,185	
NAFTA		2	1,274	
NC		17	10,828	
Pacific Islands		6	3,822	
SAARC		9	5,732	
SADC		13	8,280	
Union Européenne		20	12,739	

3.2.3 Matrice de corrélation

Le tableau ci-après indique les corrélations entre les variables à expliquer (en bleu) et les variables explicatives quantitatives (en noir) et qualitatives (en vert) de la modélisation.

Variables	Population	Population Growth Rate	Mobile phone penetration	Internet penetration	GDP/capita	IDH	Real annual GDP Growth	Legal framework	Political stability	Control of Corruption	Zone-Afrique	Zone-Amérique	Zone-Amérique Latine	Zone-Asie	Zone-Europe	Zone-Europe de l'Est	Zone-Moyen Orient	Zone-Océanie	Organisation-ASEAN	Organisation-Andean Community	Organisation-Arab Maghreb Union	Organisation-CARICOM	Organisation-CEMAC	Organisation-Central American Parliament	Organisation-Commonwealth of Independent States	Organisation-ECOWAS	Organisation-GCC	Organisation-IGAD	Organisation-MERCOSUR	Organisation-NAFTA	Organisation-NC	Organisation-Pacific Islands	Organisation-SAARC	Organisation-SADC	Organisation-Union Européenne
Population	1,000	-0,059	-0,058	0,011	-0,018	0,020	0,089	0,059	-0,124	0,000	-0,105	0,144	-0,068	0,311	0,003	-0,084	0,065	-0,051	0,040	-0,021	-0,021	-0,068	-0,048	-0,051	-0,037	-0,053	-0,052	-0,011	0,011	0,127	0,149	-0,051	0,228	-0,046	-0,060
Population Growth Rate	-0,059	1,000	-0,581	-0,569	-0,263	0,339	0,177	-0,518	-0,396	-0,452	0,517	-0,038	-0,112	0,133	-0,235	-0,452	0,156	0,045	-0,017	0,008	0,003	-0,140	0,157	0,019	-0,219	0,374	0,197	0,235	-0,070	-0,045	-0,004	0,045	-0,011	0,083	-0,447
Mobile phone penetration	-0,058	-0,581	1,000	0,787	0,650	0,436	-0,187	0,755	0,535	0,646	-0,530	0,068	0,129	0,105	0,296	0,419	0,176	-0,103	-0,106	0,023	0,029	0,127	-0,186	0,036	0,062	-0,295	0,311	-0,233	0,037	0,050	0,005	-0,103	-0,145	-0,220	0,500
Internet penetration	0,011	-0,569	0,787	1,000	0,729	0,407	-0,269	0,794	0,541	0,726	-0,475	0,177	0,092	0,071	0,394	0,343	0,055	-0,049	-0,053	0,008	-0,011	0,154	-0,174	-0,063	-0,114	-0,272	0,155	-0,156	0,050	0,131	0,109	-0,049	-0,171	-0,216	0,553
GDP/capita	-0,018	-0,263	0,650	0,729	1,000	0,347	-0,140	0,669	0,500	0,646	-0,327	0,230	-0,095	0,130	0,569	0,115	0,248	-0,012	-0,119	-0,044	-0,047	-0,020	-0,058	-0,096	-0,103	-0,201	0,465	-0,117	-0,028	0,172	-0,036	-0,012	-0,137	-0,135	0,493
IDH	-0,020	0,339	-0,436	-0,407	-0,347	1,000	-0,140	-0,475	-0,393	-0,447	0,324	-0,063	-0,122	0,034	-0,153	-0,178	0,028	-0,024	-0,005	-0,048	-0,033	-0,076	0,041	-0,021	-0,044	0,123	-0,124	0,271	-0,070	-0,072	0,060	-0,024	0,028	0,203	-0,231
Real annual GDP Growth	0,089	0,177	-0,187	-0,269	-0,140	0,140	1,000	-0,169	-0,001	-0,108	0,021	-0,091	-0,135	0,209	-0,216	-0,042	0,123	-0,071	0,069	-0,007	-0,038	-0,171	0,007	-0,039	0,182	0,024	0,163	-0,034	0,031	-0,118	0,014	-0,071	0,123	0,035	-0,192
Legal framework	0,059	-0,518	0,755	0,794	0,669	0,475	-0,169	1,000	0,689	0,919	-0,373	0,156	0,020	0,134	0,347	0,286	0,126	0,019	-0,114	-0,119	-0,005	0,158	-0,242	-0,034	-0,156	-0,227	0,232	-0,164	-0,029	0,130	0,023	0,019	-0,092	-0,046	0,516
Political stability	-0,124	-0,396	0,535	0,541	0,500	0,393	-0,001	0,689	1,000	0,726	-0,189	0,085	-0,001	0,178	0,214	0,290	0,087	0,193	-0,156	-0,188	0,017	0,160	-0,152	-0,033	-0,063	-0,124	0,104	-0,223	0,026	0,023	-0,073	0,193	-0,170	0,081	0,407
Control of Corruption	0,000	-0,452	0,646	0,726	0,646	0,447	-0,108	0,919	0,726	1,000	-0,300	0,148	0,075	0,213	0,332	0,227	0,099	0,060	-0,175	-0,062	0,006	0,170	-0,245	-0,066	-0,233	-0,185	0,258	-0,179	0,056	0,120	0,032	0,060	-0,085	-0,010	0,445
Zone-Afrique	-0,105	0,517	-0,530	-0,475	-0,327	0,324	0,021	-0,373	-0,189	-0,300	1,000	-0,055	-0,318	0,325	-0,136	-0,269	0,230	-0,136	-0,178	-0,124	0,292	-0,178	0,292	-0,148	-0,197	0,493	-0,148	0,292	-0,124	-0,078	-0,106	-0,136	-0,169	0,440	-0,261
Zone-Amérique	0,144	-0,038	0,068	0,177	0,230	0,063	-0,091	0,156	0,085	0,148	-0,055	1,000	-0,037	0,038	-0,016	-0,031	0,027	-0,016	-0,021	-0,015	-0,016	-0,021	-0,016	-0,017	-0,023	-0,027	-0,017	-0,016	-0,015	0,705	-0,028	-0,016	-0,020	-0,024	-0,031
Zone-Amérique Latine	-0,068	-0,112	0,129	0,092	-0,095	0,122	-0,135	0,020	-0,001	0,075	-0,318	-0,037	1,000	0,222	-0,093	-0,183	0,157	-0,093	-0,122	0,389	-0,093	0,560	-0,093	0,464	-0,134	-0,157	-0,101	-0,093	0,389	0,095	-0,162	-0,093	-0,115	-0,140	-0,178
Zone-Asie	0,311	-0,133	-0,105	-0,071	-0,130	0,034	0,209	-0,134	-0,178	-0,213	-0,325	-0,038	-0,222	1,000	-0,095	-0,187	0,160	-0,095	0,548	-0,086	-0,095	-0,124	-0,095	-0,103	0,234	-0,160	-0,103	-0,095	-0,086	-0,054	0,045	-0,095	0,518	-0,143	-0,182
Zone-Europe	0,003	-0,235	0,296	0,394	0,569	0,153	-0,216	0,347	0,214	0,332	-0,136	-0,016	-0,093	0,095	1,000	-0,078	0,067	-0,040	-0,052	-0,036	-0,040	-0,052	-0,040	-0,043	-0,057	-0,067	-0,043	-0,040	-0,036	-0,023	-0,069	-0,040	-0,049	-0,060	0,522
Zone-Europe de l'Est	-0,084	-0,452	0,419	0,343	0,115	0,178	-0,042	0,286	0,290	0,227	-0,269	-0,031	-0,183	0,187	-0,078	1,000	0,132	-0,078	-0,102	-0,071	-0,078	-0,102	-0,078	-0,085	0,098	-0,132	-0,085	-0,078	-0,071	-0,045	0,104	-0,078	-0,097	-0,118	0,636
Zone-Moyen Orient	-0,065	0,156	0,176	0,055	0,248	0,028	0,123	0,126	-0,087	0,099	-0,230	-0,027	-0,157	0,160	-0,067	-0,132	1,000	-0,067	-0,088	-0,061	-0,067	-0,088	-0,067	-0,073	0,141	-0,113	0,641	-0,067	-0,061	-0,038	0,289	-0,067	-0,083	-0,101	-0,129
Zone-Océanie	-0,051	0,045	-0,103	-0,049	-0,012	0,024	-0,071	0,019	0,193	0,060	-0,136	-0,016	-0,093	0,095	-0,040	-0,078	0,067	1,000	-0,052	-0,036	-0,040	-0,052	-0,040	-0,043	-0,057	-0,067	-0,043	-0,040	-0,036	-0,023	-0,069	1,000	-0,049	-0,060	-0,076

Organisation-ASEAN	0,040	-0,017	-0,106	-0,053	-0,119	0,005	0,069	-0,114	-0,156	-0,175	-0,178	-0,021	-0,122	0,548	-	-	-	1,000	-0,047	-0,052	-0,068	-0,052	-0,056	-0,075	-0,088	-0,056	-0,052	-0,047	-0,030	-0,091	-0,052	-0,064	-0,078	-0,100	
Organisation-Andean Community	-0,021	0,008	0,023	0,008	-0,044	0,048	-0,007	-0,119	-0,188	-0,062	-0,124	-0,015	0,389	0,086	-0,036	-0,071	0,061	-0,036	-0,047	1,000	-0,036	-0,047	-0,036	-0,039	-0,052	-0,061	-0,039	-0,036	-0,033	-0,021	-0,063	-0,036	-0,045	-0,054	-0,069
Organisation-Arab Maghreb Union	-0,021	0,003	0,029	-0,011	-0,047	0,033	-0,038	-0,005	0,017	0,006	0,292	-0,016	-0,093	0,095	-0,040	-0,078	0,067	-0,040	-0,052	-0,036	1,000	-0,052	-0,040	-0,043	-0,057	-0,067	-0,043	-0,040	-0,036	-0,023	-0,069	-0,040	-0,049	-0,060	-0,076
Organisation-CARICOM	-0,068	-0,140	0,127	0,154	-0,020	0,076	-0,171	0,158	0,160	0,170	-0,178	-0,021	0,560	0,124	-0,052	-0,102	0,088	-0,052	-0,068	-0,047	-0,052	1,000	-0,052	-0,056	-0,075	-0,088	-0,056	-0,052	-0,047	-0,030	-0,091	-0,052	-0,064	-0,078	-0,100
Organisation-CEMAC	-0,048	0,157	-0,186	-0,174	-0,058	0,041	0,007	-0,242	-0,152	-0,245	0,292	-0,016	-0,093	0,095	-0,040	-0,078	0,067	-0,040	-0,052	-0,036	-0,040	-0,052	1,000	-0,043	-0,057	-0,067	-0,043	-0,040	-0,036	-0,023	-0,069	-0,040	-0,049	-0,060	-0,076
Organisation-Central American Parliament	-0,051	0,019	0,036	-0,063	-0,096	0,021	-0,039	-0,034	-0,033	-0,066	-0,148	-0,017	0,464	0,103	-0,043	-0,085	0,073	-0,043	-0,056	-0,039	-0,043	-0,056	-0,043	1,000	-0,062	-0,073	-0,047	-0,043	-0,039	-0,025	-0,075	-0,043	-0,053	-0,065	-0,083
Organisation-Commonwealth of Independent States	-0,037	-0,219	0,062	-0,114	-0,103	0,044	0,182	-0,156	-0,063	-0,233	-0,197	-0,023	-0,134	0,234	-0,057	0,098	0,141	-0,057	-0,075	-0,052	-0,057	-0,075	-0,057	-0,062	1,000	-0,097	-0,062	-0,057	-0,052	-0,033	-0,100	-0,057	-0,071	-0,086	-0,110
Organisation-ECOWAS	-0,053	0,374	-0,295	-0,272	-0,201	0,123	0,024	-0,227	-0,124	-0,185	0,493	-0,027	-0,157	0,160	-0,067	-0,132	0,113	-0,067	-0,088	-0,061	-0,067	-0,088	-0,067	-0,073	-0,097	1,000	-0,073	-0,067	-0,061	-0,038	-0,117	-0,067	-0,083	-0,101	-0,129
Organisation-GCC	-0,052	0,197	0,311	0,155	0,465	0,124	0,163	0,232	0,104	0,258	-0,148	-0,017	-0,101	0,103	-0,043	-0,085	0,641	-0,043	-0,056	-0,039	-0,043	-0,056	-0,043	-0,047	-0,062	-0,073	1,000	-0,043	-0,039	-0,025	-0,075	-0,043	-0,053	-0,065	-0,083
Organisation-IGAD	-0,011	0,235	-0,233	-0,156	-0,117	0,271	-0,034	-0,164	-0,223	-0,179	0,292	-0,016	-0,093	0,095	-0,040	-0,078	0,067	-0,040	-0,052	-0,036	-0,040	-0,052	-0,040	-0,043	-0,057	-0,067	-0,043	1,000	-0,036	-0,023	-0,069	-0,040	-0,049	-0,060	-0,076
Organisation-MERCOSUR	0,011	-0,070	0,037	0,050	-0,028	0,070	0,031	-0,029	0,026	0,056	-0,124	-0,015	0,389	0,086	-0,036	-0,071	0,061	-0,036	-0,047	-0,033	-0,036	-0,047	-0,036	-0,039	-0,052	-0,061	-0,039	-0,036	1,000	-0,021	-0,063	-0,036	-0,045	-0,054	-0,069
Organisation-NAFTA	0,127	-0,045	0,050	0,131	0,172	0,072	-0,118	0,130	0,023	0,120	-0,078	0,705	0,095	0,054	-0,023	-0,045	0,038	-0,023	-0,030	-0,021	-0,023	-0,030	-0,023	-0,025	-0,033	-0,038	-0,025	-0,023	-0,021	1,000	-0,040	-0,023	-0,028	-0,034	-0,043
Organisation-NC	0,149	-0,004	0,005	0,109	-0,036	0,050	0,014	0,023	-0,073	0,032	-0,106	-0,028	-0,162	0,045	-0,069	0,104	0,289	-0,069	-0,091	-0,063	-0,069	-0,091	-0,069	-0,075	-0,100	-0,117	-0,075	-0,069	-0,063	-0,040	1,000	-0,069	-0,086	-0,105	-0,133
Organisation-Pacific Islands	-0,051	0,045	-0,103	-0,049	-0,012	0,024	-0,071	0,019	0,193	0,060	-0,136	-0,016	-0,093	0,095	-0,040	-0,078	0,067	1,000	-0,052	-0,036	-0,040	-0,052	-0,040	-0,043	-0,057	-0,067	-0,043	-0,040	-0,036	-0,023	-0,069	1,000	-0,049	-0,060	-0,076
Organisation-SAARC	0,228	-0,011	-0,145	-0,171	-0,137	0,028	0,123	-0,092	-0,170	-0,085	-0,169	-0,020	-0,115	0,518	-0,049	-0,097	0,083	-0,049	-0,064	-0,045	-0,049	-0,064	-0,049	-0,053	-0,071	-0,083	-0,053	-0,049	-0,045	-0,028	-0,086	-0,049	1,000	-0,074	-0,094
Organisation-SADC	-0,046	0,083	-0,220	-0,216	-0,135	0,203	0,035	-0,046	0,081	-0,010	0,440	-0,024	-0,140	0,143	-0,060	-0,118	0,101	-0,060	-0,078	-0,054	-0,060	-0,078	-0,060	-0,065	-0,086	-0,101	-0,065	-0,060	-0,054	-0,034	-0,105	-0,060	-0,074	1,000	-0,115
Organisation-Union Européenne	-0,060	-0,447	0,500	0,553	0,493	0,231	-0,192	0,516	0,407	0,445	-0,261	-0,031	-0,178	0,182	0,522	0,636	0,129	-0,076	-0,100	-0,069	-0,076	-0,100	-0,076	-0,083	-0,110	-0,129	-0,083	-0,076	-0,069	-0,043	-0,133	-0,076	-0,094	-0,115	1,000

3.3 Consolidation des données pour le palier 1

3.3.1. Variable « Corruption Level »

3.3.1.1 Synthèse de sélection des variables

A l'aide du logiciel XLSTAT, on effectue des simulations de modélisation, afin de sélectionner le nombre optimal de variables pour la modélisation. Le critère de décision est la valeur du R^2 ajusté.

Le R^2 (ou coefficient de détermination) donne le pourcentage de variabilité de la variable à modéliser (« Corruption Level ») expliquée par les variables quantitatives et qualitative explicatives.

Formule du R^2 :

$$R^2 = \frac{\sum_{i=1}^{i=n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2} \text{ avec } \bar{y} = \sum_{i=1}^{i=n} \frac{y_i}{n}$$

Le R^2 ajusté a la même fonction que le R^2 , mais prend en plus en compte le nombre de variables incluses dans le modèle. Il est calculé selon la formule suivante :

$$R^2 \text{ ajusté} = 1 - (1 - R^2) \frac{n-1}{n-p-1} \text{ avec } p \text{ le nombre de variables explicatives de la modélisation et } n \text{ la taille de l'échantillon.}$$

Le processus de sélection du meilleur modèle consiste à tester une à une l'ensemble des combinaisons de 1 à n variables (dans ce premier cas, $n = 12$). L'algorithme construit pour chaque combinaison la régression ANCOVA et calcule le R^2 et le R^2 ajusté associés. Si le R^2 est une fonction croissante du nombre de variables, on a vu que le R^2 ajusté en revanche se stabilise et éventuellement décroît à partir du moment où les variables ajoutées n'apportent plus aucune information supplémentaire au modèle. En choisissant le modèle ANCOVA aillant le R^2 ajusté le plus élevé, on optimise ainsi le rapport nombre de variables employées/niveau d'explication.

Pour obtenir ce modèle optimisé, on passe par un algorithme qui devra calculer :

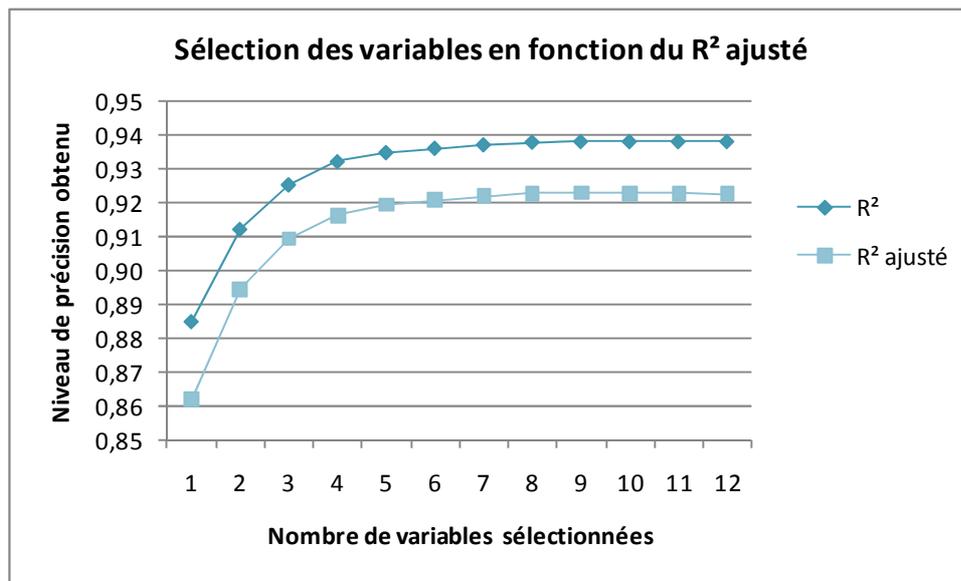
$\sum_{k=1}^{n=12} \binom{n}{k} = 2^n - 1 = 4095$ combinaisons, et pour chacune d'elles faire l'ensemble des opérations de modélisation ANCOVA. On voit ainsi que ce choix d'optimisation demande rapidement un nombre très élevé d'opérations, et qu'il sera envisageable de scinder le nombre de combinaisons à prendre en compte dans les modélisations des prochaines variables afin de réduire le temps de calcul.

La synthèse de la sélection des variables par le processus décrit donne les résultats suivants :

Nb. de variables	Variables	MCE	R ²	R ² ajusté	AIC de Akaike	SBC de Schwarz
1	Control of Corruption	0,332	0,885	0,862	-171,325	-165,213
2	Internet penetration / Control of Corruption	0,255	0,912	0,894	-211,482	-202,313
3	Internet penetration / Control of Corruption / Zone	0,228	0,925	0,909	-222,707	-192,144
4	Internet penetration / Control of Corruption / Zone / Organisation	0,229	0,932	0,916	-209,696	-136,346
5	Internet penetration / GDP/capita / Control of Corruption / Zone / Organisation	0,221	0,935	0,919	-213,950	-137,544
6	Internet penetration / GDP/capita / Legal framework / Control of Corruption / Zone / Organisation	0,219	0,936	0,921	-215,014	-135,551
7	Internet penetration / GDP/capita / Legal framework / Political stability / Control of Corruption / Zone / Organisation	0,217	0,937	0,922	-215,497	-132,978
8	Internet penetration / GDP/capita / Real annual GDP Growth / Legal framework / Political stability / Control of Corruption / Zone / Organisation	0,216	0,938	0,923	-215,507	-129,932
9	Population Growth Rate / Internet penetration / GDP/capita / Real annual GDP Growth / Legal framework / Political stability / Control of Corruption / Zone / Organisation	0,217	0,938	0,923	-213,901	-125,270
10	Population Growth Rate / Mobile phone penetration / Internet penetration / GDP/capita / Real annual GDP Growth / Legal framework / Political stability / Control of Corruption / Zone / Organisation	0,219	0,938	0,923	-211,909	-120,222
11	Population Growth Rate / Mobile phone penetration / Internet penetration / GDP/capita / IDH / Real annual GDP Growth / Legal framework / Political stability / Control of Corruption / Zone / Organisation	0,220	0,938	0,923	-209,917	-115,173
12	Population / Population Growth Rate / Mobile phone penetration / Internet penetration / GDP/capita / IDH / Real annual GDP Growth / Legal framework / Political stability / Control of Corruption / Zone / Organisation	0,222	0,938	0,922	-207,920	-110,120

Le tableau ci-dessus donne un R^2 à 93,8%. Compte tenu de l'hétérogénéité et du caractère « économique » de nos données, on considère que ce taux d'explication est très satisfaisant.

Graphiquement, on visualise que le nombre de variables sélectionnées atteint son optimum au point où la tangente au R^2 ajusté devient nulle. Compte tenu de sa construction, le R^2 ajusté se remet en effet à décroître légèrement après avoir atteint cet optimum, alors que le R^2 est une fonction strictement croissante du nombre de variables.



L'algorithme de sélection du meilleur modèle fournit également deux autres critères de sélection :

Le Critère d'Information d'Akaike (AIC), calculé selon la formule :

$$AIC = \ln\left(\frac{SCR}{n}\right) + \frac{2p}{n}$$

Avec n le nombre d'observations et p le nombre de variables explicatives du modèle.

Le Critère de Schwarz (BIC), calculé selon la formule :

$$SBC = \ln\left(\frac{SCR}{n}\right) + \frac{p \ln(n)}{n}$$

On remarque que selon le critère de minimisation de l'AIC, on choisirait le modèle à 8 variables. On éliminerait alors la variable explicative « Population Growth Rate » dont l'ajout au modèle de prédiction n'apporte que peu d'information supplémentaire. On visualise sur la courbe du R^2 que l'apport marginal d'information du passage de 8 à 9 variables est limité. On conserve néanmoins la variable dans le processus de modélisation, son « élimination » sera confirmée par un coefficient très proche de zéro dans l'équation linéaire de modélisation.

Le Critère de Schwarz est quant à lui minimisé pour un modèle avec seulement deux variables : Internet Penetration et Control of Corruption. Ces deux variables sont celles qui ont le plus fort impact sur la variable à expliquer Corruption Level. Economiquement parlant, cela sous-entend qu'un niveau de contrôle et lutte contre la corruption élevé soit synonyme d'un niveau de corruption effectif plus faible. De même, un accès généralisé de la population aux technologies de l'information (Internet Penetration) semble avoir une valeur explicative forte (par rapport aux autres variables explicatives) sur un niveau de corruption modéré. Ces deux liens de causalité nous semblent assez logiques.

3.3.1.2. Tableau d'Analyse de la Variance

Le tableau ci-dessous permet de tester si l'on peut considérer que les variables explicatives sélectionnées apportent une quantité d'information significative au modèle.

Ainsi on fait le test d'hypothèses :

$H_0 : \beta_j = \beta_{k(i,j),j} = 0 \forall (i, j, k)$ (tous les coefficients sont nuls)

H_1 : Au moins un coefficient est non nul.

On utilise pour cela le test de Fisher.

Le cas où l'hypothèse H_0 serait acceptée signifie qu'il n'y a aucune relation linéaire significative entre les variables explicatives et la variable à expliquer.

On construit le tableau d'analyse de la variance qui nous permet d'effectuer le test de Fisher.

Source	DDL	Somme des carrés	Moyenne des carrés	F*	p-value
Modèle (x_1, \dots, x_k)	k = 28	SCE = 419,442	SCE/k = 14,980	69,017	< 0,0001
Erreur	n-k-1 = 128	SCR = 27,782	SCR/(n-k-1) = 0,217		
Total corrigé	n-1 = 156	SCT = 447,224			

Le variable aléatoire F suit une loi de Fisher à k et (n-k-1) degrés de liberté, sous réserve que les résidus soient gaussiens. Sous l'hypothèse H_0 , F^* est la réalisation de F. Sous l'hypothèse H_1 , F^* devient grande.

La valeur F^* est calculée selon la formule :

$$F^* = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

Le test est construit à partir d'un seuil fixé de la façon suivante :

Soit α le niveau du test (pourcentage de rejet de H_0 à tort). On définit $f(\alpha)$ satisfaisant l'équation : $f(\alpha) = P(F > f(\alpha))$. On rejette alors H_0 lorsque $F^* > f(\alpha)$.

La colonne p-value indique la p-value associée au test de Fisher. Cela correspond la probabilité que la valeur F^* calculée dans la colonne précédente soit supérieure à une variable de Fisher, ou encore au plus petit α pour lequel on ne rejette pas l'hypothèse nulle.

La p-value associée au test de Fisher est ici inférieure à 0,0001. Cela signifie que le risque d'erreur sur la validation de la significativité des variables choisies est inférieur à 0,01%. On peut conclure à la validation de notre modélisation.

3.3.1.3. Tableaux d'analyse de significativité des variables

Le tableau d'analyse de significativité par intégration progressive permet, pour chaque variable, d'évaluer son impact lorsqu'elle est ajoutée aux variables précédentes. Les valeurs du tableau dépendent donc de l'ordre de la sélection des variables. On fait ici encore appel au F de Fisher pour tester la significativité de l'impact d'une variable sur la qualité du modèle.

Source	DDL	Somme des carrés	Moyenne des carrés	F*	p-value
Population	0	0,000			
Population Growth Rate	1	91,792	91,792	422,910	< 0,0001
Mobile phone penetration	0	0,000			
Internet penetration	1	191,470	191,470	882,154	< 0,0001
GDP/capita	1	19,837	19,837	91,395	< 0,0001
IDH	0	0,000			
Real annual GDP Growth	1	0,091	0,091	0,420	0,518
Legal framework	1	78,569	78,569	361,990	< 0,0001
Political stability	1	6,150	6,150	28,335	< 0,0001
Control of Corruption	1	24,732	24,732	113,949	< 0,0001
Zone	7	3,957	0,565	2,604	0,015
Organisation	14	2,843	0,203	0,936	0,523

Plus la p-value associée au F de Fisher est faible, plus fort sera l'impact de la variable ajoutée. On voit ici que les variables Population Growth Rate, Internet penetration, GDP/capita, Legal Framework, Political Stability et Control of Corruption ont un très fort impact sur le modèle. On peut considérer que la variable qualitative Zone apporte une information supplémentaire au modèle avec un taux d'erreur de 1,5%. Les variables Population, Mobile Phone Penetration, IDH, Real annual GDP Growth ainsi que la variable qualitative Organisation n'ont qu'un impact très limité sur la qualité d'explication de la variable Corruption Level. Selon l'analyse du tableau de significativité par intégration progressive, seule le niveau de significativité marginal est pris en compte, en fonction des variables ajoutées une par une. On peut donc supposer que la variable Mobile Phone Penetration, très corrélée avec Internet Penetration (coefficient de corrélation à 0,787) n'apporte pas suffisamment d'information supplémentaire.

Le tableau de significativité par élimination progressive fournit quant à lui une évaluation de la contribution de chaque variable au modèle en testant l'effet de sa suppression du modèle. Contrairement au tableau précédent, celui-ci ne dépend donc pas de l'ordre dans lequel les variables ont été sélectionnées. On mesure la significativité de chaque variable avec le test de Fisher.

Source	DDL	Somme des carrés	Moyenne des carrés	F*	p-value
Population	0	0,000			
Population Growth Rate	1	0,065	0,065	0,298	0,586
Mobile phone penetration	0	0,000			
Internet penetration	1	2,951	2,951	13,594	0,000
GDP/capita	1	0,774	0,774	3,566	0,061
IDH	0	0,000			
Real annual GDP Growth	1	0,361	0,361	1,662	0,200
Legal framework	1	0,359	0,359	1,655	0,201
Political stability	1	0,585	0,585	2,696	0,103
Control of Corruption	1	17,053	17,053	78,569	< 0,0001
Zone	5	1,439	0,288	1,326	0,257
Organisation	14	2,730	0,195	0,898	0,562

Ce tableau nous indique que la variable qui a le plus fort impact sur le modèle est Control of Corruption, ce qui semble logique, suivie de Internet Penetration. On remarque également que toutes les autres variables dont la probabilité associée au F de Fischer était <0,0001 sur le tableau précédent, et semblaient donc avoir un impact significatif sur l'explication du modèle, ont cette fois-ci une probabilité associée au F relativement élevée, qui laisse penser qu'elles n'apportent pas vraiment d'information supplémentaire au modèle. Le niveau d'explicativité marginal de ces variables est donc relativement faible, il vient surtout « conforter » les informations apportées par Internet Penetration et Control of Corruption.

3.3.1.4. Paramètres du modèle

Le tableau suivant fournit les détails de la modélisation, et va nous permettre d'appliquer le modèle de régression linéaire trouvé aux données manquantes. On utilise cette fois la p-value associée au test de Student afin de mesurer la valeur du coefficient de chaque variable.

On effectue, pour chaque variable, le test d'hypothèses :

$H_0 : \beta_j = 0$ (ou $\beta_{k(i,j),j} = 0$ pour les variables qualitatives)

H_1 : le coefficient est différent de 0.

$$\frac{\hat{\beta}_j}{\sigma(\hat{\beta}_j)} = t_{\hat{\beta}_j}$$

Le ratio de Student $t_{\hat{\beta}_j}$ (t empirique) suit une loi de Student à (n-k-1) degrés de liberté, sous réserve que les résidus soient gaussiens. On étudie la probabilité que $t_{\hat{\beta}_j}$ soit supérieur au t de Student théorique $t_{n-k-1}^{\alpha/2}$, au seuil $\alpha = 5\%$. Comme il s'agit d'un test bilatéral, on calcule le t empirique en valeur absolue.

Lorsque la p-value prend une valeur supérieure à 5%, on peut conclure à la non-significativité de la variable dans le modèle. Cette non-significativité peut notamment être due à :

- une absence de corrélation avec la variable à expliquer,
- une colinéarité trop élevée avec une autre variable explicative.

Le tableau ci-dessous indique les résultats obtenus pour la variable Corruption Level.

Source	Valeur	Ecart-type	t	p-value	Borne inférieure (95%)	Borne supérieure (95%)
Constante	0,780	0,282	2,761	0,007	0,221	1,339
Population	0,000	0,000				
Population Growth Rate	0,039	0,069	0,567	0,572	-0,098	0,176
Mobile phone penetration	0,000	0,000				
Internet penetration	1,510	0,400	3,779	0,000	0,719	2,300
GDP/capita	0,000	0,000	1,886	0,062	0,000	0,000
IDH	0,000	0,000				
Real annual GDP Growth	-0,018	0,014	-1,350	0,179	-0,045	0,008
Legal framework	0,007	0,005	1,440	0,152	-0,003	0,016
Political stability	0,004	0,003	1,660	0,099	-0,001	0,010
Control of Corruption	0,042	0,005	9,322	< 0,0001	0,033	0,051
Zone-Afrique	-0,297	0,412	-0,720	0,473	-1,112	0,519
Zone-Amérique	-0,021	0,560	-0,037	0,971	-1,130	1,088
Zone-Amérique Latine	-0,268	0,517	-0,519	0,605	-1,291	0,755
Zone-Asie	0,033	0,361	0,091	0,928	-0,681	0,747
Zone-Europe	0,133	0,366	0,363	0,717	-0,591	0,858
Zone-Europe de l'Est	-0,344	0,259	-1,329	0,186	-0,856	0,168
Zone-Moyen Orient	0,161	0,351	0,459	0,647	-0,534	0,857
Zone-Océanie	0,000	0,000				
Organisation-ASEAN	0,116	0,334	0,348	0,729	-0,546	0,778
Organisation-Andean Community	0,272	0,518	0,525	0,600	-0,753	1,298
Organisation-Arab Maghreb Union	0,370	0,416	0,889	0,376	-0,454	1,194
Organisation-CARICOM	0,342	0,498	0,687	0,493	-0,643	1,326
Organisation-CEMAC	0,898	0,423	2,120	0,036	0,060	1,735
Organisation-Central American Parliament	0,343	0,505	0,679	0,498	-0,657	1,343
Organisation-Commonwealth of Independent States	0,144	0,278	0,518	0,605	-0,407	0,695
Organisation-ECOWAS	0,666	0,382	1,740	0,084	-0,091	1,422
Organisation-GCC	-0,535	0,358	-1,493	0,138	-1,244	0,174
Organisation-IGAD	0,543	0,417	1,302	0,195	-0,282	1,368
Organisation-MERCOSUR	0,405	0,521	0,776	0,439	-0,627	1,436
Organisation-NAFTA	0,000	0,000				
Organisation-NC	0,063	0,243	0,260	0,796	-0,417	0,543
Organisation-Pacific Islands	0,000	0,000				
Organisation-SAARC	0,301	0,349	0,863	0,390	-0,390	0,993
Organisation-SADC	0,568	0,390	1,455	0,148	-0,205	1,340
Organisation-Union Européenne	0,000	0,000				

Les variables Internet Penetration et Control of Corruption, sont les seules dont on peut valider avec certitude la significativité dans le modèle, ainsi que la présence d'une constante positive.

Equation du modèle linéaire et valeurs des coefficients

L'équation linéaire obtenue se présente sous la forme suivante :

$$Corruption\ Level = \alpha_0 + \sum_{i=1}^{28} \alpha_i X_i$$

Dans le tableau ci-dessous, on retrouve que les deux variables quantitatives les plus impactantes sont Internet Penetration et Control of Corruption. On visualise par ailleurs que les zones Europe de l'Est, Afrique et Amérique Latine obtiennent des coefficients négatifs, que l'on peut assimiler en moyenne à une moins bonne notation en terme de niveau de corruption.

Nom	Variable associée	Coefficient	Valeur
-	-	α_0	0,7801
Population Growth Rate	X_1	α_1	0,0392
Internet penetration	X_2	α_2	1,5096
GDP/capita	X_3	α_3	0,0000
Real annual GDP Growth	X_4	α_4	- 0,0182
Legal framework	X_5	α_5	0,0069
Political stability	X_6	α_6	0,0044
Control of Corruption	X_7	α_7	0,0422
Zone-Afrique	X_8	α_8	- 0,2967
Zone-Amérique	X_9	α_9	- 0,0207
Zone-Amérique Latine	X_{10}	α_{10}	- 0,2684
Zone-Asie	X_{11}	α_{11}	0,0328
Zone-Europe	X_{12}	α_{12}	0,1331
Zone-Europe de l'Est	X_{13}	α_{13}	- 0,3438
Zone-Moyen Orient	X_{14}	α_{14}	0,1613
Organisation-ASEAN	X_{15}	α_{15}	0,1162
Organisation-Andean Community	X_{16}	α_{16}	0,2723
Organisation-Arab Maghreb Union	X_{17}	α_{17}	0,3701
Organisation-CARICOM	X_{18}	α_{18}	0,3418
Organisation-CEMAC	X_{19}	α_{19}	0,8976
Organisation-Central American Parliament	X_{20}	α_{20}	0,3433
Organisation-Commonwealth of Independent States	X_{21}	α_{21}	0,1442
Organisation-ECOWAS	X_{22}	α_{22}	0,6656
Organisation-GCC	X_{23}	α_{23}	- 0,5352
Organisation-IGAD	X_{24}	α_{24}	0,5429
Organisation-MERCOSUR	X_{25}	α_{25}	0,4045
Organisation-NC	X_{26}	α_{26}	0,0630
Organisation-SAARC	X_{27}	α_{27}	0,3015
Organisation-SADC	X_{28}	α_{28}	0,5678

3.3.1.5. Estimations

Le tableau suivant indique pour chaque observation-pays la valeur réelle de la variable Corruption, ainsi que la valeur modélisée par ANCOVA (Prédiction).

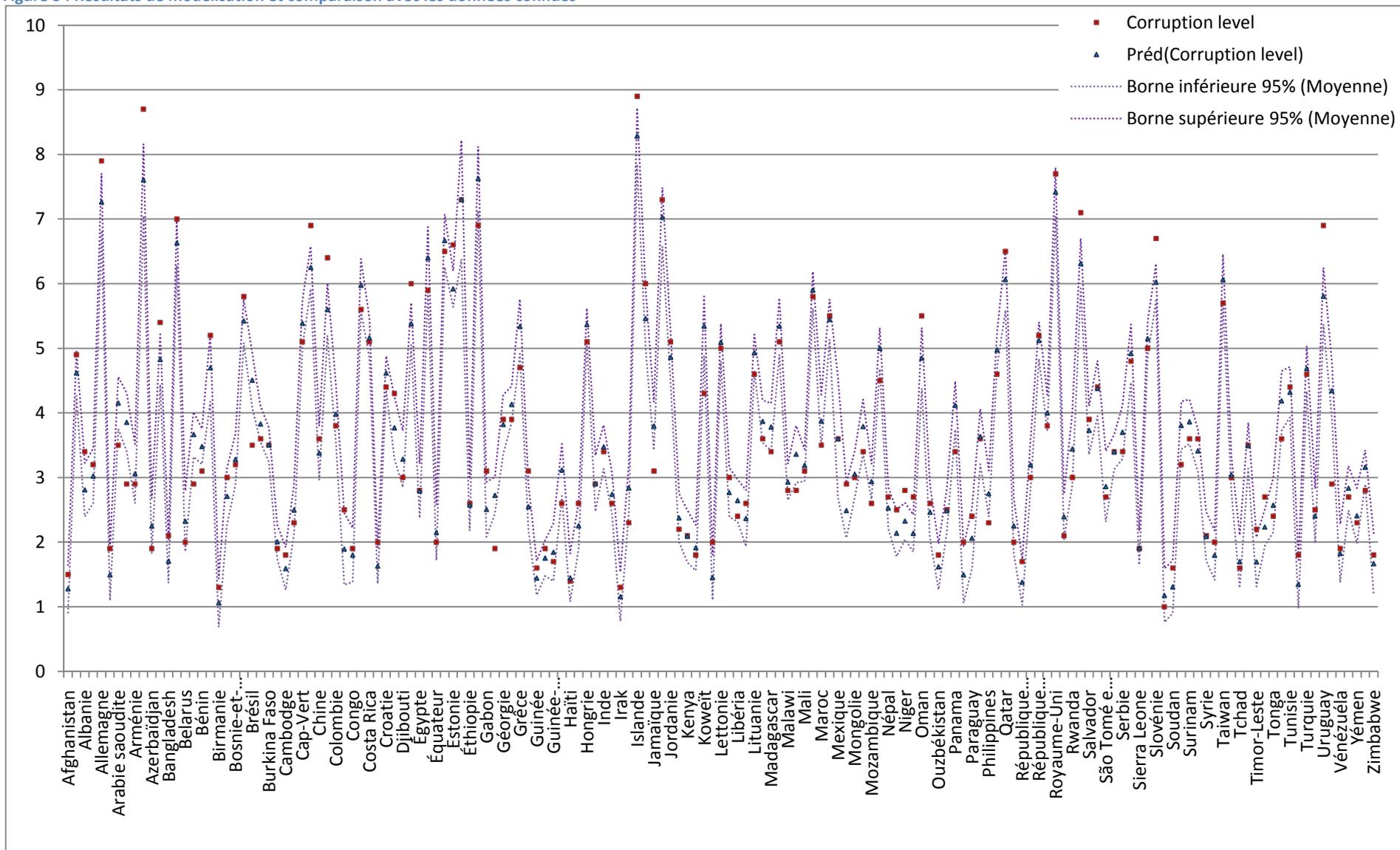
Observation	Corruption level	Prédiction	Résidu	Résidu std.
Afghanistan	1,500	1,280	0,220	0,472
Afrique du Sud	4,900	4,619	0,281	0,604
Albanie	3,400	2,809	0,591	1,269
Algérie	3,200	3,024	0,176	0,377
Allemagne	7,900	7,266	0,634	1,362
Angola	1,900	1,497	0,403	0,866
Arabie saoudite	3,500	4,154	-0,654	-1,404
Argentine	2,900	3,855	-0,955	-2,050
Arménie	2,900	3,060	-0,160	-0,343
Australie	8,700	7,607	1,093	2,347
Azerbaïdjan	1,900	2,251	-0,351	-0,753
Bahreïn	5,400	4,827	0,573	1,229
Bangladesh	2,100	1,706	0,394	0,846
Barbade	7,000	6,630	0,370	0,794
Belarus	2,000	2,325	-0,325	-0,697
Bélize	2,900	3,667	-0,767	-1,647
Bénin	3,100	3,479	-0,379	-0,813
Bhoutan	5,200	4,699	0,501	1,076
Birmanie	1,300	1,061	0,239	0,513
Bolivie	3,000	2,710	0,290	0,622
Bosnie-et-Herzégovine	3,200	3,281	-0,081	-0,174
Botswana	5,800	5,423	0,377	0,810
Brésil	3,500	4,506	-1,006	-2,159
Bulgarie	3,600	3,831	-0,231	-0,495
Burkina Faso	3,500	3,504	-0,004	-0,008
Burundi	1,900	2,006	-0,106	-0,227
Cambodge	1,800	1,591	0,209	0,449
Cameroun	2,300	2,498	-0,198	-0,426
Cap-Vert	5,100	5,387	-0,287	-0,617
Chili	6,900	6,250	0,650	1,394
Chine	3,600	3,380	0,220	0,471
Chypre	6,400	5,602	0,798	1,713
Colombie	3,800	3,982	-0,182	-0,391
Comores	2,500	1,894	0,606	1,300
Congo	1,900	1,801	0,099	0,211
Corée du Sud	5,600	5,977	-0,377	-0,810
Costa Rica	5,100	5,162	-0,062	-0,133
Côte d'Ivoire	2,000	1,632	0,368	0,790
Croatie	4,400	4,616	-0,216	-0,465
Cuba	4,300	3,770	0,530	1,137
Djibouti	3,000	3,286	-0,286	-0,614
Dominique	6,000	5,381	0,619	1,330
Égypte	2,800	2,792	0,008	0,016
Émirats arabes unis	5,900	6,399	-0,499	-1,072
Équateur	2,000	2,149	-0,149	-0,320
Espagne	6,500	6,667	-0,167	-0,358
Estonie	6,600	5,920	0,680	1,460
Etats-Unis	7,300	7,300	0,000	0,000
Éthiopie	2,600	2,575	0,025	0,053
France	6,900	7,626	-0,726	-1,559
Gabon	3,100	2,507	0,593	1,272
Gambie	1,900	2,723	-0,823	-1,767
Géorgie	3,900	3,821	0,079	0,170
Ghana	3,900	4,133	-0,233	-0,500

Observation	Corruption level	Prédiction	Résidu	Résidu std.
Grèce	4,700	5,338	-0,638	-1,369
Guatemala	3,100	2,547	0,553	1,186
Guinée	1,600	1,445	0,155	0,332
Guinée-Bissau	1,900	1,754	0,146	0,314
Guinée-Équatoriale	1,700	1,844	-0,144	-0,309
Guyana	2,600	3,114	-0,514	-1,103
Haïti	1,400	1,449	-0,049	-0,105
Honduras	2,600	2,255	0,345	0,741
Hongrie	5,100	5,369	-0,269	-0,577
Îles Salomon	2,900	2,906	-0,006	-0,013
Inde	3,400	3,476	-0,076	-0,162
Indonésie	2,600	2,741	-0,141	-0,302
Irak	1,300	1,156	0,144	0,309
Iran	2,300	2,840	-0,540	-1,159
Islande	8,900	8,285	0,615	1,319
Israël	6,000	5,462	0,538	1,155
Jamaïque	3,100	3,798	-0,698	-1,498
Japon	7,300	7,030	0,270	0,579
Jordanie	5,100	4,861	0,239	0,513
Kazakhstan	2,200	2,375	-0,175	-0,376
Kenya	2,100	2,090	0,010	0,021
Kirghizstan	1,800	1,914	-0,114	-0,244
Koweït	4,300	5,344	-1,044	-2,240
Laos	2,000	1,452	0,548	1,177
Lettonie	5,000	5,091	-0,091	-0,195
Liban	3,000	2,769	0,231	0,496
Libéria	2,400	2,644	-0,244	-0,523
Libye	2,600	2,368	0,232	0,498
Lituanie	4,600	4,932	-0,332	-0,712
Macédoine	3,600	3,868	-0,268	-0,576
Madagascar	3,400	3,782	-0,382	-0,820
Malaisie	5,100	5,342	-0,242	-0,518
Malawi	2,800	2,931	-0,131	-0,281
Maldives	2,800	3,362	-0,562	-1,206
Mali	3,100	3,189	-0,089	-0,190
Malte	5,800	5,901	-0,101	-0,217
Maroc	3,500	3,875	-0,375	-0,805
Maurice	5,500	5,445	0,055	0,119
Mexique	3,600	3,600	0,000	0,000
Moldavie	2,900	2,490	0,410	0,881
Mongolie	3,000	3,053	-0,053	-0,114
Monténégro	3,400	3,790	-0,390	-0,837
Mozambique	2,600	2,938	-0,338	-0,725
Namibie	4,500	5,001	-0,501	-1,074
Népal	2,700	2,531	0,169	0,362
Nicaragua	2,500	2,139	0,361	0,775
Niger	2,800	2,326	0,474	1,018
Nigeria	2,700	2,137	0,563	1,208
Oman	5,500	4,848	0,652	1,399
Ouganda	2,600	2,463	0,137	0,293
Ouzbékistan	1,800	1,620	0,180	0,387
Pakistan	2,500	2,487	0,013	0,029
Panama	3,400	4,117	-0,717	-1,539
Papouasie-Nouvelle-Guinée	2,000	1,494	0,506	1,086
Paraguay	2,400	2,061	0,339	0,728
Pérou	3,600	3,633	-0,033	-0,070
Philippines	2,300	2,752	-0,452	-0,970
Pologne	4,600	4,975	-0,375	-0,804
Qatar	6,500	6,065	0,435	0,933
République centrafricaine	2,000	2,251	-0,251	-0,539

Observation	Corruption level	Prédiction	Résidu	Résidu std.
République démocratique du Congo	1,700	1,380	0,320	0,687
République dominicaine	3,000	3,196	-0,196	-0,421
République tchèque	5,200	5,121	0,079	0,169
Roumanie	3,800	4,001	-0,201	-0,430
Royaume-Uni	7,700	7,418	0,282	0,605
Russie	2,100	2,394	-0,294	-0,630
Rwanda	3,000	3,443	-0,443	-0,951
Sainte-Lucie	7,100	6,312	0,788	1,692
Salvador	3,900	3,732	0,168	0,362
Samoa occidentales	4,400	4,378	0,022	0,047
São Tomé e Príncipe	2,700	2,863	-0,163	-0,349
Sénégal	3,400	3,396	0,004	0,008
Serbie	3,400	3,700	-0,300	-0,645
Seychelles	4,800	4,921	-0,121	-0,259
Sierra Leone	1,900	1,911	-0,011	-0,024
Slovaquie	5,000	5,150	-0,150	-0,322
Slovénie	6,700	6,023	0,677	1,453
Somalie	1,000	1,177	-0,177	-0,380
Soudan	1,600	1,308	0,292	0,627
Sri Lanka	3,200	3,806	-0,606	-1,302
Surinam	3,600	3,863	-0,263	-0,564
Swaziland	3,600	3,414	0,186	0,399
Syrie	2,100	2,089	0,011	0,025
Tadjikistan	2,000	1,797	0,203	0,435
Taiwan	5,700	6,063	-0,363	-0,779
Tanzanie	3,000	3,043	-0,043	-0,092
Tchad	1,600	1,698	-0,098	-0,211
Thaïlande	3,500	3,494	0,006	0,014
Timor-Leste	2,200	1,693	0,507	1,089
Togo	2,700	2,235	0,465	0,999
Tonga	2,400	2,571	-0,171	-0,367
Trinité-et-Tobago	3,600	4,188	-0,588	-1,263
Tunisie	4,400	4,319	0,081	0,173
Turkménistan	1,800	1,349	0,451	0,967
Turquie	4,600	4,686	-0,086	-0,184
Ukraine	2,500	2,405	0,095	0,203
Uruguay	6,900	5,808	1,092	2,344
Vanuatu	2,900	4,344	-1,444	-3,100
Vénézuéla	1,900	1,825	0,075	0,160
Vietnam	2,700	2,836	-0,136	-0,292
Yémen	2,300	2,408	-0,108	-0,232
Zambie	2,800	3,161	-0,361	-0,774
Zimbabwe	1,800	1,668	0,132	0,283

Le graphique ci-après (Figure 3) permet de visualiser les différents niveaux observés et modélisés de la variable Corruption Level pour chaque observation pays. On construit sur la base de la prédiction un intervalle de confiance : une borne supérieure et une borne inférieure

Figure 3 : Résultats de modélisation et comparaison avec les données connues



3.3.1.6. Tests de non violation des hypothèses sous-jacentes à la modélisation par ANCOVA

La modélisation par ANCOVA suppose, comme une régression linéaire classique, la non-violation de certaines hypothèses :

- sur la forme du modèle :

(H1) linéarité du modèle et choix des régresseurs

- sur les erreurs :

(H2) $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$ ont une moyenne nulle, c'est-à-dire $E(\varepsilon_t) = 0 \quad \forall t = 1, \dots, T$

(H3) homoscédasticité : $Var(\varepsilon_t) = \sigma^2 \quad \forall t = 1, \dots, T$

(H4) non autocorrélation : $Cov(\varepsilon_t, \varepsilon_{t'}) = 0 \quad \forall t \neq t'$

(H5) normalité : les erreurs sont distribuées suivant une loi normale.

- sur les variables explicatives :

(H6) les variables $X_1, X_2, X_3, \dots, X_p$ sont non aléatoires.

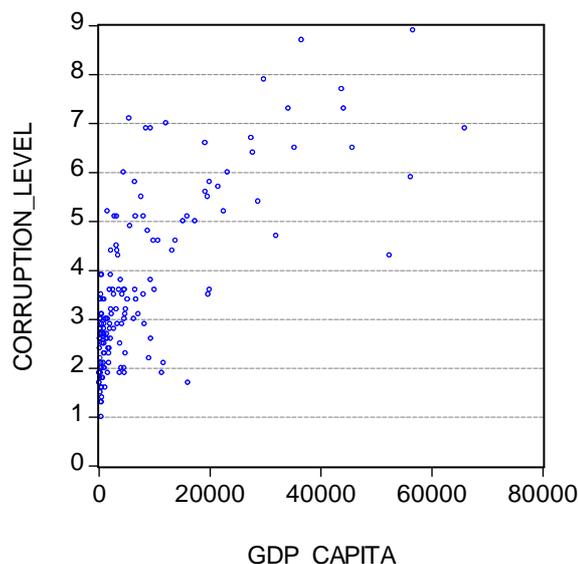
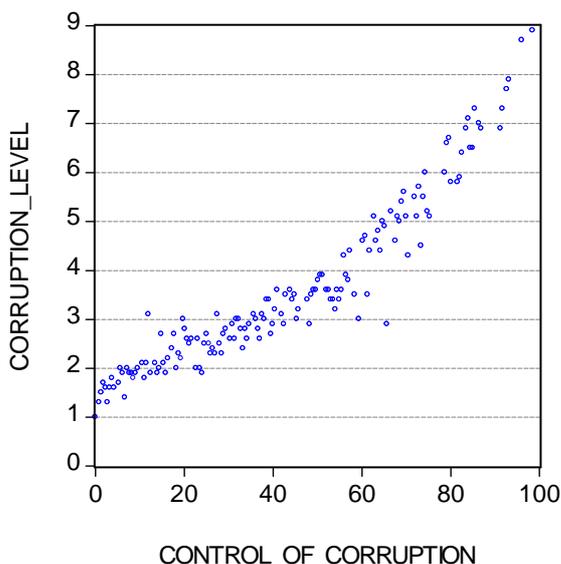
(H7) exogénéité : les variables $X_1, X_2, X_3, \dots, X_p$ sont indépendantes du terme d'erreurs

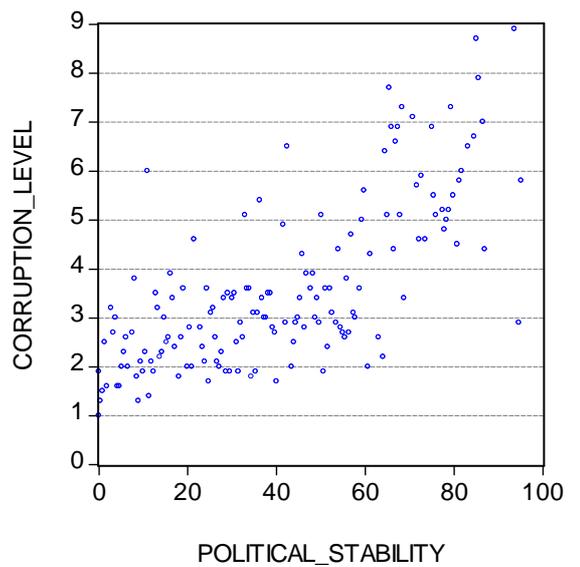
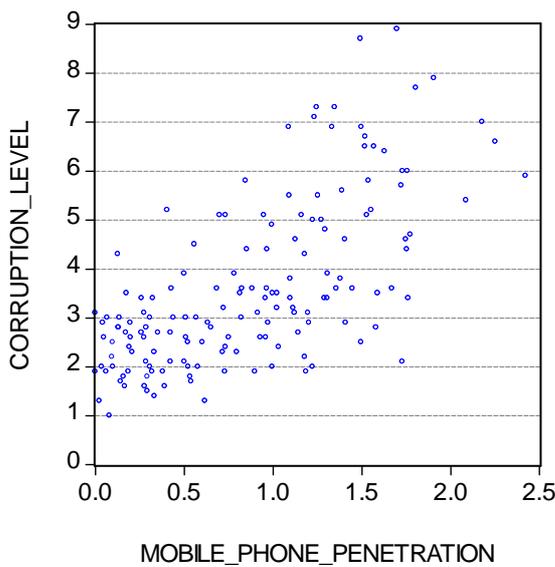
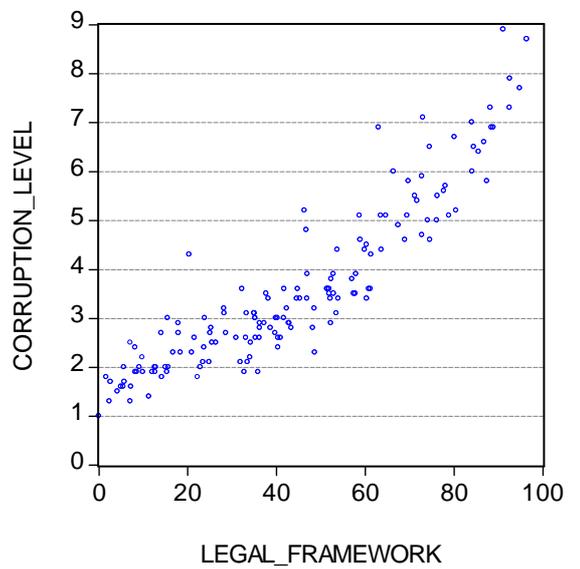
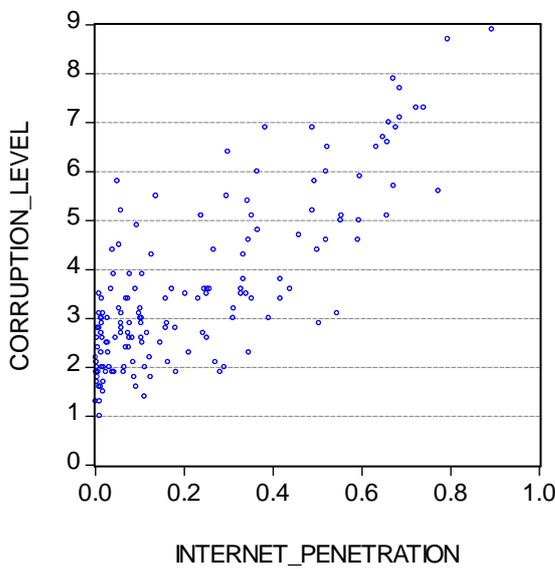
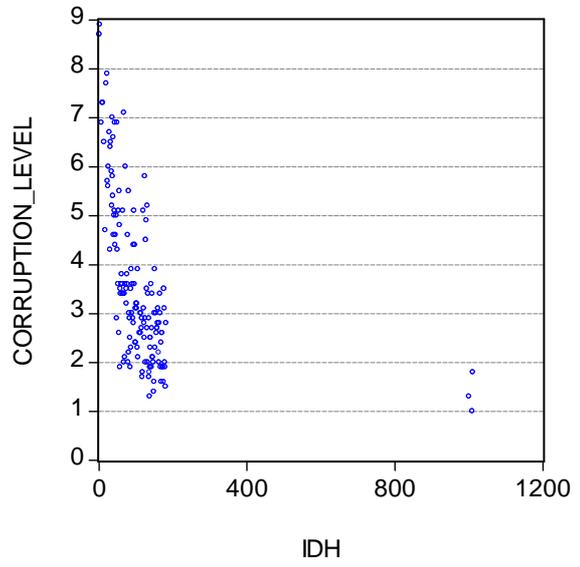
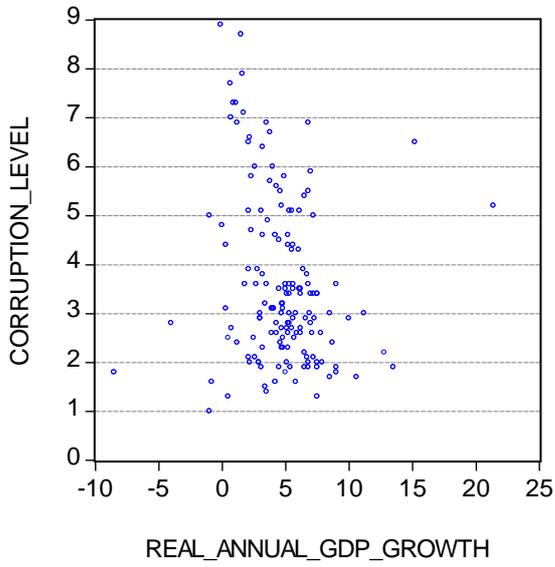
(H8) non multicollinéarité : les variables $X_1, X_2, X_3, \dots, X_p$ sont linéairement indépendantes.

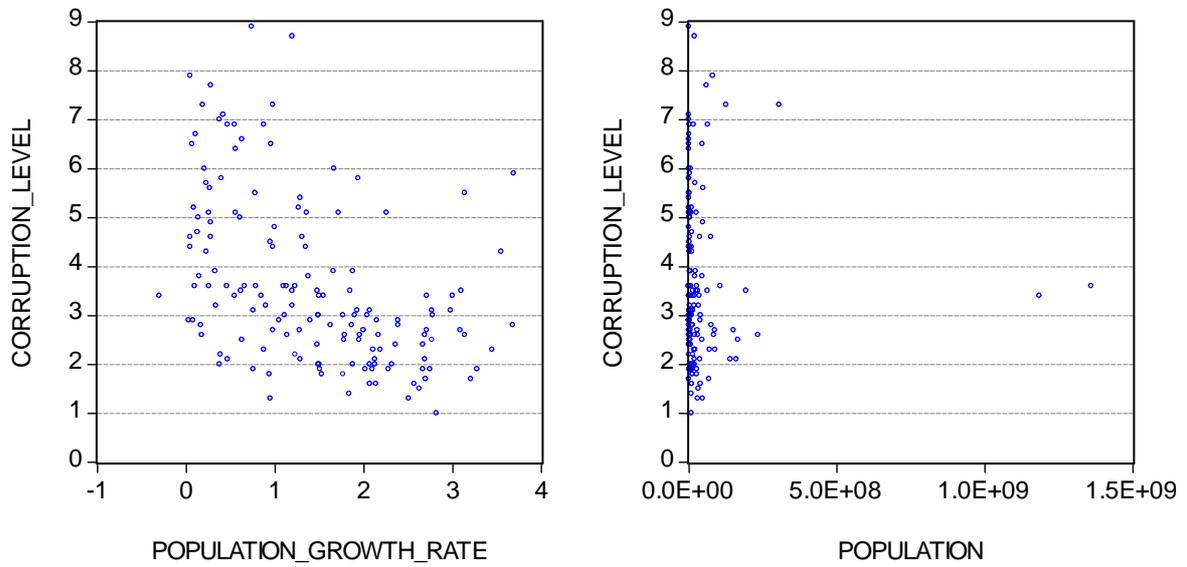
Les hypothèses H2, H6 et H7 sont par définition vérifiées par la méthode de modélisation par ANCOVA. Nous allons tester la non violation des autres hypothèses.

(H1) Linéarité du modèle et choix des régresseurs

Il s'agit d'une hypothèse de bonne spécification du modèle. On observe ci-dessous les Scatter Plots de la variable Corruption Level avec les variables explicatives. Les variables retenues dans la modélisation sont bien celles pour lesquelles on observe une relation linéaire avec la variable à expliquer : Control of Corruption, Internet Penetration, Legal Framework, Political Stability. Il existe par ailleurs une logique économique entre ces quatre variables et Corruption Level : un niveau de corruption plus faible (donc une meilleure note) serait expliqué par un bon contrôle de la corruption, un environnement politico-légal stable, un fort développement des moyens de communication.

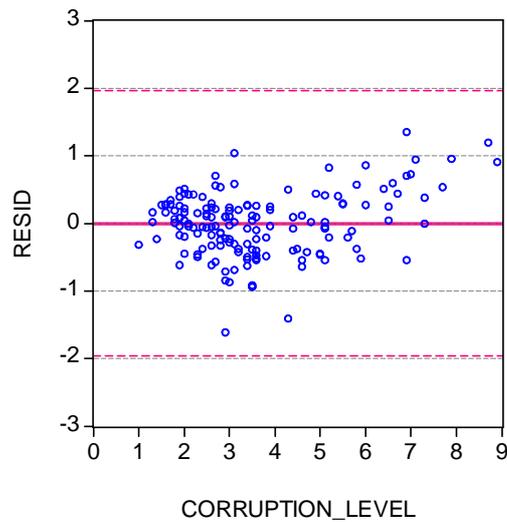






(H3) Homoscédasticité

Le graphique (scatter plot) ci-dessous, qui représente en abscisse les valeurs estimées de la variable Corruption Level, et en ordonnée les résidus associés, nous montre que les résidus ne semblent pas distribués aléatoirement. Par ailleurs, le test de White donne une probabilité de 0,03%. Cela laisse supposer qu'on doit rejeter l'hypothèse H_0 d'homoscédasticité et considérer que les données sont hétéroscédastiques.



White Heteroskedasticity Test:

F-statistic	3.079267	Probability	0.000059
Obs*R-squared	48.80595	Probability	0.000328

Compte tenu de l'hétérogénéité des pays observés, cela ne semble pas étonnant. Pour continuer la modélisation, on tient compte de l'hétéroscédasticité pour les bornes de confiance et statistiques de test en utilisant des corrections proposées par e-views : on utilise l'option « White Heteroskedasticity-Consistent Standard Errors & Covariance » qui permet de prendre en compte et de pallier la violation de l'hypothèse d'homoscédasticité en corrigeant les écarts-types des résidus pour retirer le biais. L'utilisation de cette option supplémentaire n'a pas d'impact sur les estimations et les coefficients de l'équation.

(H4) Non-autocorrélation

Pour vérifier la non-autocorrélation des résidus, on utilise le test de Durbin Watson, dont les règles de diagnostic et la table associée sont présentées ci-dessous.

Intervalle de la valeur calculée	Diagnostic
DW	
DW $\in [0 ; d_L]$	H_0 est rejetée et $\rho > 0$
DW $\in [d_L ; d_u]$	Pas de conclusion
DW $\in [d_u ; 4 - d_L]$	H_0 n'est pas rejetée
DW $\in [4 - d_u ; 4 - d_L]$	Pas de conclusion
DW $\in [4 - d_L ; 4]$	H_0 est rejetée et $\rho < 0$

Tableau 3 : Table de Durbin-Watson : Test unilatéral de $\rho=0$ contre $\rho>0$, au seuil de 5% (test bilatéral : seuil $\alpha=10\%$)

n	k' = 1		k' = 2		k' = 3		k' = 4		k' = 5		k' = 6		k' = 7		k' = 8		k' = 9		k' = 10	
	d _L	d _u																		
15	1,03	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21	0,45	2,47	0,34	2,73	0,25	2,98	0,17	3,22	0,11	3,44
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15	0,50	2,40	0,40	2,62	0,30	2,86	0,22	3,09	0,15	3,30
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10	0,55	2,32	0,45	2,54	0,36	2,76	0,27	2,97	0,20	3,20
18	1,15	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06	0,60	2,26	0,50	2,46	0,41	2,67	0,32	2,87	0,24	3,07
19	1,13	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02	0,65	2,21	0,46	2,40	0,46	2,59	0,37	2,78	0,29	2,97
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99	0,69	2,16	0,60	2,34	0,50	2,52	0,42	2,70	0,34	2,88
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96	0,73	2,12	0,64	2,29	0,55	2,46	0,46	2,63	0,38	2,81
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94	0,77	2,09	0,68	2,25	0,59	2,41	0,50	2,57	0,42	2,73
23	1,25	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92	0,80	2,06	0,71	2,21	0,63	2,36	0,54	2,51	0,46	2,67
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90	0,84	2,03	0,75	2,17	0,67	2,32	0,58	2,46	0,51	2,61
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89	0,87	2,01	0,78	2,14	0,70	2,28	0,62	2,42	0,54	2,56
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88	0,90	1,99	0,82	2,12	0,73	2,25	0,66	2,38	0,58	2,51
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86	0,92	1,97	0,84	2,09	0,77	2,22	0,69	2,34	0,62	2,47
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85	0,95	1,96	0,87	2,07	0,80	2,19	0,72	2,31	0,65	2,43
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84	0,97	1,94	0,90	2,05	0,83	2,16	0,75	2,28	0,68	2,40
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83	1,00	1,93	0,93	2,03	0,85	2,14	0,78	2,25	0,71	2,36
31	1,35	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83	1,02	1,92	0,95	2,02	0,88	2,12	0,81	2,23	0,74	2,33
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82	1,04	1,91	0,97	2,00	0,90	2,10	0,84	2,20	0,77	2,31
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81	1,06	1,90	0,99	1,99	0,93	2,08	0,86	2,18	0,79	2,28
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81	1,08	1,89	1,01	1,98	0,95	2,07	0,88	2,16	0,82	2,26
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80	1,10	1,88	1,03	1,97	0,97	2,05	0,91	2,14	0,84	2,24
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80	1,11	1,88	1,05	1,96	0,99	2,04	0,93	2,13	0,87	2,22
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80	1,13	1,87	1,07	1,95	1,01	2,03	0,95	2,11	0,89	2,20
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79	1,15	1,86	1,09	1,94	1,03	2,02	0,97	2,10	0,91	2,18
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79	1,16	1,86	1,10	1,93	1,05	2,01	0,99	2,08	0,93	2,16
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79	1,17	1,85	1,12	1,92	1,06	2,00	1,01	2,07	0,95	2,14
45	1,43	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78	1,24	1,84	1,19	1,90	1,14	1,96	1,05	2,00	1,04	2,09
50	1,50	1,58	1,48	1,63	1,42	1,67	1,38	1,72	1,34	1,77	1,28	1,82	1,25	1,87	1,20	1,93	1,16	1,98	1,11	2,04
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77	1,33	1,81	1,29	1,86	1,25	1,91	1,21	1,96	1,17	2,01
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77	1,37	1,81	1,33	1,85	1,30	1,89	1,26	1,94	1,22	1,98
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77	1,40	1,80	1,37	1,84	1,34	1,88	1,30	1,92	1,27	1,96
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77	1,43	1,80	1,40	1,84	1,37	1,87	1,34	1,91	1,30	1,95
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77	1,46	1,80	1,43	1,83	1,40	1,87	1,37	1,90	1,34	1,94
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77	1,48	1,80	1,45	1,83	1,42	1,86	1,40	1,89	1,37	1,92
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77	1,50	1,80	1,47	1,83	1,45	1,86	1,42	1,89	1,40	1,92
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78	1,52	1,80	1,49	1,83	1,47	1,85	1,44	1,88	1,42	1,91
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78	1,54	1,80	1,51	1,83	1,49	1,85	1,46	1,88	1,44	1,90
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78	1,55	1,80	1,53	1,83	1,51	1,85	1,48	1,87	1,46	1,90
150	1,72	1,75	1,71	1,75	1,69	1,77	1,68	1,79	1,66	1,80	1,65	1,82	1,64	1,83	1,62	1,85	1,60	1,86	1,59	1,88
200	1,73	1,78	1,75	1,79	1,73	1,80	1,73	1,81	1,72	1,82	1,71	1,83	1,70	1,84	1,69	1,85	1,68	1,86	1,66	1,87

Dans notre cas, le nombre d'observations n vaut 155, et le nombre de variables k' vaut 10.

On obtient $d_L = 1,59$ et $d_u = 1,88$. Pour conserver l'hypothèse H_0 de non autocorrélation des résidus, la statistique DW doit donc appartenir à l'intervalle $]1,88 ; 2,41]$. La table de statistiques descriptives suivante donne $DW = 2,30$. On ne rejette donc pas H_0 et il semble raisonnable de considérer que les résidus sont non autocorrélés.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.151956	0.193363	5.957468	0.0000
CONTROL_OF_CORRUPTION	0.039171	0.004117	9.514911	0.0000
GDP_CAPITA	1.52E-05	5.03E-06	3.021018	0.0030
IDH	4.12E-05	0.000346	0.119050	0.9054
INTERNET_PENETRATION	1.350359	0.364098	3.708779	0.0003
LEGAL_FRAMEWORK	0.010265	0.004843	2.119406	0.0358
MOBILE_PHONE_PENETRATION	-0.247358	0.129075	-1.916382	0.0573
POLITICAL_STABILITY	0.003981	0.002324	1.712495	0.0890
POPULATION	6.14E-11	2.79E-10	0.220255	0.8260
POPULATION_GROWTH_RATE	0.034535	0.055186	0.625795	0.5324
REAL_ANNUAL_GDP_GROWTH	-0.025262	0.013198	-1.914020	0.0576
R-squared	0.924203	Mean dependent var	3.612903	
Adjusted R-squared	0.918939	S.D. dependent var	1.691718	
S.E. of regression	0.481652	Akaike info criterion	1.445135	
Sum squared resid	33.40641	Schwarz criterion	1.661120	
Log likelihood	-100.9980	F-statistic	175.5807	
Durbin-Watson stat	2.302027	Prob(F-statistic)	0.000000	

Le corrélogramme des résidus suivant confirme le résultat du test de Durbin Watson.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.144	-0.144	3.2649	0.071
		2	-0.143	-0.167	6.5063	0.039
		3	0.034	-0.014	6.6944	0.082
		4	-0.063	-0.088	7.3369	0.119
		5	0.119	0.103	9.6163	0.087
		6	-0.195	-0.195	15.856	0.015
		7	-0.072	-0.103	16.720	0.019
		8	0.100	0.001	18.366	0.019
		9	0.094	0.114	19.849	0.019
		10	-0.007	0.006	19.858	0.031
		11	-0.097	-0.047	21.444	0.029
		12	0.031	-0.011	21.609	0.042
		13	-0.089	-0.149	22.981	0.042
		14	-0.034	-0.075	23.177	0.057
		15	-0.071	-0.101	24.047	0.064
		16	-0.023	-0.040	24.138	0.087
		17	0.039	-0.069	24.410	0.109
		18	0.050	0.034	24.853	0.129
		19	-0.035	-0.080	25.077	0.158
		20	-0.045	-0.078	25.444	0.185
		21	0.103	0.048	27.376	0.159
		22	-0.058	-0.052	27.999	0.176
		23	-0.102	-0.121	29.919	0.152
		24	-0.005	-0.067	29.923	0.187
		25	0.053	0.016	30.451	0.208
		26	0.104	0.024	32.507	0.177
		27	-0.120	-0.119	35.224	0.133
		28	0.019	-0.014	35.296	0.161
		29	0.024	-0.082	35.403	0.192
		30	0.030	-0.034	35.580	0.222
		31	-0.049	-0.082	36.058	0.244
		32	-0.018	0.036	36.121	0.282
		33	0.135	0.069	39.743	0.195
		34	-0.144	-0.180	43.912	0.119
		35	0.039	-0.034	44.221	0.137
		36	0.051	-0.003	44.756	0.150

(H5) Normalité des Erreurs

A partir de la différence entre la valeur réelle et la valeur modélisée pour chaque observation, on déduit la valeur des résidus, dont on veut vérifier qu'ils sont bien distribués selon une loi Normale.

On visualise ci-après le graphique de comparaison entre prédictions et valeurs réelles, encadrées par les bornes d'intervalle de confiance à 95%, ainsi que le tableau de valeurs de prédictions et des résidus.

La régression par ANCOVA implique que les résidus soient gaussiens. On vérifie que cette hypothèse est bien respectée en appliquant à nos résidus le test de Jarque-Bera.

La statistique de Jarque-Bera permet de choisir entre :

H_0 : les données suivent une loi normale

H_1 : les données ne suivent pas une loi normale

Elle se base sur les coefficients d'asymétrie (Skewness) et d'aplatissement (Kurtosis) de la distribution.

$$JB = \frac{(n - p)}{6} \times \left(Sk^2 + \frac{(Ku - 3)^2}{4} \right)$$

$n =$ nombre d'observations
où $p =$ nombre de variables explicatives
 $Sk =$ Skewness des résidus
 $Ku =$ Kurtosis des résidus

On vérifie en fait que :

Le Skewness est proche de celui de la Loi Normale (égal à 0)

$$Sk = \frac{\mu_3}{\sigma^3} \text{ où } \mu_3 \text{ est le moment centré d'ordre 3 et } \sigma \text{ l'écart - type.}$$

Le Kurtosis est proche de celui de la Loi Normale (égal à 3)

$$Ku = \frac{\mu_4}{\sigma^4} \text{ où } \mu_4 \text{ est le moment centré d'ordre 4 et } \sigma \text{ l'écart - type.}$$

L'application numérique sur les résidus de la modélisation de la variable Corruption donne les résultats suivants :

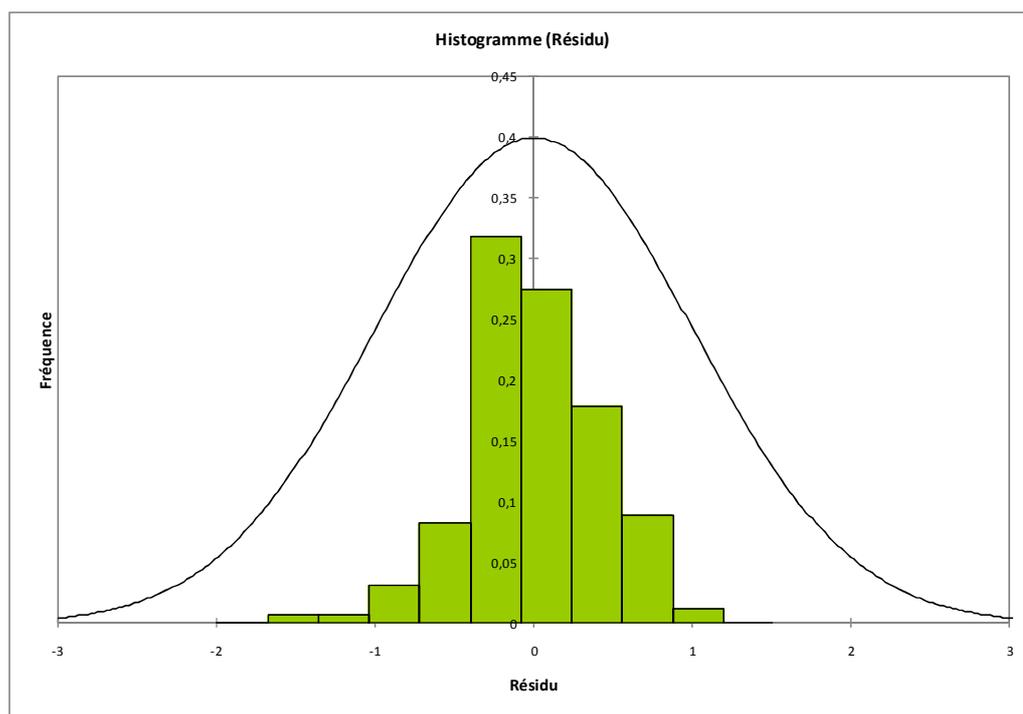
$$S_k = -0,1923$$

$$K_u = 3,4609$$

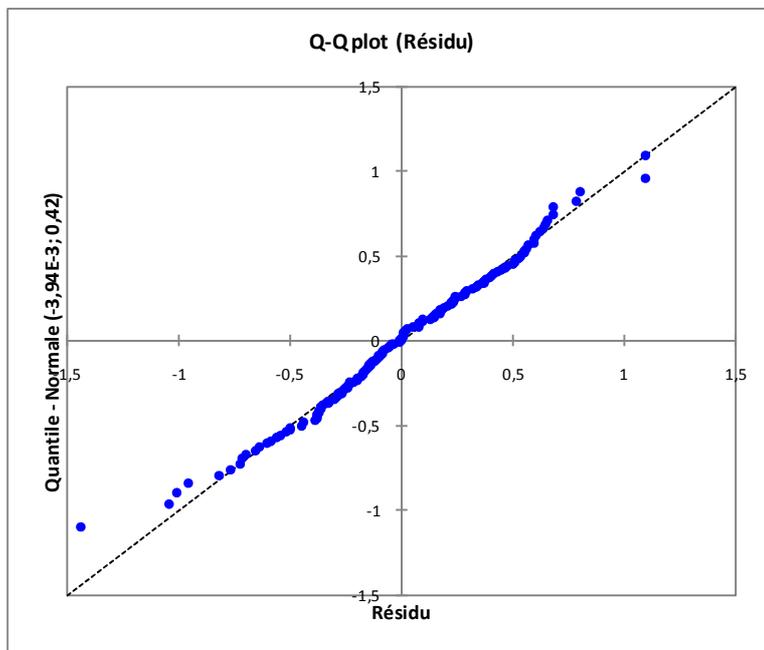
$$J_B = 2,2075$$

Or, la statistique de Jarque-Bera suit asymptotiquement une loi du Chi-deux à 2 degrés de liberté. Au seuil $\alpha=5\%$, on a $\chi^2(0,05; 2) = 5,9915$.

L'histogramme ci-dessous représente la distribution des résidus, à comparer avec la courbe de densité de la loi Normale centrée réduite.

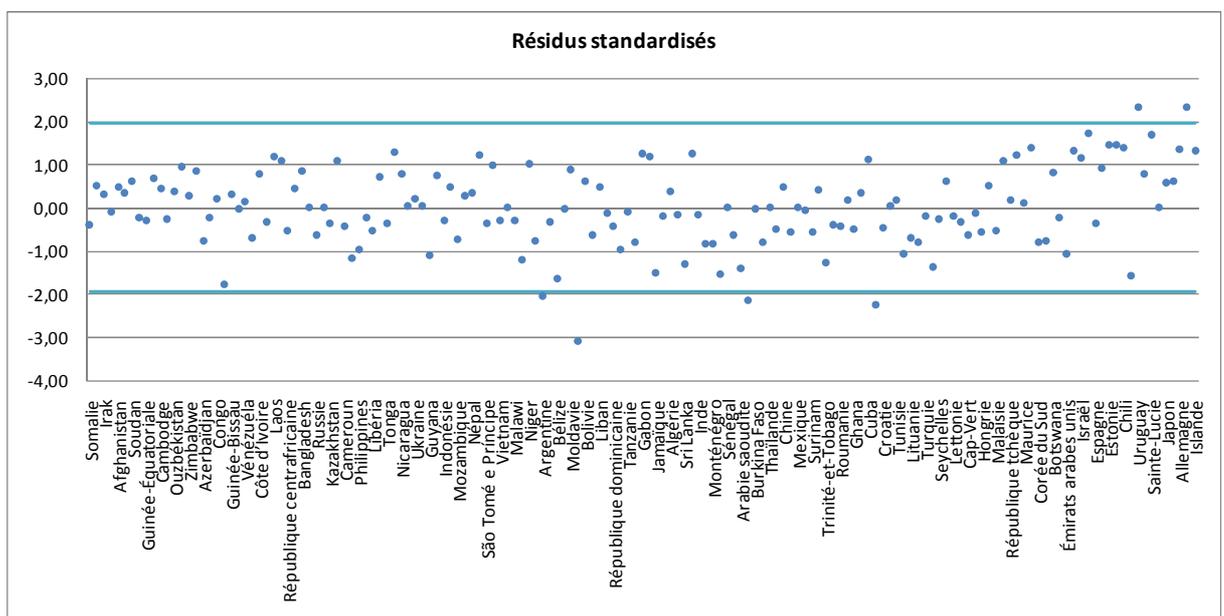


La p-value associée est de 33%, on rejette donc H_1 et on valide l'hypothèse H_0 de normalité des résidus.



Le graphe Q-Q Plot permet de comparer la distribution de nos résidus (axe vertical) à une distribution normale théorique (axe horizontal). La linéarité observée de la distribution confirme que les résidus sont gaussiens.

Pour valider le modèle de prédiction par ANCOVA, on vérifie que les résidus standardisés (centrés réduits) sont bien distribués autour de l'axe des abscisses : 95% au moins des valeurs doivent être comprises dans l'intervalle $[-1,96 ; 1,96]$. Cette vérification nous permet d'identifier des valeurs aberrantes.



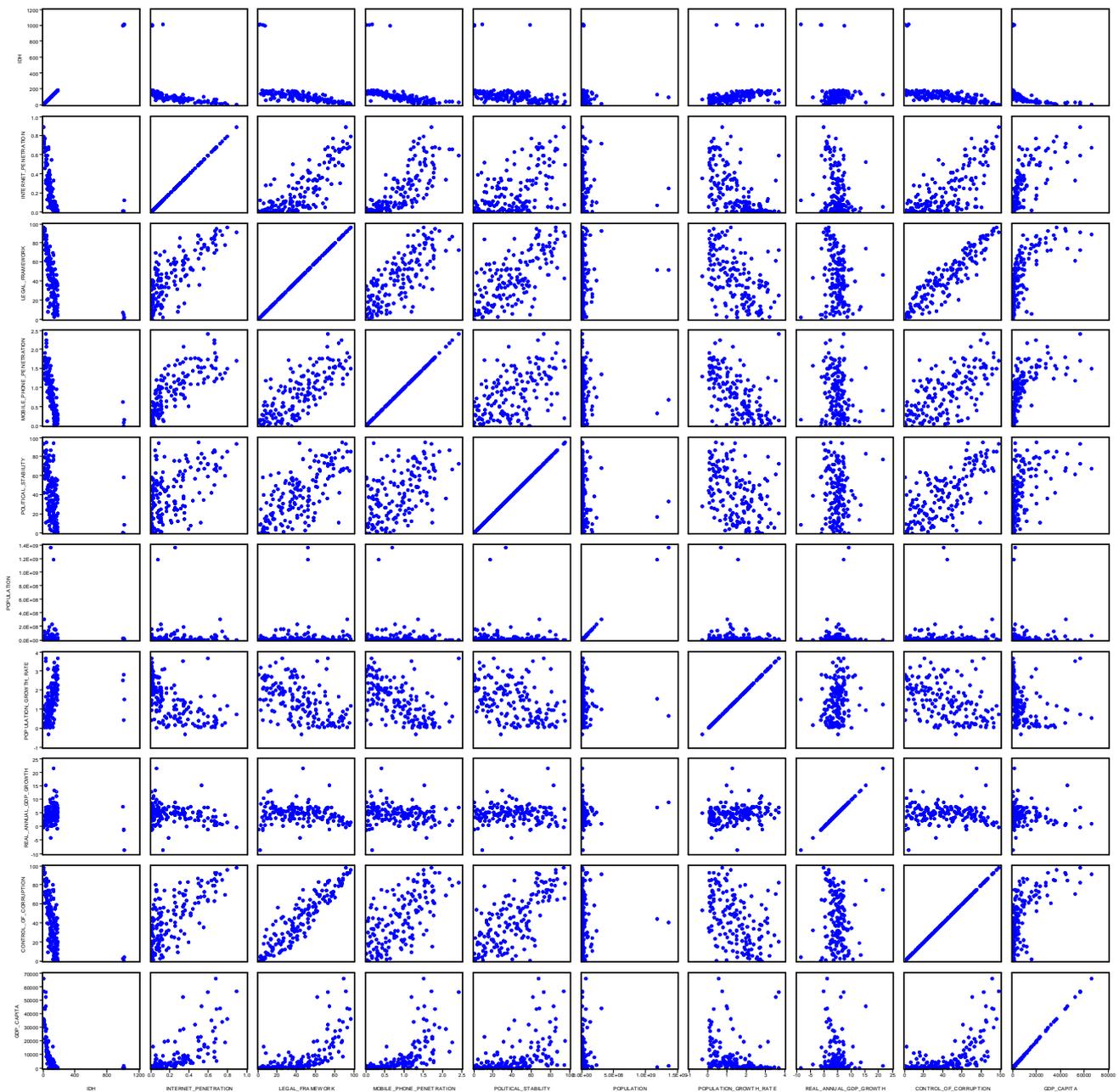
Le tableau et le graphique indiquent que seulement 6 points sont hors de l'intervalle $[-1,96 ; 1,96]$, ce qui correspond à un pourcentage de valeurs aberrantes de $6/157 = 3,82\% < 5\%$.

(H8) Non multicolinéarité

La présence de quasi-multicolinéarité rendrait les estimations instables : de petites variations dans les données pourraient faire varier fortement les estimations des coefficients.

Afin de vérifier l'absence de multicolinéarité, on construit les scatter plots deux à deux des variables explicatives. Celles-ci sont présentées ci-dessous sous la forme d'une matrice symétrique. On observe que les variables explicatives ne semblent pas linéaires entre elles, notamment par comparaison aux graphes de diagonale représentant le scatter plot d'une variable explicative par rapport à elle-même.

Figure 4: Matrice symétrique des variables explicatives

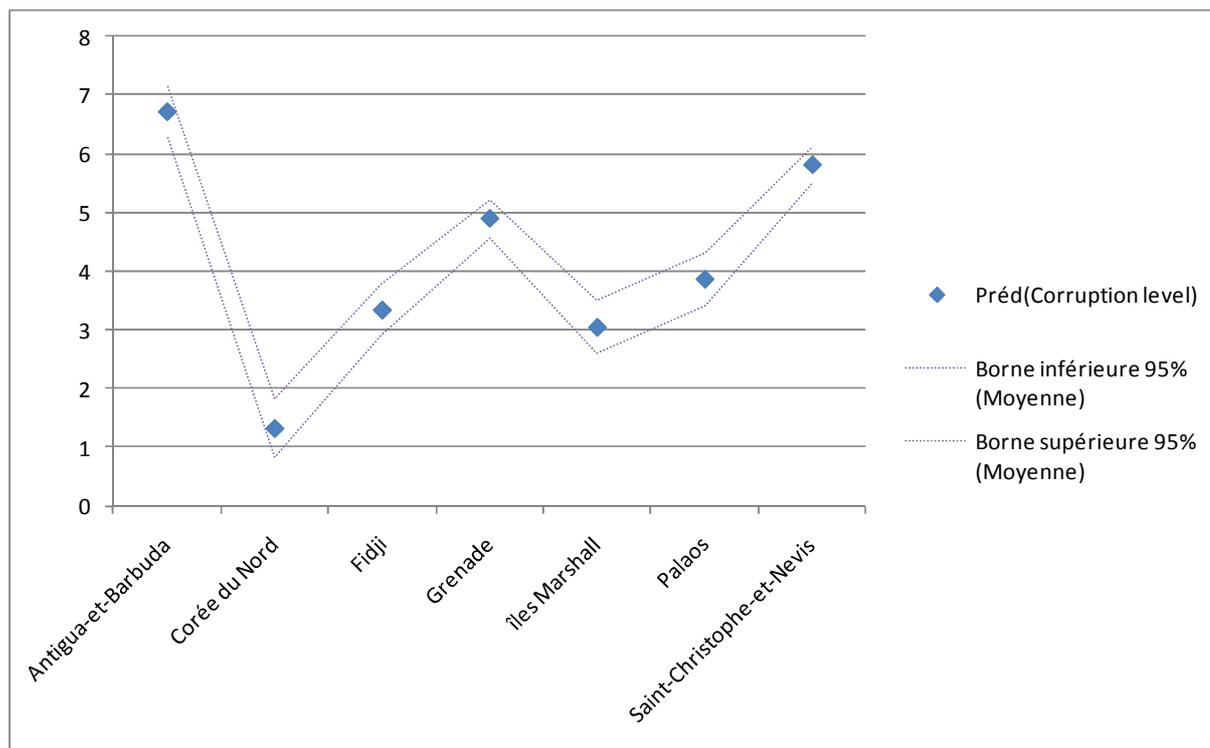


Les tests effectués sont globalement concluants, la modélisation obtenue et l'utilisation de l'ANCOVA pour la prédiction des résultats semble donc raisonnable.

3.3.1.7. Prédiction

Le tableau suivant donne les données manquantes modélisées pour la variable Corruption.

Observation	Préd(Corruption level)	Ecart-type sur la préd. (Moyenne)	Borne inférieure 95% (Moyenne)	Borne supérieure 95% (Moyenne)
Antigua-et-Barbuda	6,706	0,214	6,283	7,128
Corée du Nord	1,326	0,244	0,843	1,810
Fidji	3,338	0,217	2,908	3,768
Grenade	4,895	0,167	4,566	5,225
îles Marshall	3,046	0,228	2,595	3,497
Palaos	3,861	0,224	3,418	4,305
Saint-Christophe-et-Nevis	5,806	0,160	5,489	6,123



On considère que les intervalles de confiance sont suffisamment fins pour valider nos données. Le niveau d'erreur potentielle nous semble acceptable.

On ajoute donc les prédictions dans la base de données manquantes, et on les intègre à la régression ANCOVA de modélisation de la variable suivante : Index Economic Freedom.

On suivra la même méthodologie pour la modélisation de toutes les variables. La description des étapes de modélisation sera placée en annexe : présentation des tableaux de données permettant de construire le modèle ANCOVA et d'estimer les données manquantes.

Au final, on obtient une base de données complète sur l'ensemble des variables considérées. Cette consolidation ne fournit pas des données fiables à 100%, notamment pour les dernières variables pour lesquelles les données manquantes étaient relativement nombreuses. On considère donc ces données supplémentaires comme un indicateur qui, s'il ne permet pas de représenter la situation économique exacte de chaque pays, apporte néanmoins une information supplémentaire appréciable dans un modèle de décision dans l'incertain.

On va maintenant utiliser cette base pour construire un modèle de notation des pays comme support à la décision d'investissement.

4. Sélection

4.1. « Profiling » des pays : Application de l'ACP

4.1.1. Méthodologie de l'Analyse en Composantes Principales

4.1.1.1. Présentation

A partir de la base de données constituée, on choisit d'utiliser une méthode d'Analyse en Composantes Principales.

L'Analyse en Composantes Principales fait partie du groupe des méthodes de statistiques descriptives multidimensionnelles appelées méthodes factorielles. Elle propose de construire des relations géométriques entre les observations et les variables.

L'objectif est ici d'analyser la corrélation entre les variables, et d'identifier les pays qui se différencient fortement des autres. Il s'agit donc de détecter la présence de profils-types de pays parmi notre base, et les critères qui permettent d'identifier ces différents profils, leurs caractéristiques communes spécifiques.

L'Analyse en Composantes Principales permet en effet de :

- visualiser et analyser rapidement les corrélations entre les variables,
- visualiser et analyser les observations initialement décrites par l'ensemble des variables sur un graphique à deux ou trois dimensions, construit de manière à ce que la dispersion entre les données soit aussi bien préservée que possible,
- construire un ensemble de facteurs non corrélés (les axes de l'ACP) que l'on pourra ensuite réutiliser dans d'autres méthodes d'analyse.

4.1.1.2. Indications méthodologiques

On associe notre table de données à n observations et p variables à la matrice suivante :

$$\begin{array}{cccccc} & \mathbf{v}_1 & \dots & \mathbf{v}_j & \dots & \mathbf{v}_p \\ \mathbf{o}_1 & [x_{11} & \dots & x_{1j} & \dots & x_{1p}] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{o}_i & [x_{i1} & \dots & x_{ij} & \dots & x_{ip}] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{o}_n & [x_{n1} & \dots & x_{nj} & \dots & x_{np}] \end{array}$$

où on définit pour tout $1 < i < n$ et $1 < j < p$ les vecteurs :

$$\mathbf{o}_i = [x_{i1} \quad \dots \quad x_{ij} \quad \dots \quad x_{ip}] \text{ et } \mathbf{v}_j = \begin{bmatrix} x_{1j} \\ \dots \\ x_{ij} \\ \dots \\ x_{nj} \end{bmatrix}$$

On souhaite représenter les observations sur un espace affine : on a pour l'instant pour chaque observation p coordonnées liées à nos p variables.

On va chercher diagonaliser la matrice ci-dessus afin d'en extraire les vecteurs propres qui vont constituer nos axes. On sait qu'on aura au maximum p vecteurs propres, mais que ce nombre peut-être réduit. Afin de pouvoir visualiser les projections, on souhaite pouvoir se limiter à seulement 2 ou 3 axes.

La distance utilisée pour représenter les observations dans l'espace de l'ACP est la distance euclidienne classique. La distance entre deux observations o_i et $o_{i'}$ sera mesurée par :

$$d^2(o_i; o_{i'}) = \sum_{j=1}^p (x_{ij} - x_{i'j})^2$$

Sur la base de cette distance, les axes définis par les variables constituent une base orthogonale.

4.1.2. ACP sur base complète

On construit tout d'abord une Analyse en Composantes Principales Normée sur la base de l'ensemble des variables, pour avoir un aperçu des regroupements par axe, et du niveau de significativité de chacune.

4.1.2.1. Statistiques descriptives

Tableau 4 : Statistiques descriptives des variables quantitatives

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Population	164	20842,000	1360445010,000	40332898,555	144547921,775
Population Growth Rate	164	-0,300	3,690	1,415	0,954
Mobile phone penetration	164	0,000	2,423	0,868	0,580
Internet penetration	164	0,000	0,893	0,226	0,227
GDP/capita	164	112,070	66016,000	8120,304	12336,225
IDH	164	2,000	1012,000	131,988	177,400
Real annual GDP Growth	164	-8,500	21,400	4,703	3,311
Legal framework	164	0,000	96,433	43,407	25,327
Political stability	164	0,000	95,200	42,569	25,971
Control of Corruption	164	0,000	98,500	43,225	26,603
Corruption level	164	1,000	8,900	3,613	1,696
Economic Freedom Index	164	0,000	82,600	56,208	14,254
Vehicles per capita	164	0,000	765,000	130,134	152,729
Financial Risk	164	12,000	49,500	35,183	5,580
Economic Risk	164	18,000	49,500	36,766	5,203
Political Risk	164	23,500	87,000	64,565	10,833
Vehicle Ownership Projection	164	0,200	13,600	4,005	1,831
P&C spending/capita	164	0,000	2167,054	230,667	379,198
Penetration rate	164	0,000	0,099	0,020	0,015
P&C spending trend	164	0,000	42,567	17,120	8,884

Tableau 5 : Statistiques descriptives des variables qualitatives

Variable	Modalités	Effectifs	%	
Zone	Asie	30	18,293	
	Afrique	50	30,488	
	Europe de l'Est	21	12,805	
	Europe	6	3,659	
	Amérique Latine	31	18,902	
	Moyen Orient	16	9,756	
	Océanie	9	5,488	
	Amérique	1	0,610	
	Organisation	SAARC	9	5,488
		SADC	13	7,927
NC		17	10,366	
Arab Maghreb Union		6	3,659	
Union Européenne		20	12,195	
CARICOM		13	7,927	
GCC		7	4,268	
MERCOSUR		5	3,049	
Commonwealth of Independent States		13	7,927	
Pacific Islands		9	5,488	
ECOWAS		16	9,756	
ASEAN		10	6,098	
Andean Community		5	3,049	
CEMAC		6	3,659	
IGAD		6	3,659	
NAFTA		2	1,220	
Central American Parliament		7	4,268	

On visualise ci-après la matrice de corrélations entre les variables, découpée en trois parties : corrélations entre variables quantitatives, corrélations entre les variables quantitatives et qualitatives, corrélations entre variables qualitatives.

Tableau 6 : Matrice de corrélation entre variables quantitatives

Variables	Population	Population Growth Rate	Mobile phone penetration	Internet penetration	GDP/capita	IDH	Real annual GDP Growth	Legal framework	Political stability	Control of Corruption	Corruption	Economic Freedom Index	Vehicles per capita	Financial Risk	Economic Risk	Political Risk	Vehicle Ownership Projection	P&C spending/capita ^a	Penetration rate	P&C spending trend
Population	1	-0,054	-0,059	0,008	-0,017	-0,034	0,095	0,056	-0,133	-0,006	0,000	-0,006	0,011	0,056	0,170	-0,014	0,494	0,002	-0,070	0,013
Population Growth Rate	-0,054	1	-0,559	-0,551	-0,258	0,207	0,184	-0,503	-0,398	-0,439	-0,441	-0,291	-0,463	-0,053	0,136	-0,448	0,311	-0,305	-0,312	0,119
Mobile phone penetration	-0,059	-0,559	1	0,792	0,644	-0,401	-0,154	0,758	0,509	0,661	0,684	0,549	0,666	0,351	-0,037	0,642	-0,365	0,502	0,394	-0,183
Internet penetration	0,008	-0,551	0,792	1	0,725	-0,351	-0,236	0,796	0,522	0,731	0,801	0,514	0,795	0,213	-0,193	0,647	-0,433	0,640	0,345	-0,256
GDP/capita	-0,017	-0,258	0,644	0,725	1	-0,287	-0,123	0,668	0,479	0,641	0,714	0,431	0,823	0,387	0,002	0,574	-0,390	0,747	0,184	-0,211
IDH	-0,034	0,207	-0,401	-0,351	-0,287	1	-0,190	-0,412	-0,156	-0,384	-0,365	-0,717	-0,305	-0,288	-0,154	-0,466	0,173	-0,185	-0,195	-0,300
Real annual GDP Growth	0,095	0,184	-0,154	-0,236	-0,123	-0,190	1	-0,144	-0,038	-0,097	-0,167	0,018	-0,188	0,410	0,359	-0,033	0,248	-0,260	-0,236	0,359
Legal framework	0,056	-0,503	0,758	0,796	0,668	-0,412	-0,144	1	0,662	0,919	0,910	0,732	0,704	0,293	-0,079	0,829	-0,385	0,581	0,329	-0,346
Political stability	-0,133	-0,398	0,509	0,522	0,479	-0,156	-0,038	0,662	1	0,704	0,692	0,405	0,511	0,272	-0,057	0,840	-0,337	0,476	0,327	-0,283
Control of Corruption	-0,006	-0,439	0,661	0,731	0,641	-0,384	-0,097	0,919	0,704	1	0,943	0,664	0,654	0,287	-0,117	0,815	-0,401	0,581	0,339	-0,308
Corruption	0,000	-0,441	0,684	0,801	0,714	-0,365	-0,167	0,910	0,692	0,943	1	0,651	0,745	0,306	-0,125	0,769	-0,424	0,681	0,368	-0,349
Economic Freedom Index	-0,006	-0,291	0,549	0,514	0,431	-0,717	0,018	0,732	0,405	0,664	0,651	1	0,449	0,264	0,043	0,690	-0,314	0,284	0,142	-0,035
Vehicles per capita	0,011	-0,463	0,666	0,795	0,823	-0,305	-0,188	0,704	0,511	0,654	0,745	0,449	1	0,320	-0,086	0,621	-0,495	0,811	0,379	-0,249
Financial Risk	0,056	-0,053	0,351	0,213	0,387	-0,288	0,410	0,293	0,272	0,287	0,306	0,264	0,320	1	0,652	0,390	0,034	0,241	-0,023	-0,071
Economic Risk	0,170	0,136	-0,037	-0,193	0,002	-0,154	0,359	-0,079	-0,057	-0,117	-0,125	0,043	-0,086	0,652	1	0,051	0,270	-0,055	-0,107	0,080
Political Risk	-0,014	-0,448	0,642	0,647	0,574	-0,466	-0,033	0,829	0,840	0,815	0,769	0,690	0,621	0,390	0,051	1	-0,354	0,526	0,333	-0,212
Vehicle Ownership Projection	0,494	0,311	-0,365	-0,433	-0,390	0,173	0,248	-0,385	-0,337	-0,401	-0,424	-0,314	-0,495	0,034	0,270	-0,354	1	-0,441	-0,327	-0,028
P&C spending/capita	0,002	-0,305	0,502	0,640	0,747	-0,185	-0,260	0,581	0,476	0,581	0,681	0,284	0,811	0,241	-0,055	0,526	-0,441	1	0,566	-0,288
Penetration rate	-0,070	-0,312	0,394	0,345	0,184	-0,195	-0,236	0,329	0,327	0,339	0,368	0,142	0,379	-0,023	-0,107	0,333	-0,327	0,566	1	-0,147
P&C spending trend	0,013	0,119	-0,183	-0,256	-0,211	-0,300	0,359	-0,346	-0,283	-0,308	-0,349	-0,035	-0,249	-0,071	0,080	-0,212	-0,028	-0,288	-0,147	1

Tableau 7 : Matrice de corrélation entre variables quantitatives et qualitatives

Variables	Population	Population Growth Rate	Population penetration	Mobile phone penetration	Internet penetration	GDP/capita	IDH	Real annual GDP Growth	Legal framework	Political stability	Control of Corruption	Corruption	Economic Freedom Index	Vehicles per capita	Financial Risk	Economic Risk	Political Risk	Vehicle Ownership Projection	P&C spending/capita	Penetration rate	P&C spending trend
Zone-Asie	0,309	-0,136	-0,127	-0,086	-0,133	0,014	0,190	-0,153	-0,181	-0,236	-0,196	-0,153	-0,109	0,050	0,249	-0,174	0,361	-0,117	-0,185	0,064	
Zone-Afrique	-0,096	0,518	-0,511	-0,463	-0,320	0,180	0,047	-0,363	-0,214	-0,299	-0,293	-0,220	-0,404	-0,124	-0,060	-0,313	0,067	-0,328	-0,299	0,011	
Zone-Europe de l'Est	-0,080	-0,438	0,400	0,331	0,115	-0,164	-0,025	0,279	0,259	0,215	0,165	0,158	0,282	-0,113	-0,283	0,261	-0,245	0,094	0,176	0,074	
Zone-Europe	0,005	-0,228	0,284	0,382	0,567	-0,129	-0,201	0,340	0,196	0,320	0,402	0,188	0,457	0,055	-0,126	0,260	-0,303	0,531	0,133	-0,191	
Zone-Amérique Latine	-0,073	-0,124	0,176	0,133	-0,077	-0,142	-0,138	0,064	0,035	0,125	0,102	0,147	-0,075	0,001	0,009	0,054	-0,086	-0,023	0,336	-0,026	
Zone-Moyen Orient	-0,061	0,160	0,167	0,052	0,247	-0,008	0,133	0,123	-0,100	0,091	0,078	0,089	0,154	0,177	0,167	0,002	0,063	0,005	-0,111	0,047	
Zone-Océanie	-0,062	0,029	-0,115	-0,074	-0,031	0,233	-0,121	-0,018	0,241	0,044	0,016	-0,096	0,029	0,058	0,002	0,191	-0,017	0,183	0,066	-0,057	
Zone-Amérique	0,144	-0,036	0,065	0,172	0,230	-0,053	-0,086	0,153	0,078	0,143	0,171	0,135	0,327	-0,017	-0,057	0,130	-0,163	0,401	0,153	-0,126	
Organisation-SAARC	0,230	-0,006	-0,142	-0,168	-0,135	0,000	0,129	-0,091	-0,176	-0,086	-0,097	-0,124	-0,170	-0,131	-0,014	-0,190	0,402	-0,076	-0,108	-0,010	
Organisation-SADC	-0,042	0,088	-0,214	-0,211	-0,133	0,124	0,046	-0,045	0,063	-0,014	-0,036	-0,004	-0,185	-0,119	-0,098	-0,019	-0,040	-0,131	-0,035	0,104	
Organisation-NC	0,152	0,003	0,002	0,104	-0,034	0,015	0,028	0,022	-0,087	0,026	0,009	-0,036	0,098	-0,039	0,014	-0,019	0,066	-0,043	-0,006	-0,057	
Organisation-Arab Maghreb Union	-0,019	0,007	0,027	-0,012	-0,046	-0,041	-0,029	-0,005	0,006	0,003	-0,007	-0,036	-0,026	0,154	0,246	0,080	0,085	-0,070	-0,061	-0,119	
Organisation-Union Européenne	-0,056	-0,433	0,478	0,536	0,491	-0,202	-0,170	0,504	0,372	0,427	0,449	0,286	0,566	-0,017	-0,273	0,416	-0,403	0,446	0,165	0,008	
Organisation-CARICOM	-0,076	-0,156	0,194	0,202	0,001	-0,096	-0,183	0,203	0,209	0,233	0,229	0,191	0,011	-0,012	0,002	0,163	-0,049	0,078	0,251	-0,099	
Organisation-GCC	-0,050	0,198	0,298	0,150	0,463	-0,110	0,167	0,227	0,090	0,248	0,211	0,162	0,207	0,378	0,213	0,158	0,037	0,120	-0,089	0,021	
Organisation-MERCOSUR	0,013	-0,066	0,035	0,047	-0,027	-0,067	0,037	-0,029	0,016	0,052	0,041	-0,035	0,028	0,074	-0,054	-0,017	-0,098	-0,032	-0,020	0,138	
Organisation-Commonwealth of Independent States	-0,035	-0,226	0,022	-0,131	-0,109	0,057	0,141	-0,182	-0,058	-0,259	-0,238	-0,109	-0,070	-0,041	0,069	-0,078	-0,032	-0,121	0,015	0,275	
Organisation-Pacific Islands	-0,062	0,029	-0,115	-0,074	-0,031	0,233	-0,121	-0,018	0,241	0,044	0,016	-0,096	0,029	0,058	0,002	0,191	-0,017	0,183	0,066	-0,057	
Organisation-ECOWAS	-0,049	0,375	-0,286	-0,267	-0,199	0,063	0,037	-0,223	-0,136	-0,185	-0,169	-0,063	-0,259	-0,183	-0,196	-0,227	0,095	-0,191	-0,261	-0,180	
Organisation-ASEAN	0,043	-0,011	-0,104	-0,053	-0,118	-0,025	0,078	-0,112	-0,162	-0,174	-0,156	-0,060	-0,116	0,074	0,191	-0,140	0,169	-0,128	-0,176	-0,022	
Organisation-Andean Community	-0,019	0,012	0,021	0,007	-0,043	-0,051	0,001	-0,117	-0,190	-0,063	-0,079	-0,020	-0,080	0,106	0,131	-0,129	-0,014	-0,044	-0,022	0,020	
Organisation-CEMAC	-0,045	0,159	-0,181	-0,171	-0,057	0,014	0,015	-0,237	-0,156	-0,242	-0,174	-0,084	-0,075	0,214	0,125	-0,130	0,041	-0,105	-0,179	0,066	
Organisation-IGAD	-0,008	0,235	-0,226	-0,154	-0,116	0,183	-0,025	-0,161	-0,225	-0,177	-0,169	-0,251	-0,152	-0,162	-0,019	-0,305	-0,052	-0,081	0,009	0,062	
Organisation-NAFTA	0,128	-0,042	0,048	0,127	0,172	-0,063	-0,110	0,127	0,016	0,115	0,121	0,133	0,235	0,011	-0,038	0,133	-0,128	0,265	0,072	-0,142	
Organisation-Central American Parliament	-0,049	0,023	0,033	-0,063	-0,095	-0,033	-0,029	-0,033	-0,043	-0,068	-0,075	0,056	-0,117	-0,148	-0,052	-0,011	0,001	-0,074	0,371	-0,024	

Tableau 8 : Matrice de corrélation entre variables qualitatives

Variables	Zone-Asie	Zone-Afrique	Zone-Europe de l'Est	Zone-Europe	Zone-Amérique Latine	Zone-Moyen Orient	Zone-Océanie	Zone-Amérique	Organisation-SAARC	Organisation-SADC	Organisation-NC	Organisation-Arab Maghreb Union	Organisation-Union Européenne	Organisation-CARICOM	Organisation-GCC	Organisation-MERCOSUR	Organisation-Commonwealth of Independent States	Organisation-Pacific Islands	Organisation-ECOWAS	Organisation-ASEAN	Organisation-Andean Community	Organisation-CEMAC	Organisation-IGAD	Organisation-NAFTA	Organisation-Central American Parliament
Zone-Asie	1	-0,313	-0,181	-0,092	-0,228	-0,156	-0,114	-0,037	0,509	-0,139	0,046	-0,092	-0,176	-0,139	-0,100	-0,084	0,270	-0,114	-0,156	0,539	-0,084	-0,092	-0,092	-0,053	-0,139
Zone-Afrique	-0,313	1	-0,254	-0,129	-0,320	-0,218	-0,160	-0,052	-0,160	0,443	-0,095	0,294	-0,247	-0,194	-0,140	-0,117	-0,194	-0,160	0,496	-0,169	-0,117	0,294	0,294	-0,074	-0,194
Zone-Europe de l'Est	-0,181	-0,254	1	-0,075	-0,185	-0,126	-0,092	-0,030	-0,092	-0,112	0,109	-0,075	0,638	-0,112	-0,081	-0,068	0,090	-0,092	-0,126	-0,098	-0,068	-0,075	-0,075	-0,043	-0,112
Zone-Europe	-0,092	-0,129	-0,075	1	-0,094	-0,064	-0,047	-0,015	-0,047	-0,057	-0,066	-0,038	0,523	-0,057	-0,041	-0,035	-0,057	-0,047	-0,064	-0,050	-0,035	-0,038	-0,038	-0,022	-0,057
Zone-Amérique Latine	-0,228	-0,320	-0,185	-0,094	1	-0,159	-0,116	-0,038	-0,116	-0,142	-0,164	-0,094	-0,180	0,608	-0,102	0,367	-0,142	-0,116	-0,159	-0,123	0,367	-0,094	-0,094	0,088	0,608
Zone-Moyen Orient	-0,156	-0,218	-0,126	-0,064	-0,159	1	-0,079	-0,026	-0,079	-0,096	0,293	-0,064	-0,123	-0,096	0,642	-0,058	0,132	-0,079	-0,108	-0,084	-0,058	-0,064	-0,064	-0,037	-0,096
Zone-Océanie	-0,114	-0,160	-0,092	-0,047	-0,116	-0,079	1	-0,019	-0,058	-0,071	-0,082	-0,047	-0,090	-0,071	-0,051	-0,043	-0,071	1,000	-0,079	-0,061	-0,043	-0,047	-0,047	-0,027	-0,071
Zone-Amérique	-0,037	-0,052	-0,030	-0,015	-0,038	-0,026	-0,019	1	-0,019	-0,023	-0,027	-0,015	-0,029	-0,023	-0,017	-0,014	-0,023	-0,019	-0,026	-0,020	-0,014	-0,015	-0,015	0,705	-0,023
Organisation-SAARC	0,509	-0,160	-0,092	-0,047	-0,116	-0,079	-0,058	-0,019	1	-0,071	-0,082	-0,047	-0,090	-0,071	-0,051	-0,043	-0,071	-0,058	-0,079	-0,061	-0,043	-0,047	-0,047	-0,027	-0,071
Organisation-SADC	-0,139	0,443	-0,112	-0,057	-0,142	-0,096	-0,071	-0,023	-0,071	1	-0,100	-0,057	-0,109	-0,086	-0,062	-0,052	-0,086	-0,071	-0,096	-0,075	-0,052	-0,057	-0,057	-0,033	-0,086
Organisation-NC	0,046	-0,095	0,109	-0,066	-0,164	0,293	-0,082	-0,027	-0,082	-0,100	1	-0,066	-0,127	-0,100	-0,072	-0,060	-0,100	-0,082	-0,112	-0,087	-0,060	-0,066	-0,066	-0,038	-0,100
Organisation-Arab Maghreb Union	-0,092	0,294	-0,075	-0,038	-0,094	-0,064	-0,047	-0,015	-0,047	-0,057	-0,066	1	-0,073	-0,057	-0,041	-0,035	-0,057	-0,047	-0,064	-0,050	-0,035	-0,038	-0,038	-0,022	-0,057
Organisation-Union Européenne	-0,176	-0,247	0,638	0,523	-0,180	-0,123	-0,090	-0,029	-0,090	-0,109	-0,127	-0,073	1	-0,109	-0,079	-0,066	-0,109	-0,090	-0,123	-0,095	-0,066	-0,073	-0,073	-0,041	-0,109
Organisation-CARICOM	-0,139	-0,194	-0,112	-0,057	0,608	-0,096	-0,071	-0,023	-0,071	-0,086	-0,100	-0,057	-0,109	1	-0,062	-0,052	-0,086	-0,071	-0,096	-0,075	-0,052	-0,057	-0,057	-0,033	1
Organisation-GCC	-0,100	-0,140	-0,081	-0,041	-0,102	0,642	-0,051	-0,017	-0,051	-0,062	-0,072	-0,041	-0,079	-0,062	1	-0,037	-0,062	-0,051	-0,069	-0,054	-0,037	-0,041	-0,041	-0,023	-0,062
Organisation-MERCOSUR	-0,084	-0,117	-0,068	-0,035	0,367	-0,058	-0,043	-0,014	-0,043	-0,052	-0,060	-0,035	-0,066	-0,052	-0,037	1	-0,052	-0,043	-0,058	-0,045	-0,031	-0,035	-0,035	-0,020	-0,052
Organisation-Commonwealth of Independent States	0,270	-0,194	0,090	-0,057	-0,142	0,132	-0,071	-0,023	-0,071	-0,086	-0,100	-0,057	-0,109	-0,086	-0,062	-0,052	1	-0,071	-0,096	-0,075	-0,052	-0,057	-0,057	-0,033	-0,086
Organisation-Pacific Islands	-0,114	-0,160	-0,092	-0,047	-0,116	-0,079	1,000	-0,019	-0,058	-0,071	-0,082	-0,047	-0,090	-0,071	-0,051	-0,043	-0,071	1	-0,079	-0,061	-0,043	-0,047	-0,047	-0,027	-0,071
Organisation-ECOWAS	-0,156	0,496	-0,126	-0,064	-0,159	-0,108	-0,079	-0,026	-0,079	-0,096	-0,112	-0,064	-0,123	-0,096	-0,069	-0,058	-0,096	-0,079	1	-0,084	-0,058	-0,064	-0,064	-0,037	-0,096
Organisation-ASEAN	0,539	-0,169	-0,098	-0,050	-0,123	-0,084	-0,061	-0,020	-0,061	-0,075	-0,087	-0,050	-0,095	-0,075	-0,054	-0,045	-0,075	-0,061	-0,084	1	-0,045	-0,050	-0,050	-0,028	-0,075
Organisation-Andean Community	-0,084	-0,117	-0,068	-0,035	0,367	-0,058	-0,043	-0,014	-0,043	-0,052	-0,060	-0,035	-0,066	-0,052	-0,037	-0,031	-0,052	-0,043	-0,058	-0,045	1	-0,035	-0,035	-0,020	-0,052
Organisation-CEMAC	-0,092	0,294	-0,075	-0,038	-0,094	-0,064	-0,047	-0,015	-0,047	-0,057	-0,066	-0,038	-0,073	-0,057	-0,041	-0,035	-0,057	-0,047	-0,064	-0,050	-0,035	1	-0,038	-0,022	-0,057
Organisation-IGAD	-0,092	0,294	-0,075	-0,038	-0,094	-0,064	-0,047	-0,015	-0,047	-0,057	-0,066	-0,038	-0,073	-0,057	-0,041	-0,035	-0,057	-0,047	-0,064	-0,050	-0,035	-0,038	1	-0,022	-0,057
Organisation-NAFTA	-0,053	-0,074	-0,043	-0,022	0,088	-0,037	-0,027	0,705	-0,027	-0,033	-0,038	-0,022	-0,041	-0,033	-0,023	-0,020	-0,033	-0,027	-0,037	-0,028	-0,020	-0,022	-0,022	1	-0,033
Organisation-Central American Parliament	-0,100	-0,140	-0,081	-0,041	0,437	-0,069	-0,051	-0,017	-0,051	-0,062	-0,072	-0,041	-0,079	-0,062	-0,045	-0,037	-0,062	-0,051	-0,069	-0,054	-0,037	-0,041	-0,041	-0,023	-0,062

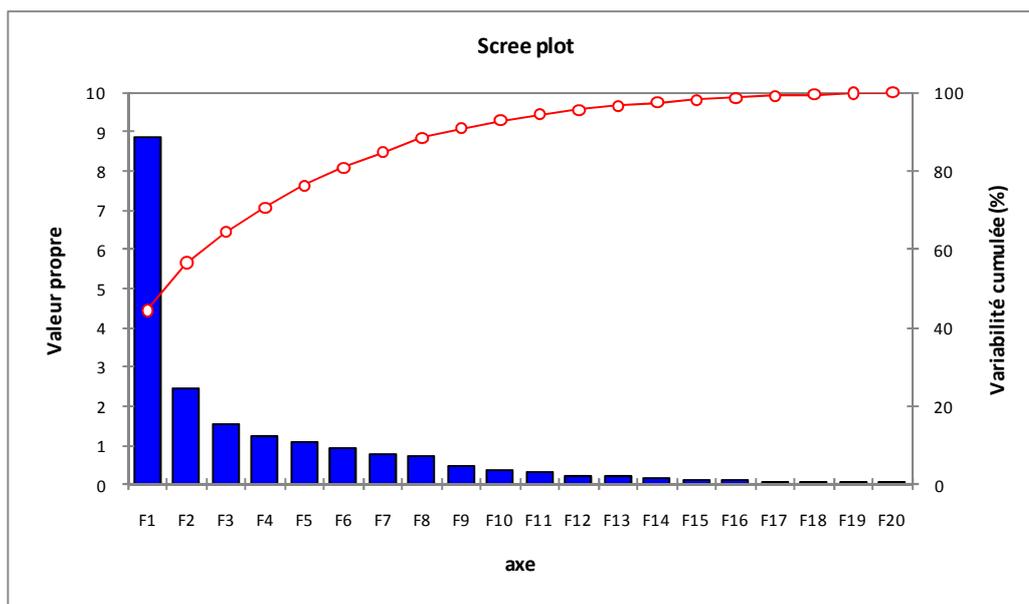
4.1.2.2. Axes

4.1.2.2.1. Scree Plot

Le tableau suivant indique pour chaque axe de F1 à 20, la valeur propre associée ainsi que le pourcentage de variabilité du modèle expliquée par cet axe. Les axes sont rangés par ordre de variabilité expliquée décroissant.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
Valeur propre	8,9	2,5	1,6	1,3	1,1	0,9	0,8	0,7	0,5	0,4	0,3	0,2	0,2	0,2	0,1	0,1	0,1	0,1	0,0	0,0
Variabilité (%)	44,4	12,3	7,8	6,3	5,5	4,6	3,9	3,7	2,5	2,0	1,6	1,2	1,0	0,9	0,7	0,5	0,5	0,4	0,2	0,1
% cumulé	44,4	56,7	64,4	70,7	76,3	80,9	84,8	88,5	90,9	92,9	94,5	95,7	96,7	97,5	98,2	98,7	99,2	99,6	99,9	100,0

Le Scree Plot donne une représentation graphique des valeurs ci-dessus.



On remarque que la variabilité cumulée obtenue avec les premiers axes n'est pas très élevée. Il faudrait garder au minimum 4 à 5 axes pour obtenir un niveau de fiabilité d'explication acceptable. Or l'utilisation de plus de 3 axes rendent les interprétations plus complexes et augmentent le risque d'erreur d'interprétation. On souhaiterait donc pouvoir réduire le nombre d'axes à conserver avant de construire les projections des observations.

4.1.2.2.2 Composition – Variables significatives

Le tableau suivant indique le pourcentage de contribution des variables pour chaque axe. Cela permet de mieux appréhender les variables qui sont les plus significatives dans le modèle, et ce que chaque axe représente.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
Population	0,02	3,74	12,87	33,47	13,581	0,101	0,196	5,879	6,237	2,280	3,736	11,374	2,045	0,021	3,064	0,167	1,017	0,028	0,154	0,022
Population Growth Rate	3,53	0,91	0,44	6,15	1,512	14,127	41,842	2,734	1,271	7,728	7,496	6,445	0,221	3,238	0,364	0,352	0,153	1,472	0,003	0,011
Mobile phone penetration	7,56	0,07	0,15	0,13	1,060	0,248	6,723	6,797	11,233	17,663	8,958	9,047	1,241	5,725	1,633	0,389	0,000	17,903	2,919	0,552
Internet penetration	8,57	0,52	0,02	0,70	1,776	2,468	4,133	0,267	3,820	1,673	0,588	0,000	0,012	54,319	5,690	6,639	0,530	6,156	0,226	1,892
GDP/capita	6,99	0,10	1,66	4,06	3,087	17,573	0,201	0,774	0,062	4,156	1,378	0,264	7,239	5,322	20,456	0,505	1,077	24,984	0,039	0,078
IDH	2,34	9,46	18,45	3,01	3,399	0,021	7,897	3,775	0,762	0,138	17,788	0,794	4,715	5,216	16,431	0,271	0,007	4,397	1,106	0,031
Real annual GDP Growth	0,38	21,15	1,11	1,23	0,757	0,467	3,410	27,105	23,093	10,161	0,534	1,781	7,386	0,187	0,124	0,406	0,053	0,093	0,155	0,416
Legal framework	9,69	0,12	0,09	3,00	2,201	0,197	0,127	0,101	0,192	1,030	5,801	0,089	2,213	0,327	0,103	0,129	12,731	0,036	61,186	0,635
Political stability	6,00	0,00	0,45	0,54	14,362	8,638	0,285	12,240	7,708	10,977	3,182	1,715	0,000	0,861	0,694	0,040	20,013	2,633	3,215	6,442
Control of Corruption	9,10	0,06	0,01	1,28	4,649	0,045	0,969	1,824	0,126	1,904	8,710	15,751	0,526	0,121	2,820	3,469	4,164	0,322	20,113	24,027
Corruption level	9,75	0,01	0,35	0,43	1,380	0,724	1,039	0,911	0,198	1,807	4,411	14,343	1,368	0,198	0,161	0,618	14,549	1,742	0,006	46,005
Economic Freedom Index	5,33	4,09	8,12	5,27	2,611	1,702	7,137	5,824	0,225	3,014	0,278	4,423	4,526	9,699	16,645	2,479	10,819	4,030	1,660	2,110
Vehicles per capita	8,15	0,22	0,82	1,86	7,301	3,614	0,919	0,639	0,130	1,276	7,871	3,213	0,023	0,034	14,904	43,556	0,422	4,715	0,139	0,185
Financial Risk	1,41	21,18	4,70	11,22	0,001	0,170	1,805	2,294	0,623	0,010	0,501	0,014	45,167	2,987	1,423	0,821	0,145	3,600	0,995	0,931
Economic Risk	0,07	22,53	4,85	6,74	0,997	4,694	0,204	12,892	11,991	0,505	2,825	1,557	21,160	6,439	0,109	0,945	0,111	1,100	0,257	0,017
Political Risk	8,33	1,53	0,08	0,30	5,671	3,130	0,426	1,225	7,540	3,679	2,569	4,587	0,131	0,231	0,454	0,733	32,243	4,023	7,242	15,878
Vehicle Ownership Projection	3,03	5,45	13,96	13,81	0,010	0,264	0,730	0,011	7,717	20,671	8,191	14,944	1,382	1,389	7,545	0,576	0,145	0,109	0,016	0,039
P&C spending/capita	6,33	1,32	3,28	5,88	11,975	0,062	4,706	3,607	0,730	0,904	4,239	3,447	0,550	2,604	1,731	34,926	1,261	11,722	0,010	0,718
Penetration rate	2,37	3,31	0,11	0,91	12,178	41,644	14,768	0,451	7,388	0,043	2,014	0,317	0,049	0,117	0,014	2,905	0,513	10,813	0,080	0,001
P&C spending trend	1,03	4,21	28,47	0,00	11,491	0,113	2,485	10,650	8,955	10,380	8,931	5,895	0,044	0,965	5,636	0,076	0,047	0,123	0,479	0,011

4.1.2.3. Choix des variables de la base réduite

A partir du tableau ci-dessus, on décide de ne prendre en compte que les variables qui semblent les plus significatives sur les axes F1 et F2.

	F1	F2
Population	0,02	3,74
Population Growth Rate	3,53	0,91
Mobile phone penetration	7,56	0,07
Internet penetration	8,57	0,52
GDP/capita	6,99	0,10
IDH	2,34	9,46
Real annual GDP Growth	0,38	21,15
Legal framework	9,69	0,12
Political stability	6,00	0,00
Control of Corruption WGI	9,10	0,06
Corruption level	9,75	0,01
Economic Freedom Index	5,33	4,09
Vehicles per capita	8,15	0,22
Financial Risk	1,41	21,18
Economic Risk	0,07	22,53
Political Risk	8,33	1,53
Vehicle Ownership Projection	3,03	5,45
P&C spending/capita	6,33	1,32
Penetration rate	2,37	3,31
P&C spending trend 5Y	1,03	4,21

On espère ainsi conserver seulement les variables qui apportent le plus d'information au modèle, et retirer celles qui peuvent constituer un bruit, afin d'optimiser le niveau de variabilité expliquée sur seulement 2 ou 3 axes.

On construit donc une base réduite à partir des variables :

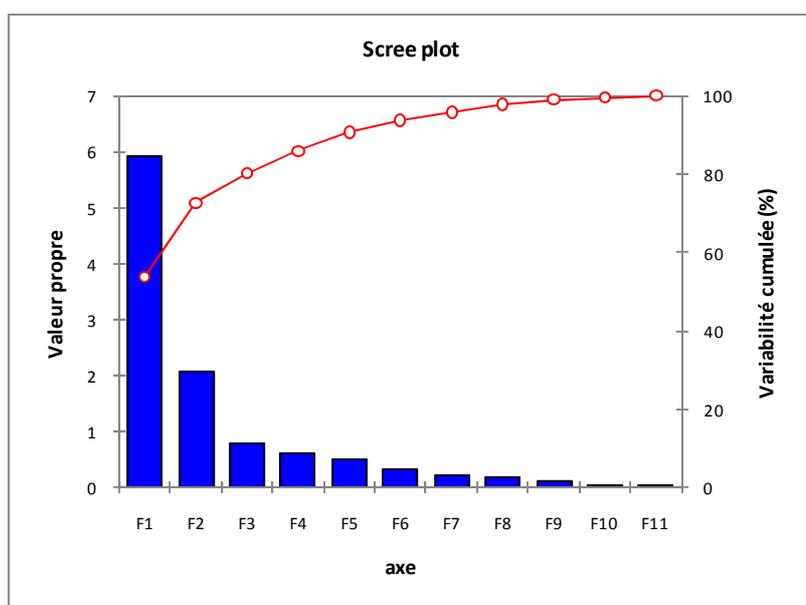
- Mobile phone penetration
- Internet penetration
- IDH
- Real annual GDP Growth
- Legal framework
- Control of Corruption
- Corruption level
- Vehicles per capita
- Financial Risk
- Economic Risk
- Political Risk

4.1.3. ACP sur base réduite

4.1.2.1. Axes

4.1.2.2.1. Scree Plot

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
Valeur propre	5,933	2,078	0,810	0,632	0,528	0,336	0,231	0,209	0,136	0,069	0,039
Variabilité (%)	53,934	18,888	7,366	5,748	4,796	3,052	2,101	1,903	1,235	0,626	0,351
% cumulé	53,934	72,822	80,189	85,936	90,732	93,784	95,885	97,787	99,023	99,649	100,000



En prenant en considération uniquement les axes F1, 2 et F3, on a maintenant une variabilité expliquée cumulée supérieure à 80%. On considère cette valeur comme suffisamment élevée pour construire des projections et analyser les regroupements de pays obtenus.

4.1.2.2.2 Variables significatives

Le tableau ci-dessous donne pour chacun des axes sa composition en fonction des 11 variables utilisées pour l'ACP.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
Mobile phone penetration	11,779	0,006	0,998	6,826	11,185	49,799	0,658	0,001	13,871	4,582	0,294
Internet penetration	13,086	1,861	0,603	1,282	13,051	0,084	3,331	7,742	57,712	0,002	1,245
IDH	4,197	5,736	47,604	35,825	0,546	2,988	1,177	1,877	0,008	0,029	0,012
Real annual GDP Growth	0,303	25,634	18,370	29,963	18,125	0,321	0,005	6,678	0,222	0,077	0,302
Legal framework	15,093	0,174	0,079	2,067	3,770	2,018	1,601	2,458	2,495	70,032	0,213
Control of Corruption	14,112	0,170	0,722	10,054	8,476	0,001	7,147	0,602	1,698	9,893	47,125
Corruption level	14,727	0,457	0,038	5,012	2,304	2,454	15,556	1,451	0,126	9,611	48,264
Vehicles per capita	11,433	0,352	4,821	0,786	17,555	42,140	4,899	1,833	15,332	0,009	0,841
Financial Risk	2,674	30,571	11,366	0,122	1,362	0,032	0,069	47,394	3,531	2,521	0,357
Economic Risk	0,039	34,343	14,633	6,680	9,988	0,055	3,645	29,007	0,312	1,297	0,001
Political Risk	12,556	0,697	0,766	1,380	13,638	0,106	61,911	0,958	4,692	1,947	1,348

On extrait les composantes des trois premiers axes, sur lesquels on souhaite projeter nos variables.

L'axe F1 est principalement construit à partir des variables Mobile phone penetration, Internet penetration, Legal Framework, Control of Corruption, Corruption, Vehicles per Capita et Political Risk.

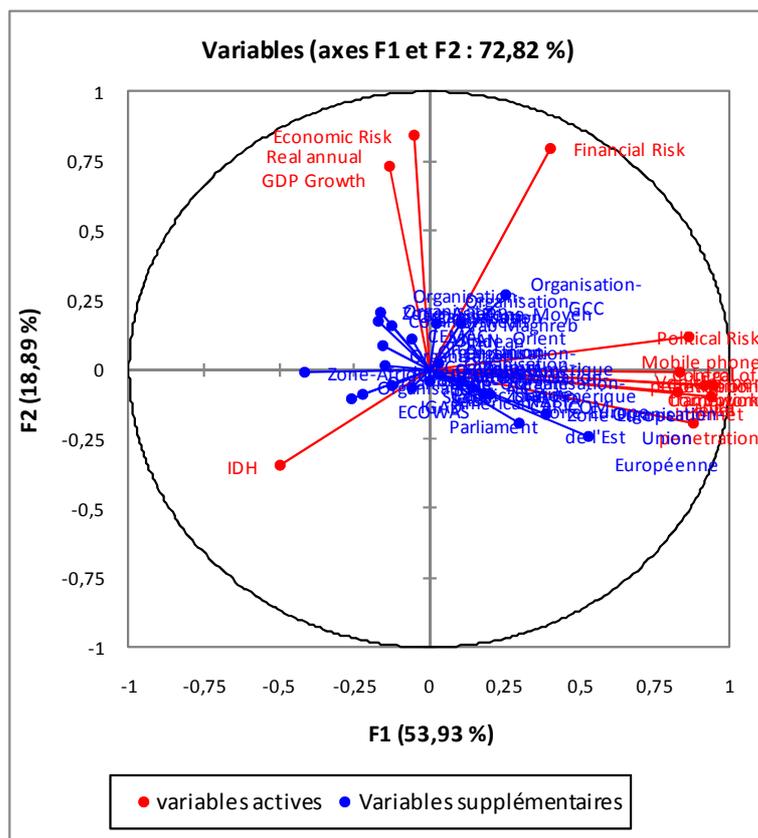
L'axe F2 est principalement construit à partir des variables Real annual GDP Growth, Financial Risk et Economic Risk.

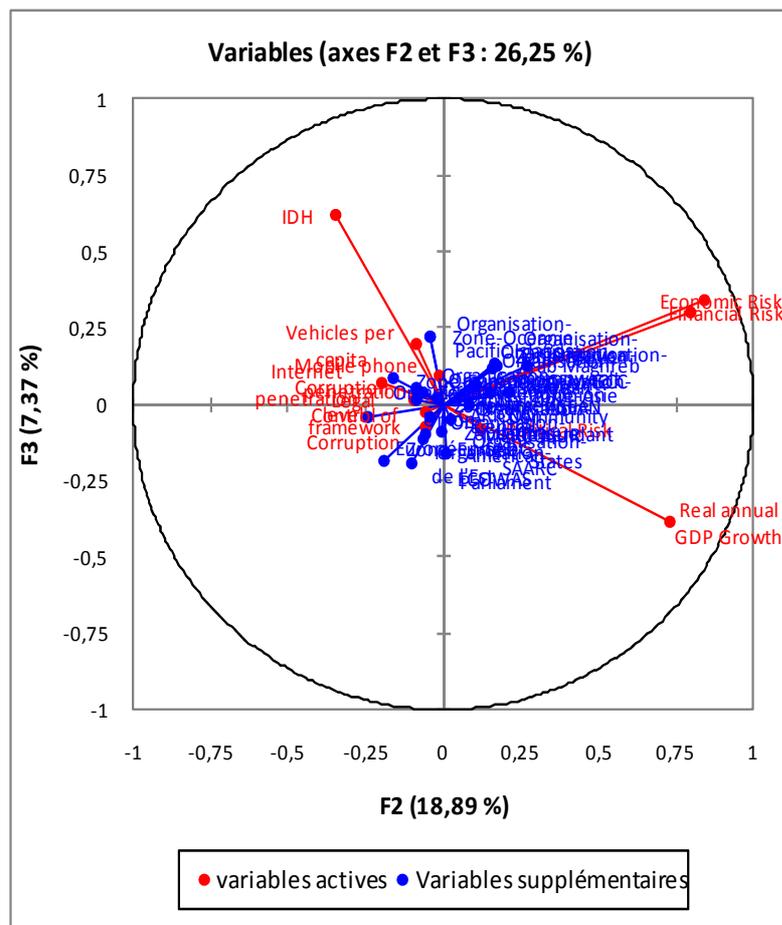
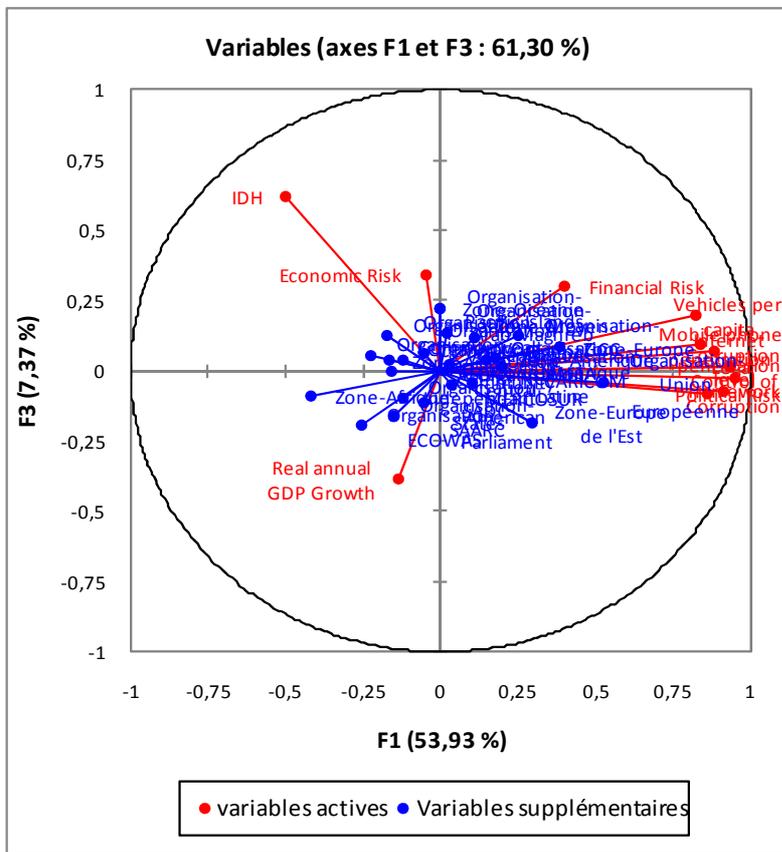
L'axe F3 est principalement construit à partir de la variable IDH, puis sur les mêmes variables significatives que l'axe F2 : Real annual GDP Growth, Financial Risk et Economic Risk.

On étudie également le tableau des corrélations entre les axes et les variables. Celui-ci va venir confirmer les corrélations qui sont les plus fortes, et nous indiquer le signe de ces corrélations, afin de déterminer si la contribution de chaque variable est positive ou négative.

	F1	F2	F3
Mobile phone penetration	0,836	-0,011	0,090
Internet penetration	0,881	-0,197	0,070
IDH	-0,499	-0,345	0,621
Real annual GDP Growth	-0,134	0,730	-0,386
Legal framework	0,946	-0,060	-0,025
Control of Corruption	0,915	-0,059	-0,077
Corruption level	0,935	-0,097	0,017
Vehicles per capita	0,824	-0,086	0,198
Financial Risk	0,398	0,797	0,303
Economic Risk	-0,048	0,845	0,344
Political Risk	0,863	0,120	-0,079

Ces valeurs de corrélations sont utilisées comme des coordonnées pour les projections des variables sur les axes F1, F2 et F3, deux à deux. On obtient les graphes suivants :





On peut interpréter l'axe F1 comme le niveau de qualité de l'environnement politico-légal. On voit bien sur la projection des corrélations que les variables que nous avons listées correspondent bien à la représentation de cet axe. On voit sur le premier graphe que la variable IDH est la plus inversement corrélée aux axes F1 et F2. Cela s'explique par le fait que cette variable est exprimée en classement, donc par valeurs décroissantes du moins développé au plus développé.

L'axe F2 indique dans quelle mesure le contexte économique est favorable dans un pays. En effet, on voit que les variables Economic Risk, Financial Risk et Real Annual GDP Growth sont les plus corrélées positivement à cet axe, alors que le rang IDH lui est corrélé négativement.

L'axe F3 nous donne donc une indication supplémentaire pour la projection : en effet, il est le plus fortement corrélé au rang IDH, c'est-à-dire qu'il va d'abord représenter une échelle du niveau de développement du plus élevé au plus faible par pays. On voit que cet axe est aussi négativement corrélé à la variable Real annual GDP Growth, et faiblement positivement corrélé aux variables Economic Risk et Financial Risk. Ainsi, l'axe F3 représente à une extrémité les pays faiblement développés (au sens de l'IDH), relativement stables économiquement, avec de faibles perspectives de croissance, et de l'autre des pays plus développés avec des perspectives de croissance importantes. Il représente donc le potentiel de croissance par pays.

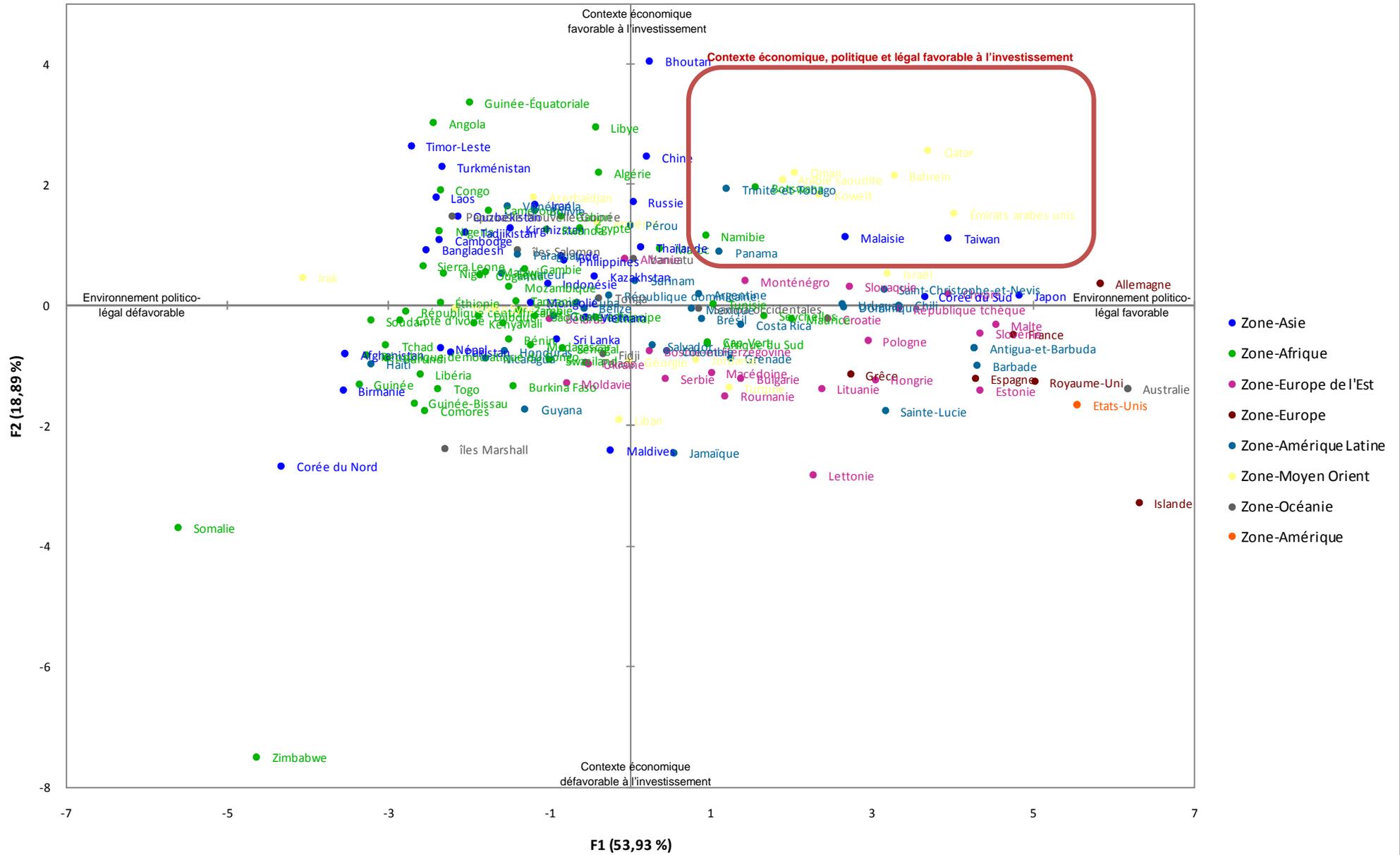
4.1.2.2. Représentation

A partir des trois axes décrits précédemment, on construit les coordonnées de chaque observation-pays et on peut ainsi projeter les données sur deux dimensions. On construit trois graphiques : F1 et F2, F1 et F3, F2 et F3.

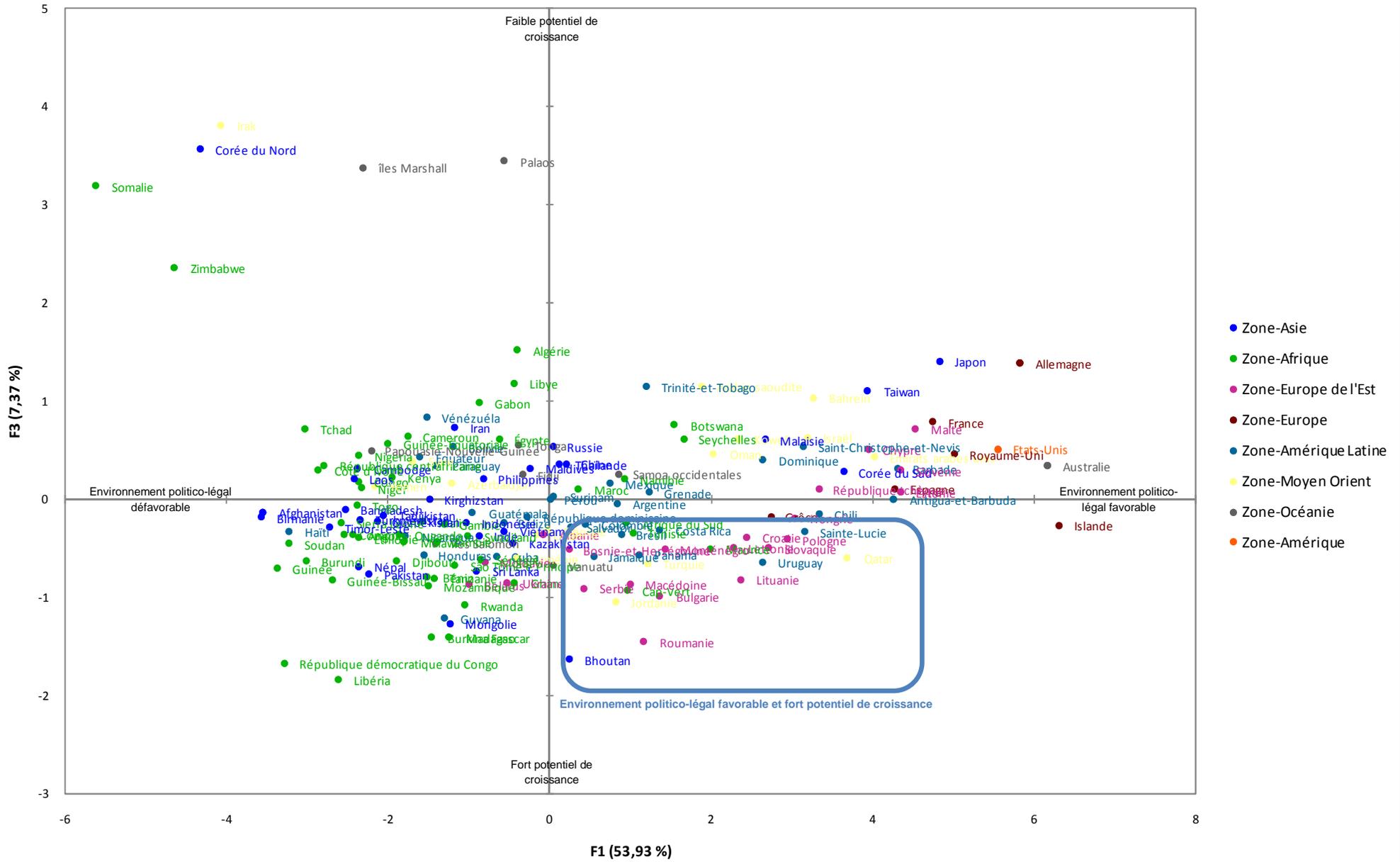
On repère ainsi des regroupements de pays et on cherche à déterminer lesquels représentent une opportunité d'investissement en assurance.

Les observations ont été colorées selon leur zone géographique. Cela permet de visualiser directement si des zones semblent particulièrement propices à l'investissement, et quels pays représentent des marchés émergents en assurance.

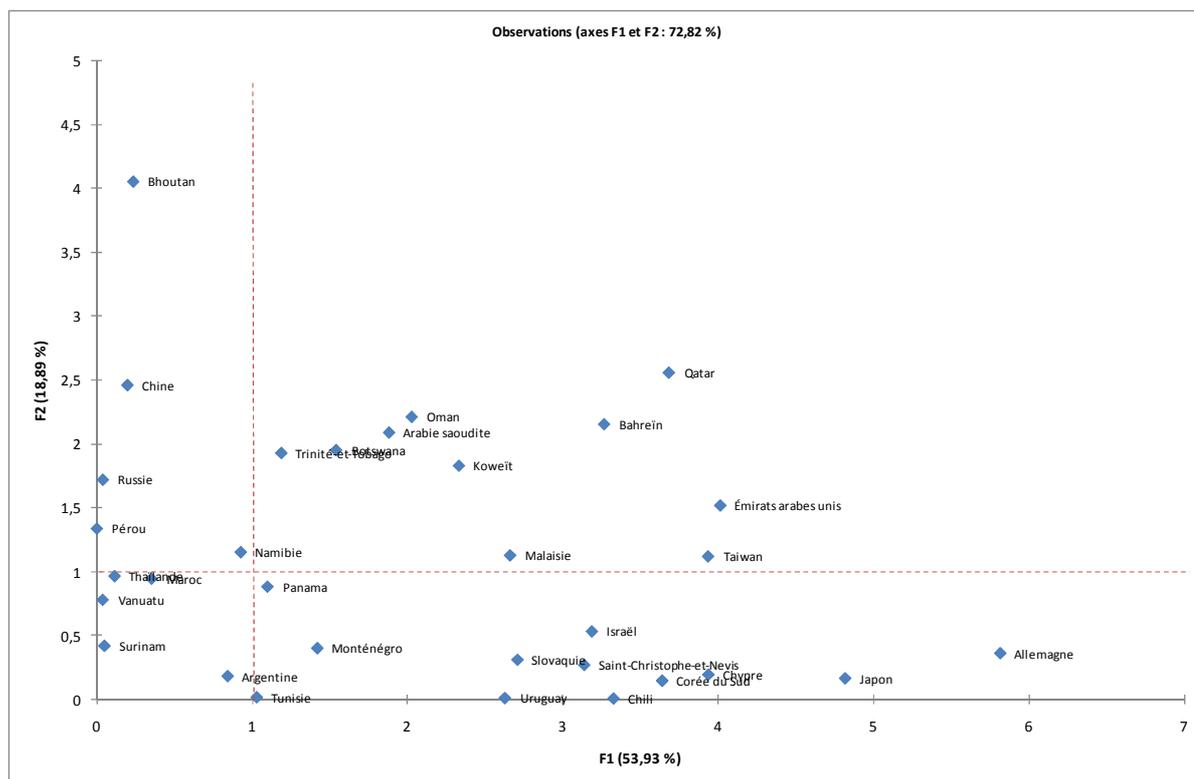
Observations (axes F1 et F2 : 72,82 %)



Observations (axes F1 et F3 : 61,30 %)



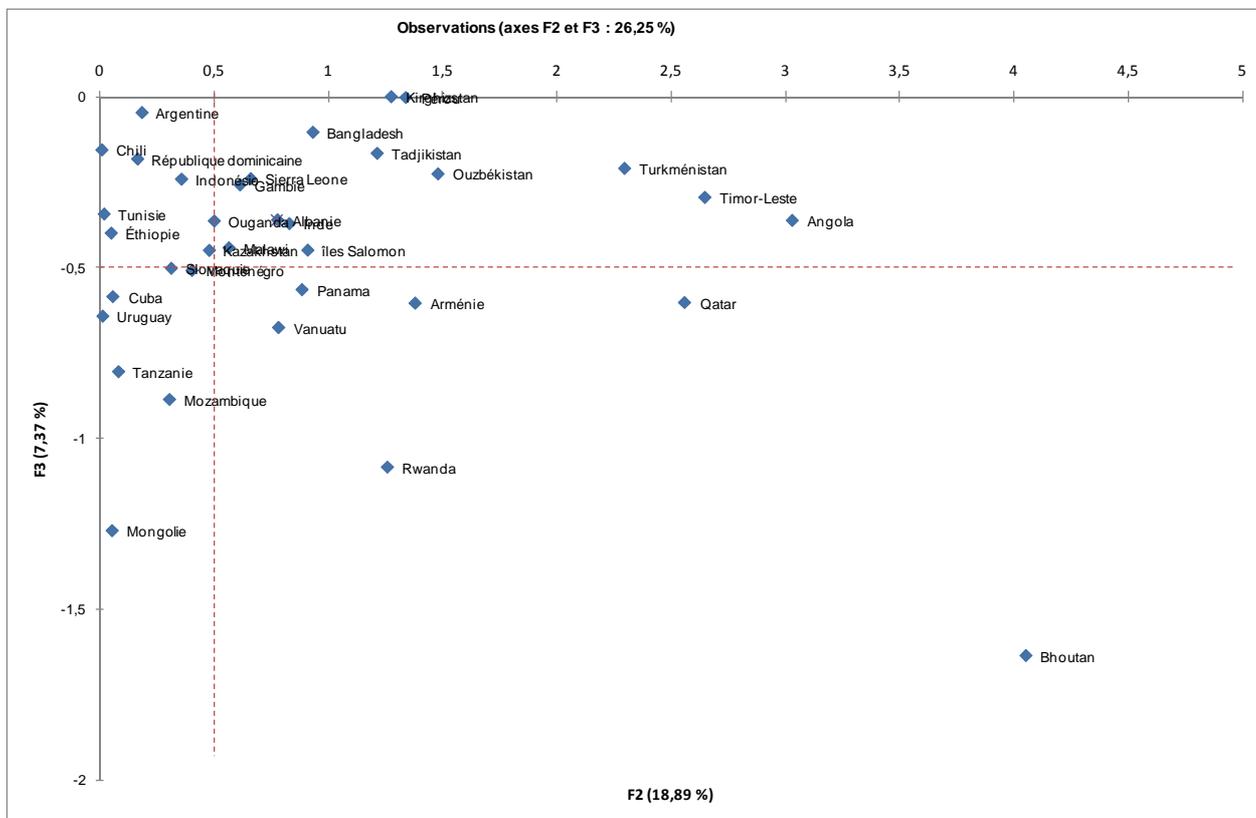
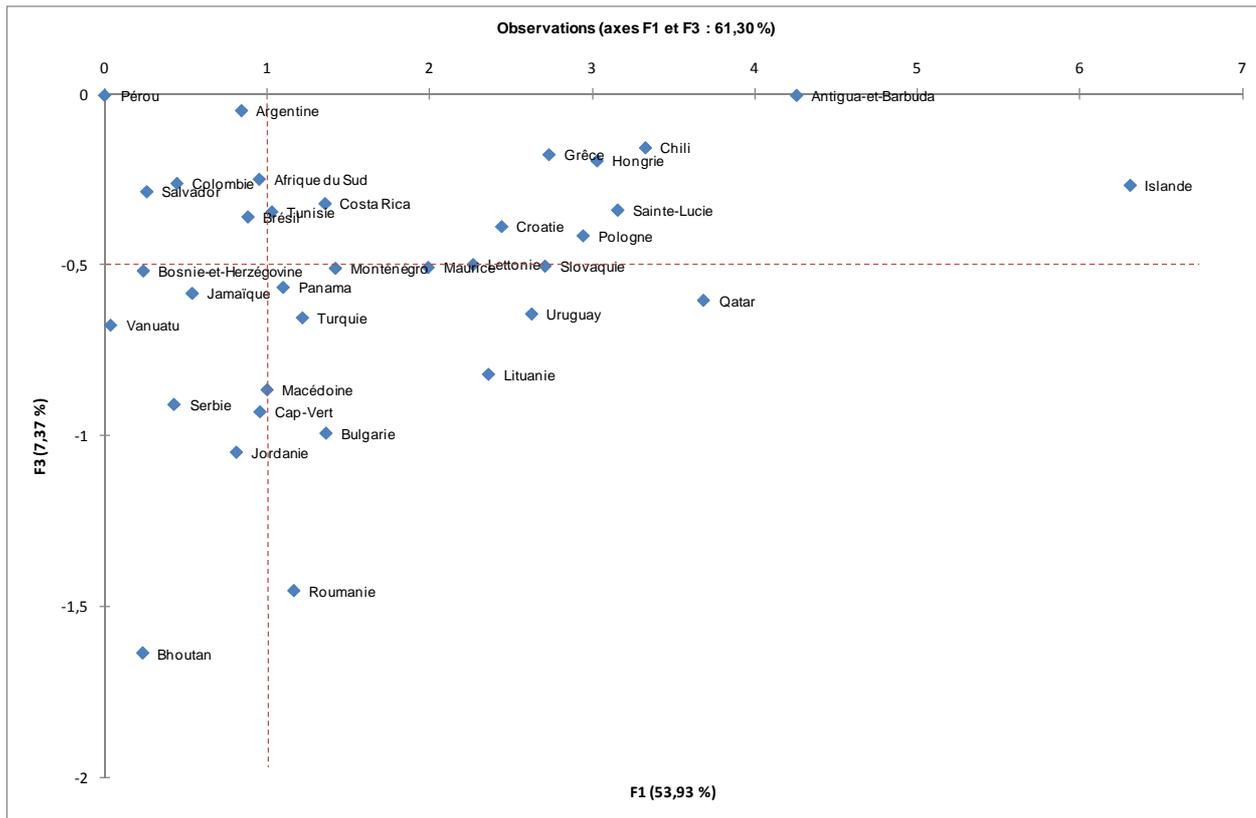
En effectuant un zoom sur les portions de graphiques qui regroupent les pays qui semblent représenter des opportunités d'investissement, on obtient les résultats suivants :



Si on se limite par exemple aux observations dont les coordonnées sur F1 et F2 sont toutes les deux supérieures à 1, on retient les pays :

- Trinité-et-Tobago
- Botswana
- Oman
- Arabie Saoudite
- Koweït
- Qatar
- Bahreïn
- Malaisie
- Taiwan
- Emirats Arabes Unis

En utilisant la même méthode sur les deux autres projections, on obtient les résultats suivants :



Au final, lorsqu'on retient uniquement les pays qui sont représentés dans les trois portions de graphes précédentes ($x_{F1} > 0$; $x_{F2} > 0$; $x_{F3} < 0$), on obtient la liste de pays suivante :

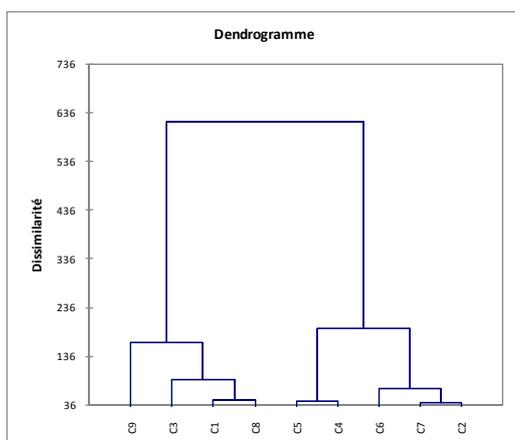
Observation	F1	F2	F3
Argentine	0,844	0,182	-0,048
Bhoutan	0,236	4,053	-1,638
Chili	3,327	0,006	-0,157
Monténégro	1,421	0,401	-0,510
Panama	1,100	0,883	-0,566
Pérou	0,001	1,337	-0,003
Qatar	3,683	2,558	-0,604
Slovaquie	2,710	0,310	-0,504
Tunisie	1,031	0,017	-0,345
Uruguay	2,628	0,010	-0,644
Vanuatu	0,039	0,780	-0,677

Cela peut constituer une première sélection de pays représentant des marchés émergents en assurance, où le contexte d'investissement semble particulièrement favorable.

4.1.4. Visualisation des regroupements par pays

Utilisation des résultats de l'ACP dans une CAH : en utilisant le tableau des coordonnées des observations-pays sur chacun des 10 axes retenus, on construit une classification ascendante hiérarchique selon la même méthodologie que dans l'étude préalable de données. La base d'analyse présente cependant une différence importante, puisqu'il s'agit d'un tableau de données et non plus d'une matrice de similarité construite sur des données binaires. On utilise cette fois-ci la méthode d'agrégation de Ward, avec un indice de dissimilarité par distance euclidienne. Cette méthode consiste à réunir les deux classes d'observations dont le regroupement fera le moins baisser l'inertie interclasse. C'est la distance de Ward qui est utilisée pour cela : la distance entre deux classes est celle de leurs barycentres au carré, pondérée par les effectifs des deux clusters. Cela aura tendance à regrouper ensemble les petites classes.

Les regroupements en eux-mêmes n'apportent pas de réelle nouveauté par rapport aux résultats obtenus lors de l'Analyse en Composantes Principales, mais permettent d'obtenir une visualisation des groupes différents, et notamment du niveau auquel les groupes se séparent les uns des autres. On retient un dendrogramme à 9 classes :



Les pays sont répartis dans les classes correspondantes dans le tableau ci-après (Tableau 9).

Tableau 9 : Répartition des pays dans les classes construites par CAH

Classe	1	2	3	4	5	6	7	8	9
Objets	27	29	21	11	18	9	13	30	6
Variance intra-classe	2,435	3,201	2,744	3,399	2,983	6,991	3,181	2,851	13,834
Distance minimale au barycentre	0,789	0,870	0,882	0,934	1,108	1,249	0,572	0,705	1,260
Distance moyenne au barycentre	1,484	1,688	1,556	1,636	1,648	2,293	1,595	1,573	3,165
Distance maximale au barycentre	2,387	2,950	2,260	3,344	2,243	4,615	2,720	3,357	5,058
	Afghanistan	Afrique du Sud	Algérie	Allemagne	Antigua-et-Barbuda	Arabie saoudite	Argentine	Belarus	Corée du Nord
	Bangladesh	Albanie	Angola	Australie	Barbade	Bahreïn	Bosnie-et-Herzégovine	Bénin	îles Marshall
	Birmanie	Arménie	Azerbaïdjan	Chypre	Chili	Bhoutan	Bulgarie	Burkina Faso	Irak
	Burundi	Bélieze	Bolivie	Espagne	Corée du Sud	Botswana	Colombie	Cuba	Palaos
	Comores	Brésil	Cambodge	Etats-Unis	Dominique	Émirats arabes unis	Croatie	Fidji	Somalie
	Côte d'Ivoire	Cap-Vert	Cameroun	France	Estonie	Koweït	Jamaïque	Gambie	Zimbabwe
	Djibouti	Chine	Congo	Islande	Grèce	Oman	Lettonie	Ghana	
	Éthiopie	Costa Rica	Équateur	Japon	Hongrie	Qatar	Liban	Guatémala	
	Guinée	Égypte	Gabon	Malte	Israël	Trinité-et-Tobago	Macédoine	Guyana	
	Guinée-Bissau	Géorgie	Guinée-Équatoriale	Royaume-Uni	Lituanie		Monténégro	Honduras	
	Haïti	Grenade	Iran	Slovénie	Malaisie		Roumanie	îles Salomon	
	Kenya	Indonésie	Kirghizstan		Pologne		Serbie	Inde	
	Libéria	Jordanie	Laos		République tchèque		Turquie	Kazakhstan	
	Malawi	Maroc	Libye		Saint-Christophe-et-Nevis			Madagascar	
	Népal	Maurice	Ouzbékistan		Sainte-Lucie			Maldives	
	Niger	Mexique	Papouasie-Nouvelle-Guinée		Slovaquie			Mali	
	Nigeria	Namibie	Paraguay		Taiwan			Moldavie	
	Ouganda	Panama	Tadjikistan		Uruguay			Mongolie	
	Pakistan	Pérou	Timor-Leste					Mozambique	
	République centrafricaine	Philippines	Turkménistan					Nicaragua	
	République démocratique du Congo	République dominicaine	Vénézuéla					Rwanda	
	Sierra Leone	Russie						São Tomé e Príncipe	
	Soudan	Salvador						Sénégal	
	Syrie	Samoa occidentales						Sri Lanka	
	Tchad	Seychelles						Swaziland	
	Togo	Surinam						Tanzanie	
	Yémen	Thaïlande						Tonga	
		Tunisie						Ukraine	
		Vanuatu						Vietnam	
								Zambie	

4.1.5. Caractéristiques des classes

Le tableau ci-dessous représente les coordonnées des barycentres de chaque classe en fonction des axes F1 à F11 :

Classe	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
1	-2,603	-0,406	-0,306	-0,301	0,071	-0,336	-0,271	0,110	0,012	0,016	0,041
2	0,405	0,334	-0,131	0,043	-0,567	0,255	0,061	0,033	0,015	0,094	-0,052
3	-1,721	1,802	0,370	-0,479	0,295	-0,150	0,087	0,000	0,113	-0,113	0,078
4	5,042	-0,850	0,564	-0,102	0,206	-1,074	0,032	-0,157	-0,193	-0,010	0,106
5	3,240	-0,323	0,041	0,029	0,150	0,352	-0,215	0,060	0,263	0,017	0,040
6	2,242	2,255	0,370	0,581	-0,039	0,252	-0,143	0,091	-0,338	-0,099	-0,081
7	1,018	-1,135	-0,619	-0,220	0,755	0,267	0,127	-0,151	0,062	0,072	-0,172
8	-1,033	-0,342	-0,588	0,127	-0,290	0,078	0,198	-0,027	-0,118	-0,028	-0,001
9	-3,585	-2,782	3,287	1,890	0,432	0,372	0,160	-0,218	0,002	-0,031	-0,024

Le barycentre de la classe 1 a pour coordonnées (-2,6 ; -0,4 ; -0,3) sur les trois premiers axes. Cela correspond à un environnement politico-légal très défavorable, un contexte économique légèrement défavorable à l'investissement, un potentiel de croissance légèrement positif.

On exclue donc les pays de la classe 1 de notre sélection d'opportunités.

Le barycentre de la classe 2 a pour coordonnées (0,4 ; 0,3 ; -0,1) sur les trois premiers axes. Cela correspond à un environnement politico-légal légèrement favorable, un contexte économique légèrement favorable à l'investissement, un potentiel de croissance légèrement positif. La classe 2 peut sembler acceptable pour des investissements mais les pays qu'elle contient ne constituent pas de vraies opportunités.

On exclue donc les pays de la classe 2 de notre sélection d'opportunités.

Le barycentre de la classe 3 a pour coordonnées (-1,7 ; 1,8 ; 0,3) sur les trois premiers axes. Cela correspond à un environnement politico-légal très défavorable, un contexte économique très favorable à l'investissement, un potentiel de croissance légèrement négatif. Les pays de la classe 3 semblent attractifs pour un investissement, mais le risque politico-légal associé est trop élevé pour qu'ils constituent une opportunité d'investissement.

On exclue donc les pays de la classe 3 de notre sélection d'opportunités.

Le barycentre de la classe 4 a pour coordonnées (5,0 ; -0,9 ; 0,6) sur les trois premiers axes. Cela correspond à un environnement politico-légal très favorable, un contexte économique très défavorable à l'investissement, un potentiel de croissance fortement négatif. Les pays de la classe 4 offrent un contexte politico-légal particulièrement stable, mais ne sont pas attractifs économiquement et financièrement. Il s'agit en fait de nos pays-témoins où l'assurance est déjà mature.

On exclue donc les pays de la classe 4 de notre sélection d'opportunités.

Le barycentre de la classe 5 a pour coordonnées (3,2 ; -0,3 ; 0) sur les trois premiers axes. Cela correspond à un environnement politico-légal très favorable, un contexte économique très légèrement défavorable à l'investissement, un potentiel de croissance neutre. Les pays de la classe 5 offrent un contexte politico-légal stable, et sont relativement attractifs économiquement et financièrement.

On inclue donc les pays de la classe 5 de notre sélection d'opportunités.

Le barycentre de la classe 6 a pour coordonnées (2,2 ; 2,3 ; 0,4) sur les trois premiers axes. Cela correspond à un environnement politico-légal très favorable, un contexte économique très favorable à l'investissement, un potentiel de croissance légèrement défavorable. Les pays de la classe 6 offrent un contexte politico-légal stable, et sont très attractifs économiquement et financièrement.

On inclue donc les pays de la classe 6 de notre sélection d'opportunités.

Le barycentre de la classe 7 a pour coordonnées (1 ; -1,1 ; -0,6) sur les trois premiers axes. Cela correspond à un environnement politico-légal très favorable, un contexte économique défavorable à l'investissement, un potentiel de croissance relativement élevé. Les pays de la classe 7 offrent un contexte politico-légal relativement stable, et un potentiel de croissance qui compense le contexte économique.

On inclue donc les pays de la classe 7 de notre sélection d'opportunités.

Le barycentre de la classe 8 a pour coordonnées (-1 ; -0,3 ; -0,5) sur les trois premiers axes. Cela correspond à un environnement politico-légal défavorable, un contexte économique neutre, un potentiel de croissance relativement élevé. Les pays de la classe 8 ne proposent pas un environnement suffisamment favorable pour être attractifs pour un investissement.

On exclue donc les pays de la classe 8 de notre sélection d'opportunités.

Le barycentre de la classe 9 a pour coordonnées (-3,6 ; -2,8 ; 3,3) sur les trois premiers axes. Cela correspond à un environnement politico-légal très défavorable, un contexte économique très défavorable à l'investissement, un potentiel de croissance très faible.

On exclue donc les pays de la classe 8 de notre sélection d'opportunités.

4.1.6. Résultats

On retient, au final :

En premier lieu, les pays de la classe 6 qui semblent les plus attractifs :

- Arabie saoudite
- Bahreïn
- Bhoutan
- Botswana
- Émirats arabes unis
- Koweït
- Oman
- Qatar
- Trinité-et-Tobago

Dans un second temps, les pays des classes 5 et 7, qui correspondent à des opportunités d'investissement supplémentaires:

Antigua-et-Barbuda
 Barbade
 Chili
 Corée du Sud
 Dominique
 Estonie
 Grèce
 Hongrie
 Israël
 Lituanie
 Malaisie
 Pologne
 République tchèque
 Saint-Christophe-et-Nevis
 Sainte-Lucie
 Slovaquie
 Taiwan
 Uruguay

Argentine
 Bosnie-et-Herzégovine
 Bulgarie
 Colombie
 Croatie
 Jamaïque
 Lettonie
 Liban
 Macédoine
 Monténégro
 Roumanie
 Serbie
 Turquie

4.2. Construction du modèle de notation

L'application de l'ACP normée utilisée précédemment nous donne une sélection de pays construite selon une optimisation purement mathématique. EMINCAP souhaite en plus se doter d'un outil de notation linéaire, utilisant la même base de données, mais en laissant la possibilité de modifier les pondérations des variables et les règles décisionnelles de seuil.

4.2.1. Construction des Clusters

Les données de la base sont maintenant complétées, mais elles ne sont pas comparables entre elles : comment travailler linéairement en associant une taille de population, un taux de croissance du PIB ou un taux de pénétration internet ?

Afin d'obtenir un modèle simple à comprendre et à utiliser, on souhaite affecter à chaque valeur de la base une note sur une échelle de 0 à 10. Etant donné que l'on travaille sur un modèle linéaire, on construit les classes de notation également de façon linéaire, à partir des rangs des valeurs aux centiles 10%, 20%, ..., 100%. Ainsi les 10% de valeurs les moins élevées d'une variable seront affectées de la note 1. Le minimum de la variable pourra obtenir la note 0. La règle de notation complète est décrite dans le tableau ci-dessous.

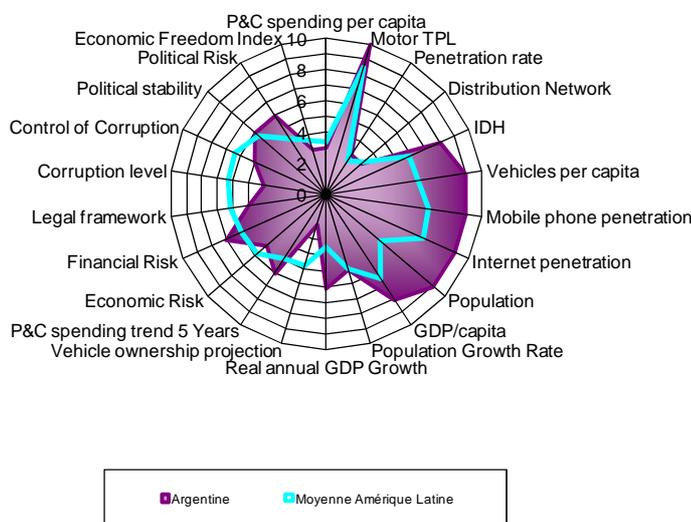
$0 < \text{Centile } (x) < 10$	Score = 1
$10 < \text{Centile } (x) < 20$	Score = 2
$20 < \text{Centile } (x) < 30$	Score = 3
$30 < \text{Centile } (x) < 40$	Score = 4
$40 < \text{Centile } (x) < 50$	Score = 5
$50 < \text{Centile } (x) < 60$	Score = 6
$60 < \text{Centile } (x) < 70$	Score = 7
$70 < \text{Centile } (x) < 80$	Score = 8
$80 < \text{Centile } (x) < 90$	Score = 9
$90 < \text{Centile } (x) < 100$	Score = 10

Pour la variable IDH et les variables assurantielles P&C Spending per Capita, Penetration Rate, les scores sont inversés : pour la première, on a vu que les valeurs étaient sous forme de rang, donc classées par ordre contraire aux autres variables. Pour les deux variables assurantielles, il s'agit ici de détecter la maturité du marché assurance sur un pays. Les pays où les taux de pénétration et de dépenses par habitant sont élevés ont dépassé le cap de la zone d'émergence en assurance.

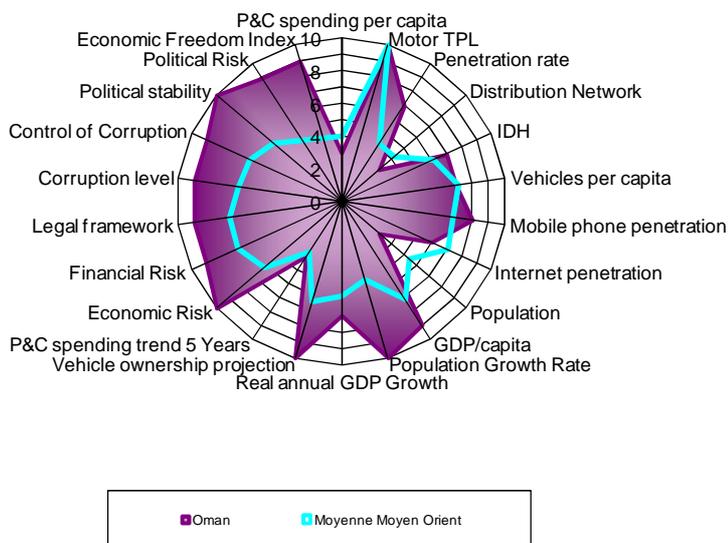
4.2.2. Représentation graphique par pays

On construit une représentation graphique dynamique par pays, sous la forme d'un radar qui vient indiquer le score du pays pour chaque variable, ainsi que la moyenne de la zone géographique associée sur cette même variable. Cela permet de visualiser directement si le pays donne une représentation graphique équilibrée, s'il semble être une opportunité d'investissement, quels sont ses points forts et ses points faibles dans le processus décisionnel d'EMINCAP. En voici quelques exemples :

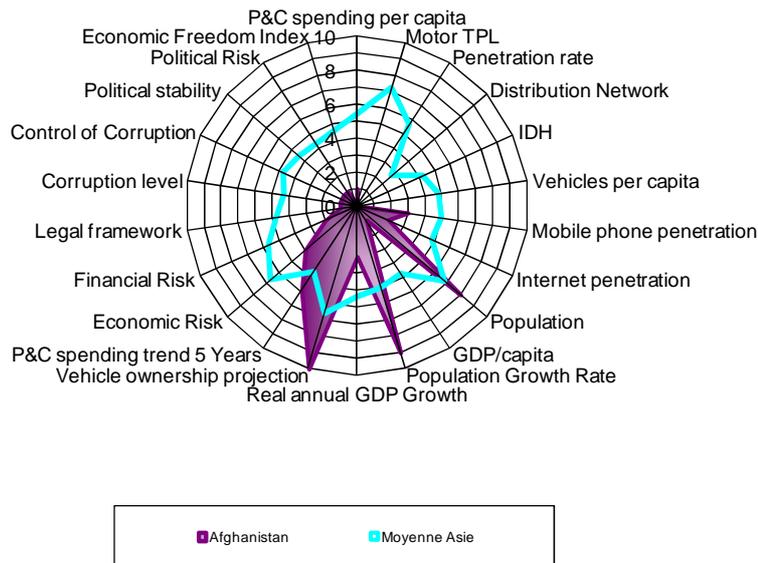
Argentine



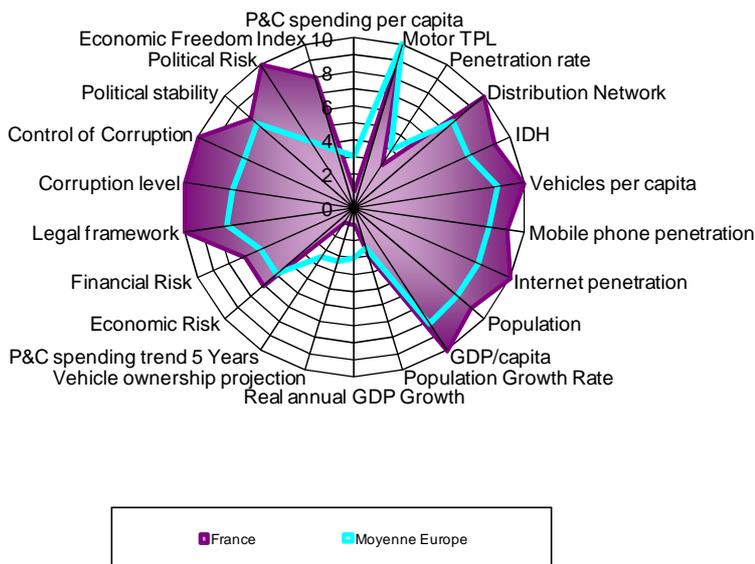
Oman



Afghanistan



France

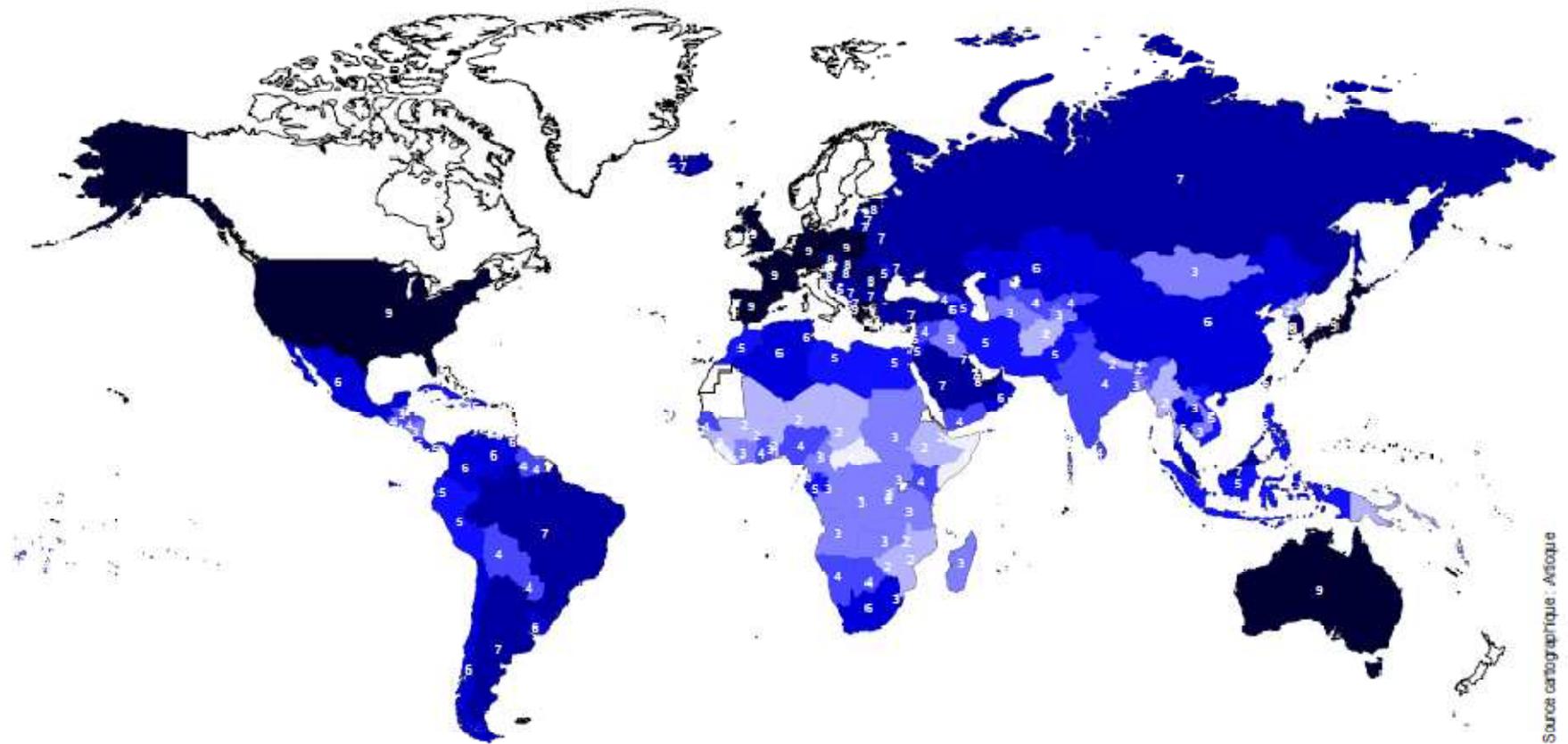


Le processus de notation revient à sélectionner les pays dont le radar est le plus équilibré et le plus couvert en violet. La ligne bleue correspond à la valeur moyenne pour la zone géographique du pays.

Le logiciel Cartes et Données permet à partir de nos quatre agrégats (tels que présentés en partie 1.2 : Insurance, Economic, Outlook et Political & Legal) de construire des cartes du monde colorées selon une échelle basée sur les moyennes de score par agrégat. Les pays ayant obtenu le meilleur score (10) sur la moyenne de chaque agrégat seront colorés en bleu foncé, ceux obtenant le moins bon score en bleu clair ; les pays non étudiés ne sont pas colorés.

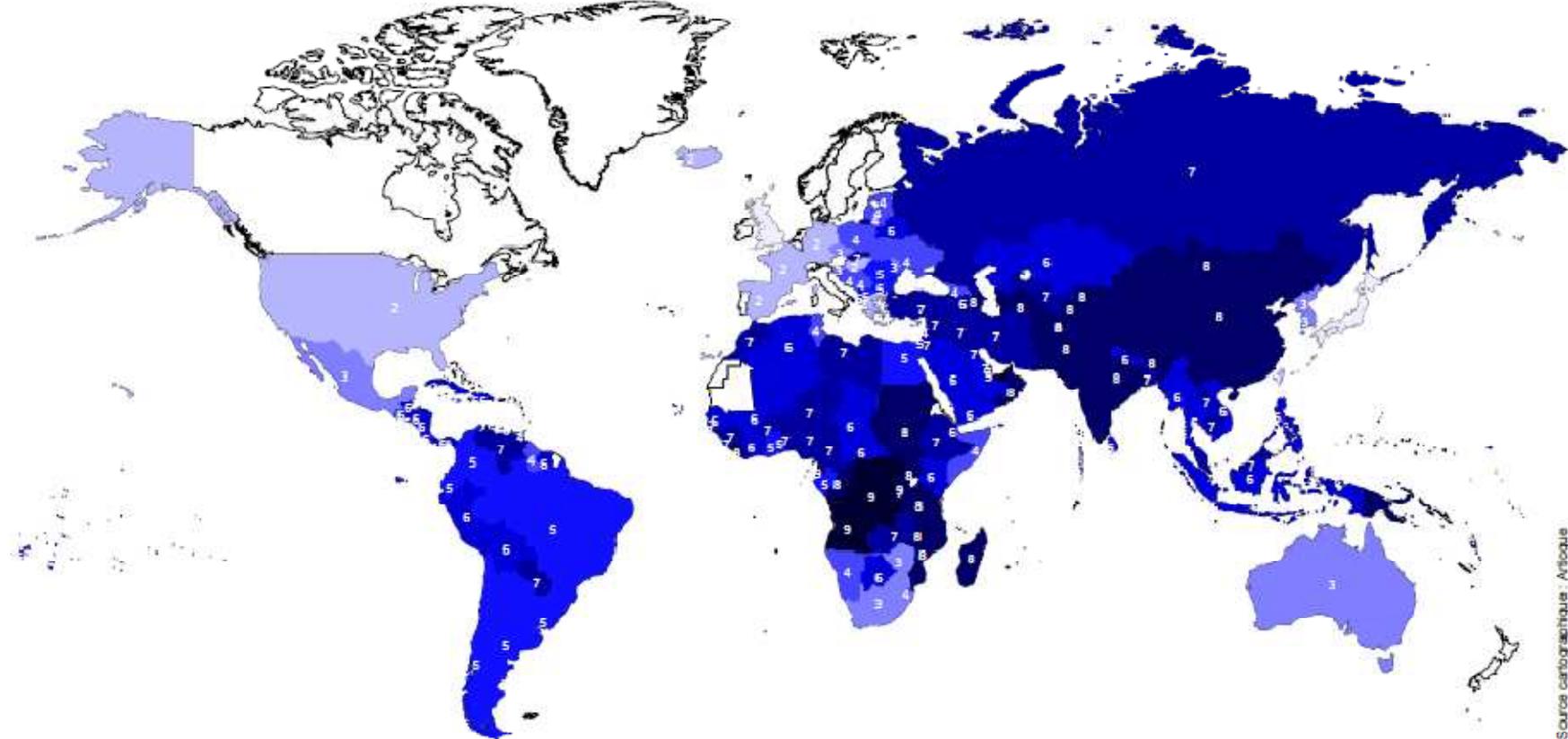
On obtient les résultats suivants :

Economic



Source cartographique : Astoquo

Outlook



4.2.3. Règles de décision

On souhaite classifier les pays sur la base d'une unique note qui permettra de décider du caractère attractif de ce pays pour un investissement en assurance.

Faire la moyenne de l'ensemble des coefficients présente un inconvénient majeur : cela a tendance à « tasser » les notations, et à affecter une note relativement bonne à des pays dont les notes par coefficients sont très hétérogènes. La construction d'une moyenne pondérée des quatre agrégats, mettant en avant les notes Insurance et Outlook (coefficientées 2 contre 1 pour Economic et Political & Legal) donne des résultats plus cohérents, sans toutefois régler entièrement la problématique d'erreur de sélection, par exemple d'un pays qui offre un fort potentiel de croissance mais présente un risque politique et économique très élevé.

Une autre méthode simple consiste à filtrer les éléments selon leur score par agrégat, en appliquant par exemple les règles suivantes :

Insurance Overview	Economic Overview	Outlook	Political & Legal Overview
> 5	> 5	> 5	> 5

4.2.4. Résultats

En appliquant la méthode de sélection par moyenne pondérée avec les coefficients fixés ci-dessus, on obtient la liste de pays suivante, obtenant une moyenne supérieure à 6/10 :

Émirats arabes unis	Inde
Oman	Malaisie
Arabie saoudite	Pérou
Koweït	Qatar
Mongolie	Tanzanie
Azerbaïdjan	Ouganda
Madagascar	Pakistan
Chine	Bahreïn
Rwanda	République démocratique du
Arménie	Congo
Lituanie	Vietnam

En appliquant la méthode de sélection par règles sur les agrégats on retient les pays suivants :

Arabie saoudite	Malaisie
Arménie	Maroc
Bahreïn	Oman
Chine	Panama
Émirats arabes unis	Pérou
Jordanie	Qatar
Kazakhstan	République dominicaine
Koweït	Slovaquie
Libye	Turquie

4.3. Synthèse des pays sélectionnés

Les différentes méthodes de modélisation utilisées proposent des sélections de pays relativement cohérentes. Le tableau ci-dessous permet de visualiser l'ensemble des pays choisis, par quelle méthode et quelles sont leurs caractéristiques propres.

Pays	Sélection par :				Scores - Aggrégats				Moyenne Pondérée
	ACP	CAH	Moyenne Pondérée	Règles Aggrégats	Insurance	Economic	Outlook	Political & Legal	
Arabie saoudite	✓	✓	✓	✓	7,33	7,86	5,75	7,50	6,92
Bahreïn	✓	✓	✓	✓	5,67	7,14	6,00	8,38	6,48
Émirats Arabes Unis	✓	✓	✓	✓	5,00	8,14	8,75	9,13	7,46
Koweït	✓	✓	✓	✓	7,00	7,00	6,00	8,50	6,92
Oman	✓	✓	✓	✓	6,67	6,14	7,75	9,25	7,37
Qatar	✓	✓	✓	✓	5,67	7,43	6,00	8,88	6,61
Malaisie	✓	✓	✓	✓	6,33	7,43	6,00	8,13	6,70
Pérou	✓	✗	✓	✓	8,00	5,57	5,75	6,63	6,62
Slovaquie	✓	!	✗	✓	5,33	8,00	6,00	8,00	6,44
Panama	✓	✗	✗	✓	5,00	5,86	5,75	6,88	5,71
Arménie	✗	✗	✓	✓	8,33	6,00	5,75	6,25	6,74
Chine	✗	✓	✓	✓	6,67	6,14	7,50	6,50	6,83
Trinité-et-Tobago	✓	✓	✗	✗	5,33	6,00	3,00	8,00	5,11
Bhoutan	✓	✓	✗	✗	6,33	2,86	7,00	8,00	6,25
Botswana	✓	✓	✗	✗	4,33	4,14	5,50	9,38	5,53
Turquie	✗	!	✗	✓	7,00	7,29	6,00	5,38	6,44
Monténégro	✓	!	✗	✗	4,67	7,14	5,00	6,75	5,54
Taiwan	✓	!	✗	✗	4,33	8,00	1,50	9,25	4,82
Uruguay	✓	!	✗	✗	6,67	6,57	4,25	8,00	6,07
Lituanie	✗	!	✓	✗	9,33	7,43	3,50	7,25	6,72
Argentine	✓	!	✗	✗	5,33	7,86	4,75	5,13	5,52
Chili	✓	!	✗	✗	5,67	6,86	4,00	8,88	5,84
République dominicaine	✗	✗	✗	✓	7,33	5,86	5,75	5,38	6,23
Libye	✗	✗	✗	✓	6,00	5,43	6,50	5,63	6,01
Maroc	✗	✗	✗	✓	6,00	5,71	6,00	6,38	6,01
Jordanie	✗	✗	✗	✓	6,00	5,00	6,25	6,25	5,96
Kazakhstan	✗	✗	✗	✓	7,67	6,14	5,50	5,38	6,31
Tunisie	✓	✗	✗	✗	6,00	6,71	3,75	6,75	5,49
Vanuatu	✓	✗	✗	✗	4,67	3,29	6,25	7,13	5,37
Madagascar	✗	✗	✓	✗	9,00	3,00	7,75	4,63	6,85
Mongolie	✗	✗	✓	✗	9,67	3,29	7,00	4,88	6,92
Ouganda	✗	✗	✓	✗	8,67	3,57	7,00	4,38	6,55
Pakistan	✗	✗	✓	✗	8,67	5,14	7,00	2,75	6,54
République démocratique du Congo	✗	✗	✓	✗	8,67	3,57	8,25	1,38	6,46
Rwanda	✗	✗	✓	✗	8,00	2,86	8,25	5,50	6,81
Tanzanie	✗	✗	✓	✗	8,33	3,71	7,00	5,13	6,58
Vietnam	✗	✗	✓	✗	8,33	5,86	5,75	4,75	6,46
Azerbaïdjan	✗	✗	✓	✗	8,67	5,14	7,00	4,75	6,87
Inde	✗	✗	✓	✗	8,33	4,14	7,00	5,50	6,72
Pologne	✗	!	✗	✗	5,00	9,14	3,00	7,25	5,40
République tchèque	✗	!	✗	✗	4,67	8,71	2,50	8,50	5,26
Roumanie	✗	!	✗	✗	6,33	8,00	4,50	5,50	5,86
Saint-Christophe-et-Nevis	✗	!	✗	✗	3,67	6,57	3,00	8,75	4,78
Sainte-Lucie	✗	!	✗	✗	3,67	6,14	2,00	7,50	4,16
Serbie	✗	!	✗	✗	6,00	7,43	3,00	4,38	4,97
Antigua-et-Barbuda	✗	!	✗	✗	0,33	7,14	3,25	8,25	3,76
Barbade	✗	!	✗	✗	3,67	7,00	1,50	8,88	4,37
Bosnie-et-Herzégovine	✗	!	✗	✗	4,67	6,57	3,00	4,38	4,38
Bulgarie	✗	!	✗	✗	5,00	7,71	4,75	6,00	5,54
Colombie	✗	!	✗	✗	6,67	6,29	4,00	4,75	5,39
Corée du Sud	✗	!	✗	✗	4,00	8,14	2,50	8,25	4,90
Croatie	✗	!	✗	✗	4,67	8,29	3,50	6,75	5,23
Dominique	✗	!	✗	✗	0,33	6,43	2,25	8,13	3,29
Estonie	✗	!	✗	✗	5,67	8,29	3,00	7,75	5,56
Grèce	✗	!	✗	✗	6,33	9,00	1,25	6,50	5,11
Hongrie	✗	!	✗	✗	5,67	8,43	1,75	7,50	5,13
Israël	✗	!	✗	✗	4,00	7,43	4,00	7,38	5,13
Jamaïque	✗	!	✗	✗	4,67	5,71	3,00	5,00	4,34
Lettonie	✗	!	✗	✗	5,00	7,71	3,75	6,38	5,26
Liban	✗	!	✗	✗	5,33	6,00	3,00	3,25	4,32
Macédoine	✗	!	✗	✗	6,33	7,14	4,50	5,38	5,70

Conclusion

Le projet EMINCAP, par son originalité et son caractère innovant en assurance, offre un vrai potentiel de recherche et de créativité dans la construction du modèle.

Cet aspect innovant constitue également une difficulté, en particulier par l'absence d'ouvrages et de méthodologies directement rattachées au sujet. Le choix de la méthode ANCOVA pour le traitement des données manquantes, ou encore de techniques descriptives telles que l'Analyse en Composante Principales, sans être arbitraire, a été retenu car il semblait acceptable en termes de tests de validation et de logique économique de répartition et de sélection des pays. Un travail complémentaire de recherche en analyse statistique et économétrique pourrait donc être effectué pour améliorer la modélisation de données comme le processus de notation et de sélection

Par ailleurs, le manque de données disponibles facilement sur le périmètre géographique de travail, ainsi que la question de la fiabilité de ces données pour un certain nombre des pays étudiés, ont posé de véritables difficultés de préparation de la base, en amont de tout choix de modélisation. L'accès à des bases de données (payantes) comme par exemple IHS Global Insight, Oxford Economics, etc. pourrait permettre de fiabiliser les sources et peut-être d'obtenir des résultats de modélisation plus précis.

La construction de notre base de données et notre modèle nous apportent plusieurs pistes en termes d'aide à la décision en investissement. Dans cette perspective, il semble important de rappeler quelques caractéristiques inhérentes au modèle, dont la prise en compte est fondamentale dans le processus décisionnel.

Tout d'abord, le choix des variables est primordial, et le système de notation est relativement sensible au choix de la pondération des coefficients, notamment en matière d'exposition aux risques.

Ensuite, le modèle construit n'est qu'une base de travail, qui nous a permis de mettre en évidence à quel point le terme de marché émergent, ou même d'opportunité d'investissement, peut rassembler des pays très hétérogènes. Il conviendra donc, sur une liste réduite de pays présélectionnés, d'approfondir l'analyse économique et financière, ainsi que l'état du marché de l'assurance en place : quels sont les concurrents présents ? Quel est le taux de sinistre en assurance automobile ? L'assurance Responsabilité Civile, obligatoire dans presque tous les pays étudiés, est-elle réellement respectée ?

Enfin, le modèle construit en plus de proposer des cibles d'investissement « attendues », les pays émergents « classiques » tels que l'Argentine, la Turquie ou la Malaisie, nous ouvre d'autres horizons : il met particulièrement en valeur les pays du Moyen-Orient, comme l'Arabie Saoudite ou les Émirats Arabes Unis plus développés économiquement, mais où l'activité d'assurance reste en grande partie à construire. L'adaptation de l'expérience française à l'« assurance Islamique », ou « Takaful », pourrait représenter une réelle opportunité d'investissement pour EMINCAP.

Bibliographie

Livres et Publications

La Société du Risque, U. Beck, éditions Champs essais

La « Société du Risque » : analyse et critique, J. Méric, Y. Pesquieux, A. Solé, éditions Economica

Econométrie, R. Bourbonnais, 6ème édition, éditions Dunod

An Exploratory Technique for Investigating Large Quantities of Categorical Data, G. V. Kass, éditions Journal of Applied Statistics, Vol. 29, No. 2 (1980), pp. 119-127

Cours de Modèles d'Aide à la Décision et Décision dans l'Incertain (ESSEC), M. Alfandari, 2007

Cours de Statistiques Multidimensionnelles (ESSEC), M. Vincenzo-Esposito, 2008

Internet

www.swissre.com : Analyses Sigma 2004 à 2008, Analyse Marchés Emergents 2008

www.genre.com : Article « Le téléphone, outil de souscription »

www.scor.fr : Marchés Emergents 2004

www.problematiques.fr : Articles "Faut-il investir dans les pays émergents ?", "Pays émergents: santé, immobilier et mobile", "Pays émergents: le mobile comme vecteur de bancarisation"

<http://www.internetworldstats.com>

<http://stats.oecd.org>

www-wds.worldbank.org, <http://ddp-ext.worldbank.org>, <http://publications.worldbank.org> : sites de la Banque Mondiale – articles, bases de données et publications

www.larousse.fr/encyclopedie

www.observateurocde.org : « Afrique : Le nouveau monde des marchés émergents »

www.prsgroup.com : indices ICRG

<http://oica.net> : Nombre de véhicules par habitant, Régulation sur l'assurance Auto RC

www.heritage.org : indice Economic Freedom

<http://fr.wikipedia.org>

<http://unstats.un.org> : site de l'ONU – rapports statistiques

Presse

Problèmes économiques n° 2962, janv.-09, "Pays Emergents, moteur de la croissance mondiale"

AGEFI Hebdo, 15/05/2008, Crédit Agricole

Les Echos, 13/07/2006, "En assurance, la croissance se trouve dans les pays émergents"

Lettre du MBA de l'ENASS, sept-05, "Madagascar et l'Assurance"

Risques n°23, juillet - septembre 1995, "L'assurance vie dans les pays en voie de développement"

L'Assurance Française n°620, déc.-90, "L'assurance vie dans les pays en voie de développement : Analyse stratégique et pistes de travail"

L'argus n°827, mars-90, "L'assurance au secours du développement"

Le Figaro, oct.-09, "La finance Islamique arrive en France"

Alternatives Economiques, oct.-09, "La finance Islamique débarque en France"

Les Echos, 02/06/2009, "La fiducie va permettre l'émission du sukuk"

Finance Grandes Ecoles, "La finance Islamique, que faut-il savoir ?"

Les Echos, 29/05/2009, "Banquiers arabes et français dialoguent sur la crise et la banque Islamique"

L'Express, 09/04/2009, "Les banquiers d'Allah"

Novethic.fr, 18/06/2009, "L'éthique de la finance Islamique"

L'Argus de l'Assurance, 28/11/2008, "Les assureurs français se convertiront-ils au takaful?"

La tribune de l'assurance, janv.-09, "L'assurance Islamiquement correcte"

Problèmes économiques n° 2962, janv.-09, "Pays Emergents, moteur de la croissance mondiale"

Logiciels

Excel 2007

XLSTAT 2010 (éditeur Addinsoft)

Cartes & Données (éditeur Articque)

Annexes