



**HEC Montréal**

**Les catastrophes naturelles ont-ils un impact sur les courus  
discrétionnaires ?**

**par**

**Yacine Ndiaye**

**Sciences de la gestion  
(Spécialisation Finance)**

*Mémoire présenté en vue de l'obtention  
du grade de maîtrise ès sciences en gestion  
(M. Sc.)*

Mars 2025

© Yacine Ndiaye, 2025

## Résumé :

La gestion des résultats financiers, une pratique répandue en comptabilité et finance, implique des ajustements intentionnels des états financiers pour présenter une performance économique plus favorable. Cette étude examine l'impact des catastrophes naturelles sur la gestion des résultats discrétionnaires dans les pays dont le PIB par habitant est inférieur à 15 000 \$ US. En s'appuyant sur les modèles de Jones modifié (Dechow et al., 1995) et la méthodologie de Roychowdhury (2006), nous analysons les accruals discrétionnaires et les anomalies dans les activités réelles. L'objectif est d'identifier si les entreprises utilisent les catastrophes naturelles comme contexte opportun pour manipuler leurs résultats financiers.

Les résultats montrent que, de manière générale, les catastrophes naturelles entraînent une augmentation des coûts discrétionnaires et des flux de trésorerie opérationnels, quelle que soit leur nature. Cette hausse des coûts discrétionnaires pourrait être liée à une intensification des dépenses en publicité et vente pour soutenir la demande, tandis que l'augmentation du CFO pourrait refléter une gestion plus stricte des liquidités.

L'impact sur les coûts de production varie selon le type de catastrophe, sans tendance uniforme. Quant aux accruals anormaux, aucune relation généralisée n'est observée, mais une augmentation est souvent constatée lorsque l'effet est significatif, suggérant que les entreprises cherchent à augmenter artificiellement leur résultat comptable. Enfin, l'impact des catastrophes est plus marqué sur les courus discrétionnaires résultants d'activité réelle que sur les ajustements comptables, probablement en raison des contraintes réglementaires limitant la manipulation des accruals. En période de crise, les entreprises privilégient donc des stratégies opérationnelles et financières immédiates plutôt que des ajustements comptables.

**Mots clés :** Catastrophe naturelle, courus discrétionnaires, modèle modifié de Jones, Roychowdhury, accruals anormaux, coûts discrétionnaires, coûts de production, flux de trésorerie opérationnels (CFO), manipulation des résultats financiers, stratégies d'ajustement comptable, manipulation des activités réelles.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>10</b>
<b>2</b>	<b>Revue de la littérature sur les courus discrétionnaires</b>	<b>12</b>
2.1	La manipulation des résultats par le biais des accruals . . . . .	14
2.2	La manipulation des activités réelles . . . . .	15
2.2.1	Réduction des dépenses discrétionnaires . . . . .	17
2.2.2	Surproduction . . . . .	17
2.2.3	Manipulation des flux de trésorerie d'exploitation (CFO) . .	18
2.3	Comparaison entre les deux approches . . . . .	19
<b>3</b>	<b>Données</b>	<b>21</b>
3.1	Collecte et nettoyage des données . . . . .	21
3.1.1	Données sur les catastrophes naturelles . . . . .	21
3.1.2	Données financières sur les entreprises . . . . .	23
3.1.3	Données économiques . . . . .	23
3.2	Variables dépendantes et indépendantes . . . . .	23
<b>4</b>	<b>Méthodologie</b>	<b>25</b>
4.1	Détermination des niveaux anormaux des indicateurs de mesure de la manipulation des résultats . . . . .	26
4.1.1	Méthode de régression . . . . .	26
4.1.2	Suppression des valeurs extrêmes des courus discrétionnaires	29
4.2	Analyse des courus discrétionnaires et des catastrophes naturelles . .	31
4.2.1	Test de normalité des courus anormaux . . . . .	31
4.2.2	Statistiques descriptives des catastrophes naturelles . . . . .	33
4.3	Approche utilisée pour notre étude . . . . .	35
<b>5</b>	<b>Résultats</b>	<b>38</b>
5.1	Influence de chaque groupe de catastrophe naturelle sur les courus discrétionnaires . . . . .	38
5.1.1	Influence de chaque groupe de catastrophe naturelle sur les coûts discrétionnaires anormales . . . . .	40
5.1.2	Influence de chaque groupe de catastrophe naturelle sur les coûts de production anormaux . . . . .	40

5.1.3	Influence de chaque groupe de catastrophe naturelle sur les CFO anormaux . . . . .	41
5.1.4	Influence de chaque groupe de catastrophe naturelle sur les accuals anormaux . . . . .	41
5.2	Influence de chaque type de catastrophe naturelle sur les courus discrétionnaires . . . . .	42
5.2.1	Influence de chaque type de catastrophe naturelle sur les coûts discrétionnaires anormaux . . . . .	43
5.2.2	Influence de chaque type de catastrophe naturelle sur les coûts de production anormaux . . . . .	44
5.2.3	Influence de chaque type de catastrophe naturelle sur les CFO anormaux . . . . .	45
5.2.4	Influence de chaque type de catastrophe naturelle sur les ac- cruals anormaux . . . . .	45
<b>6</b>	<b>Limites</b>	<b>46</b>
<b>7</b>	<b>Conclusion</b>	<b>48</b>
<b>8</b>	<b>Références</b>	<b>50</b>
<b>9</b>	<b>Annexes</b>	<b>73</b>

## Liste des tableaux

Table 1	Nombre d'observations par année . . . . .	53
Table 2	Nombre d'observations par pays . . . . .	56
Table 3	Statistiques descriptives des courus discrétionnaires en séparant les courus par pays . . . . .	57
Table 4	Statistiques descriptives des courus discrétionnaires en agrégant les courus pour tous les pays . . . . .	57
Table 5	Statistiques descriptives des courus discrétionnaires excluant les données extrêmes en les séparant par pays . . . . .	58
Table 6	Statistiques descriptives des courus discrétionnaires excluant les données extrêmes en les agrégant pour tous les pays . . . . .	58
Table 7	Résultat du test de Jarque-Bera sur les courus discrétionnaires en séparant les courus par pays . . . . .	59
Table 8	Résultat du test de Jarque-Bera sur les courus discrétionnaires en agrégant les courus pour tous les pays . . . . .	59
Table 9	Statistiques descriptives des courus discrétionnaires selon les deux configurations étudiées . . . . .	60
Table 10 a	Impact des groupes de catastrophe naturelle sur les coûts discrétionnaires anormaux selon les deux configurations étudiées . .	61
Table 10 b	Impact des groupes de catastrophe naturelle sur les coûts de production anormaux selon les deux configurations étudiées . . . .	62
Table 10 c	Impact des groupes de catastrophe naturelle sur les CFO anor- maux selon les deux configurations étudiées . . . . .	63
Table 10 d	Impact des groupes de catastrophe naturelle sur les accruals anormaux selon les deux configurations étudiées . . . . .	64
Table 11 a	Impact de chaque type de catastrophe naturelle sur les coûts discrétionnaires anormaux selon les deux configurations étudiées . .	65
Table 11 b	Impact de chaque type de catastrophe naturelle sur les coûts de production anormaux selon les deux configurations étudiées . . .	66
Table 11 c	Impact de chaque type de catastrophe naturelle sur les CFO anormaux selon les deux configurations étudiées . . . . .	67
Table 11 d	Impact de chaque type de catastrophe naturelle sur les accruals anormaux selon les deux configurations étudiées . . . . .	69

## Liste des figures

Figure 1	Évolution de la moyenne des variables dépendantes au fil des années . . . . .	70
Figure 2	Répartition des courus discrétionnaires en séparant les courus par pays . . . . .	71
Figure 3	Répartition des courus discrétionnaires en agrégeant les courus pour tous les pays . . . . .	72

## Liste des abréviations :

TA: Total accruals

$\Delta REV$ : Variation du revenu

$\Delta AR$ : Variation des comptes recevables

$\Delta PPE$ : Variation du terrain et équipements

EBEI: Bénéfice avant éléments extraordinaire

CFO: Flux de trésorerie d'exploitation

R&D: Coût de recherche et développement

SG&A: Coût de vente, général et administrative

COGS: Coût des biens vendus

$\Delta Inv$ : Variation de l'inventaire

DiscExp: Dépenses discrétionnaires

ProdExp: Dépenses de production

**À noter:** Dans le texte, le terme courus discrétionnaires désigne l'ensemble des coûts discrétionnaires anormaux, des coûts de production anormaux, des flux de trésorerie opérationnels anormaux (CFO anormaux) ainsi que des accruals anormaux. Le coût discrétionnaire anormal, également appelé dépense discrétionnaire anormale, est l'une de nos variables de mesure de l'ajustement des activités réelles des entreprises (Roychowdhury, 2006).



## **Remerciements :**

Je tiens à exprimer ma profonde gratitude au Professeur Martin Boyer, du Département de finance à HEC Montréal, pour son rôle essentiel en tant que directeur de ce mémoire. Ses conseils avisés et sa capacité à orienter ma recherche vers des questions pertinentes à mes intérêts ont été d'une valeur inestimable. De plus, je tiens à remercier M. Mohamed Jabir du Laboratoire de calcul et d'exploitation des données (LACED) de HEC Montréal pour sa disponibilité et son aide dans la collecte des données nécessaires à cette étude.

Je souhaite exprimer ma sincère gratitude envers mes parents, piliers inébranlables de soutien et de bienveillance tout au long de ma vie. Leur présence constante et leur encouragement sans faille ont été le socle sur lequel j'ai pu bâtir mes accomplissements. Leur amour inconditionnel et leur soutien inestimable ont été mes plus grands atouts, et je leur suis infiniment reconnaissant pour leur dévouement sans limites.

Enfin, je tiens à exprimer ma profonde reconnaissance à mes frères, mes sœurs et mes amis pour leur soutien indéfectible tout au long de mes études. Leur présence réconfortante et leurs encouragements ont été des piliers essentiels de ma réussite académique. Leur amitié sincère et leur soutien inconditionnel ont illuminé mon chemin et ont été une source constante de motivation et d'inspiration.

# 1 Introduction

La gestion des résultats financiers, pratique largement répandue en comptabilité et en finance, consiste pour les entreprises à ajuster leurs états financiers de manière à présenter une image plus favorable de leur performance. Ces ajustements, tels que le report de revenus non réalisés ou la réduction stratégique des charges, visent souvent à influencer la perception des investisseurs et des analystes financiers. L'objectif de cette gestion des résultats est d'améliorer les bénéfices rapportés, même si ceux-ci ne reflètent pas fidèlement la réalité économique de l'entreprise.

Face aux enjeux de transparence et de fiabilité de l'information financière, de nombreuses recherches ont exploré les motivations, les méthodes et les impacts de cette gestion, ainsi que les mécanismes de régulation. Des organismes de surveillance, comme la SEC aux États-Unis, ainsi que l'Autorité des marchés financiers AMF et l'Organisme canadien de réglementation du commerce des valeurs mobilières (OCRCVM) au Canada, renforcent leur vigilance pour encadrer ces manipulations.

Cependant, la question de savoir si les entreprises exploitent des événements exceptionnels, comme les catastrophes naturelles, pour manipuler leurs résultats financiers de manière opportuniste reste encore insuffisamment explorée. En période de crise, les pressions financières peuvent amener certaines entreprises à justifier des ajustements comptables, rendant ainsi ces manipulations plus difficiles à détecter. Cette question est d'autant plus cruciale dans les pays à faible revenu, où les entreprises peuvent être particulièrement vulnérables aux chocs extérieurs. En étudiant cette question, nous cherchons à mieux comprendre l'influence des catastrophes naturelles sur les courus discrétionnaires, un phénomène potentiellement sous-estimé. Cette analyse est cruciale pour les parties prenantes telles que les régulateurs, les investisseurs et les analystes financiers, qui doivent évaluer la solidité financière des entreprises et leur comportement en période de crise.

Pour répondre à cette question, nous appliquons le modèle développé par Dechow et al. (1995), qui repose sur les fondements du modèle de Jones (1991), pour évaluer la manipulation des accruals discrétionnaires. Parallèlement, la méthodologie de Roychowdhury (2006) est utilisée pour quantifier les ajustements liés aux activités réelles, notamment les niveaux anormaux de dépenses discrétionnaires, du coût de production et des flux de trésorerie opérationnels (CFO). Nous avons restreint notre échantillon aux pays dont le PIB par habi-

tant est inférieur à 15 000 \$ US, afin d'observer ces effets dans des économies plus vulnérables.

En combinant ces approches avec des régressions en panel, notre analyse explore les corrélations potentielles entre les courus discrétionnaires et différents types de catastrophes naturelles. Par ailleurs, nous analysons en profondeur les impacts spécifiques de chaque groupe de catastrophes pour observer si les comportements de gestion des résultats varient selon le type de crise.

De manière générale, les résultats indiquent que les catastrophes naturelles sont associées à une hausse des charges discrétionnaires et des flux de trésorerie d'exploitation, indépendamment de leur type. En revanche, les effets sur les coûts de production apparaissent hétérogènes et ne suivent pas de tendance claire. S'agissant des accruals anormaux, aucune relation systématique n'est observée. Toutefois, lorsqu'un effet significatif est détecté, il tend à se traduire par une augmentation de ces accruals. Enfin, l'impact des catastrophes est plus marqué sur les courus discrétionnaires résultants d'activité réelle que sur les ajustements comptables.

En somme, cette étude contribue à la compréhension des comportements des entreprises en situation de crise, en examinant les risques de manipulation des états financiers lors de catastrophes dans des contextes économiques plus fragiles.

## 2 Revue de la littérature sur les courus discrétionnaires

La gestion des résultats pour éviter les baisses de bénéfices ou les pertes est une pratique courante, comme le démontrent Burgstahler et Dichev (1997). Ces auteurs montrent que les entreprises, sous pression pour répondre aux attentes du marché, manipulent souvent leurs résultats afin d'éviter de déclarer des pertes. Cette stratégie, bien qu'efficace à court terme pour maintenir la confiance des investisseurs, comporte des risques à plus long terme, notamment une perte de crédibilité.

Dans ce contexte, plusieurs études ont approfondi l'analyse de la gestion des résultats, en examinant tant les accruals discrétionnaires que les manipulations par activités réelles. Teoh, Welch et Wong (1998) ont ainsi montré que, lors d'émissions d'actions secondaires (SEOs), les entreprises tendent à gonfler artificiellement leurs bénéfices pour obtenir une meilleure valorisation. Cependant, cette pratique se retourne souvent contre elles, conduisant à une sous-performance des actions après l'émission, une fois que le marché découvre la manipulation.

Poursuivant cette réflexion, Kothari, Mizik et Roychowdhury (2016) comparent deux approches de gestion des résultats : la manipulation via les accruals discrétionnaires et celle via les activités réelles. Ils concluent que, bien que plus coûteuse et difficile à détecter, la gestion par activités réelles produit des effets plus durables sur les performances financières post-émission. Les changements réglementaires ont également influencé ces pratiques. L'étude de Cohen, Dey et Lys (2008) souligne que l'adoption de la loi Sarbanes-Oxley a réduit l'utilisation des accruals discrétionnaires tout en favorisant le recours à la manipulation par les activités réelles. Ce constat met en lumière la capacité des entreprises à adapter leurs pratiques face à de nouvelles contraintes réglementaires.

Par ailleurs, Marciukaityte (2011) explore l'impact de la gestion des résultats sur les décisions de financement, notamment lors d'émissions de dettes ou d'actions. Elle démontre que les entreprises ayant recours à de telles pratiques se voient souvent imposer un coût du capital plus élevé, en raison d'une perte de confiance des investisseurs. Enfin, Cohen et Zarowin (2010) analysent spécifiquement les SEOs et mettent en évidence que le choix entre accruals et activités réelles peut avoir des répercussions significatives sur la performance à long terme des actions. Ils conclu-

ent que les manipulations par activités réelles, bien qu'avantageuses à court terme, peuvent entraîner des effets persistants sur les résultats financiers des entreprises.

Ainsi, la littérature existante met en évidence que les entreprises ajustent stratégiquement leurs résultats financiers en réponse à divers facteurs, qu'il s'agisse de pressions du marché, d'événements spécifiques comme les émissions d'actions, ou encore de changements réglementaires. Cependant, un aspect moins exploré concerne l'impact des chocs externes, tels que les catastrophes naturelles, sur ces pratiques de gestion des résultats. Dans la continuité des travaux analysant les ajustements comptables et réels des entreprises, notre étude s'intéresse à l'influence des catastrophes naturelles sur les courus discrétionnaires. En ce sens, elle apporte une contribution nouvelle en reliant la gestion des courus discrétionnaires aux effets des catastrophes naturelles, enrichissant ainsi la compréhension des stratégies d'ajustement des entreprises face aux chocs exogènes.

Pour approfondir cette analyse, il est essentiel de s'appuyer sur les deux approches majeures de la gestion des résultats, largement étudiées dans la littérature : la gestion par les accruals (ou courus) et la gestion des activités réelles. La première, principalement illustrée par le modèle de Jones modifié (Dechow et al., 1995), repose sur la détection des ajustements discrétionnaires des comptes, qui permettent aux dirigeants d'ajuster les chiffres de manière à atteindre des objectifs financiers précis. Cette méthode met en lumière l'usage des estimations comptables et des provisions pour influencer les résultats, sans modifier les opérations réelles de l'entreprise. La seconde approche, mise en avant par Roychowdhury (2006), s'intéresse à la gestion des activités réelles, où les entreprises modifient leurs décisions opérationnelles, telles que la réduction des dépenses de recherche et développement, la gestion des stocks ou l'anticipation des ventes..., pour manipuler leurs résultats. Cette forme de gestion est souvent plus coûteuse à long terme, mais moins détectable par les auditeurs et les régulateurs que la gestion basée sur les accruals (Zang, 2012).

Dans les enquêtes menées par Bruns et Merchant (1990) ainsi que par Graham, Harvey et Rajgopal (2005), les dirigeants financiers indiquent une plus grande volonté de manipuler les résultats à travers des activités réelles plutôt qu'au moyen des accruals. Même si la manipulation des activités réelles peut entraîner des coûts plus élevés pour l'entreprise à long terme, il existe des raisons de penser que les dirigeants s'attendent à supporter des coûts privés plus importants, du moins à court terme, lorsqu'ils utilisent la manipulation des accruals. En effet, la manipula-

tion comptable laisse des traces dans les états financiers, ce qui augmente le risque de détection par les auditeurs et les régulateurs, comme le soulignent Cohen, Dey et Lys (2008). Les accruals étant plus facilement assimilables à une fraude comptable, les sanctions juridiques et réputationnelles encourues sont généralement plus sévères, selon Karpoff, Lee et Martin (2008).

## 2.1 La manipulation des résultats par le biais des accruals

La méthode de la manipulation des résultats par le biais des accruals se base sur l'analyse des ajustements comptables discrétionnaires pour détecter les manipulations des résultats financiers.

Pour estimer ces accruals, le modèle cross-sectionnel de Jones est utilisé. Ce modèle a été développé par Dechow, Sloan, et Sweeney en 1995. Leur modèle est une extension du modèle original de Jones (1991) et a été conçu pour améliorer la détection de la gestion des résultats en utilisant des données transversales. Elle est plus appropriée lorsque les facteurs spécifiques à l'entreprise doivent être contrôlés. La formule utilisée est l'équation (1):

$$TA_{it}/Asset_{i,t-1} = b_1 (1/Asset_{i,t-1}) + b_2 ((\Delta REV_{it} - \Delta AR_{it}) / Asset_{i,t-1}) + b_3 (\Delta PPE_{it}/Asset_{i,t-1}) + \epsilon_{it} \quad (1)$$

avec  $TA_{it} = EBEI_{i,t} - CFO_{it}$ , où :

- $a_0, b_1, b_2, b_3$  sont les coefficients de régression
- $TA_{it}$  : Total des accruals de l'entreprise  $i$  à l'année  $t$
- $Asset_{i,t-1}$  : Total des actifs de l'entreprise  $i$  à l'année  $t-1$
- $\Delta REV_{it}$  : Variation des revenus entre les années  $t$  et  $t-1$
- $\Delta AR_{it}$  : Variation des comptes clients entre les années  $t$  et  $t-1$

- $\Delta PPE_{it}$  : Variation des immobilisations corporelles entre les années  $t$  et  $t-1$
- $EBEI_{i,t}$ <sup>1</sup> : Bénéfice avant éléments extraordinaires à l'année  $t$
- $CFO_{it}$  : Flux de trésorerie d'exploitation à l'année  $t$
- $\epsilon_{it}$  : Terme d'erreur du modèle c'est-à-dire le niveau anormal des accruals à l'année  $t$

Le modèle modifié est très utilisé pour détecter la gestion des résultats, comme le soulignent Kothari, Leone et Wasley (2005). Il permet d'estimer la portion discrétionnaire des accruals totaux, notée  $\epsilon$ . Le modèle suppose que les accruals non discrétionnaires sont influencés par les changements dans les revenus et les changements dans les actifs immobilisés. Contrairement au modèle de Jones, le modèle de Jones modifié prend en considération que les comptes recevables peuvent être manipulés par les gestionnaires, notamment en gonflant les ventes à crédit. Le modèle suppose la stationnarité entre les accruals et les variables indépendantes à travers le temps et que les revenus ne sont pas sujets à manipulation par les gestionnaires.

## 2.2 La manipulation des activités réelles

La manipulation des activités réelles implique des actions comme la modification des dépenses d'exploitation, la réduction des coûts de recherche et développement ou la gestion des niveaux de production pour influencer les résultats financiers à court terme.

La méthode de Roychowdhury permet de détecter la gestion des bénéfices réels en se basant sur l'analyse des décisions opérationnelles qui pourraient être prises pour manipuler les résultats financiers à court terme, au détriment de la performance à long terme. En effet, la gestion des bénéfices réels peut améliorer les résultats à court terme, mais elle peut également nuire aux performances à long terme, car elle peut réduire les investissements nécessaires ou engendrer des coûts plus élevés dans le futur.

---

<sup>1</sup>Les formules du EBEI et du CFO sont donnés en annexe.

Le modèle présume que les activités économiques suivent des tendances normales, et tout écart par rapport à ces tendances peut indiquer une manipulation des résultats.

Pour évaluer la présence de manipulation des activités réelles, Roychowdhury a utilisé trois indicateurs : le niveau anormal des dépenses discrétionnaires, le coût de production anormal, et le flux de trésorerie d'exploitation (CFO) anormal. Pour se faire, il a identifié trois techniques couramment employées pour manipuler les résultats financiers:<sup>2</sup>

- **La manipulation des ventes** : Les gestionnaires peuvent temporairement augmenter les ventes en offrant des remises importantes ou des conditions crédit plus avantageuses, incitant les clients à avancer leurs achats. Parmi les stratégies qui sont souvent utilisées, il y a les rabais temporaires ou le financement à taux zéro qui permettent d'augmenter les volumes de ventes à court terme, mais au prix d'une réduction des marges bénéficiaires. Ainsi, bien que les bénéfices comptables apparaissent plus élevés, ces pratiques entraînent une diminution du flux de trésorerie opérationnel en raison de la baisse des marges.
- **La réduction des dépenses discrétionnaires** : Les entreprises peuvent accroître leurs bénéfices à court terme en réduisant les dépenses discrétionnaires, telles que les dépenses en recherche et développement (R&D), en publicité et certaines charges générales et administratives (SG&A). Cette réduction des coûts entraîne une amélioration immédiate des résultats financiers, mais peut nuire à la performance future si ces dépenses sont essentielles pour générer des revenus à long terme.
- **Surproduction** : Les gestionnaires peuvent augmenter la production au-delà de la demande réelle afin de répartir les coûts fixes sur un plus grand nombre d'unités, réduisant ainsi le coût unitaire de production et améliorant les marges bénéficiaires. Toutefois, cette surproduction engendre des coûts supplémentaires de stockage et de production qui ne sont pas immédiatement compensés par des ventes, ce qui réduit les flux de trésorerie malgré l'apparente amélioration des marges.

---

<sup>2</sup>selon les définitions de Roychowdhury



### 2.2.1 Réduction des dépenses discrétionnaires

Selon Roychowdhury, les entreprises peuvent réduire les dépenses discrétionnaires pour améliorer les bénéfices à court terme. Par conséquent, selon lui, une baisse des dépenses discrétionnaires, en particulier par rapport aux revenus passés, peut signaler une manipulation des bénéfices à court terme.

Le niveau anormal des dépenses discrétionnaires peut être déterminé à l'aide de l'équation (2):

$$DiscExp_{it}/Asset_{i,t-1} = b_1 (1/Asset_{i,t-1}) + b_2 (Sales_{i,t-1}/Asset_{i,t-1}) + \epsilon_{it} \quad (2)$$

avec  $DiscExp = R\&D + Publicité + SG\&A$ , où :

- $a_0, b_1, b_2$  sont les coefficients de régression
- $DiscExp_{it}$  : Dépenses discrétionnaires de l'entreprise  $i$  à l'année  $t$
- $Asset_{i,t-1}$  : Total des actifs de l'entreprise  $i$  à l'année  $t-1$
- $Sales_{i,t-1}$  : Revenus totaux à l'année  $t-1$
- $\epsilon_{it}$  : Terme d'erreur (c'est-à-dire les dépenses discrétionnaires anormales).

À noter: Tant que SG&A est disponible, les dépenses en R&D et en publicité peuvent être mis à zéro s'ils ne sont pas disponibles.

### 2.2.2 Surproduction

Cette stratégie consiste à produire plus que nécessaire, ce qui permet de répartir les coûts fixes sur un plus grand nombre d'unités, réduisant ainsi le coût unitaire et augmentant artificiellement la marge bénéficiaire. Par conséquent, une augmentation disproportionnée de la production par rapport aux ventes pourrait indiquer une tentative de réduction artificielle du coût des biens vendus (COGS), ce qui est une forme de manipulation des bénéfices.

Le coût anormal de production peut être déterminé à l'aide de l'équation (3):

$$\begin{aligned}
Prod_{it}/Asset_{i,t-1} = & b_1 (1/Asset_{i,t-1}) + b_2 (Sales_{it}/Asset_{i,t-1}) \\
& + b_3 (\Delta Sales_{it}/Asset_{i,t-1}) + b_4 (\Delta Sales_{i,t-1}/Asset_{i,t-1}) + \epsilon_{it}
\end{aligned} \tag{3}$$

avec  $Prod = COGS + \Delta Inv$ , où:

- $a_0, b_1, b_2, b_3, b_4$  sont les coefficients de régression
- $Prod_{it}$  : Coûts de production de l'entreprise  $i$  à l'année  $t$
- $Asset_{i,t-1}$  : Total des actifs de l'entreprise  $i$  à l'année  $t-1$
- $Sales_{it}$  : Revenus totaux à l'année  $t$
- $\Delta Sales_{it}$  : Variation du revenu total entre l'année  $t$  et  $t-1$
- $\Delta Sales_{i,t-1}$  : Variation du revenu total entre l'année  $t-1$  et  $t-2$
- $\epsilon_{it}$  : Terme d'erreur (c'est-à-dire le coût anormal de production).

### 2.2.3 Manipulation des flux de trésorerie d'exploitation (CFO)

L'idée est de détecter des anomalies dans les flux de trésorerie d'exploitation qui ne seraient pas cohérentes avec les revenus ou les activités normales de l'entreprise. Par conséquent, si les entreprises manipulent leurs résultats en augmentant temporairement les ventes (par exemple en offrant des remises importantes), cela peut entraîner des flux de trésorerie d'exploitation anormalement élevés par rapport aux revenus. À l'inverse, retarder le paiement des fournisseurs ou accélérer le recouvrement des créances peut également fausser le CFO.

Le CFO anormal peut être déterminé à l'aide de l'équation (4):

$$\begin{aligned}
CFO_{it}/Asset_{i,t-1} = & b_1 (1/Asset_{i,t-1}) + b_2 (Sales_{it}/Asset_{i,t-1}) \\
& + b_3 (\Delta Sales_{it}/Asset_{i,t-1}) + \epsilon_{it}
\end{aligned} \tag{4}$$

où :

- $a_0, b_1, b_2, b_3$  sont les coefficients de régression
- $CFO_{it}$  : flux de trésorerie opérationnel de l'entreprise  $i$  à l'année  $t$
- $Asset_{i,t-1}$  : Total des actifs de l'entreprise  $i$  à l'année  $t-1$
- $Sales_{it}$  : Revenus totaux à l'année  $t$
- $\Delta Sales_{it}$  : Variation du revenu total entre l'année  $t$  et  $t-1$
- $\epsilon_{it}$  : Terme d'erreur (c'est-à-dire le niveau anormal de flux de trésorerie d'exploitation).

## 2.3 Comparaison entre les deux approches

L'utilisation de la manipulation des résultats par le biais des accruals et la manipulation des activités réelles offrent des perspectives complémentaires sur les stratégies utilisées par les entreprises pour atteindre des objectifs financiers spécifiques. La manipulation des résultats par les accruals est souvent identifiée comme un indicateur clé de la gestion des résultats. Le modèle de Jones modifié est largement accepté dans la littérature académique en raison de son cadre théorique solide et de sa capacité à fournir des comparaisons inter-études fiables. En effet, une vaste littérature soutient l'utilisation du modèle de Jones modifié, offrant ainsi une bonne comparabilité. De plus, le modèle est particulièrement précieux pour plusieurs raisons, notamment la robustesse théorique et l'accessibilité des données.

Le modèle de Jones modifié est fondé sur des principes comptables solides qui permettent de distinguer les accruals normaux des accruals discrétionnaires. Cette distinction est cruciale pour identifier les tentatives de manipulation qui respectent techniquement les règles comptables, mais qui visent à fausser l'image économique de l'entreprise. Les accruals sont dérivés directement des états financiers publiés, ce qui rend les données facilement accessibles et normalisées. Cette accessibilité permet de mener des études à grande échelle et de réaliser des comparaisons entre

différentes entreprises ou secteurs, renforçant ainsi la pertinence et l'impact des recherches.

Bien que le modèle de Jones modifié soit un outil puissant, les accruals peuvent être manipulés dans les limites des principes comptables, rendant leur détection plus complexe. Cela nécessite donc l'utilisation de techniques d'ajustement sophistiquées et l'application de jugements experts pour interpréter correctement les résultats.

La méthode de la manipulation des activités réelles capte les décisions de gestion quotidiennes, fournissant des perspectives sur la prise de décision stratégique au-delà des simples ajustements comptables. Ceci fait en sorte que cette approche est très pertinente.

Contrairement à la manipulation des accruals, qui affecte principalement les chiffres comptables, la manipulation des activités réelles permet d'examiner l'impact économique réel des décisions managériales, un point soulevé par Roychowdhury (2006). Elle offre ainsi une compréhension plus profonde des choix stratégiques effectués par les gestionnaires, tout en permettant d'étudier leur influence non seulement sur les résultats financiers à court terme, mais aussi sur la viabilité à long terme de l'entreprise, comme l'indiquent Graham, Harvey et Rajgopal (2005). De plus, les impacts de la manipulation des activités réelles peuvent s'étendre sur plusieurs périodes, ce qui la rend particulièrement pertinente pour les études longitudinales, selon Zang (2012).

Cependant, il est essentiel de noter que la détection de la manipulation basée sur les activités réelles est plus difficile. Contrairement aux accruals, facilement quantifiables, les décisions opérationnelles peuvent être justifiées par des stratégies légitimes, rendant la distinction entre manipulation et stratégie légitime plus complexe, comme le précisent Dechow, Kothari et Watts (1998). Cette complexité nécessite donc une analyse contextuelle approfondie ainsi que des modèles économétriques adaptés pour détecter ces anomalies, comme le suggèrent Cohen, Dey et Lys (2008).

En résumé, l'approche par les accruals discrétionnaires, notamment à travers le modèle de Jones modifié proposé par Dechow, Sloan et Sweeney (1995), est largement utilisée pour détecter la gestion des résultats comptables. L'avantage principal de ce modèle réside dans l'amélioration qu'il apporte au modèle initial de Jones (1991), en excluant l'effet des variations des créances clients. Cette modification permet une estimation plus précise des accruals dits normaux et facilite ainsi l'identification de leur composante discrétionnaire, souvent associée à des pratiques

de manipulation. Toutefois, cette approche présente plusieurs limites : elle repose sur des hypothèses fortes concernant le comportement normal des accruals, peut capter des variations économiques non liées à une manipulation intentionnelle, et reste peu efficace pour détecter des pratiques de gestion des résultats fondées sur des décisions opérationnelles plutôt que comptables.

C'est justement pour répondre à ces limites que Roychowdhury (2006) propose une approche fondée sur la manipulation des activités réelles. Celle-ci s'intéresse aux ajustements stratégiques des décisions d'exploitation tels que la réduction des dépenses discrétionnaires, la surproduction ou encore l'escompte sur les ventes, visant à influencer les résultats financiers. Cette méthode présente l'avantage de révéler des formes de gestion plus coûteuses et moins directement visibles dans les états financiers. Toutefois, elle repose sur des hypothèses complexes quant aux comportements opérationnels normaux des entreprises, et risque de confondre des décisions légitimes de gestion avec des pratiques opportunistes. De plus, elle est particulièrement sensible aux différences sectorielles et structurelles entre entreprises, ce qui peut nuire à la comparabilité des résultats. Ainsi, leur combinaison offre des perspectives complémentaires et permet une analyse plus riche et nuancée des stratégies de gestion des résultats.

## **3 Données**

### **3.1 Collecte et nettoyage des données**

L'objectif de cette étude est d'examiner si les catastrophes naturelles ont un impact significatif sur les courus discrétionnaires des entreprises. Pour ce faire, trois catégories principales de données ont été collectées : les données relatives aux catastrophes naturelles, les données financières des entreprises, et les données économiques.

#### **3.1.1 Données sur les catastrophes naturelles**

Les données concernant les catastrophes naturelles ont été recueillies à partir de la base de données EM-DAT. Ces données incluent le groupe de catastrophe, le sous-groupe, le type de désastre, le pays, la région, ainsi que la date de début de

l'événement. Selon EM-DAT<sup>3</sup>, est considérée catastrophe tout événement impliquant un impact inattendu et écrasant sur les êtres humains. Le critère pour qu'un événement soit considéré comme catastrophe doit également respecter au moins un de ces critères :

- au moins dix décès (y compris les morts et les disparus),
- au moins 100 personnes affectées (personnes affectées, blessées ou sans-abri),
- un appel à l'aide internationale ou une déclaration d'urgence.

Les catastrophes naturelles sont présentées sous deux grandes catégories: celles d'origine naturelle et celles d'origine technologique. Cependant, pour notre étude, seules les catastrophes de type naturel sont prises en compte. De plus, notre analyse se concentre sur cinq des six groupes définis dans la classification des catastrophes naturelles<sup>4</sup>.

Le premier groupe correspond aux catastrophes de type géophysique, qui incluent les tremblements de terre, les mouvements de terrain (secs) et les activités volcaniques. Le deuxième groupe regroupe les catastrophes de type hydrologique, comprenant les inondations, les glissements de terrain et les mouvements de terrain (humides). Le troisième groupe concerne les catastrophes de type météorologique, telles que les tempêtes et les températures extrêmes. Le quatrième groupe englobe les catastrophes de type climatologique, qui incluent les sécheresses et les feux de forêt. Enfin, le cinquième groupe correspond aux catastrophes de type biologique, comprenant les épidémies et les infestations.

Nous avons éliminé toutes les catastrophes naturelles pour lesquelles la date de début était manquante (1,95 % de l'échantillon) ou pour lesquelles le nombre d'observations était inférieur à 100 (0,02 % de l'échantillon).<sup>5</sup> Dans le cadre de cette recherche, nous considérons que la durée d'une catastrophe naturelle s'étend jusqu'à un an après sa date de début. Toutes les catastrophes de type technologique ont été regroupées dans un seul groupe nommé "autres"<sup>6</sup>.

---

<sup>3</sup>Source: <https://doc.emdat.be/docs/protocols/entry-criteria/>

<sup>4</sup>La catégorie "extra-terrestre" n'est pas incluse dans cette étude.

<sup>5</sup>Seule la catastrophe naturelle de type infestation a été exclue de l'échantillon.

<sup>6</sup>Les types d'accidents incluent : accidents ferroviaires, accidents routiers, effondrements (industriels), explosions (industrielles), accidents maritimes, incendies (divers), effondrements (divers), accidents aériens, incendies (industriels), explosions (divers), accidents divers (généraux), accidents industriels (généraux) et fuites de gaz.

### **3.1.2 Données financières sur les entreprises**

Les données financières des entreprises ont été extraites des bases de données Compustat North America et Compustat Global, étant donné que cette étude couvre des entreprises à l'échelle mondiale. Les variables spécifiques utilisées sont : l'identifiant unique de l'entreprise, le code de la devise, la date de l'exercice fiscal, les actifs totaux, le coût des biens vendus, le bénéfice avant élément extraordinaire, la variation des inventaires, les flux de trésorerie opérationnels, les immobilisations corporelles brutes de l'année, la variation des comptes recevables, les revenus totaux, les dépenses en publicité, les dépenses en recherche et développement, les dépenses de vente, générales et administratives, le code du pays où se trouve l'entreprise, la ville où se trouve l'entreprise, le secteur d'activité de l'entreprise. Les données collectées ont été nettoyées en éliminant les doublons d'entreprises et en supprimant les observations pour lesquelles la date fiscale ou la devise de présentation n'étaient pas disponibles (exclusion de 47,49 % de l'échantillon, le ramenant ainsi à 791 212 observations). Pour les données provenant de Compustat Global, lorsque le coût de publicité ou de R&D n'est pas disponible, il est considéré comme nul conformément aux critères de Roychowdhury (2006).

### **3.1.3 Données économiques**

Les données économiques utilisées dans cette étude comprennent le taux de change des devises par rapport au dollar américain pour l'année 2010 ainsi que le PIB par habitant. Les taux de change ont été extraits de la base de données Compustat Global, tandis que les données relatives au PIB par habitant proviennent de la Banque mondiale. Pour les devises ayant changé de code au fil des années, telles que le peso mexicain et le bolívar vénézuélien, les taux de change ont été ajustés afin de refléter les codes en vigueur à l'époque des observations. En raison de l'hyperinflation observée au Zimbabwe et au Soudan, les données de ces deux pays ont été exclues de notre analyse.

## **3.2 Variables dépendantes et indépendantes**

Au début du nettoyage des données, notre base comptait 1 506 747 observations. Nous avons ensuite exclu 47,49 % des observations en raison de l'absence de date fiscale ou de devise de présentation, ramenant ainsi l'échantillon à 791 212 obser-

vations réparties sur 130 pays. Après avoir éliminé les observations associées au Zimbabwe, à la Zambie, au Soudan et à l'ex-Rhodésie du Sud, nous obtenons un total de 790 789 observations couvrant 128 pays (exclusion de 0,05 % des données). Par la suite, un filtrage a été appliqué pour ne conserver que les pays dont le PIB par habitant est inférieur à 15 000 \$ US, réduisant notre échantillon à 275 387 observations issues de 97 pays (exclusion de 65,18 % des données). Les observations pour lesquelles les actifs sont égaux à zéro ont ensuite été retirées, ramenant le total à 251 205 observations toujours réparties sur 97 pays(exclusion de 8,78 % des données). L'élimination des valeurs extrêmes a permis d'obtenir un jeu de données final de 209 636 observations (exclusion de 16,55 % des données). Enfin, les catastrophes naturelles ayant moins de 30 observations ont été exclues, menant à un échantillon définitif de 209 342 observations réparties sur 55 pays (exclusion de 0,14 % des données). Ainsi, à partir de l'échantillon initial, seulement 13,89 % des données ont été utilisées. Le nombre d'observations pour chaque type de catastrophe naturelle, ainsi que pour les années sans catastrophe, est présenté en annexe dans le tableau 1. Le tableau 2, également en annexe, montre la répartition des observations par pays.

Pour préparer les données financières en vue de l'utilisation des méthodes de Jones modifiée en coupe transversale et de Roychowdhury (2006), plusieurs variables<sup>7</sup> ont été créées ou ajustées afin de fournir les indicateurs nécessaires à l'analyse. Par la suite, quatre groupes de variables spécifiques ont été élaborés pour les régressions : Les trois premiers groupes suivent la méthodologie de Roychowdhury (2006), permettant d'analyser les anomalies dans les dépenses discrétionnaires, les coûts de production et les flux de trésorerie opérationnels. Le quatrième groupe repose sur la méthode de Jones modifiée en coupe transversale, adaptée pour détecter les accruals discrétionnaires. Les quatre variables dépendantes utilisées pour déterminer les courus discrétionnaires sont le ratio des dépenses discrétionnaires par rapport aux actifs retardés, le ratio du flux de trésorerie opérationnel par rapport aux actifs retardés, le ratio des dépenses de production par rapport aux actifs retardés et le ratio des accruals par rapport aux actifs retardés. Les trois premières variables mesurent l'activité réelle, conformément à Roychowdhury (2006), tandis que la dernière évalue les ajustements comptables selon la méthode modifiée de

---

<sup>7</sup>La liste des variables dépendantes et indépendantes, ainsi que leurs formules, est fournie en annexe.



Jones.

La figure 1, présentée en annexe, illustre l'évolution moyenne de ces différentes variables au fil des années, mettant en évidence leurs tendances respectives dans le temps. Nous observons que les dépenses discrétionnaires, rapportées aux actifs retardés, restent relativement stables, bien qu'une légère tendance à la baisse soit perceptible après 2010. Concernant les coûts de production, également rapportés aux actifs retardés, des fluctuations notables apparaissent : une augmentation initiale entre 2007 et 2009, suivie d'une diminution progressive à partir de 2012, malgré une certaine volatilité. Les flux de trésorerie opérationnels, toujours en proportion des actifs retardés, montrent une stabilité globale, avec une légère hausse après 2019. Enfin, les accruals rapportés aux actifs retardés présentent un ratio généralement faible et négatif, caractérisé par une volatilité marquée, avec des pics et des creux au fil des années.

Ces différentes fluctuations pourraient indiquer que les entreprises ajustent ces variables au fil du temps en réponse à des chocs externes, tels que des catastrophes naturelles. Par exemple, une diminution des dépenses discrétionnaires pourrait représenter une réaction directe pour contrer l'impact d'un choc externe, ces coûts étant souvent plus flexibles. De même, les variations des coûts de production pourraient refléter des efforts pour relancer la production après des perturbations. De plus, la faiblesse des accruals par rapport aux actifs retardés pourrait suggérer une préférence pour des ajustements réels plutôt que comptables. Les formules des variables indépendantes pour chaque variable dépendante sont fournies en annexe. Chaque ensemble de formules correspond à la variable dépendante associée.

## 4 Méthodologie

Dans cette section, nous détaillerons la méthodologie employée pour déterminer si les catastrophes naturelles ont un impact sur les courus discrétionnaires. Nous commencerons par décrire la méthode de régression utilisée pour identifier les courus discrétionnaires, puis nous expliquerons l'exclusion des valeurs extrêmes. Ensuite, nous analyserons les courus discrétionnaires en relation avec les catastrophes naturelles, en testant notamment la somme des courus discrétionnaires et la normalité de ces courus. Nous présenterons également des statistiques descriptives sur les catastrophes naturelles pour contextualiser notre analyse. Enfin, nous expliquerons

l’approche globale de notre étude, incluant les fonctions indicatrices et les régressions utilisées pour tester nos hypothèses. Cette approche nous permettra de tirer des conclusions sur l’impact des catastrophes naturelles sur les cours discrétionnaires pour les pays ayant un PIB par habitant inférieur à 15 000 \$ US.

## 4.1 Détermination des niveaux anormaux des indicateurs de mesure de la manipulation des résultats

Afin de déterminer l’impact des catastrophes naturelles sur les cours discrétionnaires, notre méthodologie débute par l’identification des niveaux normaux des trois indicateurs, en utilisant le modèle de Jones modifié proposé par Dechow, Sloan et Sweeney (1995) ainsi que la méthode de détection de la manipulation des activités réelles développée par Roychowdhury (2006). Nous utilisons ensuite une régression PanelOLS pour calculer les niveaux anormaux de chaque indicateur. La même procédure est appliquée pour estimer les niveaux anormaux des accruals, en utilisant le modèle de Dechow (1995).

Les deux modèles employés dans notre étude apportent des perspectives riches et complémentaires, malgré leurs limites respectives. Le modèle de Jones modifié est particulièrement efficace pour isoler la part discrétionnaire des résultats comptables, mais il montre des limites lorsqu’il s’agit de détecter les manipulations fondées sur des décisions opérationnelles. En revanche, l’approche de Roychowdhury se concentre spécifiquement sur ces décisions, révélant des pratiques souvent plus subtiles et coûteuses tout en offrant une vue plus détaillée des stratégies de gestion des résultats. En combinant ces deux modèles, nous parvenons à une analyse plus complète et nuancée, tirant parti des forces de chaque méthode.

### 4.1.1 Méthode de régression

Pour déterminer les niveaux anormaux de chacun de nos cours discrétionnaires, la technique du PanelOLS a été utilisée. Dans notre modèle de régression<sup>8</sup>, nous n’avons pas inclus d’effets fixes d’entité. Tout de même, nous avons spécifié *cov - type = ‘clustered’*, ce qui signifie que nous avons utilisé une correction des erreurs types robuste en regroupant les erreurs au niveau des entités. Cette approche ajuste les erreurs types pour tenir compte de la corrélation des erreurs au

---

<sup>8</sup>Le code est fourni en annexe: page 80, bloc 9, fonction `calculate_regression_results()`

sein des mêmes entités, ce qui est crucial pour obtenir des tests statistiques valides dans les données de panel (Wooldridge, 2010). En utilisant la correction des erreurs types clusterisées, nous avons atténué les problèmes d'hétéroscédasticité potentiels, permettant ainsi de produire des intervalles de confiance plus précis et des tests statistiques plus fiables. La clusterisation des erreurs par entité aide également à gérer l'autocorrélation au sein des entités, ce qui est important dans les données de panel où les erreurs peuvent être corrélées dans le temps pour la même entité (Hsiao, 2007).

Il est important de noter que même avec ces ajustements, il subsiste toujours un risque d'endogénéité, où les variables explicatives peuvent être corrélées avec l'erreur aléatoire du modèle. Les effets fixes ne résolvent pas les problèmes d'endogénéité entre les variables indépendantes et la variable dépendante, et des méthodes supplémentaires comme les variables instrumentales peuvent être nécessaires pour obtenir des estimations non biaisées (Baltagi, 2008).

Pour conclure, deux scénarios ont été envisagés dans notre analyse. Le premier scénario considère que le niveau normal de chaque indicateur est uniquement déterminé par le pays dans lequel se trouve l'entreprise ( $\sum \epsilon_{i, pays}$ ). Le second scénario, en revanche, considère que le niveau normal de chaque indicateur dépend d'un contexte plus global, prenant en compte l'ensemble des pays étudiés ( $\sum \epsilon_i$ ). Les niveaux anormaux des dépenses discrétionnaires, des coûts de production, des flux de trésorerie et des accruals, représentés par les résidus des régressions des équations 1, 2, 3 et 4, constitueront les principales données utilisées dans la suite de notre étude pour déterminer si les catastrophes naturelles ont un impact sur les courus discrétionnaires.

### **Statistiques descriptives:**

Les statistiques descriptives pour chacun des niveaux anormaux de nos quatre indicateurs sont présentées en annexe dans le tableau 3 et le tableau 4. Les moyennes des courus varient entre les deux tableaux et présentent généralement une légère augmentation dans la configuration agrégée par rapport à celle par pays. Par exemple, la moyenne des coûts discrétionnaires anormaux est de 4,48 dans le tableau 3 (séparé par pays) contre 5,80 dans le tableau 4 (agrégé pour tous les pays). De même, la moyenne des coûts de production anormaux est de 0,02 dans le tableau

3, contre 0,28 dans le tableau 4. Pour les flux de trésorerie anormaux (CFO), la moyenne est de 0,49 dans le tableau 3, contre 0,99 dans le tableau 4. Enfin, la moyenne des accruals anormaux est de -0,32 dans le tableau 3, contre -2,09 dans le tableau 4.

Cette différence peut refléter des variations des courus entre les pays, suggérant qu'en moyenne, les courus discrétionnaires peuvent être plus élevés lorsqu'ils sont observés dans un contexte global. De plus, les résultats des courus agrégés affichent une variabilité plus élevée : 4 544,8 dans le tableau 3 contre 4 554,6 dans le tableau 4 pour les coûts discrétionnaires anormaux, 46,26 dans le tableau 3 contre 68,39 dans le tableau 4 pour les coûts de production anormaux, 433,78 dans le tableau 3 contre 464,99 dans le tableau 4 pour les CFO anormaux, et 418,78 dans le tableau 3 contre 481,59 dans le tableau 4 pour les accruals anormaux. Ces écarts-types plus élevés dans la configuration agrégée montrent une plus grande variabilité des courus discrétionnaires entre les pays, ce qui pourrait être dû aux différences dans la gestion des courus d'un pays à l'autre.

Les valeurs minimales et maximales des indicateurs diffèrent également entre les deux tableaux. La valeur maximale des coûts discrétionnaires anormaux dans le tableau 4 (2 279 643) est légèrement supérieure à celle du tableau 3 (2 271 285). Les valeurs maximales pour les coûts de production anormaux sont très proches, avec 18 672,43 dans le tableau 3 et 18 692,20 dans le tableau 4. Les flux de trésorerie anormaux (CFO) suivent une tendance similaire, avec une légère augmentation dans le tableau 4 (217 809,44) par rapport au tableau 3 (215 967,58). Cela indique que, dans un contexte agrégé, les coûts peuvent atteindre des niveaux plus élevés.

Cependant, pour les accruals anormaux, la valeur maximale du tableau 3 (15 848,02) est significativement supérieure à celle du tableau 4 (7 122,28). Cela suggère une variabilité plus importante des accruals anormaux à l'échelle individuelle, par rapport à une configuration agrégée où les valeurs extrêmes sont souvent atténuées.

D'après les formules basées sur Roychowdhury, les trois indicateurs que sont les coûts discrétionnaires anormaux, les coûts de production anormaux et les flux de trésorerie opérationnels anormaux sont calculés à partir d'un ratio fonction de l'actif total retardé. Cependant, les valeurs maximales et minimales observées apparaissent aberrantes, ce qui justifie la nécessité de leur suppression afin d'atténuer l'impact des observations extrêmes et d'assurer une analyse plus robuste et fiable.

### 4.1.2 Suppression des valeurs extrêmes des courus discrétionnaires

Pour chaque variable, nous avons éliminé les observations dont les ratios sont supérieurs à 1 ou inférieurs à -1. Étant donné que chaque variable est rapportée à l'actif total retardé, un ratio absolu supérieur à 1 indique que la dépense ou le flux considéré dépasse la valeur des actifs totaux de l'entreprise, ce qui est économiquement peu plausible et suggère une possible erreur de saisie, une valeur extrême ou un cas atypique difficilement interprétable. Ainsi, nous avons exclu ces observations afin d'éviter que ces valeurs aberrantes ne biaisent les résultats de l'analyse.

Cette approche s'inscrit dans la pratique courante en recherche comptable et financière, où les observations extrêmes sont souvent écartées pour améliorer la robustesse des modèles empiriques (Dechow et Dichev, 2002).<sup>9</sup>

En limitant les valeurs extrêmes, nous réduisons leur influence sur les statistiques descriptives (comme les moyennes et les écarts-types), ce qui rend nos résultats plus robustes et représentatifs des données centrales (Hastie et al., 2009). De plus, dans les modèles statistiques, les valeurs aberrantes peuvent fausser les résultats et conduire à des conclusions incorrectes (Rousseeuw & Leroy, 1987).

#### **Statistiques descriptives :**

Les statistiques descriptives des niveaux anormaux après exclusion des valeurs extrêmes pour chacun de nos quatre indicateurs sont présentées en annexe dans le tableau et le tableau 6. L'exclusion des variables discrétionnaires extrêmes a significativement impacté les résultats.

Un des changements majeurs est la moyenne, qui se rapproche désormais de zéro, ce qui est conforme à une situation sans manipulation de données. En effet, dans la configuration séparée par pays, la moyenne des coûts discrétionnaires anormaux est passée de 4,5 à 0,01. De même, la moyenne des coûts de production anormaux est passée de 0,03 à 0,02. La moyenne des flux de trésorerie anormaux (CFO) a elle aussi baissé de 0,5 à 0,00, et celle des accruals anormaux de -0,31 à -0,00.

La variabilité, un autre point crucial, a également été largement réduite. L'écart-type des coûts discrétionnaires anormaux est passé de 4 544,8 à 0,10. Pour les coûts de production anormaux, il a baissé de 46,3 à 0,17 ; pour les flux de trésorerie

---

<sup>9</sup>Le code est fourni en annexe: page 80, bloc 4

anormaux (CFO), de 433,8 à 0,10 ; et pour les accruals anormaux, de 418,8 à 0,12. Les valeurs extrêmes se sont également resserrées. Dans la configuration agrégée par pays, les valeurs minimales et maximales des coûts discrétionnaires anormaux sont passées de -19 289 à -5,5 et de 2 279 643 à 1,00, respectivement. Pour les coûts de production anormaux, elles sont passées de -2 237 à -18,71 et de 18 692 à 1,50 ; pour les flux de trésorerie anormaux (CFO), de -7 496 à -3,41 et de 217 809 à 1,00 ; et enfin, pour les accruals anormaux, de -220 372 à 1,98 et de 7 123 à 2,04.

En limitant les valeurs extrêmes, l'exclusion des valeurs extrêmes a permis de recentrer les moyennes des variables discrétionnaires vers zéro et de réduire considérablement leur variabilité. En comparant les statistiques descriptives des variables après exclusion des valeurs extrêmes pour les deux configurations (séparée par pays et agrégée), on retrouve une tendance similaire à celle observée avant leur exclusion, sauf en ce qui concerne les valeurs maximales.

Comme avant l'exclusion des valeurs extrêmes, les moyennes des variables diffèrent entre les deux tableaux, affichant une légère augmentation dans la configuration agrégée. La moyenne des coûts discrétionnaires anormaux est de 0,01 dans le tableau 5 (séparé par pays) contre 0,05 dans le tableau 6 (agrégé pour tous les pays). La moyenne des coûts de production anormaux est de 0,02 dans le tableau 5 contre 0,04 dans le tableau 6, celle des flux de trésorerie anormaux (CFO) et des accruals anormaux est à 0,00 dans le tableau 5 ainsi que dans le tableau 6. De plus, la variabilité des données agrégées est plus élevée pour toutes les variables : 10 % dans le tableau 5 contre 11 % dans le tableau 6 pour les coûts discrétionnaires anormaux, 17 % dans le tableau 5 contre 20 % dans le tableau 6 pour les coûts de production anormaux, 10 % dans le tableau 5 contre 11 % dans le tableau 6 pour les CFO anormaux, et 12 % dans le tableau 5 contre 12 % dans le tableau 6 pour les accruals anormaux.

Contrairement à avant l'exclusion des valeurs extrêmes, la valeur maximale pour toutes les variables, sauf pour les coûts de production anormaux, est supérieure dans le tableau 6 comparé au tableau 5. Pour les coûts de production anormaux, la valeur maximale est de 1,91 dans le tableau 5 et 1,50 dans le tableau 6. Les flux de trésorerie anormaux (CFO) suivent la tendance inverse avec une valeur maximale pareille dans les deux tableaux de 1,00. Les accruals anormaux affichent une valeur maximale de 1,68 dans le tableau 5 contre 2,04 dans le tableau 6. Les coûts discrétionnaires anormaux suivent la même tendance avec une valeur de 1,00 dans le tableau 6 qui

est légèrement supérieure à celle du tableau 5 (0,98).

Ainsi, en comparaison avec les statistiques descriptives avant l'exclusion des valeurs extrêmes, la tendance générale reste similaire pour les moyennes et les écarts-types, avec une variabilité et des moyennes plus élevées dans la configuration agrégée. Ces résultats continuent de suggérer qu'en moyenne, les cours discrétionnaires peuvent être plus élevés dans un contexte global.

## **4.2 Analyse des cours discrétionnaires et des catastrophes naturelles**

Dans le cadre de l'analyse de l'impact des catastrophes naturelles sur les accruals discrétionnaires, il est essentiel de vérifier l'hypothèse de normalité. En effet, de nombreuses méthodes statistiques utilisées, telles que les tests  $t$ , reposent sur l'hypothèse que les données suivent une distribution normale. Cette hypothèse est fondamentale pour garantir la validité des statistiques inférentielles, telles que les  $p$ -values, les intervalles de confiance et les tests d'hypothèses. Une distribution normale des résidus permet de s'assurer que les conclusions tirées des résultats de régression ou des tests de comparaison sont robustes et précises. Toutefois, l'absence de normalité pourrait ne pas poser de problème dans cette analyse, notamment en raison de la taille importante de l'échantillon. Des théorèmes statistiques comme le théorème central limite (Billingsley, 1995) garantissent que, même si les données brutes ne suivent pas une distribution normale, la distribution des moyennes d'échantillons tend vers la normalité (Kwak & Kim, 2017). En conséquence, des statistiques paramétriques comme les tests- $t$  ou les intervalles de confiance peuvent demeurer valides (Lumley & al., 2002), permettant ainsi de tirer des conclusions fiables sur l'impact des catastrophes naturelles sur les accruals discrétionnaires.

### **4.2.1 Test de normalité des cours anormaux**

Pour analyser la normalité des cours discrétionnaires, une visualisation de la distribution des résidus a été effectuée, suivie d'un test de Jarque-Bera, sur les données excluant les valeurs aberrantes.

#### **Analyse préliminaire avec histogrammes:**

Dans un premier temps, des histogrammes ont été produits à partir des cours

discrétionnaires<sup>10</sup>. La Figure 2 et la Figure 3, présentées en annexe, représentent ces visualisations selon nos deux configurations : pays non agrégés et pays agrégés.<sup>11</sup>

Dans la figure 2, les histogrammes des variables de coûts discrétionnaires anormaux, de coûts de production anormaux, de CFO anormaux, et d'accruals anormaux, la concentration des valeurs autour de zéro est marquée, avec des valeurs extrêmes peu fréquentes. Cependant, la forme des distributions semble très pointue avec des queues allongées, ce qui est un signe de distribution leptokurtique.

Une distribution normale typique aurait une forme de cloche plus étalée, avec des queues moins prononcées. Ici, la concentration excessive de valeurs autour de zéro et la présence de quelques valeurs extrêmes indiquent une déviation de la normalité, suggérant des distributions non normales pour chacune de ces variables. Dans la figure 3, les distributions pour chaque variable restent également fortement centrées autour de zéro, avec un étalement similaire aux distributions non agrégées. Le regroupement des données de tous les pays renforce l'impression de forte concentration autour de zéro, accentuant encore plus l'aplatissement central et les queues étendues. Cela est cohérent avec une distribution non normale, car une distribution normale serait plus symétrique et moins pointue au centre. En conséquence, les distributions des résidus agrégés ne semblent pas suivre la normalité non plus.

#### **Confirmation par le test de Jarque-Bera:**

Pour corroborer les résultats observés visuellement, le test de Jarque-Bera est réalisé afin de vérifier la normalité des distributions, en complément des histogrammes qui indiquent déjà une déviation probable de la normalité. Les résultats sont présentés en annexe dans les tableaux 7 et 8.<sup>12</sup> Les deux tableaux confirment l'absence de normalité des accruals discrétionnaires, avec des statistiques de Jarque-Bera très élevées et des p-values de l'ordre de 0. Le test de Jarque-Bera, qui évalue la symétrie et le kurtosis des résidus, montre que ceux-ci s'écartent significativement de la normalité. Cela est en lien avec nos conclusions après la visualisation des histogrammes.

Cette non-normalité de nos courus discrétionnaires pourrait être expliquée par la nature des données, en particulier par le fait que les valeurs des courus discrétionnaires sont très proches de zéro. Lorsque les données sont faiblement dispersées ou concentrées autour de valeurs très basses, cela peut entraîner une

---

<sup>10</sup>Excluant les valeurs aberrantes

<sup>11</sup>Le code est fourni en annexe: page 83, bloc 2

<sup>12</sup>Le code est fourni en annexe: page 83, bloc 3



déviations de la distribution normale, même si les relations sous-jacentes entre les variables sont valides.

L'absence de normalité des résidus implique que les hypothèses fondamentales sous-tendant les tests  $t$  ne sont pas respectées dans cette étude. Cela signifie que les inférences statistiques telles que la significativité des tests pourraient être biaisées, voire incorrectes. Cependant, les échantillons étant suffisamment grands, la non-normalité des courus discrétionnaires ne devrait pas affecter de manière significative nos résultats.

#### 4.2.2 Statistiques descriptives des catastrophes naturelles

Pour analyser les catastrophes naturelles, des statistiques descriptives ont été examinées selon nos deux configurations pour chaque type de catastrophe. L'analyse des statistiques descriptives permet de mieux comprendre la distribution et les caractéristiques de chaque variable de courus discrétionnaires, ce qui est essentiel pour évaluer l'impact potentiel des catastrophes naturelles. Les résultats sont compilés dans le tableau 9, présenté en annexe, selon les deux configurations étudiées. Ils présentent respectivement les coûts discrétionnaires anormaux, les coûts de production anormaux, les CFO anormaux et les accruals anormaux.

**Coûts discrétionnaires anormaux :** Les coûts discrétionnaires en période d'absence de catastrophe naturelle présentent une moyenne de 0,02 dans la configuration séparée<sup>13</sup> et de 0,03 dans la configuration agrégée<sup>14</sup>, bien plus élevée que pour la majorité des catastrophes. Par exemple, en cas de tempête, les moyennes sont de 0,01 dans la configuration séparée et de 0,02 dans la configuration agrégée. En général, les valeurs moyennes dans la configuration agrégée sont supérieures à celles de la configuration séparée pour la plupart des types de catastrophes. Cela suggère que l'agrégation des données entraîne un lissage des coûts discrétionnaires anormaux, reflétant ainsi une réaction collective plus homogène face aux catastrophes au niveau global, contrairement aux réponses spécifiques par pays.

Ainsi, en l'absence de catastrophe naturelle, les entreprises semblent disposer d'une plus grande flexibilité pour engager des coûts discrétionnaires anormaux, potentiellement pour optimiser leurs dépenses stratégiques. En revanche, cette flexibilité diminue en période de catastrophe, où les contraintes limitent ces dépenses.

---

<sup>13</sup>En séparant les courus par pays.

<sup>14</sup>En agrégeant les courus pour tous les pays.

L'effet agrégé pourrait également souligner un comportement d'atténuation où, collectivement, les entreprises se restreignent davantage lors des catastrophes.

**Coûts de production anormaux :** La moyenne des coûts de production anormaux varie énormément selon le type de catastrophe naturelle. Tout de même, en période d'absence de catastrophe naturelle, est de 0,00 dans la configuration séparée et de 0,01 dans la configuration agrégée. Pendant les catastrophes, la plupart des moyennes augmentent légèrement; par exemple, lors d'une tempête, les moyennes sont de 0,01 dans la configuration séparée et de 0,03 dans la configuration agrégée. Dans la configuration agrégée, les moyennes sont globalement plus élevées que dans la configuration séparée pour la majorité des catastrophes. Ce phénomène pourrait refléter une standardisation ou une régularisation des coûts de production au niveau global, qui atténue les variations plus spécifiques observées dans chaque pays. L'écart-type est plus élevé dans la configuration agrégée, montrant une plus grande dispersion des coûts lorsque l'on prend en compte tous les pays.

Les entreprises semblent ajuster leurs coûts de production en période de catastrophe, une tendance particulièrement visible aussi bien dans les données séparées que les données agrégées par pays. Ceci, amplifiée dans l'analyse agrégée, suggère une résilience collective face aux pressions économiques des catastrophes.

**CFO anormaux :** Les CFO anormaux sont légèrement plus élevés en période d'absence de catastrophe naturelle avec une moyenne de -0,01 aussi bien dans la configuration séparée que dans la configuration agrégée. Lors de catastrophes comme les inondations, les moyennes sont légèrement inférieures (0,00 dans les deux configurations). La configuration agrégée produit des valeurs moyennes assez semblables mais des écarts-types légèrement plus importants pour la plupart des catastrophes. Cela pourrait indiquer que, bien que les CFO restent relativement stables, les flux de trésorerie agrégés montrent une plus grande variabilité, probablement due aux réponses différentes des pays face aux catastrophes.

Le CFO semble moins sensible aux événements de catastrophe, mais l'analyse agrégée révèle une dispersion plus élevée, ce qui pourrait signifier qu'en période de catastrophe, les flux de trésorerie sont globalement résilients, bien que leur variation augmente lorsque les données sont agrégées.

**Accruals anormaux :** En période d'absence de catastrophe naturelle, les accruals anormaux affichent des moyennes de -0,01 dans la configuration séparée et -0,02 dans la configuration agrégée, globalement plus négatives que celles observées

lors des catastrophes. Par exemple, lors d'une tempête, les accruals sont moins négatifs (-0,00 dans la configuration séparée et 0,00 dans la configuration agrégée). Les valeurs dans la configuration agrégée montrent généralement des accruals plus négatifs que dans la configuration séparée. Cela suggère que, lorsqu'on agrège les données de plusieurs pays, l'effet d'ajustement comptable devient plus prononcé, probablement en raison des différences dans les régulations et pratiques de comptabilité en réaction aux catastrophes dans les différents pays.

L'ajustement des accruals est plus prononcé en période d'absence de catastrophe naturelle, ce qui pourrait être attribué à des stratégies de gestion des résultats moins prioritaires lors des catastrophes. L'agrégation des accruals renforce cette tendance, indiquant un ajustement global plus important en l'absence de catastrophe et plus de variabilité.

Ainsi, l'analyse comparant les périodes d'absence de catastrophes naturelles aux périodes de catastrophe pourrait révéler plusieurs tendances distinctes dans les comportements financiers des entreprises : une réduction générale des coûts discrétionnaires en période de crise, une diminution des coûts de production, une légère baisse des flux de trésorerie opérationnels et une augmentation des accruals anormaux. Ces tendances pourraient indiquer que les entreprises adaptent leur gestion financière en réponse aux catastrophes, limitant les dépenses discrétionnaires et manipulant potentiellement les résultats pour améliorer leurs états financiers face aux contraintes et incertitudes accrues.

Les statistiques suggèrent que l'agrégation des résultats par pays pourrait atténuer certaines variations spécifiques, entraînant ainsi des moyennes et des écarts-types plus élevés dans la configuration agrégée. Cela pourrait indiquer une homogénéisation des pratiques financières en réponse aux catastrophes. En revanche, l'analyse séparée par pays pourrait révéler des réponses plus diversifiées et adaptées aux pressions spécifiques des catastrophes, suggérant ainsi des stratégies de gestion distinctes selon les contextes nationaux.

### **4.3 Approche utilisée pour notre étude**

Deux analyses distinctes sont effectuées pour déterminer si les catastrophes naturelles ont un impact sur les courus discrétionnaires des entreprises. La première analyse évalue l'effet de groupes de catastrophes naturelles, basés sur les cinq catégories définies dans la base de données EM-DAT, en utilisant nos deux configu-

rations : l'analyse individuelle par pays et l'analyse agrégée regroupant tous les pays. La seconde analyse examine l'influence de chaque type spécifique de catastrophe naturelle sur les courus, toujours selon nos deux configurations. Nous avons regroupé les catastrophes naturelles en cinq grandes catégories, conformément aux classifications de l'EM-DAT : catastrophes hydrologiques, météorologiques, géophysiques, biologiques et climatologiques.

Pour chacun de ces groupes, une variable indicatrice a également été créée, prenant la valeur 1 lorsqu'au moins une catastrophe de ce groupe survient, et 0 sinon. En ajoutant ces nouvelles variables, nous avons ainsi créé cinq indicatrices supplémentaires, permettant une analyse plus large de l'impact des groupes de catastrophes naturelles sur les courus discrétionnaires des entreprises.

$$(5) \quad f(\text{Groupe de catastrophe}) = \begin{cases} 1, & \text{Si ce groupe de catastrophe naturelle survient} \\ 0, & \text{sinon} \end{cases}$$

Ensuite, pour chaque catastrophe naturelle, nous avons ajouté une nouvelle variable indicatrice prenant la valeur 1 lorsqu'une catastrophe de ce type se produit au cours de la période d'observation, et 0 en l'absence de cette catastrophe. Cette opération a été répétée pour chacun des 11 types de catastrophes naturelles identifiés dans les données, en plus des périodes sans catastrophe. Ainsi, 11 nouvelles variables indicatrices ont été intégrées dans notre modèle, couvrant des catastrophes comme les inondations, les tempêtes, les tremblements de terre, etc.<sup>15</sup>

$$(6) \quad f(\text{Type de catastrophe}) = \begin{cases} 1, & \text{Si ce type de catastrophe naturelle survient} \\ 0, & \text{sinon} \end{cases}$$

---

<sup>15</sup>Le code est fourni en annexe: page 82, bloc 4

Nous avons décidé de faire également une analyse agrégée. Pour ce faire, toutes les catastrophes naturelles ont été incluses, à l'exception de celles classées sous la catégorie "Autres".<sup>16</sup> Ces catastrophes, bien que répertoriées dans l'EM-DAT, sont principalement d'origine technologique (comme les accidents industriels) et ne correspondent pas aux catastrophes naturelles ciblées par cette étude. Leur inclusion aurait pu biaiser les résultats en raison de leur impact potentiellement différent sur les finances des entreprises.

Enfin, nous avons réalisé une série de régressions, en utilisant chaque type de catastrophe naturelle, l'agrégation de toutes les catastrophes naturelles, et les cinq groupes de catastrophes comme variables explicatives. Chaque régression a été effectuée deux fois : d'abord sur les données par pays, pour capter des variations spécifiques à chaque contexte national, puis sur l'ensemble des pays combinés, afin de détecter des effets globaux. Nous avons également opté pour une exclusion des catastrophes technologiques du groupe de périodes "sans catastrophe", de sorte à isoler l'effet des catastrophes naturelles en minimisant les biais.<sup>17</sup> Dans ces régressions, les variables dépendantes sont les quatre mesures des courus discrétionnaires : les coûts discrétionnaires anormaux, les coûts de production anormaux, les flux de trésorerie opérationnels anormaux et les accruals anormaux. Ces indicateurs sont des mesures clés pour évaluer comment les entreprises ajustent leur comptabilité en réponse aux pressions économiques induites par les catastrophes. Les variables explicatives sont les indicatrices de catastrophes que nous avons créées. Dans la première analyse, les 5 variables indicatrices associées aux groupes de catastrophes naturelles sont employées. Dans la seconde analyse, ce sont les 11 variables indicatrices associées aux types spécifiques de catastrophes naturelles qui sont utilisées comme variables indépendantes.

Ce cadre d'analyse nous permet de mieux comprendre si et comment les pratiques comptables des entreprises sont influencées par les catastrophes naturelles, que ce soit de manière spécifique à un type de catastrophe ou de manière générale par groupe de catastrophes. Les résultats sont compilés à la prochaine section.

---

<sup>16</sup>Le code est fourni en annexe: page 84, bloc 2

<sup>17</sup>Le code est fourni en annexe: page 84, bloc 2

## 5 Résultats

Malgré l'absence de normalité des données, la taille importante de notre échantillon permet de renforcer la validité des conclusions, comme le soutient Wooldridge (2013) à travers le théorème central limite. En effet, ce théorème stipule que, pour un grand nombre d'observations, la distribution des moyennes tend vers une distribution normale. De plus, l'exclusion des données extrêmes a permis de réduire l'influence des valeurs extrêmes, rendant ainsi les résultats plus robustes. Enfin, la méthodologie adoptée repose sur des tests et des modèles économétriques capables de traiter des données non parfaitement normales, ce qui assure la fiabilité des conclusions tirées de cette analyse.

### 5.1 Influence de chaque groupe de catastrophe naturelle sur les courus discrétionnaires

Dans cette première analyse, nous examinons l'influence de chaque groupe de catastrophe naturelle sur les courus discrétionnaires, en utilisant nos deux configurations: en séparant les courus par pays et en agrégeant les courus pour tous les pays.

La formule utilisée est l'équation (7):

$$Y_i = \beta_j \mathbb{1}_j + \epsilon_i \quad (7)$$

Avec:

- $Y_i$  l'une des variables dépendantes (coûts discrétionnaires anormaux, coûts de production anormaux, CFO anormaux, accruals anormaux)
- $\mathbb{1}_j$  est une fonction indicatrice qui prend la valeur 1 si l'observation  $i$  appartient au groupe de catastrophe  $j$ , et 0 sinon
- $\beta_j$  est le coefficient estimé, représentant l'impact du groupe de catastrophe  $j$  sur  $Y_i$ , c'est-à-dire la variation des courus discrétionnaires par rapport à une situation sans catastrophe naturelle.
- $\epsilon_i$  est le terme d'erreur.

En l'absence de manipulation des résultats,  $\sum_j \beta_j$  devrait être nulle à long terme. Cependant, comme démontré dans le test précédent, ce n'est pas le cas, ce qui pourrait indiquer une gestion des résultats par les entreprises en réponse aux catastrophes naturelles. De plus, si ce coefficient est statistiquement significatif et

différent de zéro, cela signifie que la survenance d'une catastrophe de type  $j$  est associée à des courus discrétionnaires anormaux, suggérant ainsi une potentielle manipulation des résultats financiers. Les tableaux 10a, 10b, 10c et 10d, présentés en annexe, offrent une synthèse des résultats selon les deux configurations que nous avons utilisées.

**Interprétation:**<sup>18</sup> Dans l'ensemble, l'agrégation des données conduit souvent à une amplification des effets, car elle regroupe les données de plusieurs pays, réduisant ainsi les variations dues aux particularités locales. Cela donne des résultats avec des moyennes et des écarts-types plus élevés qui renforcent la signification statistique des impacts des catastrophes sur les coûts discrétionnaires, les coûts de production, les flux de trésorerie, et les accruals.

**Agrégation de toutes les catastrophes naturelles :** L'impact des catastrophes naturelles<sup>19</sup> sur les courus discrétionnaires est statistiquement significatif pour toutes les variables étudiées, sauf pour les accruals anormaux dans la configuration séparée. En général, les coefficients associés dans les deux configurations (séparée et agrégée) montrent des variations importantes, avec des t-statistiques significatives. Sous la configuration séparée, les dépenses discrétionnaires anormales affichent une augmentation de 1,0 % par rapport aux périodes sans catastrophe. Tandis que sous la configuration agrégée, cette hausse est plus marquée (1,7 %). Concernant les coûts de production anormaux, sous la configuration séparée, nous notons une hausse de 1,3 % de ces coûts anormaux par rapport aux périodes sans catastrophe. Tandis que sous la configuration agrégée, cette augmentation est plus significative (3,8 %). Pour les flux de trésorerie anormaux, les résultats indiquent une augmentation anormale de 0,2 % des CFO sous la configuration séparée et de 0,6 % sous la configuration agrégée. Enfin, les accruals anormaux n'affichent aucune augmentation statistiquement significative des courus discrétionnaires anormaux sous la configuration séparée mais une augmentation de 0,2 % sous la configuration agrégée est notée.

---

<sup>18</sup>Les valeurs absolues seront considérés pour les comparaisons

<sup>19</sup>L'agrégation des catastrophes naturelles

### **5.1.1 Influence de chaque groupe de catastrophe naturelle sur les coûts discrétionnaires anormaux**

Les catastrophes géophysiques présentent le plus fort impact sur les coûts discrétionnaires anormaux dans la configuration séparée avec une valeur de 1,9 %, indiquant une augmentation notable des dépenses discrétionnaires en réponse à ces événements. Cette tendance se confirme également dans la configuration agrégée, où le coefficient des catastrophes géophysiques augmente à 4,3 %. En revanche, le coefficient minimum pour les dépenses discrétionnaires est noté pour les catastrophes météorologiques en configuration séparée, avec une valeur de 0,9 %. En configuration agrégée, ce minimum demeure également pour les catastrophes météorologiques (-1,4 %) <sup>20</sup>.

Les coefficients pour les dépenses discrétionnaires anormales sont positifs pour toutes les catégories de catastrophes sauf climatologiques, suggérant que les entreprises augmentent généralement leurs dépenses discrétionnaires en réponse à ces événements. Cela pourrait être motivé par une volonté de montrer un engagement envers la reprise, ou pour enregistrer des dépenses accrues justifiées par l'impact de la catastrophe.

### **5.1.2 Influence de chaque groupe de catastrophe naturelle sur les coûts de production anormaux**

Les catastrophes géophysiques affichent de nouveau le coefficient maximum sous la configuration séparée avec une valeur de 4,4 %, traduisant une hausse importante des coûts de production à la survenance de ces événements. Ce maximum est amplifié en configuration agrégée, atteignant 11,2 %. Quant au coefficient minimum statistiquement significatif, il est observé pour les catastrophes hydrologiques sous la configuration séparée, avec une valeur négative de -0,4 %. Sous la configuration agrégée, le plus faible impact est noté pour les catastrophes biologiques (-1,7 %). Ce minimum se confirme en configuration agrégée, atteignant 1,1 %.

Les catastrophes météorologiques, géophysiques et technologiques présentent des coefficients positifs, ce qui pourrait s'expliquer par les dépenses de réparation et de rétablissement liées aux infrastructures touchées par ces types d'événements. En revanche, les catastrophes hydrologiques, biologiques et climatologiques présentent

---

<sup>20</sup>Les coefficients pour les catastrophes biologiques et climatologiques ne sont pas statistiquement significatifs dans les deux configurations.



des coefficients négatifs, bien que non significatifs pour les deux derniers. Ces coûts de production réduits pourraient refléter des ajustements où les entreprises limitent les dépenses dans certains secteurs pour atténuer l'impact financier de la catastrophe.

### **5.1.3 Influence de chaque groupe de catastrophe naturelle sur les CFO anormaux**

Concernant les CFO anormaux, ce sont les catastrophes géophysiques qui présentent le plus fort impact sous la configuration séparée avec un coefficient de 0,4 %. Ce maximum est accru en configuration agrégée (1,4 %). En termes de coefficient minimum, les catastrophes météorologiques montrent la plus faible influence avec une valeur de 0,2 % en configuration séparée et 0,6 % en configuration agrégée.

Les flux de trésorerie anormaux montrent des coefficients aussi bien positifs que négatifs pour les différentes catastrophes.

Cela pourrait traduire une gestion plus prudente des liquidités en période de catastrophe. L'augmentation des flux de trésorerie opérationnels pourrait également donner l'image d'une entreprise capable de générer des liquidités malgré la crise, renforçant ainsi la confiance des investisseurs dans sa résilience financière. La baisse des flux de trésorerie opérationnels pourrait quant à elle signifier une difficulté à générer des liquidités à cause de la crise. <sup>21</sup>.

### **5.1.4 Influence de chaque groupe de catastrophe naturelle sur les accuals anormaux**

Le plus fort impact statistiquement significatif est observé pour les catastrophes géophysiques sous la configuration séparée avec une valeur de -0,2 %, ce qui indique une diminution des accuals anormaux. Cette réduction s'efface en configuration agrégée avec un coefficient de 0 %, malgré la non-significativité du résultat. Le seul coefficient statistiquement significatif sous la configuration séparée est associé aux catastrophes météorologiques avec une valeur de 0,4 %. À l'exception de ces deux coefficients et de ceux associés aux catastrophes technologiques, tous les autres coefficients ne sont pas statistiquement significatifs. Les accuals anormaux présentent

---

<sup>21</sup>Les coefficients pour les catastrophes biologiques et climatologiques ne sont pas statistiquement significatifs dans les deux configurations. Les coefficients pour les catastrophes hydrologiques sont statistiquement significatifs que dans la configuration agrégée.

des coefficients aussi bien négatifs que positifs. Les coefficients négatifs pourraient suggérer des tentatives de réduire artificiellement le résultat comptable en sous-estimant les revenus ou en augmentant les charges (par exemple, par des provisions ou des amortissements), dans le but d'améliorer artificiellement les bénéfices futurs. À l'inverse, les coefficients positifs pourraient refléter des stratégies visant à accroître artificiellement le résultat comptable en surestimant les revenus ou en diminuant les charges, afin de présenter des performances financières plus favorables à court terme, notamment pour satisfaire les attentes du marché ou atteindre des objectifs internes.

Par ailleurs, il est essentiel de noter que l'impact des différents groupes de catastrophes naturelles est plus marqué sur les courus discrétionnaires résultant de l'activité réelle que sur les ajustements comptables (accruals anormaux), probablement en raison des contraintes réglementaires limitant la possibilité de manipulation des accruals.

Pour résumer, en agrégeant tous les groupes de catastrophes naturelles, nous notons un impact significatif sur toutes les variables étudiées, tant dans les configurations séparées qu'agrégées. Certaines catégories de catastrophes ont un impact significatif sur l'ensemble des variables, tandis que d'autres n'en affectent que certaines. Les catastrophes de type météorologique et géophysique montrent un impact notable sur tous les courus discrétionnaires. Pour ces groupes, il semble que les entreprises ajustent à la fois leurs activités réelles et leurs données comptables, bien que l'impact des ajustements comptables demeure plus faible que celui des modifications d'activité réelle. En revanche, pour les catastrophes de type hydrologique, les entreprises ont tendance à ajuster principalement leurs activités réelles, avec un effet encore plus limité sur les ajustements comptables. Il est important de noter que le groupe de catastrophe de type technologique a un impact significatif sur toutes les variables étudiées.

## **5.2 Influence de chaque type de catastrophe naturelle sur les courus discrétionnaires**

Dans cette deuxième analyse, nous examinons l'influence de chaque type spécifique de catastrophe naturelle sur les courus discrétionnaires, en utilisant nos deux configurations: en séparant les courus par pays et en agrégeant les courus pour tous les pays.

La formule utilisée est l'équation (8):

$$Y_i = \beta_j \mathbb{1}_j + \epsilon_i \quad (8)$$

Avec:

- $Y_i$  l'une des variables dépendantes (coûts discrétionnaires anormaux, coûts de production anormaux, CFO anormaux, accruals anormaux).
- $\mathbb{1}_j$  est une fonction indicatrice qui prend la valeur 1 si l'observation  $i$  appartient au type de catastrophe  $j$ , et 0 sinon.
- $\beta_j$  est le coefficient estimé, représentant l'impact de la catastrophe  $j$  sur  $Y_i$ , c'est-à-dire la variation des cours discrétionnaires par rapport à une situation sans catastrophe naturelle.
- $\epsilon_i$  est le terme d'erreur.

Les tableaux 11a, 11b, 11c et 11d, présentés en annexe, offrent une synthèse des résultats selon les deux configurations que nous avons utilisées.

La configuration agrégée<sup>22</sup> tend à être plus significative car elle consolide les données de plusieurs pays, ce qui réduit les variations dues à des circonstances locales ou sectorielles spécifiques. En regroupant les observations, les effets statistiques de chaque catastrophe deviennent plus robustes, atténuant les anomalies ou variations extrêmes que l'on peut observer dans des échantillons plus restreints. Cela entraîne des moyennes et des écarts-types plus élevés dans la configuration agrégée, donnant une image plus lissée et souvent plus significative de l'impact global des catastrophes. Par ailleurs, l'agrégation pourrait aussi refléter une homogénéisation des pratiques de gestion de crise, rendant les stratégies d'ajustement des coûts et des accruals plus uniformes et donc statistiquement plus significatives lorsqu'elles sont évaluées ensemble.

### 5.2.1 Influence de chaque type de catastrophe naturelle sur les coûts discrétionnaires anormaux

Pour l'analyse des coûts discrétionnaires anormaux, le tremblement de terre est la catastrophe ayant le plus grand impact dans les deux configurations. En configuration séparée, elle atteint 2,1 % tandis qu'en configuration agrégée, elle baisse

---

<sup>22</sup>Les valeurs absolues seront considérées pour les comparaisons

légèrement à 1,5 %, suggérant que l'agrégation atténue cet effet. À l'inverse, l'inondation montre l'impact le plus faible sur les coûts discrétionnaires anormaux, avec des valeurs de 0,5 % en configuration séparée et 0,3 % en configuration agrégée. Les coefficients positifs des coûts discrétionnaires anormaux pour la plupart des catastrophes indiquent que les entreprises ont tendance à augmenter ces dépenses pendant les périodes de crise.

Cette augmentation peut refléter une stratégie de gestion des impressions où les entreprises cherchent à démontrer leur engagement dans la reprise post-catastrophe pour obtenir un soutien positif des parties prenantes. Ainsi, en augmentant ces dépenses, les entreprises peuvent renforcer leur image auprès des investisseurs, en apparence proactives dans la gestion de crise. L'augmentation pourrait également être liée à des provisions ou à des dépenses pour sécuriser l'entreprise, comme des réparations ou des adaptations structurelles.

### **5.2.2 Influence de chaque type de catastrophe naturelle sur les coûts de production anormaux**

Pour les coûts de production anormaux, l'impact le plus significatif provient toujours des tremblements de terre, atteignant 5,6 % en configuration séparée et augmentant à 12,4 % en configuration agrégée. Cet écart montre qu'une consolidation des données accroît les coûts de production lors de catastrophes majeures comme les tremblements de terre. À l'opposé, les mouvements de masse humide (-0,7 %) produisent les plus faibles impacts dans les deux configurations.

Un coefficient positif (par exemple, lors d'un tremblement de terre) signifie que les catastrophes induisent une hausse des coûts de production, souvent due à des besoins accrus en ressources pour maintenir ou adapter les opérations en période de crise. Les entreprises peuvent devoir augmenter les dépenses en logistique, réparations, ou sécurité pour garantir la continuité de leurs activités malgré les perturbations externes. À l'inverse, les coefficients négatifs, tels que ceux observés pour les mouvements de masse humides ou les mouvements de masse secs, pourraient indiquer que certaines entreprises réduisent leurs activités non essentielles ou ajustent leur production pour diminuer les coûts lors de catastrophes aux impacts temporaires ou aux effets localisés. Ces baisses de coûts peuvent aussi refléter des choix stratégiques visant à préserver les marges bénéficiaires en période de crise.

### **5.2.3 Influence de chaque type de catastrophe naturelle sur les CFO anormaux**

Concernant les flux de trésorerie opérationnels anormaux, les tremblements de terre engendrent le plus fort impact en configuration séparée, avec une augmentation de 0,5 %, et 1,5 % en configuration agrégée. Les tempêtes sont celles qui ont l'effet le plus bas dans la configuration séparée avec 0,2 %. Les inondations ont quant à elles l'effet le plus bas dans la configuration agrégée (0,3 %). Les flux de trésorerie opérationnels anormaux affichent des coefficients positifs, ce qui pourrait traduire une gestion plus prudente des liquidités en période de catastrophe. Le maintien ou l'augmentation modérée des flux de trésorerie opérationnels pourrait également donner l'image d'une entreprise capable de générer des liquidités malgré la crise, renforçant ainsi la confiance des investisseurs dans sa résilience financière. Tout de même, pour certaines des catastrophes naturelles, les coefficients sont négatifs, bien que statistiquement non significatifs. La baisse des flux de trésorerie opérationnels pourrait quant à elle signifier une difficulté à générer des liquidités à cause de la crise.

### **5.2.4 Influence de chaque type de catastrophe naturelle sur les accruals anormaux**

Enfin, pour les accruals anormaux, les températures extrêmes restent les catastrophes ayant le plus d'impact, atteignant 0,4 % en configuration séparée et 0,8 % en configuration agrégée. En revanche, les inondations montrent l'impact le moins important en configuration séparée (0,2 %), tandis que les tempêtes présentent l'impact le moins significatif en configuration agrégée (0,2 %). Les coefficients positifs pourraient indiquer que l'entreprise a choisi d'augmenter ses bénéfices en reconnaissant des revenus de manière anticipée ou en réduisant les charges. Cela peut être perçu comme une stratégie visant à améliorer les résultats financiers à court terme, en faisant apparaître des profits plus élevés.

À l'opposé, des accruals négatifs (comme pour les tremblements de terre) pourraient indiquer des tentatives de réduire artificiellement le résultat comptable en sous-estimant les revenus ou en augmentant les charges (ex. provisions, amortissements). Cela peut être stratégique pour lisser les bénéfices. De plus, en enregistrant des charges plus élevées aujourd'hui, l'entreprise peut améliorer artificiellement ses bénéfices futurs (ex. constitution de provisions excessives qui seront reprises plus

tard). Par ailleurs, il est essentiel de noter que l'impact des différents types de catastrophes naturelles est plus marqué sur les courus discrétionnaires résultant de l'activité réelle que sur les ajustements comptables (accruals anormaux), probablement en raison des contraintes réglementaires limitant la possibilité de manipulation des accruals.

Pour résumer, certaines catastrophes ont un impact significatif sur toutes les variables étudiées, tandis que d'autres n'affectent que certaines d'entre elles. Les inondations, les températures extrêmes, les tempêtes et les tremblements de terre montrent un impact notable sur toutes les variables. Pour ces catastrophes, il semble que les entreprises ajustent à la fois leurs activités réelles et leurs données comptables, bien que l'impact des ajustements comptables demeure plus faible que celui des modifications d'activité réelle. En revanche, pour les mouvements de masse (humides), les entreprises tendent à ajuster principalement leurs activités réelles, avec un recours plus limité aux ajustements comptables.

## 6 Limites

Bien que notre étude ait fourni des aperçus précieux sur l'impact des catastrophes sur les courus discrétionnaires, plusieurs limites doivent être reconnues. Nous avons constaté que nos courus discrétionnaires ne suivaient pas une distribution normale, ce qui soulève des préoccupations quant à la robustesse statistique des coefficients obtenus. La non-normalité peut affecter la validité des tests statistiques appliqués, comme les tests de signification. La décision de restreindre notre échantillon aux pays avec un PIB par habitant inférieur à 15 000 \$ US peut également limiter la généralisation des conclusions tirées. Les dynamiques économiques, et les systèmes de gestion des risques varient considérablement entre les économies en développement et les économies plus avancées. Par exemple, les entreprises dans des pays développés pourraient disposer de meilleures ressources financières, de systèmes d'assurance plus sophistiqués, et d'un cadre réglementaire plus robuste, ce qui peut influencer leur réponse face aux catastrophes. Par conséquent, les résultats de notre étude pourraient ne pas être directement applicables à des contextes où le PIB par habitant est plus élevé, et les différences de comportement en matière de courus discrétionnaires pourraient être significativement différentes.

Les deux approches utilisées dans notre étude présentent également certaines

limites. En effet, même si le modèle de Jones modifié permet d'isoler la part discrétionnaire des résultats comptables, il demeure limité face aux formes de manipulation issues de décisions opérationnelles. De plus, le modèle peut être biaisé en présence de chocs économiques non liés à une manipulation, comme l'ont souligné Healy & Wahlen (1999). L'approche de Roychowdhury cible précisément ces limites, mettant en lumière des pratiques souvent plus coûteuses et moins visibles, bien qu'elle repose sur des hypothèses plus complexes. Malgré les limites propres à chaque modèle, leur combinaison offre des perspectives complémentaires, permet une analyse plus riche et nuancée de la gestion des résultats.

L'agrégation des données dans certaines analyses, dans le but d'obtenir une vue d'ensemble, présente des inconvénients. En effet, l'agrégation peut masquer des spécificités locales et des réponses différenciées des entreprises face aux catastrophes, offrant ainsi une vision simplifiée et potentiellement trompeuse de la réalité. Par exemple, différentes régions au sein d'un même pays peuvent être affectées par des catastrophes de manière très différente en raison de divers facteurs tels que la géographie, les infrastructures locales et la capacité d'adaptation des entreprises. En omettant ces variations, nous risquons de ne pas saisir pleinement la complexité des réponses des entreprises aux catastrophes, limitant ainsi la portée de nos conclusions. Une analyse plus granulaire, prenant en compte les spécificités locales et les disparités régionales, aurait pu fournir des insights plus nuancés et représentatifs.

Étant donné que notre étude se concentre sur des pays à faible revenu, il existe un risque de biais de sélection. Les pays inclus dans notre échantillon peuvent avoir des caractéristiques particulières qui influencent les résultats, comme une vulnérabilité plus élevée aux catastrophes ou des stratégies de gestion des risques spécifiques. Ces biais peuvent fausser les conclusions et limiter la capacité à extrapoler les résultats à d'autres contextes, notamment dans les pays à revenu intermédiaire ou élevé.

Enfin, la temporalité des données utilisées dans notre analyse peut poser des défis. Les réponses des entreprises aux catastrophes peuvent évoluer dans le temps, et les effets observés dans notre étude peuvent ne pas refléter les tendances futures. Les changements dans la réglementation, les améliorations technologiques et l'évolution des comportements des investisseurs peuvent influencer les décisions de la gestion des courus discrétionnaires au fil du temps. Ainsi, il est essentiel de reconnaître que nos conclusions sont basées sur un instantané de la réalité, et que des études futures pourraient être nécessaires pour évaluer l'évolution des réponses

des entreprises face aux catastrophes naturelles.

## 7 Conclusion

Dans cette étude, nous avons cherché à comprendre l'impact des catastrophes naturelles sur les courus discrétionnaires dans les pays ayant un PIB par habitant inférieur à 15 000 \$ US de 2010. Pour ce faire, nous avons utilisé deux modèles : celui de la manipulation des activités réelles de Roychowdhury (2006) et le modèle cross-sectionnel de Dechow, Sloan et Sweeney (1995), qui est une extension du modèle original de Jones (1991).

Nous avons étudié plusieurs types de catastrophes, notamment hydrologiques, météorologiques, biologiques, géophysiques et climatologiques, ainsi qu'une configuration regroupant toutes les catastrophes. De plus, pour obtenir une vision complète et comparative, nous avons effectué notre analyse selon deux configurations distinctes : d'une part, en séparant les données par pays afin d'évaluer les impacts spécifiques de chaque type de catastrophe sur les économies individuelles ; et d'autre part, en agrégeant les données pour examiner les effets combinés au niveau global. Cette double approche nous a permis de mesurer l'influence de chaque type de catastrophe sur différents paramètres financiers liés aux courus discrétionnaires et aux décisions financières des entreprises.

Les résultats révèlent que, de manière générale, la présence d'une catastrophe naturelle entraîne une augmentation des coûts discrétionnaires anormaux et des flux de trésorerie opérationnels anormaux, indépendamment du type de catastrophe. L'augmentation des coûts discrétionnaires anormaux pourrait être le résultat d'une hausse des investissements visant à reconstruire les infrastructures endommagées ou d'une intensification des dépenses en publicité et en promotion pour stimuler la demande et maintenir l'activité de l'entreprise dans un contexte d'incertitude. De même, la hausse des flux de trésorerie opérationnels (CFO) pourrait résulter d'une gestion plus rigoureuse des liquidités, notamment par la renégociation des conditions de paiement avec les fournisseurs afin d'optimiser la trésorerie disponible.

L'impact des catastrophes naturelles sur les coûts de production varie, en revanche, selon le type de catastrophe, sans qu'une tendance claire ne se dégage. Concernant les ajustements comptables, aucune tendance généralisée n'est observée, à l'instar des coûts de production anormaux. Toutefois, pour les catastrophes



ayant un impact statistiquement significatif, une augmentation des accruals est généralement constatée. Cela pourrait suggérer que les entreprises cherchent à augmenter artificiellement leur résultat comptable en surévaluant les revenus ou en diminuant les charges, ce qui peut être interprété comme une stratégie de gestion du résultat en période de crise.

Par ailleurs, l'impact des catastrophes naturelles sur les courus discrétionnaires anormaux est plus marqué lorsque l'on utilise les variables mesurant l'activité réelle que lorsqu'on évalue les ajustements comptables des accruals. Cette différence pourrait s'expliquer par les contraintes qui limitent la gestion comptable des résultats. En effet, bien que les accruals puissent être utilisés pour lisser les bénéfices, leur manipulation reste encadrée par des normes comptables et soumise au contrôle des auditeurs. En période de crise, la marge de manœuvre pour ajuster les résultats comptables peut être réduite en raison d'une surveillance accrue des régulateurs et des investisseurs. Dans ce contexte, les entreprises privilégient une gestion proactive de leur liquidité et de leurs opérations à court terme, plutôt que de recourir à des ajustements comptables.

En somme, notre étude met en lumière les divers comportements financiers des entreprises face aux chocs externes que représentent les catastrophes naturelles, et illustre l'importance d'une analyse multi-niveaux pour bien saisir ces dynamiques. Ces conclusions offrent une base pour de futures recherches et soulignent l'importance de la prise en compte des catastrophes dans la gestion des accruals et dans les décisions financières stratégiques des entreprises.

## 8 Références

Barnett, Victor; Lewis, Toby “Outliers in Statistical Data (3rd ed.)”, John Wiley & Sons, 1994.

Billingsley, Patrick “Probability and Measure”, John Wiley & Sons, 1995.

Burgstahler, David; Dichev, Ilia “Earnings Management to Avoid Earnings Decreases and Losses”. *Journal of Accounting and Economics*, 24(1), 1997, p. 99-126.

Cohen, Daniel A.; Zarowin, Paul “Accrual-Based and Real Earnings Management Activities around Seasoned Equity Offerings”, *Journal of Accounting and Economics*, Vol. 50, no 1, 2010, p. 2-19

Cohen, Daniel A.; Dey, Aiysha; Lys, Thomas Z. “Real and Accrual-Based Earnings Management in the Pre- and Post-Sarbanes-Oxley Periods”. *The Accounting Review*, 83(3), 2008, p. 757-787.

Dechow, Patricia M.; Dichev, Ilia D. “The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors”, *The Accounting Review*, 77(s-1), 2002, p. 35-59.

Dechow, Patricia M.; Sloan, Richard G.; Sweeney, Amy P. “Detecting earnings management”, *Accounting Review*, Vol. 70, no 2, 1995, p. 193-225.

Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome “The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Springer, 2009.

Healy, Paul M.; Wahlen, James M. “A Review of the Earnings Management Literature and Its Implications for Standard Setting”, *Accounting Horizons*, 13(4), 1999, pp. 365-383.

Jones, Jennifer “Earnings Management During Import Relief Investigations”, *Journal of Accounting Research*, 29(2), 1991, p. 193-228.

Karpoff, Jonathan M.; Lee, D. Scott; Martin, Gerald S. “The Consequences to Managers for Financial Misrepresentation”, *Journal of Financial Economics*, 88(2), 2008, p. 193-215.

Kothari, Suresh P.; Mizik, Natalie; Roychowdhury, Sugata “Managing for the Moment: The Role of Earnings Management via Real Activities versus Accruals in SEO Pricing”. *The Accounting Review*, 91(2), 2016, p. 579-608.

Kothari, Suresh P.; Leone, Andrew J.; Wasley, Charles E. “Performance matched discretionary accrual measures”. *Journal of Accounting and Economics*, 39(1), 2005, pp. 163–197.

Kwak, Sang Gyu; Kim, Jong Hae “Statistical Data Preparation: Management of Missing Values and Outliers”, *Korean Journal of Anesthesiology*, Vol. 70, no 4, 2017, p. 407-411.

Lumley, Thomas; Diehr, Paula; Emerson, Scott; Chen, Lu “The Importance of the Normality Assumption in Large Public Health Data Sets”, *Annual Review of Public Health*, Vol. 23, no 1, 2002, p. 151-169.

Marciukaityte, Donata “Earnings Management in Corporate Financing Decisions”. *Journal of Corporate Finance*, 17(4), 2011, p. 1113-1130.

Roychowdhury, Sugata “Earnings Management through Real Activities Manipulation”, *Journal of Accounting and Economics*, Vol. 42, no 3, 2006, p. 335-370.

Rousseeuw, Peter J.; Leroy, Annick M. “Robust Regression and Outlier Detection”, Wiley-Interscience, 1987.

Teoh, Siew Hong; Welch, Ivo; Wong, Tjun Jen “Earnings Management and the Underperformance of Seasoned Equity Offerings”. *Journal of Financial Economics*, 50(1), 1998, p. 63-99.

Wooldridge, Jeffrey M. "Introductory Econometrics: A Modern Approach". 5th ed.  
South-Western Cengage Learning, 2013.

Liste des tableaux :

Le tableau 1 présente le nombre d’observations par type de catastrophe naturelle et par année, pour les pays avec un PIB/habitant < 15 000 \$US.

Obs. / Année	Sécheresse	Trembl. de terre	Épidémie	Temp. extrême	Inondation	Mvt. masse sec	Mvt. masse humide	Pas de Catastro- phe	Catastrophe tech- nologique	Tempête	Act. vol- canique	Feu de forêt
2007	0	0	0	0	6	0	0	3	20	2	0	0
2008	0	0	0	4	178	0	289	57	138	51	1	0
2009	0	106	7	1	776	0	74	325	2124	1009	0	0
2010	0	0	0	1438	825	0	158	973	3622	1087	0	0
2011	35	0	332	2440	2029	0	0	1299	2768	1146	0	112
2012	0	356	149	3143	2024	0	203	2178	3421	538	0	0
2013	47	1386	23	387	1279	2197	6	1412	2530	2753	0	0
2014	24	125	99	546	1276	0	0	1781	7366	1726	0	3
2015	23	49	36	0	1894	0	0	1854	9223	667	0	2
2016	9	6118	71	5735	931	0	0	1895	2621	841	0	4
2017	0	0	3	46	1667	0	352	1736	10874	274	0	178
2018	13	1631	8	393	1572	0	0	1608	9353	379	444	0
2019	9	0	288	9	807	0	503	1888	9171	2528	0	0
2020	0	68	113	30	1946	0	0	2323	7596	3168	230	0
2021	0	554	0	0	2969	0	0	2290	4853	8347	0	9
2022	41	1886	2	194	1589	0	5492	1723	5305	864	80	0
2023	0	619	0	222	402	0	46	1772	11282	2748	0	0
2024	0	4	0	63	15	0	0	1291	300	786	0	0

Tableau 1: Nombre d’observations par année

Le nombre d'observations pour chaque type de catastrophe naturelle par pays, est présenté dans le tableau 2 pour les pays ayant un PIB par habitant inférieur à 15 000 \$ US.

Obs. / Pays	Sécheresse	Trembl. de terre	Épidémie	Temp. extrême	Inondation	Mvt. masse sec	Mvt. masse humide	Pas de Catastro- phe	Catastrophe tech- nologique	Tempête	Act. vol- canique	Feu de forêt
Argentina	0	0	0	57	3	0	39	0	71	0	0	0
Burkina Faso	4	0	0	0	9	0	0	5	18	0	0	0
Bangladesh	0	468	0	373	278	0	0	2	2383	0	0	0
Bulgaria	0	0	0	39	424	0	0	286	73	0	0	0
Brazil	0	0	318	0	128	0	0	0	552	0	0	0
Botswana	23	0	0	0	62	0	0	171	0	0	0	0
Chile	0	0	0	0	134	0	0	0	170	0	0	112
China	0	0	0	5128	0	2197	5432	117	30616	5875	0	0
Ivory Coast	0	0	0	0	173	0	0	0	125	0	0	0
Colombia	0	95	48	0	131	0	156	8	284	0	1	0
Cyprus	0	0	0	16	0	0	0	57	22	0	0	9
Ecuador	0	6	0	0	23	0	6	3	17	0	0	0
Egypt	0	0	0	0	211	0	0	0	1868	0	0	0
Estonia	0	0	0	7	0	0	0	45	7	0	0	0
Guernsey	0	0	0	0	0	0	0	803	0	0	0	0
Ghana	0	0	37	0	73	0	0	0	120	19	0	0
Croatia	0	68	0	148	634	0	0	61	80	0	0	98
Hungary	0	16	0	14	10	0	0	6	12	0	0	0

Obs. / Pays	Sécheresse	Trembl. de terre	Épidémie	Temp. extrême	Inondation	Mvt. masse sec	Mvt. masse humide	Pas de Catastro- phe	Catastrophe tech- nologique	Tempête	Act. vol- canique	Feu de forêt
Indonesia	0	2060	0	0	2657	0	428	97	382	0	438	0
India	0	5224	0	3894	0	0	289	95	32101	2721	0	0
Jamaica	21	0	0	0	0	0	0	388	0	161	0	0
Jordan	0	0	0	278	406	0	0	1050	989	0	0	0
Kazakhstan	0	0	0	0	39	0	0	0	43	0	0	0
Kenya	35	0	116	0	121	0	0	1	364	0	0	0
Lebanon	0	0	0	0	0	0	0	45	9	32	0	0
Liberia	0	0	5	0	4	0	0	8	16	0	0	0
Sri Lanka	0	0	248	0	1785	0	0	0	254	1120	0	0
Lithuania	0	0	0	24	47	0	0	80	0	0	0	0
Morocco	0	0	0	248	75	0	0	0	536	101	0	0
Mexico	0	119	0	117	87	0	0	0	695	0	0	0
Marshall Is- lands	75	0	0	0	70	0	0	433	0	47	0	0
Mauritius	0	0	0	0	39	0	0	417	26	279	0	0
Malawi	0	0	9	0	56	0	0	5	0	67	0	0
Malaysia	0	0	0	0	4840	0	0	784	4831	0	0	0
Namibia	24	0	0	0	87	0	0	34	0	0	0	0
Nigeria	0	0	199	0	125	0	0	19	1349	0	0	0

Obs. / Pays	Sécheresse	Trembl. de terre	Épidémie	Temp. extrême	Inondation	Mvt. masse sec	Mvt. masse humide	Pas de Catastro- phe	Catastrophe tech- nologique	Tempête	Act. vol- canique	Feu de forêt
Pakistan	0	0	0	0	617	0	448	32	2569	689	0	0
Panama	19	0	0	0	73	0	0	0	12	12	0	0
Peru	0	0	0	0	436	0	75	4	624	0	80	0
Philippines	0	0	0	0	1423	0	203	50	256	1156	230	0
Papua New Guinea	0	9	0	0	28	0	14	16	19	12	6	0
Poland	0	0	0	1383	175	0	0	0	19	0	0	0
Palestine	0	0	0	69	13	0	0	229	0	35	0	0
Romania	0	0	0	198	566	0	0	40	0	102	0	0
Russia	0	0	0	73	0	0	0	0	159	1	0	0
Serbia	0	0	0	104	201	0	0	27	0	0	0	0
Thailand	0	0	0	570	3059	0	0	713	3484	1344	0	0
Tunisia	0	0	0	0	80	0	0	144	603	0	0	76
Turkey	0	0	0	0	138	0	29	0	188	0	0	0
Tanzania	0	0	0	0	82	0	0	0	43	29	0	0
Uganda	0	0	37	0	29	0	0	0	58	0	0	0
Ukraine	0	0	0	61	33	0	0	0	139	0	0	0
Venezuela	0	0	14	0	14	0	0	0	15	0	0	0
Vietnam	0	0	0	0	657	0	0	0	1379	2304	0	0
South Africa	0	0	0	264	456	0	0	0	1571	571	0	0

Tableau 2: Nombre d'observations par pays



Les statistiques descriptives des coûts discrétionnaires associés à la manipulation des activités réelles, selon le modèle de Roychowdhury (2006) (coûts discrétionnaires anormaux, coûts de production anormaux et CFO anormaux), ainsi que des accruals anormaux basés sur le modèle de Jones modifié (Dechow, Sloan et Sweeney, 1995), sont présentées dans les tableaux 3 et 4 selon deux configurations : l'agrégation des données de tous les pays et une configuration où les pays sont analysés séparément.

	<b>Coûts discrétionnaires anormaux</b>	<b>Coûts de production anormaux</b>	<b>CFO anormal</b>	<b>Accruals anormaux</b>
<b>Nombre d'observations</b>	251 168	251 168	251 168	251 168
<b>Moyenne</b>	4,48	0,03	0,49	-0,32
<b>Écart-type</b>	4 544,84	46,26	433,78	418,78
<b>Min</b>	-36 142,52	-7 266,62	-16 060,39	-208 227,80
<b>1 %</b>	-0,41	-0,97	-1,01	-1,24
<b>25 %</b>	-0,02	-0,15	-0,05	-0,06
<b>Médiane</b>	0,00	-0,04	0,00	0,02
<b>75 %</b>	0,06	0,03	0,08	0,16
<b>99 %</b>	0,70	0,96	0,55	4,56
<b>Max</b>	2 271 285,00	18 672,43	215 967,58	15 848,02

Tableau 3: Statistiques descriptives des courus discrétionnaires en séparant les courus par pays

	<b>Coûts discrétionnaires anormaux</b>	<b>Coûts de production anormaux</b>	<b>CFO anormal</b>	<b>Accruals anormaux</b>
<b>Nombre d'observations<sup>23</sup></b>	251 205	251 205	251 205	251 205
<b>Moyenne</b>	5,80	0,28	0,99	-2,09
<b>Écart-type</b>	4 554,57	68,39	464,99	481,59
<b>Min</b>	-19 288,77	-2 237,39	-7 495,95	-220 371,59
<b>1 %</b>	-31,41	-0,91	-0,62	-1,38
<b>25 %</b>	-5,46	-0,24	-0,03	-0,13
<b>Médiane</b>	-0,04	-0,08	0,01	0,00
<b>75 %</b>	0,01	-0,00	0,09	0,13
<b>99 %</b>	0,52	0,86	0,46	1,13
<b>Max</b>	2 279 643,00	18 692,20	217 809,44	7 122,28

Tableau 4: Statistiques descriptives des courus discrétionnaires en agréant les courus pour tous les pays

<sup>23</sup>La différence avec le tableau 3 provient de l'exclusion de certains pays ayant un nombre insuffisant d'observations pour la régression.

Les statistiques descriptives des coûts discrétionnaires associés à la manipulation des activités réelles, selon le modèle de Roychowdhury (2006) (coûts discrétionnaires anormaux, coûts de production anormaux et CFO anormal), ainsi que des accruals anormaux basés sur le modèle de Jones modifié (Dechow, Sloan et Sweeney, 1995), suite à l'exclusion des données extrêmes, sont présentées dans les tableaux 5 et 6 selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, pays}$ ).

	Coûts discrétionnaires anormaux	Coûts de pro- duction anor- maux	CFO anormal	Accruals anor- maux
<b>Nombre</b>	209 303	209 303	209 303	209 303
<b>Moyenne</b>	0,01	0,02	0,00	-0,00
<b>Écart-type</b>	0,10	0,17	0,10	0,12
<b>Minimum</b>	-4,86	-12,70	-1,68	-1,83
<b>1%</b>	-0,17	-0,42	-0,33	-0,38
<b>25%</b>	-0,03	-0,05	-0,03	-0,04
<b>Médiane</b>	-0,00	0,00	-0,00	0,00
<b>75%</b>	0,03	0,08	0,04	0,04
<b>99%</b>	0,40	0,46	0,30	0,34
<b>Maximum</b>	0,98	1,91	1,00	1,68

Tableau 5: Statistiques descriptives des courus discrétionnaires excluant les données extrêmes en les séparant par pays

	Coûts discrétionnaires anormaux	Coûts de pro- duction anor- maux	CFO anormal	Accruals anor- maux
<b>Nombre</b>	209 342	209 342	209 342	209 342
<b>Moyenne</b>	0,02	0,04	0,00	-0,00
<b>Écart-type</b>	0,11	0,20	0,11	0,12
<b>Minimum</b>	-5,50	-18,71	-3,41	-1,98
<b>1%</b>	-0,17	-0,44	-0,34	-0,39
<b>25%</b>	-0,02	-0,03	-0,03	-0,04
<b>Médiane</b>	0,00	0,00	0,00	0,00
<b>75%</b>	0,04	0,12	0,05	0,04
<b>99%</b>	0,44	0,47	0,31	0,36
<b>Maximum</b>	1,00	1,50	1,00	2,04

Tableau 6: Statistiques descriptives des courus discrétionnaires excluant les données extrêmes en les agréant pour tous les pays

Les résultats du test de Jarque-Bera des coûts discrétionnaires associés à la manipulation des activités réelles, selon le modèle de Roychowdhury (2006) (coûts discrétionnaires anormaux, coûts de production anormaux et CFO anormaux), ainsi que des accruals anormaux basés sur le modèle de Jones modifié (Dechow, Sloan et Sweeney, 1995), sont présentés dans les tableaux 7 et 8 selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, pays}$ ).

	Moyenne	Écart-type	Asymétrie	Aplatissement	t_stat	n	jb_stat	jb_p-value	Résultat
<b>Coût discrétionnaire anormal</b>	0,01	0,10	0,33	60,26	53,69	209 303	$3,17 \times 10^7$	0,00	Non-normale
<b>Coût de production anormal</b>	0,02	0,17	-7,98	380,93	41,36	209 303	$1,27 \times 10^9$	0,00	Non-normale
<b>CFO anormal</b>	0,00	0,10	-0,57	12,81	4,61	209 303	$1,44 \times 10^6$	0,00	Non-normale
<b>Accruals anormaux</b>	-0,00	0,12	-0,29	14,27	-13,86	209 303	$1,78 \times 10^6$	0,00	Non-normale

Tableau 7: Résultat du test de Jarque-Bera sur les courus discrétionnaires en séparant les courus par pays

	Moyenne	Écart-type	Asymétrie	Aplatissement	t_stat	n	jb_stat	jb_p-value	Résultat
<b>Coût discrétionnaire anormal</b>	0,02	0,11	-0,92	124,24	77,42	209 342	$1,35 \times 10^8$	0,00	Non-normale
<b>Coût de production anormal</b>	0,04	0,20	-15,99	1038,82	84,10	209 342	$9,42 \times 10^9$	0,00	Non-normale
<b>CFO anormal</b>	0,00	0,11	-0,89	19,54	19,80	209 342	$3,36 \times 10^6$	0,00	Non-normale
<b>Accruals anormaux</b>	-0,00	0,12	-0,23	14,43	-10,03	209 342	$1,82 \times 10^6$	0,00	Non-normale

Tableau 8: Résultat du test de Jarque-Bera sur les courus discrétionnaires en agréant les courus pour tous les pays

		Nombre		Coûts discr. anorm.		Coûts prod. anorm.		CFO anormaux		Accruals anormaux	
		<i>Séparé</i>	<i>Agrégé</i>	<i>Séparé</i>	<i>Agrégé</i>	<i>Séparé</i>	<i>Agrégé</i>	<i>Séparé</i>	<i>Agrégé</i>	<i>Séparé</i>	<i>Agrégé</i>
<b>Pas de catastrophe</b>	$\mu$	26 394	26 408	0,02	0,03	0,00	0,01	-0,01	-0,01	-0,01	-0,02
	$\sigma$			0,13	0,14	0,16	0,18	0,12	0,12	0,15	0,15
<b>Inondation</b>	$\mu$	22 177	22 185	0,01	0,01	-0,00	0,01	-0,00	0,00	0,00	0,00
	$\sigma$			0,09	0,09	0,14	0,16	0,09	0,10	0,10	0,11
<b>Tempête</b>	$\mu$	28 913	28 914	0,01	0,02	0,01	0,03	0,00	0,01	-0,00	0,00
	$\sigma$			0,10	0,11	0,17	0,20	0,11	0,11	0,10	0,11
<b>Épidémie</b>	$\mu$	1 128	1 131	0,00	0,00	-0,00	-0,01	-0,00	-0,00	-0,00	-0,00
	$\sigma$			0,09	0,09	0,13	0,15	0,09	0,10	0,11	0,12
<b>Mouv. de masse (humide)</b>	$\mu$	7 119	7 123	0,01	0,01	-0,01	-0,01	-0,00	-0,00	-0,00	-0,00
	$\sigma$			0,10	0,10	0,13	0,14	0,10	0,11	0,10	0,10
<b>Température extrême</b>	$\mu$	14 647	14 651	0,01	0,01	0,01	0,02	0,00	0,00	0,00	0,01
	$\sigma$			0,10	0,11	0,16	0,17	0,10	0,11	0,11	0,11
<b>Tremblement de terre</b>	$\mu$	12 902	12 902	0,02	0,05	0,06	0,12	0,01	0,01	-0,00	-0,00
	$\sigma$			0,10	0,11	0,22	0,27	0,11	0,12	0,12	0,12
<b>Feu de forêt</b>	$\mu$	308	308	-0,00	-0,00	-0,01	-0,01	0,00	0,00	0,01	0,00
	$\sigma$			0,07	0,07	0,10	0,11	0,09	0,09	0,09	0,09
<b>Sécheresse</b>	$\mu$	197	201	0,00	-0,01	-0,00	-0,01	-0,00	0,00	0,00	0,01
	$\sigma$			0,07	0,08	0,09	0,11	0,07	0,09	0,10	0,11
<b>Activité volcanique</b>	$\mu$	754	755	-0,00	-0,00	-0,01	-0,01	-0,00	-0,00	-0,00	-0,00
	$\sigma$			0,08	0,08	0,11	0,12	0,08	0,09	0,10	0,10
<b>Mouv. de masse (sec)</b>	$\mu$	2 197	2 197	0,01	0,01	-0,01	-0,01	0,00	0,00	0,01	0,01
	$\sigma$			0,08	0,08	0,13	0,13	0,09	0,09	0,09	0,09
<b>Autres</b>	$\mu$	92 567	92 567	0,01	0,02	0,02	0,04	0,00	0,01	-0,00	-0,00
	$\sigma$			0,09	0,09	0,18	0,20	0,10	0,11	0,12	0,12

Tableau 9: Statistiques descriptives des courus discrétionnaires selon les deux configurations étudiées

Les statistiques descriptives (moyenne et écart-type) des coûts discrétionnaires associés à la manipulation des activités réelles, selon le modèle de Roychowdhury (2006) (coûts discrétionnaires anormaux, coûts de production anormaux et CFO anormaux), ainsi que des accruals anormaux basés sur le modèle de Jones modifié (Dechow, Sloan et Sweeney, 1995), sont présentées dans le tableau 10 selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, pays}$ ).

Nous testons la variation des coûts discrétionnaires anormaux (dépenses anormales en *R&D, Publicité, et SG&A*) en présence d'un groupe de catastrophe naturelle de type  $j$ . Le modèle estimé est le suivant :

$$Y_{\text{coûts discrétionnaires anormaux}} = \beta_j \mathbb{1}_j + \epsilon_i.$$

Les coefficients de la régression univariée, ainsi que les statistiques  $t$  associées, sont présentés dans ce tableau, selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, pays}$ ). Nous observons que les coûts discrétionnaires augmentent de façon statistiquement significative en présence des groupes de catastrophe naturelle.

	$Y_{\text{coûts discrétionnaires anormaux}}$	
	Séparé	Agrégé
$\mathbb{1}_{Hydrologique}$	0,006 <sup>a</sup> (8,897)	0,008 <sup>a</sup> (11,808)
$\mathbb{1}_{Météorologique}$	0,009 <sup>a</sup> (16,814)	0,014 <sup>a</sup> (24,435)
$\mathbb{1}_{Biologique}$	0,005 (1,157)	0,004 (1,031)
$\mathbb{1}_{Géophysique}$	0,019 <sup>a</sup> (19,731)	0,043 <sup>a</sup> (41,358)
$\mathbb{1}_{Climatologique}$	-0,000 (-0,020)	-0,003 (-0,534)
$\mathbb{1}_{Technologique}$	0,011 <sup>a</sup> (33,408)	0,017 <sup>a</sup> (45,730)
$\mathbb{1}_{Toutes\ les\ groupes\ de\ catastrophe}^{24}$	<b>0,010<sup>a</sup></b> <b>(27,444)</b>	<b>0,017<sup>a</sup></b> <b>(44,646)</b>
Légende:	b: $p < 5\%$ ; a: $p < 1\%$ . Les t-stats sont entre parenthèses.	

Tableau 10 a: Impact des groupes de catastrophe naturelle sur les coûts discrétionnaires anormaux selon les deux configurations étudiées

<sup>24</sup>L'agrégation de toutes les catastrophes naturelles inclut tous les types de catastrophes,

Nous testons la variation des coûts de production anormaux en présence d'un groupe de catastrophe naturelle de type j. Le modèle estimé est le suivant :

$$Y_{\text{coûts de production anormaux}} = \beta_j \mathbb{1}_j + \epsilon_i.$$

Les coefficients de la régression univariée, ainsi que les statistiques t associées, sont présentés dans ce tableau selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, \text{pays}}$ ). Nous observons que, contrairement aux coûts discrétionnaires anormaux, les coûts de production anormaux augmentent ou diminuent dépendamment du groupe de catastrophes naturelles.

	<b>Y<sub>Coûts de production anormaux</sub></b>	
	Séparé	Agrégé
$\mathbb{1}_{Hydrologique}$	-0,004 <sup>a</sup> (-4,715)	0,011 <sup>a</sup> (11,179)
$\mathbb{1}_{Météorologique}$	0,014 <sup>a</sup> (17,669)	0,030 <sup>a</sup> (33,374)
$\mathbb{1}_{Biologique}$	-0,004 (-0,767)	-0,006 (-1,128)
$\mathbb{1}_{Géophysique}$	0,044 <sup>a</sup> (30,734)	0,112 <sup>a</sup> (66,445)
$\mathbb{1}_{Climatologique}$	-0,007 (-0,930)	-0,005 (-0,690)
$\mathbb{1}_{Technologique}$	0,023 <sup>a</sup> (40,055)	0,044 <sup>a</sup> (66,735)
$\mathbb{1}_{Toutes \text{ les groupes de catastrophe}}$	<b>0,013<sup>a</sup></b> <b>(23,622)</b>	<b>0,038<sup>a</sup></b> <b>(57,908)</b>
Légende:	b: p < 5 % ; a: p < 1 % Les t-stats sont entre parenthèses.	

Tableau 10 b: Impact des groupes de catastrophe naturelle sur les coûts de production anormaux selon les deux configurations étudiées

---

à l'exception des catastrophes de type technologique, dans une seule régression.

Nous testons la variation des CFO anormaux en présence d'un groupe de catastrophe naturelle de type j. Le modèle estimé est le suivant :

$$Y_{CFO\ anormaux} = \beta_j \mathbb{1}_j + \epsilon_i.$$

Les coefficients de la régression univariée, ainsi que les statistiques t associées, sont présentés dans ce tableau selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, pays}$ ). Nous observons que les CFO anormaux augmentent de façon statistiquement significative en présence des groupes de catastrophes naturelles.

	$Y_{CFO\ anormaux}$	
	Séparé	Agrégé
$\mathbb{1}_{Hydrologique}$	-0,001 (-1,261)	0,003 <sup>a</sup> (4,191)
$\mathbb{1}_{Météorologique}$	0,002 <sup>a</sup> (4,259)	0,006 <sup>a</sup> (10,766)
$\mathbb{1}_{Biologique}$	-0,002 (-0,505)	-0,002 (-0,689)
$\mathbb{1}_{Géophysique}$	0,004 <sup>a</sup> (4,554)	0,014 <sup>a</sup> (15,284)
$\mathbb{1}_{Climatologique}$	-0,000 (-0,013)	0,004 (0,791)
$\mathbb{1}_{Technologique}$	0,003 <sup>a</sup> (8,509)	0,007 <sup>a</sup> (19,201)
$\mathbb{1}_{Toutes\ les\ groupes\ de\ catastrophe}$	<b>0,002<sup>a</sup></b> <b>(4,319)</b>	<b>0,006<sup>a</sup></b> <b>(16,963)</b>
Légende:	b: p < 5 % ; a: p < 1 % Les t-stats sont entre parenthèses.	

Tableau 10 c: Impact des groupes de catastrophe naturelle sur les CFO anormaux selon les deux configurations étudiées

Nous testons la variation des accruals anormaux en présence d'un groupe de catastrophe naturelle de type  $j$ . Le modèle estimé est le suivant :

$$Y_{Accruals\ anormaux} = \beta_j \mathbb{1}_j + \epsilon_i.$$

Les accruals anormaux représentent la partie des ajustements comptables qui ne peuvent pas être expliquée par la variation des revenus et des immobilisations corporelles. Un coefficient négatif indique que l'entreprise adopte une gestion comptable prudente ou manipule ses résultats à la baisse. Les coefficients de la régression univariée, ainsi que les statistiques  $t$  associées, sont présentés dans ce tableau selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, pays}$ ). Nous observons que, contrairement aux coûts discrétionnaires anormaux et aux CFO anormaux, les accruals anormaux augmentent ou diminuent dépendamment du groupe de catastrophes naturelles.

	$Y_{Accruals\ anormaux}$	
	Séparé	Agrégé
$\mathbb{1}_{Hydrologique}$	0,001 (1,635)	0,001 (1,315)
$\mathbb{1}_{Météorologique}$	0,001 (1,407)	0,004 <sup>a</sup> (6,033)
$\mathbb{1}_{Biologique}$	-0,004 (-0,808)	-0,002 (-0,520)
$\mathbb{1}_{Géophysique}$	-0,002 <sup>b</sup> (-2,249)	0,000 (0,249)
$\mathbb{1}_{Climatologique}$	0,006 (0,868)	0,005 (0,787)
$\mathbb{1}_{Technologique}$	-0,004 <sup>a</sup> (-9,920)	-0,002 <sup>a</sup> (-5,490)
$\mathbb{1}_{Toutes\ les\ groupes\ de\ catastrophe}$	<b>0,000</b> <b>(0,910)</b>	<b>0,002<sup>a</sup></b> <b>(5,349)</b>
Légende:	b: $p < 5\%$ ; a: $p < 1\%$ Les t-stats sont entre parenthèses.	

Tableau 10 d: Impact des groupes de catastrophe naturelle sur les accruals anormaux selon les deux configurations étudiées



Nous testons la variation des coûts discrétionnaires anormaux (dépenses anormales en *R&D, Publicité, et SG&A*) en présence d'un type de catastrophe naturelle de type  $j$ . Le modèle estimé est le suivant :

$$Y_{\text{coûts discrétionnaires anormaux}} = \beta_j \mathbb{1}_j + \epsilon_i.$$

Les coefficients de la régression univariée, ainsi que les statistiques  $t$  associées, sont présentés dans ce tableau selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_i, \text{pays}$ ). Nous observons que les coûts discrétionnaires augmentent de façon significative en présence des catastrophes naturelles.

	<b>Y<sub>coûts discrétionnaires anormaux</sub></b>	
	Séparé	Agrégé
$\mathbb{1}_{\text{Inondation}}$ 22 177	0,005 <sup>a</sup> (7,14)	0,003 <sup>a</sup> (3,904)
$\mathbb{1}_{\text{Tempête}}$ 28 913	0,010 <sup>a</sup> (14,26)	0,007 <sup>a</sup> (9,830)
$\mathbb{1}_{\text{Épidémie}}$ 1 128	0,005 (1,16)	-0,003 (-0,689)
$\mathbb{1}_{\text{Mouvement de masse (humide)}}$ 7 119	0,007 <sup>a</sup> (4,45)	-0,002 (-1,337)
$\mathbb{1}_{\text{Température extrême}}$ 14 647	0,005 <sup>a</sup> (4,91)	0,004 <sup>a</sup> (4,294)
$\mathbb{1}_{\text{Tremblement de terre}}$ 12 902	0,021 <sup>a</sup> (19,77)	0,015 <sup>a</sup> (14,072)
$\mathbb{1}_{\text{Feu de forêt}}$ 308	-0,001 (-0,12)	0,003 (0,406)
$\mathbb{1}_{\text{Sécheresse}}$ 197	0,001 (0,12)	0,002 (0,267)
$\mathbb{1}_{\text{Activité volcanique}}$ 754	-0,001 (-0,11)	-0,002 (-0,387)
$\mathbb{1}_{\text{Mouvement de masse (sec)}}$ 2 197	0,008 <sup>a</sup> (2,84)	0,002 (0,852)
$\mathbb{1}_{\text{Autres}}$ 92 567	0,011 <sup>a</sup> (33,41)	0,007 <sup>a</sup> (19,201)
$\mathbb{1}_{\text{Toutes les catastrophes}}^{25}$ 90 342	<b>0,010<sup>a</sup></b> <b>(27,444)</b>	<b>0,017<sup>a</sup></b> <b>(44,646)</b>
Légende:	b: $p < 5 \%$ ; a: $p < 1 \%$ . Les $t$ -stats sont entre parenthèses.	

Tableau 11 a: Impact de chaque type de catastrophe naturelle sur les coûts discrétionnaires anormaux selon les deux configurations étudiées

<sup>25</sup>L'agrégation de toutes les catastrophes naturelles inclut tous les types de catastrophes, à l'exception des catastrophes de type "Autres", dans une seule régression.

Nous testons la variation des coûts de production anormaux en présence d'un type de catastrophe naturelle de type j. Le modèle estimé est le suivant :

$$Y_{\text{coûts de production anormaux}} = \beta_j \mathbb{1}_j + \epsilon_i.$$

Les coefficients de la régression univariée, ainsi que les statistiques t associées, sont présentés dans ce tableau selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, \text{pays}}$ ). Nous observons que, contrairement aux coûts discrétionnaires anormaux, les coûts de production anormaux augmentent ou diminuent dépendamment du type de catastrophes naturelles.

	<b>Y<sub>Coûts de production anormaux</sub></b>	
	Séparé	Agrégé
$\mathbb{1}_{\text{Inondation}}$ 22 177	-0,003 <sup>a</sup> (-3,10)	0,013 <sup>a</sup> (11,171)
$\mathbb{1}_{\text{Tempête}}$ 28 913	0,011 <sup>a</sup> (11,31)	0,035 <sup>a</sup> (31,442)
$\mathbb{1}_{\text{Épidémie}}$ 1 128	-0,004 (-0,77)	-0,006 (-1,128)
$\mathbb{1}_{\text{Mouvement de masse (humide)}}$ 7 119	-0,007 <sup>a</sup> (-3,94)	-0,007 <sup>a</sup> (-3,625)
$\mathbb{1}_{\text{Température extrême}}$ 14 647	0,008 <sup>a</sup> (6,05)	0,016 <sup>a</sup> (10,968)
$\mathbb{1}_{\text{Tremblement de terre}}$ 12 902	0,056 <sup>a</sup> (35,12)	0,124 <sup>a</sup> (66,151)
$\mathbb{1}_{\text{Feu de forêt}}$ 308	-0,008 (-0,85)	-0,010 (-1,011)
$\mathbb{1}_{\text{Sécheresse}}$ 197	-0,005 (-0,42)	-0,010 (-0,772)
$\mathbb{1}_{\text{Activité volcanique}}$ 754	-0,007 (-1,20)	-0,007 (-1,048)
$\mathbb{1}_{\text{Mouvement de masse (sec)}}$ 2 197	-0,010 <sup>a</sup> (-2,96)	-0,010 <sup>b</sup> (-2,681)
$\mathbb{1}_{\text{Autres}}$ 92 567	0,023 <sup>a</sup> (40,06)	0,044 <sup>a</sup> (66,735)
$\mathbb{1}_{\text{Toutes les catastrophes}}$ 90 342	<b>0,013<sup>a</sup></b> <b>(23,622)</b>	<b>0,038<sup>a</sup></b> <b>(57,908)</b>
Légende:	b: p < 5 % ; a: p < 1 % . Les t-stats sont entre parenthèses.	

Tableau 11 b: Impact de chaque type de catastrophe naturelle sur les coûts de production anormaux selon les deux configurations étudiées

Nous testons la variation des CFO anormaux en présence d'un type de catastrophe naturelle de type j. Le modèle estimé est le suivant :

$$Y_{CFO \text{ anormaux}} = \beta_j \mathbb{1}_j + \epsilon_i.$$

Les coefficients de la régression univariée, ainsi que les statistiques t associées, sont présentés dans ce tableau selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, \text{pays}}$ ). Nous observons que les CFO anormaux augmentent de façon statistiquement significative en présence des catastrophes naturelles.

	$Y_{CFO \text{ anormaux}}$	
	Séparé	Agrégé
$\mathbb{1}_{Inondation}$ 22 177	-0,000 (-0,68)	0,003 <sup>a</sup> (3,904)
$\mathbb{1}_{Tempête}$ 28 913	0,002 <sup>b</sup> (2,44)	0,007 <sup>a</sup> (9,830)
$\mathbb{1}_{Épidémie}$ 1 128	-0,002 (-0,50)	-0,003 (-0,689)
$\mathbb{1}_{Mouvement \text{ de masse (humide)}}$ 7 119	-0,002 (-1,14)	-0,002 (-1,337)
$\mathbb{1}_{Température \text{ extrême}}$ 14 647	0,001 (1,45)	0,004 <sup>a</sup> (4,294)
$\mathbb{1}_{Tremblement \text{ de terre}}$ 12 902	0,005 <sup>a</sup> (4,99)	0,015 <sup>a</sup> (14,072)
$\mathbb{1}_{Feu \text{ de forêt}}$ 308	0,002 (0,27)	0,003 (0,406)
$\mathbb{1}_{Sécheresse}$ 197	-0,003 (-0,36)	0,002 (0,267)
$\mathbb{1}_{Activité \text{ volcanique}}$ 754	-0,001 (-0,23)	-0,002 (-0,387)
$\mathbb{1}_{Mouvement \text{ de masse (sec)}}$ 2 197	0,002 (0,89)	0,002 (0,852)
$\mathbb{1}_{Autres}$ 92 567	0,003 <sup>a</sup> (8,51)	0,007 <sup>a</sup> (19,201)
$\mathbb{1}_{Toutes \text{ les catastrophes}}$ 90 342	<b>0,002<sup>a</sup></b> <b>(4,319)</b>	<b>0,006<sup>a</sup></b> <b>(16,963)</b>
Légende:	b: p < 5 % ; a: p < 1 % . Les t-stats sont entre parenthèses.	

Tableau 11 c: Impact de chaque type de catastrophe naturelle sur les CFO anormaux selon les deux configurations étudiées

Nous testons la variation des accruals anormaux en présence d'un type de catastrophe naturelle de type  $j$ . Le modèle estimé est le suivant :

$$Y_{Accruals\ anormaux} = \beta_j \mathbb{1}_j + \epsilon_i.$$

Les accruals anormaux représentent la partie des ajustements comptables qui ne peuvent pas être expliquée par la variation des revenus et des immobilisations corporelles. Un coefficient négatif indique que l'entreprise adopte une gestion comptable prudente ou manipule ses résultats à la baisse. Les coefficients de la régression univariée, ainsi que les statistiques  $t$  associées, sont présentés dans ce tableau selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, \text{pays}}$ ). Nous observons que, contrairement aux coûts discrétionnaires anormaux et les CFO anormaux, les accruals anormaux varient dépendamment du type de catastrophes naturelles.

	<b>Y</b> <i>Accruals anormaux</i>	
	Séparé	Agrégé
$\mathbb{1}_{Inondation}$ 22 177	0,002 <sup>a</sup> (2,588)	-0,000 (-0,070)
$\mathbb{1}_{Tempête}$ 28 913	-0,001 (-0,861)	0,002 <sup>b</sup> (2,003)
$\mathbb{1}_{Épidémie}$ 1 128	-0,004 (-0,808)	-0,002 (-0,520)
$\mathbb{1}_{Mouvement de masse (humide)}$ 7 119	-0,002 (-1,406)	-0,003 (-1,569)
$\mathbb{1}_{Température extrême}$ 14 647	0,004 <sup>a</sup> (3,216)	0,008 <sup>a</sup> (6,762)
$\mathbb{1}_{Tremblement de terre}$ 12 902	-0,004 <sup>a</sup> (-3,380)	-0,003 <sup>b</sup> (-2,329)
$\mathbb{1}_{Feu de forêt}$ 308	0,006 (0,741)	0,004 (0,471)
$\mathbb{1}_{Sécheresse}$ 197	0,005 (0,459)	0,006 (0,528)
$\mathbb{1}_{Activité volcanique}$ 754	-0,002 (-0,287)	-0,002 (-0,363)
$\mathbb{1}_{Mouvement de masse (sec)}$ 2 197	0,007 <sup>b</sup> (2,219)	0,007 <sup>b</sup> (2,171)
$\mathbb{1}_{Autres}$ 92 567	-0,004 <sup>a</sup> (-9,920)	-0,002 <sup>a</sup> (-5,490)
$\mathbb{1}_{Toutes les catastrophes}$ 90 342	<b>0,000</b> <b>(0,910)</b>	<b>0,002<sup>a</sup></b> <b>(5,349)</b>
Légende:	b: p < 5 % ; a: p < 1 % . Les t-stats sont entre parenthèses.	

Tableau 11 d: Impact de chaque type de catastrophe naturelle sur les accruals anormaux selon les deux configurations étudiées

## Liste des figures

La Figure 1 illustre l'évolution moyenne des coûts discrétionnaires associés à la manipulation des activités réelles, selon le modèle de Roychowdhury (2006) (coûts discrétionnaires anormaux, coûts de production anormaux et CFO anormaux), ainsi que des accruals anormaux basés sur le modèle de Jones modifié (Dechow, Sloan et Sweeney, 1995) au fil des années, mettant en évidence leurs tendances respectives dans le temps.

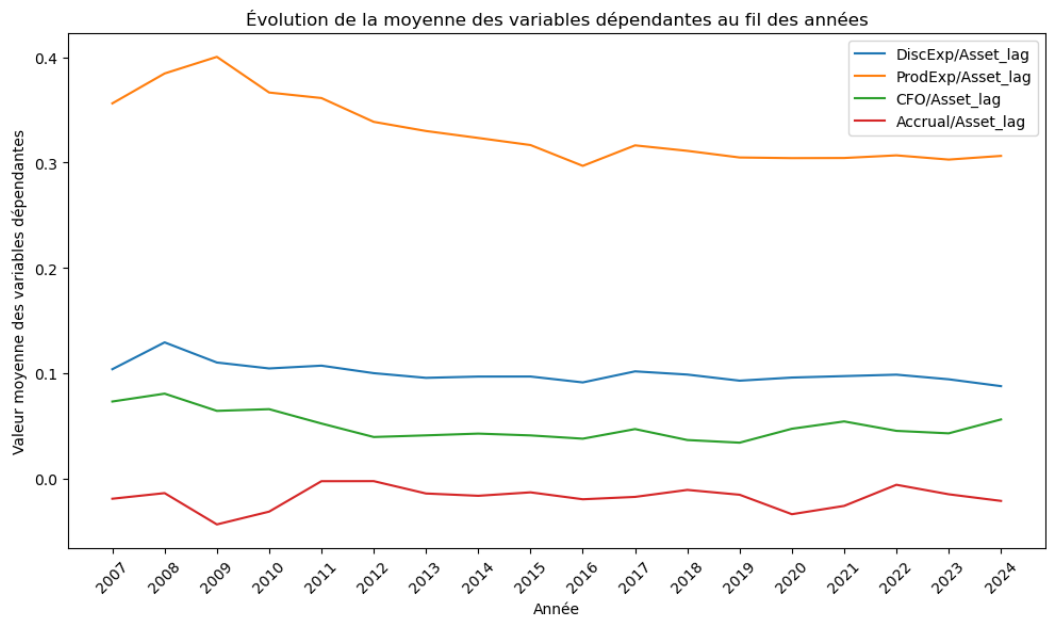


Figure 1: Évolution de la moyenne des variables dépendantes au fil des années

La Figure 2 et la Figure 3 ont été produites à partir des coûts discrétionnaires associés à la manipulation des activités réelles, selon le modèle de Roychowdhury (2006) (coûts discrétionnaires anormaux, coûts de production anormaux et CFO anormaux), ainsi que des accruals anormaux basés sur le modèle de Jones modifié (Dechow, Sloan et Sweeney, 1995), suite à l'exclusion des valeurs extrêmes, selon deux configurations : l'agrégation des données de tous les pays ( $\sum \epsilon_i$ ) et une configuration où les pays sont analysés séparément ( $\sum \epsilon_{i, pays}$ ).

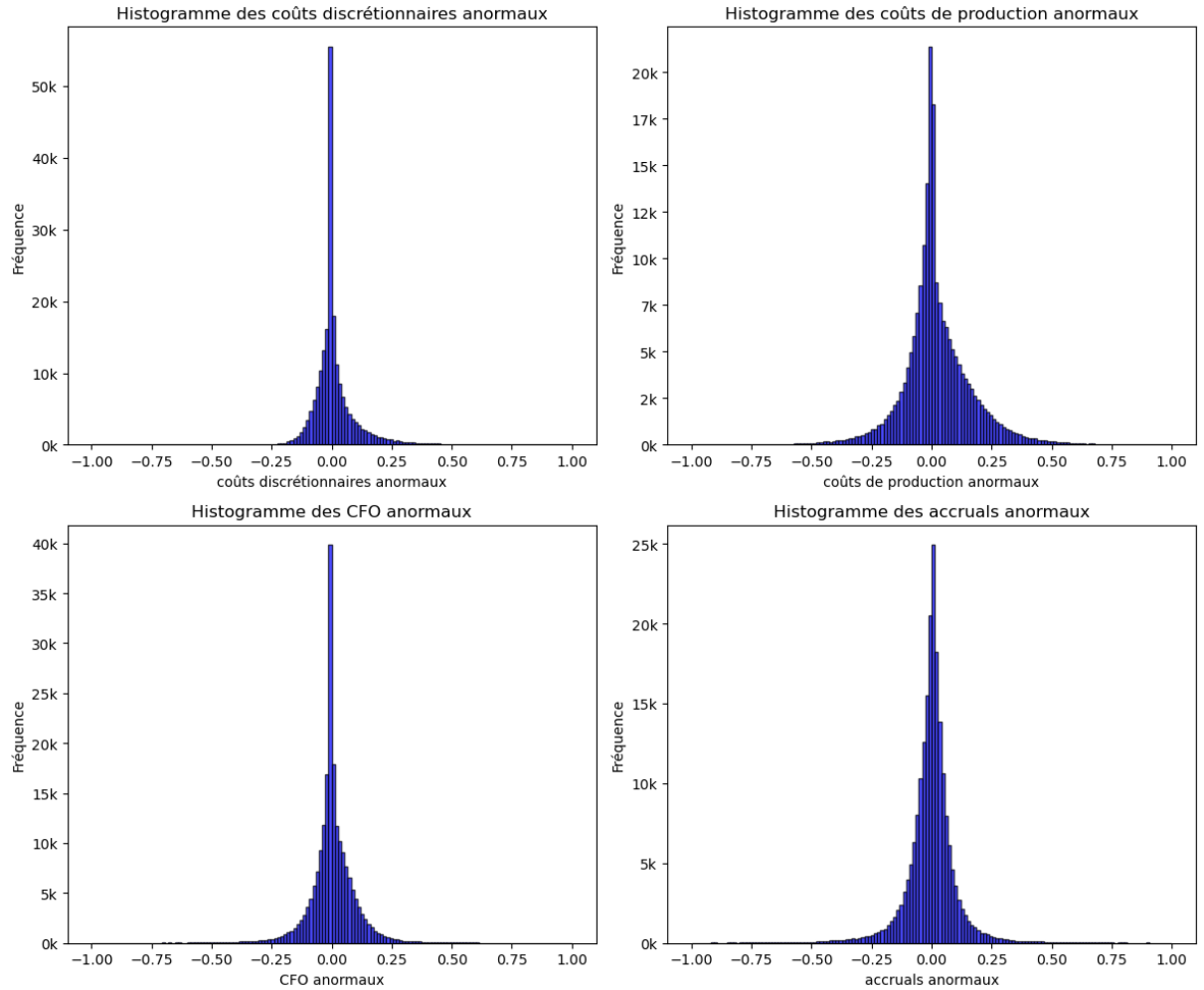


Figure 2: Répartition des courus discrétionnaires en séparant les courus par pays

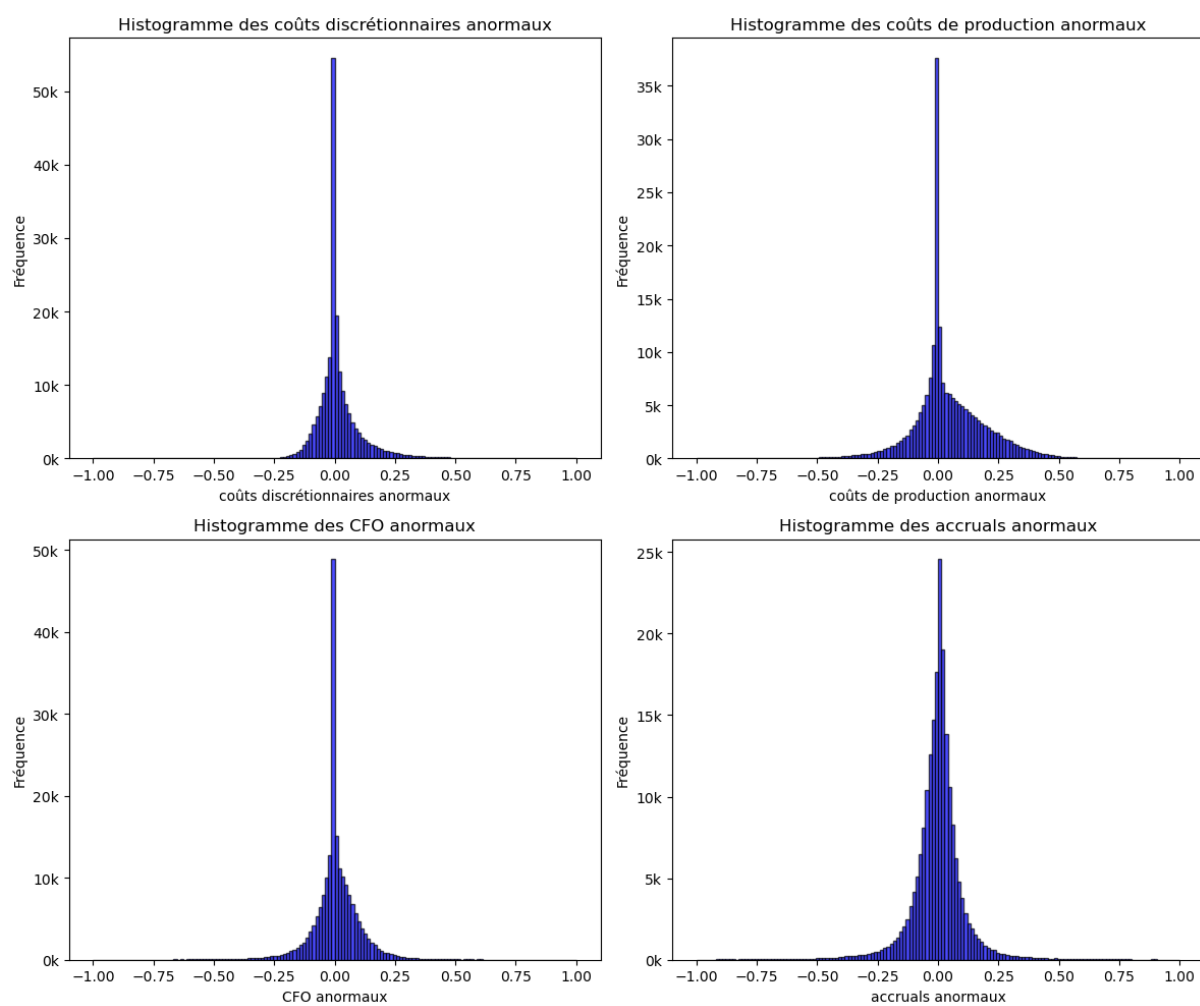


Figure 3: Répartition des courus discrétionnaires en agrégeant les courus pour tous les pays



## 9 Annexes

Liste des équations permettant de déterminer le niveau anormal des courus discrétionnaires :

$$TA_{it}/Asset_{i,t-1} = b_1 (1/Asset_{i,t-1}) + b_2 ((\Delta REV_{it} - \Delta AR_{it}) / Asset_{i,t-1}) \\ + b_3 (\Delta PPE_{it}/Asset_{i,t-1}) + \epsilon_{it}$$

avec:  $TA_{it} = EBEI_{i,t} - CFO_{it}$

$$DiscExp_{it}/Asset_{i,t-1} = b_1 (1/Asset_{i,t-1}) + b_2 (Sales_{i,t-1}/Asset_{i,t-1}) + \epsilon_{it}$$

avec:  $DiscExp = R\&D + Publicité + SG\&A$

$$Prod_{it}/Asset_{i,t-1} = b_1 (1/Asset_{i,t-1}) + b_2 (Sales_{it}/Asset_{i,t-1}) \\ + b_3 (\Delta Sales_{it}/Asset_{i,t-1}) + b_4 (\Delta Sales_{i,t-1}/Asset_{i,t-1}) + \epsilon_{it}$$

avec:  $Prod = COGS + \Delta Inv$

$$CFO_{it}/Asset_{i,t-1} = b_1 (1/Asset_{i,t-1}) + b_2 (Sales_{it}/Asset_{i,t-1}) \\ + b_3 (\Delta Sales_{it}/Asset_{i,t-1}) + \epsilon_{it}$$

Où:

$a_0, b_1, b_2, b_3, b_4$  sont les coefficients de régression

$\epsilon_{it}$  est le niveau anormal de la variable dépendante.

### Formule de EBEI et du CFO

Le EBEI est calculé comme suit :

**EBEI** = Revenus—Coût des ventes—Dépenses d'exploitation—Charges d'intérêts—Impôts

où :

- Revenus : chiffre d'affaires total de l'entreprise,
- Coût des ventes : coûts directement liés à la production des biens ou services vendus,
- Dépenses d'exploitation : frais généraux, de vente et administratifs,
- Charges d'intérêts : intérêts sur la dette de l'entreprise,
- Impôts : impôts sur le revenu hors éléments extraordinaires.

Le flux de trésorerie opérationnel est calculé comme suit :

$$\mathbf{CFO} = \mathbf{EBEI} + \mathbf{Amortissements} - \mathbf{\Delta BFR}$$

où :

- EBEI : Résultat avant éléments extraordinaires,
- Amortissements : dépréciation des actifs fixes,
- $\Delta\text{BFR}$  : Variation du besoin en fonds de roulement (BFR), définie comme suit :

$$\Delta\text{BFR} = (\Delta\text{Créances clients} + \Delta\text{Stocks}) - \Delta\text{Dettes fournisseurs}$$

## Liste des variables dépendantes et indépendantes :

Les colonnes des données financières ont été renommées pour une meilleure clarté, les actifs totaux étant désignés sous *Asset*, les ventes sous *Sales*, les dépenses en recherche et développement sous *R\&D<sub>exp</sub>*, les dépenses en ventes et administratives sous *SGA<sub>exp</sub>*, et les dépenses publicitaires sous *Ads<sub>exp</sub>*.

Ensuite, les dépenses discrétionnaires des entreprises ont été calculées en sommant les dépenses en R\&D, en SG\&A et en publicité pour chaque observation, produisant ainsi une nouvelle colonne intitulée *DiscExp*.

Les dépenses de production normales ont été calculées en additionnant le coût des biens vendus et la variation des inventaires, et stockées dans la colonne *ProdExp*.

Pour estimer le total des accruals *TA<sub>it</sub>*, la différence entre le bénéfice avant éléments extraordinaires et le flux de trésorerie d'exploitation a été calculée.

Les colonnes représentant les actifs et les ventes ont été décalées d'une période pour créer des colonnes *Asset<sub>lag</sub>* et *Sales<sub>lag</sub>*, permettant ainsi l'analyse des

variations par rapport aux périodes précédentes. La variation des ventes  $\Delta Sales$  a été calculée en soustrayant les ventes de la période précédente des ventes actuelles, et une version retardée de cette variation a également été créée  $\Delta Sales_{lag}$ .

## **Variables dépendantes ou indicateurs de mesure de la manipulation des résultats :**

**Variable dépendante 1 : Variables de régression des coûts discrétionnaires anormaux :**

$(DiscExp/Asset_{lag})$  : le ratio des dépenses discrétionnaires par rapport aux actifs retardés.

**Variable dépendante 2 : Variables de régression du flux de trésorerie opérationnel anormal (CFO) :**

$(CFO/Asset_{lag})$  : le ratio du flux de trésorerie opérationnel par rapport aux actifs retardés.

**Variable dépendante 3 : Variables de régression des coûts de production anormaux :**

$(ProdExp/Asset_{lag})$  : le ratio des dépenses de production par rapport aux actifs retardés.

**Variable dépendante 4 : Variables de régression des accruals anormaux:**

$(Accrual/Asset_{lag})$  : le ratio des accruals par rapport aux actifs retardés.

## **Variables indépendantes :**

**Ensemble 1 : Variables indépendantes de régression des coûts**

**discrétionnaires anormaux :**

$(Var - 1/Asset_{lag})$  : l'inverse des actifs retardés

$(Var - Sales_{lag}/Asset_{lag})$  : le ratio des ventes retardées par rapport aux actifs retardés.

**Ensemble 2 : Variables indépendantes de régression des coûts de production anormaux :**

$(Var - 1/Asset_{lag})$  : l'inverse des actifs retardés ( $1/Asset_{lag}$ )

$(Var - Sales/Asset_{lag})$  : le ratio des ventes actuelles par rapport aux actifs retardés

$(Var - \Delta Sales/Asset_{lag})$  : le ratio de la variation des ventes par rapport aux actifs retardés

$(Var - \Delta Sales_{lag}/Asset_{lag})$  : le ratio de la variation retardée des ventes par rapport aux actifs retardés.

**Ensemble 3 : Variables indépendantes de régression du flux de trésorerie opérationnel anormal (CFO) :**

$(Var - 1/Asset_{lag})$  : l'inverse des actifs retardés

$(Var - Sales/Asset_{lag})$  : le ratio des ventes actuelles par rapport aux actifs retardés

$(Var - \Delta Sales/Asset_{lag})$  : le ratio de la variation des ventes par rapport aux actifs retardés.

**Ensemble 4 : Variables indépendantes de régression des accruals anormaux :**

$(Var - 1/AL)$  : l'inverse des actifs retardés ( $1/Asset_{lag}$ )

$(\Delta Sales - \Delta AR)/Asset_{lag}$  : une variable indiquant la différence entre la variation des ventes et la variation des créances par rapport aux actifs retardés

$(PP\&E/Asset_{lag})$  : le ratio des actifs corporels par rapport aux actifs retardés

## Nettoyage des données

```
In [1]: import numpy as np
import pandas as pd
import statsmodels.api as sm
from linearmodels.panel import PanelOLS
from scipy import stats
import warnings
# Suppress SettingWithCopyWarning
warnings.filterwarnings("ignore")

In [ ]: '''
Il lit un fichier CSV et stocke les données dans un DataFrame.
Ensuite, il sélectionne les colonnes pertinentes.
Il supprime les lignes avec des valeurs manquantes dans les colonnes 'Start Year' et 'Start Month'.
Les colonnes 'Start Year' et 'Start Month' sont converties en entiers.

Pour chaque ligne, il détermine le dernier jour du mois à partir de l'année et du mois donnés.
Les valeurs manquantes dans la colonne 'Start Day' sont remplies avec le dernier jour du mois correspondant.
Les colonnes individuelles pour l'année, le mois et le jour de début sont supprimées.
Enfin, une nouvelle colonne 'Start Date' est créée en fusionnant les colonnes 'Start Year', 'Start Month' et 'Start Day'.'''
# Read disaster file into a DataFrame
disaster_data=pd.read_csv(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\disaster_data.csv",on_bad_lines='warn')
# Select columns of interest
disaster_data_selected = disaster_data[['Disaster Group', 'Disaster Subgroup', 'Disaster Type',
                                         'ISO', 'Country', 'Subregion', 'Region', 'Location',
                                         'Start Year', 'Start Month', 'Start Day']]

# Drop rows with NaN values in 'Start Year' and 'Start Month' columns
disaster_data_selected.dropna(subset=['Start Year', 'Start Month'], inplace=True)

# Convert 'Start Year' and 'Start Month' to integers
disaster_data_selected['Start Year'] = disaster_data_selected['Start Year'].astype(int)
disaster_data_selected['Start Month'] = disaster_data_selected['Start Month'].astype(int)
# Determine the last day of each month
last_day_of_month = disaster_data_selected.apply(lambda x: pd.Timestamp(year=x['Start Year'], month=x['Start Month'], day=1) + pd.offsets.MonthEnd(0), axis=1)

# Fill NaN values in 'Start Day' column with the corresponding last day of the month
disaster_data_selected['Start Day'].fillna(last_day_of_month.dt.day, inplace=True)
disaster_data_selected['Start Day'] = disaster_data_selected['Start Day'].astype(int)

# Create a new column 'Start Date' by merging 'Start Year', 'Start Month', and 'Start Day'
disaster_data_selected['Date'] = disaster_data_selected.apply(lambda x: pd.Timestamp(year=x['Start Year'], month=x['Start Month'], day=x['Start Day']), axis=1)

# Drop the individual date columns if needed
disaster_data_selected.drop(['Start Year', 'Start Month', 'Start Day'], axis=1, inplace=True)

In [3]: '''Le code lit les fichiers CSV contenant les données financières des entreprises nord-américaines et mondiales dans des DataFrames distincts.
Il définit une liste des colonnes financières pertinentes.
Il supprime les lignes où les colonnes 'fdate' ou 'curcd' contiennent des valeurs NaN.
Il remplit les valeurs manquantes avec 0 pour les frais de pub.
Il combine les données des entreprises nord-américaines et mondiales dans un seul DataFrame 'financial_data' en utilisant la fonction 'concat' de Pandas.'''
# Read NA F/S and Global F/S data file into a DataFrame
# Read the CSV file for north american companies
NA_data = pd.read_csv(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\NA_data.csv",on_bad_lines='warn')
# Select columns needed
NA_data=NA_data[['gvkey', 'curcd', 'fdate', 'at', 'cogs', 'ib', 'invch',
                 'oancf', 'ppeg', 'recch', 'revt', 'sale', 'xad', 'xrd', 'xsga',
                 'fic', 'gsector']]
# Financial columns
financial_columns = ['at','cogs', 'ib', 'invch', 'oancf', 'ppeg', 'recch', 'revt', 'sale', 'xad', 'xrd', 'xsga']
# Drop rows where 'fyear' or 'curcd' contains NaN
NA_data = NA_data.dropna(subset=['fdate', 'curcd'])
# Fill NaN values with 0 for specified columns
NA_data[financial_columns] = NA_data[financial_columns].fillna(0)

# Read the CSV file for global companies
global_data = pd.read_csv(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\global_data.csv", on_bad_lines='warn')
# Rename 'xcom' column to 'xad'
global_data = global_data.rename(columns={'xcom': 'xad'})
# Fill the 'xad' column with 0 for all values because non existant on computstat
global_data['xad'] = 0
# Select columns needed
global_data=global_data[['gvkey', 'curcd', 'fdate', 'at', 'cogs', 'ib', 'invch',
                        'oancf', 'ppeg', 'recch', 'revt', 'sale', 'xad', 'xrd', 'xsga',
                        'fic', 'gsector']]
# Drop rows where 'fdate' or 'curcd' contains NaN
global_data = global_data.dropna(subset=['fdate', 'curcd'])
# Fill NaN values with 0 for specified columns
global_data[financial_columns] = global_data[financial_columns].fillna(0)

# Combine north america and global data
financial_data = pd.concat([NA_data, global_data], ignore_index=True)

In [ ]: '''Cette section vise à importer les données de GDP per capita(current currency, international currency) ayant comme source la base de données de la banque mondiale
pour les années 2023 et 2024, le GDP de 2022 est considéré'''
gdp_data = pd.read_csv(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\gdp_capita.csv", skiprows=4,on_bad_lines='warn')

gdp_data.drop(columns=['Unnamed: 68'], inplace=True)
# Convert 'fdate' column to datetime format
financial_data['fdate'] = pd.to_datetime(financial_data['fdate'])
# Extract year from 'fdate' and create a new column 'year_of_fdate'
financial_data['year'] = financial_data['fdate'].dt.year
# Melt the DataFrame to have ISO code and year in separate rows
gdp_data = gdp_data.melt(id_vars=['Country Name', 'Country Code', 'Indicator Name', 'Indicator Code'], var_name='year', value_name='GDP_per_capita')
# Drop unnecessary columns
gdp_data.drop(columns=['Indicator Name', 'Indicator Code'], inplace=True)
gdp_data['year'] = pd.to_datetime(gdp_data['year'])
# Filter the financial data to keep rows with dates after the earliest year
gdp_data = gdp_data[gdp_data['year'].dt.year >= financial_data['fdate'].dt.year.min()]
gdp_data['year'] = gdp_data['year'].dt.year

# Iterate over unique ISO codes in gdp_data
for iso in gdp_data['Country Code'].unique():
    # Get the GDP data for the current ISO code
    iso_data = gdp_data[gdp_data['Country Code'] == iso]

    # Replace NaN values with previous year's value
    iso_data['GDP_per_capita'] = iso_data['GDP_per_capita'].fillna(method='ffill')
```

```
# Update the GDP data with the modified values
gdp_data.loc[iso_data.index, 'GDP_per_capita'] = iso_data['GDP_per_capita']

# Replace all NaN values with 0 in the 'GDP_per_capita' column
gdp_data['GDP_per_capita'].fillna(0, inplace=True)
# Update the year to 2024
gdp_data_2024 = gdp_data[gdp_data['year'] == 2023].copy()
gdp_data_2024['year'] = 2024
# Concatenate the original DataFrame with the new DataFrame for the year 2024
gdp_data = pd.concat([gdp_data, gdp_data_2024], ignore_index=True)
```

```
In [ ]: '''un fichier excel contenant les currency code de differents pays sont lu afin de translate le GDP per capita en $US de 2010'''
# Read the CSV file for north american companies
code= pd.read_excel(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\code.xlsx")
# Select columns needed
# Assuming your DataFrame is named 'code'
code['country'] = code['country'].apply(lambda x: x.title() if len(x.split()) <= 2 else x)
# Merge financial_data with gdp_data on 'curcd' and 'datadate'
financial_data = pd.merge(financial_data, gdp_data, left_on=['fic', 'year'], right_on=['Country Code', 'year'], how='left')
# Drop unnecessary columns from the merged data
financial_data.drop(columns=['Country Code', 'year'], inplace=True)
```

```
In [ ]: # Drop the 'country' column from the merged DataFrame if needed
financial_data.drop(columns=['Country Name'], inplace=True)
```

```
In [ ]: '''Le code prend des données financières mondiales et les traduit en USD de 2010 en utilisant les taux de change moyens pour chaque devise.
Il met à jour les codes de devise, supprime 'ZWL', 'ZMW', 'SDG', 'ZWE', filtre et nettoie les données, puis les associe aux données de catastrophe correspondantes.'''
# Translation of global financial data into 2010 USD
# Read the CSV file for currencies from Jan 1st 2010 to dec 31 2010
currency_data = pd.read_csv(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\currency_data.csv", on_bad_lines='warn')
# Calculate the average exchange rate for each currency
average_exc_rate = currency_data.groupby('curcd')['exratd_toUSD'].mean()
average_exc_rate_dict = average_exc_rate.to_dict()

# dictionary for updating currency codes after checking for unmatched
currency_updates = {'ROL': 'RON', 'MPX': 'MXN', 'BRR': 'BRL', 'GHC': 'GHS', 'SDG': 'SDG', 'XEU': 'EUR', 'BRE': 'BRL', 'RUR': 'RUB', 'ZWL': 'ZWL', 'VES': 'VEF', 'TRL': 'TRY', 'ARA': 'AR'}

# Update currency codes in global_data
financial_data['curcd'] = financial_data['curcd'].map(currency_updates).fillna(financial_data['curcd'])
# Filter out rows where the currency code is not 'ZWL', 'ZMW', 'SDG', or 'ZWE'
financial_data = financial_data[financial_data['curcd'] != 'ZWL']
financial_data = financial_data[financial_data['curcd'] != 'ZMW']
financial_data = financial_data[financial_data['curcd'] != 'SDG']
financial_data = financial_data[financial_data['curcd'] != 'ZWE']
# Financial columns
financial_columns = ['at','cogs', 'ib', 'invch', 'oancf', 'ppeg', 'recch', 'revt', 'sale', 'xad', 'xrd', 'xsga', 'GDP_per_capita']
# Map the average exchange rates to global_data based on the currency code
financial_data['average_exchange_rate'] = financial_data['curcd'].map(average_exc_rate_dict)
for col in financial_columns:
    financial_data[col] = financial_data[col] * financial_data['average_exchange_rate']
```

```
In [10]: # Drop both 'average_exchange_rate' and 'curcd' columns since translation is completed with 2010 USD
financial_data = financial_data.drop(columns=['average_exchange_rate', 'curcd'])
# Sort the dataframe using 'fyear' and 'gvkey'
financial_data = financial_data.sort_values(by=['fdate', 'gvkey'])
# Fill NaN values with 0 for specified financial columns
financial_data[financial_columns] = financial_data[financial_columns].fillna(0)
# Rename columns for merge with disaster data
financial_data = financial_data.rename(columns={'fdate': 'Date', 'fic': 'ISO'})
# Filter the disaster data based on conditions
disaster_data_selected = disaster_data_selected[disaster_data_selected['Date'] >= financial_data['Date'].min()]
disaster_data = disaster_data_selected.drop(['Disaster Subgroup', 'Country', 'Subregion', 'Region', 'Location'], axis=1)
```

```
In [ ]: '''Calcul des données financières pour la regression'''
# Calcul des variables afin de faciliter les regressions après le merge avec le disaster dataframe
reg_data=financial_data.copy()
# Lag 'Asset' and 'Sales' columns and store them in new columns
# Sort the dataframe by gvkey and Date
reg_data = reg_data.sort_values(by=['gvkey', 'Date'])
# Calculate lagged columns for assets and sales by grouping them into same company
reg_data['Asset_lag'] = reg_data.groupby('gvkey')['at'].shift(1)
reg_data['Sales_lag'] = reg_data.groupby('gvkey')['sale'].shift(1)
# Rename columns for clarity
reg_data = reg_data.rename(columns={'at': 'Asset', 'revt': 'Sales', 'xrd': 'R&D_exp', 'xsga': 'SGA_exp', 'xad': 'Ads_exp'})
# Calculating discretionary expenses
reg_data['DiscExp'] = reg_data[['R&D_exp', 'SGA_exp', 'Ads_exp']].sum(axis=1, skipna=True)
# Create a new column for the change in sales
reg_data['Delta_Sales'] = reg_data['Sales'] - reg_data['Sales_lag']
reg_data['Delta_Sales_lag'] = reg_data['Delta_Sales'].shift(1)
# calculating normal production cost
reg_data['ProdExp'] = reg_data[['cogs', 'invch']].sum(axis=1, skipna=True)
# Estimating the accrual earning
reg_data['accrual'] = reg_data['ib'] - reg_data['oancf']
# 1. Create variable of regression of abnormal discretionay cost
reg_data['Var_1/AL'] = 1/reg_data['Asset_lag']
reg_data['Var_SL/AL'] = reg_data['Sales_lag']/reg_data['Asset_lag']
reg_data['DiscExp/Asset_lag'] = reg_data['DiscExp']/reg_data['Asset_lag']
# 2. Create variable of regression of abnormal CFO
reg_data['Var_S/AL'] = reg_data['Sales']/reg_data['Asset_lag']
reg_data['Var_DS/AL'] = reg_data['Delta_Sales']/reg_data['Asset_lag']
reg_data['CFO/Asset_lag'] = reg_data['oancf']/reg_data['Asset_lag']
reg_data = reg_data.drop(['sale', 'Ads_exp', 'R&D_exp', 'SGA_exp'], axis=1)
# 3. Create variable of regression of abnormal PRODUCTION cost
reg_data['Var_DSL/AL'] = reg_data['Delta_Sales_lag']/reg_data['Asset_lag']
reg_data['ProdExp/Asset_lag'] = reg_data['ProdExp']/reg_data['Asset_lag']
# Estimating the accrual earning
reg_data['accrual'] = reg_data['ib'] - reg_data['oancf']
reg_data['Var_2'] = (reg_data['Delta_Sales'] - reg_data['recch'])/reg_data['Asset_lag']
reg_data['Var_3'] = reg_data['ppeg']/reg_data['Asset_lag']
reg_data['Accrual/Asset_lag'] = reg_data['accrual']/reg_data['Asset_lag']
# Drop columns
reg_data=reg_data.drop(['Asset', 'cogs', 'ib', 'invch', 'oancf', 'ppeg', 'recch', 'Sales', 'DiscExp', 'Asset_lag', 'Sales_lag', 'Delta_Sales', 'Delta_Sales_lag', 'ProdExp', 'accrual', ], axis=1)
```

```
In [12]: '''
Le code remplace les valeurs infinies par NaN, supprime les lignes contenant des valeurs NaN dans le DataFrame reg_data,
et extrait l'année et le mois à partir de la colonne 'Date' dans les DataFrames reg_data et disaster_data.
Enfin, il ajuste l'année de fin pour les données de catastrophe pour correspondre à un an après la date de début.'''
# Replace 'inf' with np.nan
reg_data.replace([np.inf, -np.inf], np.nan, inplace=True)
```

```

In [ ]: # Drop rows with NaN values
reg_data=reg_data.dropna()
# Extract year and month from both datasets
reg_data['Year'] = pd.to_datetime(reg_data['Date']).dt.year
reg_data['End Year'] = reg_data['Year']
reg_data['Month'] = pd.to_datetime(reg_data['Date']).dt.month
disaster_data['Year'] = pd.to_datetime(disaster_data['Date']).dt.year
disaster_data['End Year'] = disaster_data['Year']+1
disaster_data['Month'] = pd.to_datetime(disaster_data['Date']).dt.month
print(financial_data.shape)
financial_data['ISO'].nunique()

In [ ]: '''
Le code fusionne les données en fonction des années et des mois correspondants, en utilisant la clé 'ISO' et 'Year'.
Il élimine les doublons basés sur certaines colonnes spécifiques. Ensuite, il supprime les colonnes 'Year' et 'Month' redondantes.
Enfin, il renomme certaines colonnes pour une meilleure clarté et libère de la mémoire en supprimant le DataFrame reg_data'''
# Merge data based on matching years and months
data = pd.merge(reg_data, disaster_data, how='left', on=['ISO', 'Year']).drop_duplicates(subset=['Var_1/AL', 'Var_SL/AL',
'DiscExp/Asset_lag', 'Var_S/AL', 'Var_DS/AL', 'CFO/Asset_lag',
'Var_DSL/AL', 'ProdExp/Asset_lag', 'Var_2', 'Var_3',
'Accrual/Asset_lag'])
# Drop the 'Year' and 'Month' columns
data.drop(['Year', 'Month_x', 'Month_y'], axis=1, inplace=True)

# Rename columns
data.rename(columns={'Date_x': 'FY Date', 'Date_y': 'Disaster date', 'End Year_x': 'End Year'}, inplace=True)

In [ ]: '''Fusion des DataFrames sur les colonnes 'ISO' et 'End Year', élimination des doublons.
Remplacement des valeurs NaN dans certaines colonnes avec celles des colonnes correspondantes après fusion.
Suppression des colonnes inutiles.
Renommage des colonnes pour plus de clarté.'''
# Step 1: Merge dataframes based on 'ISO' and 'End Year'
data = pd.merge(data, disaster_data, on=['ISO', 'End Year'], how='left', suffixes=('_data', '_disaster')).drop_duplicates(subset=['Var_1/AL', 'Var_SL/
'DiscExp/Asset_lag', 'Var_S/AL', 'Var_DS/AL', 'CFO/Asset_lag',
'Var_DSL/AL', 'ProdExp/Asset_lag', 'Var_2', 'Var_3',
'Accrual/Asset_lag'])

# Step 2: Replace NaN values in 'Disaster Group', 'Disaster Type', and 'Disaster date'
data['Disaster Group_data'].fillna(data['Disaster Group_disaster'], inplace=True)
data['Disaster Type_data'].fillna(data['Disaster Type_disaster'], inplace=True)
data['Disaster date'].fillna(data['Date'], inplace=True)

# Step 3: Drop unnecessary columns
data.drop(['Disaster Group_disaster', 'Disaster Type_disaster', 'Date', 'End Year', 'End Year_y', 'Month', 'Year'], axis=1, inplace=True)

# Rename columns
data.rename(columns={'Disaster Group_data': 'Disaster Group', 'Disaster Type_data': 'Disaster Type'}, inplace=True)

In [ ]: '''Création d'une nouvelle colonne 'Disaster' dans le DataFrame data. Si la valeur de la colonne 'Disaster Group' n'est pas NaN,
la valeur de la colonne 'Disaster' est définie sur "Disaster", sinon elle est définie sur "No Disaster".
Remplacement des valeurs NaN dans la colonne 'Disaster Type' par "No Disaster".
Conversion des colonnes de date en objets datetime pour la colonne 'FY Date'.'''
# Create a new column 'Disaster' in data. Set to "Disaster" if 'Disaster Group' is not NaN, otherwise set to "No Disaster"
data['Disaster'] = data['Disaster Group'].apply(lambda x: 'Disaster' if pd.notna(x) else 'No Disaster')

# Replace NaN values in 'Disaster Type' column with "No Disaster"
data['Disaster Type'].fillna('No Disaster', inplace=True)
# Convert date columns to datetime objects
data['FY Date'] = pd.to_datetime(data['FY Date'])

In [ ]: data = data.sort_values(by='ISO', ascending=True)

In [ ]: # Exporter Les données nettoyées dans un fichier CSV
data.to_csv(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\merged_data_with_GDPperCapita_new_new.csv", index=False)

In [ ]: # S'assurer que les colonnes sont bien au format string
data['ISO'] = data['ISO'].astype(str)
data['Disaster Type'] = data['Disaster Type'].astype(str)

# Regrouper par pays et type de catastrophe, puis compter les observations
disaster_counts_by_country = data.groupby(['ISO', 'Disaster Type']).size().unstack(fill_value=0)

# Afficher le tableau
disaster_counts_by_country

```



## Analyse selon la configuration 1

```
In [1]: import numpy as np
import pandas as pd
import statsmodels.api as sm
from linearmodels.panel import PanelOLS
from scipy import stats
from scipy.stats import t
from scipy.stats import ttest_ind
import warnings
import matplotlib.pyplot as plt
# Suppress SettingWithCopyWarning
warnings.filterwarnings("ignore")
from scipy.stats import norm
import matplotlib.pyplot as plt
from matplotlib.ticker import FuncFormatter
from scipy.stats import ttest_1samp, jarque_bera, skew, kurtosis

In [2]: # Read disaster file into a DataFrame
data=pd.read_csv(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\merged_data_with_GDPperCapita_new.csv",on_bad_lines='warn')
# Filter the DataFrame to contain only rows where GDP_per_capita is Lower than 15000
data = data[data['GDP_per_capita'] < 15000]

In [3]: # Dropping rows where 'Disaster Type' is 'Infestation'
data = data[data['Disaster Type'] != 'Infestation']

In [ ]: #Suppression of outliers
data = data[data['DiscExp/Asset_lag'] <= 1]
data = data[data['DiscExp/Asset_lag'] >= -1]
data = data[data['ProdExp/Asset_lag'] <= 1]
data = data[data['ProdExp/Asset_lag'] >= -1]
data = data[data['CFO/Asset_lag'] <= 1]
data = data[data['CFO/Asset_lag'] >= -1]
data = data[data['Accrual/Asset_lag'] <= 1]
data = data[data['Accrual/Asset_lag'] >= -1]

In [6]: # Define the List of disaster types that should not be changed
disaster_keep = ['Flood', 'Storm', 'Epidemic', 'Extreme temperature',
                 'Mass movement (dry)', 'Mass movement (wet)', 'Earthquake',
                 'Wildfire', 'Drought', 'Volcanic activity', 'No Disaster']

# Replace disaster types not in the List with 'Others'
data['Disaster Type'] = data['Disaster Type'].apply(lambda x: x if x in disaster_keep else 'Others')
# Donne Le nom des éléments contenu dans la colonne Disaster pour séparer Les statistiques par La suite
disaster_counts = data['Disaster Type'].value_counts()

In [ ]: #Ajustement taille et désastre
iso = data['ISO'].unique()
for iso in iso:
    num_observations = data[data['ISO'] == iso].shape[0]
    if num_observations < 30:
        data = data[data['ISO'] != iso]
disaster_types = data['Disaster Type'].unique()
for disaster_type in disaster_types:
    num_observations = data[data['Disaster Type'] == disaster_type].shape[0]
    if num_observations < 200:
        data.loc[data['Disaster Type'] == disaster_type, 'Disaster Type'] = 'Others'
# Sort the DataFrame by the 'Disaster' and 'ISO' columns in descending order
data = data.sort_values(by=['Disaster', 'ISO'], ascending=False)
# Nombre d'observation
print(data.shape)
disaster_counts = data['Disaster Type'].value_counts()

In [8]: # Function to classify disasters
def classify_disaster(cata):
    if cata in ['Earthquake', 'Mass movement (dry)', 'Volcanic activity']:
        return 'Géophysique'
    elif cata in ['Flood', 'Mass movement (wet)']:
        return 'Hydrologique'
    elif cata in ['Storm', 'Extreme temperature']:
        return 'Météorologique'
    elif cata in ['Drought', 'Wildfire']:
        return 'Climatologique'
    elif cata == 'Epidemic':
        return 'Biologique'
    elif cata == 'No Disaster':
        return 'Sans Catastrophe'
    else:
        return 'Technologique'

# Apply the classification function to the DataFrame
data['Groupe'] = data['Disaster Type'].apply(classify_disaster)

# Donne Le nom des éléments contenu dans la colonne Disaster type pour séparer Les statistiques par La suite
# 'Disaster Type' à changer à 'Groupe' pour L'analyse selon Les groupes de catastrophe naturel
disaster_types = data['Disaster Type'].unique()

In [9]: def calculate_regression_results(data, disaster_type, type):
    '''La fonction calculate_regression_results prend des données, un liste d'element d'une colonne précis et la colonne spécifique,
    effectue des régressions sur ces données pour chaque unité statistique,
```

```

effectue des régressions sur ces données pour chaque unité statistique,
et retourne les résidus des régressions en ne considérant pas d'effets fixes et en excluant les unités statistiques avec des exceptions.'''
# Subset the data for the specified disaster type
reg_data = data[data[type] == disaster_type]

# Initialize DataFrame with all unique ISOs
reg_results = pd.DataFrame({'ISO': reg_data['ISO'].unique(),
                           'Abn_disc_exp': [[] for _ in range(len(reg_data['ISO'].unique()))],
                           'Abn_cost_prod': [[] for _ in range(len(reg_data['ISO'].unique()))],
                           'Abn_cash_flow': [[] for _ in range(len(reg_data['ISO'].unique()))],
                           'Abn_accrual': [[] for _ in range(len(reg_data['ISO'].unique()))])

# List to store ISOs with exceptions
iso_exceptions = []

for idx, iso in enumerate(reg_results['ISO']):
    try:
        iso_data = reg_data[reg_data['ISO'] == iso]
        iso_data['FY Date'] = pd.to_datetime(iso_data['FY Date'])
        iso_data.set_index(['gvkey', 'FY Date'], inplace=True)

        # Regression for abnormal discretionary expense
        X1 = iso_data[['Var_1/AL', 'Var_SL/AL']]
        y1 = iso_data['DiscExp/Asset_lag']
        model1 = PanelOLS(y1, X1, entity_effects=False, drop_absorbed=True, check_rank=False).fit(cov_type='clustered')
        reg_results.at[idx, 'Abn_disc_exp'] = model1.resids.tolist()

        # Regression for abnormal production expense
        X2 = iso_data[['Var_1/AL', 'Var_SL/AL', 'Var_DS/AL', 'Var_DSL/AL']]
        y2 = iso_data['ProdExp/Asset_lag']
        model2 = PanelOLS(y2, X2, entity_effects=False, drop_absorbed=True, check_rank=False).fit(cov_type='clustered')
        reg_results.at[idx, 'Abn_cost_prod'] = model2.resids.tolist()

        # Regression for abnormal cash flow
        X3 = iso_data[['Var_1/AL', 'Var_SL/AL', 'Var_DS/AL']]
        y3 = iso_data['CFO/Asset_lag']
        model3 = PanelOLS(y3, X3, entity_effects=False, drop_absorbed=True, check_rank=False).fit(cov_type='clustered')
        reg_results.at[idx, 'Abn_cash_flow'] = model3.resids.tolist()

        # Regression for abnormal accrual
        X4 = iso_data[['Var_1/AL', 'Var_2', 'Var_3']]
        y4 = iso_data['Accrual/Asset_lag']
        model4 = PanelOLS(y4, X4, entity_effects=False, drop_absorbed=True, check_rank=False).fit(cov_type='clustered')
        reg_results.at[idx, 'Abn_accrual'] = model4.resids.tolist()

    except (ZeroDivisionError, ValueError, TypeError, Exception):
        iso_exceptions.append(iso)

# Remove ISOs with exceptions
reg_results = reg_results[~reg_results['ISO'].isin(iso_exceptions)]
#print(iso_exceptions)
return reg_results

```

In [255.. '''Ce code crée un dictionnaire vide reg\_results\_dict et itère à travers les différents types de désastres. Pour chaque type de désastre, il appelle la fonction calculate\_regression\_results pour calculer les résultats de la régression spécifiques à ce type de désastre, puis stocke ces résultats dans le dictionnaire sous une clé basée sur le type de désastre.'''

```

reg_results_dict = {}
# Iterate over unique disaster types
for disaster_type in disaster_types:
    # Call the function to calculate regression results for the current disaster type
    # 'Disaster Type' à changer à 'Groupe' pour l'analyse selon les groupes de catastrophe naturel
    reg_results = calculate_regression_results(data, disaster_type, 'Disaster Type')
    # Store the results in the dictionary with a key based on the disaster type
    reg_results_dict[f"{disaster_type}"] = reg_results

```

In [256.. '''Ce code itère à travers un dictionnaire reg\_results\_dict contenant des DataFrames de résultats de régression pour s'il y'a désastre ou non. Pour chaque DataFrame, il extrait et aplatit les valeurs de chaque colonne, puis stocke ces valeurs aplaties dans un dictionnaire residuals\_by\_disaster organisé par type de désastre et nom de colonne d'accruals.'''

```

# Dictionary to store flattened values for each column and each key
residuals_by_disaster = {}

# Iterate over each key (disaster type) and DataFrame in reg_results_dict
for key, df in reg_results_dict.items():
    # Dictionary to store flattened values for the current key
    flattened_values_by_column = {}

    # Iterate over each column in the DataFrame
    for column in df.columns[1:]:
        # Extract values for the current column and convert to List
        values = df[column].tolist()
        # Flatten the List of Lists into a single List
        flattened_values = [item for sublist in values for item in sublist]
        # Store the flattened List in the dictionary with the column name as key
        flattened_values_by_column[column] = flattened_values

    # Store flattened_values_by_column dictionary in flattened_values_by_key_and_column with key as key
    residuals_by_disaster[key] = flattened_values_by_column

```

In [ ]: # Assuming residuals\_by\_disaster contains flattened values for each column and each disaster type

```

# Initialize a List to store all flattened values for concatenation
all_flattened_values = []

# Iterate over each key (disaster type) and its corresponding flattened values

```

```

for flattened_values in residuals_by_disaster.values():
    # Convert flattened values dictionary to a DataFrame and append to the list
    df_flattened = pd.DataFrame(flattened_values)
    all_flattened_values.append(df_flattened)

# Concatenate all DataFrames into a single DataFrame
combined_df = pd.concat(all_flattened_values, ignore_index=True)

# Calculate descriptive statistics for the combined DataFrame
descriptive_stats = combined_df.describe()

# Display the descriptive statistics
print("Descriptive Statistics for All Disaster Types:\n", descriptive_stats)

```

```

In [ ]: # Calculate the 1st and 99th percentiles
percentiles = combined_df.quantile([0.01, 0.99])

```

```

In [ ]: # Assuming residuals_by_disaster contains flattened values for each column and each disaster type
# Dictionary to store descriptive statistics for each disaster type
descriptive_stats_by_disaster = {}

# Iterate over each disaster type and its corresponding flattened values
for disaster_type, flattened_values in residuals_by_disaster.items():
    # Create a DataFrame from the flattened values
    df_flattened = pd.DataFrame(flattened_values)

    # Calculate descriptive statistics for each column
    descriptive_stats = df_flattened.describe()

    # Store the descriptive statistics in the dictionary with the disaster type as key
    descriptive_stats_by_disaster[disaster_type] = descriptive_stats

# Display the descriptive statistics for each disaster type
for disaster_type, stats in descriptive_stats_by_disaster.items():
    print(f"Descriptive Statistics for {disaster_type}:\n{stats}\n")

```

```

In [ ]: # Initialize an empty DataFrame to store all the data
all_data = pd.DataFrame()

# Iterate over each disaster type in the dictionary
for disaster_type, data in residuals_by_disaster.items():
    # Convert the nested dictionary into a DataFrame
    df = pd.DataFrame(data)

    # Add a new column for the disaster type
    df['Disaster_Type'] = disaster_type

    # Append the data to the all_data DataFrame
    all_data = pd.concat([all_data, df], ignore_index=True)

# Create indicator columns for each unique disaster type
for disaster in all_data['Disaster_Type'].unique():
    all_data[f'Disaster_{disaster.replace(" ", "_")}'] = all_data['Disaster_Type'].apply(lambda x: 1 if x == disaster else 0)

```

```

In [ ]: # List of disaster columns to sum
# 'disaster_cols' à changer à 'disaster_cols_gr' pour l'analyse selon les groupes de catastrophe naturel
'''disaster_cols_gr = ['Disaster_Sans_Catastrophe', 'Disaster_Technologique', 'Disaster_Hydrologique', 'Disaster_Météorologique',
'Disaster_Biologique', 'Disaster_Géophysique',
'Disaster_Climatologique']'''
disaster_cols = ['Disaster_No_Disaster', 'Disaster_Flood', 'Disaster_Others', 'Disaster_Storm',
'Disaster_Epidemic', 'Disaster_Mass_movement_(wet)',
'Disaster_Extreme_temperature', 'Disaster_Earthquake', 'Disaster_Wildfire',
'Disaster_Drought', 'Disaster_Volcanic_activity', 'Disaster_Mass_movement_(dry)']

# Create a new column 'Disaster' by summing up the disaster indicator columns
all_data['Disaster'] = all_data[disaster_cols[2:]].sum(axis=1)

```

```

In [ ]: # Dynamically get the column names from the first disaster type
columns = list(next(iter(residuals_by_disaster.values())).keys())
# Initialize a dictionary to store aggregated residuals for each column
aggregated_residuals = {column: [] for column in columns}

# Aggregate residuals for each column across all disaster types
for disaster_type, residuals_dict in residuals_by_disaster.items():
    for column in columns:
        if column in residuals_dict:
            aggregated_residuals[column].extend(residuals_dict[column])

# Initialize a DataFrame to store the sum test results
sum_test_results = pd.DataFrame(columns=['sum', 'z_statistic', 'p_value', 'n', 'result', 'std'], index=columns)

# Perform the sum = 0 test for each column
for column, residuals in aggregated_residuals.items():
    # Calculate the sum of residuals
    sum_residuals = np.sum(residuals)

    # Number of observations
    n = len(residuals)

    # Standard error for sum: std deviation * sqrt(n)
    std_error_sum = np.std(residuals, ddof=1) * np.sqrt(n)

    # Calculate z-statistic (sum divided by standard error)
    z_statistic = sum_residuals / std_error_sum

```

```

# Calculate p-value for the z-statistic
p_value = 2 * (1 - norm.cdf(abs(z_statistic))) # two-tailed

# Determine the result based on the p-value
result = 'Non rejet de H0' if p_value >= 0.05 else 'Rejet de H0'
# Calculer l'écart-type (std)
std_residuals = np.std(residuals, ddof=1) # ddof=1 pour l'échantillon
# Erreur standard pour la somme : écart type * sqrt(n)
std_error_sum = std_residuals * np.sqrt(n)

# Store the results in the DataFrame
sum_test_results.loc[column] = [sum_residuals, z_statistic, p_value, n, result, std_residuals]

# Display the results DataFrame for a quick overview
print("\n Sum = 0 Test Summary:")
sum_test_results

```

```

In [ ]: # Custom function to divide y-axis values by 1000
def thousands_formatter(x, pos):
    return f'{int(x / 1000)}k'

# List of columns to plot histograms for
columns_to_plot = list(next(iter(residuals_by_disaster.values())).keys())

# Dictionary to rename columns in the plot titles
column_names = {
    'Abn_disc_exp': 'coûts discrétionnaires anormaux',
    'Abn_cost_prod': 'coûts de production anormaux',
    'Abn_cash_flow': 'CFO anormaux',
    'Abn_accrual': 'accruals anormaux'
}

# Create a figure with 2x2 subplots
fig, axs = plt.subplots(2, 2, figsize=(12, 10))

# Iterate over each column (up to 4)
for i, column in enumerate(columns_to_plot[:4]): # Limit to the first four columns

    # Aggregate residuals for the current column across all disaster types
    aggregated_residuals = []
    for residuals_dict in residuals_by_disaster.values():
        aggregated_residuals.extend(residuals_dict[column])

    # Select the subplot for the current histogram
    ax = axs[i // 2, i % 2] # This maps 0, 1, 2, 3 to the grid of 2x2

    # Create histogram
    ax.hist(aggregated_residuals, bins=150, color='blue', edgecolor='black', alpha=0.7, range=(-30, 30))
    # Set the title using the renamed column name
    ax.set_title(f"Histogramme des {column_names.get(column, column)}")
    ax.set_xlabel(column_names.get(column, column))
    ax.set_ylabel("Fréquence")

    # Apply the custom y-axis formatter
    ax.yaxis.set_major_formatter(FuncFormatter(thousands_formatter))

# Adjust Layout to avoid overlapping of subplots
plt.tight_layout()

# Display the plot
plt.show()

```

```

In [ ]: # Dynamically get the column names from the first disaster type
columns = list(next(iter(residuals_by_disaster.values())).keys())

# Initialize a dictionary to store aggregated residuals for each column
aggregated_residuals = {column: [] for column in columns}

# Aggregate residuals for each column across all disaster types
for disaster_type, residuals_dict in residuals_by_disaster.items():
    for column in columns:
        if column in residuals_dict:
            aggregated_residuals[column].extend(residuals_dict[column])

# Initialize a DataFrame to store the test results
t_test_results = pd.DataFrame(columns=['mean', 'std', 'skewness', 'kurtosis',
                                       't_statistic', 't_p_value', 'n', 't_result',
                                       'jb_statistic', 'jb_p_value', 'jb_result'],
                              index=columns)

# Perform the tests and calculate descriptive statistics for each column
for column, residuals in aggregated_residuals.items():
    # Calculate descriptive statistics
    mean = np.mean(residuals)
    std = np.std(residuals, ddof=1) # ddof=1 for sample standard deviation
    skewness = skew(residuals)
    kurtosis_value = kurtosis(residuals)

    # Perform the one-sample t-test against 0
    t_statistic, t_p_value = ttest_1samp(residuals, 0)

    # Count the number of observations

```

```

n = len(residuals)

# Determine the t-test result based on the p-value
t_result = 'Non rejet de H0' if t_p_value >= 0.05 else 'Rejet de H0'

# Perform the Jarque-Bera test for normality
jb_statistic, jb_p_value = jarque_bera(residuals)

# Determine the normality result based on the Jarque-Bera p-value
jb_result = 'Normal' if jb_p_value >= 0.05 else 'Non-Normal'

# Store the results in the DataFrame
t_test_results.loc[column] = [mean, std, skewness, kurtosis_value,
                              t_statistic, t_p_value, n, t_result,
                              jb_statistic, jb_p_value, jb_result]

# Display the results DataFrame for a quick overview
print("\nFull Test Summary (including descriptive statistics):")
t_test_results

```

```

In [ ]: # Columns to use as dependent variables (Y)
dependent_vars = ['Abn_disc_exp', 'Abn_cost_prod', 'Abn_cash_flow', 'Abn_accrual']

# List of disaster columns to sum
'''disaster_cols = ['Disaster_Sans_Catastrophe', 'Disaster_Technologique', 'Disaster_Hydrologique', 'Disaster_Météorologique',
                   'Disaster_Biologique', 'Disaster_Géophysique',
                   'Disaster_Climatologique', 'Disaster']'''

disaster_cols = ['Disaster_No_Disaster', 'Disaster_Flood', 'Disaster_Others', 'Disaster_Storm',
                 'Disaster_Epidemic', 'Disaster_Mass_movement_(wet)',
                 'Disaster_Extreme_temperature', 'Disaster_Earthquake', 'Disaster_Wildfire',
                 'Disaster_Drought', 'Disaster_Volcanic_activity', 'Disaster_Mass_movement_(dry)']

# Dictionary to store the DataFrame results for each dependent variable
regression_dfs = {}

for y_col in dependent_vars:
    # Initialize a DataFrame to store the results for the current dependent variable
    results_df = pd.DataFrame(columns=['Coefficient', 'std', 'tstat', 'R2'])
    for x_col in disaster_cols:
        filtered_data = pd.concat([all_data[all_data['Disaster_No_Disaster'] == 1], all_data[all_data[x_col] == 1]], axis=0).drop_duplicates()
        X = filtered_data[x_col] # Directly use the independent variable without adding a constant
        y = filtered_data[y_col]

        model = sm.OLS(y, X).fit() # Fit the model without a constant

        # Extract the coefficient, std, t-statistic, and R-squared
        coefficient = model.params[0] # Coefficient of the disaster column
        std_error = model.bse[0] # Standard error of the coefficient
        tstat = model.tvalues[0] # t-statistic of the disaster column
        r2 = model.rsquared # R-squared of the model

        # Append the results to the DataFrame
        results_df.loc[x_col] = [coefficient, std_error, tstat, r2]

    # Store the DataFrame for the current dependent variable in the dictionary
    regression_dfs[y_col] = results_df
    print(f"Results for {y_col}:")
    print(regression_dfs[y_col], "\n")

```

## Analyse selon la configuration 2

```
In [1]: import numpy as np
import pandas as pd
import statsmodels.api as sm
from linearmodels.panel import PanelOLS
from scipy import stats
from scipy.stats import t
from scipy.stats import ttest_ind
import warnings
import matplotlib.pyplot as plt
# Suppress SettingWithCopyWarning
warnings.filterwarnings("ignore")
from scipy.stats import norm
import matplotlib.pyplot as plt
from matplotlib.ticker import FuncFormatter
from scipy.stats import ttest_1samp, jarque_bera, skew, kurtosis

In [2]: # Read disaster file into a DataFrame
data=pd.read_csv(r"C:\Users\yacia\OneDrive - HEC Montréal\HEC\Données\merged_data_with_GDPperCapita_new.csv",on_bad_lines='warn')
# Filter the DataFrame to contain only rows where GDP_per_capita is Lower than 15000
data = data[data['GDP_per_capita'] < 15000]

In [3]: # Dropping rows where 'Disaster Type' is 'Infestation'
data = data[data['Disaster Type'] != 'Infestation']

In [ ]: #Suppression of outliers
data = data[data['DiscExp/Asset_lag'] <= 1]
data = data[data['DiscExp/Asset_lag'] >= -1]
data = data[data['ProdExp/Asset_lag'] <= 1]
data = data[data['ProdExp/Asset_lag'] >= -1]
data = data[data['CFO/Asset_lag'] <= 1]
data = data[data['CFO/Asset_lag'] >= -1]
data = data[data['Accrual/Asset_lag'] <= 1]
data = data[data['Accrual/Asset_lag'] >= -1]

In [6]: # Define the List of disaster types that should not be changed
disaster_keep = ['Flood', 'Storm', 'Epidemic', 'Extreme temperature',
                 'Mass movement (dry)', 'Mass movement (wet)', 'Earthquake',
                 'Wildfire', 'Drought', 'Volcanic activity', 'No Disaster']

# Replace disaster types not in the List with 'Others'
data['Disaster Type'] = data['Disaster Type'].apply(lambda x: x if x in disaster_keep else 'Others')
# Donne Le nom des éléments contenu dans la colonne Disaster pour séparer Les statistiques par La suite
disaster_counts = data['Disaster Type'].value_counts()

In [ ]: #Ajustement taille et désastre
iso = data['ISO'].unique()
for iso in iso:
    num_observations = data[data['ISO'] == iso].shape[0]
    if num_observations < 30:
        data = data[data['ISO'] != iso]
disaster_types = data['Disaster Type'].unique()
for disaster_type in disaster_types:
    num_observations = data[data['Disaster Type'] == disaster_type].shape[0]
    if num_observations < 200:
        data.loc[data['Disaster Type'] == disaster_type, 'Disaster Type'] = 'Others'
# Sort the DataFrame by the 'Disaster' and 'ISO' columns in descending order
data = data.sort_values(by=['Disaster', 'ISO'], ascending=False)
# Nombre d'observation
print(data.shape)
disaster_counts = data['Disaster Type'].value_counts()

In [8]: # Function to classify disasters
def classify_disaster(cata):
    if cata in ['Earthquake', 'Mass movement (dry)', 'Volcanic activity']:
        return 'Géophysique'
    elif cata in ['Flood', 'Mass movement (wet)']:
        return 'Hydrologique'
    elif cata in ['Storm', 'Extreme temperature']:
        return 'Météorologique'
    elif cata in ['Drought', 'Wildfire']:
        return 'Climatologique'
    elif cata == 'Epidemic':
        return 'Biologique'
    elif cata == 'No Disaster':
        return 'Sans Catastrophe'
    else:
        return 'Technologique'

# Apply the classification function to the DataFrame
data['Groupe'] = data['Disaster Type'].apply(classify_disaster)

# Donne Le nom des éléments contenu dans la colonne Disaster type pour séparer Les statistiques par La suite
# 'Disaster Type' à changer à 'Groupe' pour L'analyse selon Les groupes de catastrophe naturel
disaster_types = data['Disaster Type'].unique()

In [9]: def calculate_regression_results(data, disaster_type, type):
    '''La fonction calculate_regression_results prend des données, un liste d'element d'une colonne précis et la colonne spécifique,
    effectue des régressions sur ces données pour chaque unité statistique,
```

```

effectue des régressions sur ces données, et retourne les résidus des régressions en ne considérant pas d'effets fixes
et en excluant les unités statistiques avec des exceptions.'''
# Subset the data for the specified disaster type
reg_data = data[data[type] == disaster_type]

# Initialize a DataFrame for residuals
reg_results = pd.DataFrame({
    'Abn_disc_exp': [],
    'Abn_cost_prod': [],
    'Abn_cash_flow': [],
    'Abn_accrual': []
})

# Initialize a DataFrame for regression statistics
stats_df = pd.DataFrame({
    'Abn_disc_exp_F_stat': [],
    'Abn_disc_exp_R2_adj': [],
    'Abn_disc_exp_T_stat': [],
    'Abn_cost_prod_F_stat': [],
    'Abn_cost_prod_R2_adj': [],
    'Abn_cost_prod_T_stat': [],
    'Abn_cash_flow_F_stat': [],
    'Abn_cash_flow_R2_adj': [],
    'Abn_cash_flow_T_stat': [],
    'Abn_accrual_F_stat': [],
    'Abn_accrual_R2_adj': [],
    'Abn_accrual_T_stat': []
})

try:
    # Ensure the date column is in datetime format
    reg_data['FY Date'] = pd.to_datetime(reg_data['FY Date'])
    reg_data.set_index(['gvkey', 'FY Date'], inplace=True)

    # Regression for abnormal discretionary expense
    X1 = reg_data[['Var_1/AL', 'Var_SL/AL']]
    y1 = reg_data['DiscExp/Asset_lag']
    model1 = PanelOLS(y1, X1, entity_effects=False, drop_absorbed=True, check_rank=False).fit(cov_type='clustered')
    reg_results['Abn_disc_exp'] = model1.resids.tolist()
    stats_df['Abn_disc_exp_F_stat'] = model1.f_statistic_robust.stat
    stats_df['Abn_disc_exp_R2_adj'] = model1.rsquared

    # Regression for abnormal production expense
    X2 = reg_data[['Var_1/AL', 'Var_SL/AL', 'Var_DS/AL', 'Var_DSL/AL']]
    y2 = reg_data['ProdExp/Asset_lag']
    model2 = PanelOLS(y2, X2, entity_effects=False, drop_absorbed=True, check_rank=False).fit(cov_type='clustered')
    reg_results['Abn_cost_prod'] = model2.resids.tolist()
    stats_df['Abn_cost_prod_F_stat'] = model2.f_statistic_robust.stat
    stats_df['Abn_cost_prod_R2_adj'] = model2.rsquared

    #print(disaster_type)
    #print(model2.summary)
    # Regression for abnormal cash flow
    X3 = reg_data[['Var_1/AL', 'Var_SL/AL', 'Var_DS/AL']]
    y3 = reg_data['CF0/Asset_lag']
    model3 = PanelOLS(y3, X3, entity_effects=False, drop_absorbed=True, check_rank=False).fit(cov_type='clustered')
    reg_results['Abn_cash_flow'] = model3.resids.tolist()
    stats_df['Abn_cash_flow_F_stat'] = model3.f_statistic_robust.stat
    stats_df['Abn_cash_flow_R2_adj'] = model3.rsquared

    # Regression for abnormal accrual
    X4 = reg_data[['Var_1/AL', 'Var_2', 'Var_3']]
    y4 = reg_data['Accrual/Asset_lag']
    model4 = PanelOLS(y4, X4, entity_effects=False, drop_absorbed=True, check_rank=False).fit(cov_type='clustered')
    reg_results['Abn_accrual'] = model4.resids.tolist()
    stats_df['Abn_accrual_F_stat'] = model4.f_statistic_robust.stat
    stats_df['Abn_accrual_R2_adj'] = model4.rsquared
    # Print the model summary table

except Exception as e:
    print(f"Error encountered: {e}")

return reg_results, stats_df

```

In [255...] '''Ce code crée un dictionnaire vide reg\_results\_dict et itère à travers les différents types de désastres. Pour chaque type de désastre, il appelle la fonction calculate\_regression\_results pour calculer les résultats de la régression spécifiques à ce type de désastre, puis stocke ces résultats dans le dictionnaire sous une clé basée sur le type de désastre.'''

```

reg_results_dict = {}
stats_df_dict = {}
# Iterate over unique disaster types
for disaster_type in disaster_types:
    # Call the function to calculate regression results for the current disaster type
    # 'Disaster Type' à changer à 'Groupe' pour l'analyse selon les groupes de catastrophe naturel
    reg_results, stats_df = calculate_regression_results(data, disaster_type, 'Disaster Type')
    # Store the results in the dictionary with a key based on the disaster type
    reg_results_dict[f"{disaster_type}"] = reg_results
    stats_df_dict[f"{disaster_type}"] = stats_df

```

In [256...] '''Ce code itère à travers un dictionnaire reg\_results\_dict contenant des DataFrames de résultats de régression pour s'il y'a désastre ou non. Pour chaque DataFrame, il extrait et aplatit les valeurs de chaque colonne, puis stocke ces valeurs aplaties dans un dictionnaire residuals\_by\_disaster organisé par type de désastre et nom de colonne d'accruals.'''

```

# Dictionary to store flattened values for each column and each key
residuals_by_disaster = {}

```

```

# Iterate over each key (disaster type) and DataFrame in reg_results_dict
for key, df in reg_results_dict.items():
    # Dictionary to store values for the current key
    values_by_column = {}
    # Iterate over each column in the DataFrame
    for column in df.columns:
        # Extract values for the current column and convert to List
        values = df[column].tolist()
        # Store the list of values in the dictionary with the column name as key
        values_by_column[column] = values
    # Store values_by_column dictionary in residuals_by_disaster with key as the disaster type
    residuals_by_disaster[key] = values_by_column

```

```

In [ ]: # Assuming residuals_by_disaster contains flattened values for each column and each disaster type
# Initialize a list to store all flattened values for concatenation
all_flattened_values = []

# Iterate over each key (disaster type) and its corresponding flattened values
for flattened_values in residuals_by_disaster.values():
    # Convert flattened values dictionary to a DataFrame and append to the list
    df_flattened = pd.DataFrame(flattened_values)
    all_flattened_values.append(df_flattened)

# Concatenate all DataFrames into a single DataFrame
combined_df = pd.concat(all_flattened_values, ignore_index=True)

# Calculate descriptive statistics for the combined DataFrame
descriptive_stats = combined_df.describe()

# Display the descriptive statistics
print("Descriptive Statistics for All Disaster Types:\n", descriptive_stats)

```

```

In [ ]: # Calculate the 1st and 99th percentiles
percentiles = combined_df.quantile([0.01, 0.99])

```

```

In [ ]: # Assuming residuals_by_disaster contains flattened values for each column and each disaster type
# Dictionary to store descriptive statistics for each disaster type
descriptive_stats_by_disaster = {}

# Iterate over each disaster type and its corresponding flattened values
for disaster_type, flattened_values in residuals_by_disaster.items():
    # Create a DataFrame from the flattened values
    df_flattened = pd.DataFrame(flattened_values)

    # Calculate descriptive statistics for each column
    descriptive_stats = df_flattened.describe()

    # Store the descriptive statistics in the dictionary with the disaster type as key
    descriptive_stats_by_disaster[disaster_type] = descriptive_stats

# Display the descriptive statistics for each disaster type
for disaster_type, stats in descriptive_stats_by_disaster.items():
    print(f"Descriptive Statistics for {disaster_type}:\n{stats}\n")

```

```

In [ ]: # Initialize an empty DataFrame to store all the data
all_data = pd.DataFrame()

# Iterate over each disaster type in the dictionary
for disaster_type, data in residuals_by_disaster.items():
    # Convert the nested dictionary into a DataFrame
    df = pd.DataFrame(data)

    # Add a new column for the disaster type
    df['Disaster_Type'] = disaster_type

    # Append the data to the all_data DataFrame
    all_data = pd.concat([all_data, df], ignore_index=True)

# Create indicator columns for each unique disaster type
for disaster in all_data['Disaster_Type'].unique():
    all_data[f'Disaster_{disaster.replace(" ", "_")}'] = all_data['Disaster_Type'].apply(lambda x: 1 if x == disaster else 0)

```

```

In [ ]: # List of disaster columns to sum
# 'disaster_cols' à changer à 'disaster_cols_gr' pour l'analyse selon les groupes de catastrophe naturel
'''disaster_cols_gr = ['Disaster_Sans_Catastrophe', 'Disaster_Technologique', 'Disaster_Hydrologique', 'Disaster_Météorologique',
'Disaster_Biologique', 'Disaster_Géophysique',
'Disaster_Climatologique']'''
disaster_cols = ['Disaster_No_Disaster', 'Disaster_Flood', 'Disaster_Others', 'Disaster_Storm',
'Disaster_Epidemic', 'Disaster_Mass_movement_(wet)',
'Disaster_Extreme_temperature', 'Disaster_Earthquake', 'Disaster_Wildfire',
'Disaster_Drought', 'Disaster_Volcanic_activity', 'Disaster_Mass_movement_(dry)']

# Create a new column 'Disaster' by summing up the disaster indicator columns
all_data['Disaster'] = all_data[disaster_cols[2:]].sum(axis=1)

```

```

In [ ]: # Dynamically get the column names from the first disaster type
columns = list(next(iter(residuals_by_disaster.values())).keys())
# Initialize a dictionary to store aggregated residuals for each column
aggregated_residuals = {column: [] for column in columns}

# Aggregate residuals for each column across all disaster types
for disaster_type, residuals_dict in residuals_by_disaster.items():

```



```

        for column in columns:
            if column in residuals_dict:
                aggregated_residuals[column].extend(residuals_dict[column])
# Initialize a DataFrame to store the sum test results
sum_test_results = pd.DataFrame(columns=['sum', 'z_statistic', 'p_value', 'n', 'result', 'std'], index=columns)

# Perform the sum = 0 test for each column
for column, residuals in aggregated_residuals.items():
    # Calculate the sum of residuals
    sum_residuals = np.sum(residuals)

    # Number of observations
    n = len(residuals)

    # Standard error for sum: std deviation * sqrt(n)
    std_error_sum = np.std(residuals, ddof=1) * np.sqrt(n)

    # Calculate z-statistic (sum divided by standard error)
    z_statistic = sum_residuals / std_error_sum

    # Calculate p-value for the z-statistic
    p_value = 2 * (1 - norm.cdf(abs(z_statistic))) # two-tailed

    # Determine the result based on the p-value
    result = 'Non rejet de H0' if p_value >= 0.05 else 'Rejet de H0'
    # Calculer l'écart-type (std)
    std_residuals = np.std(residuals, ddof=1) # ddof=1 pour l'échantillon
    # Erreur standard pour la somme : écart type * sqrt(n)
    std_error_sum = std_residuals * np.sqrt(n)

    # Store the results in the DataFrame
    sum_test_results.loc[column] = [sum_residuals, z_statistic, p_value, n, result, std_residuals]

# Display the results DataFrame for a quick overview
print("\n Sum = 0 Test Summary:")
sum_test_results

```

```

In [ ]: # Custom function to divide y-axis values by 1000
def thousands_formatter(x, pos):
    return f'{int(x / 1000)}k'

# List of columns to plot histograms for
columns_to_plot = list(next(iter(residuals_by_disaster.values())).keys())

# Dictionary to rename columns in the plot titles
column_names = {
    'Abn_disc_exp': 'coûts discrétionnaires anormaux',
    'Abn_cost_prod': 'coûts de production anormaux',
    'Abn_cash_flow': 'CFO anormaux',
    'Abn_accrual': 'accruals anormaux'
}

# Create a figure with 2x2 subplots
fig, axs = plt.subplots(2, 2, figsize=(12, 10))

# Iterate over each column (up to 4)
for i, column in enumerate(columns_to_plot[:4]): # Limit to the first four columns

    # Aggregate residuals for the current column across all disaster types
    aggregated_residuals = []
    for residuals_dict in residuals_by_disaster.values():
        aggregated_residuals.extend(residuals_dict[column])

    # Select the subplot for the current histogram
    ax = axs[i // 2, i % 2] # This maps 0, 1, 2, 3 to the grid of 2x2

    # Create histogram
    ax.hist(aggregated_residuals, bins=150, color='blue', edgecolor='black', alpha=0.7, range=(-30, 30))
    # Set the title using the renamed column name
    ax.set_title(f"Histogramme des {column_names.get(column, column)}")
    ax.set_xlabel(column_names.get(column, column))
    ax.set_ylabel("Fréquence")

    # Apply the custom y-axis formatter
    ax.yaxis.set_major_formatter(FuncFormatter(thousands_formatter))

# Adjust layout to avoid overlapping of subplots
plt.tight_layout()

# Display the plot
plt.show()

```

```

In [ ]: # Dynamically get the column names from the first disaster type
columns = list(next(iter(residuals_by_disaster.values())).keys())

# Initialize a dictionary to store aggregated residuals for each column
aggregated_residuals = {column: [] for column in columns}

# Aggregate residuals for each column across all disaster types
for disaster_type, residuals_dict in residuals_by_disaster.items():
    for column in columns:
        if column in residuals_dict:
            aggregated_residuals[column].extend(residuals_dict[column])

```

```

# Initialize a DataFrame to store the test results
t_test_results = pd.DataFrame(columns=['mean', 'std', 'skewness', 'kurtosis',
                                       't_statistic', 't_p_value', 'n', 't_result',
                                       'jb_statistic', 'jb_p_value', 'jb_result'],
                              index=columns)

# Perform the tests and calculate descriptive statistics for each column
for column, residuals in aggregated_residuals.items():
    # Calculate descriptive statistics
    mean = np.mean(residuals)
    std = np.std(residuals, ddof=1) # ddof=1 for sample standard deviation
    skewness = skew(residuals)
    kurtosis_value = kurtosis(residuals)

    # Perform the one-sample t-test against 0
    t_statistic, t_p_value = ttest_1samp(residuals, 0)

    # Count the number of observations
    n = len(residuals)

    # Determine the t-test result based on the p-value
    t_result = 'Non rejet de H0' if t_p_value >= 0.05 else 'Rejet de H0'

    # Perform the Jarque-Bera test for normality
    jb_statistic, jb_p_value = jarque_bera(residuals)

    # Determine the normality result based on the Jarque-Bera p-value
    jb_result = 'Normal' if jb_p_value >= 0.05 else 'Non-Normal'

    # Store the results in the DataFrame
    t_test_results.loc[column] = [mean, std, skewness, kurtosis_value,
                                t_statistic, t_p_value, n, t_result,
                                jb_statistic, jb_p_value, jb_result]

# Display the results DataFrame for a quick overview
print("\nFull Test Summary (including descriptive statistics):")
t_test_results

```

```

In [ ]: # Columns to use as dependent variables (Y)
dependent_vars = ['Abn_disc_exp', 'Abn_cost_prod', 'Abn_cash_flow', 'Abn_accrual']

# List of disaster columns to sum
'''disaster_cols_gr = ['Disaster_Sans_Catastrophe', 'Disaster_Technologique', 'Disaster_Hydrologique', 'Disaster_Météorologique',
                      'Disaster_Biologique', 'Disaster_Géophysique',
                      'Disaster_Climatologique', 'Disaster']'''

disaster_cols = ['Disaster_No_Disaster', 'Disaster_Flood', 'Disaster_Others', 'Disaster_Storm',
                 'Disaster_Epidemic', 'Disaster_Mass_movement_(wet)',
                 'Disaster_Extreme_temperature', 'Disaster_Earthquake', 'Disaster_Wildfire',
                 'Disaster_Drought', 'Disaster_Volcanic_activity', 'Disaster_Mass_movement_(dry)']

# Dictionary to store the DataFrame results for each dependent variable
regression_dfs = {}

for y_col in dependent_vars:
    # Initialize a DataFrame to store the results for the current dependent variable
    results_df = pd.DataFrame(columns=['Coefficient', 'std', 'tstat', 'R2'])
    for x_col in disaster_cols:
        filtered_data = pd.concat([all_data[all_data['Disaster_No_Disaster'] == 1], all_data[all_data[x_col] == 1]], axis=0).drop_duplicates()
        X = filtered_data[x_col] # Directly use the independent variable without adding a constant
        y = filtered_data[y_col]

        model = sm.OLS(y, X).fit() # Fit the model without a constant

        # Extract the coefficient, std, t-statistic, and R-squared
        coefficient = model.params[0] # Coefficient of the disaster column
        std_error = model.bse[0] # Standard error of the coefficient
        tstat = model.tvalues[0] # t-statistic of the disaster column
        r2 = model.rsquared # R-squared of the model

        # Append the results to the DataFrame
        results_df.loc[x_col] = [coefficient, std_error, tstat, r2]

    # Store the DataFrame for the current dependent variable in the dictionary
    regression_dfs[y_col] = results_df
    print(f"Results for {y_col}:")
    print(regression_dfs[y_col], "\n")

```

## Définition des catastrophes naturelles

**Geophysical hazard** : Un danger provenant de la terre solide. Ce terme est utilisé de manière interchangeable avec le terme “danger géologique”.

- ***Earthquake (tremblement de terre)*** : Mouvement soudain d’un bloc de la croûte terrestre le long d’une faille géologique, accompagné de secousses.
- ***Mass movement (dry) (mouvement de masse sec)*** : Tout type de mouvement descendant de matériaux terrestres dans des conditions hydrologiques sèches.
- ***Volcanic activity (activité volcanique)*** : Un type d’événement volcanique près d’une ouverture/vent dans la surface de la Terre, y compris les éruptions de lave, de cendres, de vapeur chaude, de gaz et de matériaux pyroclastiques.

### Catastrophes hydrologiques

- ***Hydrological hazard*** : Un danger causé par l’occurrence, le mouvement et la distribution des eaux de surface et souterraines, douces et salées.
- ***Flood (inondation)*** : Terme général pour le débordement d’eau d’un cours d’eau sur des terres normalement sèches dans la plaine inondable (inondation fluviale), des niveaux plus élevés que la normale le long de la côte (inondation côtière) et dans les lacs ou réservoirs ainsi que l’accumulation d’eau à ou près du point où la pluie est tombée (crues soudaines).
- ***Mass movement (wet) (mouvement de masse humide)*** : Types de mouvements de masse qui se produisent lorsque de fortes pluies ou une fonte rapide de neige/glace envoient de grandes quantités de végétation, de boue ou de roches en bas d’une pente, entraînées par des forces gravitationnelles.

### Catastrophes météorologiques

- ***Meteorological hazard*** : Un danger causé par des conditions météorologiques extrêmes de courte durée, de micro à méso-échelle, qui durent de quelques minutes à quelques jours.
- ***Extreme temperature (température extrême)*** : Terme général pour les variations de température au-dessus (chaleur extrême) ou en dessous (froid extrême) des conditions normales.
- ***Storm (tempête)*** : À compléter.

### Catastrophes climatologiques

- ***Climatological hazard*** : Un danger causé par des processus atmosphériques de longue durée, de méso à macro-échelle, allant de la variabilité climatique intra-saisonnière à multi-décennale.
- ***Drought (sécheresse)*** : Une période prolongée de précipitations inhabituellement faibles qui produit une pénurie d'eau pour les personnes, les animaux et les plantes. La sécheresse se développe lentement, parfois même sur des années, et son apparition est généralement difficile à détecter. Elle n'est pas uniquement un phénomène physique car ses impacts peuvent être exacerbés par les activités humaines et les demandes en eau. La sécheresse est donc souvent définie à la fois conceptuellement et opérationnellement. Les définitions opérationnelles de la sécheresse, c'est-à-dire le degré de réduction des précipitations qui constitue une sécheresse, varient selon la localité, le climat et le secteur environnemental.
- ***Wildfire (feu de forêt)*** : Tout incendie non contrôlé et non prescrit de plantes dans un environnement naturel tel qu'une forêt, une prairie, un maquis ou une toundra, qui consomme des combustibles naturels et se propage en fonction des conditions environnementales (par exemple, le vent ou la topographie). Les feux de forêt peuvent être déclenchés par la foudre ou des actions humaines.

#### Catastrophes biologiques :

- ***Biological hazard*** : Un danger causé par l'exposition à des organismes vivants et/ou à leurs substances toxiques (par exemple, venin ou moisissure) ou à des maladies vectorielles qu'ils peuvent transporter. Exemples : faune venimeuse et insectes, plantes toxiques, proliférations d'algues et moustiques transportant des agents pathogènes tels que des parasites, des bactéries ou des virus (par exemple, le paludisme).
- ***Epidemic (épidémie)*** : Une augmentation inhabituelle, souvent soudaine, du nombre de cas d'une maladie infectieuse déjà présente dans la région (par exemple, grippe ) ou l'apparition d'une maladie infectieuse auparavant absente de la région (par exemple, peste ou polio).

#### Catastrophes extra-terrestres

- ***Extra-terrestrial hazard*** : Un danger causé par des astéroïdes, des météoroïdes et des comètes lorsqu'ils passent près de la Terre, entrent dans l'atmosphère terrestre et/ou frappent la Terre, et par des changements dans

les conditions interplanétaires qui affectent la magnétosphère, l'ionosphère et la thermosphère de la Terre.

### **Définition des catastrophes technologiques**

**Transport accident (accident de transport)** : Terme utilisé pour décrire les accidents technologiques de transport impliquant des modes de transport mécanisés. Il comprend quatre sous-types de catastrophes (c'est-à-dire air, eau, rail et route).

**Industrial accident (accident industriel)** : Terme utilisé pour décrire les accidents technologiques de nature industrielle/impliquant des bâtiments industriels (par exemple, usines).

**Miscellaneous accident (accident divers)** : Terme utilisé pour décrire les accidents technologiques de nature non industrielle ou de transport (par exemple, impliquant des maisons).

**Collapse (effondrement)** : Accident impliquant l'effondrement d'un bâtiment ou d'une structure. Peut impliquer des structures industrielles ou domestiques/non industrielles.

**Explosion (explosion)** : Explosions impliquant des bâtiments ou des structures. Peut impliquer des structures industrielles.

**Fire (incendie)** : Incendie urbain impliquant des bâtiments ou des structures. Peut impliquer des structures industrielles.