RISQUE ASSOCIE A L'UTILISATION DE LA LOI DE BENFORD POUR DETECTER DES VENTES FRAUDULEUSES DE BIENS INNOVANTS A LA MODE

Adrien Bonache, Allocataire moniteur normalien, Université Montpellier II-MRM CREGOR COST, bonache@rip.ens-cachan.fr

Karen Moris, ATER, Université de Bourgogne LEG-FARGO, karenmoris@yahoo.fr

Résumé : En principe, une non-adéquation à la loi de Benford serait synonyme de fraudes. Le présent papier montre qu'il n'est pas toujours possible de détecter des fraudes sur les volumes de ventes à l'aide de cette loi. Pour ce faire, nous utilisons les ventes de consoles, en volume, au Japon (depuis 1989), aux États-Unis, en France, en Allemagne et au Royaume-Uni (depuis 2000). Après une brève revue de littérature et une présentation de notre méthode, nous testons l'adéquation à la loi de Benford de nos 56 séries de ventes en volume via des statistiques du chi-deux, puis à une analyse des biais et de leur significativité. Malgré l'absence de fraude, il en ressort une non-adéquation à la loi de Benford de nos séries de vente de biens à la mode. Ainsi, pour des ventes de consoles, il est possible que l'utilisation de la loi de Benford ne soit pas efficace pour repérer d'éventuelles fraudes sur les volumes vendus.

Mots clés : Loi de Benford, ventes de biens à la mode, détection de fraudes, audit, système dynamique non-linéaire

Abstract: Benford's law has been promoted as providing the auditors with a turnkey solution for fraud detection. The purpose of this paper is to show it is not always possible to detect fraudulent sales with that law. We use sales in volume of game consoles in Japan (since 1989), in United-States, in France, in Germany and in United-Kingdom (since 2000). After reviewing briefly the literature and our study design, the chi-square test and the bias analysis were used to measure the goodness-of-fit to Benford's law. Despite the absence of actual fraud, these sale series of fashion goods are not significantly in conformity with Benford's law. Thus, for the detection of fraudulent sales in this sector, the Benford's law could be useless.

Key words: Benford's law, fashion sales, detection of fraud, audit, non-linear dynamical system.

«Une vente frauduleuse importante au cours d'un exercice peut être détectée avec les procédures classiques d'analyse en comparant les ventes avec le coût des marchandises vendues. Cependant, des ventes faiblement frauduleuses mais nombreuses ne peuvent pas être détectées par les méthodes traditionnelles d'analyse. Une procédure de détection statistique de fraudes basée sur la loi de Benford peut éventuellement trouver de telles manipulations, même si la proportion des données fictives est faible (10% des montants analysés). Le test d'adéquation à cette loi est insensible à l'ampleur de l'erreur.
Ce test est également indépendant de la relation du poste comptable étudié avec d'autres postes. Tel que décrit dans l'exemple ci-dessus, une procédure traditionnelle d'examen analytique peut typiquement analyser la relation entre les ventes et le coût des marchandises vendues. Une procédure de détection statistique de fraudes basée de Benford n'utilise pas les relations entre postes comptables, fournissant ainsi un angle «différent» pour enquêter sur des données. » (Busta et Weinberg, 1998)

L'utilisation de la loi de Benford pour détecter des fraudes dans des données économiques et sociales trouve son origine dans un article de Varian (1972). Son hypothèse est que les fraudeurs auraient tendance à manipuler les chiffres selon une loi uniforme. Dit autrement, ils utiliseraient les chiffres dans les mêmes proportions, en ayant recours à autant de 1 que de 2, de 3... et de 9. Or, cette tendance irait à l'encontre d'une régularité empirique constatée sur de nombreuses données expérimentales: la loi de Benford. Il est possible de la définir simplement et *a priori* de la façon suivant : « La loi de Benford rend compte de la fréquence d'apparition du premier chiffre d'un nombre dans une série de nombres » (Bonache *et al.*, 2010). Cette loi prévoit que le premier chiffre significatif d'un nombre tiré de manière aléatoire suit une loi logarithmique et non une loi uniforme. C'est à dire que, dans la nature, le chiffre 1 apparaît plus souvent que le chiffre 2, qui lui-même apparaît plus souvent que le chiffre 3... en première position d'un nombre. Partant de cette constatation, le fisc américain, certains auditeurs et des chercheurs en comptabilité ont vu un moyen de déceler des fraudes fiscales et comptables en analysant leur adéquation ou non à la loi de Benford.
Cependant, cette loi serait-elle utilisable pour repérer des manipulations dans tous les secteurs d'activité? Et peut-on l'appliquer pour repérer des fraudes sur tous les postes comptables?

Sur le plan de la méthode, l'intérêt de se poser la question de l'efficacité de la loi de Benford sur certains secteurs n'est pas nul tant pour le praticien que pour le chercheur. En effet, pour le chercheur, il est intéressant de se demander si la non-adéquation de la loi de Benford ne traduit qu'une fraude ou si elle est imputable à la distribution de la variable analysée en dehors de toute manipulation. Inversement, l'adéquation à la loi de Benford à une série de mesures ne permet pas de conclure totalement à l'absence de fraude. La connaissance des secteurs où la loi de Benford peut ne pas s'appliquer ou ne s'applique pas avec certitude pourrait rendre plus efficace son utilisation par les auditeurs et les agents du fisc. Et paradoxalement, si dans un secteur d'activité, la distribution d'une variable ne devrait pas suivre la loi de Benford, les auditeurs et agents du fisc pourraient quand même vérifier les comptes au regard de cette loi : la découverte d'une adéquation à cette loi pourrait alors signifier non pas l'absence de fraude, mais bien la présence d'une fraude. Cela peut être le cas d'une fraude faite par une personne qui connaît l'existence de cet outil de contrôle.

Enfin, l'étude d'un secteur d'activité où la loi de Benford ne s'applique pas pour déceler les fraudes aurait un intérêt théorique. En effet, toute loi et tout contrôle se doivent d'être délimités dans leur applicabilité avant même de les mettre en pratique. Cette contribution participe à l'affinement d'une définition en creux des domaines où l'on peut appliquer cette loi pour déceler les fraudes, puisqu'elle se base sur l'analyse d'un secteur particulier et d'un type de données précises. Cela permettrait d'éviter un risque de première espèce pour les institutions spécialisées dans la révélation des manipulations comptables: déceler une fraude dans une entreprise ne maquillant pas ses comptes.

Précisons, avant toute chose, pourquoi l'étude porte sur un secteur de biens à la mode. Un certain nombre de travaux ont mis en avant que les achats de ce type de biens par les consommateurs se faisaient par des effets d'imitation ou de snobisme (pour un état de l'art voir Miller *et al*. (1993)). Puis dans des tentatives de modélisations indépendantes, Granovetter et Soong (1986) et Nakayama et Nakamura (2004) soulignent que ces interdépendances entre les acheteurs, donnant lieu à des effets d'imitation et de snobisme, pouvaient aboutir à un comportement complexe des ventes de biens innovants à la mode. Un bien à la mode est alors défini comme un produit qu'« un consommateur achète en fonction du nombre d'acheteurs passés » (Granovetter et Soong, 1986, p.83).

L'objet de cette contribution est de montrer que les comptes de certaines entreprises peuvent ne pas suivre la loi de Benford bien qu'ils ne soient pas falsifiés. Dans une posture épistémologique falsificationniste, l'analyse fait ressortir un secteur où la recherche du risque de fraude avec la loi de Benford amènerait un auditeur ou un agent du fisc à prendre le risque de se tromper. Et cela, malgré le fait qu'il soit admis que ce type de données serait en adéquation avec cette loi. En bref, la finalité est d'infirmer la loi de Benford sur un cas précis, sans infirmer globalement cette loi pouvant, par ailleurs, servir à déceler des fraudes sur les volumes de ventes de manière efficace.

Il n'est pas question ici d'autres moyens de détection de fraudes statistique ou non : détection de seuils dans la distribution des ventes, analyses de corrélations, comparaisons de distributions, etc. (Bolton et Hand, 2002). Les analyses de corrélations ne sont utilisables lorsqu'on a accès à d'autres postes comptables que les seules ventes. Elles permettent alors de voir la cohérence du tout et de détecter une absence de corrélations attendues et la présence de corrélations inattendues avant une investigation en profondeur des postes comptables suspects. Cependant, suivant Busta et Weinberg (1998), cette technique ne permet pas de voir les fraudes « faibles mais nombreuses ». Les analyses de seuils, typiquement utilisées pour repérer la gestion des résultats, ne peuvent être mobilisées pour la détection de ventes frauduleuses tant qu'un seuil n'est pas attendu pour les ventes. Cependant, suivant Carslaw (1988) et Thomas (1989), il est possible de s'attendre à une tendance à vouloir atteindre ou dépasser un point de référence. Par exemple, un directeur des ventes d'un magasin pourrait dire qu'il a vendu 100009 produits alors qu'il n'en a vendu que 99994. Cela peut notamment venir de l'existence de primes d'objectifs par seuil. Cependant, si les ventes étudiées sont chaotiques, il est possible que des seuils et des discontinuités apparaissent dans la distribution sans fraude (cf. figure en annexe A). Concernant les comparaisons de distributions (tests de comparaison de moyennes, de variances, de rangs…), suivant les produits, les moyennes et les variances des ventes diffèrent. On ne peut ainsi utiliser des tests de comparaison de moyennes ou de variances (tests de Student, de Kruskal et Wallis ou de Wilcoxon et Mann-Whitney).

Analytiquement, Granovetter et Soong (1986) et Nakayama et Nakamura (2004) ont mis en avant que les ventes dans ce secteur pouvaient avoir un comportement chaotique du fait des interactions entre les consommateurs. Au moyen de séries simulées à partir de dynamiques non-linéaires connues, Tolle *et al*. (2000) font ressortir que des variables ayant un comportement complexe pouvaient très bien être en adéquation avec la loi de Benford ou ne pas l'être. De plus, pour certaines simulations, les nombres obtenus étaient distribués suivant une loi uniforme ; cela est, d'après Varian (1972), caractéristique d'une fraude alors qu'en l'occurrence ce n'était pas le cas. Ainsi, si les ventes de produits à la mode ont un comportement chaotique, il se peut que l'usage de la loi de Benford ne soit pas efficace (voire dangereux) dans le repérage de fraudes dans ce secteur.

Afin de développer les nombreux intérêts de cette étude, nous avons retenu la question de recherche suivante: les ventes de produits à la mode suivent-elles la loi de Benford?

Dans un premier temps, l'étude de cette question débute par une clarification conceptuelle de la loi de Benford et l'illustration de son efficacité par quelques travaux en comptabilité l'utilisant pour déceler des fraudes (1.1.). Puis, l'origine de nos données est précisée, ainsi que le canevas de l'étude (1.2.). Dans un second temps, sont présentés les résultats (2.1.) et une discussion sur les limites et les recherches futures (2.2.).

1 Éléments de clarification théoriques et méthode d'investigation

Avant de présenter la méthode et les résultats, réduisons toute équivoque autour de la notion principale de l'analyse : la loi de Benford. Un résumé de ce qui a déjà été fait sur le sujet permet ensuite de faire ressortir l'originalité de l'analyse qui suit et de ses conclusions.

1.1 La loi de Benford et la recherche en comptabilité, contrôle et audit

1.1.1 Formalisation d'une loi issue d'une découverte fortuite

L'objet de ce papier n'étant pas de faire l'historique de la découverte de cette loi, le lecteur est renvoyé au papier de Geyer et Mathieu (2008) et à l'ouvrage de Glück (2008) pour ce point précis. Soulignons tout de même que cette « trouvaille » a été révélée avant même que la communauté des chercheurs ne s'en empare. On pourrait dire que Newcomb (1881) a « trouvé » une solution de manière fortuite en attente d'un problème. Puis, Benford (1938) a retrouvé cette solution et a eu l'idée de l'appliquer à un « tas de problèmes ». Bien plus tard, les chercheurs en comptabilité ont utilisé cette « loi » dans leurs travaux sur les fraudes comptables.

Formellement, comment cette « loi » empirique peut-elle se traduire? À l'origine, Newcomb a proposé une distribution d'utilisation du chiffre c en première position prenant la forme suivante: $\log_{10}(1+1/c)$. Par exemple, cela signifie que le chiffre 5 est utilisé dans une série statistique quelconque suivant la loi de Benford dans près de 8% des cas en première position [$\log_{10}(1+1/5)\approx 0{,}07918$].

C'est bien plus tard que la démonstration mathématique est réalisée par Hill (1995). Comme l'a souligné Newcomb, le premier chiffre d'un nombre sera c_1 avec une probabilité de $\log_{10}(1+1/c_1)$, c_1 (c_1=1,...,9) étant le chiffre le plus à gauche. Selon Hill, si la série étudiée suit une loi de Benford, le second chiffre est c2 avec une probabilité égale à :

$\sum_{c1=1}^{9} \log_{10}(1+\frac{1}{c1c2})$, $c_1\,c_2$ ne signifie pas ici « c_1 multiplié par c_2 », mais représente un nombre composé de c_1 en guise de premier chiffre et c_2 en seconde position (c_2=0,...9).

Au delà de deux chiffres, on peut démontrer par récurrence que la probabilité que le n-ième chiffre soit c_n, est égale à :

$$\sum_{c1=1}^{9}\sum_{c2=0}^{9}\dots\sum_{ci=0}^{9}\dots\sum_{cn=0}^{9}\log_{10}(1+\frac{1}{c1c2\dots ci\dots cn}).$$

Pour les cinq premiers chiffres d'un nombre, la table de la loi de Benford est la suivante. Nous avons limité le tableau à cinq chiffres. Les chiffres en sixième « position » semblent distribués de manière uniforme comme ceux de la cinquième.

Il apparaît que la loi de Benford ne s'apparente pas à la loi uniforme sur la table 1 pour les quatre premières colonnes. Mais, au delà de ce constat, il faudrait questionner la différence significative de ces deux lois, pour chaque position, en fonction de la quantité de nombres entrant dans un test d'adéquation (encadré 1).

Table 1
Loi de Benford pour les cinq premiers chiffres d'un nombre

chiffre/position	Première	Deuxième	Troisième	Quatrième	Cinquième
0	-	0,119679	0,10178	0,1002	0,1
1	0,301030	0,113890	0,10138	0,1001	0,1
2	0,176091	0,108821	0,10097	0,1001	0,1
3	0,124939	0,104330	0,10057	0,1001	0,1
4	0,096910	0,100308	0,10018	0,1	0,1
5	0,079181	0,096677	0,09979	0,1	0,1
6	0,066947	0,093375	0,09940	0,1	0,1
7	0,06	0,090352	0,09902	0,1	0,1
8	0,051153	0,087570	0,09864	0,1	0,1
9	0,045757	0,084997	0,09827	0,1	0,1

Encadré 1.
Loi de Benford, Fraude et taille critique de l'échantillon

Proposition. Si une distribution de chiffres, peu importe la position, n'est pas significativement en adéquation avec la loi de Benford; alors on admettra que la série n'est pas en adéquation à la loi de Benford. Il y a une preuve statistique significative de fraude.
Réciproque. Si toutes les distributions de chiffres sont significativement en adéquation à la loi de Benford, alors on admettra que la série est en adéquation avec la loi de Benford. Il y a absence de preuves statistiques de fraudes, mais il n'y a pas preuve d'absence de fraudes.
Taille de la base (quantité de nombres testés) au-delà de laquelle la différence entre la loi de Benford et la loi uniforme est significative, pour un risque de première espèce de 1% (5%):
-45 pour la première position (resp. 35),
-1798 pour la seconde position (resp. 1405),
-172255 pour la troisième position (resp. 134513),
-30644971 pour la quatrième position (resp. 11800330).

Comme présenté dans le point suivant, l'objet des travaux sur les fraudes comptables utilisant la loi de Benford est de savoir si les données dont on attend un « maquillage » sont effectivement manipulées. Autrement dit, ils cherchent à montrer l'utilité d'une telle loi pour la détection de fraude. En effet, dans le point qui suit, sont présentées des études empiriques prenant comme données des chiffres comptables transmis par des entreprises (Carslaw, 1988; Nigrini, 1999; Thomas, 1989), des contribuables (Nigrini, 1996) ou des employés à l'administration, mais aussi des observations expérimentales (Geyer et Mathieu, 2008). Cela permettra de positionner notre étude par rapport aux travaux antérieurs et de mettre en évidence son « originalité ».

1.1.2 Études réalisées sur la confirmation de l'utilité de la loi de Benford dans la détection de fraudes et enjeux de notre positionnement

Le point commun de toutes les applications antérieures en comptabilité semble résider dans la confirmation du fait que la loi de Benford est un bon moyen pour détecter des fraudes. Dans son étude sur des données néo-zélandaises, Carslaw (1988) postule que les managers ont une propension à arrondir leurs résultats pour atteindre un point de référence du type $Nx10^k$ (où N et k sont des entiers naturels). Il teste cela par le truchement d'une analyse du second chiffre de différents résultats transmis. La tendance des managers à vouloir atteindre ou dépasser un point de référence diminuerait la fréquence d'apparition du chiffre 9 en seconde position d'un résultat comptable.

Pour généraliser ce résultat à d'autres pays, Thomas (1989) a entrepris une démarche similaire en prenant comme source le bénéfice par action et des données trimestrielles comptables tirés de la base de données COMPUSTAT concernant les États-Unis. Le bénéfice par action révèlerait une manipulation flagrante, car un excès de 1 et de 5 et un manque de 9 en seconde position apparaît si l'on prend comme référence la loi de Benford. Ce résultat permet d'étendre aux États-Unis les résultats obtenus par Carslaw (1988) en Nouvelle-Zélande.

Des analyses similaires ont été menées dans différents pays et ont confirmé cette tendance: Finlande (Niskanen et Keloharju, 2000), Japon (Skousen *et al*., 2004), Royaume-Uni (Van Caneghem, 2004). Plus largement, Kinnunen et Koskela (2002) ont cherché le pays champion de la fraude comptable en étudiant les résultats de 22000 entreprises sur 18 pays entre 1995 et 1999. Il semblerait que les entreprises de tous les pays manipulent leurs résultats mais que l'Espagne, Hong-Kong et Singapour le font plus que les autres.

Christian et Gupta (1993) analysent les données des agents payant des impôts aux USA. Ils cherchent à déceler d'éventuelles diminutions du montant imposable en s'intéressant aux deux derniers chiffres des revenus imposables. Ils posent l'hypothèse, cohérente avec la loi de Benford, d'une distribution uniforme des derniers chiffres du revenu imposable en cas d'absence de manipulation. Leur papier conclut qu'il existerait une preuve de manipulations, puisque la distribution des deux derniers chiffres n'est pas uniformément distribuée sur l'intervalle [00; 99]. Or, l'encadré 1 semble montrer que les chiffre en cinquième et sixième positions sont distribués uniformément s'il suivent la loi de Benford et cela significativement si la taille de la base n'est pas très importante.

Dans la même veine, Nigrini (1996) travaille sur des données relatives à la déclaration d'impôt sur le revenu des contribuables. Il trouve alors des différences notables entre les distributions théoriques des sommes perçues par les contribuables et celles qu'ils ont payées et les deux distributions empiriques de ces mêmes sommes. Ce résultat semble cohérent avec la propension des particuliers payant l'impôt à diminuer les sommes reçues et à sur-évaluer celles payées.

Une étude récente plus originale, présentée au Vème colloque Oriane à Bayonne, vise à étudier dans un cadre expérimental, et donc contrôlé, les manipulations des résultats comptables par des étudiants (Geyer et Mathieu, 2008). Il leur était alors demandé en tant que « chef-comptable d'une multinationale » de « falsifier ce bilan en mettant un résultat positif », car le résultat réel était « catastrophique ». Seul le premier chiffre des résultats falsifiés était distribué conformément à la loi de Benford. Les chiffres suivant avaient un excès significatif de 0 (pour une erreur de première espèce de 1%) en comparaison avec la distribution théorique de Benford.

Dans toutes ces études, lorsqu'il y a des fraudes sur un type de données particulières, la sur-pondération significative touche toujours le même chiffre et la sous-pondération touche significativement un autre pour la plupart des séries. Pour tester la qualité d'une base de données, on peut à défaut de trouver une parfaite adéquation à la loi de Benford admettre qu'il n'y a pas de fraude quand les biais significatifs ne touchent pas toujours les mêmes chiffres (cette hypothèse *H3* est précisée ensuite). De plus, dans la littérature, la détection de fraude s'attache à révéler un arrondi se traduisant par une faible présence du chiffre 9 en première et seconde position d'un chiffre comptable et une forte présence du chiffre 1 en première position. Une hypothèse envisageable est qu'une possibilité de fraude détectée par la loi de

Benford prend la forme d'une sur-pondération de 1 en première position, une sous-pondération de 9 significative en première et seconde position et/ou une sur-pondération de zéro en seconde position (cette hypothèse *H2* est détaillée ensuite).

À l'issue de cette brève revue de la littérature, il apparaît qu'il existe, à notre connaissance, un « biais de confirmation » dans les travaux réalisés jusqu'à présent. En effet, il est toujours confirmé que la loi de Benford permettrait de déceler des fraudes comptables ou fiscales en cas de non adéquation de la distribution des séries étudiées avec cette dernière. Certains articles s'en tiennent à mettre en garde contre le risque de première espèce (Cleary et Thibodeau, 2005) ou sur le risque du choix d'un mauvais logiciel pour vérifier l'adéquation de chiffres comptables à la loi de Benford (Debreceny *et al.*, 2005; Ettredge et Srivastava, 1999; Weidenmier *et al.*, 2004).

Pour aider les praticiens utilisateurs ou désirant utiliser cet outil, il peut être, à présent, intéressant d'expliquer les limites de l'utilisation de cette loi. La limite principale mise en avant dans la littérature est celle portant sur l'inefficacité de cet outil dans le cas où une série étudiée comporte une limite minimale ou maximale. Geyer et Mathieu (2008) proposent un exemple fondamental et « clarificateur » : le cas des remboursements de frais de repas. Si une université rembourse au maximum 30€ de « frais de bouche » par jour à un congressiste, alors il y aura plus de remboursements inférieurs et égaux à 30€ que de montants supérieurs à 30€. Au delà de cette limite triviale, n'existe-il pas des limites sectorielles?

Table 2
Utilité de la loi de Benford pour détecter des fraudes

Quand la loi de Benford semble utile	**Exemples**
Ensemble de nombres qui résultent de combinaison mathématique d'autres nombres- résultat venant de deux distributions	Chiffres d'affaires (ventes*prix) Achats de matières premières et marchandises (achats*prix)
Données sur le volume des transactions	Décaissements, ***ventes***, dépenses
Ensemble de données « large » -plus on a de données, plus l'analyse est pertinente	Transactions durant un exercice comptable
Comptes qui semblent s'y conformer : quand la moyenne de la série est supérieure à la médiane et la *skewness* (coefficient d'asymétrie) est positive	La plupart des séries de nombres comptables
Quand la loi de Benford semble inutile	**Exemples**
Séries de données constituées de nombres attribués arbitrairement	Numéros de chèque, de facture, code postal
Nombres fixés pour influencer la « pensée » humaine	Prix fixés à des seuils psychologiques (1,99$), retraits aux guichets automatiques bancaires (20, 40…euros)
Comptes où transitent des gros montants à destinations d'un nombre précis d'entreprises	Un compte créé spécifiquement pour enregistrer une opération de refinancement de 100$
Comptes avec une borne minimum et maximum	Ensemble d'actifs qui doivent dépasser un seuil pour être enregistré
Là où aucune transaction n'est enregistrée	Vols, pots-de-vin, « contrat à l'amiable »

Source: Durtschi, Hillison et Pacini (2004, p.24)

De manière plus systématique, Durtschi *et al.* (2004) recensent les séries où la loi de Benford pourrait s'appliquer et où elle serait inefficace pour détecter la présence de fraudes. Ils ont effectué cet état des lieux en parcourant la littérature sur le sujet (table 2).

A priori, d'après ce tableau, la loi de distribution des ventes, lorsqu'elles sont exprimées en volume et non en valeur (car il y a le risque de modification par des prix psychologiques:1,99€ par exemple), pourrait être en adéquation avec la distribution exprimée par la loi de Benford. De plus, pour cela, la médiane doit être inférieure à la moyenne et la *skewness* doit être positive. Dans la partie suivante, nous montrons, en théorie, que ce n'est pas forcément le cas pour certains secteurs d'activité avant de l'illustrer avec un cas empirique.

1.2 Présentation et justification de l'échantillon et de la méthode

1.2.1 Choix du secteur d'activité et description de notre base de données

Dans certains secteurs, des limites à l'application de la loi de Benford peuvent exister. C'est le cas des secteurs où les ventes ou achats des entreprises peuvent être le résultat d'une dynamique non linéaire du fait d'externalités dans le comportement des agents. En effet, Tolle *et al*. (2000) partent de simulations pour montrer que certains systèmes non-linéaires peuvent faire sortir des séries ne suivant pas la loi de Benford. Ces résultats expérimentaux sont confirmés, formellement par Berger *et al*. (2004) et Berger (2005). Ils arrivent au résultat que les séries découlant de processus linéaires convergeant donnent des chiffres s'approchant de la loi de Benford, mais pour les processus non linéaires ils admettent qu'« il peut y avoir des exceptions ».

Or, les achats de biens innovants à la mode pourraient présenter un processus non linéaire : les acheteurs prendraient leur décision en fonction du comportement des autres (Dosi et Metcalfe, 1991). Cette idée est formalisée par Granovetter et Soong (1986) dans le cadre de modèles à seuil et par Nakayama et Nakamura (2004) dans le cadre d'un modèle de physique statistique du comportement des consommateurs. Sans entrer dans le détail de ces modèles, le seul fait qu'il puisse y avoir un processus non linéaire sous-jacent devrait, en théorie, suffire pour que les ventes réelles ne suivent pas toujours la loi de Benford. C'est ce que nous tentons de montrer avec des séries de ventes de biens innovants à la mode. L'hypothèse associée à cette possibilité est la suivante:

H1: les ventes de consoles de jeux vidéo ne suivent pas toutes la loi de Benford.

Dans la littérature, le secteur des nouvelles technologies est cité pour être un secteur engendrant des dynamiques de type « bandwagon effects », propres aux phénomènes de mode (Rohlfs, 2001). Cet effet pourrait être renforcé pour le secteur des jeux vidéo, car impulsé par des effets de réseau associés aux nouvelles technologies (Crandall et Sidak, 2006). De plus, il est possible de légitimer le choix de ce secteur par l'existence d'une base de données possédant des séries de ventes longues : « un ensemble de données large- plus il y a de données, plus l'analyse est pertinente » (Durtschi *et al.*, 2004, p.24).

Mais, comment s'assurer de la propreté des données utilisées ? Autrement dit, comment s'assurer que des séries de données n'étant pas en adéquation avec la loi de Benford n'ont pas été manipulées ?

L'analyse de bases de données permettrait d'avoir une démarche d'investigation qui fasse consensus, puisque l'ensemble des études réalisées précédemment ont utilisé des bases de données d'entreprises ou publiques, à l'exception de Geyer et Mathieu (2008). De plus, compte tenu de la forme de notre question de recherche, de la nécessité d'utiliser des données passées et du faible contrôle que nous possédons sur les ventes de produits à la mode, il serait préférable, selon Yin (2003, p.3-8), de procéder à une analyse utilisant une base de données numérique plutôt qu'une étude de cas.

Nous avons opté pour une base de données où les ventes enregistrées sont triangulées pour améliorer leur fiabilité. Cela permet d'éviter, autant que faire se peut, de travailler avec des données truquées. Nous nous sommes assurés auprès de l'organisme qui regroupe les données relatives aux ventes de la qualité de ces données : dans un premier temps, un décompte est réalisé chaque fin de semaine par un sondage auprès de commerçants sélectionnés aléatoirement dans chaque pays. Dans un second temps, il s'assure que ces montants correspondent aux estimations des produits envoyés par les entreprises productrices à ces commerçants sondés.

Cette base comprend les ventes hebdomadaires de consoles de jeux vidéo d'avril 1989 à mai 2009 au Japon et de novembre 2000 à mai 2009 pour tous les autres pays. La suite de ce papier se concentrera exclusivement sur les données américaines, japonaises, britanniques, allemandes et françaises pour des questions pratiques et de clarté. Ce choix n'a pas diminué, pour autant, la portée de nos résultats puisqu'ils reposent sur suffisamment de données : les données de cinq pays et de neuf consoles de jeux par pays (et vingt pour le Japon). Avant d'aller plus loin dans l'explication de notre méthode, notons que le nombre de semaines de mise en vente est très variable suivant les consoles et les pays. Pour prendre un exemple, la X-Box ne s'est vendue que pendant 195 semaines au Japon contre 299 aux USA.

Le fait qu'il s'agisse de séries de ventes longues permet, d'après le tableau de Durtschi, Hillison et Pacini (2004, p. 24), de penser qu'il pourrait y avoir une adéquation significative par rapport à la loi de Benford. Mais en plus de ce critère, il faudrait aussi que la moyenne dépasse la médiane pour toutes les séries étudiées et que la *skewness* soit positive (distribution étalée à droite). Les tableaux suivants (tables 3) permettent de s'assurer que nos ventes respectent bien ces deux conditions: pour toutes les consoles, la moyenne dépasse la médiane et la *skewness* est positive.

Tables 3

Statistiques descriptives des séries de ventes classées par pays

Japon	Wii	PS3	X360	PSP	DS	SC	XB	GC	GBA	WSC	PS2	WS	PockSt	DC	NGP	N64	PS	SAT	SNES	GB
mediane	50038	17765	3712	41766	84662	761	802	5279	33601	2550	35773	4631	12759	6608	1744	9169	19952	16509	30077	30034
moyenne	64402	24239	5934	53559	116761	2191	2450	12307	46916	5654	48158	8752	26822	12302	2540	17707	33413	24543	34436	39323
skewness	2,98	1,56	3,58	2,30	1,97	5,59	11,79	4,37	4,17	8,11	6,19	4,85	1,45	3,20	4,40	5,08	2,28	2,62	2,04	4,15

Allemagne	**Wii**	**PS3**	**X360**	**PSP**	**DS**	**XB**	**GC**	**GBA**	**PS2**
mediane	17415	9655	4279	7178	19984	2281	2500	6498	9378
moyenne	19570	10116	5477	9001	24399	3071	3713	9352	13382
skewness	2,42	1,95	2,43	3,26	2,16	2,74	4,36	2,30	2,78
USA	**Wii**	**PS3**	**X360**	**PSP**	**DS**	**XB**	**GC**	**GBA**	**PS2**
médiane	135296	54637	65459	59348	124557	38326	23330	76540	77205
moyenne	181685	66555	96745	78551	149649	53467	38148	115194	113432
skewness	3,09	1,91	2,81	2,95	2,86	3,66	3,83	2,93	3,70
Roy-Uni	**Wii**	**PS3**	**X360**	**PSP**	**DS**	**XB**	**GC**	**GBA**	**PS2**
mediane	40023	14213	13644	12377	37067	5847	2632	10193	16100
moyenne	43313	18250	19045	17281	43828	8823	4685	14876	23074
skewness	1,56	4,86	2,47	5,17	2,37	2,64	4,98	1,99	3,33

France	Wii	PS3	X360	PSP	DS	XB	GC	GBA	PS2
mediane	24444	10869	5333	7915	28026	2930	3621	7045	9852
moyenne	27139	12537	7575	11439	33114	3599	5111	9931	13755
skewness	1,98	3,04	2,70	5,88	2,40	2,92	3,29	2,42	3,23

Wii est la dernière console de Salon de Nintendo, PS3=Playstation 3 (Sony), X360 (Microsoft), PSP=Playstation Portable (Sony), DS= Dual Screen (Nintendo), GB=Game Boy (Nintendo), SNES=Super Nintendo Entertainment System (Nintendo), SAT= Sega Saturne (Sega), PS=Play Station (Sony), N64=Nintendo 64 (Nintendo), NGP=Neo Geo Pocket (constructeur japonais SNK), DC=Dreamcast (Sega), PockSt=PocketStation (Sony), WS=WonderSwan (Bandaï), PS2=PlayStation2 (Sony), WSC=WonderSwan Color (Bandaï), GBA=Game Boy Advanced (Nintendo), GC=Game Cube (Nintendo), XB=Xbox (Microsoft), SC=Swan Color (Bandaï). La skewness ou coefficient d'asymétrie est le rapport du moment centrée d'ordre trois au cube de l'écart type. La moyenne est un moment non centré d'ordre un, la médiane est une caractéristique de forme divisant en deux parties égales la distribution des ventes.

La suite du papier expose précisément le canevas d'investigation permettant de répondre à la question de recherche.

1.2.2 Éléments de méthodes

En rupture avec le consensus scientifique de son époque, Karl Popper a établi qu'il ne suffisait que d'un seul cas pour montrer qu'une théorie ou une loi pouvait être fausse. Pour renforcer la validité d'un tel résultat, il nous semble prudent de trianguler les données et les méthodes. Par exemple, lorsqu'on obtient un résultat surprenant en physique, il convient de refaire les calculs plusieurs fois avant de l'annoncer à la communauté scientifique. De même, en astronomie, quand on observe un élément nouveau (exo-planète, présence d'eau sur une autre planète...), il est souhaitable, dans la mesure du possible, de changer la lunette ou l'observateur pour avoir confirmation de cette « découverte » (Chalmers, 1987).

La méthode de notre étude a consisté à tester l'adéquation des séries de ventes de consoles de jeux vidéo par rapport à la loi de Benford au moyen d'une analyse globale de la distribution grâce à des tests du chi-deux d'adéquation pour chaque chiffre (premier, deuxième..., cinquième), pour chaque série de ventes (Game Boy, ...,Wii) dans chacun des pays retenus. Par ailleurs, la Z-statistique de Fleiss (198, p.13) a été utilisée pour chercher d'éventuelles sur-pondérations significatives d'un ou plusieurs chiffre(s) par rapport aux autres chiffres.

Ces deux techniques sont complémentaires pour valider nos hypothèses.

La première, le test du chi-deux, permet de s'assurer que les distributions des chiffres des séries étudiées peuvent ne pas être en adéquation avec la loi de Benford. Ces tests d'adéquation sur nos données nous ont permis de tester l'hypothèse H1 pour chaque série. H1 peut être formalisée au regard de ce test de la façon suivante:

H1: $D>\chi^2_{1-\alpha}(n-1)$, avec D la statistique de test du chi-deux d'adéquation calculée et $\chi^2_{1-\alpha}(n-1)$ le fractile d'ordre 1-α de la loi du chi-deux à n-1 degré de liberté, n le nombre de chiffres possibles sur cette position (9 en première position et 10 en seconde), α l'erreur de première espèce.

La deuxième, le calcul de la Z-statistique associée à l'analyse du signe du biais, a permis de s'assurer qu'il existait, premièrement, des cas où un investigateur analysant, par exemple, la distribution du second chiffre pouvait trouver des résultats s'apparentant à une fraude (sur-proportion de 0 ou de 5 et sous-représentation des 9, comme dans l'étude de Thomas (1989) et Carslaw (1988)). Deuxièmement, ce test a permis de montrer que la base n'était pas trafiquée car les biais révélés par rapport à la loi de Benford ne concernaient pas toujours les mêmes chiffres. En effet, si nous avions toujours eu en seconde position, par exemple, une sur-pondération de 0 au dépend des 9, illustrant par la même une inadéquation à la loi de Benford, alors nous aurions pu mettre en doute la fiabilité de notre base.

Somme toute, cette Z-statistique permettra de tester deux hypothèses:

H2: Il existe des possibilités que la loi de Benford tranche en faveur d'une présence de fraudes s'il y a sur-pondération de 1 et sous-pondération de 9.

H3: Nos données ne présentent pas les mêmes distorsions (ce qui serait caractéristique d'une base de données non frauduleuses).

Plus formellement au regard de la statistique Z, H2 et H3 peuvent être présentées de la façon suivante:

H2: L'ensemble des cas de sur-pondération significatif de chiffre 1 en première position et de chiffre 0 en seconde position et de sous pondération de 9 en première et deuxième position est non vide. Soit $\text{Prob}[(Z(1,+,1)>z_{1-\alpha})\cap(Z(2,+,0)>z_{1-\alpha})\cap(Z(1,-,9)>z_{1-\alpha})\cap(Z(2,-,9)>z_{1-\alpha})]\neq 0$ avec Z(i,j,k) la statistique Z associé au chiffre k à la i-ème position avec un biais de signe j={+,-} et $z_{1-\alpha}$ étant le fractile d'ordre 1-α de la loi normale centrée réduite.

H3: $\text{Prob}[(Z(1,+,1)> z_{1-\alpha})\cup(Z(1,-,9)> z_{1-\alpha})]\neq 1$ et $\text{Prob}[(Z(2,+,0)>z_{1-\alpha})\cup(Z(2,-,9)>z_{1-\alpha})]\neq 1$; Cela signifie que tous les biais significatifs ne sont pas concentrés sur des cas de fraudes *a priori* et qu'ils sont répartis aléatoirement entre les chiffres d'une position donnée.

Ces précisions méthodologiques ayant été apportées, nous présentons maintenant les résultats. Une discussion sur les limites de ces « découvertes » suivra ainsi que des pistes de recherches futures à mener pour poursuivre ce travail de délimitation en creux du domaine d'applicabilité de la loi de Benford afin de détecter des fraudes.

2 Résultats et discussions autour de l'applicabilité de la loi de Benford

Nous présenterons, tout d'abord, nos principaux résultats. Leur présentation sera particulièrement synthétique: pour des raisons de clarté et de lisibilité, nous ne ferons apparaître qu'une sélection de tableaux de calculs de statistique du chi-deux pour certains pays et certaines consoles. Nous ferons de même pour les calculs de la statistique Z. Puis ces résultats seront discutés et des pistes de recherches futures exposées.

2.1 Les ventes réelles peuvent ou non être en adéquation avec la loi de Benford

Concernant l'adéquation de nos séries à la loi de Benford, il nous a semblé particulièrement utile de présenter les statistiques du chi-deux calculées résultant de chaque test d'adéquation à la loi de Benford. Nos résultats pour chaque pays apparaissent dans les tables 4.
Ces tableaux de résultats font apparaître en italique (et en gras) les statistiques du chi-deux indiquant une bonne adéquation des données pour la position (du nombre de ventes hebdomadaires en volume) indiquée à la loi de Benford pour une erreur de première espèce de 1% (respectivement 5%).

Tables 4

Résultats de tests d'adéquation par pays, console et position

Tableau de résultats de tests d'adéquation, USA					
console/position	première	deuxième	troisième	quatrième	cinquième
PS2	212,05	***10,68***	***1,32***	***7,86***	***10,97***
GBA	40,47	***9,71***	***1,54***	***4,17***	***11,96***
GC	45,94	***3,52***	***0,26***	***5,28***	***2,83***
XB	79,7	***4,22***	***0,05***	***13,52***	***7,74***
DS	51,67	19,94	***0,28***	***4,74***	***9,38***
PSP	192,9	***6,12***	***8,32***	***11,97***	***7,43***
X360	126,95	***5,01***	***0,75***	***7,14***	***3,17***
PS3	62,17	*18,01*	***0,46***	***1,44***	22,25
wii	127,59	22,93	***2,99***	***14,13***	***13,52***
	Ddl=7			Ddl=8	
	X²(5%)=14,07			X²(5%)=15,51	
	X²(1%)=18,48			X²(1%)=20,09	

Tableau de résultats de tests d'adéquation, Royaume Uni					
console/position	première	deuxième	troisième	quatrième	cinquième
PS2	158,2	22,81	***11,51***	***7,32***	***3***
GBA	***13,09***	67,07	42,04	23,65	***5,19***
GC	***1,81***	***7,28***	***9,51***	***14,9***	21,29
XB	108,87	***6,23***	***8,75***	***8,06***	***3***
DS	68,51	***8,1***	***7,21***	***14,25***	*18,38*
PSP	109,11	20,89	***8,02***	***10,52***	***1,66***
X360	49,32	***7,58***	***8,08***	***5,29***	***12,49***
PS3	38,54	***14,44***	***6,93***	***10,77***	***9***
Wii	58,79	***10,3***	***7,38***	***7,86***	*16,04*
	Ddl=7			Ddl=8	
	X²(5%)=14,07			X²(5%)=15,51	
	X²(1%)=18,48			X²(1%)=20,09	

tableau de résultats de tests d'adéquation, Japon					
console/position	première	deuxième	troisième	quatrième	cinquième
GB	206,31	156,24	159,01	115,4	70,23
SNES	51,01	***11,02***	138,04	***5,83***	***10,78***
SAT	***11,57***	***10,31***	***4,11***	***12,05***	***8,58***
PS	44,3	34,34	35,61	97,36	143,62
N64	18,6	***9,98***	***2,79***	***8,29***	***9,24***
NGP	19,03	***11,63***	***9,71***	***4,03***	***11***
DC	***12,08***	***6,53***	***10,32***	***6,75***	22,56
PockSt	***7,36***	***12,18***	***9,48***	***8,32***	***7,21***
WS	***3,47***	***5,02***	***5,08***	***4,74***	***6,75***
PS2	32,11	***5,27***	***9,4***	***3,91***	***5,57***
WSC	21,06	***7,48***	***9,64***	*16,56*	***8***
GBA	34,78	***11,65***	***10,48***	***11,21***	***10,31***
GC	***9,35***	*18,77*	***11,49***	***5,47***	*19,89*
XB	25,59	***9,37***	***8,52***	***8,64***	***7***
SC	***10,38***	*16,87*	***6,57***	***11,84***	***8***
DS	19,34	***4,68***	***13,54***	*17,93*	***13,73***
PSP	71,64	***4,13***	***13,77***	***7,89***	***2,51***
X360	***6,48***	***5,42***	***8,44***	***5,65***	***5***
PS3	*14,53*	***8,79***	***4,62***	***8,02***	***14,4***
Wii	30,82	***7,01***	***7,37***	***8,42***	***7,45***
	Ddl=7			Ddl=8	
	X²(5%)=14,07			X²(5%)=15,51	
	X²(1%)=18,48			X²(1%)=20,09	

Tableau de résultats de tests d'adéquation, France					
console/position	première	deuxième	troisième	quatrième	cinquième
PS2	146,4	***12,94***	***7,82***	***9,73***	***7,72***
GBA	*17,07*	20,52	***9,88***	***2,79***	***3,68***
GC	58,82	***5,79***	***3,44***	*16,31*	***10,67***
XB	60,22	***7,63***	***4,57***	***2,64***	*17,75*
DS	43,52	***8,75***	***6,22***	***6,45***	***4,04***
PSP	93,29	***8,79***	***21,2***	***6,36***	***9,57***
X360	21,12	***8,05***	***9,1***	***4,64***	***8,4***
PS3	***13,43***	***14,65***	***9,26***	***10,02***	***7,67***
Wii	64,85	***5,64***	***10,07***	*15,65*	***4,71***
	Ddl=7 X²(5%)=14,07 X²(1%)=18,48		Ddl=8 X²(5%)=15,51 X²(1%)=20,09		

tableau de résultat Allemagne					
console/position	première	deuxième	troisième	quatrième	cinquième
Wii	***13,74***	***10,71***	***13,94***	***9,71***	*19,5*
PS3	18,97	24,74	***6,63***	***7,81***	***8,22***
X360	*17,22*	***15,28***	***9,27***	***8,82***	***4,2***
PSP	98,66	***11,99***	***9,55***	***11,04***	***5,27***
DS	45,11	***4,95***	***7,88***	***7,42***	***4,16***
XB	75,6	***13,83***	***8,63***	***8,48***	***6***
GC	29,13	21,06	***13,95***	***11,87***	***8,29***
GBA	20,9	***8,55***	***8,2***	***4,59***	***12,89***
PS2	250,79	20,11	***5,92***	***5,46***	***8,99***
	Ddl=7 X²(5%)=14,07 X²(1%)=18,48		Ddl=8 X²(5%)=15,51 X²(1%)=20,09		

X²(5%) est le fractile d'ordre 0,05 de la loi du chi-deux. On retient pour la première position un degré de liberté de 8 (9-1) et pour les autres positions un degré de liberté de 9 (10-1).

Il apparaît que les cas de vente en conformité avec la loi de Benford sont extrêmement rares. Il existe, par exemple, le cas des ventes de « Sega Saturne » (SAT) au Japon dont tous les chiffres suivent une distribution significativement en adéquation avec la loi de Benford, pour une erreur de première espèce de 5% ou 1%. Il n'existe que deux séries (ventes hebdomadaires de Game Boy et Playstation au Japon, GB-Japon et PS-Japon) où il résulte une non adéquation significative à la loi de Benford pour toutes de positions des chiffres composants les ventes hebdomadaires, pour une erreur de première espèce de 1%. Globalement, on s'aperçoit que bien souvent c'est le premier chiffre des séries de ventes par semaine de console qui ne semble pas significativement distribué conformément à la loi de Benford (dans 26 cas sur 56 contre 11 sur 56 pour le second chiffre, 4 pour le troisième chiffre, 3 et 5 pour les deux derniers chiffres pour un risque de première espèce de 5%).

Concernant notre hypothèse H1, au regard de ces tableaux, il est possible d'admettre que les ventes de console de jeux vidéos ne suivent pas toujours significativement une loi de Benford, même s'il existe deux cas pour lesquels la série de ventes est significativement en adéquation avec cette loi. En principe (encadré 1), on pourrait admettre la preuve statistique d'une présence de fraudes. Est-ce possible avec des sources de données triangulées?

Mais, cela ne serait-il pas dû à la présence de manipulations comptables dans les séries? Même s'il y a triangulation, il convient de tester cette éventualité en regardant par exemple la distribution du premier chiffre et du second chiffre pour quelques pays de notre base. Ne pouvant pour un impératif de concision présenter tous nos tableaux, nous allons regarder les biais dans les distributions des deux premiers chiffres (cf. tables 5 en annexe). Ce choix peut se justifier par le fait que, dans l'analyse précédente, il est apparu que la distribution du premier chiffre était significativement différente de la loi de Benford pour de nombreux cas. De plus, pour déceler d'éventuelles fraudes, la littérature montre qu'il est particulièrement intéressant de se focaliser sur les deux premiers chiffres pour découvrir d'éventuels arrondis.

Nous avons limité la présentation des résultats aux pays occidentaux pour ce qui concerne l'analyse des biais en faveur d'un chiffre et son degré de significativité. Les résultats sont exposés dans les deux tableaux suivants (tables 5 en annexe et table 6 page suivante). Nous indiquons les biais de sur-pondération par (+) ou de sous-pondération par (-) du chiffre en question, sur la série étudiée. Les biais significatif pour une erreur de première espèce de 5% (1%) apparaissent en italique (et en gras).

On voit dans les tableaux, en annexe, qu'il n'apparaît aucun biais systématique pour toutes les séries que ce soit pour la distribution du chiffre en première position (position 1) ou en seconde position (position 2). Cela semble valable pour les quatre pays pour lesquels les résultats sont présentés. Ainsi, il semblerait *a priori* qu'il n'y ait pas de fraudes dans nos séries puisque si manipulation frauduleuse il y avait, elle se traduirait par une distribution similaire des biais entre les séries étudiées. De plus, l'apparition de biais significatifs (en gras et/ou en italique) semble être totalement aléatoire. Ainsi, il apparaît des cas sans biais significatifs (Game Cube Royaume-Uni pour les deux premiers chiffres, par exemple).

Partant du résultat d'une distribution sans biais systématiques en faveur de certains chiffres, nous acceptons l'hypothèse 3: tous les biais significatifs ne sont pas concentrés sur des cas de fraudes *a priori* et ils sont répartis aléatoirement entre les chiffres d'une position donnée.

Ce résultat, d'une éventuelle distribution aléatoire du signe du biais et du degré de significativité qui lui est associé, fait apparaître une nouvelle possibilité: une non-adéquation à la loi de Benford mais sans qu'il y ait manipulation comptable en réalité.

En effet, bien que nous n'ayons pas constaté dans notre échantillon une série ayant l'apparence d'une série frauduleuse (chiffre 0 en seconde position sur-pondéré et chiffre 9 sous-pondéré significativement), ces possibilités de fraudes pourraient apparaître compte tenu de l'aléa concernant les biais et leur significativité présents dans les résultats. Dans notre échantillon, il arrive que le chiffre zéro soit significativement sur-pondéré et le chiffre 9 sous-pondéré non significativement en seconde position. C'est le cas des ventes de PSP (PlayStation Portable) en Allemagne. Une autre série de vente présente en seconde position le chiffre 9 sous-pondéré significativement et le chiffre zéro sur-pondéré significativement. C'est le cas des ventes de Game Boy Advance (GBA) en Allemagne.

Des résultats similaires sont apparus lors de l'étude des distributions Japonaises (table 6).

Table 6

Mauvaises détections de fraude avec la loi de Benford sur les données Japonaises

Japon	N64		PockSt		NGP	
chiffre2	biais	Z	biais	Z	biais	Z
0	+	***2,19***	+	***2,08***	+	0,2
1	+	0,07	+	1,72	-	0
2	-	0,01	+	0,25	+	0,06
3	+	0,71	-	0,56	-	0,99
4	-	2,1	-	1,42	-	0,57
5	-	0,08	+	0,54	+	0,14
6	-	0,08	-	0,63	+	1,13
7	+	0,01	-	0,5	+	1,27
8	+	0,38	+	0,03	+	1,41
9	-	0,27	-	0,81	-	***2,03***

Finalement, bien que les données ne semblent pas trafiquées (acceptation H3), il semble alors fortement probable que l'on puisse avoir des séries où un test d'adéquation de la loi de Benford puisse nous faire trancher en faveur de la présence de fraudes. Cependant, cela ne permet pas de valider clairement l'hypothèse 2 en attendant de trouver une série ayant tous les « stigmates » d'une fraude sans être manipulée en réalité.

Avant de passer à la conclusion, nous discuterons ce travail et mettrons en avant de possibles recherches futures permettant de généraliser et de mieux cerner les conditions d'applicabilité de la loi de Benford.

2.2 Limites et pistes de recherches futures

Nos résultats ont montré que l'on pouvait avoir des ventes ne suivant pas la loi de Benford sans fraudes sous-jacentes. La non-adéquation avec la loi de Benford n'est pas suffisante pour qu'il y ait fraude avérée. En effet, pour qu'une fraude soit avérée, il faudrait que les distributions des chiffres soient proches de celles de fraudes couramment observées dans la littérature (deuxième chiffre présentant peu de 9 et beaucoup de 0...). Cependant, nos données laissent penser qu'un tel cas serait possible même si elles n'en donnent pas un en particulier.

De plus, nous avons, certes, testés nos hypothèses sur un panel large comportant beaucoup de produits dans un certain nombre de pays, mais ces données portent sur un seul secteur précis. Ainsi, la généralisation de nos résultats à tous les produits à la mode s'avère impossible.

Par voie de conséquence, il serait intéressant de prendre de nombreuses séries dans d'autres secteurs comme celui de l'habillement, ou d'autres produits à la mode connaissant un large succès (vente de PC, téléphone mobile...) pour connaître les possibilités de généralisation de nos résultats. Cela permettra de savoir si la possible inefficacité de la loi de Benford pour repérer des fraudes se retrouve pour toutes les ventes de biens à la mode ou seulement pour ceux innovants.

Au delà de cette possibilité pour les biens à la mode, il existe dans la littérature économique et « managériale » d'autres types de produits dont les ventes pourraient avoir, en théorie, une dynamique non-linéaire. Par exemple, pour définir en creux les secteurs où l'on peut appliquer la loi de Benford pour repérer d'éventuelles fraudes, il serait bienvenu, dans la ligné de notre travail, de regarder du côté des ventes de produits entraînant une addiction. Selon Feichtinger *et al*. (1995) la vente de ces biens peut avoir une dynamique non-linéaire complexe.

La posture épistémologique annoncée en début d’article peut aussi poser problème : peut-on tester avec des données empiriques un « modèle » statistique ? Premièrement, suivant Dufour (2000), un modèle statistique est logiquement infalsifiable : en se basant sur des hypothèses probabilistes, rien n’est impossible. Il existe juste des occurrences plus probables. Secondement, les tests d’hypothèses reposent sur des données non-expérimentales : elles sont issues d’observations non-contrôlées. Le problème de Duhem-Quine de la présence d’hypothèses d’auxiliaires dans les résultats est donc présent : 1) Pour appliquer notamment un test du chi-deux, il faut supposer que la base de données vérifie certaines hypothèses théoriques, 2) Pour interpréter les résultats, il a été notamment supposé que les ventes de biens innovants à la mode étudiées sont chaotiques. Doit-on alors admettre que « la seule méthode qui nous assure de ne pas se tromper, c’est celle [qui] consiste à ne rien dire » (Dufour, 2000, p.8) ?

3 Conclusion

Finalement, après une définition formelle de la loi de Benford et un bref état de l'art de son application en comptabilité en matière de détection des fraudes, la méthode de recherche a été présentée et justifiée: l'utilisation des statistiques du chi-deux d'adéquation et des Z-statistiques pour chaque série de données de ventes de biens innovants à la mode dans plusieurs pays. Les résultats attestent qu'il serait possible que des ventes non frauduleuses dans le secteur des consoles de jeux vidéos ne suivent pas la loi de Benford. Cependant, nous n'avons pas montré un cas précis s'apparentant parfaitement à une fraude selon le test d'adéquation à la loi Benford (sur-pondération du chiffre 1 et sous pondération du 9e en première position et sur pondération de 0 et sous pondération de 9 en seconde position) sans en être une en réalité. Des recherches futures pourraient détecter une telle série.

Ainsi, dans les années à venir, il serait bienvenu de chercher à répliquer ces conclusions pour savoir si elles sont valables pour toutes les ventes de produits à la mode et non pour quelques secteurs innovants particuliers. Une telle investigation pourrait aussi être entreprise pour les ventes de produits présentant des risques d'addictions (café, thé, cigarette, alcool...).

Les résultats étant établis sur des ventes en volume, il serait intéressant de voir si ces conclusions tiennent sur des séries simulées et sur des données réelles de ventes en valeur.

Références

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the Americal Philosophical Society* 78(4):551-572.

Berger, A. (2005). Multi-dimensional dynamical systems and Benford’s law. *Discrete and Continuous Dynamical Systems* 13(1):219-237.

Berger, A., Bunimovitch, L., Hill, T. (2004). One-dimensional dynamical systems and Benford’s Law. *American Mathematical Society* 357(1): 197-219.

Bolton, R.J., Hand, D.J. (2002). Statistical fraud detection: A review. *Statistical Science* 17(3): 235-249

Bonache, A.B., Moris, K., Maurice, J. (2010). Détection de fraudes et loi de Benford : quelques risques associés. *Revue française de comptabilité* 431 :2-5.

Carslaw, C.(1988). Anomalies in income numbers: Evidence of goal-oriented behavior. *The Accounting Review* 63(2): 321-327.

Chalmers, A.F. (1987). *Qu'est ce que la science? Popper, Kuhn, Lakatos, Feyerabend.* Paris: La Découverte Livre Poche.

Christian, C., Gupta, S. (1993). New evidence on ‘secondary evasion’. *The Journal of the American Taxation Association* 15(1): 72-93

Cleary, R., Thibodeau, J. (2005). Applying digital analysis using Benford's law to detect fraud: the dangers of type I errors. *Auditing: A Journal of Practice and Theory* 24(1): 77-81

Cohen, M., March, J., Olsen, J. (1972). A garbage can model of organizational choice. *Administrative science quarterly* 17(1):1-25

Crandall, R.W., Sidak, J.G. (2006). Video games: Serious business for America’s economy. *Entertainment Software Association*.

Debreceny, R., Lee, S., Neo. W., Toh, J. (2005). Employing generalized audit software in the financial services sector. *Managerial Auditing Journal* 20(6): 605-618

Dosi, G., Metcalfe, J. (1991), Approaches de l'irréversibilité en théorie économique. In *Les Figures de Irréversibilité en Economie* (Eds, Chavrence, B., Godard, O.). Paris: EHESS.

Dufour, J.M. (2000). Econométrie, théorie des tests et philosophie des sciences. *Académie des lettres et des sciences humaines de la Société royale du Canada*. Montréal. septembre 2000.

Durtschi, C., Hillison, W., Pacini, C. (2004), The effective use of Benford’s law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting* 5(1):17-43.

Ettredge, M., Srivastava, R. (1999). Using digital analysis to enhance data integrity. *Issues in Accounting Education* 14(4):675-690.

Feichtinger, G., Prskawetz, A., Herold, W., Zinner P. (1995). Habit formation with threshold adjustment: addiction may imply complex dynamics. *Journal of Evolutionary Economics* 5(2):157-172.

Fleiss J. (1981), *Statistical Methods for rates and proportions.* New York: Willey.

Geyer, D., Mathieu, J.-P. (2008). La détection des risques de fraude comptable par la loi de Benford. In *Méthodes et thématiques pour la gestion des risques* (Ed., Guillon, B.), Paris:L'Harmattan.

Glück, M. (2008), *Die Benford-verteilung-Anwendung auf reale Daten der Marktforschung.* GRIN Verlag

Granovetter, M., Soong, R. (1986). Threshold models of interpersonal effects in consumer demand. *Journal of Economic Behavior and Organization* 7(1): 83-99.

Hill, T. (1995). A statistical derivation of the significant-digit law, *Statistical Science* 10(4): 354-363.

Kinnunen, J., Koskela, M. (2002). Who is Miss World in cosmetic earnings management? A cross-national comparison of small upward rounding of net income numbers among eighteen countries. *Journal of International Accounting Research* 2(1):39-68.

Miller, C., McIntyre, S., Mantrala, M. (1993). Toward formalizing fashion theory. *Journal of Marketing Research* XXX:142-157.

Nakayama, S., Nakamura, Y. (2004). A fashion model with social interaction. *Physica A: Statistical Mechanics and its Applications* 337(3):625-634.

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics* 4(1): 39-40.

Nigrini, M. (1996). A taxpayer compliance application of Benford’s law. *Journal of the American taxation Association* 18(1):72-91.

Niskanen, J., Keloharju. M. (2000). Earnings cosmetics in a tax-driven accounting environment: evidence from Finnish public firms. *European Accounting Revie*w 9(3): 443-452.

Rohlfs, J.H. (2001). *Bandwagon effects in high technology industries*. MIT Press, Cambridge.

Skousen. C., Guan. L., Wetzel. T. (2004). Anomalies and unusual patterns in reported earnings: Japanese managers round earnings. *Journal of International Financial Management and Accounting* 15(3):212-234.

Thomas. J. (1989). Unusual patterns in reported earnings. *The Accounting Review* 64(4): 773-787.

Tolle. C., Budzien. J., LaViolette. R. (2000). Do dynamical systems follow Benford’s law?. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 10(2): 331-336.

Van Caneghem. T. (2004). The impact of audit quality on earnings rounding-up behaviour: some UK evidence. *European Accounting Review* 13(4): 771-786.

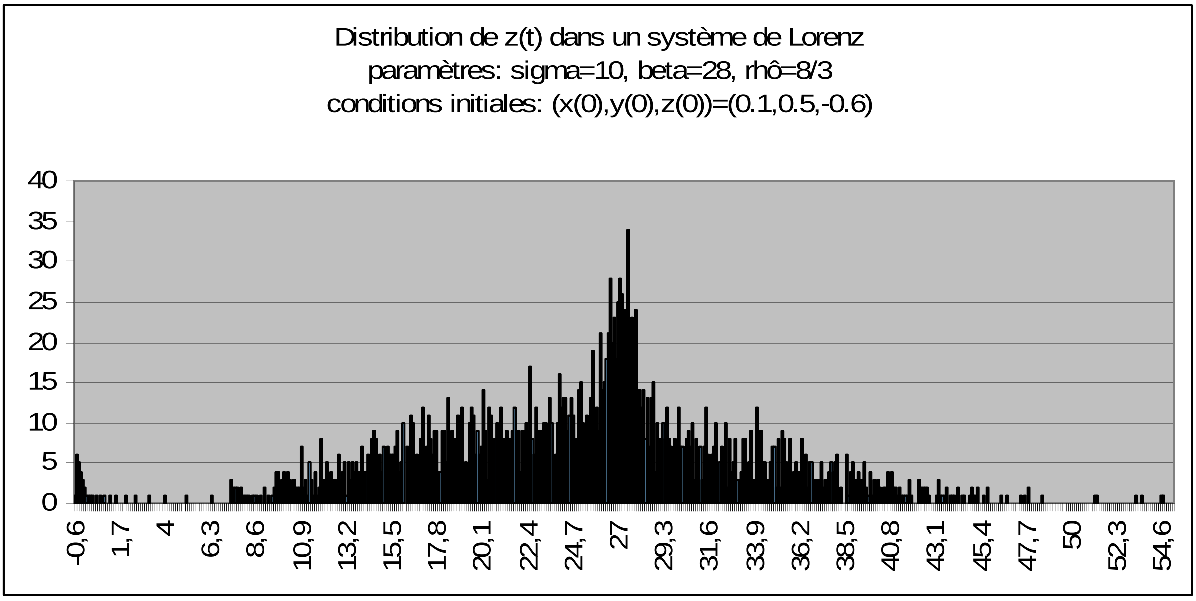
Varian. H. (1972). Benford’s law. *The American statistician* 26(3):64-66.

Weidenmier. M., Herron. T., Building. G. (2004). Selecting an Audit Software Package for Classroom Use. *Journal of Information Systems* 18(1):95-110.

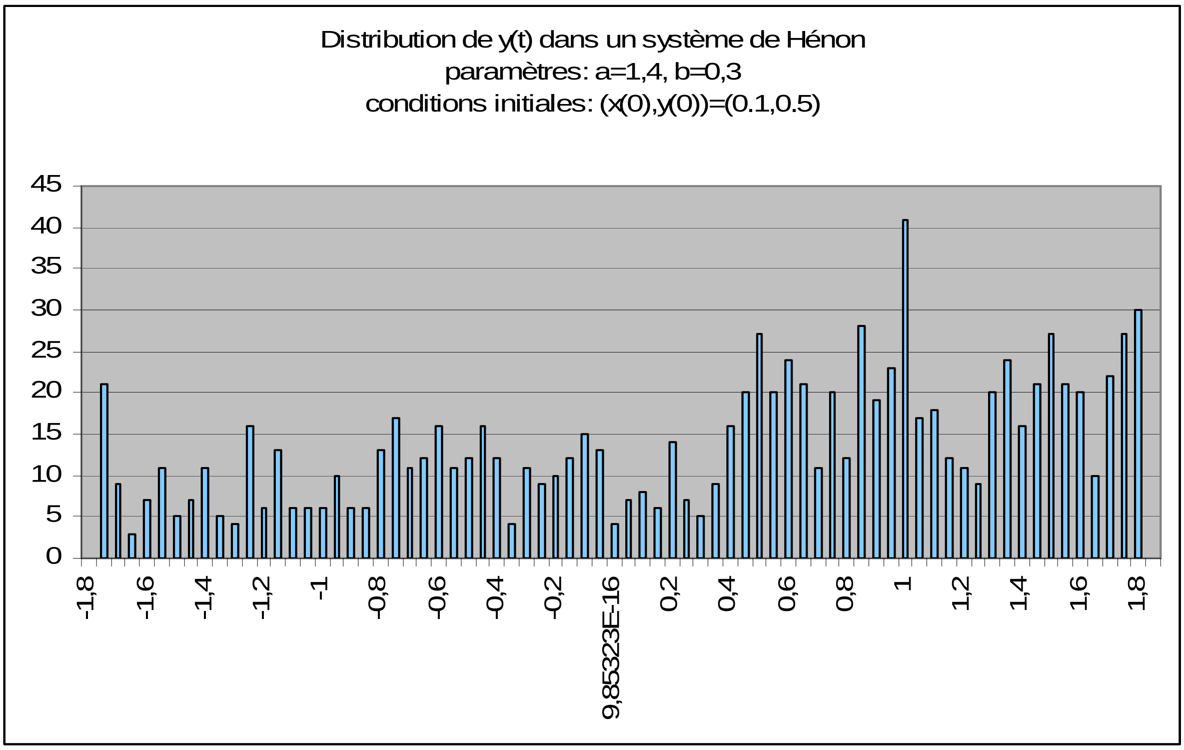
Yin. R.K.(2003). *Case study research: Design and methods.* California:Sage Publications, Inc.

Annexes

A. Systèmes dynamiques chaotiques et existences de seuils dans les distributions



Il apparaît des seuils (l'aspect n'est pas « lisse ») sur cet histogramme, alors que cette série est une simulation d'un système de Lorenz. Il n'y a donc pas de manipulations.



B. Tables 5

Biais significatifs par rapport à la loi de Benford des chroniques de ventes de jeux vidéo dans les pays occidentaux

Roy-Uni	PS2		GBA		GC		XB		DS		PSP		X360		PS3		Wii	
position 1	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z
1	+	**9,67**	+	**2,03**	-	0,09	-	**2,78**	-	1,71	+	**8,99**	+	**5,28**	+	**5,28**	-	***2,36***
2	-	4	+	1,23	-	0,01	-	**3,85**	-	**3,67**	-	**4,39**	-	1,12	-	***1,97***	-	0,59
3	-	**4,54**	-	1,14	+	1,12	-	**3,42**	+	**5,87**	-	**3,8**	-	**3,09**	-	**2,67**	+	**2,67**
4	-	**4,37**	+	0,28	-	0	+	**3,13**	+	**4,04**	-	**2,94**	-	**3,72**	+	1,24	+	**5,86**
5	-	**2,71**	-	0,29	-	0,36	+	**6,54**	+	**0,6**	-	**2,58**	-	0,98	-	0,78	+	1,19
6	-	**2,85**	-	0,06	-	0,36	+	1,21	-	0,26	-	0,08	+	0,05	-	1,47	-	1,74
7	-	0,62	-	1,21	-	0,05	+	**5,08**	-	1,75	-	0,49	+	1,36	-	0,15	-	1,43
8	+	**3,89**	-	1,94	+	0,02	-	1,39	-	1,4	+	***2,21***	+	1,18	-	1,35	-	1,99
9	+	**3,28**	-	1,05	+	0,13	-	0,01	-	1,1	+	0,61	-	0,58	-	0,24	-	1,38

Roy-Uni	PS2		GBA		GC		XB		DS		PSP		X360		PS3		Wii	
Position 2	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z
0	-	0,76	-	**6,72**	+	0,04	+	0,95	-	***2,38***	+	1,03	-	0,39	-	0,49	+	0,42
1	-	1,29	-	0,04	-	1,05	-	0,28	-	0,02	+	1,54	+	1,26	-	0,31	-	0,77
2	-	1,59	+	1,13	+	0,09	+	0,97	+	0,22	+	1,56	+	0,06	+	0,16	-	1,75
3	-	1,62	+	0,01	+	0,43	-	1,4	+	0,22	+	1,56	+	0,5	+	0,01	+	1,01
4	-	0,74	+	**4,58**	-	0,81	-	0,18	+	0,87	-	0,64	-	0,31	-	0,49	+	1,18
5	+	1,28	-	0,07	+	0,22	-	0,19	-	0,55	-	0,48	+	0,35	+	**3,18**	+	1,03
6	+	1,22	+	1,82	-	0,03	-	1,07	-	0,04	-	0,58	-	1,3	-	1,24	+	0,56
7	+	0,46	-	0,04	+	1,9	+	0,37	+	0,47	-	3,22	-	0,13	+	0,19	-	0,87
8	+	0,16	-	1,01	-	1,04	+	0,32	+	0,62	-	0,31	+	1,06	-	1,06	-	0,46

FRANCE	PS2		GBA		GC		XB		DS		PSP		X360		PS3		Wii	
position 1	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z
1	+	***2,47***	-	0,34	-	**3,65**	-	0,22	-	0,42	-	1,46	+	0,15	+	**2,91**	-	0,67
2	-	**5,52**	-	0,76	-	0,01	+	***1,97***	+	2,37	-	**4,8**	-	**3,12**	-	1,76	+	**7,32**
3	-	**4,84**	+	0,66	+	**2,83**	+	**6,09**	+	**4,6**	-	**2,94**	+	**3,46**	-	1,25	-	0,07
4	-	**3,1**	-	1	+	**5,31**	-	0,21	+	0,11	+	0,46	+	0,57	+	0,08	-	**2,62**
5	-	0,26	+	0,91	+	1,27	-	1,91	-	1,94	+	0,61	-	0,72	-	0,81	-	***2,47***
6	+	0,09	-	***2,44***	+	0,64	-	***2,49***	-	1,64	+	**3,94**	+	0,17	-	0,35	-	1,4
7	+	**5,08**	+	0,41	-	***2,47***	-	***2,31***	-	1,48	+	**6,6**	-	0,1	+	0,84	-	1,45
8	+	**4,74**	+	1,26	-	1,55	-	***2,23***	-	1,73	+	1,86	-	0,04	+	0,14	+	0,02
9	+	**6,23**	+	***2,32***	-	1,83	-	1,34	-	***2,41***	+	0,59	-	0,23	-	0,26	-	0,54

FRANCE	PS2		GBA		GC		XB		DS		PSP		X360		PS3		Wii	
Position 2	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z
0	+	0,41	+	**3,03**	+	1,2	+	1,37	+	0,99	+	0,71	+	0,59	-	0,09	-	0,68
1	+	***1,99***	-	0,89	+	0,73	-	0,48	+	0,06	+	0,37	-	1,13	+	0,11	-	0,8
2	+	0,69	+	0,78	-	0,41	+	0,37	+	1,28	+	0,14	+	1,72	+	0,74	+	0,8
3	+	1,95	-	0,17	-	1,01	-	0,19	-	0,25	+	1,76	-	0,02	+	***2,46***	+	0,98
4	-	1,54	-	1,37	+	0,23	-	1,41	-	0,29	+	0,3	-	0,09	+	0,75	+	0,11
5	+	0,06	+	1,4	+	0,43	+	0,22	-	***2,18***	-	0,5	-	0,05	-	1,36	+	0,7
6	-	0,6	-	1,54	+	0,19	+	0,09	+	0,76	-	0,85	+	0,74	-	1,91	+	0,07
7	-	0,06	-	1,18	-	1,18	-	0,49	+	0,09	-	0,09	-	0,68	+	0,49	-	0,12
8	-	0,86	-	0,24	+	0,52	-	0,01	+	0,36	+	0,18	+	0,77	-	0,07	+	0,46
9	-	1,02	+	1,48	-	0	+	1,58	-	0,11	-	1,76	-	0,98	-	0,66	-	1,03

All	PS2		GBA		GC		XB		DS		PSP		X360		PS3		Wii	
position 1	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z
1	+	**0,74**	-	2,58	-	0,36	-	**0,36**	+	**0,17**	-	**2,88**	+	**2,27**	+	**3,12**	+	**0,01**
2	-	**5,87**	+	**0,37**	+	**1,97**	+	7,48	+	***4,86***	-	**4,04**	-	**0,36**	-	0,26	+	***0,77***
3	-	4,69	+	***1,48***	+	**3,59**	+	**0,87**	+	1,11	-	2,07	-	***1,98***	-	***2,4***	-	**2,76**
4	-	**5,07**	-	**0,08**	+	0,1	-	**2,5**	-	1,04	-	**1,49**	+	0,57	-	***0,4***	-	**0,52**
5	-	1,29	-	1,3	-	1,06	-	**0,31**	-	**2,69**	+	**3,81**	+	0,66	-	**1,51**	+	***0,83***
6	+	**2,43**	-	1,14	-	***2,88***	-	3,23	-	***3***	+	**6,25**	+	0,77	-	1,11	+	0,38
7	+	**9,43**	+	2,04	-	***0,85***	-	1,52	-	**2,35**	+	**4,13**	-	0,26	+	***0,43***	+	***0,45***
8	+	***7,68***	+	1,5	-	**0,98**	-	1,95	+	0,12	+	0,88	-	0,89	+	0,35	+	1,64
9	+	**4,13**	+	**1,8**	-	1,23	-	**1,93**	+	0,02	+	0,94	-	2,38	+	0,65	-	**0,54**

All	PS2		GBA		GC		XB		DS		PSP		X360		PS3		Wii	
Position 2	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z
0	+	1,48	+	1,42	-	1,51	-	0,87	+	0,32	+	**2,62**	-	0,15	+	1,32	-	0,39
1	+	**2,81**	-	0,04	-	**2,83**	-	1,64	-	0,69	+	1,51	+	0,05	+	**2,95**	-	0,1
2	+	1,27	+	0	-	0,21	+	***2,15***	+	0,41	-	0,55	+	0,99	+	0,43	-	1,2
3	+	0,38	+	0,72	+	***2,28***	-	0,09	+	0,41	-	1,31	+	1,45	+	1,22	+	0,1
4	-	0,58	+	0,43	+	0,44	+	0,43	-	0,51	-	0,66	-	1,09	-	1,15	-	0,63
5	-	1,44	+	0,3	+	1,5	+	0,02	+	0,12	-	0,5	+	0,83	+	0,25	-	0,51
6	-	1,39	+	0,89	+	0,62	-	0,65	+	0,29	-	0,11	-	1,06	-	0,61	+	0,53
7	-	1,19	+	0,15	+	0,14	+	1,66	-	0,74	+	1,55	+	0,37	-	1,83	+	1,9
8	+	0,64	-	0,82	+	0,08	+	0,1	-	0,36	-	0,59	+	1,3	-	2,76	-	0,17

USA	PS2		GBA		GC		XB		DS		PSP		X360		PS3		wii	
position 1	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z
1	-	**3,48**	-	1,02	-	0,96	-	**5,49**	+	**6,13**	-	**8,13**	-	**4,72**	-	**4,74**	+	**9,41**
2	-	**6,54**	-	**3,31**	+	**4,7**	-	0,17	-	***2,57***	-	**4,48**	-	**5,07**	+	1,01	-	***2,25***
3	-	1,83	-	***2,31***	+	**2,94**	+	**5,27**	+	0,08	+	0,21	-	1,5	-	***2,49***	-	**4,35**
4	-	**3,27**	+	**3,49**	-	1,09	+	**2,84**	+	0,74	+	**4,63**	+	***2,12***	+	***2,2***	-	**2,93**
5	+	1,14	-	0,2	-	0,4	+	**3,17**	-	**2,85**	+	**7,04**	+	**7,92**	+	**4,78**	-	***2,18***
6	+	**7,79**	+	1,56	-	***2,16***	+	1,5	-	***2,07***	+	**7,23**	+	**3,2**	+	1,81	-	1,58
7	+	**8,93**	+	1,04	-	***2,31***	-	0,95	-	**2,2**	+	**2,71**	+	**3,29**	+	***1,98***	-	***2,16***
8	+	***2,35***	+	1,22	-	**2,38**	-	1,78	-	0,36	-	0,42	+	0,49	-	1,6	-	1,81
9	+	**3,48**	+	**2,78**	-	0,73	-	**2,82**	-	0,31	-	1,38	+	0,14	+	0,1	+	**4,26**

USA	PS2		GBA		GC		XB		DS		PSP		X360		PS3		wii	
Position 2	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z	biais	Z
0	+	1,18	+	0,15	-	0,99	+	0,03	-	0,4	+	0,1	-	0	+	0,2	+	0,01
1	-	1,47	-	0,25	+	0,25	-	0,08	-	0,74	-	0,03	-	1,34	-	***2,2***	+	**2,59**
2	-	0,24	+	1,75	+	0,38	-	0,17	+	0,55	+	0,54	-	0,06	-	0,63	-	0,55
3	-	0,87	+	0,16	+	1,19	+	1,22	+	1,86	+	0,76	+	0,5	+	1,56	+	0,71
4	-	0,13	+	0,95	+	0,09	-	0,46	+	1,88	+	1,43	+	1,19	+	1,14	+	3,04
5	+	0,28	+	1,02	-	0,15	+	0,72	+	0,98	-	0,7	+	0,35	+	1,3	-	1,21
6	+	1,51	-	1,83	+	0,25	+	0,33	+	0,71	+	0,64	+	0,25	+	1,14	-	1,05
7	-	0,4	+	0,09	+	0,25	+	0,52	-	**3,04**	-	0,88	+	0,4	+	0,66	-	0,9
8	-	0,2	-	0,19	+	0,05	-	0,12	-	0,83	-	0,74	-	0,13	-	1,43	-	1,65
9	+	1,87	-	0,96	-	0,57	-	1	-	0,23	-	0,37	-	0,69	-	1,34	-	0,17