

Le commerce électronique de biens culturels : contributions empiriques

Bora Eang

▶ **To cite this version:**

Bora Eang. Le commerce électronique de biens culturels : contributions empiriques. Economies et finances. Télécom ParisTech, 2012. Français. NNT : 2012ENST0076 . pastel-00990172

HAL Id: pastel-00990172
https://pastel.hal.science/pastel-00990172

Submitted on 13 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2012-ENST-076

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Sciences Économiques et Sociales »

présentée et soutenue publiquement par

Bora EANG

le 14 décembre 2012

LE COMMERCE ÉLECTRONIQUE DES BIENS CULTURELS : CONTRIBUTIONS EMPIRIQUES

Directeur de thèse : **David BOUNIE**
Co-encadrement de la thèse : **Patrick WAELBROECK**

Jury

M. Alain RALLET, Professeur, Université Paris Sud		Rapporteur
M. Marvin SIRBU, Professeur, Carnegie Mellon University		Rapporteur
M. Raphaël GOUMAIN, Responsable Marketing, Google		Examinateur

THÈSE

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

Table des matières

REMERCIEMENTS 7

CHAPITRE 1 : INTRODUCTION : LE COMMERCE ELECTRONIQUE DE BIENS CULTURELS 9

1.1 Objectifs de la thèse et contributions à la littérature 12

1.1.1 Théorie de la longue traîne et cannibalisation des ventes de produits culturels 12

1.1.2 La dispersion des prix sur Internet 17

1.2 Nos papiers de recherche 19

1.3 Références 20

CHAPITRE 2 : MARCHE INTERNET ET RESEAUX PHYSIQUES. UNE COMPARAISON DES VENTES DE LIVRES EN FRANCE 23

2.1. Introduction 23

2.2 Méthodologie et description des données 26

2.3 Comparaison des ventes dans les réseaux physiques et en ligne 30

2.3.1. Caractérisation de la distribution des ventes 30

2.3.2. Les marchés des livres en ligne et hors ligne sont-ils comparables ? 35

2.4 Interactions entre les catalogues de livre en ligne et des réseaux physiques semaine après semaine 38

2.5 Conclusion 44

2.6 Références 44

2.7 Annexe 45

CHAPITRE 3 : LES PLATEFORMES DE VENTES SUR INTERNET : UNE OPPORTUNITE POUR LES INDUSTRIES CULTURELLES ? 47

3.1 Introduction 47

3.2 Les meilleures ventes de livres, CD et DVD sur Amazon et Marketplace 49

3.2.1 Mesure des ventes 51

3.2.2 Taille et structure des marchés 52

3.2.3 Les meilleures ventes de produits sur Amazon et Marketplace 53

3.2.3.1 Des catalogues de produits différents 54

3.2.3.2 L'importance des titres du répertoire ancien 55

3.3 Les meilleures ventes de livres sur Amazon, Marketplace et les réseaux de distribution physique 56

3.4 Conclusion 59

3.5 Références 61

3.6 Annexe 62

CHAPITRE 4: SUPERSTARS AND OUTSIDERS IN ONLINE MARKETS, AN EMPIRICAL ANALYSIS OF ELECTRONIC BOOKS 64

4.1 Introduction 64

4.2 Related Literature 66

4.3 Data and methodology 68

4.3.1 Data collection 69

4.3.2 Descriptive statistics 70

4.4 Econometric Methodology and Results 72

4.4.1 Superstar Titles 73

4.4.2 Digital Outsiders 76

4.5 Conclusion 78

4.6 References 80

CHAPITRE 5 : DISPERSION DES PRIX SUR INTERNET : UNE ANALYSE EMPIRIQUE SUR LES MARCHES DES LIVRES, CD ET DVD EN FRANCE 83

5.1 Introduction 83

5.2 Description des données 86

5.3 Dispersion des prix affichés ou postés: stratégies de prix et dimension temporelle 89

5.3.1 Distribution des changements de prix sur Marketplace 91

5.4 Dispersion des prix de vente : caractéristiques des produits et des vendeurs 96

5.4.1 Détermination d'une transaction 96

5.4.2 Comparaison des prix postés et des prix de vente 98

5.4.3 Déterminants des prix de vente : caractéristiques des vendeurs et du temps 99

5.5 Conclusion 101

5.6 Références 102

CHAPITRE 6: ONLINE PRICE DISPERSION, AN INTERNATIONAL COMPARISON 105

6.1 Introduction 105

6.2 Data Set and Definitions 107

6.3 Posted Price Dispersion – Time Dimension 111

6.3.1 Distribution of Price Changes on Amazon Marketplace 113

6.4 Price, Seller and Item Characteristics 118

6.4.1 Definition of a Transaction 118

6.4.2 Transaction and Posted Prices: A Comparison 121

6.5 Determinants of Transaction Prices: Seller Characteristics and Time 121

6.6 Conclusions 123

6.7 References 125

6.8 Appendix 127

CHAPITRE 7: CONCLUSION 129

A mes Parents,

REMERCIEMENTS

Je tiens à adresser en premier lieu mes remerciements les plus sincères et chaleureux à mes directeurs de thèse, David Bounie et Patrick Waelbroeck, sans qui cette thèse n'aurait tout simplement pas été possible. Durant ces années, j'ai eu l'occasion de bénéficier de leurs conseils et remarques avisés qui m'ont toujours été très précieux dans la réalisation de ce travail. Je garde en moi le souvenir de discussions passionnantes et riches tant sur le plan académique que sur le plan humain. Je leur suis également très reconnaissant pour le dévouement et la confiance qu'ils m'ont tous deux accordé. Aussi, je ne peux exprimer ici, toute ma gratitude pour leur infaillible optimisme qui m'a toujours permis d'avancer malgré les petits obstacles du quotidien.

Je remercie aussi très sincèrement Marvin Sirbu, Alain Rallet et Raphael Goumain pour l'honneur qu'ils m'ont fait en acceptant de participer à ce jury de thèse et d'évaluer ce travail.

Je suis également reconnaissant à mon Ecole Doctorale l'EDITE et Telecom ParisTech pour m'avoir attribué une allocation de thèse. Je remercie à ce propos Florence Besnard pour sa compréhension et son aide précieuse durant chaque année de réinscription.

Ces années de thèse n'auraient sûrement pas été aussi agréables sans mes collègues de bureau et du département avec qui je me suis attaché. Je leur exprime ici, mes remerciements les plus chaleureux pour les rires et les bons moments partagés. Je pense notamment à Laurie Marrauld pour son thé et ses biscuits, Yassine Bouhdaoui et ses compétences mathématiques, Caroline Guillot pour sa bonne humeur, Germain Gaudin pour son accueil dans le département et ses suggestions d'articles, Laurent Camus pour sa curiosité, Thomas Houy et son humour toujours présent et enfin Isabelle Garron pour m'avoir si gentiment ouvert sa porte sur l'art.

Merci à Marie Baquero, Corinne Chevalier et Marie-Josée Vatin pour leur aide administrative tout au long de mon doctorat.

Enfin, je n’oublie pas mes parents et mon frère auxquels je dois tout. Merci d’avoir été là et de m’avoir soutenu après toutes ces longues années d’études.

CHAPITRE 1 : INTRODUCTION : LE COMMERCE ELECTRONIQUE DE BIENS CULTURELS

L'utilisation généralisée d'Internet depuis une quinzaine d'années s'est accompagnée d'une augmentation des ventes de biens et services en ligne (commerce électronique). En 2011, le chiffre d'affaires du commerce sur Internet s'est élevé à 256 milliards de dollars aux Etats-Unis , à 59,4 milliards d'euros au Royaume-Uni et à 38,7 milliards d'euros en France Les taux de croissance du commerce électronique se situent aujourd'hui entre 12% et 24% selon les pays et le commerce sur Internet représente désormais près de 7,9 % du commerce de détail en France et 12% au Royaume-Uni.

Le développement du commerce et des échanges en ligne font aujourd'hui régulièrement les titres de la presse. Dans ces derniers, Internet est tour à tour présenté comme un canal de distribution prometteur qui offre de nouvelles opportunités d'achat pour les consommateurs, des succès entrepreneuriaux aux méthodes de management efficaces, de perspectives nouvelles pour améliorer l'efficacité des administrations (administration électronique), etc. Mais, le succès d'Internet a des revers. Aux nouvelles opportunités d'achat sont opposées les fraudes sur les paiements, le vol des données sensibles, l'uniformatisation des goûts, la concurrence déloyale, etc. Ces questions, au cœur des préoccupations quotidiennes des consommateurs, des entreprises et des administrations, font désormais l'objet de nombreuses recherches académiques en sciences sociales et constituent un champ de recherches fécond et innovant en économie, l'économie d'Internet et du numérique.

Notre travail de recherche s'inscrit précisément dans ce champ de recherches et se focalise sur un thème particulier, le commerce électronique de biens culturels. Nos recherches montrent que l'Internet crée à la fois des menaces et des opportunités pour les acteurs des industries culturelles; menaces à travers l'existence de nouveaux canaux de distributions concurrents aux réseaux de distribution physique et opportunités pour les auteurs d'ancien et nouveaux produits de niches ainsi que pour les vendeurs joignant des plateformes de ventes telles qu'Amazon Marketplace.

Les études relatives au commerce électronique de biens culturels sont anciennes et nombreuses pour plusieurs raisons.

Tout d'abord, les biens culturels à l'image des CD, DVD et livres ont très tôt été commercialisés sur Internet par des sites de commerce électronique pionniers tels qu'Amazon, eBay, etc. Le succès rapide de ces sites offraient aux consommateurs de nouveaux outils pour explorer de nouveaux produits, talents et genres (musicaux) et ce, très souvent, à des prix plus attractifs que dans le commerce traditionnel (Chevalier et Golsbee, 2003). Cette première révolution technologique qui a consacré Internet en tant que nouveau canal de distribution de biens physiques s'est également accompagnée du développement de réseaux de partage sur lesquels les biens culturels étaient échangés gratuitement (piratage). Cette première révolution technologique a donné naissance à un ensemble de recherches empiriques pour évaluer les effets économiques d'un nouveau canal de distribution sur les ventes dans les réseaux traditionnels (Balasubramanian, 1998), le piratage des œuvres culturelles (Peitz and Waelbroeck, 2006), le développement des communautés électroniques (Curien et al., 2001), etc.

Mais une deuxième révolution technologique est en passe aujourd'hui de modifier à nouveau en profondeur les logiques de consommation et de production des biens culturels : les œuvres numériques. Internet offre la possibilité de commercialiser directement des contenus culturels numériques qui deviennent des produits concurrents aux supports physiques traditionnels. Cette deuxième révolution technologique suscite de nouvelles interrogations et méfiances de la part des éditeurs et distributeurs qui voient dans cette dématérialisation, une réelle menace pesant sur leur modèle économique historique basé sur la vente de supports physiques. L'industrie de la musique, première à subir les changements provoqués par la première révolution technologique, est aujourd'hui très affectée par ces évolutions et reste encore aujourd'hui à la recherche d'un modèle économique capable de coexister avec une offre quasi gratuite sur internet. Selon l'IFPI (International Federation of the Phonographic Industry), les ventes de musique préenregistrée ont baissé de 3% dans le monde en dépit de la croissance des ventes de musique légale en ligne. De même, l'industrie du film et de la vidéo n'est pas épargnée par cette révolution numérique. Le circuit traditionnel de la vidéo a dû faire

face à l'arrivée de distributeurs en ligne tel qu'Amazon pour la vente de DVD, mais aussi d'un nouveau type de concurrents comme iTunes Video Store ou Netflix pour la vidéo à la demande, de Youtube ou Dailymotion pour les sites de partage et, dernièrement, des chaînes télévisées pour la télévision de rattrapage (arte+7, M6 replay). Enfin, le secteur du livre hier encore épargné par la révolution numérique, doit aujourd'hui s'organiser devant la numérisation de ses contenus et la distribution des livres numériques (e-books). L'arrivée des tablettes numériques multifonctions permettant la lecture d'e-books à l'instar de l'iPad d'Apple et du Kindle d'Amazon, remettent en question le rôle des maisons d'édition du livre en permettant à des auteurs de s'autoproduire et de commercialiser directement ses œuvres sur les plateformes. Les livres électroniques en constante croissance, représentent désormais 8,3% des ventes totales de livres aux Etats-Unis en 2010 (contre 3,2% en 2009) et Amazon annonçait dès 2011 vendre plus de livres électroniques sous format Kindle que de livres imprimés.

La deuxième raison qui explique le développement des études sur les biens culturels est relative à l'accès gratuit et public aux caractéristiques des produits et de vente sur les sites de commerce électronique. Il est désormais possible de suivre les ventes de milliers de produits, leurs prix, leurs évaluations par les consommateurs, etc. et ce sur des dizaines de mois. La disponibilité de ces données permet donc aux économistes de mesurer et de tester les prédictions de modèles microéconomiques sur de nombreux sujets tels que les stratégies de tarification des firmes sur les marchés (Ghose and Sundararajan, 2006), le fonctionnement des marchés électroniques (Ellison and Ellison, 2005), le rôle de la réputation (Resnick and Zeckhauser, 2002; Bajari and Hortacsu, 2004; Houser and Wooders, 2006), les externalités de réseau (Rochet et Tirole, 2006), le rôle de l'information dans la formation de la demande (Acquisti and Varian, 2005) et son impact sur les marchés (prix, qualité, diversité, concurrence entre firmes (open source), etc.). A titre d'illustration, les données accessibles sur Amazon permettent de constituer des données de panel sur des milliers de titres CD, DVD et livres sur les douze dernières années et disponibles sur plusieurs marchés (France, Etats-Unis, Japon, Allemagne, Chine, etc.).

Enfin, les études sur le commerce électronique de biens culturels font suite aux nombreux débats de société sur les impacts économiques de l'internet sur différentes industries culturelles à l'image de celle de la musique, du cinéma et de la vidéo et du livre. Les autorités publiques ont été contraintes d'intervenir sur les marchés pour promouvoir la diversité culturelle (loi sur le prix unique du livre électronique), préserver les droits d'auteur (loi Création et Internet), etc. Dans le contexte particulier du droit d'auteur, une institution exclusivement dédiée à la diffusion des œuvres et la protection des droits sur internet (Hadopi) a été créée en France en 2009. L'Hadopi a pour missions de promouvoir le développement de l'offre légale et d'observer l'utilisation licite et illicite des œuvres sur Internet ; de protéger les œuvres à l'égard des atteintes aux droits qui leur sont attachés dans le cadre de la réponse graduée et de réguler l'usage des mesures techniques de protection.

1.1 Objectifs de la thèse et contributions à la littérature

L'objectif de notre thèse est de mobiliser un ensemble de données originales sur le fonctionnement des marchés électroniques de biens culturels pour éclairer les débats économiques et publics sur deux thèmes particuliers : la dispersion des prix et la théorie de la longue traîne. Ces thématiques ont donné lieu à de nombreuses controverses et sont au cœur de nos contributions. Nous abordons successivement les résultats de la littérature sur la théorie de la longue traîne et la dispersion des prix et nous discutons ensuite nos contributions.

1.1.1 Théorie de la longue traîne et cannibalisation des ventes de produits culturels

Selon Chris Anderson (2005), l'essor du commerce en ligne s'accompagne d'une mutation économique des industries culturelles qui évolueraient d'une économie de produits « superstar » à une économie de produits de « niche ». En d'autre termes, comparativement aux marchés traditionnels caractérisés par des ventes élevées d'un très faible nombre de titres et des ventes très faibles d'un très grand nombre de produits dits « supertars », la distribution des ventes sur Internet tendrait à être moins concentrée autour de produits « superstars » et à être caractérisée par une « longue traîne » agrégeant un grand nombre de produits de niche, peu profitable individuellement mais profitable dans leur

ensemble. Cette prédiction s'appuie sur des caractéristiques particulières d'Internet : la contrainte physique d'espace de stockage ne s'impose pas à la vente en ligne contrairement aux enseignes physiques et, de plus, l'utilisation de plus en plus courante d'outils de recherche sur Internet devrait faciliter l'exploration et la découverte de produits de niche, et donc rendre leur commercialisation profitable pour les distributeurs de contenus culturels. Cette prédiction s'appuie notamment sur une étude empirique de Brynjolfsson et al. (2003) qui montre qu'une proportion significative des ventes sur Amazon.com provient de produits de niche non disponibles dans les magasins traditionnels.

Dans un premier article intitulé « Les plateformes de ventes sur Internet : une opportunité pour les industries culturelles », nous étendons la recherche Brynjolfsson et al. (2003) au cas des plateformes de vente. Depuis cette recherche pionnière, de nouveaux intermédiaires sont apparus sur le marché pour organiser les ventes de produits culturels : les plateformes de vente. Ces dernières telle qu'Amazon MarketPlace, eBay, etc. mettent en relation des vendeurs professionnels et particuliers pour commercialiser des produits culturels neufs et d'occasion. Ces plateformes entrent parfois directement en concurrence avec les sites web qui les hébergent à l'image de la plateforme Amazon MarketPlace. Dans ce contexte, se pose la question de l'impact de ces nouveaux intermédiaires sur la commercialisation de biens culturels : ces intermédiaires commercialisent-ils plus des produits de niche que les sites web et que les marchés traditionnels ? Nous avons étudié en détails les meilleures ventes de livres, CD et DVD sur le site français d'Amazon et nous les avons comparé à celles de la plateforme de vente Amazon Marketplace et à celles des meilleures ventes dans les réseaux traditionnels hors Internet. Nous montrons que sur la période de Mars à Août 2006 le catalogue des meilleures ventes d'Amazon et de Marketplace diffère principalement selon deux dimensions. Premièrement, le catalogue des livres, CD et DVD en commun dans les meilleures ventes respectives n'est que de l'ordre de 20%. Deuxièmement, au moins 25% des meilleures ventes de Marketplace ont des dates de sorties antérieures à celles d'Amazon. Ces résultats nous permettent de conclure que la demande de produits culturels sur la plateforme se porte sur des produits du répertoire ancien qui ne sont plus (ou peu) commercialisés sur Amazon et dans les réseaux physiques. En moyenne, 36 % des meilleures ventes d'Amazon font également parties des meilleures ventes hebdomadaires des réseaux

physiques, alors que ce pourcentage n'est que de 18 % pour les meilleures ventes de Marketplace. En croisant les trois marchés, nous observons qu'en moyenne 11% des livres sont communs aux trois réseaux de distribution.

Contribution n°1 : Les plateformes de vente sur Internet augmentent les ventes de produits de niche qui se caractérisent par des dates de parution antérieures aux produits vendus sur les sites web commerciaux et les réseaux traditionnels (magasin). Les plateformes de vente augmentent la durée de vie des produits et constituent une opportunité pour les industries culturelles. Ce résultat est vérifié sur le marché du livre mais également sur le marché des CD et DVD en France.

Ce premier résultat a des implications importantes pour la théorie de la longue traîne. Autrement dit, ce résultat semble valider empiriquement que le réseau Internet favorise la commercialisation plus importante de produits de niche et valide en partie la thèse d'une cannibalisation des ventes de produits culturels entre les marchés Internet et les marchés traditionnels. Selon la théorie de la longue traîne, nous devrions donc assister sur un marché en particulier, celui du livre par exemple, à un aplatissement global de la distribution des ventes de livres traduisant ainsi une cannibalisation des ventes de produits superstars au profit des produits de niche. La théorie de la longue traîne reste aujourd'hui cependant controversée. Tout d'abord, peu d'études semblent confirmer son existence. Ainsi, Elberse et Oberhozer-Gee (2006) montrent que l'essor du commerce électronique de DVD et VHS aux Etats-Unis a augmenté à la fois le phénomène de superstars et développé la longue traîne sans toutefois rendre profitable un modèle d'affaires basé sur son exploitation. Dans le cas français, Benghozi et Benhamou (2008) confirment ce résultat pour les CD et les DVD pour la période de 2002 à 2005. Benhamou (2008) observe également des ventes très concentrées lors de la rentrée littéraire 2005 : sur les 693 titres étudiés et appartenant essentiellement à la catégorie littéraire roman français (hors format poche), 16% d'entre eux se sont partagés 83% des ventes totales. Outre les débats sur l'existence ou non d'une longue traîne, peu d'études se sont intéressées aux interactions entre les marchés : est-ce que le marché Internet influence les ventes de livre papier dans les réseaux traditionnels ?

Dans un deuxième article intitulé, « Marché Internet et réseaux physiques : comparaison des ventes de livres en France », nous comparons les meilleures ventes de livres en France dans les réseaux traditionnels et le réseau Internet (Amazon.fr) sur une période de 22 semaines (de mars à août 2006). L'objectif de ce travail est, d'une part, de caractériser la distribution des ventes de livres dans les marchés hors-ligne et en ligne et, d'autre part, de modéliser la dynamique des interactions entre ces deux marchés. Nos analyses montrent premièrement que 20% seulement des livres appartenant aux tops 100 respectifs sont communs sur l'ensemble de la période. Les catalogues en ligne et hors-ligne sont donc différents et suggèrent que i. les marchés sont plutôt complémentaires et ii. que les produits superstars diffèrent en partie entre les marchés. Par ailleurs, environ 50% des meilleures ventes (top 100) d'Amazon ne se retrouvent pas dans le top 500 des meilleures ventes de livres sur le marché physique. Nous en concluons que pour ces livres, l'effet d'expansion de marché domine l'effet de cannibalisation. Deuxièmement, sur les 20% de livres communs aux deux plates-formes, nous observons pour 58% d'entre eux des décalages temporels d'entrée des livres dans les tops des ventes. Les livres concernés appartiennent à des genres littéraires précis. Nous utilisons un modèle vectoriel autorégressif (VAR) et montrons que les ventes en librairies ont un effet positif sur les ventes en ligne alors que les ventes en ligne n'ont qu'un effet très marginal sur les ventes réalisées dans les réseaux physiques. L'effet de prescription d'Internet sur les ventes physiques semble encore limité.

Contribution n°2 : Les meilleures ventes de titres superstars diffèrent entre les marchés Internet et les réseaux traditionnels (mais ne permettent pas de conclure à une hausse des ventes de produits superstars) et confirment un effet d'expansion pour les produits de niche. Internet a cependant peu d'influence sur les ventes de titres dans les magasins traditionnels.

Les deux premiers travaux de recherche portaient sur l'impact d'Internet sur les ventes de biens physiques. Internet offre désormais la possibilité non pas d'acheter des biens physiques mais des biens numériques. Le cas du livre est illustratif à cet égard. Le développement du livre électronique est redouté par les acteurs de l'industrie du livre pour de multiples raisons. Bien que limité pour l'heure en France, les derniers chiffres publiés par l'Association of American Publishers confirment l'engouement du public pour les livres numériques et Amazon a confirmé il y a quelques mois que les

ventes de kindles sur Amazon.com dépassaient désormais les ventes de livres papier. Compte tenu des contraintes budgétaires des consommateurs, le livre numérique peut cannibaliser les ventes de livre papier. Cette crainte d'une substitution du livre numérique au livre papier est en outre renforcée par l'arrivée de nouveaux entrants sur le marché du livre numérique venus de l'Internet ou de l'informatique tels qu'Amazon, Apple ou Google qui bouleversent les rapports de force entre les éditeurs, auteurs et distributeurs (libraires, grandes surfaces spécialisées) et, de fait, les modèles économiques de ces acteurs. Enfin, à l'image de l'industrie de la musique, les auteurs et les éditeurs appréhendent une augmentation du piratage qui pourrait engendrer à terme une baisse des revenus.

De nombreuses études, relayées par la presse, ont analysé et commenté les impacts possibles du livre numérique sur l'industrie. Oestreicher-Singer et Sundararajan (2010) s'interrogent sur la cannibalisation des ventes de livre papier et sur les mesures tarifaires appropriées mais ne mesurent pas directement les effets potentiels de substitution entre les formats. Hu et Smith (2011) est la seule recherche empirique à s'intéresser à l'impact de la distribution d'e-books sur les ventes de livres papiers. Les auteurs étudient les données de ventes de livres papier et électronique d'un éditeur qui décide à une date donnée de stopper la vente de livres électroniques sur le site d'Amazon.com et, peu de temps après (8 semaines), de revendre à nouveau des livres électroniques sur Amazon.com. Les livres étudiés ne sont pas des titres superstars car ils suppriment de leur échantillon les livres dont le nombre de copies hebdomadaires est très élevés (supérieur à 1600). Les auteurs montrent en particulier que les livres papier et électronique sont des biens complémentaires : une absence de vente de livres électroniques ne semble pas avoir d'incidence sur les volumes de vente de livre électronique papier (ce qui suggère une élasticité prix croisée négative).

Dans un troisième article intitulé « Superstars and Outsiders in Online Markets: An Empirical Analysis of Electronic Books », nous analysons la liste des tops 100 des meilleures ventes de livre papier et numérique sur la période de Novembre 2007 à Juillet 2010 sur le site d'Amazon.com. Nos analyses montrent que la part de recouvrement entre les marchés sur l'ensemble de la période est faible et que la demande de livre numérique semble porter sur des titres qui ne font pas partie des meilleures ventes dans les réseaux de vente traditionnels. Ce premier résultat conforte l'idée d'un effet

d'expansion du marché du livre. Les best-sellers papier qui connaissent également un succès numérique sont publiés essentiellement par des auteurs à succès et possèdent des prix de vente inférieurs à leurs équivalents papier. L'apparition d'outsiders numériques, i.e. de titres numériques qui existent en format papier mais qui n'apparaissent pas dans les meilleurs classements des ventes de livre papier, s'explique par une différence de prix élevée avec le format papier et par une date de publication ancienne ; dans ce dernier cas, les réseaux de distribution numériques permettent de relancer la popularité des livres du répertoire ancien. Enfin, les résultats des estimations mettent en évidence une élasticité-prix croisée positive entre les livres papiers et électronique et confirme une substituabilité (cannibalisation) entre les deux formats pour les titres superstar.

Contribution n°3 : Les titres superstars papier et électronique sont des produits substituts.

1.1.2 La dispersion des prix sur Internet

Un nombre important de comparateur de prix tels que kelkoo.fr, twenga.fr ou achetezfacile.com, permettent aujourd'hui aux internautes de consulter en ligne un vaste catalogue de produits triés par catégories de prix et de comparer immédiatement ces différentes offres. En toute logique, cet accès à une information quasi exhaustive sur l'offre des produits existants devraient réduire les écarts de prix entre les différentes offres en ligne d'un même produit puisque tous les internautes ont la possibilité d'acheter au meilleur prix. En d'autres termes, un produit de qualité homogène vendu sur internet ne devrait posséder qu'un seul prix puisque tous les internautes peuvent choisir la meilleure offre disponible, c'est-à-dire l'offre proposant le prix le plus bas. Cette théorie d'une convergence des prix est d'autant plus justifiée que le coût d'ajustement des prix sur un site Web est beaucoup plus faible que le coût d'impression d'un nouveau catalogue papier. Pourtant la plupart des études observent une forte dispersion des prix en ligne qui persiste dans le temps (Baye, Morgan and Scholten, 2004; Pan et al., 2004).

Plusieurs études empiriques ont été réalisées pour résoudre ce paradoxe apparent. Claye et al., (2002), Brynjolfsson and Smith (2000) et Brynjolfsson et al., (2010) montrent que les coûts de recherche peuvent expliquer la dispersion des prix même si un internaute utilise des sites de

comparaison de prix. De même, Resnick and Zeckhauser (2002), Bajari and Hortacsu (2004), Houser and Wooders (2006) affirment que la réputation en ligne permet d'expliquer une partie de l'augmentation du prix et des ventes d'un produit sur un site tel qu'eBay même si ce résultat est plus difficile à observer sur d'autres sites tels qu'Amazon (Ghose et al., 2009). Spulber (1995), Claye et al., (2002) et Baye, Morgan and Scholten (2004) trouvent également que la dispersion des prix décroît avec le nombre de vendeurs sur un marché. Enfin, Baye and Morgan (2001) avancent que la dispersion des prix pourrait être liée à l'adoption de stratégies de tarification aléatoire de certains vendeurs. En effet, les vendeurs d'un bien homogène avec coût de recherche choisissent une stratégie mixte de prix qui consiste à choisir un prix tiré de manière aléatoire de la distribution de prix d'équilibre.

Dans un quatrième article intitulé « Une analyse empirique de la dispersion des prix sur Internet », nous avons collecté et suivi dans le temps toutes les offres de livres, CD et DVD neufs qui sont entrés dans le classement des cent meilleures ventes journalières sur la plateforme de vente française Amazon Marketplace sur la période de mars à novembre 2006. L'utilisation de cette plateforme est intéressante car les coûts de recherche sont très limités (comparaison très simple des offres), les offres peuvent être suivies dans le temps (dimension temporelle des changements de prix) et les caractéristiques de la réputation des vendeurs peut être suivie dans le temps. Cette collecte de données permet donc de tester la plupart des arguments avancés pour expliquer la dispersion des prix. Les résultats des traitements montrent en particulier en utilisant les données de prix postés que les vendeurs modifient très peu leurs prix dans le temps et que ces changements sont faibles. Deuxièmement en comparant les prix postés aux prix de transaction, nous montrons que la dispersion des prix en ligne croît avec le nombre de vendeurs dans le temps parce que de nouveaux acteurs entrent sur le marché à des prix plus faibles que les prix moyens sur le marché. Enfin, nous montrons qu'un vendeur reconnu qui dispose d'un large catalogue de produits peut tarifer son bien au-delà du prix de ses concurrents ; la réputation d'un vendeur influence les prix de transaction.

Contribution n°4 : En l'absence de coûts de recherche, la dispersion des prix résulte ne résulte pas de stratégies de prix aléatoire et n'augmente pas avec la baisse du nombre de vendeurs dans le

temps. Nos résultats confirment en revanche un effet de réputation des vendeurs capturé par la taille du catalogue des vendeurs.

Dans un dernier travail intitulé « Online Price Dispersion : An International Comparison », nous confirmons l'existence de ces contributions. Nous montrons en outre l'existence d'un nouvel effet : le degré de maturité des plateformes de vente en ligne. Ainsi, en comparant les analyses effectuées en France à celles du Royaume-Uni et des Etats-Unis, nous mettons en évidence que la dispersion des prix croît également avec la structure des marchés liées étroitement au degré de maturité des marchés. Ainsi, en 2006, date à laquelle les relevés de données ont été effectués, le marché du commerce électronique en France est moins développé que celui du Royaume-Uni et des Etats-Unis. Dans ce contexte, nous montrons, dans le prolongement de nos premiers travaux que les marchés peu matures sur lesquels le nombre de commerçants est structurellement moins important que la dispersion des prix est plus faible. En revanche, dans les marchés plus matures, la dispersion des prix est plus importante.

Contribution n°5 : La dispersion des prix est liée au degré de maturité des marchés.

1.2 Nos papiers de recherche

Ce manuscrit est articulé autour de cinq papiers de recherche. Trois de ces papiers ont été publiés et les deux autres ont été soumis à des revues internationales et ont le statut de document de travail. Les articles de recherche sont :

1 Bounie, D., Eang, B. et Waelbroeck, P., (2010a), « Marché Internet et réseaux physiques : comparaison des ventes de livres en France, » Revue d'économie politique, 120(1): 141-163, 2010.

2 Bounie, D., Eang, B., et Waelbroeck, P., (2010b), « Les plateformes de ventes sur Internet : une opportunité pour les industries culturelles ? » Revue économique, 62(1): 101-112, 2011.

3 Bounie, D., Eang, B., Sirbu, M. and Waelbroeck, P. (2012), « Superstars and Outsiders in Online Markets: An Empirical Analysis of Electronic Books », mimeo.

4 Bounie, D., Eang, B., Sirbu, M. and Waelbroeck, P., (2011), « Une analyse empirique de la dispersion des prix sur Internet", Revue Française d'Économie, 4(25): 121-146.

5 Bounie, D., Eang, B., Sirbu, M. and Waelbroeck, P., (2012), « Online Price Dispersion: An International Comparison », mimeo.

1.3 Références

Acquisti, A. and Varian H., 2005. Conditioning Prices on Purchase History. Marketing Science. 24(3), 1-15.

Baye, M. R., and Morgan, J., 2001. Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets. American Economic Review. 91(3), 454-474.

Baye, M. R., Morgan, J., and Sholten, P., 2004. Price Dispersion in The Small And in The Large: Evidence From An Internet Price Comparison Site. Journal of Industrial Economics. 52(4), 463-496.

Bajari, P. and Hortacsu A., 2004. Economic Insights from Internet Auctions. Journal of Economic Literature. 42(2), 457-86.

Benghozi, P-J., Benhamou, F. 2008. Longue traîne : levier numérique de la diversité culturelle? Culture Prospective. 2008-1.

Benhamou, F., Éléments de réflexion et propositions à propos de la diversité culturelle. Les cas du livre et du cinéma en salles, rapport pour le Groupe d'analyse stratégique des industries culturelles (Gasic), Paris, DEPS, Ministère de la Culture et de la Communication, janvier 2008.

Brousseau E. et Curien N. 2007. Internet and Digital Economics. Principles, Methods and Applications. Cambridge University Press.

Balasubramanian, S. 1998. Mail Versus Mall: A Strategic Analysis of Competition Between Direct Marketers and Conventional Retailers. Marketing Science. 17(3), 181-195.

Brynjolfsson, E. and Smith, M., 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. Management Science, 46(4), 563-585.

Brynjolfsson, E., Hu, Y., Smith M.D, 2003. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. Management Science, 49(11).

Brynjolfsson, Erik, A A. Dick, and M. D. Smith. 2010. A Nearly Perfect Market? Differentiation vs. Price in Consumer Choice. Quantitative Marketing and Economics, 8(1), 1-33.

Chevalier, J., Goolsbee A. 2003. Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. Quantitative Marketing and Economics. 1(2), 203-222.

Clay, K., Krishnan, R., and Wolff, E. 2002. Prices and Price Dispersion on the Web: Evidence from the Online Book Industry. Journal of Industrial Economics, 49(4), 521-539

Curien, N., Fauchart E., Laffond G., Lesourne J., Lainé J. et Moreau F. 2001. "Forums de consommation sur Internet : un modèle évolutionniste. Revue Economique, 52: 119-135.

Elberse, A., Oberholzer-Gee, F. 2007. Superstars and underdogs: An examination of the long tailphenomenon in video sales. Marketing Science Institute, 4, 49-72.

Ellison, G., and Ellison, S. 2005. Lessons about Markets from the Internet. Journal of Economic Perspectives, 19(2), 139-158.

Greffe X. et Sonnac N. 2008. Culture Web - Création, contenus, économie numérique. Dalloz.

Ghose, A., and Sundararajan, A. 2006. Evaluating Pricing Strategy Using eCommerce Data: Evidence and Estimation Challenges. Statistical Science, 21(2).

Ghose, A., Ipeirotis, P. G., and Sundararajan, A. 2009. The Dimensions of Reputation in Electronic Markets. Working Paper.

McKnight L.W. et Bailay J. 1997. Internet Economics. MIT Press.

Houser D. and Wooders J. 2006. Reputation in Auctions: Theory, and Evidence from eBay. Journal of Economics and Management Strategy. 15, 353-369.

Hu E. et Smith M.D. 2003. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. Management Science, 49(11).

Hu E. et Smith M.D. 2011. The Impact of Ebook Distribution on Print Sales: Analysis of a Natural Experiment, mimeo available at http://ssrn.com/abstract=1966115.

Oestreicher-Singer G., Sundararajan A. 2010. Are Digital Rights Valuable? Theory and Evidence from Ebook Pricing. CeDER Working Paper No. 06-01 Working Paper Series.

Pan, X., Ratchford, B. T., and Shankar, V. 2004. Price Dispersion on the Internet: A Review and Directions for Future Research. Journal of Interactive Marketing, 18(4), 116-135.

Peitz M. et Waelbroeck P. 2006. Piracy of Digital Products: A Critical Review of the Theoretical Literature. Information Economics and Policy, 18(4), 449-476.

Resnick, P. and Zeckhauser R. 2002. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System, in Baye M.R. (ed.), The Economics of the Internet and E-Commerce, Advances in Applied Microeconomics, Amsterdam, Elsevier Science, 11: 127-157.

Rochet J. C et Tirole J. 2006. Two-Sided Markets: A Progress Report. The RAND Journal of Economics, 35(3), p. 645-667.

Rosen, S. 1981. The Economics of Superstars. American Economic Review. 71 (5), 845-858.

Spulber D. 1995. Bertrand Competition when Rivals' Costs are Unknown. Journal of Industrial Economics, 43, 1-12.

CHAPITRE 2 : MARCHE INTERNET ET RESEAUX PHYSIQUES. UNE COMPARAISON DES VENTES DE LIVRES EN FRANCE[1]

Bounie D.[2], Eang B.[3] et Waelbroeck P.[4]

2.1. Introduction

Nous comparons les ventes de livres en ligne aux ventes en magasin (hors-ligne) sur la période de mars à août 2006. Cette recherche est motivée par deux questions portant sur l'économie des biens culturels.

La première est liée à la théorie de la longue traîne qui reste aujourd'hui controversée. Selon cette théorie, les outils de recherche et de communication dont disposent les internautes devraient leur permettre d'explorer et de découvrir plus facilement des produits de niche ou des produits dits « obscurs » qui réalisent peu de ventes sur les marchés physiques. En cumulant les ventes issues de ces produits, les sites de commerce électronique seraient alors en mesure de générer des profits suffisants pour conserver et stocker ces produits. En d'autres termes, comparativement aux marchés physiques caractérisés par des ventes élevées d'un très faible nombre de titres et des ventes très faibles d'un très grand nombre de produits (loi de Pareto), la distribution des ventes sur Internet posséderait une structure différente caractérisée par une queue de distribution moins épaisse (avec moins de produits « superstars »). Les plates-formes de vente en ligne, telles qu'Amazon par exemple, soutiennent même l'idée que le commerce électronique constituerait un formidable outil d'accès à la culture, de réduction des prix de vente et de découverte de nouveaux auteurs. L'achat en ligne permettrait aux consommateurs d'acheter plus de livres neufs et favoriserait également une plus grande diversité dans la consommation de biens culturels : il s'agirait d'un effet d'expansion du marché lié à l'existence

[1] Nous tenons à remercier très vivement la société TITE-LIVE qui édite le palmarès des meilleures ventes de livres en France aux professionnels du secteur à travers Edistat. Nous remercions en particulier Elise Tamisier qui nous a permis d'exploiter les données collectées auprès des réseaux physiques de distribution des livres en France.
[2] Telecom ParisTech, Département d'Economie et de Sciences Sociales (SES).
[3] Telecom ParisTech, Département d'Economie et de Sciences Sociales (SES).
[4] Telecom ParisTech, Département d'Economie et de Sciences Sociales (SES).

d'un longue traîne. En dépit de l'intérêt des économistes pour ces questions, peu d'études semblent confirmer l'existence de cette longue traîne (Elberse and Oberzholzer-Gee (2006) s'oppose à Brynjolfsson et *al.* (2003)). Dans le cas français, Benghozi et Benhamou (2008) montrent que même si les albums musicaux et les DVD les moins connus ont vu leurs ventes augmenter au cours des dernières années, le phénomène est de trop faible ampleur pour rentabiliser un modèle d'affaires basé sur la longue traîne.

La deuxième question est liée à la cannibalisation des ventes de biens par le réseau Internet. Le développement de sites de ventes de biens d'occasion sur Internet offre aux consommateurs la possibilité de consulter et de comparer à moindres coûts les prix de biens neufs et d'occasion par rapport à ceux proposés dans les réseaux de distribution traditionnel. La simplicité de l'achat en ligne et la disponibilité d'un catalogue quasi exhaustif ont déjà suscité aux Etats-Unis la méfiance des associations d'auteurs et d'éditeurs qui craignent une « cannibalisation » des ventes de biens neufs par les ventes en ligne, une baisse à terme des revenus des éditeurs et des auteurs et, *in fine*, une réduction des financements liés à la découverte de nouveaux talents. Des premiers éléments de réponse ont été apportés dans le cadre d'un travail économétrique réalisés sur la plate-forme de vente de livres neuf et occasion d'Amazon.com aux Etats-Unis. Les auteurs, Ghose et al. (2006), concluent que les livres d'occasion de qualité dégradée sont des substituts très imparfaits aux livres neufs dans la mesure où seuls 16% des livres d'occasion cannibaliseraient des ventes de livres neufs.

Notre étude poursuit ces deux axes de recherche en analysant les effets d'expansion de marché et de cannibalisation sur le marché du livre en France. L'objectif de ce travail est, d'une part, de caractériser la distribution des ventes de livres dans les marchés hors-ligne et en ligne et, d'autre part, de modéliser la dynamique des interactions entre ces deux marchés. Pour étudier les interactions entre les marchés, nous disposons de deux ensembles de données. D'une part, nous disposons des tops 100 et 500 des meilleures ventes hebdomadaires de livres réalisées auprès d'un échantillon représentatif de commerces en France en 2006 répartis en trois secteurs de distribution : librairies, grandes surfaces spécialisées, grandes surfaces alimentaires. D'autre part, nous disposons du top 100 des meilleures ventes de livres sur Amazon de mars à octobre 2006. Nous sommes donc en mesure de caractériser les

distributions des ventes de livres sur les marchés ainsi que la dynamique des interactions entre les marchés.

Ce travail est le premier à caractériser les lois de distribution des ventes de livres en France et à s'interroger sur la dynamique des interactions entre les marchés physiques et en ligne de livres neufs.[5] Nos analyses montrent premièrement que les distributions des ventes de livre sur les marchés physiques et en ligne sont différentes en France. La comparaison des marchés montre que 20% seulement des livres appartenant aux tops 100 respectifs sont communs sur l'ensemble de la période. Les catalogues en ligne et hors-ligne sont donc différents et suggèrent que les marchés sont plutôt complémentaires. Il existe donc une longue traîne, pas uniquement en terme de catalogue, mais également en terme de ventes, puisque nous analysons les meilleures ventes d'Amazon. Par ailleurs, environ 50% des meilleures ventes (top 100) d'Amazon ne se retrouvent pas dans le top 500 des meilleures ventes de livres sur le marché physique. Nous en concluons que pour ces livres, l'effet d'expansion de marché domine l'effet de cannibalisation. Deuxièmement, sur les 20% de livres communs aux deux plates-formes, nous observons pour 58% d'entre eux des décalages temporels d'entrée des livres dans les tops des ventes. Les livres concernés appartiennent à des genres littéraires précis. Nous utilisons un modèle vectoriel autorégressif (VAR) et montrons que les ventes en librairies ont un effet positif sur les ventes en ligne alors que les ventes en ligne n'ont qu'un effet très marginal sur les ventes réalisées dans les réseaux physiques. L'effet de prescription d'Internet sur les ventes physiques semble encore limité.

Il est intéressant de mettre en perspective nos résultats par rapport à ceux de Benhamou (2008) qui étudie la rentrée littéraire en 2005 de 693 titres appartenant essentiellement à des romans littéraires (romans de la littérature française qui ne sont pas au format de poche). L'auteur montre que les ventes sont très concentrées : 16% des titres représentent 83% des ventes. Notre étude complète ce travail en établissant que ce genre littéraire est le moins susceptible de générer un phénomène de longue traîne par rapport à d'autres genres comme celui des sciences, religion et politique.

[5] Nous rappelons que Ghose et *al*. (2006) étudient l'effet de cannibalisation du marché du neuf par le marché de l'occasion (qualité différente du neuf).

L'article est organisé de la manière suivante. Dans la section 2, nous décrivons les données Dans la section 3, nous comparons la distribution des ventes physiques à celle des ventes sur Amazon MarketPlace et nous identifions les livres qui appartiennent à la fois au top 100 des ventes en ligne et à celui des ventes physiques sur la même période. Dans la section 4, nous analysons les interactions entre le marché physique et le marché en ligne.

2.2 Méthodologie et description des données

Nous utilisons deux bases de données originales sur les 100 et 500 meilleures ventes de livres en France dans les réseaux physiques et sur les 100 meilleures ventes de livres sur Internet. Nous avons suivi plus précisément l'évolution de ces meilleures ventes de mars à août 2006, soit une période de 22 semaines.

La première base de données relative aux meilleures ventes de livres sur Internet (ou en ligne) a été constituée à partir des données publiques sur l'un des premiers sites de ventes de biens culturels en France : le site Amazon.fr. L'étude de la plate-forme Amazon est désormais devenue une référence dans les milieux académiques internationaux (Ghose et *al*., 2006) dans la mesure où le site est implanté dans de nombreux pays, et dans la mesure également où il héberge une plate-forme d'achats/ventes de livres neuf et d'occasion entre professionnels et/ou particuliers (MarketPlace). Cette plate-forme est intéressante pour les particuliers car le prix d'un livre neuf commercialisé par Amazon peut être directement comparé aux prix des livres neufs et d'occasion proposés sur la plate-forme par d'autres vendeurs et, ensuite, achetés en ligne. L'originalité de notre approche est d'avoir collecté les données du site Amazon.fr de manière quotidienne sur la période allant de mars à août 2006. Nous avons plus précisément identifié l'entrée d'un produit dans le top 100 des meilleures ventes de livre d'Amazon et nous avons suivi l'évolution de son offre sur la plate-forme au cours de 22 semaines. Dans cette base, chaque livre est décrit par trois ensembles d'information. Le premier est lié aux caractéristiques du livre telles que l'ISBN, le titre, l'éditeur, la langue, la date de sortie, le genre littéraire, le classement des ventes sur le site Amazon (SalesRank), la note moyenne attribuée au livre par les consommateurs, le nombre de notes, le prix proposé par Amazon, etc. Le deuxième

ensemble d'information a trait aux caractéristiques des différents exemplaires à la vente d'un même livre disponible sur la plate-forme MarketPlace comme la qualité du livre (état de cet exemplaire : neuf, comme neuf, etc.), le prix, etc. Enfin, la troisième source d'information concerne le vendeur de chaque exemplaire sur la plateforme d'occasion comme le nom du vendeur, la réputation du vendeur, le nombre d'exemplaires vendus par le vendeur, etc.

La deuxième base de données a trait au classement des 500 meilleures ventes de livres dans les réseaux physiques en France sur la période de mars à août à 2006. Les ventes sont estimées à partir des relevés de caisses issus d'un panel de près de 1200 magasins en France métropolitaine (hors Corse). Ce panel est représentatif du poids des différents circuits de distribution dans la vente de livres neufs aux particuliers (hors Club, VPC et Internet). La représentativité de l'échantillon est calculée selon les données du DEP-Ministère de la Culture (TN-Sofres pour l'Observatoire de l'économie du livre). Pour chaque livre appartenant aux 100 meilleures ventes de livre, nous disposons d'informations sur le genre littéraire (en 7 classes), le prix conseillé, l'éditeur, l'auteur, le total des ventes hebdomadaires et le total des ventes par réseau de distribution. Trois réseaux de distribution sont en particulier distingués : les librairies, les grandes surfaces alimentaires et les grandes surfaces spécialisées.

L'objectif est de comparer le top 100 des meilleures ventes de livres en ligne au top 100 des meilleures ventes de livres en magasins physiques sur une période de 22 semaines allant de Mars à Août 2006.[6] Dans la mesure où nous disposons des meilleures ventes de livres en ligne par semaine dans les réseaux physiques et par jour sur Amazon, nous avons harmonisé les données en calculant le classement moyen des ventes hebdomadaires d'un livre en ligne à partir de ses classements journaliers.[7] Cette opération effectuée, nous avons ensuite recodé les classements moyens pour créer un top 100 hebdomadaire (pour que le top 100 prenne des valeurs allant de 1 à 100 sur la semaine).

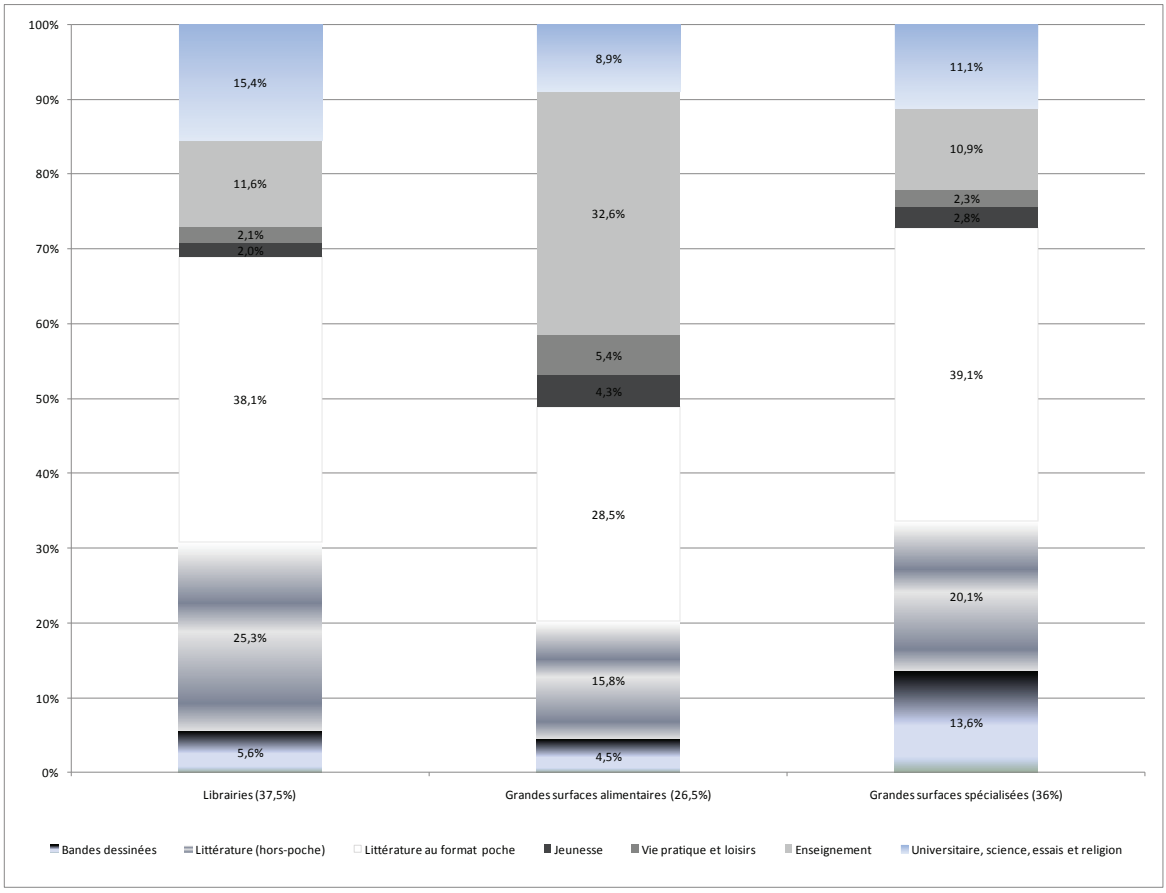
L'analyse descriptive des données montre tout d'abord que 488 livres différents sont entrés dans le top 100 hors-ligne pour un volume total de ventes de 11 980 647 exemplaires. En comparaison, 765

[6] Ces semaines d'observation ne tiennent pas compte des périodes de vente exceptionnelles de fin d'année.

[7] Par exemple, si un livre est numéro 1 le premier jour, numéro 2 le 2ème jour, numéro 3 le 3ème jour, etc., nous calculons le rang moyen de ce livre sur la semaine. Pour les livres qui entrent dans le top 100 au cours de la semaine, nous calculons la moyenne sur la durée de présence hebdomadaire.

livres sont entrés dans le top 100 sur Amazon.fr. La dynamique d'entrée-sortie des livres semble donc beaucoup forte sur Internet. Les ventes de livres hors-ligne se répartissent principalement sur 7 genres littéraires.[8] Les genres les plus vendus sont la littérature au format de poche (36%), la littérature (hors-poche) (21%), les manuels d'enseignement (17%), les livres « universitaire, science, essais et religion » (12%), les bandes-dessinées (8%), les livres de la « vie pratique et loisirs » (3%) et, enfin, la littérature jeunesse (3%) (voir Tableau 9 en annexe). Nous observons dans le Graphique 1 que les ventes de livres hors-ligne varient sensiblement par réseau de distribution. Ainsi, les manuels de l'enseignement sont distribués en plus grande proportion dans les grandes surfaces alimentaires (32,6%) (devant la littérature au format de poche (28,5%)) alors que les bandes-dessinées sont en grande partie vendues en grande surface spécialisées (13,6%). Proportionnellement, le genre « Universitaire, essais et religion » est plus distribué dans les librairies (15,4%).

[8] Il y avait initialement 8 genres. Nous avons fusionné les genres "Autres" et "Vie pratique et loisirs " en raison des très faibles effectifs.



Graphique 1 : Pourcentage des ventes par réseau de distribution

La dynamique des entrées-sorties dans le top 100 en ligne se traduit directement, ainsi que l'indique le Tableau 1, par une durée de présence moyenne des livres dans le top 100 beaucoup plus faible. Ainsi, alors que la durée moyenne hebdomadaire d'un livre dans le top 100 hors-ligne est de 4,8 semaines, elle n'est que de 2,5 semaines en ligne, soit la moitié. Le Tableau 1 donne les statistiques descriptives pour les variables les plus importantes. Les bandes-dessinées restent en moyenne 2,3 semaines dans le top 100 sur Amazon.fr contre 2,7 semaines sur les marchés physiques. Globalement, la répartition des genres est similaire en ligne et hors-ligne. Nous constatons cependant que la littérature au format de poche se caractérise par des ventes très importantes dans les marchés physiques et sur une période de temps beaucoup plus longue par rapport par exemple aux bandes dessinées.

Réseaux physiques								
	Bandes dessinées	**Littérature (hors-poche)**	**Jeunesse**	**Enseignement**	**Universitaire, science, essais et religion**	**Littérature au format poche**	**Vie pratique et loisirs**	**Total**
Durée moyenne dans le top 100 (en semaine)	2,7	5,5	3	3,3	4,9	7,1	5,3	4,8
Nb de livres	78	78	32	103	56	126	15	488
Prix moyen	8,5	19,7	9,2	8,9	16,5	6,7	10,8	11
Classement moyen	50,6	47,3	63,5	49,3	52,7	51,2	52	50,9
Vente moyenne	4640	5799	3635	5988	5286	4823	4586	5155
Total des ventes	983 738	2 505 168	345 321	2 023 896	1 453 607	4 302 021	366 896	11 980 647
Internet (Amazon.fr)								
Durée moyenne dans le top 100 (en semaine)	2,3	3,6	2,3	1,9	2,7	3,6	2,4	2,5
Nb de livres	67	155	60	34	131	226	91	765
Prix moyen	8,8	15,2	7,1	15,7	16,8	5,8	16,2	13,7
Classement moyen	58,9	56,0	66,0	61,4	50,1	60,9	54,6	56,9

Tableau 1 : Statistiques descriptives

2.3 Comparaison des ventes dans les réseaux physiques et en ligne

Dans cette partie, nous cherchons premièrement à caractériser les distributions des ventes de livres dans les marchés physiques et sur Internet. Deuxièmement, nous identifions les livres qui sont communs aux meilleures ventes des deux marchés. A l'exception du travail de Benghozi et Benhamou (2008) sur le marché des DVD et CD, il n'existe pas de travaux similaires à notre connaissance sur le marché du livre.

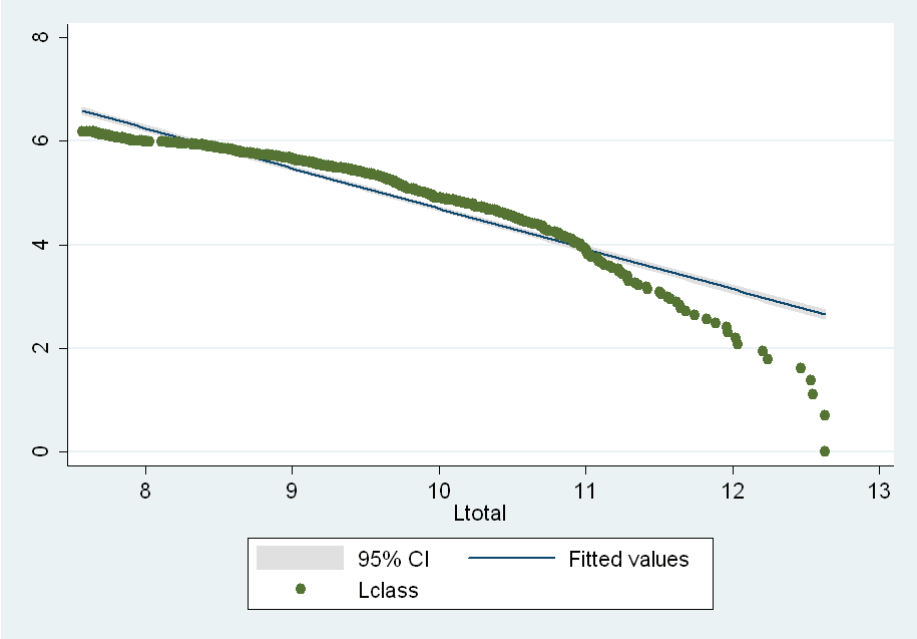
2.3.1. Caractérisation de la distribution des ventes

Pour connaître le type de distribution des ventes dans les réseaux hors-ligne et, en particulier, pour estimer la loi de distribution des ventes, nous calculons les ventes totales pour chaque livre sur l'ensemble de la période. Les résultats pourront être comparés à ceux de Ghose et Sundararajan (2006) qui suggèrent que la distribution des ventes en ligne suit une loi de Zipf avec une queue de distribution

de type Pareto (c'est-à-dire proportionnelle à une loi de type puissance). Il est facile de tester si les données sont générées par ce type de distributions, car les lois de Zipf et de Pareto possèdent comme propriété une relation linéaire entre le logarithme du classement des ventes et le logarithme des ventes (en unités). Pour faire cette comparaison, nous calculons le classement des ventes (allant de 1 à 488) et nous faisons une régression du logarithme de ce classement sur le logarithme des ventes. Nous obtenons le résultat suivant :

$$\text{Log(classement)} = 12.44 - 0.77 \log(\text{ventes}),\ R^2 = 0.89,\ n = 488 \quad (1)$$
$$(0.12)\ \ (0.01)$$

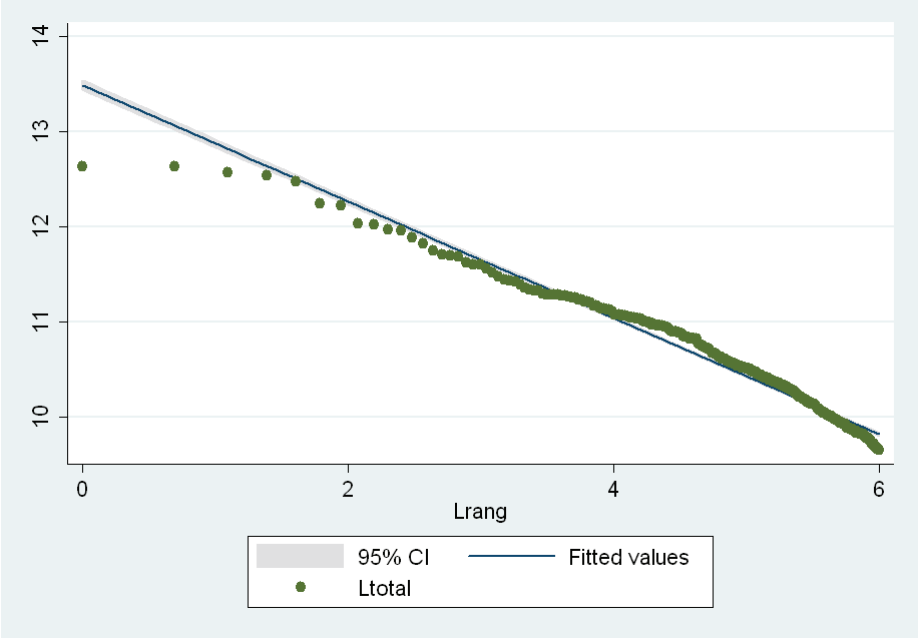
La qualité de l'ajustement indique une bonne approximation par une loi de type puissance (Pareto). Cependant le Graphique 2 montre une déviation par rapport à la loi de Pareto pour les meilleures ventes (à la droite du graphique).



Graphique 2 : Régression log du classement – log des ventes sur la totalité des livres hors-ligne

Si l'on estime à nouveau l'équation (1) uniquement sur les classements des livres les plus faibles (c'est-à-dire sur les meilleure ventes, ici inférieur à 6 en logarithme), nous obtenons une pente plus forte avec un coefficient égal à -1.63. Même si la qualité de l'ajustement linéaire est dans ce cas de 0.97 dans son ensemble, on constate une déviation par rapport à la droite de régression pour les

meilleures ventes (Graphique 3 avec le logarithme des ventes en ordonnée, le logarithme des ventes en abscisse et une pente de -0.6128 = 1/-1.63).



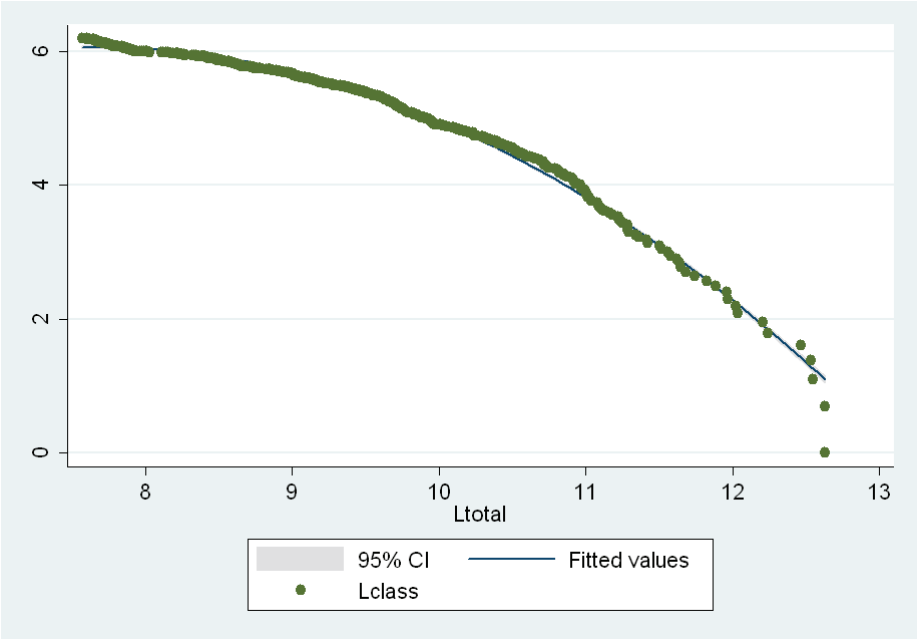
Graphique 3 : Régression log du classement – log des ventes sur les meilleures ventes de livres hors-ligne

Nous avons donc rajouté un terme quadratique. Les résultats de la régression indiquent un effet non linéaire qui diminue l'importance des valeurs extrêmes par rapport à la loi de Pareto.

$$\text{Log(classement)} = -5.7 + 3.07\ \log(\text{ventes}) - 2.01\ \log(\text{ventes})^2,\ R^2 = 0.99,\ n = 488$$
$$(0.21)\ (0.02)\qquad (0.05)$$

L'ajustement quadratique est représenté dans le Graphique 4 et est nettement meilleur que l'ajustement linéaire.[9]

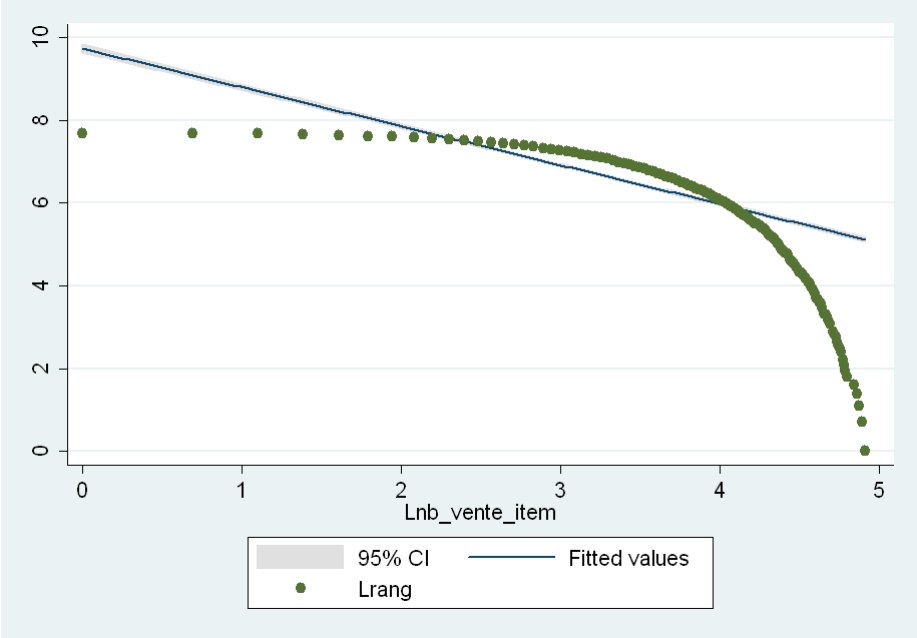
[9] La caractérisation exacte de cette loi est l'objet d'un travail de recherche indépendant.



Graphique 4 : Régression log du classement – log des ventes au carré sur les meilleures ventes de livres hors-ligne

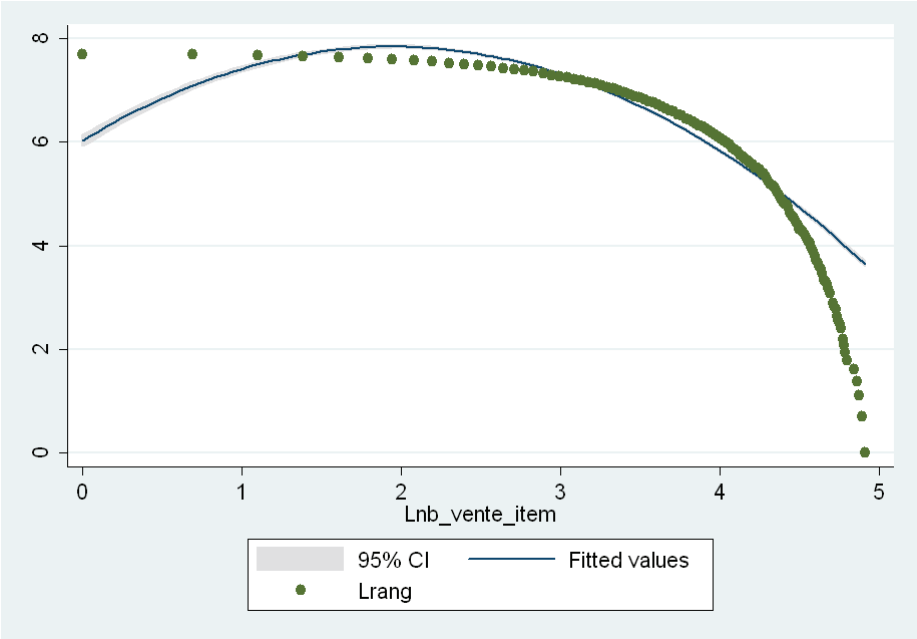
La caractérisation de la distribution des ventes de livres en ligne est moins aisée dans la mesure où il n'existe pas de données agrégées publiques sur le marché du livre. Pour contourner ce problème, certains auteurs ont utilisé des procédures artificielles pour évaluer les ventes de livres en ligne. Goolsbee et Chevalier (2005) utilisent par exemple une procédure qui consiste à acheter une petite quantité d'un livre sur un intervalle de temps très bref pour mesurer la manière dont le classement des ventes augmente. En utilisant cette procédure, ils déduisent que le coefficient de Pareto des ventes en ligne est de l'ordre de -0.82. De manière similaire, Brynjolfsson, Smith et Hu (2003) utilisent les ventes d'un libraire particulier pendant trois semaines pour estimer les coefficients de la régression log-log à 10.526 et -0.87 (en utilisant le classement des ventes Amazon). Mais utiliser cette stratégie pour estimer la distribution des ventes en ligne semble hasardeux dans la mesure où la relation entre le logarithme des ventes et le logarithme du classement des ventes semble linéaire. Nous avons opté pour une autre approche. Nous pouvons calculer une mesure approchée du nombre de transactions associées à chaque livre sur la plateforme MarketPlace qui rassemble des vendeurs professionnels sur le site d'Amazon. Pour ce faire, nous avons relevé l'ensemble des offres postées des vendeurs de livres neufs identifiés dans le top 100 en ligne sur la même période. A partir de ces relevés, nous pouvons tenter de caractériser les ventes de livres. Nous considèrerons, à la lumière du travail de Smith et *al*. (2008),

qu'une transaction a lieu si un livre disponible à la date *t* chez un vendeur disparait à la date *t+1* et ne revient plus ou revient au plus tôt à la date *t+2*.[10] A l'évidence, les ventes sur la plateforme MarketPlace ne sont pas nécessairement corrélées avec celles d'Amazon ou celles des réseaux physiques, car les internautes qui achètent par l'intermédiaire de MarketPlace souhaitent des prestations différentes en termes de disponibilité immédiate et de qualité de service. Il est néanmoins instructif de caractériser la distribution des ventes sur la plateforme MarketPlace. Pour ce faire, nous estimons les mêmes paramètres que ceux définis pour les ventes de livres dans les réseaux physiques. Les résultats des estimations suggèrent des queues de distribution moins épaisses (Graphique 5 et Graphique 6). Ceci illustre le potentiel de la longue traîne sur le marché des ventes en ligne où l'on observe des ventes plus égalitaires et moins concentrées que sur les marchés physiques.



Graphique 5 : Régression log du classement – log des ventes sur la totalité des livres en ligne (ajustement linéaire)

[10] Pour éviter d'attribuer une transaction lorsqu'il y a un bug de mise à jour du programme automatique qui construit les données, nous avons supprimé les jours pour lesquels moins de 97% des livres n'ont pas été mis à jour.



Graphique 6 : Régression log du classement – log des ventes sur les meilleures ventes de livres en ligne (ajustement quadratique)

Ce travail pourra servir de référence pour les recherches futures sur la comparaison de distributions des ventes en ligne et hors-ligne.

2.3.2. Les marchés des livres en ligne et hors ligne sont-ils comparables ?

Nous comparons à présent les catalogues de livres des meilleures ventes en ligne et dans les réseaux physiques. L'objectif est de déterminer, d'une part, si les préférences des consommateurs en ligne et hors ligne sont identiques et si, d'autre part, il existe une cannibalisation des ventes entre les marchés. Cette question, à notre connaissance, n'a pas été étudiée dans la littérature qui s'est concentrée principalement sur l'effet des livres d'occasion (qualité différente de l'état neuf) sur le marché du livre neuf (Ghose, Smith et Telang, 2006). Nous comparons plus précisément les catalogues sur deux dimensions temporelles : les catalogues sur l'ensemble de la période et les évolutions des catalogues semaine après semaine.

Sur un échantillon total de 1055 livres uniques commercialisés dans les réseaux physiques et sur le site d'Amazon.fr dans les tops 100 respectifs sur l'ensemble de la période d'observation, nous

trouvons que 198 titres sont communs aux deux marchés, soit 19% de livres en commun. Ce résultat est surprenant dans la mesure où il s'agit des cent meilleures ventes de livres en France. Nous observons dans le Tableau 2 que ces livres sont en majorité du genre « Littérature au format de poche » (27,8%), « Littérature (hors-poche) » (22,2%), « Universitaire, science, essais et religion » (19,7%) et « bandes dessinées » (18,7%).

Genre	Nombre de livres	%
Bandes dessinées	37	18,7%
Littérature (hors-poche)	44	22,2%
Jeunesse	7	3,5%
Enseignement	12	6,1%
Universitaire, science, essais et religion	39	19,7%
Littérature au format poche	55	27,8%
Vie pratique et loisirs	4	2,0%
Total	198	100,0%

Tableau 2 : Genre littéraire des livres communs aux tops en ligne et hors-ligne

La singularité de ce résultat est confirmée par l'étude du nombre de livres communs qui appartient au top 100 Amazon.fr et au top 500 des meilleures ventes de livres dans les réseaux physiques. Dans le Tableau 3, nous constatons que 50% des livres qui appartiennent au top 100 d'Amazon.fr sont présents dans le top 500 des meilleures ventes de livres dans les réseaux physiques. La répartition par genre littéraire est sensiblement équivalente entre le top 100 et le top 500 des meilleures ventes de livres dans les réseaux physiques. Cette analyse des catalogues semble donc confirmer dans un premier temps que les préférences des consommateurs en ligne et hors ligne sont très différentes et que la demande en ligne se porte sur des titres qui ne font pas partis des meilleures ventes dans les réseaux physiques.

	Part des livres entrés dans le top 100 Amazon		
Genre littéraire	**qui entre dans le Top 100 hors-ligne sur l'ensemble de la période**	**qui entre dans le Top 500 hors-ligne sur l'ensemble de la période**	**qui n'entre pas dans le Top 500 hors-ligne sur l'ensemble de la période**
Bandes dessinées	18,7%	15,4%	1,9%
Littérature (hors-poche)	21,7%	19,0%	21,6%
Littérature au format poche	27,8%	29,5%	29,6%
Jeunesse	3,5%	6,2%	9,6%
Vie pratique et loisirs	2,0%	5,6%	18,7%
Enseignement	6,1%	4,9%	4,0%
Universitaire, science, essais et religion	20,2%	19,5%	14,7%
Nombre de livres	198	390	375

Tableau 3 : Etude du Top 100 Amazon.fr et du Top 100 et 500 dans les réseaux physiques

Pour poursuivre cette analyse, il est intéressant d'étudier le volume total des ventes effectuées par les livres présents dans le top 100 Amazon.fr et dans les tops 100 et 500 des ventes sur les marchés physiques. Tout d'abord, l'analyse respective des tops 100 montre que si le nombre commun de livres sur les deux marchés est faible (198 livres (19%)), en revanche, les ventes de ces livres représentent 68% des ventes totales de livres sur les réseaux physiques. Une partie significative des ventes dans les réseaux physiques peut donc exister également sur le réseau Internet. L'analyse du top 100 des ventes sur Amazon.fr et du top 500 des ventes dans les réseaux physiques montre que les ventes totales des 391 livres communs s'élèvent à 48% des ventes dans les réseaux physiques. Nous pouvons donc conclure que 50% des livres qui entrent dans le top 100 Amazon.fr sur la période et qui n'entrent pas dans le top 500 hors-ligne réalisent peu de ventes dans les réseaux physiques puisqu'un livre classé au rang 500 hors-ligne réalise en moyenne moins de 700 ventes par semaine et sort rapidement du top 500. Ce résultat confirme donc l'idée qu'une proportion significative des ventes de livres sur le site Amazon.fr correspond à des produits de niche dans les réseaux physiques qui génèrent peu de ventes. Le marché électronique de livres sur Internet semble donc en partie être un marché complémentaire plutôt qu'un marché substitut.

Ceci ne permet cependant pas de conclure que les livres qui font beaucoup de ventes dans les réseaux physiques font également beaucoup de ventes en ligne ni que les livres qui font peu de ventes

hors-ligne ne font pas de ventes sur le site Amazon.fr. En effet, lorsque nous étudions le catalogue des livres présents dans le top 100 Amazon.fr qui ne figurent pas dans le top 500 hors-ligne, nous constatons que les préférences des consommateurs s'orientent sur des genres peu commercialisés dans les réseaux physiques ; par exemple, 19% des livres présents dans le top 100 Amazon.fr qui ne figurent pas dans le top 500 hors-ligne ont trait à la « Vie pratique et loisirs » (santé, tourisme, etc.). Nous pouvons donc supposer que les ventes de certains genres littéraires en ligne sont plus importantes en proportion que dans les réseaux physiques. Pour donner un ordre d'idée, il est intéressant de remarquer par exemple que le classement moyen des livres présents dans le top 100 Amazon.fr qui ne figurent pas dans le top 500 hors-ligne est de 57, c'est-à-dire que ces livres se situent dans la moyenne des classements des livres les plus vendus sur Amazon.fr.

Pour conclure sur ce point, nous pouvons mettre en perspective nos résultats par rapport à ceux de Benhamou (2008) qui étudie la rentrée littéraire en 2005 de 693 titres appartenant essentiellement à des romans littéraires (romans de la littérature française qui ne sont pas au format de poche). L'auteur montre que les ventes sont très concentrées : 16% des titres représentent 83% des ventes. Notre étude complète ce travail en établissant que ce genre littéraire est le moins susceptible de générer un phénomène de longue traîne par rapport à d'autres genres, sachant que le genre littérature française ne représente que 20% des meilleures ventes de nos données.

2.4 Interactions entre les catalogues de livre en ligne et des réseaux physiques semaine après semaine

Au-delà de l'étude du catalogue sur l'ensemble de la période d'étude, il est intéressant de caractériser les cycles de vie des marchés en étudiant par exemple le sous-ensemble des livres communs aux deux marchés semaine après semaine. Cette étude peut permettre de préciser les effets temporels des effets potentiels de cannibalisation.

Une première analyse descriptive montre que sur les 198 livres, 84 sont présents durant la même semaine. En d'autres termes, pour 42% des livres communs aux deux marchés, nous observons une synchronisation des cycles de vie des marchés. Pour les autres, nous observons que 53 livres entrent

d'abord dans le top hors-ligne (soit 26,7%) et ensuite dans le top en ligne avec un retard moyen de 3,8 semaines. Enfin, nous constatons que 61 livres entrent en premier dans le top 100 des meilleures ventes de livres en ligne et ensuite dans le top 100 hors-ligne avec un retard moyen de 2,2 semaines (30,8%). En d'autres termes, nous constatons que la non-synchronisation des catalogues dans le temps est à l'avantage des marchés en ligne pour 30,8% des livres communs et précédent l'entrée dans le top 100 hors-ligne en moyenne de 2 semaines.

Il est intéressant de caractériser les deux sous-échantillons de livres qui apparaissent de manière décalée dans le temps. Pour les livres qui entrent d'abord dans le top hors-ligne et qui ensuite apparaissent dans le top en ligne, nous constatons que ce sont principalement des livres du genre « Littérature au format poche » et « Littérature (hors-poche) ». En d'autres termes, les livres de ces genres littéraires sont préférentiellement achetés en premier dans les réseaux physiques puis sur Internet. Remarquons également que ces genres littéraires font partis des genres les plus commercialisés dans les réseaux physiques.

Genre	Nombre de livres	%
Bandes dessinées	7	13%
Littérature (hors-poche)	13	25%
Jeunesse	1	2%
Enseignement	3	6%
Universitaire, science, essais et religion	6	11%
Littérature au format poche	20	38%
Vie pratique et loisirs	3	6%
Total	53	100%

Tableau 4 : Genre littéraire des livres communs qui entrent en premier hors ligne

Il est intéressant de remarquer également que parmi certains livres au format poche, une partie est en premier achetée sur Internet. Cet effet est sans doute du à l'existence de sous-catégories dans les livres de poche qui pour certains constituent des rééditions ou bien des premières éditions. Nous observons également avec surprise que les bandes dessinées font parties également des genres littéraires qui sont d'abord achetés sur Internet puis dans les réseaux physiques.

Genre	Nombre de livres	%
Bandes dessinées	13	21,3%
Littérature (hors-poche)	11	18,0%
Jeunesse	2	3,3%
Enseignement	7	11,5%
Universitaire, science, essais et religion	13	21,3%
Littérature au format poche	14	23,0%
Vie pratique et loisirs	1	1,6%
Total	61	100,0%

Tableau 5 : Genre littéraire des livres communs qui entrent en premier en ligne

Nous pouvons enfin caractériser les deux sous-échantillons par rapport au réseau de distribution (librairies, grandes surfaces spécialisées, grandes surfaces alimentaires). Nous constatons que lorsque les livres entrent d'abord dans le top 100 hors-ligne et ensuite dans le top 100 en ligne, les ventes totales de ces livres sont plus importantes dans les librairies (45%) et dans les grandes surfaces spécialisées (37%). En revanche, lorsque les livres entrent en premier dans le top en ligne et sont ensuite vendus dans les réseaux physiques, ces livres sont vendus en plus grande proportion dans les grandes surfaces spécialisées. En croisant ces résultats par rapport aux précédents sur les genres littéraires, nous constatons donc que les librairies ont un effet de prescription sur les livres au format poche et sur les livres de la littérature (hors-poche) alors que le réseau Internet semble être un relai de ventes pour les bandes dessinées, les livres universitaires, sciences, essais et religion ainsi qu'une partie des livres au format poche pour les grandes surfaces spécialisées et les librairies.

	Librairies	Grandes surfaces spécialisées	Grandes surfaces alimentaires	Total
Le offline arrive plus tard				
Volume des ventes	755145	855232	454936	2065313
%	37	41	22	100
Le online arrive plus tard				
Volume des ventes	1028309	851177	427778	2307264
%	45	37	19	100
Total des ventes	4489104	3174137	4317406	11980647

Tableau 6 : Décalage dans le temps des catalogues et réseau de distribution

Ces effets sur les décalages temporels entre les catalogues peuvent être estimés plus systématiquement dans le cadre d'un modèle vectoriel autorégressif (VAR). Pour ce faire, nous faisons une régression de la variation hebdomadaire (en pourcentage) du logarithme du classement d'un livre sur Amazon.fr sur sa variation de la semaine précédente et la variation des ventes en pourcentage de ce même livre sur le marché physique (semaine précédente pour éviter les problèmes d'endogénéité). Soit y1*it* les ventes en ligne et y2*it* les ventes hors-ligne du livre *i* à la date *t* (pour *i* = 1, ..., *N* et *t* = 2, ..., *T*). Nous considérons le modèle vectoriel autorégressif (VAR) suivant :

$$y_{1it} = \alpha_{1t} + \beta_{11}\, y_{1it-1} + \beta_{12}\, y_{2it-1} + \varepsilon_{1it}$$

$$y_{2it} = \alpha_{2t} + \beta_{21}\, y_{1it-1} + \beta_{22}\, y_{2it-1} + \varepsilon_{2it}$$

Nous prenons le modèle en différence :

$$\Delta y_{1it} = \beta_{11}\, \Delta y_{1it-1} + \beta_{12}\, \Delta y_{2it-1} + \eta_{1it}$$

$$\Delta y_{2it} = \beta_{21}\, \Delta y_{1it-1} + \beta_{22}\, \Delta y_{2it-1} + \eta_{2it} \quad (2)$$

Où $\Delta y_{jit} = \Delta y_{jit} - \Delta y_{jit-1}$ et $\eta_{jit} = \varepsilon_{jit} - \varepsilon_{jit-1}$ (j=1,2) sont i.i.d. de moyenne nulle et de variance σ^2 constante. Les paramètres de l'équation (2) sont estimés par la méthode des moindres carrés ordinaires.

Si l'on suppose une relation « linéaire » entre le logarithme du classement et le logarithme des ventes (voir la section précédente), la variable dépendante est le négatif de la variation des ventes en ligne (en pourcentage). Le modèle est pris en différence et donc élimine les effets fixes propres à chaque livre. Etant donné que certains livres ne restent que quelques semaines dans le top 100, nous avons restreint l'échantillon aux livres qui sont restés plus de 4 semaines dans notre échantillon. Ceci élimine une variabilité artificielle liée aux livres peu populaires. Au final, nous disposons de 340 observations.

	Δlog (class.)$_t$				Δ ventes $_t$			
	(A)		(B)		(C)		(D)	
Variable	Coeff.	Ecart-type	Coeff.	Ecart-type	Coeff.	Ecart-type	Coeff.	Ecart-type
Δlog (class.)t-1	-0,322	0,072	-0,323	0,072	-0,056	0,035	-0,071	0,037
Δlog (class.)t-2							-0,073	0,045
Δventes t-1	-0,214	0,079	-0,215	0,079	0,043	0,05	-0,078	0,056
Δventes t-2			0,012	0,071				
constante	0,087	0,017	0,087	0,017	-0,54	0,11	-0,047	0,012
R^2	0,06		0,06		0,01		0,02	
# obs.	340		340		385		343	

Tableau 7 : Résultats d'estimation du modèle VAR.

Les résultats de la régression dans le Tableau 7 indiquent qu'une augmentation des ventes en ligne de 1% réduit les ventes en ligne de 0.32% % la semaine suivante (interprétation : classement relatif) et augmente les ventes en magasins physiques de 0.21%. A l'opposé, l'estimation de la seconde équation du modèle (2), indique que l'effet des ventes en ligne sur les ventes physiques est significativement positif mais de faible amplitude (colonnes C et D).

Cette estimation peut être complétée par une étude des réseaux de distribution. Dans le Tableau 8, nous synthétisons les résultats de l'estimation des paramètres du modèle (2) dans lequel on cherche à expliquer l'effet de la variation des ventes en ligne d'une semaine sur les ventes physiques par réseau de distribution. Nous considérons un ou deux retards (modèle A et B) dans la spécification du modèle (2).

	Δlog (class.)t			
	(A)		(B)	
Variable	Coeff.	Ecart-type	Coeff.	Ecart-type
Δ log (class.)t-1	-0,296	0,072	-0,305	0,072
Δ ventes Libraires t-1	-0,278	0,091	-0,383	0,102
Δ ventes GSA t-1	-0,063	0,046	-0,061	0,046
Δ ventes GSS t-1	0,142	0,092	0,212	0,097
Δ ventes Libraires t-2			-0,225	0,100
Δ ventes GSA t-2			-0,001	0,002
Δ ventes GSS t-2			0,179	0,091
constante	0,091	0,017	0,092	0,017
R^2	0,07		0,09	
# obs.	340		340	

Tableau 8 : Estimations par réseau de distribution.

Lorsque l'on détaille les résultats par réseau de distribution (Tableau 8), on constate que le lien positif entre l'augmentation des ventes de livres en ligne et l'augmentation des ventes de livres hors-ligne joue principalement à travers les ventes en librairies (effet cumulé de 0.6 dans la colonne B). Nous pouvons interpréter ce résultat de deux manières. Premièrement, un lecteur intéressé par un ouvrage peut se déplacer chez le libraire qui doit commander le livre pour raison d'indisponibilité et décider de commander plutôt en ligne (problème d'inventaire, en particulier pour les bandes dessinées qui se vendent en série). Deuxièmement, un lecteur peut avoir découvert un livre chez le libraire et en parler à ses amis qui le commandent en ligne (bouche à oreille : en particulier pour les livres universitaires, les essais, etc. qui sont surreprésentés en termes de ventes totales chez les libraires). Enfin, si l'on rajoute un délai supplémentaire pour les réseaux de distribution dans le modèle (2), on met en évidence un effet négatif des grandes surfaces spécialisées sur les ventes en ligne (avec un effet cumulé assez fort de 0.4, voir Tableau 8). Ceci pourrait s'expliquer par une concurrence plus forte entre les grandes surfaces spécialisées et les sites de commerces électroniques de produits culturels.

2.5 Conclusion

Cet article avait pour objectif de comparer les ventes de livres en ligne (sur Amazon.fr) aux ventes en magasin (hors-ligne) sur la période de mars à août 2006 en France. Notre travail montre premièrement qu'il existe une différence entre les meilleures ventes en ligne et celles réalisées dans les réseaux physiques : nos résultats indiquent que l'effet positif d'expansion liée à la longue traîne domine l'effet négatif de cannibalisation des ventes physiques par les ventes en ligne. Nous observons également des décalages temporels d'entrée des livres dans les tops des ventes. Les librairies semblent jouer un rôle prescripteur pour les ventes de livres du genre « Universitaire, essais, science et religion». Nos résultats pourront être répliqués dans d'autres pays comme l'Angleterre ou l'Allemagne qui différent sur le plan institutionnel et sur les usages d'Internet en vue de déterminer quels sont les réseaux de distribution les plus prescripteurs pour les ventes en ligne.

2.6 Références

Benghozi, P-J., Benhamou, F. (2008), "Longue traîne : levier numérique de la diversité culturelle ?", Culture Prospective, 2008-1

Benhamou, F., Éléments de réflexion et propositions à propos de la diversité culturelle. Les cas du livre et du cinéma en salles, rapport pour le Groupe d'analyse stratégique des industries culturelles (Gasic), Paris, DEPS, Ministère de la Culture et de la Communication, janvier 2008.

Brynjolfsson, E., Hu, Y., Smith, M. (2003), "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," Management Science, Vol. 49 (11), 1580-1596

Elberse, A., Oberholzer-Gee, F. (2006), "Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales." Harvard Business School Working Paper, No. 07-015

Ghose, A., Smith, M., Telang, R. (2006), "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," Information Systems Research, 17(1), 3-19.

Ghose, A., and A. Sundararajan. 2006. Evaluating Pricing Strategy Using eCommerce Data: Evidence and Estimation Challenges, *Statistical Science*, 21(2), 131-142.

Goolsbee, A., Chevalier, J. (2005), "Measuring Prices and Price Competition Online: Amazon vs. Barnes and Noble," *Quantitative Marketing and Economics* I, Vol. 2, June 2003

Smith, M., Telang, R. (2008), "Internet Exchange for Used Digital Goods," Working paper, Carnegie Mellon University

2.7 Annexe

	Librairies 37,5%	Grandes surfaces alimentaires 26,5%	Grandes surfaces spécialisées 36%	Total
Bandes dessinées	5,6%	4,5%	13,6%	8,2%
Littérature (hors-poche)	25,3%	15,8%	20,1%	20,9%
Littérature au format poche	38,1%	28,5%	39,1%	35,9%
Jeunesse	2,0%	4,3%	2,8%	2,9%
Vie pratique et loisirs	2,1%	5,4%	2,3%	3,0%
Enseignement	11,6%	32,6%	10,9%	16,9%
Universitaire, science, essais et religion	15,4%	8,9%	11,1%	12,1%
Total (%)	100,0%	100,0%	100,0%	100,0%
Total (nombre)	4 489 104	3 174 137	4 317 406	11 980 647

Tableau 9 : Pourcentage des ventes par genre littéraire et réseau de distribution

ISBN	Titre	Genre	Auteur	Date de parution	Prix	Total des ventes	Ventes librairies	Ventes grandes surfaces alimentaires	Ventes grandes surfaces spécialisées
226614434	Da Vinci Code	Littérature format poche	Dan Brown	19/04/2005	7	306643	88114	96856	121673
287858233	Dans les bois éternels	Littérature	Fred Vargas	28/04/2006	18	306227	179599	32694	93934
208068948	La tragédie du Président. Scènes de la vie politique 1986-2006	Religion et Politique	Giesbert Franz-Olivier	10/03/2006	20	280905	146138	41747	93020
222110764	Mes amis mes amours	Littérature	Marc Levy	05/07/2006	21	278183	83768	83983	110432
226614852	Vous revoir	Littérature format poche	Marc Levy	17/05/2006	6,8	260319	74942	74831	110546
226615832	Sauve-moi	Littérature format poche	Guillaume Musso	21/03/2006	6,8	207753	67482	42069	98202
224667031	La touche étoile	Littérature	Benoîte Groult	05/04/2006	18	200239	113907	30105	56227
222617316	Deux petites filles en bleu	Littérature	Mary Higgins Clark	18/05/2006	22	168797	47174	60946	60677
284563287	Témoignage	Essais	Nicolas Sarozy	15/07/2006	17	166869	69017	42716	55136
206711821	France	Vie pratique et loisirs	Michelin		4	157986	30600	87610	39776

Tableau 10 : Top 10 des meilleures ventes de livres en magasin

CHAPITRE 3 : LES PLATEFORMES DE VENTES SUR INTERNET : UNE OPPORTUNITE POUR LES INDUSTRIES CULTURELLES ?[11]

Bounie D.[12], Eang B.[13] et Waelbroeck P.[14]

3.1 Introduction

De nombreuses plateformes de ventes de livres, CD et DVD sur Internet, à l'image d'Amazon Marketplace, se sont développées au cours de ces dernières années. Ces plateformes ont pour fonction de mettre directement en relation des professionnels et des particuliers intéressés par des produits neufs et d'occasion. Malgré le développement de ces plateformes, peu d'études se sont intéressées à ces marchés en ligne qui pourtant suscitent de nombreuses interrogations. Par exemple, les ventes de biens d'occasion sur ces plateformes sont-elles susceptibles de cannibaliser les ventes de biens neufs dans les réseaux de distribution traditionnels ? Pourquoi les consommateurs achètent-ils des produits neufs sur la plateforme alors qu'ils peuvent soit s'adresser à une grande enseigne Internet réputée comme Amazon soit faire leurs achats dans un magasin physique ?

Des premiers éléments de réponse ont été apportés. Ghose, Smith et Telang (2006) ont ainsi étudié l'effet de cannibalisation des ventes de produit neufs par les ventes de biens d'occasion sur la période 2002-2003 et ont montré que l'effet de cannibalisation était faible, de l'ordre de 15%. Cependant, il n'existe pas de travaux qui s'interrogent sur les caractéristiques des produits culturels neufs achetés par les consommateurs sur ces plateformes.

[11] Nous tenons à remercier très vivement la société TITE-LIVE qui édite le palmarès des meilleures ventes de livres en France aux professionnels du secteur à travers Edistat. Nous remercions en particulier Elise Tamisier qui nous a permis d'exploiter les données collectées auprès des réseaux physiques de distribution des livres en France. Enfin, nous remercions Marvin Sirbu, Christine Halmenschlager et Abel François pour leurs commentaires sur une première version de l'article.

[12] Département d'Economie et de Sciences Sociales, Telecom ParisTech.

[13] Département d'Economie et de Sciences Sociales, Telecom ParisTech.

[14] Département d'Economie et de Sciences Sociales, Telecom ParisTech.

Dans cet article, nous étudions les ventes de livres, CD et DVD neufs échangés sur la plateforme Marketplace directement rattachée au site français d'Amazon. Nous argumentons que les consommateurs utilisent la plateforme pour acheter des produits dont les dates de sortie et de parution sont antérieures à celles disponibles d'une part sur Amazon et d'autre part dans les réseaux de distribution physique. En définitive, la plateforme de vente constitue une réelle opportunité pour les industries culturelles en organisant la commercialisation de produits plus anciens (de répertoire) qui ne font pas ou plus l'objet de ventes auprès des enseignes Internet et des commerces dans les réseaux de distribution physique.

Pour mettre en évidence ces résultats, nous procédons en deux étapes. Dans un premier temps, nous comparons les ventes de livres, CD et DVD sur Amazon et Marketplace sur la période de mars à août 2006. Après avoir mesuré les ventes sur chaque marché, nous évaluons le marché du livre sur Marketplace à 0,1% du marché français du livre dans les réseaux physiques (en comparaison, Amazon représentait 6% du marché physique en 2006). En restreignant l'analyse aux cent meilleures ventes hebdomadaires de produits échangés sur ces plateformes, nous montrons que la demande sur la plateforme de vente Marketplace se porte sur des produits plus anciens. Nous concluons cette partie en montrant que les catalogues des meilleures ventes hebdomadaires de produits sur Amazon et Marketplace sont distincts : en moyenne sur la période, seulement 20% des produits sont communs aux deux plates-formes. Dans un deuxième temps, nous étendons notre analyse en comparant, sur la même période, les ventes de livres sur Amazon, sur Marketplace et dans les réseaux de distribution physique. Nous montrons que la demande de produits sur Marketplace diffère également de celle des réseaux physiques.

Notre étude contribue à la littérature sur le commerce électronique de produits culturels sur trois points. Premièrement, nous utilisons une mesure de vente qui permet de classer les meilleures ventes sur Marketplace et de comparer ainsi le catalogue des meilleures ventes à ceux des réseaux physiques et d'Amazon pour des produits neufs équivalents. Deuxièmement, notre analyse est à notre connaissance la première à mettre en évidence que les meilleures ventes de produits culturels sur la plateforme porte sur des titres du répertoire ancien qui ne constituent plus des succès dans les grandes

enseignes Internet ou dans les réseaux physiques de distribution. Troisièmement, notre étude est la première à comparer directement trois produits culturels différents sur la même période. Les études précédentes s'intéressent principalement au commerce électronique du livre d'occasion aux Etats-Unis ainsi qu'aux logiciels, aux DVD et aux CDs de manière indépendante. Nos résultats s'observent sur les trois marchés et constituent des faits stylisés robustes.

Notre recherche poursuit les travaux de Bounie, Eang et Waelbroeck (2010) qui analysent la dynamique des ventes de livres en France et l'essor du commerce électronique. L'originalité de ces travaux repose sur l'exploitation d'un nombre conséquent de données publiques de plus en plus riches et accessibles sur Internet mais qui nécessitent des capacités informatiques et des traitements statistiques très longs. Notre recherche démontre qu'aujourd'hui la limite de la recherche sur le commerce électronique de produits culturels n'est plus le manque de données, mais le traitement de cette abondance d'informations. Cet article s'insère dans une logique de débat de politique économique et de présentation de recherches potentiellement intéressantes.

L'article est structuré en trois parties. Premièrement, nous comparons les meilleures ventes de livres, CD et DVD sur Amazon et Marketplace. Dans une deuxième partie, nous restreignons l'analyse au marché du livre et nous comparons la plateforme Marketplace aux réseaux de distribution physique. Enfin, nous concluons cet article.

3.2 Les meilleures ventes de livres, CD et DVD sur Amazon et Marketplace

L'étude de la plateforme Amazon est désormais devenue une référence dans les milieux académiques internationaux (Ghose, Smith et Telang, 2006) dans la mesure où le site est implanté dans de nombreux pays, et dans la mesure également où il héberge une plateforme d'achats/ventes de livres neuf et d'occasion entre professionnels et/ou particuliers (Marketplace). Cette plateforme est intéressante pour les internautes car le prix d'un livre neuf commercialisé par Amazon peut être directement comparé aux prix des livres neufs et d'occasion proposés sur la plateforme par d'autres vendeurs. L'internaute peut ainsi acheter le produit sur Marketplace en utilisant l'interface réputée d'Amazon.

L'originalité de notre approche est d'avoir collecté, à l'aide de scripts automatisés, les données du site Amazon.fr de manière quotidienne sur la période de mars à août 2006 sur les marchés du livre, du CD et du DVD, soit une période de 22 semaines[15].

Nous avons plus précisément identifié l'entrée d'un produit dans le top 100 des meilleures ventes d'Amazon et nous avons suivi l'évolution de son offre sur Amazon ainsi que les offres postées par les vendeurs sur la plateforme Marketplace au cours de 22 semaines. Dans cette base de données, chaque produit est décrit par trois ensembles d'information. Le premier est lié aux caractéristiques du produit telles que l'identifiant (ISBN ou ASIN), le titre, l'éditeur, la langue, la date de sortie, le classement des ventes sur le site Amazon (SalesRank), la note moyenne attribuée au livre par les consommateurs, le nombre de notes, le prix proposé par Amazon, etc. Le deuxième ensemble d'informations a trait aux caractéristiques des différents exemplaires à la vente d'un même livre disponible sur la plateforme Marketplace comme la qualité du livre (état de cet exemplaire : neuf, comme neuf, etc.), le prix, etc. Enfin, la troisième source d'information concerne le vendeur de chaque exemplaire sur la plateforme d'occasion comme le nom du vendeur, la réputation du vendeur, le nombre d'exemplaires vendus par le vendeur, etc.

Avant de commenter plus en détails les marchés, il est important de souligner que le travail de collecte et de traitement des données est extrêmement lourd informatiquement puisque les bases de données sont composées environ de 2 millions d'observations pour le marché des livres, de plus de 2,3 millions d'observations pour le marché des CD et de plus de 620 000 observations pour le marché des DVD. Le Tableau 1 indique que les marchés de produits culturels ont des caractéristiques différentes. Ainsi, le nombre d'objets mis en vente est le plus important sur le marché des CD. Les livres et les DVD suivent. Le nombre de DVD proposés sur la plateforme sur la période d'étude est plus de deux

[15] Cette période d'observation connait un effet de saisonnalité lié au mois de juin qui enregistre des ventes élevées des « cahiers de vacances » dans les grandes surfaces alimentaires principalement. En outre, notre période d'observation ne contient pas le mois de décembre propice aux ventes de produits superstars. Le premier phénomène est de nature à modifier les classements des meilleures ventes de livres dans les grandes surfaces alimentaires uniquement et n'affecte donc pas les comparaisons des meilleures ventes de livres entre Amazon et Marketplace. Il est intéressant toutefois de noter que les ventes de livres dans les grandes surfaces alimentaires sur la période ne représentent que 27% des ventes totales de livres dans les réseaux physiques et que les ventes de manuels scolaires ne représentent à leur tour que 27% des ventes de livres dans ces magasins. Ce premier effet de saisonnalité a donc peu de chance d'affecter sensiblement nos résultats. Le deuxième phénomène concerne tous les marchés et tous les canaux de distribution et ne porte donc pas atteinte à nos analyses.

fois plus faible que celui des CD. Le nombre d'offres postées par les vendeurs suit le même ordre. Ces marchés diffèrent également par le nombre de vendeurs : il y a proportionnellement plus de vendeurs par produit sur le marché des CD que sur ceux des livres et des DVD, même si le nombre total de vendeurs est plus faible pour les CD. Les vendeurs sont donc moins nombreux mais proposent un catalogue plus étoffé sur le marché des CD.

Marchés	**Livre**	**CD**	**DVD**
Nombre d'observations	1 944 177	2 348 686	622 289
Nombre de produits	2 474	2 993	1 367
Nombre de vendeurs	1 164	686	1 120

Tableau 1 : Description des données

3.2.1 Mesure des ventes

Pour comparer les meilleures ventes de produits sur Amazon et la plateforme Marketplace, nous procédons en deux étapes. Premièrement, nous disposons des cent meilleures ventes journalières de produits sur Amazon que nous agrégeons à la semaine afin d'obtenir un classement hebdomadaire[16].

Deuxièmement, le classement des meilleures ventes sur Marketplace n'est pas communiqué par Amazon. Nous proposons donc de mesurer les ventes d'un produit à partir de la définition suivante. Pour une offre postée, nous considérons comme une vente à la date *t* une situation où cette offre est disponible à la date *t*, disparaît en *t+1* puis réapparaît en t+2 ou après[17]. Par exemple, si une offre d'un vendeur n'apparaît plus dans la base de données en t+1 par exemple, nous supposons que l'offre postée par le marchand a été acceptée par un internaute et nous la considérons donc comme une vente. Cette méthode permet d'obtenir 312 000 ventes sur un ensemble total de 4 915 000 offres postées sur les marchés du livre, CD et DVD (soit 6,35 %). A partir de cette mesure des ventes, nous sommes en mesure de calculer le nombre de ventes par produit par jour et par semaine et de construire le classement des cent meilleures ventes de produits sur chaque marché.

[16] Pour les produits qui rentrent dans notre base de données durant un jour de la semaine, nous calculons le rang moyen uniquement sur les jours où ils sont présents durant la semaine. Nous utilisons un classement hebdomadaire pour le comparer ultérieurement au marché physique pour lequel nous disposons seulement des ventes de produits à la semaine et également pour réduire les problèmes liés à la volatilité des classements journaliers.

[17] Notre mesure de vente peut servir de base de comparaison pour les recherches futures. Par la suite, nous agrégeons les ventes pour obtenir des classements relatifs.

Avant de poursuivre, il est important de noter que cette définition d'une vente a des limites puisqu'elle peut soit surestimer soit sous-estimer les ventes. D'un côté, nous pouvons surestimer les ventes lorsqu'un vendeur retire une offre de son catalogue qu'il a vendue par ailleurs par un autre moyen (réseau physique ou vente directe, etc.)[18]. D'un autre côté, notre définition peut sous-estimer les ventes : tant qu'une offre est affichée, nous ne pouvons pas identifier une vente. Ainsi un vendeur disposant d'un stock important d'un produit peut en vendre le jour *t* et poster à nouveau une autre offre au même prix le jour suivant. Dans ce cas, notre méthode ne permet pas d'identifier cette vente. Cette définition pénaliserait donc les meilleurs vendeurs qui ajustent régulièrement leurs offres sur Marketplace. Nous montrons dans la suite cependant que notre mesure de vente n'est pas affectée par ce dernier problème.

3.2.2 Taille et structure des marchés

Le calcul des ventes sur les trois marchés de la plateforme Marketplace est synthétisé dans le Tableau 2. Nous observons que le marché du CD avec près de 130 000 ventes est une fois et demi plus développé que le marché du livre et trois plus développé que le marché du DVD. Avec un nombre estimé de 89 000 ventes, le marché du livre sur Marketplace représenterait environ 0,1% du marché physique, estimation qui pourra servir de référence pour des recherches futures.

Livre		**CD**		**DVD**	
Nb. total de ventes	Nb. vendeurs	Nb. total de ventes	Nb. Vendeurs	Nb. total de ventes	Nb. vendeurs
89 447	703	130 534	486	41 954	757

Tableau 2 : Taille des marchés du livre, CD et DVD.

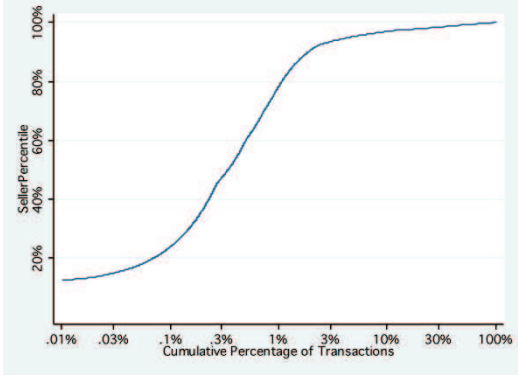
Nous remarquons ensuite que les ventes de livres sont concentrées sur un nombre de marchands plus restreint contrairement au marché du CD et du DVD par exemple (voir Graphique 1Tableau 2). Ainsi, 10% des vendeurs sur le marché du livre réalisent environ 99% des ventes alors que 40% des marchands de DVD sur Marketplace réalisent 99% des ventes[19]. Ces chiffres contredisent en

[18] Il faut remarquer que même dans ce cas, notre mesure capture une vente mais non par l'intermédiaire de Marketplace.

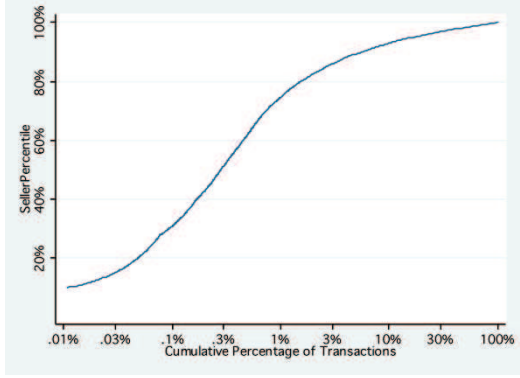
[19] Ces chiffres sont similaires pour les Etats-Unis et le Royaume Uni, voir Bounie et *al*. (2010).

conséquence l'intuition selon laquelle les marchés sur Internet seraient très concurrentiels. L'annexe 1 fournit plus de détails sur les caractéristiques des vendeurs.

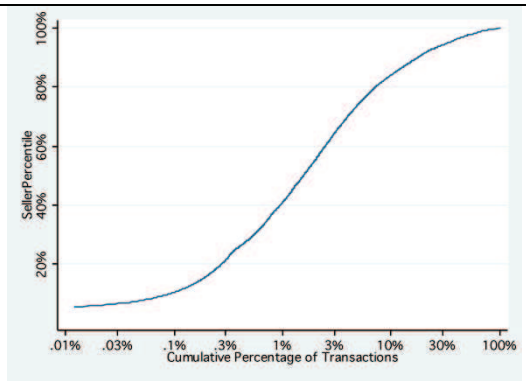
Livre



CD



DVD



Graphique 1 : Distribution des transactions sur les vendeurs

3.2.3 Les meilleures ventes de produits sur Amazon et Marketplace

Nous argumentons dans cette section que les internautes achètent sur Marketplace des produits différents de ceux d'Amazon. Ces produits sont pour la plupart des produits du répertoire ancien, *i.e.* dont la date de parution ou de sortie est ancienne. Cette demande pour des anciens titres structure les catalogues des meilleures ventes de produits sur Amazon et Marketplace qui sont distincts.

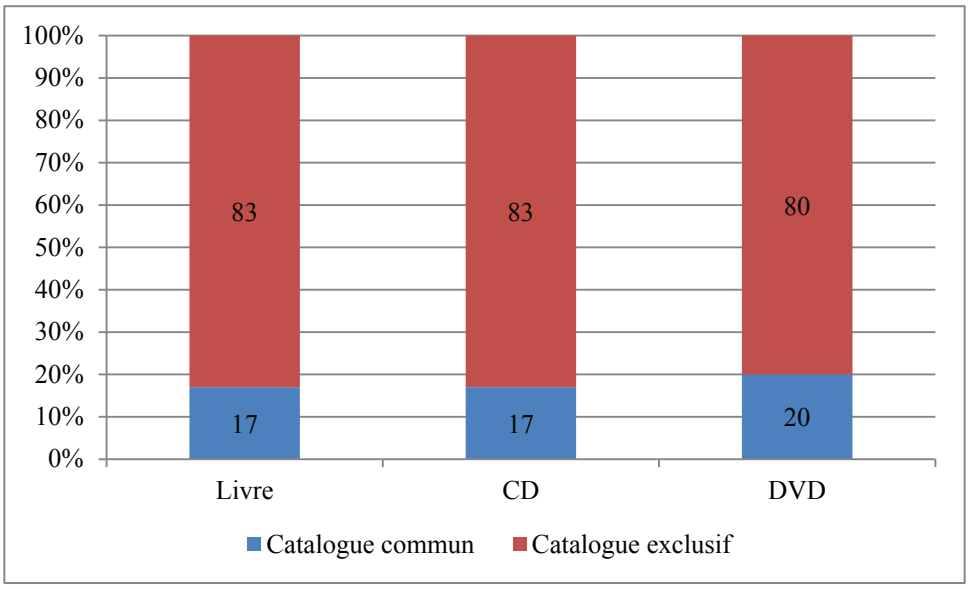
3.2.3.1 Des catalogues de produits différents

Dans cette partie, nous comparons les catalogues des cent meilleures ventes sur Amazon et Marketplace. Au préalable, il est utile de remarquer que le nombre total de produits qui entre dans le top 100 des meilleures ventes hebdomadaires sur l'ensemble de la période est largement supérieur sur Marketplace sur tous les marchés.

	Livre	CD	DVD
Nombre de produits dans le TOP 100 Amazon	762	830	633
Nombre de produits dans le TOP 100 Marketplace	932	858	662

Tableau 3 : Nombre total de livres, CD et DVD des cent meilleures ventes hebdomadaires

Nous représentons dans le Graphique 2, le nombre moyen de produits appartenant simultanément au top 100 des meilleures ventes d'Amazon et de Marketplace. Ainsi, en moyenne, sur les 22 semaines d'observation, il y a 17 livres, 17 CD et 20 DVD qui partagent le top 100 sur les deux plates-formes. On peut donc conclure de l'analyse de ce graphique que la demande sur la plateforme Marketplace est en moyenne à plus de 80% différente de celle d'Amazon et, en conséquence, que la demande de produits culturels achetés sur la plateforme se porte sur des titres très différents. En première analyse les risques de cannibalisation des ventes de biens neufs sur Amazon par les ventes de biens neufs sur la plateforme sont très limités.



Graphique 2 : Pourcentage du nombre moyen de livres, CD et DVD appartenant simultanément aux cent meilleures ventes sur Amazon et Marketplace

Les consommateurs achètent alors de préférence des produits qui ont des caractéristiques différentes sur Amazon. Nous nous intéressons en particulier à l'une de ces caractéristiques dans la partie qui suit.

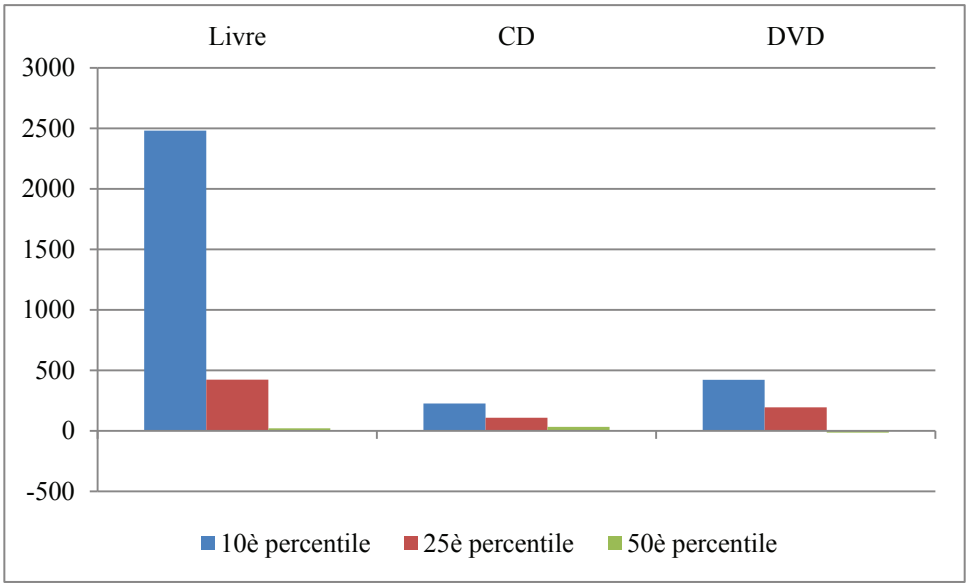
3.2.3.2 L'importance des titres du répertoire ancien

Nous nous intéressons dans cette section à un élément important de différenciation entre les plateformes : les dates de sortie des meilleures ventes de produits. Nous souhaitons en particulier vérifier si les produits consommés sur la plateforme sont des produits du répertoire ancien.

Pour ce faire, nous calculons chaque semaine les 10ème, 25ème et 50ème percentiles des dates de sortie des cent meilleures ventes de livres, CD et DVD sur Amazon et sur la plateforme Marketplace en excluant les produits communs à Amazon et à Marketplace (qui ont forcément des dates de sortie identiques). Nous calculons ensuite la moyenne de ces percentiles hebdomadaires sur l'ensemble des semaines.

Cette analyse montre que jusqu'au 25ème percentile, les dates de sortie des cent meilleures ventes sur la plateforme Marketplace sont nettement inférieures à celles d'Amazon (Graphique 3). Ces

résultats sont particulièrement prononcés pour les livres et les CD et, dans une moindre mesure, pour les DVD qui représentent un support plus récent.[20] Notre analyse confirme donc que les meilleures ventes de livres, CD et DVD sur la plateforme Marketplace constituent des produits du répertoire ancien ou de stock épuisé qu'Amazon ne peut sans doute plus garantir dans des délais raisonnables.



Graphique 3 : Différence en nombre de jours des percentiles des dates de sortie des meilleures ventes sur Amazon et sur Marketplace sur l'ensemble de la période

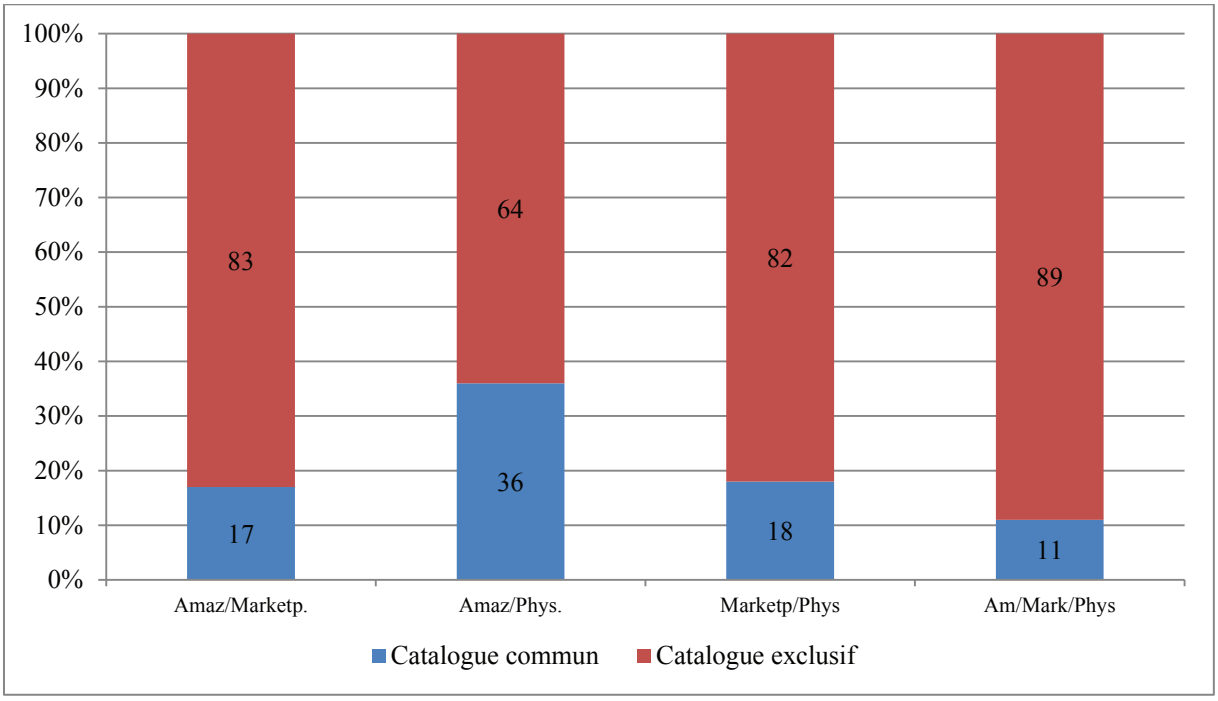
Dans la section suivante, nous prolongeons l'analyse en comparant les cent meilleures ventes de livre sur Amazon, Marketplace et les réseaux de distribution physique.

3.3 Les meilleures ventes de livres sur Amazon, Marketplace et les réseaux de distribution physique

Dans cette partie, nous souhaitons vérifier si les meilleures ventes sur la plateforme Marketplace coïncident ou non avec les meilleures ventes sur les marchés physiques. Nous souhaitons en particulier confirmer le résultat précédent qui montre que la demande sur la plateforme se porte sur des titres du répertoire ancien.

[20] Une nouvelle édition d'un film ancien au format DVD reçoit automatiquement un nouveau numéro ASIN. Ce qui peut expliquer pourquoi certains DVD d'Amazon ont des dates de sorties plus récentes que sur Marketplace.

Les données de vente par titre enregistrées dans les réseaux physiques sont rares. Nous limitons donc notre analyse au marché du livre pour lequel nous disposons des meilleures ventes de livres hebdomadaires dans les réseaux traditionnels (librairies, grandes surfaces alimentaires et grandes surfaces spécialisées) en France sur la période de mars à août à 2006. Les ventes sont estimées à partir des relevés de caisses issus d'un panel de près de 1200 magasins en France métropolitaine (hors Corse). Ce panel est représentatif du poids des différents circuits de distribution dans la vente de livres neufs aux particuliers (hors Club, vente par correspondance et Internet). La représentativité de l'échantillon est calculée selon les données du DEP-Ministère de la Culture (TN-Sofres pour l'Observatoire de l'économie du livre). Pour chaque livre appartenant aux 100 meilleures ventes de livre, nous disposons d'informations sur le genre littéraire, le prix conseillé, l'éditeur, l'auteur et le total des ventes hebdomadaires.



Graphique 4 : Nombre de produits communs (en %) dans les catalogues des cent meilleures ventes de livres sur Amazon, Marketplace et les réseaux physiques (en moyenne par semaine)

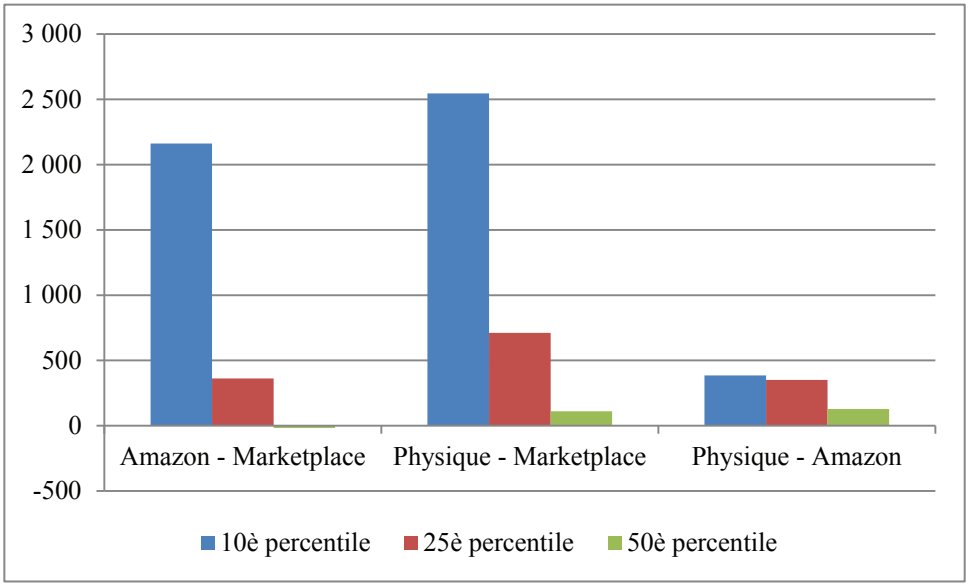
L'analyse descriptive des données montre tout d'abord que 488 livres différents sont entrés dans le top 100 hors-ligne pour un volume total de ventes de 11 980 647 exemplaires. En comparaison, 762 livres sont entrés dans le top 100 sur Amazon.fr et 932 sur Marketplace. La dynamique d'entrée-sortie des livres semble donc beaucoup forte sur Internet.[21] Dans le Graphique 4, nous comparons les cent meilleures ventes hebdomadaires dans les trois canaux de distribution, Amazon, Marketplace et les réseaux physiques, sur l'ensemble de la période. Nous constatons que les catalogues des cent meilleures ventes de produits sur Amazon est plus proche de celui des réseaux physiques que du catalogue de Marketplace : en moyenne, 36 % des meilleures ventes d'Amazon font également parties des meilleures ventes hebdomadaires des réseaux physiques, alors que ce pourcentage n'est que de 18 % pour les meilleures ventes de Marketplace. En croisant les trois marchés, nous observons qu'en moyenne 11% des livres sont communs aux trois réseaux de distribution. A nouveau, nous pouvons conclure que les listes des meilleures ventes sont très différentes sur les trois réseaux de distribution.

Ces différences de catalogue se retrouvent lorsque l'on analyse les dates de sortie des meilleures ventes de livres sur les trois réseaux de distribution. A l'image des analyses précédentes, nous calculons les 10ème, 25ème et 50ème percentiles des dates de sortie des cent meilleures ventes de livre sur Amazon, Marketplace et les réseaux physiques. Nous calculons ensuite la moyenne de ces percentiles hebdomadaires sur l'ensemble de la période. Le Graphique 5 confirme le résultat établi dans la section précédente : les dates de sortie des meilleures ventes de livre sur Amazon, Marketplace et les réseaux physiques sont globalement très différentes.

Tout d'abord, nous constatons que les dates de sortie des livres sur Internet en général sont plus anciennes que celles sur les réseaux physiques. Ainsi, au 10è percentile, il existe plus de 2500 jours (7 ans) d'écart entre les dates de sortie des meilleures ventes de livres sur les réseaux physiques et sur la plateforme Marketplace. Ensuite, nous remarquons qu'il existe entre une et deux années de décalage au 25ème percentile entre les réseaux physiques d'une part et Amazon et Marketplace d'autre part. En d'autres termes, nos analyses montrent, d'une part, qu'une large partie des meilleures ventes de livres

[21] Sur un ensemble total de 2509 livres rentrés dans le top 100 quotidien, seuls 762 sont retenus lorsque nous agrégeons les meilleures ventes à la semaine. En effet, un certains nombre de livres entre dans le top 100 un jour et puis ressortent quelques jours après et n'apparaissent donc pas dans les 100 meilleures ventes de la semaine.

sur Amazon constitue également des livres du répertoire ancien (par rapport aux réseaux physiques) et, d'autre part, qu'une large partie des meilleures ventes sur la plateforme sont également des titres d'un répertoire plus ancien par rapport à Amazon et par rapport aux réseaux physiques.



Graphique 5 : Différence en nombre de jours des percentiles des dates de sortie des cent meilleures ventes sur Amazon, Marketplace et les réseaux physiques

3.4 Conclusion

Il est difficile aujourd'hui d'acheter des produits culturels neufs du répertoire ancien en raison des ruptures de stock des éditeurs et de l'espace limité des rayons dans les grandes surfaces spécialisées. Avec le développement des plateformes de vente sur Internet telles qu'Amazon Marketplace, la question de la dispersion dans l'espace et dans le temps de l'offre de produits culturels se pose de manière moins saillante, ce qui rend plus attractif le marché des produits neufs et d'occasion pour les consommateurs et les offreurs.

Nous avons comparé le catalogue des cent meilleures ventes de produits culturels neufs sur la plateforme de vente Marketplace à celui de la grande enseigne Internet Amazon. Nous avons montré qu'ils diffèrent principalement selon deux dimensions. Premièrement, le catalogue des livres, CD et DVD en commun dans les meilleures ventes respectives n'est que de l'ordre de 20%. Deuxièmement,

au moins 25% des meilleures ventes de Marketplace ont des dates de sorties inférieures à celles d'Amazon. Ce résultat a été également vérifié lorsque nous comparons la plateforme de vente Marketplace aux physiques (librairies, grandes surfaces alimentaires et grandes surfaces spécialisées).

Le développement des plateformes de vente de produits culturels constitue donc une opportunité pour les industries culturelles pour plusieurs raisons. Premièrement, elles concourent à accroître la variété des produits culturels achetés par les consommateurs puisque les meilleures ventes de produits culturels sont très différentes entre les réseaux de distribution. Deuxièmement, elles contribuent à une plus grande différenciation des réseaux de distribution en termes de variété des produits culturels offerts. Nous avons ainsi montré qu'un élément important de différenciation entre les réseaux de distribution physique et Internet reposait sur la dimension temporelle : les réseaux physiques semblent très liés à l'actualité des sorties de nouveaux titres alors que les plateformes de ventes sur Internet constituent un canal privilégié pour commercialiser des produits du répertoire plus ancien. Troisièmement, l'existence de plateformes de vente comme Marketplace permet de décentraliser une partie du problème de gestion du stock croissant des produits culturels. Amazon semble déjà tirer parti de cette stratégie en certifiant les vendeurs de réputation incontestable et en leur déléguant une partie de leurs commandes. Enfin, les plateformes de ventes tendent à allonger la durée de vie des produits culturels, ce qui constitue un avantage pour les auteurs et les artistes qui bénéficient d'une fenêtre d'exposition beaucoup plus longue.

Ces résultats posent cependant de nouvelles questions auxquelles les recherches futures devront répondre. Tout d'abord, le nouveau canal de distribution que représentent les plateformes de ventes a une incidence sur la variété des produits culturels proposée aux consommateurs. De fait, les produits du répertoire ancien, même s'ils sont certainement distribués dans les réseaux physiques, connaissent un succès relatif plus important en ligne. Mais quel est l'impact de la consommation de produits du répertoire ancien sur les achats de biens neufs récents par les consommateurs ?[22] Compte tenu des contraintes budgétaires des consommateurs, ces achats de produits du répertoire ancien peuvent-ils se

[22] Ce phénomène va dans le sens du modèle d'affaires de la longue traîne qui exploite la commercialisation des produits de niche en ligne.

substituer à terme aux achats de livres neufs plus récents ? En d'autres termes, s'il existe une augmentation des ventes pour ces livres sur les plateformes, il reste à démontrer qu'elle est supérieure à une éventuelle réduction des ventes sur les marchés physiques. Combien de librairies adoptent ces plateformes de ventes en ligne ? Ne risque-t-on pas une exclusion des petites librairies du commerce électronique ? Ensuite, nos résultats révèlent une modification du cycle de vie des produits qui apparait plus long sur Internet. Dès lors, les mesures de succès et d'audience des produits culturels devraient également s'adapter à cette évolution et proposer des mesures cumulées sur les différents réseaux de distribution. Enfin, il reste à caractériser les livres qui appartiennent au répertoire ancien et à expliquer leur succès. Ce dernier résulte-t-il d'un effet d'exposition entre différents médias (comme la programmation à la télévision d'un roman adapté d'un livre que les lecteurs souhaiteraient découvrir) ? S'agit-il plutôt d'un effet de bouche-à-oreille retardé ? Ou bien le succès s'explique-t-il par la révélation tardive d'un auteur (comme la découverte d'un ancien livre d'un auteur nouvellement révélé) ? Ou, enfin, s'agit-il de la redécouverte de livres qui s'inscrivent dans des séries (comme les bandes dessinées, les suites, etc.) ?

3.5 Références

Anderson, C., 2004, "The Long Tail", Wired, October.

Bounie D., B. Eang, M. Sirbu et P. Waelbroeck, 2010, "Online Price Dispersion: What can we learn from Amazon Marketplace?", mimeo.

Bounie D., B. Eang et P. Waelbroeck, 2010, "Caractérisation et dynamique des ventes de livres en France: l'essor du commerce électronique", *Revue d'Economie Politique*, à venir.

Ghose A., M. Smith et R. Telang, 2006, "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact", *Information Systems Research*, 17(1), 3-19.

3.6 Annexe

Statistiques des vendeurs sur Amazon et Marketplace

Cette annexe caractérise les meilleurs vendeurs, *i.e.* ceux qui réalisent plus de dix ventes sur la période de l'étude.

Nous remarquons que les meilleurs vendeurs réalisent quasi l'entièreté des ventes (dernière ligne du tableau 3). Il est intéressant également de constater que les marchands qui ne font pas parti des meilleurs vendeurs tarifient en moyenne leur bien en-dessous du prix d'Amazon (14% livre ; 42% CD et 9% DVD). Cette tarification est sans doute en relation avec leur réputation puisque nous observons que la réputation des vendeurs professionnels est supérieure de 1 point (sur 5) sur la plateforme du livre et du DVD.

Produit culturel	**Livre**	**CD**	**DVD**
Nombre de produits	2 474	2993	1367
Nombre de vendeurs			
Nombre moyen de vendeurs par produit	8,97	10,31	5,93
Nombre moyen de « meilleurs vendeurs » par produit	8,87	10,24	5,75
Nombre moyen des autres vendeurs par produit	0,1	0,07	0,18
Note des vendeurs			
Note moyenne de tous les vendeurs	4,52	4,55	4,54
Note moyenne des « meilleurs vendeurs »	4,53	4,56	4,56
Note moyenne des autres vendeurs	3,3	3,93	3,58
Nombre de note des vendeurs			
Nombre moyen de notes de tous les vendeurs	8 788	31 595	6 606
Nombre moyen de notes des « meilleurs vendeurs »	8 842	31 853	6 799
Nombre moyen de notes des autres vendeurs	91	99	270
Prix des produits			
Prix moyen d'un produit sur Amazon	11,54	14,82	17,96
Prix moyen d'un produit sur Marketplace	11,95	15,22	18,18
Prix moyen d'un produit vendu par les « meilleurs vendeurs »	11,96	15,24	18,18
Prix moyen d'un produit vendu par les autres vendeurs	10,42	10,7	16,66
Ventes			
% des ventes réalisées par les « meilleurs vendeurs »	98,7	99,4	96,8
% des ventes réalisées par les autres vendeurs	1,3	0,6	3,2

Tableau 4 : Statistiques des vendeurs sur Amazon et Marketplace

Nous constatons également que le prix moyen d'un produit sur Marketplace est quasi-identique à celui d'Amazon sur tous les marchés[23]. Ce résultat n'est pas surprenant pour le marché du livre neuf puisque celui-ci est réglementé en France par la loi du prix unique et il est donc normal que nous n'observions quasi aucune différence entre le prix moyen d'Amazon et le prix moyen Marketplace puisque le rabais maximum autorisé par rapport au prix éditeur est de 5%.[24] En revanche, ce résultat est plus surprenant pour les CD et DVD car on aurait pu penser que les internautes allaient sur Marketplace pour trouver de meilleurs prix.

[23] Le calcul de la moyenne est effectué de la manière suivante. Pour chaque objet, nous déterminons le prix moyen d'Amazon et celui des meilleurs vendeurs sur Marketplace dans le temps. Nous calculons ensuite la moyenne pour tous les produits.

[24] La loi n°81-766 relative au prix du livre du 10 août 1981 établit dans son article 1er que « les détaillants doivent pratiquer un prix effectif de vente au public compris entre 95 pour cent et 100 pour cent du prix fixé par l'éditeur ou l'importateur ». En outre, l'article 5 de la loi précise que les détaillants peuvent pratiquer des prix inférieurs au prix de vente au public mentionné à l'article 1er sur les livres édités ou importés depuis plus de deux ans et, dont le dernier approvisionnement remonte à plus de six mois ».

CHAPITRE 4: SUPERSTARS AND OUTSIDERS IN ONLINE MARKETS, AN EMPIRICAL ANALYSIS OF ELECTRONIC BOOKS

Bounie D.[25], Eang B.[26], Sirbu M.[27] and Waelbroeck P.[28]

4.1 Introduction

The book industry is facing a new revolution. According to most experts, electronic books (e-books) will shake the industry in the same way that the printing press did.[29] Recent numbers published by the *Association of American Publishers* show that 49.5 million e-books have been sold in the US in 2010 for a value of USD 441 million (an increase of 291% compared to 2009). Sales of e-books now represent 8.3% of total sales of books in the US in 2010 (against 3.2% in 2009).[30] Amazon currently sells more e-books for Kindle, their proprietary e-book reader, than print books.[31] Many actors in the industry view this major technological change with apprehension for various reasons. First, sales of e-books could threaten retail stores selling print books, given the income constraint of consumers. Secondly, this potential cannibalization of print sales by e-books is reinforced by the entry of new online retailers such as Amazon, Apple or Google. The bargaining power between traditional retail stores and new online merchants is changing the way companies conduct business. For instance, new technologies allow writers to publish e-books directly without signing with a publisher.[32] Finally,

[25] Telecom ParisTech, Department of Economics and Social Sciences, 46 rue Barrault, 75634 Paris Cedex 13, France; Tel.: + 33 1 45 81 73 32; Email address: david.bounie@telecom-paristech.fr; Corresponding author.
[26] Telecom ParisTech, Department of Economics and Social Sciences, 46 rue Barrault, 75634 Paris Cedex 13, France; Email address: bora.eang@telecom-paristech.fr.
[27] Carnegie Mellon University, Department of Engineering and Public Policy and Tepper School of Business, 5000 Forbes Avenue Pittsburgh, PA 15213-3891; Email address: sirbu@cmu.edu.
[28] Telecom ParisTech, Department of Economics and Social Sciences, 46 rue Barrault, 75634 Paris Cedex 13, France; Email address: patrick.waelbroeck@telecom-paristech.fr.

[29] See for instance The NewYork Times website, "Using E-Books to Sell More Print Versions", June 26, 2011; or, The Wall Street Journal, "How the E-Book Will Change the Way We Read and Write", April 20, 2009; etc.
[30] The press release is available at the following address: http://www.publishers.org/press/24/ (last visit: 20/05/2011).
[31] A press release available at the following address outlines: "Since April 1 [2010], for every 100 print books Amazon.com has sold, it has sold 105 Kindle books" (last visit: 20/05/2011).
[32] See two illustrative papers retrieved from the Guardian.co.uk, 12 January 2012. "Amanda Hocking, The writer who made millions by self-publishing online" and guardian.co.uk, 8 February 2012, "Self-published

editors and publishers fear illegal copying on the Internet of copyrighted material that could reduce their income.

This article directly addresses the issue of the cannibalization: do e-books cannibalize or instead increase print sales? In this study, we focus on the characteristics of books that are best sellers in both print and electronic formats, best sellers in print but not as e-books, and successful as e-books but either currently less successful in print or have no print equivalent. We refer to the first group as "superstars", the second as "print preferred" and the third as "digital outsiders". Paper may be preferred for reasons of color, graphics, or need to navigate non-linearly, a style to which the current generations of e-book readers are not well adapted. To do so, we have run automated scripts to collect all titles that have appeared in the monthly top 100 list of best-selling print and electronic books from the US Amazon websites. The period of the sample ranges from November 2007 to July 2010 (33 months). There are 1861 unique titles in the set of print and electronic books that we compare. The focus on the US market is justified by the faster development of the US e-book market compared to other countries. Using econometric methods, we find that the cannibalization of print sales by sales of electronic books is more likely to occur for superstar titles written by successful authors. However, we find that a new segment of successful electronic titles that are not best-selling in print format emerge (digital outsiders); these books would probably have been unpopular without the new Kindle store and therefore this new distribution channel creates a market expansion effect. These digital outsiders are characterized not only by lower prices but also by older release dates. They also include titles that are only released in electronic format. We then argue that electronic books increase the duration of popularity of older print releases.

The remainder of the article is organized in four sections. We review the literature in Section 2. In Section 3, we describe the dataset and compare the best selling books in print and electronic formats. In Section 4, we explain the probability that a best selling print book is also best selling in its

---

ebook author becomes Amazon's top seller; "Amazon's success with self-published authors in the UK follows the US arm of the retailer's announcement last year that two self-published authors, John Locke and Amanda Hocking, had sold more than 1m books on the Kindle. Hocking went on to sign a reported $2m deal with St Martin's Press, while Locke has signed up with Simon & Schuster."

electronic version by a set of book characteristics. We also characterize digital outsiders and show that they either correspond to electronic books that do not exist in print formats or to old releases that Internet users want to rediscover through the electronic store. Section 5 concludes the article.

4.2 Related Literature

The issue of cannibalization or disintermediation has been a central question since the early days of the Internet and has taken a new turn in recent years. The Internet first created a new sales channel for *physical* cultural products and rapidly a growing empirical literature has emerged to assess how the Internet has impacted the sales of *physical* goods on off-line channels (Balasubramanian, 1998). This research led to the formulation of the "long tail" theory (Anderson, 2005). According to the latter, cultural industries which are characterized by superstar products and "winner-takes-all" phenomenon (Rosen, 1981) – *i.e.* a small portion of artists/writers/films account for most sales in the market – should be impacted by the Internet, which lower search costs and allows internet users to share information within online communities. As a result, the Internet should make cumulative sales of "niche" (obscure or unknown) products profitable and thus flatten the distribution of total sales of cultural products (Brynjolfsson *et al*., 2003). This prediction however remains controversial in the economics literature (Elberse and Oberholzer-Gee, 2007). In the specific context of the book market, Chevalier and Goolsbee (2003) and Ghose, Smith, and Telang (2006) have found that online channels result in a relatively small cannibalization of print sales.[33] Likewise, Bounie et *al.* (2011) have compared the top 100 weekly best-selling books in France via three different distribution channels: Amazon, Amazon Marketplace and physical stores during the period from March to August 2006. They show that on average, only 11% of all books ever entered all three lists of weekly Top 100 best selling items. Old releases are more successful in the electronic markets (Amazon and Amazon Marketplace) than in traditional physical stores.

[33] Ghose, Smith and Telang (2006) have especially assessed the impact of used print books on the sales of new print books on Amazon MarketPlace during the period 2002-2003. These authors show that used print books are imperfect substitutes to new ones in the sense that only 16% of used books cannibalize sales of new print books.

The issue of cannibalization has now taken a new turn with the increasing availability of *digital* cultural goods. Indeed, new formats have appeared on the Internet such as digital music (iTunes, Spotify, etc.) or online films and movie downloads (iTunes, Amazon Instant Video, Google Play, Netflix, Youtube, etc.). Likewise, the development of e-books has led to an increasing popularity of the Kindle store. The emergence of these new formats raises a new question: do *electronic* formats cannibalize or instead increase the sales of *physical* cultural products (that can be sold through physical or online channels)? Indeed, consumers not interested in print formats could now be interested in purchasing new electronic formats. This market expansion could not only be the result of decreasing search costs and online communities but could also be driven by a new demand for electronic formats (increased portability, library of books always available, bookmarks and annotations, etc.). The long tail theory should therefore be extended to account not only for the effect of online distribution channels on print sales but also the impact of new electronic formats on the variety of cultural products purchased by Internet users. Several empirical papers have addressed this question. Focusing respectively on the introduction of online newspapers and music, Deleersnyder *et al*. (2002) and Biyalogorsky and Naik (2003) find a relatively small cannibalization effect on physical newspaper circulation or record sales even though these findings are currently being debated; Koh *et al.* 2010 show for instance that digital music substitutes for physical album sales whereas Dang Nguyen *et al.* (2012) find that consuming music as streams has no impact on CDs purchases.[34] Finally, focusing on videos and TV, Waldfogel (2009) shows that the Youtube platform that allows users to stream videos online has only a small negative impact on television viewing and Danaher *et al*. (2010) find that the iTunes distribution channel has no significant statistical impact on DVD sales.

To the best of our knowledge, few studies have analyzed the market expansion/cannibalization issues in the book industry due to the emergence of an electronic format.[35] Jiang and Katsamakas (2010) use a game theoretical model to study the effects of the entry of an e-book seller on prices and total book readership. They find that prices in the book market may increase after the entry of an e-

[34] Zentner (2008) also shows that online music purchases are related to a change in the number of music stores in the United States for the years 1998 and 2002.

[35] Huang and Hsieh (2012) show that consumers' perceived innovative attributes not only directly affect their acceptance behaviour, but also influence behaviour via their perception of switching costs.

book seller and that the total readership may decrease, if the e-book seller is owned by one of the print book sellers. In an empirical context, Oestreicher-Singer and Sundararajan (2010) examine the e-book industry but only focus on the pricing strategy of e-book publishers and their choices of technological protection (with or without piracy). The closest paper to our research is probably Hu and Smith (2011). They study the impact of the decision of a publisher to stop releasing Kindle e-books on the Amazon website. This leads to a natural experiment where a title is initially only available in print format and later on available in both print and electronic formats. These titles can be compared to a control group of titles that were available in both formats during the full period of observation. They find that delaying the release of e-books causes an insignificant change in overall hardcover sales but a significant decrease in e-book sales, total sales, and likely total revenue and profit to the publisher. Unfortunately, they only have one to eight weeks of observations of titles that were delayed in electronic format compared to two months of observations for titles in the control group, which makes a comparison based on sales difficult to interpret. Moreover, they find a negative cross-price elasticity between books in print and electronic formats, which suggests that the formats are complements. On the contrary, as we will see, we estimate a positive cross-price elasticity, which implies that both formats are rather substitutes. Finally, they restrict their sample to avoid outliers with unusually large sales (average weekly sales higher than 1600 copies), while we specifically focus on these superstar titles. In addition, they do not analyze digital outsiders that are only available in electronic formats, whereas, they are a main focus of this article.

4.3 Data and methodology

Electronic books refer both to a title and to a reader of specific formats such as the Kindle reader.[36] In this article, we use the word e-book for the content. Several formats coexist, proprietary or not: ".azm", ".epub", ".pdf", ".txt", etc. Each format has its own specific features (reading images, printing content, bookmarking, searching indexes, etc.). We use data from the Kindle store owned by

[36] Two types of readers are available: the first one, called e-Reader, is exclusively dedicated to the reading of electronic texts such as the Amazon Kindle, the Sony Reader, etc. The second one is a multifunctional "touch pad" which allows reading electronic text files as well as playing videos, listening music, etc. The well known iPad by Apple is an example. Many e-book sellers also provide software based readers for PCs and smartphones.

Amazon.com that are readily available from the Internet. Indeed, Amazon keeps archives of the monthly top 100 best-selling print books (since 2000) and electronic books (since 2007). We ran automated PHP scripts to collect characteristics of print and electronic books (including rank, author, ISBN and so on) that entered at least once in the respective monthly top 100 list. Top print books were followed from July 2000 to July 2010, and top e-books from November 2007 (first available date of the archives) to July 2010. Amazon is the leading seller of e-books in the US.

Overall, the dataset covers a period of 121 months for print books and 33 months for e-books. In the rest of the article, we mainly focus on the period where both print books and e-books are available, *i.e.* November 2007 to July 2010.

4.3.1 Data collection

The characteristics of the books are the following: title, author, ISBN (for the print format) or ASIN (for the electronic format) which stands for Amazon Standard Identification Number, monthly sales rank (from 1 to 100), genre, publication date, observation date, the average rating of customers, and price. It is important to stress several data issues related to online data collection of book characteristics.

First, the same book can have a different title between its print and electronic versions; for instance, “Harry Potter 7” and “Harry Potter seven”. These two titles are the same and we treat them as such in our database by recoding them to a single title. Secondly, we face a similar problem for the name of the author of a title that might be spelled differently in each format: “J.K. Rowling” and “Joanne Rowling” for instance. We also recoded these fields to a single field. Thirdly, there are different versions of the same book like “soft paperback”, “hard paperback” versions, limited edition, audiobook edition, etc. These versions have different unique identifiers (ISBN or ASIN) but correspond to the same content. In what follows, we treat these different versions as a single informational content, which is what we are eventually interested in. At this point, we have a unique identifier for a book that is the combination of the author's name and title. Fourthly, the publication date is sometimes misreported on Amazon websites. Indeed, Amazon.com reports a publication date

of a specific version of a print book or an e-book that depends on the edition. While this is a minor issue for recent titles, we need to correct publication dates of books that were reprinted or that exist in different versions. We define the publication date of a title as the earliest publication date of existing versions sold on the Amazon website. For books in the public domain, we use the publication date of the first edition collected from Wikipedia. We have excluded books that were published before the 19th century to avoid biases due to a small number of observations. For reference, 7 out of 2340 unique titles can be considered "historical". Finally, books are assigned to genre categories on Amazon. We have grouped these genres in 6 categories for e-books: Practical (including self-help and hobbies) essays, science and biography (non-fiction); books for young readers; guidebooks and how-to; fiction; reference books and textbooks. Print books have an additional category: "Comics and graphic novels". A title that belongs to several categories is assigned to the category under which it had the highest sales rank. When Amazon does not assign a category to a title, we used keywords to assign a category. To sum up, the categories that we have assigned to books does not necessarily match the categories listed on the website.

4.3.2 Descriptive statistics

The data collection process resulted in two data sets: one for e-books and one for prints. Overall, 1,244 e-books appeared at least once in the top 100 list of best-selling titles versus 1,097 print books. There are only a few cases of multiple versions of the same book as the total number of unique titles is fairly similar (1,238 for e-books and 1,041 for print books). As a consequence, the top 100 list of best-selling electronic titles is characterized by a higher turnover. In this paper, we are interested in both superstar and long tail properties of electronic commerce in order to test any cannibalization or market expansion effects. To do that, we therefore merge the two data sets on e-books and prints into one single data set and decide to create three categories of titles: "superstars", *i.e.* titles that enter in both top 100 list of print and electronic books, "print preferred", *i.e.* best sellers in print but not as e-books, and "digital outsiders" which are successful as e-books but either currently less successful in print or have no print equivalent. These three categories are mutually exclusive. We provide descriptive statistics in Table 1.

First, among the 1,861 unique titles in our dataset, 33.5% are classified as print preferred (623), 44.1% are digital outsiders (820) and 22.4% (418) can be considered as superstar titles according to our previous definition. Thus the extent of cannibalization is rather limited in the sense that 77.6% of all titles are only best-selling in a given format (print or electronic).

	Print prefered	Digital outsiders	Superstar		Total
Format	-	-	**Print**	**Kindle**	
Number of unique books	623	820	418		1861
Number of unique authors	523	601	288		-
Number of books by author	1,5	2,0	2,7		-
Average date of release	2004	1999	2007		-
Minimum date of release	1900	1811	1905		-
Average rank of entry	59,8	53,1	51,1	44,4	-
Average rank of exit	64,6	59,9	59,9	65,4	-
Average best rank	51,3	48,8	38,6	35,5	-
Average lifespan (in month)	2,7	2,0	3,5	3,5	11,7
Average number of comments by book	4,22	3,8	3,9	3,9	15,82
Average Amazon price	16,8	10,3	15	12	54,1
Average Amazon price: hardcover (n=402)	18,5		16,5		35
Average Amazon price: paperback (n=184)	12,9		9,8		22,7
Categories					
Comics & Graphic Novels	11	0	0		11
Business & Investing	88	17	39		144
Non fiction	136	55	94		285
Children books	76	31	22		129
Fiction	65	634	226		925
Professional and Technical, Reference	38	9	0		47
Practical	201	57	36		294
Other	8	17	1		26

Table 1 : Descriptive statistics

Secondly, we count 79 digital outsiders (or 4.2% of all unique titles) that are only available in kindle edition (without a print equivalent). Out of these 79 titles, we find two categories of titles: classics and pure e-books. The first category of titles (27 e-books) are old titles released in print with a specific edition and which have not the exact counterpart on Amazon kindle store. The second category (54 titles), are specific to the kindle edition and often correspond to “Harlequin-like” e-books (see the Appendix A for a representative sample).

Thirdly, the number of books by author is the largest for superstar titles. Moreover, we define the best rank as the best monthly sales rank achieved by a book during the period of observation. This best

rank is on average lower (better) for superstar titles than for print preferred and digital outsider titles. Similarly, superstar titles have a slightly lower rank of entry on average (average rank of entry) and stay longer in the monthly top 100 list (average lifespan).

Fourthly, digital outsiders are on average older with an average release date in 1999 compared with print preferred titles (2004) and superstar titles (2007). If we analyze more closely the oldest titles in each category, we again find a digital outsider title first published in 1811 that made it to the monthly top 100 list of best selling electronic books, earlier than the oldest print preferred title (1900) or the oldest superstar title (1905).

Fifthly, print preferred and digital outsiders belong to specific genres, print books having more varied genres. The "fiction" category (291 out of 1,041 or 28%) is the main category of print books, followed by "practical" (23%), and "non-fiction" (22%). Electronic books are predominantly "fiction" (860 out of 1238, or 69%), followed by "non-fiction" (12%) and "practical" (8%). The category "Professional, technical and reference" is more popular among print preferred books than digital outsiders. A really small fraction of print preferred titles correspond to "Comics", currently unavailable in electronic format.

Finally, electronic books are on average cheaper than print books: USD 10.3 for digital outsiders vs. USD 16.8 for print preferred titles. Similarly, superstar titles cost more in print format (USD 15) than in electronic format (USD 10).

To sum up, we found that print preferred and digital outsider titles have different characteristics in terms of date of release and genres. In the following section, we use econometric methods to analyze the factors that determine the probability that a title is a superstar or digital outsider.

4.4 Econometric Methodology and Results

The methodology is the following. We run two sets of regressions. In the first set, we explain the probability that a best-selling print title in our dataset is a superstar, *i.e.* that it also enters the best selling list of electronic titles. In the second set of regressions, we explain the probability that a best-

selling electronic title is a digital outsider, *i.e.* that it does not enter the monthly list of best-selling print titles. Since we have three mutually exclusive categories (print preferred/digital outsiders/superstars), we discriminate among best-selling print titles between superstars and print preferred, and among best-selling electronic titles between superstars and digital outsiders.

The dependent variable is a binary variable y_i that is equal to 1 if the best-selling print (electronic) title is a superstar (a digital outsider) and to 0 otherwise. We use a linear probability model that assumes that the probability of observing the dependent variable equal to 1 is a linear function of a set of explanatory variables.[37] We specify the probability of observing y_i as:

$$p_i = P(y_i = 1) = x_i'\beta + \varepsilon_i$$

where y_i equals 1 with probability p_i and 0 with probability $1 - p_i$, x_i is the set of explanatory variables and ε_i is an unobserved variable with mean 0 and variance σ^2. We estimate parameters of the model by OLS.

4.4.1 Superstar Titles

In this section, we analyze the factors that determine whether a successful print book is also successful in electronic format. We are especially interested in best-selling print books that enter the monthly top 100 list of best-selling e-books. We run a regression on the sample of print books that have an e-book equivalent and that entered the monthly top 100 list at some date during our period of observation.[38]

We use most characteristics of a book as explanatory variables. In addition, we create the variable *DIFFPRICE* that computes the difference between the price of the print book and the price of the electronic version. The price of a print book can be related to its softcover or hardcover format. As a result, the price difference that we compute can either be attributed to a difference between a

[37] The linear probability model has the advantage of giving directly the partial effects and their standard deviations. We have checked the robustness of the results by estimating probit models as well. Results are similar and available upon request.

[38] There are 1,041 unique print books among which 267 do not exist in electronic format, leading to a sample of 774 observations. Finally, 4 titles have no average rating. The final sample is then 770.

paperback and an e-book or between a hardcover and a e-book version.[39] Therefore, this variable is not homogenous among observations. Nevertheless, we expect that a larger price difference increases the probability that a print title is very successful in the Kindle store. Next, we compute the number of different books published by the same author over the period of observation. We expect that the more published the author the higher the probability that a print book is also successful in electronic format. We also use the rating of a title by customers (between 0 and 5) and hypothesize that higher ratings should increase the probability that the dependent variable is equal to 1. We finally introduce a variable that computes the best rank reached by a title on the Amazon website. This variable is defined between 1 and 100, 1 being the best possible rank. All three of the preceding variables capture a superstar effect in the print book market. Next, we control for the number of days that a title has been available on the shelves, *TOM* for time on market, by calculating the difference between the day the title exited the monthly top 100 list and the first publication date. Finally, we use several binary variables to account for the different genres that we defined in the previous section; the base category is "fiction"; the other categories are "non-fiction", "reference and textbooks", "practical"; "guidebooks and how-to", "young readers".

In model A, we explain the probability that a best-selling print book is best-selling in electronic format, using all print titles that exist in electronic format (n=770). In model B, we use the same sample but use as dependent variable the probability that a best-selling print book reaches a rank of 33 or lower in a monthly top 100 list of best-selling electronic books. In model C, we restrict the sample to titles for which both print and electronic prices exist. This constraint on prices reduces the number of observations to 332. Table 2 gives the estimation results that confirm our interpretation of the descriptive statistics in the previous section.[40] The lowest (best) ranking has a negative effect on the probability that a successful print book is also successful in electronic format. For instance a variation of -45 ranks (from top 46 to 1), the probability increases by 18% (-45×-0.004) in Model A (the results are globally very close to those of Model B).

[39] For a small number of cases, the softcover and the hardcover formats of the same title entered together in the monthly top 100 list of best-selling books. In this case, we computed the average price of both formats to obtain the price of the print version.

[40] We provide in Table 5 (Appendix B) the correlation between the explanatory variables.

Dependent variable	Probability that a best-selling print book is best-selling in electronic format	Probability that a best-selling print book reaches a rank of 33 or lower in a monthly top 100 list of best-selling electronic books	
Model	A	B	C
	Coeff.	Coeff.	Coeff.
Minimum Ranking	-0.004***	-0.003***	-0.006***
Fiction (reference)			
Non-fiction	-0.254***	-0.236***	-0.256***
Reference and textbooks	-0.534***	-0.241***	-0.298
Practical	-0.480***	-0.260***	-0.267***
Guidebooks and how-to	-0.523***	-0.230***	0.0226
Young readers	-0.313***	-0.232***	-0.204**
Tom	-0.0002**	-0.0001	-0.0002
Average rating	-0.064**	-0.039	-0.036
Nb. of books by author	0.0099*	0.0173***	0.0188***
Diffprice			0.0153***
Constant	1.226***	0.602***	0.669***
R^2	0.298	0.253	0.249
N	770	770	332

Table 2 : Estimation results: – Probability that a best-selling print book is best-selling in electronic format

This superstar effect is confirmed by looking at the effect of the number of books by an author on the probability of e-book success: for an author with 12 books, this probability increases by 11% (12×0.009). Fiction is the most common genre for electronic books and this translates to negative coefficients for all other genres (from -25% to -52%). In Model C, we only keep observations for which *DIFFPRICE* is not missing. We find a positive and significant effect of the price difference on the probability of e-book success; for instance, for the peak of the distribution at USD 5, the probability increases by 7.5% (5×.015): an e-book that cost USD 5 less than its print version increase the probability of being a digital outsider by 7.5%.[41] In other words, increasing the price of the print format increases the sales (rank) in the electronic format, which implies a positive cross-price elasticity: the two formats are competing and there is cannibalization of sales of print titles by sales of electronic titles. Finally, the time on market of books does not significantly influence the probability of e-book success.

[41] This effect would be reinforced if we only included hardcovers that have a higher price than paperbacks.

4.4.2 Digital Outsiders

In this section, we analyze the factors that influence the probability that a successful electronic title does not appear in the monthly top 100 list of best-selling print books. The sample of books is the set of all electronic books that exist in print format.

In addition to the explanatory variables used in the previous section, we include in the regression the following variables. First, we create two variables that capture the price difference between the electronic version of a book and the paperback version (*DIFFPRICE1*)[42]. Moreover, we construct four variables that are related to the first date of publication of a title. The first variable, TOM, is computed as the difference between the month the book entered the top 100 list and its first publication date (Model A in Table 3). Next, we create 3 binary variables that account for contemporary books (8.4%), old books (6.3%) and public domain (3.5%). The first two binary variables are included in Model B and the public domain binary variable is used in Model C.

Estimation results are reported in Table 3.[43]

[42] We also run the same regression by using the price difference between the electronic book and its *hardcover* version. The results remain the same and are not reproduced in this version. Diffprice1 is limited to 774 observations since only 774 e-books had a price for its kindle version and its paperback version at the time of the data collection.

[43] We provide in Table 6 (Appendix B) the correlation between the explanatory variables.

Dependent variable	Probability of being an outsider	Probability of being an outsider	Probability of being an outsider
Model	A	B	C
	Coeff.	Coeff.	Coeff.
Minimum ranking	0.0046***	0.0046***	0.0046***
Fiction (reference)			
Non-fiction	-0.3620***	-0.3556***	-0.3603***
Reference and textbooks	0.3077	0.3115*	0.3062*
Practical	-0.1565***	-0.1570***	-0.1525***
Guidebooks and how-to	-0.5158***	-0.5223***	-0.5165***
Young readers	-0.1798194***	-0.1813***	-0.1856***
Nb. e-books by author	-0.0270***	-0.0270***	-0.0269***
Diffprice1	-0.0083***	-0.0094***	-0.0082***
Tom	0.0002***		
Contemporary (2000-2005)		0.1061**	
Old (before 2000)		0.2148***	
Public domain			0.3001***
Constant	0.5510***	0.5442***	0.5535***
R^2	0.223	0.223	0.222
N	774	774	774

Table 3 : Estimation results – Probability of being an outsider

First, we find that decreasing the best rank by one (increasing its popularity) has a negative effect on being a digital outsider. In other words, digital outsiders are more likely to be in the middle or the bottom of the charts. Secondly, digital outsiders are more likely to belong to the fiction category since all the coefficients with the other categories are negative. Next, the number of electronic books by authors increases the probability that a best-selling electronic book is also a best-selling print book (negative effect of this variable on the probability of being a digital outsider). In addition, the price difference between the electronic and the paperback version decreases the probability of being an outsider. In other words, when the price of a paperback is lower than the price of a Kindle version, the probability of reaching a high rank in the top 100 list of best-selling print books increases. This again translates a positive cross-price elasticity between the two markets. Finally, old titles and books that have been on the market for a long time have a lower probability of becoming high sellers. In the three specifications (Models A to C), the variables related to the first publication date decrease the

probability to become a print best-seller. As a consequence, estimation results bring evidence that new electronic distribution channels favor the renewal of old titles.

4.5 Conclusion

In this article, we have compared the best-selling list of print and electronic books sold on the US Amazon website from November 2007 to July 2010. Many people from the industry have feared the cannibalization of print sales by electronic books. If there is cannibalization we should see high cross-price elasticity between electronic and print. Our estimation results show high cross-price elasticity for a portion of the top selling print books. However, the Internet channel also creates a market expansion effect for titles that we referred to as digital outsiders. The latter fall into one of two categories: e-books that are not popular in print or are older and e-books that do not have a print equivalent (in our dataset, 4.2% of all titles). The first category corresponds to best-selling electronic books that never entered the monthly top 100 list of best-selling print books during the observation period of our study. It includes print titles from the public domain or released before 2000; for these titles, the electronic format increases the duration of popularity of print books and contributes to the renewal of old titles. There are two possible reasons why we are seeing older books appearing as best selling e-books. First, people who already own paper copies of these books are now buying e-book versions because they value unique properties of the e-book version (searchability, portability, ease of bookmarking/annotating). If this is the explanation, it is a short-lived (e.g. 5-10 year) phenomenon that will disappear once paper has been replaced by e-books. This phenomenon would be similar to converting one's CD library to iTunes. Secondly, there is a continuing interest in the classics. Readers who are new to these books are more likely to acquire their first edition as an e-book (lower cost) rather than as print, thus raising the ranking of the e-book sales so they appear in the top 100. This phenomenon should be ongoing, and is another instance of cannibalization. Indeed, Amazon offers public domain titles as e-books at a price of zero. The second category of digital outsiders corresponds to electronic books written in majority by contemporary authors specializing in the romance genre (Harlequin-like). Online distribution of electronic books allows them to cheaply self-release new titles

and to test the potential readership of each title, leading to a print version for successful electronic titles.

The existence of digital outsiders has both research and strategic implications.

First, our results suggest a new direction for future research on the research on the long tail theory. Current research has mainly focused on the Internet as a new channel for distributing physical products in niche markets. Our results show that titles published before 2000 and new titles published exclusively in electronic format extend the standard view of the long tail theory by showing that a significant proportion of titles are specific to a single format. We have already commented on the fact that there are 79 electronic titles that are not available in print but also that 267 print titles do not exist in electronic format. Secondly, the increasing commercial duration of a title has implications for marketing research, requiring new ways to measure cumulative audiences and readership of titles that are sold in different formats over a long period of time.

Secondly, our results have managerial implications for writers, publishers and online intermediaries. On the one hand, the new electronic format creates an opportunity for emerging niche or independent authors to reach their audience as well as to test new markets and pricing strategies. Successful writers of best-sellers in print format can see the electronic format as a mean to increase revenues from new readers who have acquired an e-book reader. New readers are also exposed to older titles that benefit from the increasing life "on the shelves". On the other hand, online intermediaries such as Amazon see the new electronic format as an opportunity to publish authors who may not have signed a contract with a traditional publisher, as well as a new way to differentiate their services by offering both print and electronic versions of a title. It remains to be seen who will provide the traditional publisher/editor services if authors directly go to digital. One possibility is that e-book distributors, such as Amazon, will integrate backwards into editing/publishing, capturing more of the value added. However, editing services are far from an e-book distributor's core competencies, and it is more likely that either traditional publishers, or new online intermediaries will eventually come to dominate over e-book distributors. In any event it seems probable that traditional publishers are likely

to lose bargaining power as authors find that they can self-publish their electronic books using online intermediaries.

4.6 References

Anderson, C. 2005. The Long Tail. Wired, October.

Balasubramanian, S. 1998. Mail Versus Mall: A Strategic Analysis of Competition Between Direct Marketers and Conventional Retailers. Marketing Science. 17(3), 181-195.

Biyalogorsky, E., Naik P., 2003. Clicks and Mortar: The Effect of On-line Activities on Off-line Sales. Marketing Letters. 14(1), 21-32.

Bounie, D., Eang, B., Waelbroeck P., 2011. Les plateformes de ventes sur Internet: une opportunité pour les industries culturelles. Revue Economique. 62(1), 101-112.

Brynjolfsson, E., Hu, Y., Smith M.D, 2003. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. Management Science, 49(11).

Chevalier, J., Goolsbee A., 2003. Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. Quantitative Marketing and Economics. 1(2), 203-222.

Danaher, B., Dhanasobhon, S., Smith, M.D., Telang, R. 2010. Converting Pirates without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy. Marketing Science. 29(6), 1138-1151.

Dang Nguyen, G., Dejean, S., Moreau, F., 2012, Are streaming and other music consumption modes substitutes or complements?. Available at SSRN: http://ssrn.com/abstract=2025071.

Deleersnyder, B., Geyskens, I., Gielens, K., Dekimpe, M.G, 2002. How Cannibalistic is the Internet Channel? A study of the newspaper industry in the United Kingdom and The Netherlands. International Journal of Research in Marketing. 19(4), 337-348.

Elberse, A., Oberholzer-Gee, F., 2007. Superstars and underdogs: An examination of the long tailphenomenon in video sales. Marketing Science Institute, 4, 49-72.

Ghose, A., Smith, M.D., Telang, R. 2006. Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact. Information Systems Research. 17(1), 3-19.

Hu, Yu Jeffrey and Smith, Michael D. 2011. The Impact of Ebook Distribution on Print Sales: Analysis of a Natural Experiment. Available at SSRN: http://ssrn.com/abstract=1966115.

Huang, L.Y., Hsieh, Y.J., 2012. Consumer electronics acceptance based on innovation attributes and switching costs: The case of e-book readers. Electronic Commerce Research and Applications. forthcoming.

Jiang, Y., Katsamakas, E., 2010. Impact of e-book technology: Ownership and market asymmetries in digital transformation. Electronic Commerce Research and Applications. 9, 386-399.

Koh, B., Murthi, B.P.S., Raghunathan, S., 2010. Shift in Demand for Music: Causal Effect of Online Music Piracy and Digital Music on Album Sales. Mimeo.

Oestreicher-Singer G., Sundararajan A., 2010. Are Digital Rights Valuable? Theory and Evidence from Ebook Pricing. CeDER Working Paper No. 06-01 Working Paper Series.

Rosen, S., 1981. The Economics of Superstars. American Economic Review. 71 (5), 845-858.

Waldfogel, J., 2009. Lost on the web: Does Web Distribution Stimulate or Depress Television Viewing? Information Economics and Policy. 21(2), 158-168.

Zentner A., 2008. Online Sales, Internet Use, File Sharing, and the Decline of Retail Music Specialty Stores. Information Economics and Policy. 20(3), 288-300.

Appendix A

Title	Writer	Publication year	Category	Best rank	Average customer review	Average lifespan (in month)
72 Hours	Shannon Stacey	2006	fiction	3	3	1
My Soul To Lose	Rachel Vincent	2009	children's book	5	3.5	9
The Babysitter'S Code	Laura Lippman	2008	fiction	6	2	1
Icy Heat: A Heat Series Story	Leigh Wyndfield	2008	fiction	7	4	1
When Night Falls	Margaret Daley	2009	fiction	10	2	6
Haley'S Cabin	Anne Rainey	2007	fiction	12	2.5	1
Talking With The Dead	Shiloh Walker	2006	fiction	14	3.5	2
Thin Blood	Vicki Tyley	2010	fiction	15	4	2
Look What Santa Brought	Annmarie McKenna	2007	fiction	16	3	1
Believe	Daniel Oran	2007	fiction	24	4.5	2

Table 4 : Example of e-books not available in print on Amazon

Appendix B

Unique print titles	Minimum ranking	Tom	Diffprice	Average rating	Nb. of books by author
Minimum ranking	1.0000	-	-	-	-
Tom	0.0038	1.0000	-	-	-
Diffprice	0.0544	-0.1846	1.0000	-	-
Average rating	0.0297	0.0874	-0.1435	1.0000	-
Nb. of books by author	-0.1971	-0.0373	-0.0777	-0.1806	1.0000

Table 5 : Correlation between explanatory variables (unique print titles; n=770)

Unique electronic titles	Minimum ranking	Nb. of books by author	Diffprice1	Tom
Minimum ranking	1.0000	-	-	-
Nb. of books by author	-0.1092	1.0000	-	-
Diffprice1	0.0560	-0.0272	1.0000	-
Tom	0.0237	-0.0223	-0.1734	1.0000

Table 6 : Correlation between explanatory variables (unique electronic titles; n=774)

CHAPITRE 5 : DISPERSION DES PRIX SUR INTERNET : UNE ANALYSE EMPIRIQUE SUR LES MARCHES DES LIVRES, CD ET DVD EN FRANCE

Bounie D.[44], Eang B.[45], Sirbu M.[46] and Waelbroeck P.[47]

5.1 Introduction

La distribution de produits culturels sur Internet s'est considérablement développée en France au cours des dernières années avec des enseignes telles que PriceMinister, Amazon, Cdiscount et des plateformes d'échange que sont eBay et Amazon Marketplace.[48] La recherche en économie s'est très tôt intéressée à ce phénomène pour deux principales raisons. La première a trait à l'impact du développement des ventes de biens culturels sur Internet sur les canaux de distribution traditionnels et sur les stratégies de tarification des firmes sur les différents marchés (Larribeau et Pénard, 2002). La seconde est liée à l'accès gratuit et public d'une quantité importante de données sur les transactions, les prix, les qualités des offres (biens culturels neufs et d'occasion), le nombre et la réputation des marchands, etc. Ces données sont particulièrement précieuses parce qu'elles sont difficiles et coûteuses à recueillir sur les marchés physiques et surtout parce que leur collecte permet de tester un certain nombre de théories sur les stratégies de tarification des firmes (Acquisti et Varian, 2005 ; Ghose et Sundararajan, 2006 et 2009), le fonctionnement des marchés (Ellison et Ellison, 2005), la diversité culturelle et la viabilité des modèles d'affaires des firmes (Brynjolfsson et al., 2010b ; Ghose et Gu, 2006), les biens collectifs numériques (Curien et al., 2007), etc. Ces études sont aujourd'hui de

[44] Telecom ParisTech, Department of Economics and Social Sciences.
[45] Telecom ParisTech, Department of Economics and Social Sciences.
[46] Carnegie Mellon University, Department of Engineering and Public Policy and Tepper School of Business.
[47] Telecom ParisTech, Department of Economics and Social Sciences.

[48] Le classement des sites de e-commerce en France réalisé par Médiamétrie et commandité par la Fédération du ecommerce et de la vente à distance (FEVAD) montre qu'eBay, PriceMinister, Amazon et Cdiscount font partie des cinq plus fortes audiences en 2010. Les plateformes d'échange de biens telles qu'eBay et Amazon Marketplace organisent la rencontre de consommateurs et marchands qui commercialisent des biens culturels neufs et d'occasion.

plus en plus nombreuses et constituent un champ de recherche à part entière : l'économie d'Internet (Brousseau et Curien, 2007).

Cette recherche s'inscrit précisément dans ce contexte. Les manuels de microéconomie précisent que le résultat du fonctionnement de marchés en concurrence sur des biens homogènes devrait conduire les producteurs à fixer les mêmes prix (loi du prix unique). Pourtant, après de multiples travaux sur différents produits et marchés traditionnels, la plupart des études empiriques infirment ce résultat (Baye, Morgan et Scholten, 2004 ; Pan et al., 2004) : la dispersion des prix constitue la règle plutôt que l'exception. Le développement du commerce sur Internet qui se caractérise par une absence de barrières géographiques et qui offre aux internautes la possibilité de comparer en temps réel les prix d'un produit identique à l'aide de moteurs de recherche et de comparateurs de prix a très vite constitué une nouvelle opportunité d'étude de la dispersion des prix pour les économistes. En théorie, les prix d'un produit de qualité comparable commercialisés par des firmes sur Internet devraient être identiques puisque tous les internautes peuvent acheter au meilleur prix.[49] Ce raisonnement semble d'autant plus valide que les coûts d'ajustement des prix sur un site web sont plus faibles que ceux liés à l'impression d'un catalogue ou à l'étiquetage des prix en magasin (Brynjolfsson et Smith, 2000). Or, à nouveau, non seulement les études empiriques confirment l'existence d'une forte dispersion des prix, mais cette dispersion est persistante et ne diminue pas au fil du temps.

De nombreux travaux ont tenté d'apporter des éléments de réponse à ce phénomène en se concentrant principalement sur les coûts de recherche et de traitement de l'information par les consommateurs. En effet, Varian (1980) a très tôt montré qu'en présence de coûts de recherche d'information pour les consommateurs (contraintes géographiques, etc.), les firmes sont incitées à pratiquer des prix différents à l'équilibre. Prolongeant ce travail, Clay et al., (2002), Brynjolfsson et Smith (2000), Brynjolfsson et al., (2010a) ont montré que les coûts de recherche pouvaient également expliquer en partie la dispersion des prix sur Internet même lorsque les internautes utilisent des

[49] Il existe même des sites de comparatifs de prix qui fournissent une liste des différentes offres disponibles en ligne. Par exemple kelkoo.fr permet de consulter un vaste catalogue de produits par catégorie de prix et de comparer directement les offres, frais de port compris.

comparateurs de prix. Mais les auteurs reconnaissent tout de même qu'une part non négligeable de la dispersion des prix provient d'autres dimensions.

Dans cet article, nous nous intéressons à un ensemble d'autres facteurs liés à la dimension temporelle, aux stratégies des vendeurs, à la structure du marché et à la différenciation des vendeurs.[50] En particulier, nous tentons de mettre en évidence que la dispersion des prix est liée en partie à la stratégie de tarification des vendeurs qui tirent profit de leur réputation et de la structure de marché qui en découle. Par exemple, la réputation d'un vendeur acquise sur la plateforme peut lui permettre de fixer des prix plus élevés (prime à la réputation), ce qui crée de la dispersion des prix et concentre les ventes sur un petit de nombre de vendeurs jugés fiables par les internautes. Le rôle de la réputation est donc un élément important pour comprendre l'évolution de la dispersion des prix.

Pour mettre en évidence ces effets, nous avons collecté et suivi dans le temps toutes les offres de livres, CD et DVD qui sont entrés dans le classement des cent meilleures ventes journalières sur la plateforme de vente française Amazon Marketplace sur la période de mars à novembre 2006. Les internautes qui utilisent la plateforme Marketplace peuvent en effet comparer les prix des différentes offres d'un livre, CD ou DVD commercialisé par de multiples vendeurs (y compris Amazon). Les offres que nous étudions portent exclusivement sur des biens neufs : nous excluons les offres de biens d'occasion afin d'étudier des biens homogènes.

Notre étude contribue à la littérature existante sur plusieurs points. Premièrement, en utilisant des données de prix affichés par les vendeurs (appelés prix postés), nous montrons que les vendeurs modifient très peu leurs prix et lorsqu'ils le font les variations de prix constatées sont très faibles. Ceci remet en question les explications liées à la tarification stratégique des vendeurs. Deuxièmement, nous construisons une méthode pour identifier les prix de transaction à partir des prix postés et nous comparons les prix postés aux prix de transaction ; nous montrons que les prix de transaction ne correspondent pas aux prix les plus faibles. Ce résultat conforte les analyses de Brynjolfsson et al.,

[50] Notre article ne traite pas des coûts de recherche d'information sur les produits qui s'avèrent difficiles à quantifier. En outre, la nature de nos données conduit à minimiser ce problème car les consommateurs disposent en temps réel de l'ensemble des prix et des informations disponibles sur un produit donné ainsi que sur les vendeurs.

(2010a) et s'opposent à celles de Ghose (2006). Troisièmement, nous quantifions une prime à la réputation des vendeurs qui explique de manière significative la dispersion des prix même si les caractéristiques des vendeurs n'expliquent globalement qu'une faible partie de la dispersion des prix en ligne. Ce travail est complémentaire à ceux traditionnellement réalisés sur les sites d'enchère en ligne tels qu'eBay (Resnick et Zeckhauser, 2002 ; Houser et Wooders, 2006). De même, il étend le travail de Ghose et al., (2009) sur le rôle de la réputation des vendeurs sur Amazon étudié dans le cadre de prix postés de produits neufs et d'occasion en analysant des prix de transaction. Quatrièmement, nous montrons que la dispersion des prix postés en ligne augmente avec le nombre de vendeurs. Ce résultat semble intuitif mais va à l'encontre des résultats établis dans la littérature. En effet, Spulber (1995) et Baye, Morgan et Scholten (2004) ont montré que la dispersion des prix décroît avec le nombre de vendeurs.

L'article est organisé comme suit. Dans une deuxième section, nous présentons et décrivons les données. Dans une troisième section, nous étudions la dispersion des prix sur les offres proposées par les vendeurs. Dans une quatrième section nous analysons la dispersion des prix de transaction et nous estimons les déterminants des prix de transaction. Enfin, nous concluons cet article dans la section 5.

5.2 Description des données

L'objectif principal de cet article est d'analyser la dispersion des prix des produits culturels neufs vendus sur les plateformes de ventes Amazon Marketplace.[51] Amazon Marketplace est la plateforme de vente d'Amazon qui permet à des vendeurs professionnels et des particuliers de proposer aux internautes des offres à côté de celles d'Amazon. En d'autres termes, lorsqu'Amazon commercialise un livre par exemple, des vendeurs professionnels et des particuliers ont la possibilité de proposer à la vente leurs exemplaires.[52] Le site Amazon et les vendeurs Marketplace sont indépendants et concurrents.[53]

[51] Nous écartons de notre analyse les produits d'occasion afin de comparer les prix de produits homogènes.

[52] Une commission variable en fonction du prix du produit commercialisé est prélevée par Amazon au vendeur qui a conclu la vente.

[53] Des filiales d'Amazon peuvent opérer sur la plateforme Marketplace.

Pour analyser la dispersion des prix, nous avons relevé quotidiennement à l'aide de scripts PERL automatisés, tous les livres, CD et DVD qui sont entrés dans le classement des cent meilleures ventes sur Amazon et nous avons collecté toutes les offres associées à ces produits et proposées par les vendeurs sur les plateformes Marketplace. Le suivi des offres a débuté en mars 2006 et s'est achevé en novembre 2006, soit une période d'observation de 260 jours. Nous avons continué à suivre tous ces produits, même s'ils ne faisaient plus partie des cent meilleures ventes d'Amazon par la suite. Au final, nous disposons de trois sources d'information sur i. les caractéristiques du produit i (un identifiant unique (ISBN, ASIN)), titre, type de produit (couverture cartonnée pour un livre, édition double pour un DVD, etc.) ; ii. les caractéristiques de l'offre k du produit i à la date t sur la plateforme (prix, classement de vente, prix d'Amazon, état du produit (neuf ou occasion)) ; iii. enfin, les caractéristiques du vendeur j qui a proposé l'offre k pour le produit i à la date t sur la plateforme Marketplace (note laissée par les internautes, nombre de notes reçues, prix, etc.).

Avant de commenter plus en détails les données de prix, il est important de souligner quelques problèmes liés à la collecte de données sur Internet. Premièrement, certains vendeurs postent des offres de prix au-dessus de 3 millions d'euros par exemple. Il s'agit vraisemblablement d'une erreur de saisie. Pour supprimer l'influence des ces offres extrêmes sur le calcul d'une mesure de dispersion de prix, nous avons éliminé toutes les offres de prix deux fois supérieures au prix médian (de chaque produit) dans nos données. Cette correction permet en outre de régler en partie le problème des offres à prix très élevées sur la plateforme qui correspondent à des produits uniques (signés par l'auteur) ou à des produits de collection (par exemple des premières éditions de livre ou de DVD). Deuxièmement, certains produits sont proposés à 1 centime sur les plateformes de livres et de CD. Ces offres augmentent de manière artificielle la dispersion des prix, en particulier si cette dernière est mesurée à partir du prix le plus faible comme par exemple la différence entre l'offre la plus élevée et l'offre la plus faible ou encore, la différence entre les deux offres les plus faibles. Nous avons décidé de supprimer aussi ces offres qui représentent une partie très marginale de nos données. Troisièmement, environ 10% des offres correspondent à des offres proposées pour un produit par le même vendeur le

même jour, et ce à des prix différents. Nous traitons ces offres comme si elles correspondaient à une seule offre (au prix le plus faible) avec plusieurs exemplaires en stock.

Le Tableau 1 fournit une vue générale du nombre d'offres disponibles pour les différents produits. Il montre que le turnover dans le Top 100 est le plus important sur le marché du CD (3029 titres pour 4,6 millions d'offres) suivi par celui du livre (2505 titres pour 3,4 millions d'offres) et du DVD (1383 titres pour 1,2 millions d'offres).

Produits	**Livre**	**CD**	**DVD**
Nombre d'offres	3 486 379	4 591 715	1 259 779
Nombre de prix à 1 centime	10	1	0
Nombre de prix > 2 écarts-type	85 416	113 049	31 673
Nombre de produits	2 505	3 029	1 383
Nombre de vendeurs	1 170	691	1 128
Nombre moyen quotidien de vendeurs par produit	**7,9**	**9,8**	**5,5**
Nombre moyen de meilleurs vendeurs	5,1	5	2,2
Nombre moyen d'autres vendeurs	2,9	4,8	3,2
Note moyenne des vendeurs	**4,5**	**4,5**	**4,5**
Note moyenne des meilleurs vendeurs	4,5	4,5	4,5
Note moyenne des autres vendeurs	4,5	4,5	4,5
Nombre moyen de notes	**8 843**	**34 779**	**7 701**
Nombre moyen de notes des meilleurs vendeurs	10 773	63 477	4 297
Nombre moyen de notes des autres vendeurs	2 376	4 901	8 588
Prix moyen d'un produit	**11,72**	**14,71**	**18,71**
Prix moyen du produit d'un meilleur vendeur	11,79	14,57	18,76
Prix moyen du produit d'un autre vendeur	11,66	14,67	17,94

Tableau 1: Description des données

En outre, la plateforme de CD est la plus concurrentielle en France avec en particulier un nombre moyen de vendeurs par produit et par jour plus important que sur les autres marchés. En outre, le marché du CD est très concentré puisque le nombre de vendeurs est plus de une fois et demie inférieur à celui observé sur les marchés des livres et DVD. Nous pouvons distinguer deux types de vendeur que nous qualifions de « meilleur vendeur » et « d'autre vendeur ». Un meilleur vendeur appartient par définition au top 10 des vendeurs par marché et dispose du plus large catalogue de produits offerts (par opposition un autre vendeur n'appartient pas à ce top 10). Nous constatons que pour les livres, les meilleurs vendeurs reçoivent plus de notes et postent des prix plus élevés ; pour les CD, les meilleurs

vendeurs ont beaucoup plus de notes mais ont des prix plus faibles, alors que l'inverse est vrai pour les DVD.

L'analyse statistique de nos données est rendue difficile d'une part par le nombre d'offres suivies (plus de 9 millions) et d'autre part le fait que chaque produit est caractérisé par trois dimensions. La première dimension correspond au produit lui-même. La deuxième dimension correspond au temps puisque nous suivons chaque produit à partir du moment où il entre dans le top 100 jusqu'à la fin de la période d'observation. La troisième dimension est liée au nombre de vendeurs proposant à la vente un produit particulier sur la plateforme. La notion de dispersion de prix dépend donc de la dimension que nous étudions. Dans cet article, nous nous intéressons à la dispersion des prix générée par les différentes offres de vendeurs. Cependant, pour être complet, nous analysons tout d'abord la manière dont le prix des produits évolue au cours du temps.

5.3 Dispersion des prix affichés ou postés: stratégies de prix et dimension temporelle

Pour étudier l'effet de la dimension temporelle sur la dispersion des prix, nous construisons un panel équilibré de produits en vue d'éliminer une dispersion artificielle des prix dans le temps due aux entrées et sorties de produits. Pour ce faire, nous gardons tous les produits disponibles à partir du début de la troisième semaine d'observation (deuxième dans le cas de CD) et nous les suivons jusqu'à la fin de la période d'observation.[54],[55]

Pour chaque catégorie de produit (livre, CD, DVD), nous donnons plusieurs statistiques. Nous calculons tout d'abord le prix moyen d'un produit de la manière suivante : nous calculons la moyenne journalière (dans le temps) pour chaque paire (produit / vendeur), puis la moyenne par vendeur et, enfin, la moyenne par produit[56]. Nous utilisons une procédure similaire pour calculer l'écart-type et le

[54] Certains produits sont entrés puis sortis du Top 100 durant la première semaine de la période; par exemple, 201 livres sont entré dans le top 100 durant la première semaine, et seuls 199 sont restés la semaine d'après.

[55] Pour être plus précis, la première observation pour le marché du livre est le 72ème jour, pour le CD le 84ème jour et pour le DVD le 72ème jour (ces jours correspondent tous à des lundis).

[56] Il s'agit du prix hors frais de livraison. Nous ne connaissons pas d'articles qui modélisent différents frais de livraison. Cependant ces frais de livraison sont standardisés et ne sont pas nécessairement supérieurs à ceux

prix minimum moyen ainsi que l'écart-type du nombre de vendeurs. Le Tableau 2 résume les statistiques descriptives pour les prix affichés par les vendeurs sur Marketplace.

Nous pouvons effectuer quelques comparaisons intéressantes. Tout d'abord, le nombre total de vendeurs est similaire entre les marchés, autour de 330, mais les offres proposées par les vendeurs sont plus de trois fois plus nombreuses sur les marchés du livre et du CD que sur le marché du DVD.

Ensuite, nous constatons qu'un grand nombre de vendeurs n'offrent qu'un seul produit ; ce constat est particulièrement vrai pour le marché du livre où 62% des vendeurs ne possèdent qu'une offre. Nous en concluons qu'une plus grande proportion de particuliers poste des offres sur cette plateforme. Cette observation affecte sans doute la durée des offres par les vendeurs sur ces marchés puisque nous constatons que la durée médiane d'une offre par un vendeur est deux à trois plus élevée sur le marché du livre que sur les marchés des CD et DVD ; elle s'élève plus précisément à 19 semaines sur le marché du livre et respectivement à 11 et 5 semaines sur les marchés des CD et DVD. Enfin, l'analyse des prix indique que le prix moyen d'un produit sur Amazon est assez concurrentiel puisqu'il est légèrement inférieur à ceux des vendeurs sur la plateforme Marketplace. Soulignons également pour conclure que les écarts de prix sur le marché du livre sont plus restreints en raison de la loi du prix unique qui limite les rabais sur le prix éditeur d'un livre neuf à 5 % au maximum. En effet, le prix minimum moyen est plus proche du prix moyen pour les livres en France que pour les autres produits ou les autres pays. Il est intéressant également de souligner que le coefficient de variation pour les CD est presque deux fois supérieur à celui du livre et du DVD. Le marché du CD a également le plus grand nombre de vendeurs pour chaque item.

d'Amazon (voir le site du fonctionnement de la plateforme Marketplace). De plus, les frais de livraison sont clairement affichés dans la colonne Prix + Frais de livraison sur la page Marketplace.

Produits	Livre	CD	DVD
Nombre d'offres	459 249	442 520	133 779
Nombre d'offres à 1 centime	10	0	0
Nombre d'offres > à 2 écarts-type	30 342	9 194	2 943
Nombre de produits	199	177	97
Nombre de vendeurs	334	310	364
Nombre moyen de vendeurs par jour (par offre)	9,4	11,2	6,3
Ecart-type du nombre de vendeurs	3,8	5,8	3,9
% de vendeurs offrant un seul produit	66	42	46
Durée médiane en nombre de jours d'une offre par un marchand	134	78	40
% des offres proposées par des vendeurs qui disposent de plus 75% du catalogue	17	26	10
Prix moyen d'Amazon	10,79	15,37	19,24
Prix moyen des offres sur Marketplace	10,93	16,38	20,01
Ecart type du prix des offres sur Marketplace	2,01	4,58	3,68
Prix minimum moyen des offres sur Marketplace	8,95	11,16	16,62
Prix médian des offres sur Marketplace	10,45	15,72	19,65

Tableau 2: Statistiques descriptive sur le panel équilibré

Dans la partie qui suit, nous nous intéressons à la distribution des changements de prix sur les marchés dans le temps ainsi qu'à la dispersion des prix dans le temps.

5.3.1 Distribution des changements de prix sur Marketplace

La première source de dispersion des prix avancée par Baye et Morgan (2001) est due en particulier à l'utilisation de stratégies mixtes ou aléatoires par les vendeurs dans le temps.

5.3.1.1 Variabilité des prix des vendeurs au cours du temps

Pour étudier les changements de prix des vendeurs, nous pouvons tout d'abord calculer le nombre de fois qu'un vendeur augmente, diminue ou laisse inchangé le prix d'un produit mis en vente d'un jour à l'autre.[57] Le Tableau 3 résume ces changements de prix d'une part pour Amazon et d'autre part pour les vendeurs sur Marketplace.

Le premier constat est qu'Amazon ne change pas ses prix dans 90% des cas et que la distribution des changements de prix est relativement symétrique : en d'autres termes, il existe autant

[57] Dans quelques cas très restreints, certaines observations sur les prix sont manquantes dans notre base de données. Nous utilisons alors la dernière observation dans nos calculs.

d'augmentations de prix que de baisses. En revanche, la distribution des changements de prix des vendeurs est plus asymétrique avec un plus grand nombre de baisses que de hausses sur les marchés du CD et DVD (ce qui traduit donc une baisse de prix sur la période d'observation). De plus, les prix sont plus variables sur la plateforme que sur le site Amazon. Nous observons plus de changements de prix pour les CD que pour les livres et les DVD. Nous devons tout de même souligner que dans 86% des cas, les prix restent inchangés. Ainsi, les stratégies de prix individuelles ne peuvent pas être la source principale de la persistance de dispersion de prix. Au contraire, la dispersion des prix peut provenir justement du fait que les vendeurs n'ajustent pas régulièrement leur prix.

	Livre	**CD**	**DVD**
Amazon			
Constant	99,25	90,06	92,54
Baisse	0,34	4,56	3,52
Hausse	0,41	5,38	3,94
Vendeurs			
Constant	98,81	86,35	95
Baisse	0,71	8,13	3,67
Hausse	0,48	5,52	1,33

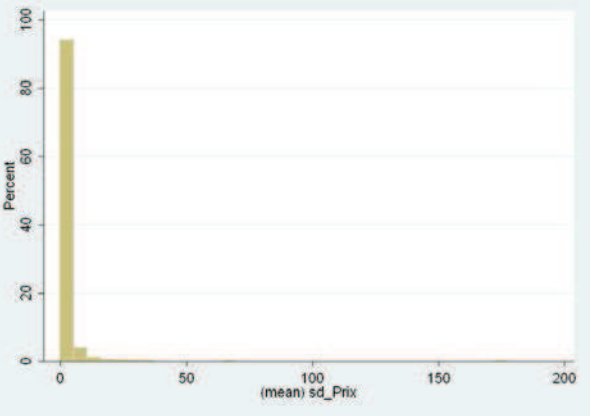
Tableau 3: Distribution des changements de prix d'Amazon (en %)

Pour tester l'hypothèse selon laquelle les vendeurs utilisent des stratégies de prix aléatoires, nous calculons pour chaque produit la distribution des prix pour tous les vendeurs sur l'ensemble de la période d'observation. Nous divisons ensuite ces distributions en quartiles. Nous examinons combien de fois un vendeur change son prix en le faisant passer d'un quartile de la distribution à un autre au cours de l'ensemble de la période d'observation. Nous trouvons que, dans la grande majorité des cas, les vendeurs ne changent pas du tout leur prix ou en tout cas pas suffisamment pour changer de quartile de la distribution des prix dans lequel ils se trouvent. La probabilité d'offrir un produit à la fois dans le quartile le plus haut et le plus bas de la distribution des prix est inférieur à 6% (Tableau 5). Nous concluons donc que les stratégies de tarification des vendeurs semblent avoir un effet marginal sur la dispersion des prix.

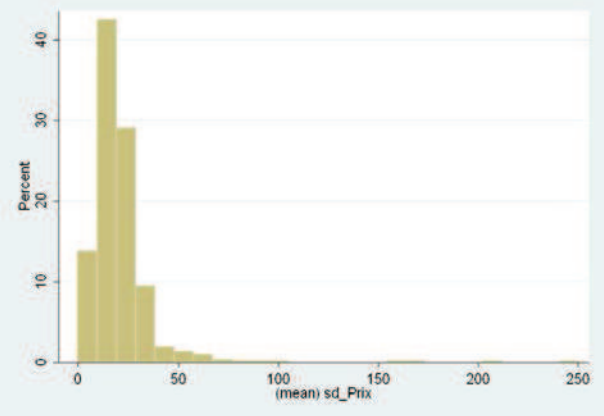
	Livre	CD	DVD
% des prix des vendeurs qui restent dans le même quartile sur l'ensemble de la période	70,7	58,4	66,8
% des prix des vendeurs qui apparaissent au moins un jour dans le quartile le plus élevé et le plus faible	2,4	5,7	4,8

Tableau 4: Fréquence à laquelle les vendeurs modifient leur prix de manière significative

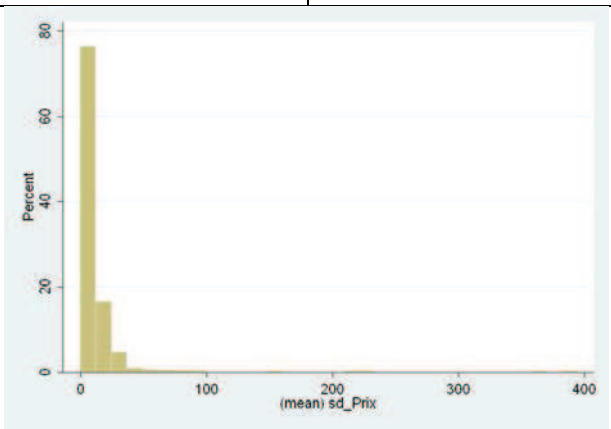
Une autre manière de vérifier si les vendeurs changent leurs prix de manière conséquente est de calculer l'évolution de l'écart-type des prix (moyenne sur l'ensemble des vendeurs) au cours du temps (en utilisant la semaine comme unité de temps) pour chaque produit. Nous calculons ensuite la moyenne pour l'ensemble des produits et représentons les histogrammes des écart types sur chaque marché. L'écart-type moyen des changements de prix est de l'ordre de 25 centimes, ce qui remet également en question l'hypothèse de prix aléatoire des vendeurs. Les observations dans le Graphique 1 correspondent aux écarts-type des offres de prix par un vendeur unique durant une semaine. Nous calculons ensuite la moyenne de ces données sur tous les vendeurs pour chaque produit. Nous prenons ensuite la moyenne sur tous les produits. Le résultat de ces doubles moyennes correspond à l'axe des x, (mean) sd_Prix.



Livre



CD



DVD

Graphique 1: Histogrammes des écart-type moyens des prix en centimes

5.3.1.2 Evolution de la dispersion des prix au cours du temps

Nous avons déjà commenté le fait que les prix diminuent dans le temps. Nous étudions à présent l'évolution de la dispersion des prix et nous la mettons en parallèle avec l'évolution du nombre de vendeurs. De manière plus formelle, nous étudions la façon dont le nombre de vendeurs augmente en utilisant un modèle de panel avec effets fixes en utilisant le nombre de jours depuis qu'un produit est apparu dans le TOP 100 (et donc dans nos données) ainsi qu'une constante comme variables explicatives. De manière générale, nous trouvons un effet positif qui traduit une tendance positive du nombre de vendeurs par produit. Par exemple, après 216 jours passés dans nos données, il y a aura un vendeur supplémentaire pour les livres en France (Tableau 6)

Produit	**Livre**	**CD**	**DVD**
Variable dépendante	Nombre de vendeurs	Nombre de vendeurs	Nombre de vendeurs
Nombre de jours	0,00462***	0,00114***	-0,00271***
	[0,00011]	[0,00015]	[0,00021]
Constante	9,05909***	11,11373***	7,42871***
	[0,01622]	[0,02038]	[0,03061]
R^2	0,03438	-0,00296	0,00381
N	47 468	39 202	18 869

Tableau 5: Nombre de vendeurs et nombre de jours (avec effets fixes)

Pour déterminer la part relative du temps et du nombre de vendeurs sur l'explication de l'évolution de la dispersion des prix, nous estimons les paramètres d'un modèle où la variable dépendante est le coefficient de variation (exprimé en pourcentage) et les variables explicatives sont le temps et le nombre de vendeurs. Dans le Tableau 6, nous avons utilisé un modèle avec effets fixes à nouveau. Il ressort des estimations que la dispersion des prix est influencée de manière positive par ces deux facteurs et que le nombre de vendeurs a un effet plus important. Ainsi, si le nombre de vendeurs augmente d'un écart-type au dessus de la moyenne (4 vendeurs pour les DVD en France), nous obtenons que le coefficient de variation augmente de 4,8%.

Produit	**Livre**	**CD**	**DVD**
Variable dépendante	Coefficient de Variation	Coefficient de Variation	Coefficient de Variation
Nombre de jours	0,00521***	0,00807***	0,00684***
	[0,00027]	[0,00042]	[0,00073]
Nombre de vendeurs	0,42187***	0,48292***	1,19991***
	[0,01166]	[0,01527]	[0,02541]
Constante	14,97511***	22,20086***	11,48760***
	[0,11394]	[0,18352]	[0,22584]
R^2	0,03884	0,03041	0,10735
N	46 690	38 527	17 866

Tableau 6: Coefficient de Variation (%), nombre de jours et nombre de vendeurs (avec effets fixes)

Cette partie a permis de montrer que les vendeurs ne changent pas souvent leur prix et lorsqu'ils le font les changements de prix sont de faible amplitude.

Une limite importante de ces dernières analyses porte toutefois sur la nature des données observées. A l'image de la plupart des travaux dans ce domaine, la question de la dispersion des prix est appréhendée, en raison de l'absence de données sur les ventes, sur les prix des offres affichées. Ces offres ne donnent pourtant pas nécessairement lieues à des ventes. Or, il est possible que la dispersion des prix observée sur les prix affichés ou postés disparaisse lorsque nous étudions les prix de vente des produits. Dans la partie qui suit, nous nous concentrons sur les prix de vente et nous montrons qu'il existe toujours un phénomène de dispersion.

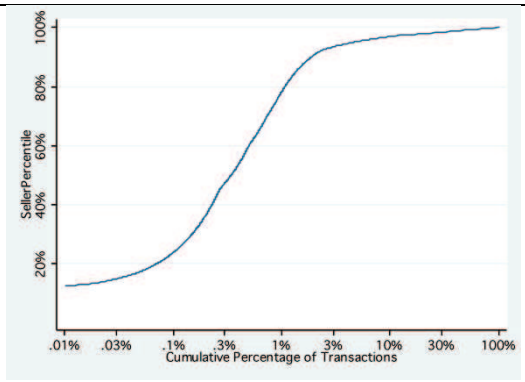
5.4 Dispersion des prix de vente : caractéristiques des produits et des vendeurs

Nous commençons cette section par préciser la manière dont nous mesurons une transaction. Ensuite, nous comparons la dispersion des prix de vente à celle des prix postés présentée dans la section précédente. Enfin, nous estimons l'effet de la réputation d'un vendeur et des caractéristiques des produits sur les prix de vente en utilisant une régression hédonique.

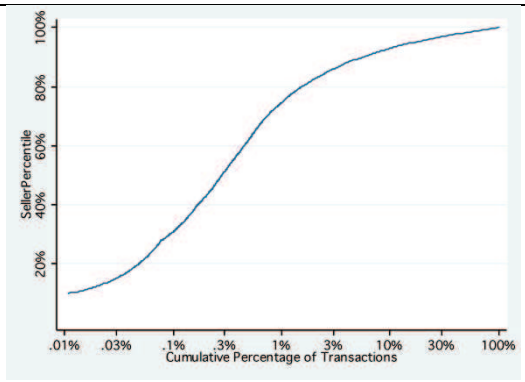
5.4.1 Détermination d'une transaction

Nous déterminons si une transaction (vente) a eu lieu de la manière suivante. Nous observons des offres de prix pour les vendeurs d'un produit particulier à chaque date. Parmi les offres proposées, certaines se concrétisent par une transaction. Nous considérons qu'une offre proposée par un vendeur *j* à la date *t* résulte d'une transaction dans deux cas : si l'offre n'est plus jamais disponible après la date *t* ce qui signifie que le vendeur ne poste plus d'offre concernant ce produit après la date *t*) ou si l'offre disparaît après la date *t* et réapparaît à une date ultérieure à *t+2*. Par exemple, un produit est proposé par un vendeur sur la plateforme à la date *t*, n'est plus proposé à la vente en *t+3* et, ai à nouveau proposé par ce même vendeur à la vente en *t+5*. En d'autres mots, ce produit n'était pas disponible durant 2 jours *(t+3 et t+4)* et nous considérons donc qu'une vente a eu lieu. En outre, nous considérons qu'il y a également une vente si le produit était à nouveau proposé à la vente par ce même vendeur en *t+4* (auquel cas l'item n'était pas disponible pendant un jour).

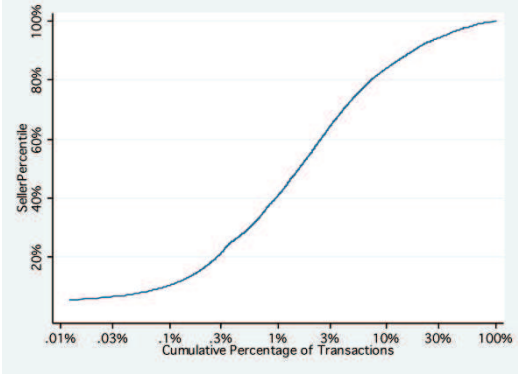
La distribution des transactions est représentée dans le Graphique 2 pour chaque marché. Ces graphiques confirment qu'un petit nombre de vendeurs réalise la plus grande partie des transactions sur tous les marchés. Ainsi, 95% des transactions sur le marché du livre sont réalisées par 5% des vendeurs.



Livre



CD



DVD

Graphique 2: Distribution des transactions par vendeur (échelle log)

Avant de poursuivre, il est important de noter que cette définition d'une vente a des limites puisqu'elle peut conduire soit à surestimer soit à sous-estimer les ventes. D'un côté, nous pouvons surestimer les ventes lorsqu'un vendeur retire une offre de son catalogue qu'il a vendue par ailleurs par un autre moyen (réseau physique ou vente directe, etc.). D'un autre côté, notre définition peut sous-estimer les ventes : tant qu'une offre est affichée, nous ne pouvons pas identifier une vente. Ainsi un vendeur disposant d'un stock important d'un produit peut en vendre le jour *t* et poster à nouveau une autre offre au même prix le jour suivant. Dans ce cas, notre méthode ne permet pas d'identifier

cette vente. Cette définition pénaliserait donc les meilleurs vendeurs qui ajustent régulièrement leurs offres sur Marketplace. Néanmoins, nous venons de montrer qu'un petit nombre de vendeurs réalisent la grande majorité des transactions ; dès lors, notre définition ne semble pas outre mesure affectée par ce phénomène.

5.4.2 Comparaison des prix postés et des prix de vente

Le Tableau 7 résume quelques statistiques descriptives sur les prix des transactions. Nous calculons l'écart-type des prix de transaction pour chaque produit sur l'ensemble de la période et le comparons à l'écart-type des prix postés. Le tableau montre qu'il subsiste une dispersion des prix substantielle et que le prix de transaction moyen est au-dessous du prix d'Amazon alors que les prix postés sont en moyenne supérieurs aux prix d'Amazon; ceci limite la portée des études qui font l'hypothèse que les prix de transaction sont concentrés autour du prix posté le plus faible, telles que Baye, Morgan et Scholten (2004). Ces comparaisons varient d'un marché à l'autre : par exemple, à cause de la loi du prix unique, la dispersion des prix est beaucoup plus faible sur la plateforme des livres que sur les autres marchés.

Produit	**Livre**	**CD**	**DVD**
Prix moyen	10,36	17,56	19,29
Prix moyen d'Amazon	10,68	16,48	19,66
Ecart type des prix	1,53	4,25	3,13
Prix médian	10,28	17,4	19,1
Prix minimum	9,71	14,53	17,61
Prix maximum	11,24	21,17	21,39
Percentile 75 (prix)	10,88	19,57	20,59
Percentile 95 (prix)	11,23	21,1	21,39
Nombre de ventes	14 167	17 268	6 935
Nombre de vendeurs	320	275	348
Nombre de produits	201	129	97

Tableau 7: Description des prix de vente[58]

Dans la partie suivante, nous analysons les déterminants des prix de transaction en étudiant la prime de réputation d'un vendeur ainsi que les caractéristiques des produits mis en vente.

[58] Le nombre total d'observations change marginalement car certains livres n'ont pas reçu de transactions.

5.4.3 Déterminants des prix de vente : caractéristiques des vendeurs et du temps

Dans cette partie, nous expliquons les prix de transaction (p_{ij}) d'un produit *i* vendu par le vendeur *j* sur la plateforme Marketplace par la réputation des vendeurs et les caractéristiques du produit. On peut envisager plusieurs variables pour capturer la réputation d'un vendeur. A l'image de Ghose et al. (2009), nous utilisons la note moyenne d'un vendeur et le nombre de notes qu'il a reçues par le passé. Ces variables servent de proxy pour la réputation et les quantités vendues par le vendeur (les vendeurs qui gèrent un stock important de livres, CD et DVD peuvent plus facilement internaliser les coûts et offrir des réductions intéressantes). Nous considérons également une variable que nous pensons corrélée avec la réputation : le nombre de ventes réalisées par un vendeur. Nous considérons deux mesures de la taille d'un vendeur : le nombre de produits différents offerts par le vendeur sur un marché particulier et le nombre total de transactions de ce vendeur. Ces deux mesures sont très fortement corrélées (avec un coefficient de corrélation typiquement supérieur à 0,9) et nous utilisons seulement la premièrement variable liée à la taille du catalogue. En revanche, ces deux variables sont très peu corrélées aux variables traditionnelles de la réputation que sont la note et le nombre de notes (0,2). Enfin, nous contrôlons pour la durée de présence d'un produit sur le marché (le nombre de jour depuis sa date de parution) afin de capturer une tendance de prix dans le temps.

Pour déterminer l'effet de ces variables sur le prix de transaction, nous estimons les paramètres d'un modèle avec effets fixes sur l'ensemble du panel non équilibré sur une période d'observation de 22 semaines. Les résultats des estimations sont résumés dans le Tableau 8. Les régressions permettent de tirer plusieurs enseignements.

	Livre	CD	DVD
Variable dépendante	Prix	Prix	Prix
Note	4,21	0,337	-3,59
Nombre de notes	-0,156	-6,35***	-16,0**
Catalogue	-0,011***	0,0551***	0,0550***
Time on Market (TOM)	0,130***	-0,0188	1,66***
TOM²	-0,00001	0,00005**	-0,00030***
Effet fixe par produit	Oui	Oui	Oui
Constante	956***	1170***	794***
R^2 (within)	0,0023	0,0111	0,00861
N	48 954	82 762	29 287
Nombre de produits	2505	3020	1381
Réputation Premium pour un écart type de changement dans le catalogue (€)	-0,08	0,47	0,23

Tableau 8: Résultats des estimations

Tout d'abord, la mesure traditionnelle de réputation liée à la note moyenne d'un vendeur n'influence pas significativement le prix de vente et constitue donc une variable *proxy* très imparfaite de la réputation d'un vendeur. En revanche, le nombre de notes a un effet négatif sur les prix des CD et DVD alors que la taille du catalogue (et le nombre de transactions) a un effet positif sur les prix des livres, CD et DVD. La taille du catalogue semble donc mieux capturer la réputation d'un vendeur : en particulier, l'augmentation d'un écart-type de la taille du catalogue peut être associée à une prime de réputation qui va de 23 à 47 centimes d'euros sur les marchés des CD et DVD. Cependant, il est à noter que les vendeurs qui disposent d'un large catalogue ne peuvent exploiter leur réputation pour accroître les prix des livres puisque l'effet est négatif. Ce résultat est probablement attribuable à la réglementation du prix du livre qui limite les augmentations de prix et la concurrence en prix entre les différents vendeurs. Plus généralement, nous pouvons conclure que les variables de réputation d'un vendeur semblent capturer différents effets et que les études qui utilisent le nombre de notes comme une variable proxy du nombre de transactions sont incorrectes car leurs effets sont opposés. Cette différence de signe est sans doute liée à la possibilité donnée aux consommateurs de réagir négativement aux transactions conclues avec certains marchands.

Ensuite, la durée de présence d'un produit sur le marché (mesurée en nombre de jours depuis la date de sortie sur le marché) semble avoir un effet ambigu sur les CD et sur les livres et les DVD. Ce résultat s'explique par le fait que la date de transaction peut précéder la date de parution en raison des possibilités de préachat de produits culturels (la date de sortie officielle est donc postérieure à la date d'achat et la valeur de la variable est donc négative). Dans d'autres cas, des classiques de plus de 40 ans apparaissent dans les classements des ventes avec une durée sur le marché de plus de 40 ans.

Enfin, il est à noter que la qualité des ajustements sur les différentes plateformes est très faible et qu'une partie substantielle de la variance des prix demeure inexpliquée par nos variables relatives aux caractéristiques des vendeurs et des produits. Des travaux complémentaires sont donc nécessaires pour déterminer l'origine de la dispersion des prix.

5.5 Conclusion

Au final, notre contribution permet de tirer plusieurs enseignements. Premièrement, il existe de la dispersion des prix sur Internet et celle-ci augmente avec le nombre de vendeurs. Deuxièmement, les explications liées aux stratégies individuelles de prix des vendeurs ne sont pas convaincantes. Troisièmement, la réputation d'un vendeur semble jouer un rôle important même si le mécanisme qui lie la tarification à la réputation nécessite d'être approfondi. En particulier, nous n'expliquons pas les origines de la prime à la réputation. Il est important de souligner que la notion de réputation est difficile à capturer par les variables traditionnelles liées à la note et au nombre de notes d'un vendeur. Ces dernières n'ont pas toujours les signes attendus et ne sont pas robustes. Dans cet article, nous avons exploré des mesures complémentaires comme la taille du catalogue et un effet de clientèle mais d'autres mesures devraient être étudiées.

Enfin, pour conclure, il est important de souligner qu'il peut persister des problèmes informationnels pour les consommateurs qui peuvent prendre plusieurs formes et qui n'ont pas été capturés ni dans notre recherche ni dans les recherches existantes. Tout d'abord, les consommateurs ont des préférences variées pour certaines caractéristiques liées au meilleur prix, à la réputation d'un vendeur, etc. Ces caractéristiques de la demande évoluent sans doute au cours du temps et les premiers

consommateurs n'ont pas les mêmes caractéristiques que les consommateurs plus tardifs puisque les biens que nous étudions sont durables. Ensuite, il existe sans doute des variations temporelles liées au jour de la semaine. Les consommateurs du « week end » peuvent avoir des caractéristiques différentes de celles des consommateurs de la semaine et qui seraient susceptibles d'influencer la dispersion des prix (est-ce que les consommateurs les plus informés sont ceux du week-end auquel cas les prix devraient converger vers le prix minimum plus rapidement durant le week-end ?).

5.6 Références

Acquisti, A., et Varian H., (2005), "Conditioning Prices on Purchase History", Marketing Science, 24(3): 367-381.

Baye, M.R., et Morgan, J., (2001), "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets", American Economic Review, 91(3): 454-474.

Baye, M.R., Morgan, J., et Sholten, P., (2004), "Price Dispersion In The Small And In The Large: Evidence From An Internet Price Comparison Site", Journal of Industrial Economics, 463-496.

Brousseau, E. et Curien N., (2007), Internet and Digital Economics, Cambridge University Press.

Brynjolfsson, E. et Smith, M., (2000), "Frictionless Commerce? A Comparison of Internet and Conventional Retailers", Management Science, 46(4).

Brynjolfsson, E., Dick A.A., et Smith M.D., (2010a), "A Nearly Perfect Market? Differentiation vs. Price in Consumer Choice," Quantitative Marketing and Economics, November, 8(1): 1-33.

Brynjolfsson, E., Hu Y., et M. D. Smith, (2010b), "The Longer Tail: The Changing Shape of Amazon's Sales Distribution Curve", available at SSRN: http://ssrn.com/abstract=1679991.

Carlson, J.A. et Pescatrice, D.R., (1980), "Persistent Price Distributions", Journal of Economics and Business, 33(1): 21-27.

Clay, K., Krishnan, R., et Wolff, E., "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," The Journal of Industrial Economics, 49(4): 521-539.

Curien, N., Fauchart E., Laffond G. et Moreau, F., (2007), Online Consumer Communities: Escaping the Tragedy of the Digital Commons, in Brousseau E. et N. Curien (eds), Internet and Digital Economics, Cambridge University Press.

Ellison, G., et Ellison, S. (2005), "Lessons about Markets from the Internet", Journal of Economic Perspectives, 19(2): 139-158.

Ghose, A. (2006), "Goodbye Price Dispersion? New Evidence from Transaction Prices in Electronic Markets", mimeo.

Ghose, A., et Gu, B. (2006), "Search Costs, Demand Structure and Long Tail in Electronic Markets: Theory and Evidence", Working Papers 06-19, NET Institute.

Ghose, A., et Sundararajan, A. (2006), "Evaluating Pricing Strategy Using eCommerce Data: Evidence and Estimation Challenges", Statistical Science, 21(2).

Ghose, A., Ipeirotis, P. G., et Sundararajan, A. (2009), "The Dimensions of Reputation in Electronic Markets", NYU Center for Digital Economy Research Working Paper No. CeDER-06-02. Available at SSRN: http://ssrn.com/abstract=885568.

Houser, D., et Wooders, J., (2006), "Reputation in Auctions: Theory, and Evidence from eBay", Journal of Economics and Management Strategy, 15: 353-369.

Larribeau, S. et Pénard, T., (2002), "Le commerce électronique en France: un essai de mesure sur le marché des CD", Économie et Statistiques, 355-356: 27-46.

Pan, X., Ratchford, B.T., et Shankar, V., (2004), "Price Dispersion on the Internet: A Review and Directions for Future Research", Journal of Interactive Marketing, 18(4): 116-135.

Resnick, P. et Zeckhauser R., (2002), “Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System”, in Baye M.R. (ed.), The Economics of the Internet and E-Commerce, Advances in Applied Microeconomics, Amsterdam, Elsevier Science, Vol. 11: 127-157.

Spulber, D., (1995), “Bertrand Competition when Rivals’ Costs are Unknown”, Journal of Industrial Economics, 43: 1-12.

Varian, H., (1980), “A Model of Sales”, American Economic Review, 70(4): 651-659.

CHAPITRE 6: ONLINE PRICE DISPERSION, AN INTERNATIONAL COMPARISON

Bounie D.[59], Eang B.[60], Sirbu M.[61] and Waelbroeck P.[62]

6.1 Introduction

Most studies on online markets find significant and persistent price dispersion regardless of the product categories (Baye, Morgan and Scholten, 2004; Pan et al., 2004). By contrast, in theory, the price of an homogenous product offered by several sellers should quickly converge to the best offer. The fact that menu costs are much smaller online than the cost of printing new catalogs strengthens the theoretical argument for price convergence. Further, the availability of price comparison sites considerably lowers search costs online.

Several theoretical explanations have been proposed to solve this apparent puzzle. First, Varian (1980) has argued that search costs related to geographical barriers lead firms to post different prices. Secondly, Baye and Morgan (2001) explain how price dispersion can emerge due to search costs when where firms compete on price with homogeneous goods. These models posit that consumers purchase at the lowest posted price and that online price dispersion decreases with the number of sellers. Thirdly, Pan et al. (2004) have suggested that non-price competition such as reputation and obfuscation strategies can lead to price dispersion.

The empirical literature on online price dispersion has looked at these factors. Claye et al., (2002), Brynjolfsson and Smith (2000) and Brynjolfsson et al., (2010) have shown that search costs can explain price dispersion even when internet users search through comparison sites. Researchers have

59 Department of Economics and Social Sciences, Telecom ParisTech.
60 Department of Economics and Social Sciences, Telecom ParisTech.
61 Department of Engineering and Public Policy and Tepper School of Business, Carnegie Mellon University.
62 Department of Economics and Social Sciences, Telecom ParisTech.

found in online auctions such as eBay (Resnick and Zeckhauser, 2002; Bajari and Hortacsu, 2004; Houser and Wooders, 2006) that reputation increases sales probability and the price at which an item is sold. The reputation effect seems more difficult to prove in other contexts; Ghose et al. (2009) looked at the impact of vendor reputation on markets for both new and used goods on Amazon Marketplace (U.S.) and found an inconsistent effect of reputation on prices.

In addition to reputation and search costs, we believe that two additional dimensions are important to analyze: the time dimension and the maturity of the market. First, price dispersion could persist over time due to seller strategy for instance if they do not change or, conversely, if they substantially change their prices. Moreover, price dispersion could exist due to the entry of new sellers at different prices on the market. Secondly, in less mature markets, large sellers could have more market power and discourage entry which in turn reduces price dispersion. The opposite could be true in mature online markets.

To study the impact of these factors on price dispersion, we have collected data on all *new* books, CDs and DVDs entering the daily TOP 100 list in the US, the UK and France Amazon Marketplace platforms from March 2006 to November 2006. Unlike many shopbots, the Amazon Marketplace by default lists vendors from lowest total cost (price plus shipping) to high. This dataset has several interesting features. First, the search costs are low on the platform. Secondly, the time dimension of the data set can be used to study pricing strategies over time (how often vendors change their prices and the characteristics of entry of new sellers). Thirdly, using the cross variation in the characteristics of the sellers, we construct three measures of reputation that we use to estimate a reputation premium across markets. Fourth, we propose a method for inferring at what price a transaction takes place (search-based theories posit that transactions take place at the lowest prices) and find that transactions are not always at the lowest price. Finally, by comparing three platforms in different countries, we can analyze how price dispersion is sensitive to the maturity of the market.

This study enriches the growing empirical literature on the nature of competition in online markets that it is interested in topics such as pricing theories (Acquisti and Varian, 2005; Ghose and

Sundararajan, 2006), market structure and functioning (Ellison and Ellison, 2005), cultural variety and business models (Ghose and Gu, 2006), online communities and infomediaries (Caillaud and Jullien, 2003), etc. First, using posted prices, we show that vendors make few modifications to their prices over time (once every 10 to 20 days), and those variations that they do make are relatively small (on average 15 cents). This challenges explanations of price dispersion based on some form of mixed or random pricing strategy by vendors (Baye and Morgan, 2001). Secondly, we compare posted prices to transaction prices and we show that transaction prices do not necessarily correspond to the lowest posted price but rather are centered around the median price. This result confirms the findings of Brynjolfsson et al. (2010) and challenges findings by Ghose and Yao (2007). Thirdly, we show that online price dispersion increases with the number of sellers mainly because new sellers price their product significantly lower that the current average price. This intuitive result challenges studies by Spulber (1995), Claye et al., (2002) and Baye, Morgan and Scholten (2004) that find that price dispersion decreases with the number of sellers.[63] Fourth, we show that there is a significant reputation premium of about 10 to 15%. A well established seller can charge a price 10 to 20% higher than the average seller. Finally, our estimation results support the idea that price dispersion is sensitive to the maturity of the market. Consistent with our finding that price dispersion increases with the number of sellers, less mature markets, such as France, with fewer sellers exhibit lower price dispersion. We also find that large sellers account for a greater proportion of transactions in France than in the U.S.

The article is organized as follows. In Section 2 we describe our dataset. Section 3 examines posted price dispersion while Section 4 analyzes transaction price dispersion. Section 5 presents our conclusions.

6.2 Data Set and Definitions

Our data set consists of posted offers compiled from Amazon Marketplaces (Amazon.com, Amazon.uk and Amazon.fr) for CDs, DVDs and book markets. Amazon Marketplace is Amazon's

[63] The last study uses a controversial measure of price dispersion based on the difference between the two lowest posted prices that statistically decreases with the number of sellers and posted prices.

online marketplace that allows sellers to offer their new items alongside Amazon's offerings. Amazon merchants and marketplace sellers are independent[64].

The data are gathered daily using automated Perl scripts. We study each item (a CD, a DVD and a book) entering in the TOP 100 on the Amazon websites and analyze the posted offers for each item on the respective Amazon Marketplaces. The data were collected over 240 days between March 2006 and November 2006. We follow each item entering the TOP 100 any time in the first 22 weeks, from its first appearance until the end of the study interval, even if the item later passes out of the TOP 100. As a result of this procedure, the number of different items we follow in each category depends only on the rate of turnover in the top 100, and is not in any way representative of the market size. Each item is described by:

- Characteristics of the item *i* (CD, DVD and Book): ID (ISBN, etc.), title, type of the product (ex. hardcover, etc.), etc.
- Characteristics of a posted offer *k* of an item *i* at time *t* (price, sales rank, Amazon price, state of the good (new or used), etc.) on the Amazon marketplace.
- Characteristics of a seller *j* who posted an offer *k* of an item *i* at time *t* (rating score, price, etc.) on the marketplace platform.

The main goal of our study is to analyze price dispersion for new goods on the Amazon marketplaces; the study of used goods in the Amazon marketplace is outside the scope of this paper.

There are several data issues to deal with when we analyze posted price dispersion. First, some sellers post prices above $3 million in the US DVD market. This is obviously a bug. As a general rule, we dropped items for which the price is greater than twice the median price of all items sold on the market in our dataset. We provide in Table 1 an overview of the number of offers available in the different markets. Secondly, some items that are selling way above the Amazon price on the marketplace platform are in fact unique items (signed by the author) or collectible items (such as first edition books), even for new products. We cannot construct an automated rule to identify these items.

[64] Some Amazon.com subsidiaries may operate on Marketplace.

Therefore, some prices above the average prices will be driven by product heterogeneity rather than by price dispersion. Thirdly, some items are offered at one cent on the book and DVD markets. Including these items would artificially increase price dispersion if these offers are bogus, especially if the measure of price dispersion is based on the minimum price. We decided to drop these offers as well (see Table 1) that represent a very small portion of all offers. Fourth, about 10% of offers in some categories are the same product by the same seller on the same day and at the same or different prices. We treat these as though they were a single offer (at the lowest of the prices) with multiple copies in stock.

Table 1 shows that the local platforms (US, UK and FR) have different characteristics. The US online book, CD and DVD markets are the largest in terms of the number of sellers, followed respectively by the UK and France. The turnover in the Top 100, and thus the number of distinct items followed, is considerably larger in France than in the U.S. and the U.K. This is due to the fact that the online market share of Amazon is smaller in France than in the other countries; furthermore, Amazon Marketplace itself is less developed in France, as there are generally fewer sellers per item than in the U.K. or the U.S. It appears that Amazon is used more frequently for items other than bestsellers, resulting in high volatility of the TOP 100 list. We defined "large sellers" arbitrarily as the top 10 in each category by range of products offered. These large sellers sometimes show higher ratings and, except for French DVDs, always more ratings than other sellers. Large sellers offer, on average, lower prices than other sellers for books in the UK, CDs in the U.K. and France and DVDs in the U.S.

		Book			CD			DVD	
Markets	**US**	**UK**	**FR**	**US**	**UK**	**FR**	**US**	**UK**	**FR**
Number of observations	4,925,495	3,309,467	3,486,379	3,170,116	3,653,848	4,591,715	4,130,331	2,239,062	1,259,779
Number of one cent offers	148	2,362	10	0	1	1	2,253	19	0
Number of offers greater than twice the median price	230,461	246,092	85,416	99,767	62,53	113,049	39,456	167,184	31,673
Number of items	819	972	2,505	563	758	3,029	685	790	1,383
Number of sellers	19,057	6,189	1,17	4,925	3,548	691	8,208	5,392	1,128
Daily average number of sellers per item[65]	**33.1**	**19.4**	**7.9**	**35.3**	**28.3**	**9.8**	**35.6**	**17.2**	**5.5**
Daily average number of large sellers	7.7	7.7	5.1	8.1	7.4	5.0	8.0	6.5	2.2
Daily average number of other sellers	25.5	11.7	2.9	27.2	20.8	4.8	27.5	10.7	3.2
Average rating of sellers	**4.5**	**4.6**	**4.5**	**4.6**	**4.7**	**4.5**	**4.5**	**4.6**	**4.5**
Average rating of large sellers	4.6	4.7	4.5	4.6	4.7	4.5	4.7	4.7	4.5
Average rating of other sellers	4.5	4.6	4.5	4.6	4.7	4.5	4.5	4.6	4.5
Average number of ratings	**27,348**	**38,392**	**8,843**	**25,891**	**31,156**	**34,779**	**24,078**	**38,776**	**7,701**
Average number of ratings for large sellers	69,413	82,77	10,773	74,592	75,755	63,477	74,717	43,382	4,297
Average number of ratings for other sellers	15,635	12,114	2,376	12,543	16,424	4,901	10,244	37,234	8,588
Average product price (in local currency)	**16.66**	**9.49**	**11.72**	**14.95**	**11.08**	**14.71**	**27.37**	**20.70**	**18.71**
Average product price for large sellers	16.68	8.60	11.79	15.14	10.97	14.57	26.98	20.62	18.76
Average product price for other sellers	16.64	10.18	11.66	14.98	11.55	14.67	27.61	20.53	17.94

Table 1 - Dataset Overview

There was surprisingly little overlap between countries in the items appearing in the TOP 100 and thus entering our dataset, even between the U.S. and the U.K. where there is not a language barrier (Table 2).

[65] To get daily averages, we first compute a daily average across sellers for each day and each distinct product. Because we have an unbalanced panel over time, we compute a time average for each product and then an average over products, weighting each product equally. We do the same separately for large and small subsets of sellers. Because large and small sellers may not have offers for equal numbers of days, the separate results for number of vendors may not total to the average number of all vendors. Similar problems affect the computation of average prices.

	US/UK	US/FR	UK/FR	UK/US/FR
Book	30	23	15	3
	3.7%	2.8%	1.5%	0.1%
CD	71	108	190	47
	12.6%	19.2%	25.1%	1.1%
DVD	146	74	158	52
	10.6%	10.8%	20.0%	1.8%

Table 2 – Overlap between TOP 100 Goods across Countries

6.3 Posted Price Dispersion – Time Dimension

To study price dispersion over time, we construct a balanced panel in order to avoid overweighting those items that enter the data collection earlier vs. later. For our balanced panel we keep all those items that were in our dataset beginning Monday of the second week of observation and follow only these items until the end of the sample (some items have entered or dropped out of the top 100 during the first week of the period; for instance, in France, 201 different books had appeared in the top 100 by the time we fixed our sample, of which 199 remained). We have 3 dimensions to any given item in our dataset. The first dimension is the product dimension: there are different books with different ISBN, CDs with different ASIN, etc. The second dimension is the time dimension since we follow each product over time. The third dimension is related to the number of sellers of a particular item on the Amazon marketplace. The meaning of price dispersion varies according to the dimension that we focus on. In the next section we consider how price dispersion across sellers varies over time.

For each item category (book, CD, DVD) and market (US, UK, FR), we construct several useful statistics. For the average price in a market, we first compute the daily (time) average price for each pair (item, seller) and then average over sellers and then over items (in all cases prices are in local currency: dollars, pounds and euros respectively). We use a similar procedure for the standard deviation and average minimum of prices and for average and standard deviation of number of sellers. We compute the percentiles of prices using the week as the time unit. Table 3 summarizes the descriptive statistics for prices posted by sellers on the marketplace and the last 4 lines represent the total number of offers, sellers and items over the whole period and the (daily) average number of sellers per item.

We can make several interesting international comparisons. The U.S. Marketplace generally has the largest number of sellers followed by the UK and then France. We also observe that most sellers in Amazon Marketplace offer only a single item for sale. The book markets in the U.S. and the U.K. have the most sellers, whereas in France it is the DVD market; for all goods, France has many fewer sellers than in the U.S. or the U.K. The median number of days that a seller offers an item online is relatively short for all goods in the US—two to three weeks—but is more than four months for books in France. Finally, we note that while the majority of offers in the U.S. and the U.K. are from "full line" sellers that offer virtually all of the items covered in the dataset for sale, very few French sellers offer a full catalog. (An examination of the full distribution shows that with the exception of Amazon itself, a typical "large" seller in France offers no more than about half the catalog.)

Looking at prices (all prices are in local currency), we note that CDs and DVDs are far more expensive in the UK, and the U.S. has the lowest prices for these goods. Book prices are comparable across the three countries.

France has a law limiting discounting of books to no more than 5% off of the publisher's list price. As a result, we find that Amazon does not offer a significant discount compared to other sellers, and the average minimum price offer tends to be closer to the average price offer for books in France than for other goods or books in other countries. We also note that the smallest coefficient of variation is in the French DVD and book markets (both .184)—with the two lowest average numbers of sellers, and the largest is in the French CD market (.280).

		Book			CD			DVD	
Markets	**US**	**UK**	**FR**	**US**	**UK**	**FR**	**US**	**UK**	**FR**
Number of observations	1,176,823	615,97	445,37	959,27	932,44	440,25	1,003,551	378,65	128,47
Number of one cent offers	0	53	10	0	0	0	164	3	0
Number of offers greater than twice the median price	52,57	47,876	29,615	24,252	11,699	9,163	11,918	24,472	2,066
Number of items	98	107	199	115	122	177	98	99	97
Total unique sellers	9,961	2,666	323	2,725	1,764	310	4,176	1815	347
Average number of sellers	51.1	24.2	9.5	40.6	34.2	11.2	44.5	17.9	6.3
Standard deviation of number of sellers	29.7	6.1	3.9	8.2	10.9	5.8	11.6	5.6	3.9
Percent of sellers offering only one item	73	77	65	59	62	43	60	69	46
Median total days seller has item on offer	9	24	132	23	44	77	12	15	40
% of offers accounted for by sellers offering>75% of catalog	47	69	16	70	52	26	70	57	10
Average Amazon Price	13.58	7.01	10.79	12.72	9.02	15.38	24.66	16.85	19.34
Average price	15.09	8.03	10.93	15.14	11.07	16.39	26.58	21.34	20.06
Standard deviation of prices	3.60	2.20	2.01	3.43	3.23	4.58	5.06	5.43	3.63
Average minimum price	9.51	5.03	8.94	9.60	6.22	11.15	18.60	13.96	16.67
Median price	14.73	7.76	10.45	14.84	10.87	15.72	26.25	20.77	19.70
75th percentile	16.91	9.28	12.11	16.31	13.24	19.58	28.53	24.79	22.41

Table 3 – Descriptive Statistics of Balanced Sample

To summarize, we observe that first the average number of sellers per item is the largest in the US, then in the UK and then in France. Secondly, the number of book offers in France is greater than the number of offers for CDs and DVDs, while the US secondary markets for the three product categories are fairly similar in terms of offers (the CD market is smaller in terms of the number of sellers). Thirdly, the Amazon marketplace for CDs is very active in the UK. Finally, the average Amazon price seems fairly competitive, as it is lower than the average price posted by the other sellers on the marketplace platform.

6.3.1 Distribution of Price Changes on Amazon Marketplace

6.3.1.1 Individual Seller Price Variability over Time

The first source of price dispersion is the individual price variability arising from the fact that a seller might increase or decrease its price over time. In Table 4 we compute the number of times

sellers increase/decrease the price of its items or leave them unchanged from one day to another.[66] We report the results of our computation separately for Amazon and for the other sellers.

		Book			CD			DVD		
		US	UK	FR	US	UK	FR	US	UK	FR
Amazon	Flat	96.62	96.93	99.25	92.43	90.51	90.06	91.57	87.02	92.54
	Decrease	1.37	1.47	0.34	3.62	4.61	4.56	4.18	6.46	3.52
	Increase	2.01	1.60	0.41	3.95	4.88	5.38	4.25	6.52	3.94
Sellers	Flat	93.51	91.49	98.81	91.63	87.83	86.35	89.92	87.81	95.00
	Decrease	3.88	5.05	0.71	5.00	6.96	8.13	6.12	7.29	3.67
	Increase	2.61	3.46	0.48	3.37	5.21	5.52	3.96	4.90	1.33

Table 4 – Distribution of Amazon Price Changes (in %)

The distribution of price change is fairly symmetric for Amazon. Price remains the same 90% of the time. If we assume that the time between price changes is exponentially distributed, these results suggest a mean time between price changes from a low of every 8 days for UK DVDs, to more than 100 days for books in France. Across product markets, prices are far more stable for books than for CDs or DVDs. The distribution of price changes is more asymmetric, with more price decreases than price increases (thus prices decrease over time on the platform). Given the relatively infrequent price changes, particularly for books, individual pricing strategies cannot be the main source of persistent price dispersion. Instead, persistent price dispersion is found even when sellers do not adjust their price very often.

To test random pricing by the sellers, we computed for each product the distribution of prices across all sellers and all days and broke the distribution into four quartiles. We then ask how often the seller of a given item changes prices so as to move from one quartile to another over the 269 days of observations. We found that, the vast majority of the time, sellers do not change their prices at all or sufficiently to alter their quartile rank. Offering prices in both the highest and lowest quartiles over the observation period generally happened less than 6% of the time (Table 5). We conclude from these data that sellers do not practice random pricing.

[66] In some cases, we have some missing observations and we use the last valid observation in our computations.

	Book			CD			DVD		
	US	UK	FR	US	UK	FR	US	UK	FR
Percent of sellers x items where seller's price remains in same quartile for entire observation period	85.9	74.9	70.7	79.4	68.6	58.4	75.5	71.1	66.8
% where seller's price appears for at least one day in the highest and in lowest quartile	1.5	3.3	2.4	2.2	3.6	5.7	3.6	6.0	4.8

Table 5 – Frequency at which Sellers Price Ranking Changes Significantly

To further check whether sellers propose large price variations, we compute the standard deviation of their prices over time (weeks) for each item. We then average over the items and plot the histogram of standard deviations over the different items in each market. The results are plotted in Figure 1 for the US book market and in the Appendix for the other markets. We find that the average of the standard deviation of price changes ((mean) sd_Prix, in Figure 1) is about 15 cents in the US book marketplace, which strongly challenges the assumption of random pricing by seller.

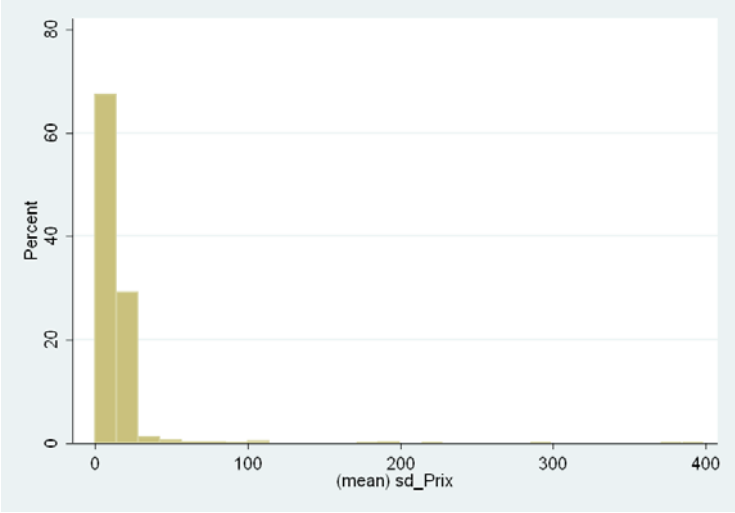


Figure 1 – Distribution of Price Changes (standard deviation, in cents)

6.3.1.2 Price Dispersion over Time

We also have found that price dispersion increases over time. That is, the longer an item is available for sale in our balanced sample, the greater the daily price dispersion. In searching for an explanation, we noted that the number of sellers for an item also tends to increase over time.

We investigated how the number of sellers for an item changes over time using a fixed effects panel regression of the number of sellers per item per day versus elapsed days since an item first appears in the Top 100 and thus enters our dataset. In general we find a positive relationship. For example, after 200 days on the market, on average the number of sellers for a book in the U.S. will have increased by roughly 14 sellers whereas the overall mean is 50 sellers (Table 6).

	Book			**CD**			**DVD**		
Markets	**US**	**UK**	**FR**	**US**	**UK**	**FR**	**US**	**UK**	**FR**
Dependent variable	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers
Elapsed days	0.06972***	0.01052***	0.00462***	0.04035***	0.03331***	0.00114***	0.05770***	0.01350***	-0.00271***
	[0.00069]	[0.00027]	[0.00011]	[0.00056]	[0.00039]	[0.00015]	[0.00065]	[0.00035]	[0.00021]
Constant	42.55418***	22.83876***	9.05909***	36.45592***	30.98434***	11.11373***	36.85958***	17.04289***	7.42871***
	[0.10505]	[0.04027]	[0.01622]	[0.07631]	[0.05478]	[0.02038]	[0.09913]	[0.05090]	[0.03061]
R-squared	0.2998	0.05272	0.03438	0.17769	0.20992	-0.00296	0.25261	0.06151	0.00381
N	23,445	26,352	47,468	23,345	26,712	39,202	23,086	20,959	18,869

* p<0.05, ** p<0.01, *** p<0.001

Table 6 – Number of Sellers vs. Number of Days in the Market (Fixed Effects)

To determine the relative importance of time versus number of sellers on price dispersion, we regressed the coefficient of variation (expressed as a percentage) on both time and number of sellers. In Table 7 we again use a fixed effects panel regression to show that price dispersion is positively influenced by number of sellers and by elapsed time, with number of sellers having, in general, the larger effect. For example, if the number of sellers is larger by one standard deviation above the average (e.g. 29 sellers for books in the U.S. or 4 sellers for DVDs in France – *Cf.* Table 3) we find the coefficient of variation increases by 9 and 4.8 percent respectively. Previous empirical studies that have fixed the number of sellers investigated have missed this phenomenon.

	Book			CD			DVD		
Markets	**US**	**UK**	**FR**	**US**	**UK**	**FR**	**US**	**UK**	**FR**
Dependent variable	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)
Elapsed days	-0.00270***	0.01977***	0.00521***	-0.00090**	0.02033***	0.00807***	-0.00392***	0.02003***	0.00684***
	[0.00040]	[0.00037]	[0.00027]	[0.00030]	[0.00033]	[0.00042]	[0.00031]	[0.00069]	[0.00073]
Number of Sellers	0.31173***	0.54646***	0.42187***	0.26538***	0.20389***	0.48292***	0.19560***	0.37063***	1.19991***
	[0.00323]	[0.00853]	[0.01166]	[0.00381]	[0.00494]	[0.01527]	[0.00331]	[0.01406]	[0.02541]
Constant	8.58682***	13.42176***	14.97511***	12.19362***	20.07548***	22.20086***	12.56333***	19.26073***	11.48760***
	[0.14756]	[0.20261]	[0.11394]	[0.14865]	[0.16035]	[0.18352]	[0.13474]	[0.26852]	[0.22584]
R-squared	0.34448	0.25348	0.03884	0.19345	0.27219	0.03041	0.14327	0.08345	0.10735
N	23,362	26,324	46,69	2,3105	26,619	38,527	22,729	20,508	17,866

* p<0.05, ** p<0.01, *** p<0.001

Table 7 – Coefficient of Variation (%) vs. Elapsed Days and Number of Sellers (Fixed Effects)

We hypothesize that this increase in price dispersion with number of sellers (Table 7) comes about because new sellers enter the market with offers at substantially lower prices than the median. We analyzed the price offer of sellers entering the market after our first day of observation, and compare their offer price to the median of all sellers for that item on the previous day. We find that, with the exception of the French book market, new sellers are more likely to enter the market at prices below the median, and, again with the exception of France, typically with a double digit percent discount from the prior day's median price (Table 8).

	Book			CD			DVD		
Market	US	UK	FR	US	UK	FR	US	UK	FR
Probability new entrant's price is below prior day's median price	0.79	0.71	0.47	0.73	0.67	0.57	0.68	0.71	0.72
mean percent difference from median for new sellers	-22	-16	-3	-18	-14	0.4	-14	-11	-8

Table 8 -- New Sellers Probability of Discount from Median and Mean Percent Discount

6.4 Price, Seller and Item Characteristics

We begin with the definition of a transaction on Amazon Marketplace. Next we compare transaction prices with posted prices discussed in the previous section. Finally we assess the effect of a seller's reputation and item characteristics on the prices of items that he or she sells using a simple hedonic regression.

6.4.1 Definition of a Transaction

We compute transactions on the Amazon marketplace as follows. We observe posted offers by sellers for a particular item at a given time. Among the posted offers, some will lead to transactions. A posted offer by a seller *j* at date *t* will be considered as a transaction in two cases: if the posted offer is no longer available after *t*, meaning that the seller does not propose it at any time after *t*; or if the posted offer disappeared after *t* and reappeared some day after *t+2*.[67]

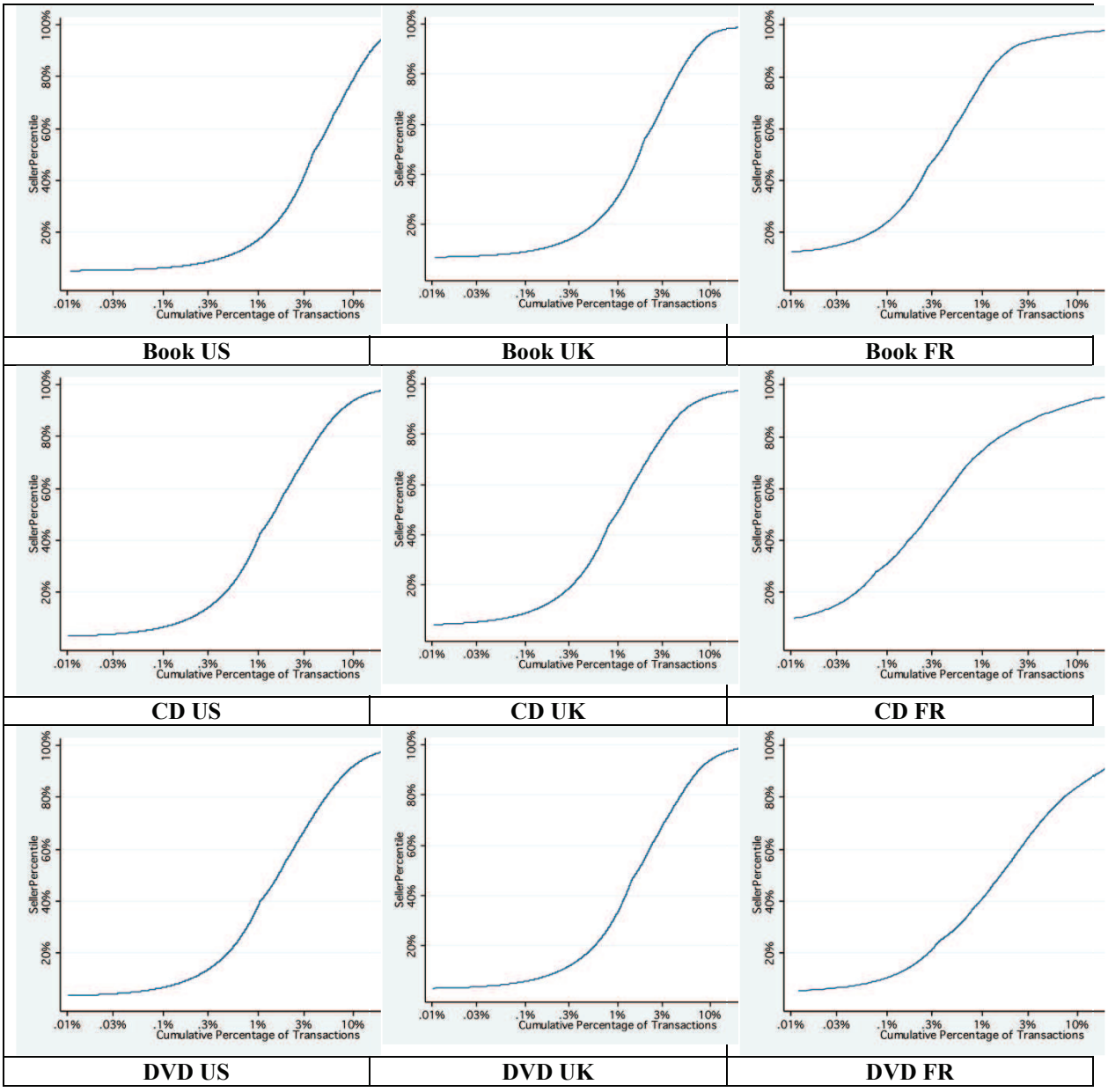
An illustration of the distribution of transactions is given in Figure 2 for every market. The figures confirm that a small percentage of sellers account for the vast majority of the transactions in all markets. Given the large number of sellers on Amazon marketplace, even 1% of sellers can mean anywhere from 6 to 190 firms, so no one firm may have significant market power.

This method has several shortcomings since it can overestimate or underestimate sales. First, the method can overestimate sales since a seller can drop his offer from the catalogue without necessarily selling it (due to pricing concerns for instance) or, conversely, the seller can cancel his offer on the Amazon marketplace because he sold the item by way of an alternative channel (physical, other website, etc.). However, in our analysis, the disappearance of this offer will be considered as a sale from the Amazon marketplace. Second, the method can underestimate sales: as long as an offer posted by a seller is available on the platform, we consider it as a posted price and not as a transaction. Now, a seller for instance can sell the item at date *t* and propose another offer at date *t+1*. Similarly, a seller may have multiple offers (at the same or different prices) for the same item posted on Amazon

[67] To control for possible bugs during the updating of the script, we eliminated from the database dates for which more than 7% of the items where not updated due to technical difficulties.

Marketplace. Yet, in the last two cases, as long as at least one posted offer is available, we will not record it as a transaction, thus underestimating transactions. Nevertheless, we have shown that large sellers account for most transactions so that the extent of such an underestimation is limited.

Figure 2 Distribution of Transactions over Sellers (log scale)



6.4.2 Transaction and Posted Prices: A Comparison

Table 9 (constructed in the same way as Table 3) reports descriptive statistics of transaction prices as defined above. We can compute the standard deviation of transaction prices for each item over the period and compare the results with the standard deviation of posted prices. Table 8 shows that transaction prices, as we measure them, are on average higher than Amazon's price for the same good; this challenges studies that assume that transactions are concentrated around the lowest posted price offers such as Baye *et al.* (2004). Note the small spread between average minimum and average maximum price for French books, due to the French law which limits book discounts.

	Book			CD			DVD		
Market	US	UK	FR	US	UK	FR	US	UK	FR
Average price	15.13	9.47	11.81	13.86	11.35	14.86	25.30	20.08	17.51
Average Amazon price	14.71	8.21	11.49	12.70	9.10	13.57	24.43	16.52	16.98
Average Median price	14.99	9.45	11.78	13.75	11.31	14.74	25.10	19.98	17.37
Average Minimum price	13.82	8.87	11.64	12.48	10.07	13.57	23.16	18.27	16.15
Average Maximum price	16.90	10.20	12.06	15.94	12.80	16.48	28.38	22.19	19.19
Number of transactions	150,202	94,06	89,447	129,111	124,998	130,534	177,055	108,508	41,954
Number of sellers	11,803	3,596	626	3,292	2,386	452	4,885	3,43	721
Number of items	806	923	2,356	561	706	2,854	667	760	1,308

Table 9 – Transaction Price Dispersion

In the next part, we analyze the determinants of transaction prices by looking at the reputation premium of a seller and at item characteristics.

6.5 Determinants of Transaction Prices: Seller Characteristics and Time

We now explain transaction prices (p_{ij}) of item *i* by seller *j* observed on the marketplace using the reputation of the sellers and characteristics of the items. There are several variables that capture the reputation of a seller. As in Ghose *et al.* (2009), we use both the average rating of a seller and the number of ratings that he or she received in the past. These serve as proxies for both reputation and size of operations (large sellers can better internalize fixed costs and offer better discounts). We

consider an additional reputational variable that we hypothesize is related to good will: the size of the seller. We considered two measures of size: the number of different items offered by the seller in a particular market, and the total transactions by the seller. We find these two measures to be highly correlated (typically > 0.9) and in the following use only the catalog size. Finally, we control for the time since release date ("time on market") of an item to detect any trend in prices over time.

We note that Amazon Marketplace does not provide ratings for Amazon itself, and ratings are missing for many other sellers as well—for Amazon.fr we are missing ratings for as many as 40% of the transactions. The shorter lifetime of the Amazon.fr marketplace compared to Amazon US probably accounts for fewer vendors having been rated. The distribution of missing values across vendor size appears to mirror the distribution for all transactions, so we believe that these missing values do not bias our regressions.

To estimate the effects of these variables, we run regressions with fixed item effects using the full unbalanced dataset but limited to 22 weeks because of gaps in the data collection after that date which could have introduced spurious transactions. Results are summarized in Table 10.

		Book			CD			DVD	
Market	US	UK	FR	US	UK	FR	US	UK	FR
Dependant variable	Price	Price	Price	Price	Price	Price	Price	Price	Price
Rating	28.4***	-5.41	4.21	-2.48	5.3	0.337	8.467	23.8	-3.59
Number of ratings	0.139*	-0.084	-0.156	0.421***	-0.884**	-6.35***	0.166	-3.19***	-16.0**
Catalog	0.516***	0.156***	-0.011***	0.838***	0.501***	0.0551***	1.15***	0.484***	0.0550***
Time on Market (TOM)	-0.215***	-0.170***	0.130***	0.0922**	-0.353***	-0.0188	-0.498***	0.129	1.66***
TOM²	0.00007***	-0.00002	-0.00001	0.00004**	-0.00002	0.00005**	-0.00009	0.00011	-0.00030***
Fixed Effects by item	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1010	883***	956***	1070***	1090	1170***	2220	1430***	794***
R^2 (within)	0.101	0.0434	0.0023	0.178	0.102	0.0111	0.162	0.0381	0.00861
N	149350	71191	48954	128189	105513	82762	176926	101067	29287
Number of items	792	753	1295	536	643	1726	663	707	834
Reputation Premium (in local currency) for one SD change in catalog	1.70	0.56	-0.08	1.70	1.17	0.47	0.41	1.36	0.23

Table 10 – Estimation Results

First, we find that traditional measures of reputation such as rating are not significant, except in the U.S. book market. Likewise, number of ratings is ambiguous sometimes increasing and sometimes decreasing the price premium when it is significant at all. Secondly, catalog and number of transactions are strongly correlated and better capture seller reputation: an increase of one standard deviation in catalog size is associated with price increases from .41 to 1.70 in local currency. Thirdly, the French markets differ from the UK and the US in several respects. As noted earlier, French law limits discounting for books; thus French booksellers with a large catalog are not undercut by smaller sellers as in the other markets. In addition we note that the R^2 values are smaller in all the French markets, suggesting that none of our variables have explanatory power in France. By contrast, in the U.S. CD market we explain 18% of the variance in price (ratings alone, without catalog size, explain no more than 1%). Fourthly, time since release date (ToM, measured in days) appears to be associated with declining prices for books (except in France, which is a special case because of the law), while having an ambiguous effect in the CD and DVD markets. We note that in several cases, the transaction date precedes the official release date, producing a negative value for ToM. In other cases, 40 year old classics appear in the top 100, producing ToM values over 15000. Obviously the variance of ToM across transactions will be greatest for new releases.

6.6 Conclusions

In this paper we examined comparable online marketplaces for new cultural goods in three countries over an eight-month period in 2006. We look at the structure of the online marketplace across countries and goods, and examine dispersion among posted and transaction prices.

Our first observation is that online marketplaces differ significantly across countries. The U.S., U.K. and France show widely differing roles for large online merchants, who carry an extensive catalog: there are many such merchants in the U.S., somewhat fewer in the U.K., and very few in France. In all countries, however, there is significant participation by small merchants who sell but a single good from within our sample set. These sellers appear and disappear rapidly; while the daily

number of sellers per good is modest (averaging 5-50 depending upon the market), more than 19,000 unique vendors appeared in the U.S. book market over our 8 month observation period.

When we examine how firms price, we uncover several findings. First, most sellers make infrequent changes to their posted prices. Moreover, changes that they do make are typically small. Thus over an eight month period changes in a vendor's posted price will result in a change of quartile ranking less than 1/3 of the time (with the exception of French CDs). In other words, vendors who enter the market with a high price will keep it high, while low price vendors tend to keep their price low over time. Even Amazon, the leader, makes relatively few day-to-day price changes, though it makes as many up as down. These findings contradict the hypothesis of random pricing put forward by other authors.

We also examine price dispersion, using coefficient of variation of posted prices as our measure, and find significant dynamics related to time and number of sellers. In general, the number of sellers for an item increases over time (even as its sales rank may drop). For example, after 200 days, the number of sellers of books in the U.S. will have increased by almost 25%. And price dispersion increases with both time and number of sellers. Thus an increase of one standard deviation in the number of sellers leads to as much as a 9% increase in the coefficient of variation. We show that this increase in dispersion is the result of new sellers entering the market at substantially lower prices than the previous day's median price. Previous studies that have fixed the number of sellers have missed this important effect on price dispersion. Finally, our results provide no support for traditional measures of reputation such as ratings and number of ratings and support new measures such as size of catalog and number of transactions.

To conclude, our results confirm that there are other dimensions than price competition in these markets. Reputation seems to be an important factor that questions two key assumptions of the search models: the symmetry of the sellers and homogenous goods. First, we have shown that most transactions are carried out by large sellers while a large number of small sellers realize at most two transactions. Second, it is hard to disentangle the product/seller couple which leads to product

differentiation among sellers with different degrees of reputation. As a consequence, we do not observe many transaction prices at the bottom of the price distribution, *i.e.* the best offers do not attract all the buyers. Further research needs to better characterize competition in these markets as well as seek other factors that create persistent price dispersion. The promising directions are the dynamics of entry and exit in these marketplaces (time dimension) and reputation. Since this initial attempt to compare posted prices with transaction prices shows significant discrepancies with theoretical predictions, future work should focus more on actual transactions.

6.7 References

Acquisti, A. and Varian H., 2005. Conditioning Prices on Purchase History. *Marketing Science*, 24(3), 1-15.

Bajari, P. and Hortacsu A., 2004. Economic Insights from Internet Auctions. *Journal of Economic Literature*, 42(2), 457-86.

Baye, M. R., and Morgan, J., 2001. Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets. *American Economic Review*, 91(3), 454-474.

Baye, M. R., Morgan, J., and Sholten, P., 2004. Price Dispersion in The Small And in The Large: Evidence From An Internet Price Comparison Site. *Journal of Industrial Economics*, 52(4), 463-496.

Brynjolfsson, E. and Smith, M., 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science*, 46(4), 563-585.

Brynjolfsson, Erik, A A. Dick, and M. D. Smith, (2010), "A Nearly Perfect Market? Differentiation vs. Price in Consumer Choice," *Quantitative Marketing and Economics*, 8(1), 1-33.

Caillaud, B. and Jullien B., 2003. Chicken & Egg: Competition among Intermediation Service Providers. *RAND Journal of Economics*, 34(2), 309-28.

Clay, K., Krishnan, R., and Wolff, E., 2002. Prices and Price Dispersion on the Web: Evidence from the Online Book Industry. *Journal of Industrial Economics*, 49(4), 521-539

Ellison, G., and Ellison, S., 2005. Lessons about Markets from the Internet. *Journal of Economic Perspectives*, 19(2), 139-158

Ghose, A., and O. Y., 2007. Goodbye Price Dispersion? New Evidence from Transaction Prices in Electronic Markets. Third Research Symposium on Statistical Challenges in E-Commerce Research (SCECR), University of Connecticut, May.

Ghose, A., and Gu, B., 2006. Search Costs, Demand Structure and Long Tail in Electronic Markets: Theory and Evidence. Working Papers 06-19, NET Institute.

Ghose, A., and Sundararajan, A., 2006. Evaluating Pricing Strategy Using eCommerce Data: Evidence and Estimation Challenges. *Statistical Science*, 21(2).

Ghose, A., Ipeirotis, P. G., and Sundararajan, A., 2009. The Dimensions of Reputation in Electronic Markets. *Working Paper*.

Houser D. and Wooders J., 2006. Reputation in Auctions: Theory, and Evidence from eBay. *Journal of Economics and Management Strategy*, 15, 353-369.

Pan, X., Ratchford, B. T., and Shankar, V., 2004. Price Dispersion on the Internet: A Review and Directions for Future Research. *Journal of Interactive Marketing*, 18(4), 116-135.

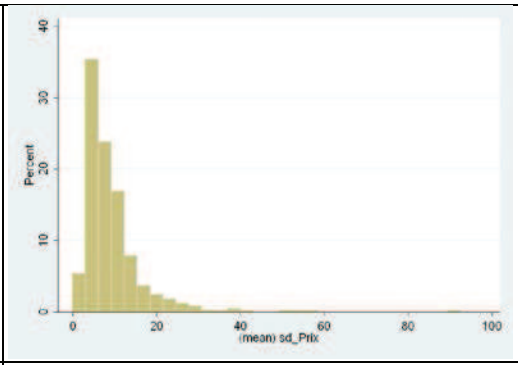
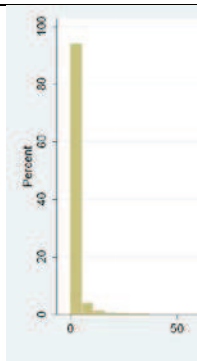
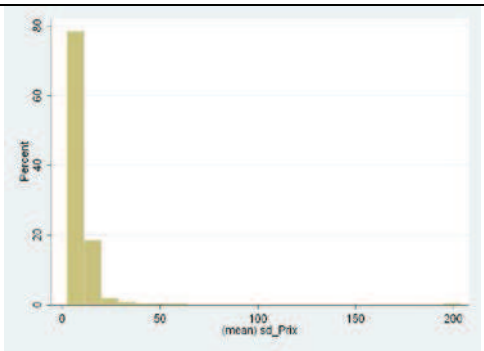
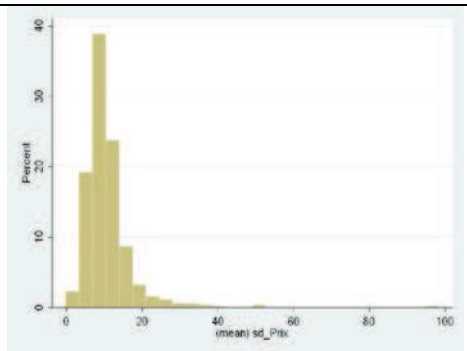
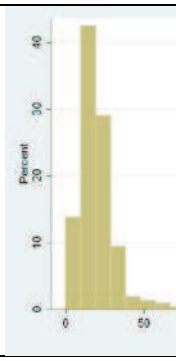
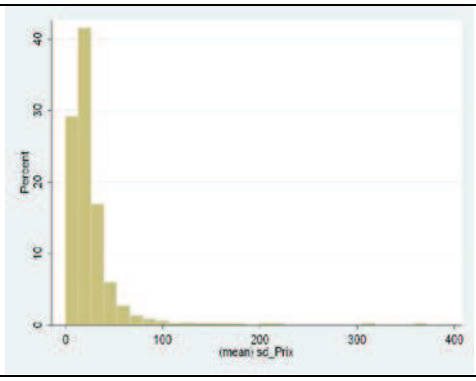
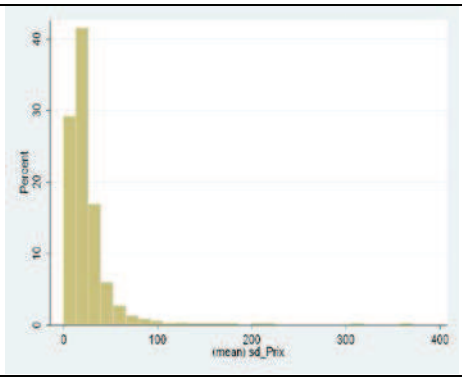
Resnick, P. and Zeckhauser R., 2002. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System, in Baye M.R. (ed.), *The Economics of the Internet and E-Commerce*, Advances in Applied Microeconomics, Amsterdam, Elsevier Science, 11: 127-157.

Spulber D., 1995. Bertrand Competition when Rivals' Costs are Unknown. *Journal of Industrial Economics*, 43, 1-12.

Varian, H., 1980. A Model of Sales. *The American Economic Review*. 70, 651-659.

6.8 Appendix

Histograms of Price Changes in Cents (Standard Deviation Averaged over all Sellers, across Items)

[In the text]		
Book US	Book UK	
		
CD US	CD UK	
		
DVD US	DVD UK	

CHAPITRE 7: CONCLUSION

Nos études empiriques sur le commerce électronique de biens culturels ont mis en avant les opportunités des plateformes de vente en ligne pour les différents acteurs des industries culturelles. Les stratégies adoptées par Amazon par exemple en développant la place de marché Marketplace concourrent à enrichir la variété des produits culturels offerts et achetés puisque les meilleures ventes de produits culturels sont très différentes entre les différents réseaux de distribution. De plus, ces mêmes plateformes permettent aux consommateurs l'accès rapide à des produits de niche plus difficilement accessibles dans les réseaux de distribution physique. Nous montrons à travers nos articles qu'une différenciation majeure entre la distribution en ligne et la distribution physique repose sur la date de sortie : les réseaux physiques semblent très liés à l'actualité des sorties de nouveaux titres alors que les plateformes de ventes sur Internet constituent un canal privilégié pour commercialiser des produits du répertoire plus ancien. Ce phénomène s'amplifiera dans les années futures grâce à l'allongement de la durée de vie des produits culturels sur les plateformes de ventes. Il est important de souligner que la création de plateformes de vente comme Amazon Marketplace (aujourd'hui répliquée par Barnes & Noble et la Fnac) permet à la fois aux grandes enseignes Internet mais aussi aux librairies physiques de résoudre le problème de gestion du stock croissant des produits culturels. Les grandes enseignes Internet peuvent ainsi déléguer une certaine partie de la demande à d'autres professionnels du secteur ou des particuliers tout en jouant le rôle de portail sur la toile. De l'autre côté, les librairies et autre magasins physiques peuvent écouler des produits se vendant moins bien en rayon sans pour autant baisser significativement le prix de ces derniers. Enfin, si les stratégies des vendeurs en ligne se sont d'abord portées sur les biens culturels physiques, il est à noter que l'environnement technologique actuel pousse désormais les consommateurs vers des offres de plus en plus dématérialisées. La dématérialisation des produits culturels remet en question aujourd'hui le rôle de chaque acteur dans la chaîne de la valeur des biens culturels qu'il s'agisse des artistes, des éditeurs ou des distributeurs.

Articulée autour de cinq papiers de recherche, cette thèse contribue à améliorer la compréhension du commerce électronique de biens culturels sur plusieurs dimensions.

Tout d'abord, la comparaison des ventes de livres en ligne (sur Amazon.fr) aux ventes en magasin (hors-ligne) sur la période de mars à août 2006 en France permet de retenir deux leçons sur la demande des biens culturels en ligne. Premièrement, nous montrons que les meilleures ventes en ligne diffèrent de celles réalisées dans les réseaux physiques. Cette différentiation dans la demande indique que l'effet positif d'expansion liée à la longue traîne domine l'effet négatif de cannibalisation des ventes physiques par les ventes en ligne. De plus, l'observation de décalages temporels d'entrée des livres dans les tops des ventes met en évidence le caractère prescripteur des librairies pour les ventes en ligne du genre « Universitaire, essais, science et religion». Ce même effet n'est cependant toujours pas vérifié dans le sens contraire.

Ensuite, la comparaison du catalogue des meilleures ventes de produits culturels neufs sur la plateforme de vente Marketplace à celui de la grande enseigne Internet Amazon souligne deux tendances importantes sur les plateformes de ce type. Premièrement, l'effet de cannibalisation des ventes reste relativement faible puisque la partie commune des livres, CD et DVD dans les meilleures ventes respectives n'est que de l'ordre de 20%. Deuxièmement, les plateformes en ligne constituent un nouveau canal de distribution permettant aux vendeurs de Marketplace d'écouler leur bas de catalogue. Nous observons ainsi qu'au moins 25% des meilleures ventes de Marketplace ont des dates de sorties antérieures à celles d'Amazon.

Nos travaux sur la dynamique et la dispersion des prix des livres, CD et DVD à travers plusieurs pays sur une période de 8 mois en 2006, permettent de tirer plusieurs enseignements. Premièrement, la dispersion des prix sur Internet augmente avec le nombre de vendeurs et dans le temps. Nous observons en outre que l'augmentation de la dispersion des prix résulte de l'entrée de nouveaux vendeurs à des prix légèrement inférieurs que le prix médian du jour précédant l'entrée. Deuxièmement, les explications liées aux stratégies aléatoires de prix des vendeurs ne sont pas convaincantes car la majorité d'entre eux n'effectuent pas de changements de prix au cours du temps

ou le font de façon très minime. Troisièmement, si la réputation d'un vendeur semble jouer un rôle important dans la décision d'un internaute d'effectuer une transaction, nous n'observons pas d'effet significatif de la note du vendeur ou de son nombre de notes sur la tarification d'un produit. Dès lors, les origines de la prime à la réputation d'un vendeur restent toujours à ce jour inexpliquées. Enfin, la comparaison de plusieurs plateformes nationales permet d'analyser de façon globale les résultats obtenus même s'il existe des différences notables entre pays et notamment dans la répartition entre petits et grands vendeurs selon les marchés.

Le cas de la vente de biens culturels numériques abordé sous l'angle de la vente de livres électroniques esquisse des résultats qui devront être confirmés dans les années futures. L'analyse de la vente d'ebooks par Amazon montre que l'effet de cannibalisation des titres papiers par les titres numériques ne concerne qu'une portion des titres phares. Notre étude met à jour l'émergence d'un nouveau segment de livres électroniques nommés digital outsiders qui est à l'origine d'un effet d'expansion de marché. Ces digital outsiders sont scindés en deux catégories. La première, englobe des titres appartenant au domaine public ou possédant une date de publication antérieure à 2000 et dont les équivalents papiers ne sont plus vraiment disponibles ou ne se vendent plus. La seconde catégorie, constituée de pure players, concerne elle des titres d'auteurs contemporains qui ont fait le choix de ne pas être édités sous forme papier ou qui n'ont pas trouvé de maison d'édition. Nous observons dans ce cas un nombre important de livres de type Arlequin dont les auteurs ont décidé de se passer des maisons d'édition pour tester eux-mêmes le succès potentiel de leurs titres sur la toile.

Au final, nos recherches permettent de renouveler un certain nombre de résultats établis dans la littérature économique même si nos travaux ne sont pas également exempts de toutes critiques.

Tout d'abord le travail de comparaison entre les meilleures ventes en ligne et physiques mérite d'être étendu à d'autres pays comme l'Angleterre ou l'Allemagne qui diffèrent de la France sur le plan de l'accès à la culture et des usages d'Internet. De même, cette comparaison permettrait une meilleure compréhension de la demande en ligne si elle était appliquée à d'autres biens culturels comme les CD et DVD. De plus, si nous avons montré qu'il existe des différences entre les produits achetés en ligne

et dans les magasins physiques, il reste à démontrer que l'augmentation des ventes pour ces produits en ligne est supérieure à une éventuelle réduction des ventes sur les marchés physiques.

Le succès et l'augmentation de la durée de vie en ligne des produits du répertoire ancien mérite aussi d'être approfondie. En effet, nous n'avons pas étudié l'impact d'une telle consommation sur les achats de produits neufs récents. Qu'il en soit des livres, CD ou DVD, la demande pour des titres plus anciens est-elle motivée par un désir d'exploration de la part des consommateurs ou par une contrainte budgétaire de ces derniers considérant comme trop chers les titres récents sortis sur les nouveaux supports (Blu-Ray, livres numériques) ? De plus, si le caractère explorateur des consommateurs en ligne s'avère vérifié, il reste à savoir si la consommation de titres anciens provient d'un effet d'exposition entre différents médias (comme la programmation à la télévision d'un roman adapté d'un livre que les lecteurs souhaiteraient découvrir), d'un effet de bouche-à-oreille retardé, ou encore de la révélation tardive d'un auteur (comme la découverte d'un ancien livre d'un auteur nouvellement révélé) ? Ces questions posent un problème au niveau de la mesure cumulée du succès et d'audience des produits culturels à travers les différents réseaux de distribution. En effet, il est encore difficile de mesurer le succès cumulé d'un produit sur une longue période d'observation à partir d'archives des meilleures ventes. Le film « Harry Potter et le prince de sang-mêlé » illustre un cas concret de problème de mesure de succès cumulé. Présent depuis 761 jours dans les 100 meilleures ventes quotidiennes de Blu-ray sur Amazon.fr et situé à la 66e position à l'heure où nous écrivons ce paragraphe, le film ne fait plus partie des 100 meilleures ventes annuelles de 2012 sur le même site. De manière plus générale, l'accès aux classements des meilleures ventes ne permet pas de déterminer précisément le nombre de ventes de produits limitant de fait la construction d'une mesure de succès cumulé.

L'effet de la réputation des vendeurs, abordée dans le cas de la dispersion des prix, reste difficile à capturer par les variables traditionnelles liées à la note et au nombre de notes de ces derniers. Ces variables n'ont pas toujours démontré leur significativité. Si des mesures complémentaires comme la taille du catalogue et un effet de clientèle ont été testées, d'autres mesures comme l'ancienneté des vendeurs sur la plateforme où même leur ordre d'apparition dans les offres méritent d'être étudiées.

Enfin, nos résultats obtenus sur l'analyse des livres numériques doivent s'étendre sur une période d'observation plus longue. Le succès relativement récent des livres numériques fait toujours l'objet d'une guerre de pouvoir entre les différents acteurs de la chaîne (auteurs, éditeurs, distributeurs et autres intermédiaires). L'absence de format universellement adopté et l'incompatibilité entre les supports limitent nos résultats à un modèle économique (celui d'Amazon) qui n'est pas assuré de dominer la distribution des livres numériques dans les années futures. Nous pensons que des études menées simultanément sur d'autres plateformes comme l'iTunes Store ou la Fnac et sur d'autres contenus (magazines, journaux, BD) permettraient de mieux caractériser la demande de l'édition numérique et de vérifier ou non l'effet de cannibalisation de ces derniers sur les ventes papiers.

Enfin nous souhaitons souligner que les résultats issus de cette thèse reposent avant tout sur l'exploitation et le traitement d'un nombre conséquent de données publiques de plus en plus riches et accessibles sur Internet. Ces résultats n'auraient pas été possibles sans l'utilisation de traitements statistiques et de recoupement de données qui se sont avérés très longs. La limite de nos travaux de recherche sur le commerce électronique de produits culturels ne repose plus désormais sur le manque de données accessibles sur la toile mais sur le traitement et l'exploitation d'une abondance d'informations parfois difficiles à croiser. Pour autant, l'extension de ce type de travaux permettrait de mieux comprendre comment fonctionne la dynamique du bouche à oreille en ligne et comment les coûts de recherche d'information et de transaction liés à l'achat de produits dans les réseaux de distribution physique sont levés par la mise en relation directe entre vendeurs et consommateurs en ligne.

Pour conclure, nous souhaitons insister sur les mutations sans précédent que connaissent aujourd'hui les industries culturelles. Si sur le plan technologique ces industries ont radicalement changé en deux décennies, il n'en est pas de même sur le plan économique et culturel. Fondées à l'origine grâce au contrôle de la reproductibilité à travers l'utilisation de supports, les industries culturelles doivent aujourd'hui s'adapter à une demande de produits culturels dématérialisés. Cette remise en question de leur modèle économique ouvre le champ de nouvelles possibilités de recherche. Tout d'abord, les meilleures ventes de musique enregistrée sous format MP3 sur iTunes ou Amazon et

basées sur la vente de CD pourraient être comparées aux titres les plus écoutés sur les sites de streaming comme Spotify, Deezer ou encore Youtube. Ce type de comparaison qui pourrait tout aussi bien s'appliquer aux séries télévisées ou aux bibliothèques proposant des livres numériques, permettrait de mieux comprendre les modes de consommation des internautes et la variété effectivement consommée dans le cas de deux modèles économiques différents. De même, la théorie de la longue traîne issue du commerce électronique de biens physiques et prédisant un rééquilibre des ventes vers des produits de niche reste encore à être vérifiée dans cet environnement complètement dématérialisé. Si Internet reste un formidable outil d'accès à la culture vers des contenus de plus en plus variés, il reste d'un point de vue académique à prouver que cet accès ne favorise pas seulement un nombre limité d'artistes.