

Substitution et complémentarité de la consommation de musique enregistrée à l'ère du numérique

Marc Ivaldi, Ambre Nicolle, Frank Verboven et Jiekai Zhang

Août 2022

Résumé

Les nouveaux canaux de consommation de musique enregistrée se substituent-ils aux canaux traditionnels, ou sont-ils complémentaires ? Cette étude s'appuie sur une base de données fournie par GfK contenant les ventes et les consommations en flux (*streams*) hebdomadaires de 30 millions de produits musicaux observés entre 2014 et 2017 pour la France. Les résultats issus de l'analyse économétrique indiquent que les consommations sur les plateformes de diffusion en flux (streaming) se substituent aux téléchargements, au niveau des titres. De même, au niveau des albums, les résultats mettent en lumière que les téléchargements se substituent aux ventes physiques. Au niveau plus agrégé des artistes, les téléchargements se substituent aux ventes physiques, alors que les consommations en flux tendent à avoir un effet promotionnel sur celles-ci. Cette complémentarité du streaming et des ventes sur le canal traditionnel est principalement portée par les genres les plus populaires. La plupart des résultats obtenus sont robustes que l'on considère seulement les titres et albums les plus populaires - ou bien que l'on inclut les produits de niche, c'est-à-dire ceux qui appartiennent à la longue traîne (*long tail*). Les variations de prix sont ensuite exploitées pour estimer l'impact du streaming sur le canal de vente numérique, ainsi que l'impact de chacun des deux nouveaux canaux de consommation (téléchargement et streaming) sur l'ancien canal de vente physique. Au niveau du produit, c'est-à-dire pour les titres et les albums, l'étude met en évidence l'existence d'un effet de substitution entre les nouveaux et les anciens canaux, ce qui a déjà été documenté dans d'autres études académiques. Les résultats suggèrent qu'un millier de streams se substitue à 1.21 téléchargements pour les titres et un album téléchargé se substitue à 19.7 albums physiques. Au niveau des artistes, il est trouvé un effet de substitution entre les téléchargements numériques et les ventes physiques. Dans le même temps, il existe cependant un effet de complémentarité entre le streaming et les ventes physiques. Tous genres confondus, un téléchargement se substitue à 3.4 ventes physiques (en équivalent titres), alors qu'un millier de streams stimulerait les ventes physiques de 14.8 unités équivalentes aux titres. Cet effet de complémentarité du streaming est principalement porté par les genres Pop et Variété. Cela suggère que les artistes qui sont positionnés dans des "segments" spécifiques ont bénéficié de l'introduction du canal de streaming (au moins en termes d'unités consommées). Cette conclusion complète certains résultats récents de la littérature. Enfin, au niveau du marché, les résultats ne sont pas concluants.

Introduction

Après plus d'une décennie de déclin, l'industrie musicale renoue avec la croissance. Depuis 2015, ses revenus mondiaux sont à nouveau à la hausse et, depuis 2017, la majeure partie est issue des services de streaming, qui ont timidement débuté en 2005. L'industrie de la musique a connu d'importants changements depuis le début des années 2000, notamment causés par la numérisation qui a drastiquement accéléré ces transformations, en impactant simultanément la nature des produits, leur distribution et leur mode de consommation. Ces transformations sont communes à de nombreuses industries culturelles (livre, cinéma, jeux vidéo, etc.) et ont attiré l'attention des chercheurs de tous horizons académiques. En économie, les principales questions étudiées par les universitaires concernent les prix des produits, le partage des revenus, mais aussi de la substitution entre les canaux de consommation. Bien qu'il existe aujourd'hui une littérature croissante visant à faire la lumière sur la question de la substitution entre les canaux, il n'existe pas d'approche systématique qui prend en compte l'ensemble des canaux (légaux) de consommation. Cette étude a pour objectif d'en proposer une.

Ce faisant, cette étude cherche à contribuer à la définition de politique économique pour l'industrie musicale. En effet, ce qu'elle documente est intrinsèquement lié à des questions portant sur la demande dans cette industrie, aux revenus des producteurs et distributeurs de musique, mais aussi à la structure du marché. En effet, le remplacement de certains produits par d'autres produits – et de façon plus pertinente dans ce cas étudié, de certains canaux de distribution par d'autres - peut entraîner des changements significatifs dans la demande et les revenus des différents acteurs de l'industrie. En termes de structure de marché, les distributeurs de musique mais aussi les artistes sont impactés par la nouvelle donne numérique. Alors que les majors et les grands détaillants constituaient les acteurs clés de l'industrie de la musique enregistrée avant la numérisation, quelques plateformes dominantes (YouTube, Spotify, Apple) occupent désormais une grande partie de « l'espace économique » et concentrent une large part des revenus générés. En ce qui concerne les artistes, la numérisation est souvent décrite comme

ayant favorisé les "stars", même si elle permet à un grand nombre d'artistes de "niche" de se faire connaître et de vendre leurs produits, même dans des volumes limités.

En termes de partage de revenus, la numérisation de l'industrie a augmenté le nombre d'artistes liés à une maison de disques par un contrat de licence, dans lequel le producteur n'est en charge que de la promotion et de la distribution, par opposition à une forme de contrat plus traditionnelle, c'est-à-dire les contrats d'artistes où la maison de disques prend également en charge la production de l'album ou du single.¹ De plus, aujourd'hui, de plus en plus d'artistes utilisent les plateformes de streaming comme principaux canaux de promotion de leurs œuvres. Il n'est cependant pas clair si ce changement de méthode de promotion de la musique profite ou non à l'ensemble des artistes. Etudier l'impact de la numérisation sur la popularité des artistes (mesurée par la consommation de leur musique) offre un éclairage sur cette question. Enfin, la mise en évidence de l'hétérogénéité entre les genres est importante pour comprendre comment la variété de la consommation de produits culturels peut être touchée par la numérisation.

En utilisant un jeu de données unique mis à notre disposition par l'institut d'étude de marché GfK et couvrant la quasi-totalité du marché français de la musique enregistrée entre 2014 et 2017, l'étude cherche à documenter s'il existe un effet de substitution ou de complémentarité entre les anciens et les nouveaux modes de consommation de la musique enregistrée - ce qui sera nommé par la suite les canaux de consommation. Les canaux de consommation traditionnels correspondent aux ventes physiques (CD et vinyles), tandis que les nouveaux correspondent aux canaux numériques, représentés par les ventes numériques (téléchargements) et les consommations en flux (streaming).

L'étude s'inscrit dans la continuité de la littérature qui étudie l'impact du numérique sur l'économie, en particulier dans les industries culturelles. Du côté de l'offre, des travaux antérieurs suggèrent que la numérisation a entraîné une augmentation de la qualité² et une plus grande variété de contenus offerte.³ Du côté de la demande, certains travaux universitaires soulignent la façon dont la numérisation des contenus et l'émergence de nouveaux canaux de consommation ont conduit à une augmentation de la quantité consommée⁴. Certaines études empiriques suggèrent qu'il y a eu une augmentation de la variété consommée⁵, tandis que d'autres travaux académiques apportent un éclairage plus nuancé sur ce point.⁶

La présente étude contribue plus spécifiquement à la littérature qui se penche sur la substitution ou la complémentarité entre les canaux de consommation de musique, à la fois entre les canaux traditionnels et les nouveaux canaux et entre les canaux numériques eux-même. Deux effets principaux peuvent théoriquement exister. D'une part, on peut observer un effet de substitution où les consommations ou les revenus générés sur un canal remplacent ceux générés sur un autre canal. Cela arrive quand les canaux sont considérés comme substituables et les produits identiques aux yeux des consommateurs. D'autre part, on peut observer un effet de complémentarité où les consommations sur un canal stimulent les ventes sur un autre canal. On parle aussi d'effet promotionnel entre canaux. Cela peut être lié à des effets de réseaux (*network effects*). Un produit grandement diffusé sur un canal (par exemple fortement partagé sur des réseaux de pair à pair) peut attirer l'attention des consommateurs qui va mécaniquement augmenter la demande sur un autre canal, par exemple le canal légal.⁷ Cela peut également être lié à un effet d'échantillonnage (*sampling effect*). Les consommateurs testent le produit sur un canal (par exemple en téléchargeant illégalement le titre ou en l'écoutant sur une plateforme de streaming), évalue sa qualité et décide ou non d'acheter une version légale ou physique.⁸

Les travaux de recherche en économie ont plutôt mis en évidence un effet de substitution important concernant les contenus piratés et les contenus légaux – mais certains autres travaux apportent une nuance à cette observation. Liebowitz (2016)⁹ estime que la quasi-intégralité de perte de revenus de l'industrie de la musique observée aux Etats-Unis sur la période 1999-2005 peut

¹ Bearing Point, « Evaluation de la répartition des rémunérations entre producteurs phonographiques et artistes-interprètes ». Etude pour la Direction Générale des Médias et des Industries Culturelles, 2017.

² Joel WALDFOGEL, "Copyright protection, technological change, and the quality of new products: Evidence from recorded music since Napster", *The Journal of Law and Economics*, 2012, Luis Aguiar et Joel Waldfogel, "Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music", *Journal of Political Economy*, 2018c.

³ Michael LUCA et Craig MCFADDEN. "How Streaming Is Changing Music (Again)." *Harvard Business Review*, 2016, et Luis Aguiar et Joel Waldfogel (2018c) cité précédemment.

⁴ Hannes DATTA, George KNOX et Bart J. BRONNENBERG, "Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery", *Marketing Science*, 2018.

⁵ Voir les articles de Luca et Mc Fadden, 2016 et Datta et al. 2018 cités plus haut.

⁶ Tobias KRETSCHMER et Christian PEUKERT, "Video killed the radio star? Online music videos and recorded music sales", *Information Systems Research*, 2020.

⁷ Christian PEUKERT, Jörg CLAUSSEN et Tobias KRETSCHMER, "Piracy and box office movie revenues: Evidence from Megaupload." *International Journal of Industrial Organization*, 2017.

⁸ Martin PEITZ et Patrick WAELBROECK, "Why the music industry may gain from free downloading—the role of sampling", *International Journal of Industrial Organization*, 2006.

⁹ Stan J. LIEBOWITZ, "How much of the decline in sound recording sales is due to file-sharing?", *Journal of Cultural Economics*, 2016.

être expliquées par l'essor du piratage. Aguiar and Martens (2016)¹⁰ utilisent des données de pays européens pour 2011 et ne trouvent pas d'effet significatif du piratage sur les ventes de produits musicaux. Finalement, on peut citer les travaux de Peukert, Claussen et Kretschmer (2017) précédemment évoqués qui mettent en évidence une complémentarité des contenus piratés et des revenus légaux dans l'industrie cinématographique.

Concernant l'impact du streaming, une littérature plus récente s'y penche. Par exemple, Wlömert et Papies (2016)¹¹ mettent en évidence l'existence d'une cannibalisation entre les consommations sur les plateformes de streaming et les revenus issus de la vente de produits physiques. Hiller (2016)¹² trouve un effet significatif de substitution entre les consommations de musique ayant lieu sur YouTube et les ventes physiques. En revanche, Kretschmer et Peukert (2020) soutiennent que YouTube a généré des externalités positives sur les ventes physiques (avec 20 % de revenus supplémentaires, toutes choses égales par ailleurs, pour les titres disponibles sur la plateforme). La présente étude est proche de celle d'Aguiar et Waldfogel (2018b)¹³ qui analyse l'impact du streaming sur les ventes physiques et le piratage. Les chercheurs s'appuient sur des données hebdomadaires de plus de 21 pays, observées entre avril et décembre 2013. Cependant, leurs données de streaming ne couvrent que le top 50. Ils exploitent également les ventes agrégées de titres et d'albums, en format numérique et physique, pour la période 2012-2013 pour les États-Unis, qui est antérieure au véritable boom de la consommation en streaming. En France, ce boom a eu lieu en 2015, ce qui est couvert par les données sur lesquelles s'appuie la présente étude.

En résumé, cet aperçu de la littérature indique qu'il existe une littérature importante mais encore quelque peu fragmentée sur la question de la substitution des consommations dans l'industrie musicale. Les conclusions de cette littérature sont parfois contradictoires, ce qui peut en partie être dû à l'étude de périodes différentes mais aussi à l'hétérogénéité des méthodologies adoptées. Grâce aux données détaillées fournies par GfK, il est possible d'étudier la globalité du marché en prenant en compte les ventes physiques et numériques, ainsi que le streaming - ce qui a rarement été possible jusque-là. En outre, la base de données GfK inclut non seulement les titres les plus populaires mais aussi la myriade de produits plus confidentiels - ce qui permet d'analyser l'effet de substitution ou de complémentarité qui a lieu dans la longue traîne, ce qui est aussi rare dans les études menées jusque-là.

En termes d'approches empiriques, les travaux antérieurs peuvent être classés en trois catégories : (i) les travaux basés sur les données individuelles des consommateurs de musique, (ii) les travaux qui exploitent des chocs exogènes, et (iii) les travaux utilisant des données agrégées au niveau des produits et des artistes.

Tout d'abord, un pan de la littérature exploite la disponibilité de données au niveau des individus. Par exemple, Waldfogel (2010)¹⁴ utilise des données d'enquête pour évaluer la disposition à payer de consommateurs pour des produits illégaux et légaux. Il constate que le taux de substitution des ventes se situe entre -0,15 et 0,3, ce qui signifie qu'un titre piraté réduit l'achat d'un tiers à un sixième de titre. Wlömert et Papies (2016) utilisent un design quasi expérimental et des données d'enquête pour une importante population de consommateurs de musique, ce qui leur permet d'intégrer des effets fixes individuels dans leur analyse empirique. Aguiar et Martens (2016) exploitent des données individuelles de navigation en ligne (*clickstream data*) et estiment un modèle qui capture à la fois des effets fixes pour les individus et les pays.

Ensuite, plusieurs travaux académiques s'appuient sur des données de ventes agrégées et exploitent des chocs exogènes pour estimer des modèles de doubles différences (*difference-in-differences models*). Hong (2013) exploite l'introduction de Napster - un logiciel internet pionnier de partage de fichiers en pair à pair - en 1999, resté dominant jusqu'à sa fermeture en 2001 après avoir rencontré des difficultés juridiques liées à la violation des droits d'auteur. Hiller (2016) utilise le choc Warner en 2009, au cours duquel tous les contenus produits par Warner ont été soudainement retiré de la plateforme YouTube pendant une période de neuf mois. Kretschmer et Peukert (2020) utilisent le choc GEMA en 2009. Cette année, un conflit juridique a éclaté entre la société de perception des droits d'auteur qui représente les artistes en Allemagne (GEMA) et YouTube. Il a entraîné le blocage de presque toutes les vidéos musicales dans ce pays pendant plusieurs mois.

Finalement, certains articles se sont appuyés sur des données de ventes agrégées au niveau des produits et des artistes. Par exemple, on peut citer les travaux d'Aguiar et Waldfogel (2018b) qui disposent de données au niveau des chansons et des artistes pour plusieurs pays. Cela leur permet d'estimer l'impact du streaming sur d'autres canaux musicaux en utilisant des effets fixes par chanson et par pays, et donc de tenir compte des chocs communs dans la popularité des chansons entre les pays. La présente étude appartient à cette troisième catégorie. La base de données GfK contient des données détaillées au niveau produit, c'est-à-dire au niveau des titres et albums, ce qui permet d'inclure des effets fixes utiles pour capturer des effets inobservés qui leur sont spécifiques. Cependant, contrairement à l'étude menée par Aguiar et Waldfogel (2018b), les données mises à disposition pour

¹⁰ Luis AGUIAR et Bertin MARTENS, "Digital music consumption on the internet: evidence from clickstream data", *Information Economics and Policy*, 2016.

¹¹ Nils WLÖMERT et Dominik PAPIES, "On-demand streaming services and music industry revenues—Insights from Spotify's market entry", *International Journal of Research in Marketing*, 2016.

¹² Scott R. HILLER, "Sales displacement and streaming music: Evidence from YouTube", *Information Economics and Policy*, 2016.

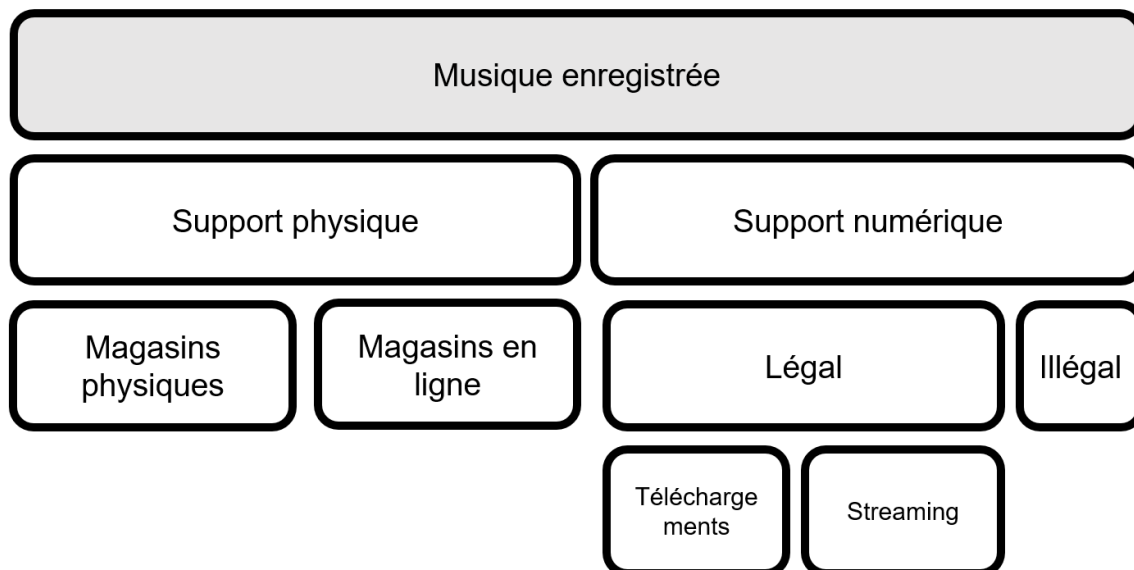
¹³ Luis AGUIAR et Joel WALDFOGEL, "As streaming reaches flood stage, does it stimulate or depress music sales?", *International Journal of Industrial Organization*, 2018b.

¹⁴ Joel WALDFOGEL, "Music file sharing and sales displacement in the iTunes era" *Information economics and policy*, 2010.

cette étude ne couvrent qu'un seul pays, la France, et il n'est donc pas possible d'exploiter des variations entre les pays. Par conséquent, la stratégie empirique développée dans cette étude s'appuie sur la variation des prix des produits.

Structure du marché

Schéma 1: Vue d'ensemble des canaux de consommations



Le schéma 1 présente la structuration du marché entre les différents canaux de distribution de la musique. Les produits de musique enregistrée sont généralement vendus sous trois formes : (i) en tant que produits physiques (CD et vinyle) vendus par des détaillants en magasin ou en ligne ; (ii) en tant que produits numériques, c'est-à-dire de la musique téléchargée à partir de plateformes et de sites web ; (iii) en tant que flux diffusés par des plateformes de streaming, proposé gratuitement (c'est-à-dire contre l'écoute ou de visionnage de publicités) ou contre le paiement d'un abonnement mensuel. Le schéma 1 propose une vue d'ensemble des différents canaux de consommation.

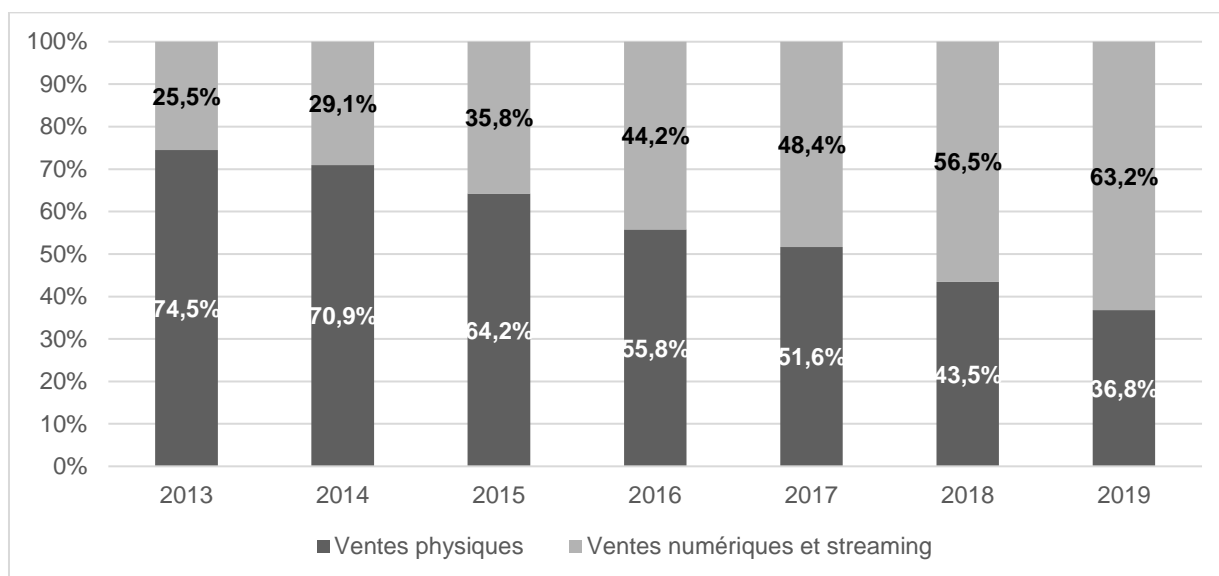
Le marché des produits physiques est relativement fragmenté, les détaillants étant de taille très variable. Par exemple, les supermarchés, les magasins indépendants spécialisés et les « places de marché » telles que eBay et Amazon proposent leurs produits sur le marché physique. Le marché des produits numériques (c'est-à-dire les téléchargements) est beaucoup plus concentré, avec seulement quelques plateformes proposant ces produits aux consommateurs. On peut citer ici iTunes, Amazon Music et Soundcloud. Le marché du streaming est encore plus concentré : un petit nombre d'acteur offre soit des services audio (Spotify, Deezer, Apple Music, Amazon Music), soit des services vidéo (YouTube), soit des services radio (Pandora et Napster). Par conséquent, il est beaucoup plus courant d'observer une variation des prix des produits physiques entre les détaillants et dans le temps. Pour les produits numériques, on observe souvent des prix focaux, tels que 1,29€ pour un titre et 9,99€ pour un album en France. Les prix des abonnements aux services de streaming sont également relativement homogènes dans le temps et entre les plateformes, avec un prix focal de 9,99€ par mois en France.

Un autre canal de consommation, qui a pris son essor avec le numérique, est également à mentionner ici : le piratage (notamment via les réseaux de partage de pair à pair). Ces consommations sont très difficiles à observer et à mesurer (notamment au niveau des titres) et la présente étude ne permettra pas de commenter sur son impact – elle se concentre sur les canaux légaux pour lesquels des données détaillées et fiables sont disponibles.

Si l'on regarde les revenus générés par l'industrie au niveau mondial,¹⁵ on peut observer que les ventes physiques (à l'exclusion des droits d'exécution et de la synchronisation) représentaient pratiquement 100 % des revenus de la musique enregistrée en 2001. Les revenus générés par le téléchargement sont observés à partir de 2004 et ont progressivement augmenté pour atteindre 35 % des revenus de l'industrie en 2013. Depuis, on observe un déclin de sa part de marché qui a atteint environ 9 % en 2019. Le streaming a timidement commencé en 2005 pour atteindre une part de marché de 8% en 2012. Son essor a véritablement eu lieu en 2016 lorsqu'il atteint une part de marché de 35%. S'en suit une augmentation continue d'au moins 10 points de pourcentage tous les ans, pour atteindre 66 % du marché en 2019.

¹⁵ International Federation of the Phonographic Industry, "Global Music Report", 2019.

Graphique 1 : Part de marché de la musique enregistrée en France



Source : D'après les données publiées par le Syndicat National de l'Édition Phonographique (2014-2020).

Si on se penche sur la situation française plus en détail, on observe que la part des revenus générés par les téléchargements et le streaming est passée d'environ 30% en 2014 à 50% en 2017, comme documenté dans les rapports du Syndicat national de l'édition phonographique (Snep) et résumé dans le graphique 1. En 2019, le numérique représentait 65% des revenus de l'industrie. On a donc assisté, au cours des dernières années, à un basculement majeur, couvert par les données GfK sur lesquelles s'appuient la présente étude.

Données

La principale source de données utilisées pour cette étude provient de GfK. Elle contient des informations hebdomadaires de volumes et de prix pour la quasi-totalité des produits (titres et albums) vendus ou consommés en flux en France entre 2014 et 2017.¹⁶ En outre, les données contiennent les caractéristiques des produits, telles que le nom de l'artiste, de l'éditeur et du label, son genre et sous-genre musical principal et sa semaine de sortie. Plusieurs types de produits sont observés dans cette base de données : les produits physiques (albums ou singles), les produits numériques, c'est-à-dire les téléchargements légaux (titres ou albums) et les streams. Les produits physiques sont des CDs ou des vinyles, vendus dans les supermarchés, les magasins spécialisés et les magasins en ligne (Fnac, Amazon, etc). Les produits numériques sont des téléchargements qui ont lieu sur des plateformes légales telles que iTunes et Qobuz. Les streams sont diffusés par partir de plateformes audio telles que Spotify, Deezer ou Qobuz ; de plateformes vidéo telles que YouTube et Dailymotion, ou de radio (Napster). Même si les quantités pour chaque type de streaming (streaming audio premium, audio gratuit et vidéo gratuite) sont observées dans la base de données GfK, elles sont regroupées pour l'analyse car les quantités observées sont fortement corrélées entre elles. Pour les produits physiques et numériques, la base de données permet d'observer à la fois les quantités et les revenus générés, et permet donc de calculer le prix moyen du produit pour une semaine donnée. Pour le streaming, un prix unitaire ne peut être calculé de la même manière puisque les utilisateurs paient un abonnement mensuel qui leur permet de consommer autant de musique que souhaité ou acceptent une écoute interrompue par des publicités – dans les deux cas, une quantité consommée ne peut que difficilement être associée à un prix. Enfin, ces données sont complétées par un jeu de données publié par l'Agence nationale des fréquences radioélectriques (ANFR) qui contient des informations sur le nombre d'antennes 2G, 3G et 4G sur le territoire français entre 1997 et 2019.

Période étudiée

Pour les ventes physiques et numériques, l'ensemble de données original couvre la période 2006 - 2018. Les données sur le streaming commencent pendant la dernière semaine de 2014 et se terminent en 2018. Cependant, GfK ne collecte plus les chiffres du streaming vidéo depuis janvier 2018 – la période d'analyse est donc limitée à la période 2014 – 2017 pour s'assurer que les consommations sur tous les canaux sont simultanément observées. Aussi, il est important de noter que le jeu de données initial est constitué d'un grand nombre de données en coupe au niveau hebdomadaire. Elles sont fusionnées grâce au nom du titre ou de l'album - et celui de l'artiste, de façon à obtenir un jeu de données de panel. 56 des 208 semaines sont supprimées

¹⁶ Les données couvrent la quasi-totalité du marché de la musique numérique (environ 99 % pour les ventes numériques et tous les streams, selon une discussion informelle avec GfK) et également la quasi-totalité du marché physique (environ 95 %).

car les données fournies par GfK étaient incomplètes. La base de données finale contient 152 semaines, qui correspondent à la période qui va de la fin de l'année 2014 à la fin de l'année 2017.

Interpolations

La création du jeu de données de panel met en évidence le fait que certains produits ne sont pas vendus ou consommés chaque mois depuis la date de lancement. Comme GfK a indiqué enregistrer pratiquement toutes les ventes physiques et numériques, l'absence de ventes d'un produit au cours d'une semaine donnée est codée avec une valeur zéro si elle a lieu après la semaine où la première vente a lieu. Pour ces produits dont les ventes sont nulles pendant certaines semaines, le prix moyen n'est pas observé. Une simple interpolation linéaire est alors utilisée pour imputer un prix. En ce qui concerne le streaming, l'interprétation d'une quantité manquante prend un autre sens. En effet, pour apparaître dans le panel de GfK, un titre doit être streamé au moins 100 fois sur les plateformes de streaming audio et au moins 1000 fois sur les plateformes de streaming vidéo sur une semaine donnée. Toutes les consommations inférieures à ce seuil entrent dans un panier qui est fourni par GfK. Bien que ce panier permette de calculer le nombre total d'unités consommées sur le marché, les données sont partiellement censurées au niveau des titres. Pour traiter ce problème, l'approche retenue dans cette étude est d'utiliser une interpolation linéaire pour les titres avec des valeurs manquantes. Ceci est appliqué seulement pour les titres et albums pour lesquels il est raisonnable de faire l'hypothèse de disponibilité du produit lors de la semaine en question. Cette interpolation a lieu au niveau du titre lorsque cela est possible, au niveau de l'artiste sinon.

Types de produit

Pour les ventes physiques, GfK code explicitement si un produit est un album ou un single. Pour les ventes numériques, cette information n'est que partiellement codée. Lorsqu'elle est manquante, le produit est défini comme un titre si le prix est au maximum de 3€ et comme un album si le prix est supérieur à 4,9€. Les produits dont le prix se situe entre les deux seuils, qui représentent un nombre négligeable d'observations (moins de 0,5 %), sont supprimés. En ce qui concerne les consommations en flux, il s'agit toujours de titres dans le jeu de données GfK. Un petit nombre d'observations contenant un prix aberrant est également supprimé : les produits dont le prix est inférieur à 0,9€ ou supérieur à 60€ sur le canal physique (ce qui élimine les éditions spéciales très haut de gamme); et les produits dont le prix est supérieur à 30€ pour les albums numériques et supérieur à 5€ pour les titres numériques (lorsque cela a été codé par GfK). Enfin, les observations présentant une variation de prix aberrante (au-delà de +100% du prix d'un produit donné) sont supprimées.

Statistiques descriptives

L'ensemble de données final est un panel de 29 672 540 observations, couvrant 4 330 868 produits uniques de 894 134 artistes uniques observés pendant 152 semaines. Les albums représentent 25% de tous les produits et sont vendus soit sous forme physique, soit sous forme numérique (donc de téléchargement). Les titres constituent les 75% restants des produits. Ils sont consommés soit sous forme de vente numérique, soit en flux. Pour rappel, les singles vendus sous forme physique (qui pourraient être apparentés à des titres) sont exclus de l'analyse.

Volumes et parts de marché par canal de consommation

Comme le jeu de données disponible pour cette étude couvre la quasi-totalité du marché français entre 2014 et 2017, une description de l'évolution des parts de marché peut être faite pour chaque canal : ventes physiques, ventes numériques et flux. Comme l'information sur le prix moyen d'un stream n'est pas disponible, les statistiques descriptives porteront sur les parts de marché en volume plutôt qu'en valeur. Pour agréger les ventes sur les titres et les albums, les ventes d'albums sont converties en unités équivalentes aux titres, en supposant qu'il y a 10 titres par album, ce qui est une hypothèse courante dans la littérature.¹⁷

Le volume total des ventes numériques est en moyenne de 827 196 unités équivalentes à des titres par semaine. Ce chiffre est en effet considérablement inférieur au volume total des ventes physiques, qui s'élève en moyenne à 3.9 millions par semaine, et au volume total de streams d'en moyenne 535 millions d'unités par semaine. Par rapport aux chiffres mondiaux, le marché français semble se différencier, surtout concernant le nombre de ventes numériques qui représentent une faible part du volume chaque semaine. En outre, le marché physique représente toujours une part importante des revenus dans le pays. Ceci a été décrit comme une " exception française" par le Snep (SNEP, 2018 ; SNEP, 2019 et SNEP, 2020).

Cet aperçu semble indiquer que le phénomène économique le plus pertinent à étudier est l'impact du streaming sur les ventes physiques. Toutefois, les ventes numériques génèrent toujours des revenus importants pour le secteur par rapport au streaming, et c'est pour cette raison que l'étude de l'impact du streaming sur les ventes numériques présente également un intérêt économique. Il convient de noter qu'il n'est pas possible d'évaluer directement l'impact du streaming sur les ventes physiques au niveau des titres car les ventes physiques de titres (singles) sont pratiquement inexistantes. Cette partie de l'analyse sera donc proposée au niveau de l'artiste.

Ventes et variation des prix

¹⁷ Voir par exemple Aguiar et Waldfogel (2018c).

Tableau 1: Volumes, flux et prix par semaine, au niveau produit, 2014-2017

	Moyenne	Ecart type	Minimum	Maximum	Unités et euros Nombre d'observations
Volumes hebdomadaires des ventes et des flux					
Albums physiques (CDs et vinyles)	9	187	0	111 677	6 273 804
Titres numériques	2	18	0	8 149	18 520 524
Albums numériques	2	24	0	20 832	4 890 143
Consommations en flux (streams)	10 839	63 618	100	12 019 576	7 509 941
Prix des produits					
Albums physiques	13,5	7	0,9	60	6 273 804
Titres numériques	1,3	0,2	0,2	5	18 520 519
Albums numériques	10	2,1	0,5	30	4 890 143

Source : GfK/TSE

Le tableau 1 présente les statistiques récapitulatives des principales variables. La partie supérieure du tableau présente les statistiques concernant les volumes de ventes hebdomadaires, ventilées par canal (physique, numérique et streaming) et par format (titre et album). On peut observer que le volume des ventes numériques est en moyenne très faible, avec de nombreuses valeurs nulles. Les ventes physiques et les volumes de streaming sont en moyenne beaucoup plus élevés. Alors que le volume moyen des ventes d'albums semble être d'une ampleur comparable à celui des flux, il est en fait beaucoup plus élevé une fois converti en unités équivalentes aux titres (c'est-à-dire après avoir été multiplié par 10). Une représentation des volumes au cours de la période étudiée permet de constater que le volume total des ventes physiques reste relativement stable, à l'exception des traditionnels pics de fin d'année autour de Noël. Le volume total des ventes numériques est en baisse constante sur la période étudiée et a été divisé par deux entre la fin de 2014 et la fin de 2017. Quant au nombre de streams, il est en constante hausse sur la période : il commence aux environs de 600 millions de streams en 2014 pour atteindre environ 1,4 milliard en 2017. Un choc est observé pour le nombre de streams comptabilisé par GfK au milieu de l'année 2016 – cela est dû à un changement dans la méthode de construction des données. Pour tenir compte de cet événement, des effets fixes temporels seront introduits dans les estimations économétriques. En moyenne, un album physique est vendu environ 9 fois par semaine, tandis qu'un album numérique est vendu deux fois par semaine. Les chansons disponibles sur les plateformes de streaming sont streamées 11 000 fois par semaine en moyenne.

La partie inférieure du tableau 1 présente les statistiques concernant les prix, qui sont des déterminants clés des volumes de vente. Le prix d'un titre numérique (c'est-à-dire le téléchargement d'une chanson) est en moyenne de 1,3€, avec un écart type de 0,2€. Le prix d'un album numérique est en moyenne de 10€, avec un écart-type de 2€. Le prix d'un album physique est en moyenne de 13,5€, avec un écart-type de 7€. Une représentation des prix dans un histogramme permet d'observer que les prix des titres et des albums numériques apparaissent comme "standard" et ne varient que très peu selon les produits et dans le temps. En effet, la grande majorité des titres et albums numériques sont vendus respectivement à 1,29€ et 9,99€. Néanmoins, il existe des variations autour de ces valeurs de prix, en particulier pour les albums. Les prix des produits physiques présentent des variations beaucoup plus importantes. Bien qu'il existe des prix focaux pour les albums physiques à 6,99€, 10€ et 14€, il reste une variation considérable autour de ces valeurs focales.

Tops et genres

La base de données permet d'observer une forte concentration des ventes et des streams sur un nombre restreint de titres et d'albums, ce qui est typique des industries culturelles où coexistent les produits très populaires (les *Hits*) et les produits de niche, plus confidentiels (qui appartiennent à la longue traîne, ou *long tail*).¹⁸ La disponibilité d'informations sur les ventes des produits faisant partie de cette longue traîne est rare – c'est pour cela que la présente étude propose une analyse additionnelle qui

¹⁸ Chris ANDERSON, "The long tail: Why the future of business is selling less of more", Hachette Books, 2006.

considèrera en particulier les 50, 200, 1000 ou 5 000 produits les plus populaires, de façon à comparer les résultats obtenus avec les données couvrant l'ensemble des produits.

Dans la base de données, on observe que près de 27% des produits disponibles appartiennent au genre "Pop". Cinq autres genres représentent, chacun, entre 3 et 7 % des produits disponibles : Rock, Musique urbaine, Variété, Electro/Dance et Musique classique. Ces genres (ceux qui viennent d'être listés ainsi que la Pop) représentent à eux seuls 60% des unités équivalentes aux titres consommées. Pour chaque niveau d'agrégation des données (au niveau du produit et de l'artiste), une analyse sera effectuée pour chacun de ces genres afin de commenter l'hétérogénéité de la substitution ou de la complémentarité des canaux de consommation.

Méthode

Sur la base des données hebdomadaires au niveau des produits, l'objectif est d'identifier dans quelle mesure les canaux de consommation de musique sont substitués ou compléments, du point de vue des consommateurs. Plus précisément, il s'agit ici de mesurer l'impact des nouveaux canaux musicaux sur les canaux traditionnels, c'est-à-dire à la fois l'impact du canal de streaming sur le canal de vente numérique mais aussi l'impact des canaux de vente numérique et de streaming sur le canal physique.

Deux approches sont retenues dans cette étude : une approche désagrégée, au niveau produit, et une approche plus agrégée au niveau artiste et marché. L'analyse au niveau produit sera conduite d'une part au niveau des titres et d'autre part au niveau des albums, puisque les deux types de produits ne sont pas disponibles de la même manière sur tous les canaux de consommation, comme précédemment documenté. Au niveau titre, l'analyse se concentre sur l'impact du streaming sur les ventes numériques. Au niveau album, l'analyse se porte sur l'impact du canal numérique sur le canal physique. L'analyse agrégée permettra de 'mélanger' les titres et albums en calculant les ventes et streams totaux réalisés, une semaine donnée, par un artiste, ou, alternativement, par tous les artistes (analyse au niveau du marché). Pour cela, il est nécessaire de convertir les ventes d'albums en ventes d'unités équivalentes aux titres (*track-equivalent units* dans la littérature) – il s'agit alors simplement de convertir chaque album en 10 titres. Cette approche permet de mesurer l'impact des canaux numériques et de streaming sur le canal physique dans une seule et unique régression.

Pour cela, nous utilisons un modèle qui met en relation les quantités vendues sur un canal de consommation donné avec celles vendues sur les autres canaux. De façon formelle, on note q_{it}^c la quantité de produit i vendue ou consommée en flux au temps t sur le canal c . Un produit i peut désigner un titre ou un album. Dans l'analyse au niveau de l'artiste, i fera référence à l'artiste (et les quantités seront le total des unités équivalentes à un titre pour l'ensemble des titres et des albums). Le canal c peut désigner le canal des ventes physiques, le canal des ventes numériques (c'est-à-dire les téléchargements) ou le canal du streaming. Comme suggéré dans d'autres travaux académiques, il est particulièrement pertinent de mesurer l'impact des consommations ayant lieu sur les nouveaux canaux de consommations sur les canaux plus anciens. On peut donc noter c soit O pour les anciens canaux (*old*), soit N pour les nouveaux (*new*). Le modèle de régression linéaire suivant est considéré :

$$q_{it}^O = \alpha_0 + \alpha_1 q_{it}^N + \mu_i + \theta_t + \varepsilon_{it}^O \quad (1)$$

où μ_i correspond à un effet fixe invariant dans le temps pour le produit i , θ_t un effet fixe pour la semaine t et ε_{it}^O est le terme d'erreur. L'intérêt se porte principalement sur le coefficient α_1 . Si $\alpha_1 < 0$, alors on peut conclure que les consommations ayant lieu sur le nouveau canal se substituent aux consommations ayant lieu sur le canal traditionnel. Si, au contraire $\alpha_1 > 0$, alors on peut conclure que les consommations ayant lieu sur le nouveau canal sont complémentaires aux consommations sur le canal ancien.

Le « sens » et la « force » de cette relation sont estimés par le biais de différentes méthodes statistiques. D'abord, on considère la méthode des Moindres Carrés Ordinaires (MCO), qui est la plus simple. Deuxièmement, on considère la méthode des Moindres Carrés Ordinaires utilisée avec un modèle qui inclut des effets-fixes individuels pour les produits ou les artistes. Cette approche permet de capturer l'hétérogénéité non observable, c'est-à-dire l'ensemble des caractéristiques propres aux titres, albums et artistes qu'il n'est pas possible d'introduire dans la régression économétrique car difficilement observables ou mesurables. Finalement, la méthode des Doubles Moindres Carrés Ordinaires (Doubles MCO) est considérée. Cette méthode, qui repose sur une régression en deux étapes, permet de bien définir les liens de causalité entre les variations observées sur les différents canaux de consommations. Les prix des produits sur le canal physique et numérique seront utilisés, dans les régressions de première étape, comme variables instrumentales. L'annexe propose une description détaillée des approches économétriques retenues et des hypothèses sur lesquelles elles reposent.

Résultats

La première sous-section présente les résultats de l'estimation de l'équation (1) au niveau le plus désagrégé, c'est-à-dire lorsque les produits i se réfèrent soit aux titres, soit aux albums. Étant donné que les titres ne sont disponibles que sur les canaux numériques et de streaming, et que les albums ne sont disponibles que sur les canaux physiques et numériques, ce niveau de désagrégation ne permet pas de considérer l'impact du numérique et du streaming sur le canal physique. Il sera donc étudié dans la deuxième sous-section à un niveau d'analyse plus agrégé : celui des artistes, où les titres et les albums peuvent être « mélangés » en utilisant des unités équivalentes aux titres. Les résultats de l'analyse au niveau de l'ensemble du marché français sont également présentés dans cette deuxième sous-section.

Analyse au niveau des titres et des albums

Trois modèles sont estimés ici : le premier est estimé par MCO sans effets-fixes individuels pour les produits, le deuxième est estimé par MCO avec effets-fixes individuels pour les produits et le troisième est estimé par Doubles MCO avec effets-fixes individuels pour les produits. Les trois régressions incluent des effets-fixes temporels. Les résultats correspondant à ces estimations sont présentés respectivement dans les colonnes (1), (2) et (3) du tableau 2 présenté ci-dessous.

Les deux premières régressions (colonnes 1 et 2) suggèrent un impact positif du streaming sur les ventes numériques. On peut remarquer que l'inclusion des effets fixes produits ne réduit pas l'association positive estimée. Cependant, ces régressions ignorent le problème d'endogénéité abordé précédemment et, par conséquent, aucun effet causal ne peut être conclu à ce stade. En effet, la complémentarité mise en lumière par ces régressions peut être due à la présence de chocs de demande sur certains produits certaines semaines, qui influencent à la fois la demande sur les canaux de vente numérique et de streaming. Les résultats de la troisième régression (colonne 3) sont obtenus après l'utilisation d'un instrument de prix pour le nombre de consommations en flux. Comme cela peut être anticipé, les résultats de la régression de première étape (présentés dans la seconde partie du tableau sous l'intitulé « Régression de première étape ») montrent que le volume de streams dépend positivement du prix du titre numérique, c'est-à-dire que lorsque le prix d'un téléchargement augmente pour un titre, les consommations en flux de ce même titre tend à augmenter.¹⁹ Les résultats obtenus pour la régression principale (celle de deuxième étape) suggèrent que le streaming a un impact négatif sur les ventes de titres numériques. Un millier de consommations en flux supplémentaire entraîne une diminution de 1,21 téléchargement. Cela confirme l'existence d'une substitution entre les ventes numériques et les streams qui est l'effet de substitution communément évoqué dans la littérature.

Tableau 2: Résultats de l'estimation au niveau titre

Régressions principales			
Variable à expliquer : nombre de téléchargements			
	MCO	MCO avec effets fixes individuels	Doubles MCO avec effets fixes individuels
Nombre de streams (en milliers)	0.17*** (0.01)	0.20*** (0.02)	-1.21*** (0.19)
Effets fixes semaine-année	Oui	Oui	Oui
Effets fixes titres	Non	Oui	Oui
Instruments	Non	Non	Oui
Régressions de première étape			
Variable à expliquer : nombre de streams (en milliers)			
Prix du téléchargement			5.75*** (0.82)
Nombre d'observations	6 352 183	6 352 183	6 352 178
Nombre de titres	231 740	231 740	231 740

Note : Les écart types sont groupés au niveau titre et sont indiquées entre parenthèses. Dans les tableaux, *** indique un niveau de significativité à 0,1 %, ** de 1 % et * de 5 %.

Source : GfK/TSE

Cette analyse est répétée en prenant en compte des sous-échantillons de produits populaires (Top 5000, Top 1000, Top 200 et Top 50). Le tableau 3 présente les résultats obtenus avec l'approche des Doubles MCO. Les résultats sont comparables à ceux obtenus en prenant en compte l'intégralité des produits. En effet, on trouve qu'il existe une substitution des téléchargements par les streams. Lorsqu'on se concentre sur les titres les plus populaires, cette substitution apparaît comme plus importante en termes de magnitude. Par exemple, pour le Top 200, les résultats de l'estimation montrent que mille streams supplémentaires entraînent une réduction des téléchargements de 2,6 unités. Pour le Top 50, le coefficient n'est pas significatif.

Tableau 3: Résultats des régressions sur sous-échantillons de titres

Régressions principales					
Variable à expliquer : nombre de téléchargements					
	Ensemble des titres	Top 5000	Top 1000	Top 200	Top 50
Nombre de streams (en milliers)	-1.21***	-1.07***	-1.14***	-2.58**	13.06

¹⁹ Notez que la régression de première étape n'inclut que l'effet de prix croisé. En effet, ici, il n'est pas possible d'évaluer l'impact du prix du stream sur le volume de streams car ce prix ne peut être défini de façon claire. Par conséquent, seuls les prix des titres numériques peuvent être utilisés comme instruments pour le nombre de streams.

	(0.19)	(0.18)	(0.23)	(1.04)	(33.29)
Effets fixes semaine-année	Oui	Oui	Oui	Oui	Oui
Effets fixes titres	Oui	Oui	Oui	Oui	Oui
Instruments	Oui	Oui	Oui	Oui	Oui
Régressions de première étape					
Variable à expliquer : nombre de streams (en milliers)					
Prix du téléchargement	5.75*** (0.81)	27.08*** (3.90)	57.64*** (10.21)	51.64** (19.59)	-14.03 (36.25)
Nombre d'observations	6 352 178	709 608	143 279	29 151	7 361
Nombre de titres	231 740	28 825	7 753	2 328	844

Source : GfK/TSE

Concernant les différents genres, les résultats indiquent que l'effet de substitution estimé est légèrement plus faible pour la Pop et la Musique Urbaine, alors qu'il semble être plus fort pour l'Electro et la Musique Classique. Cependant, en raison de la taille réduite des échantillons, les estimations sont également moins précises, et il convient donc d'être prudent avant de conclure à de fortes différences entre les genres. Dans l'ensemble, les résultats par genre indiquent que la substituabilité des canaux est un phénomène commun, et qu'il ne se limite pas à certains genres spécifiques.

L'analyse est également effectuée au niveau des albums plutôt que des titres. Dans ce cas, le nombre de ventes physiques est régressé économétriquement sur le nombre de ventes numériques (téléchargements). Cette régression n'inclut donc pas le streaming comme variable explicative car les consommations en flux sont toujours enregistrées au niveau du titre dans la base de données GfK. Les estimations par MCO, avec ou sans effets-fixes-produits suggèrent à nouveau un effet positif et significatif des ventes d'albums numériques sur les ventes physiques, ce qui suggère l'existence d'une complémentarité entre les "anciens" et les "nouveaux" canaux. Cependant, comme soulevé concernant l'analyse au niveau des titres, cela reflète probablement la présence de chocs communs non observés qui agissent à la fois sur les ventes physiques et les téléchargements. L'approche par variable instrumentale est adaptée à ce problème : dans un premier temps, il s'agit de régresser les ventes d'albums numériques par les prix des albums numériques et des albums physiques. Cette régression de première étape indique que le prix des albums numériques a un effet négatif et significatif sur la demande d'albums numériques, alors que le prix des albums physiques ne présente pas d'effet statistiquement significatif. En utilisant les deux prix comme instruments pour les ventes numériques, on identifie un effet négatif et significatif des ventes d'albums numériques sur les ventes physiques, ce qui implique qu'il existe un effet de substitution entre les canaux numériques et physiques au niveau des albums.

Au niveau album, comme pour le niveau titre, les analyses sont conduites sur des sous-échantillons, avec les produits les plus populaires et pour différents genres. Comparé à l'analyse au niveau des titres, celle-ci s'appuie sur moins d'observations, et, conséquemment, les estimations ont tendance à être moins précises, mais elles permettent globalement des conclusions comparables. Pour tous les sous-échantillons, à l'exception du Top 50, la première étape estime un effet négatif et statistiquement significatif du prix de l'album numérique sur les ventes d'albums numériques. Les résultats obtenus en deuxième étape sont très comparables et statistiquement significatifs pour les sous-échantillons contenant les Top albums 5000, 1000 et 200. Ce n'est que lorsque l'échantillon se limite aux albums du Top 50 qu'il n'est plus possible d'observer d'effet significatif. Cela peut être attribué au fait que le prix est un instrument faible dans ce cas, comme l'indique sa non-significativité dans la régression de première étape. Pour les genres, on documente un effet de substitution statistiquement significatif pour le genre Pop. Pour les autres genres, des effets négatifs sont obtenus, mais ils ne sont pas significatifs.

Analyse agrégée

Les analyses au niveau des titres et des albums révèlent des résultats intéressants sur la mesure dans laquelle les nouveaux canaux supplantent les anciens. Cependant, elles ne permettent pas de comparer directement la manière dont les ventes numériques et le streaming se répercutent sur les ventes physiques, car les ventes physiques de titres sont pratiquement inexistantes et le streaming d'albums n'est pas observé. Pour permettre une telle comparaison, une analyse à des niveaux de données plus agrégés sont considérés. Au niveau de l'artiste, les ventes d'albums peuvent être converties en « unités équivalentes à des titres » en utilisant le facteur de conversion discuté précédemment de 10 titres par album. Une analyse agrégée n'est pas seulement informative car elle permet de comparer l'impact relatif du canal numérique et du streaming sur les ventes physiques. Elle peut également intégrer l'impact d'éventuels effets de débordement entre les différents titres et albums d'un même artiste. En utilisant une approche similaire, une analyse au niveau du marché français est proposée à la fin de cette section.

Analyse au niveau des artistes

Comme dans l'analyse précédente au niveau des titres et des albums, les régressions MCO et MCO avec effets fixes individuels suggèrent un impact positif du canal numérique et du streaming sur les ventes dans le canal physique. La troisième approche, celle qui s'appuie sur l'estimateur des Doubles MCO est celle qui est retenue ici encore. Lors de la première étape, deux régressions de la demande sont estimées : une pour les ventes numériques (équivalentes aux titres) et une pour les flux. Les

deux demandes peuvent dépendre des prix des produits numériques et des produits physiques (et encore une fois pas du prix des flux, car le prix marginal d'un flux est nul dans le cadre des modèles d'abonnement des plateformes). Les effets de prix estimés dans les deux régressions de première étape ont le signe attendu et sont statistiquement significatifs. La demande pour les produits des artistes sur le canal numérique dépend négativement du prix sur le canal physique, et positivement du prix sur le canal physique. En outre, les flux générés par les artistes dépendent positivement des prix dans les canaux numériques et physiques. La régression de deuxième étape permet de montrer que le canal de vente numérique a un impact négatif et significatif sur le canal de vente physique, similaire à l'effet de substitution précédemment estimé au niveau produit. Par conséquent, même après avoir pris en compte les effets de débordement entre les titres et les albums d'un même artiste, les ventes numériques semblent diminuer les ventes physiques. En revanche, les streams ont un impact positif sur les ventes physiques au niveau de l'artiste. Cela indique un effet de complémentarité entre les canaux de consommation. Une possible explication pour ce résultat est que le streaming de certaines chansons fournit aux utilisateurs de nouvelles informations et les encourage à acheter des produits physiques du même artiste (y compris d'autres titres ou albums que ceux qu'ils ont écoutés en streaming). Comme dans l'analyse au niveau des titres et des albums, une analyse dédiée à différents sous-échantillons est considérée : différents groupes d'artistes les plus populaires et différents genres. La conclusion générale qui peut être tirée des résultats obtenus est les effets identifiés précédemment sont également observés mais les coefficients sont beaucoup moins précis – et parfois même non significatifs. Ils sont présentés plus en détail dans le paragraphe qui suit.

Pour les régressions basées sur différentes définitions des artistes les plus populaires, les coefficients associés aux prix, estimés dans les régressions de première étape, ont le signe attendu et sont généralement significatifs. Les régressions de deuxième étape montrent que les téléchargements ont un impact négatif sur les ventes de produits physiques, alors que le streaming semble, au contraire, les stimuler. Cependant, les estimations obtenues ici sont imprécises, peut-être parce que les instruments de prix ne sont pas suffisamment forts dans ces sous-échantillons beaucoup plus petits.

Quant aux régressions pour différents genres, les effets de prix estimés dans les régressions de première étape ont à nouveau le signe attendu et sont généralement significatifs, ce qui suggère que l'approche des variables instrumentales fonctionne également correctement avec des données agrégées au niveau de l'artiste. Néanmoins, l'impact estimé des nouveaux canaux sur les canaux physiques est parfois estimé de manière imprécise. Dans les cas où des estimations statistiquement significatives sont obtenues, elles sont conformes à celles obtenues avec des régressions au niveau des genres : pour le genre variété, les ventes numériques diminuent les ventes physiques ; pour les genres Variété et Pop, les consommations en flux augmentent les ventes physiques. Notons que cet effet promotionnel du streaming est conforme aux résultats de Kretschmer et Peukert (2020), qui ont également identifié un effet positif du streaming (vidéo) sur les ventes de musique. Le tableau 4 présente ces résultats.

Tableau 4: Résultat des régressions au niveau artiste pour différents genres

Régressions principales						
Variable à expliquer : nombre de ventes physiques (en équivalent titres)						
	Pop	Musique Urbaines	Electro	Rock	Variété Française et Internationale	Musique Classique
Nombre de téléchargements (en équivalent titres)	-0.65 (5.35)	0.64 (2.90)	-188.19 (6455.47)	-6.95 (5.15)	-18.78* (5.91)	6.13 (67.15)
Nombre de streams (en milliers)	19.76** (7.04)	4.44 (2.30)	87.40 (2949.65)	-26.80 (40.58)	49.30** (14.72)	437.85 (3697.86)
Effets fixes semaine-année	Oui	Oui	Oui	Oui	Oui	Oui
Effets fixes artistes	Oui	Oui	Oui	Oui	Oui	Oui
Instruments	Oui	Oui	Oui	Oui	Oui	Oui
Régressions de première étape						
Variable à expliquer : nombre de téléchargements (en équivalent titres)						
Prix moyen du téléchargement (pour l'artiste)	-11.65** (3.89)	-13.78 (23.52)	1.56 (21.09)	-18.43*** (2.83)	-26.04*** (6.01)	-10.85*** (0.91)
Prix moyen du produit physique (pour l'artiste)	14.95** (2.47)	7.35 (3.94)	0.80 (0.96)	1.91 (2.10)	21.00* (7.08)	0.81 (0.43)
Variable à expliquer : nombre de streams (en milliers)						
Prix moyen du téléchargement (pour l'artiste)	8.03*** (2.13)	50.25 (34.85)	4.77 (3.17)	0.88 (1.66)	-0.26 (4.25)	0.20 (0.17)

Prix moyen du produit physique (pour l'artiste)	3.34* (1.53)	8.76 (5.75)	1.76** (0.68)	1.49 (0.79)	18.15* (7.18)	-0.03 (0.16)
Nombre d'observations	161 942	62 155	41 379	106 891	55 608	79 201
Nombre d'artistes	6 773	1 678	1 938	3 717	1 643	3 633

Source : GfK/TSE

Analyse au niveau du marché

Par souci d'exhaustivité l'analyse a été menée au niveau le plus agrégé des données : le marché français dans sa globalité – lorsqu'est testée la dépendance des ventes physiques en équivalents-titres aux téléchargements et aux *streams*. Les différentes approches retenues donnent toutes des résultats non significatifs, et les régressions de première étape semblent faibles à ce niveau d'agrégation. Les résultats sont peu sensibles à des définitions alternatives de calcul des prix moyens utilisés comme instruments.

Conclusion

En utilisant un jeu de données unique qui couvre la quasi-totalité du marché français 2014-2017 de la musique enregistrée, l'étude visait à tester l'existence d'un effet de substitution ou de complémentarité entre les anciens et les nouveaux canaux de distribution. La variation des prix des produits a été utilisée comme instrument pour estimer l'effet causal des consommations ayant lieu sur les nouveaux canaux de consommation sur le canal traditionnel.

Au niveau du produit, c'est-à-dire pour les titres et les albums, les résultats soulignent qu'il existe un effet de substitution entre les nouveaux et les anciens canaux, conformément aux résultats déjà mis en lumière dans la littérature. Au niveau des artistes, les résultats mettent en lumière l'existence d'un effet de substitution entre les canaux numériques et physiques. Cependant, dans le même temps, l'analyse montre qu'il existe un effet de complémentarité entre le streaming et les ventes physiques, qui est principalement porté par les genres Pop et Variété. Cela suggère que les artistes positionnés dans des "segments" spécifiques ont bénéficié de l'introduction du canal de streaming. Cette conclusion complète certains résultats récents de la littérature empirique (par exemple, Kretschmer et Peukert (2020)). Enfin, au niveau du marché, les résultats obtenus ne sont pas concluants.

Les éléments mis en lumière dans cette étude semblent apporter d'une contribution utile à la politique économique. En effet, ils sont cruciaux pour comprendre la façon dont la demande a évolué dans cette industrie – ce qui est intrinsèquement lié aux revenus et pouvoirs de négociation des différents acteurs, artistes, producteurs, distributeurs... mais aussi les plateformes de streaming dont le pouvoir croissant a attiré l'attention, à la fois des universitaires, mais aussi des régulateurs, au niveau national, européen et mondial.

Annexe

Le modèle économétrique

Une approche typique pour mesurer l'impact d'un nouveau canal musical sur un canal existant consiste à régresser économétriquement les ventes du canal existant sur les ventes du nouveau canal. Cela entraîne un problème d'endogénéité classique dans ce type de situation : les ventes d'un titre (ou d'un album ou toutes celles réalisées par un artiste) peuvent être soumises à un même choc commun (un effet inattendu de popularité par exemple), de sorte que toute relation positive entre les ventes sur le canal existant et le nouveau canal peut simplement capturer ce choc commun plutôt qu'un effet de complémentarité entre les deux canaux de vente. Pour aborder cette question, plusieurs articles ont exploité des expériences naturelles, comme le retrait temporaire d'une partie des contenus disponibles sur des plateformes de streaming.²⁰ D'autres articles ont utilisé des données de panel provenant de pays différents et ont introduit des effets fixes par titre et par période pour contrôler la popularité mondiale commune d'un titre ou d'un artiste.²¹ Cependant, ces approches ne sont pas toujours réalisables pour des questions de disponibilité de données. Par exemple, les données de GfK utilisées pour cette étude ne couvrent pas une période où un choc de ce type a eu lieu et ne sont disponibles que pour la France. En outre, les travaux précédents reposent toujours sur certaines hypothèses (par exemple, l'hypothèse selon laquelle l'indisponibilité d'un canal de consommation n'est pas corrélée avec sa popularité dans la première approche, et l'hypothèse selon laquelle les chocs de ventes locaux ne sont pas corrélés entre les pays, conditionnellement aux effets fixes de produits et de périodes). Dans la présente étude, une alternative est proposée : les prix hebdomadaires de chaque produit sont exploités pour mettre en œuvre la méthode des variables instrumentales afin d'identifier l'impact causal des nouveaux canaux sur le canal traditionnel.

De façon formelle, on note q_{it}^c la quantité de produit i vendue ou consommée en flux au temps t sur le canal c . Un produit i peut désigner un titre ou un album. Dans l'analyse au niveau de l'artiste, i fera référence à l'artiste (et les quantités seront le total des unités équivalentes à un titre pour l'ensemble des titres et des albums). Le canal c peut désigner le canal des ventes physiques, le canal des ventes numériques (c'est-à-dire les téléchargements) ou le canal du streaming. Comme suggéré dans d'autres travaux académiques, il est particulièrement pertinent de mesurer l'impact des consommations ayant lieu sur les nouveaux canaux de consommations sur les canaux plus anciens. On peut donc noter c soit O pour les anciens canaux (*old*), soit N pour les nouveaux (*new*). Le modèle de régression linéaire suivant est considéré :

$$q_{it}^O = \alpha_0 + \alpha_1 q_{it}^N + \mu_i + \theta_t + \varepsilon_{it}^O \quad (1)$$

où μ_i correspond à un effet fixe invariant dans le temps pour le produit i , θ_t un effet fixe pour la semaine t et ε_{it}^O est le terme d'erreur. L'intérêt se porte principalement sur le coefficient α_1 . Si $\alpha_1 < 0$, alors on peut conclure que les consommations ayant lieu sur le nouveau canal se substituent aux consommations ayant lieu sur le canal traditionnel. Si, au contraire $\alpha_1 > 0$, alors on peut conclure que les consommations ayant lieu sur le nouveau canal sont complémentaires aux consommations sur le canal ancien.

Dans le cadre de cette étude, ce modèle de régression peut être mis en œuvre comme suit :

- Si les produits se réfèrent à des titres, alors q_{it}^O se réfère à la quantité vendue sur le canal numérique et q_{it}^N se réfère à la quantité de consommations en flux. Les « titres » physiques, qui correspondent aux singles sont ignorés ici.
- Si les produits se réfèrent à des albums, alors q_{it}^O se réfère à la quantité vendue sur le canal physique, et q_{it}^N se réfère à la quantité vendue sur le canal numérique. Les consommations en flux sont ignorées ici car elles correspondent, dans la base de données GfK toujours à des titres et jamais à des albums.
- Enfin, au niveau d'analyse plus agrégé, si les i se réfèrent à des artistes, alors q_{it}^O se réfère à la quantité totale équivalente à un titre de l'artiste vendue sur le canal physique, et q_{it}^N correspond à la fois à la quantité totale vendue par l'artiste sur le canal numérique et la quantité totale de consommations en flux générées par cet artiste.

Cette équation peut être estimée avec l'estimateur des Moindres Carrés Ordinaires (MCO). Cependant, malgré l'introduction d'un grand nombre d'effets fixes, il est possible qu'un problème d'endogénéité subsiste, c'est-à-dire que les variables explicatives soient corrélées au terme d'erreur, ce qui est une violation d'une hypothèse importante assurant la convergence de l'estimateur MCO – et par conséquent la possibilité d'une interprétation causale de la relation mise en lumière par la régression. Il est donc nécessaire d'utiliser la méthode des variables instrumentales. Pour cela, les prix des produits seront utilisés dans des régressions de première étape, et le modèle sera estimé avec l'estimateur des Doubles Moindres Carrés Ordinaires.

Effectivement, pour estimer α_0 et α_1 dans l'équation (1), l'approche la plus simple consisterait à ignorer les effets fixes du produit et du temps (donc $\mu_i = \theta_t = 0$) et à utiliser les MCO pour l'estimation. Cela conduirait toutefois une grave violation d'hypothèse de stricte exogénéité des variables explicatives nécessaire à une estimation non biaisée et précise des coefficients. Effectivement, il y a ici un problème d'endogénéité car les chocs affectant la demande du produit i sur le canal traditionnel (ε_{it}^O) sont probablement fortement corrélés à la demande du même produit sur le canal plus récent (q_{it}^N). L'inclusion des effets fixes pour les produits et les périodes de temps peut atténuer ce problème. Cependant, il est probable qu'il subsiste une forte corrélation entre le terme d'erreur du modèle ε_{it}^O et la variable explicative principale q_{it}^N . Cela peut être le cas si un titre ou un album jouit d'une couverture médiatique importante en France – ce qui affectera les ventes et les streams sur tous les canaux. Ainsi, le problème d'endogénéité n'est pas totalement écarté et il est nécessaire de faire appel à la méthode des variables instrumentales. De bons instruments pour q_{it}^N seraient des variables ayant un pouvoir explicatif sur q_{it}^N mais qui n'influencent pas directement la variable à expliquer du

modèle, q_{it}^O . Les prix des produits sur les différents canaux de vente, c'est à dire p_{it}^O et p_{it}^N , sont de bons candidats ici. Cependant, leur utilisation comme instruments nécessite qu'ils ne soient pas corrélés avec le terme d'erreur, c'est-à-dire qu'ils ne déterminent pas la demande qui constitue la variable à expliquer. Cette hypothèse n'est pas déraisonnable : d'une part, la régression inclut, par ailleurs, un très grand nombre d'effets fixes qui capturent l'hétérogénéité inobservée constante dans le temps ; d'autre part, on peut observer que les prix des produits ont tendance à être concentrés autour d'un nombre limité de valeurs – des valeurs *focales*- tant pour les produits physiques que pour les produits numériques. La variation de prix résiduelle qui peut être observée pour un produit donné a donc tendance à être discrète et n'est pas manifestement due à des chocs soudains et non observés de la demande.