
“How Have Video-on-Demand Platforms Shaped Our Preferences? Endogenous Preferences in a Cultural Market”

Bernat Mallén Alberdi

UBIREA

**Institut de Recerca en Economia
Aplicada Regional i Pública**
UNIVERSITAT DE BARCELONA

WEBSITE: www.ub-irea.com • CONTACT: irea@ub.edu

The logo for the AQR group, featuring a green circle and the letters 'AQR' in a bold, sans-serif font.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • CONTACT: aqr@ub.edu

Universitat de Barcelona

Av. Diagonal, 690 • 08034 Barcelona

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

IREA Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. For that reason, IREA Working Papers may not be reproduced or distributed without the written consent of the author. A revised version may be available directly from the author.

Any opinions expressed here are those of the author(s) and not those of IREA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

Abstract

In this paper, I investigate the effect of the introduction of video-on-demand (VoD) platforms on the preferences for different language versions of movies. By using survey data gathered from 2014 to 2019 in the bilingual region of Catalonia (Spain), I explore whether the exposure to VoD affected the likelihood of preferring the Catalan, Spanish or original versions. I found a negative effect on the preference for Catalan and Spanish versions, although this was not significant in all the specifications. Regarding the impact of VoD exposure on the preference for original versions, I identified a positive, significant and very robust effect in all the specifications. The effect is heterogeneous and varies depending on an individual's language, education level and age. These results prove that people adapt their preferences to what they experience, so the introduction of a new technology into a market (in the case of this paper, the movie market) can rapidly change the preferences of consumers, who accustom, or accommodate, their preferences to the new paradigm.

JEL Classification: L82, O33, Z13

Keywords: Multilingualism, Video-on-demand, Dubbing, Original version, Preferences.

Bernat Mallén Alberdi: AQR-IREA, Universitat de Barcelona. Email: bernatmallen@ub.edu

1. Introduction

This work analyses whether the introduction of video-on-demand (VoD) platforms generated an endogenous change in users' preferences in terms of the different language versions of movies. The empirical setup exploits the launch of Netflix in Spain in late 2015, which was the starting point of the VoD expansion, by using repeated cross-section data about the preferred language version when watching a film, as well as the use of VoD.

There is a vast body of literature on endogenous preferences and habit formation. Bowles (1998) asserted that policies and constitutions do affect preferences and that this is crucial to welfare analysis; indeed, the welfare analysis should take into account that the policy itself can change the preferences of the individuals and thus the surpluses that were considered a priori. Rational choice theory defines how an individual can rationally behave, given his or her [exogenous] preferences, but says little about where those preferences come from (Dietrich & List, 2013). However, most of the studies on endogenous preferences are theoretical and focus on welfare analysis or how economic institutions determine the formation of the preferences. To the best of my knowledge, there has been little research that empirically analyses how preferences change and adapt to new products and paradigms.

This paper also contributes to another emerging strand of the literature on the effects of television on social phenomena. Television influences violent crime (Dahl & DellaVigna, 2009), voting turnout (Gentzkow, 2006), political patterns in the US (DellaVigna & Kaplan, 2007) and international policy (Eisensee & Strömberg, 2007). The study by Micola et al. (2019) is in this regard the closest work to this paper, since it considers the effect of the type of version on English skills attainment. Specifically, the authors examine the influence of television translation techniques on the worldwide distribution of English-speaking skills.

In this paper, the case of the new paradigm in the movie market, VoD platforms, is analysed. VoD platforms have become the main source of movie consumption in the recent years, and experienced a boost due to the pandemic lockdown. Unlike cinemas, VoD

platforms allow users to choose from different language versions, as well as subtitles. Cultural goods, such as movies, deliver their content to the consumers through a particular language; the language of these cultural goods will be a determinant factor in the decision of the consumer (Caminal, 2010). Hence, in this context, the language should be regarded as a relevant product characteristic.

The technological differences between VoD and movie theatres define the provision of each language version, including the original version with subtitles. If we focus on dubbing, VoD platforms face the same problem of movie theatres: the need to incur the dubbing cost to offer a film in more than one language. However, if a one-screen movie theatre wants to offer a movie in a minority language, then it needs to separate the audience by alternating different linguistic versions, and hence the local linguistic minority has to be sizeable. Similarly, in a multiscreen movie complex a minimum demand for a second language is also required in order to allocate an entire screen to a version in that language. If the linguistic minority is very small at each location, then no movie theatre may want to show the movie in the minority language. In contrast, because of its centralized nature, the VoD platform can still profit from serving the linguistic minority. Such a phenomenon was theorized by Chris Anderson in his book *Long Tail* (2013). According to the long tail theory:

For many product categories, smart technology is transforming mass markets into millions of small niche markets. Although each of these niche markets may be small, when all the various niches are combined, the volume of business is actually greater than the traditional mass market successes. Thus, the great commercial opportunity of the future isn't catering to the "short head" of the demand curve where multiple copies of the same product are sold. Instead, the real opportunity to move forward lies in serving the "long tail" of the same curve – the millions of various niche markets. (Anderson, C. 2013: p. 2)

Quan and Williams (2018) provided a recent empirical example of enhanced product variety thanks to the online market. Hence, in our case the “smart” technology of VoD could allow for increasing the linguistic diversity, especially for those languages that did not reach critical mass when movies were more often consumed in cinemas. Regarding subtitled original versions, the cost of them is negligible compared to dubbing, so technological change should imply a great increase in this type of version.

In reality, what we observe is not precisely that. On the one hand, the availability of original versions has indeed rocketed and virtually all the movies on VoD platforms can be watched in their original versions with many options for subtitles. On the other hand, minority languages are even less present in the VoD compared to the movie theatres. The reason for that might be that although the long tail effect works in favour of the provision of minority languages in VoD, the bundling effect makes the local markets and small niches irrelevant: since the user does not buy each film but the whole subscription (bundling), the globalized VoD platforms can overlook the long tail. Nonetheless, the reasons behind the actual supply of different versions in VoD is beyond the scope of this paper, which treats their provision of language versions as exogenous.

The sudden change in the availability of language versions in the movie market represents the ideal setup for exploring the endogeneity of users’ preferences. In monolingual markets in which dubbing was (and is) the norm, such as Spain, France, Italy or Germany, VoD platforms have dramatically increased the exposure to original versions with subtitles (in their own language). In bilingual markets with minority languages, in turn, the increase in diversity is not comparable; while the availability of original versions with subtitles have clearly increased as well, access to local-language versions has been reduced. Moreover, as usually happens with innovations, there was no regulation on this matter until very recently in many countries.

In the European Union (EU), in late 2018 the European Parliament approved the first directive on the VoD platforms, which should be steadily applied by each member state in their respective parliaments. In the case of Spain, for instance, the EU directive was implemented in 2022 through the “Ley General de Comunicación Audiovisual” (General Law of Audiovisual Communication) that established quotas for production in the different official languages of Spain,¹ among other obligations. Overall, VoD platforms have allowed viewers to consume movies in foreign-language and subtitled versions, but may have reduced the diversity (at least in the first unregulated years) regarding local languages.

This case study considers the bilingual region of Catalonia in Spain. As explained, in bilingual regions the impact of VoD platforms have been twofold. Hence, this makes the study of this particular case more complex and richer. The main hypothesis to be tested in this paper is that a greater exposure to a language version increases the preference for that version. Consumers adapt their tastes to what is offered in the market. More specifically, I expect to find an increase in the preference for the original version and a decrease in that for the Catalan version. The effect on the preference for the Spanish versions is unclear, since the Spanish versions are still always present but, unlike the movie theatres, they are non-exclusive: the consumer cannot choose the language version in movie theatres, as it offered in one version or another. Instead, VoD can offer the full availability of two versions for all the films in the catalogue.

Furthermore, the competition between versions must be understood in two categories: the horizontal and vertical differentiations. The vertical differentiation refers to the quality difference between the two dimensions: dubbing vs subtitled original versions. If the exposure to VoD increases the preference for original versions, it should do so at the expense of the preference for dubbing, either Catalan or Spanish, or both. The horizontal

¹ Spanish is the official language of Spain among other regionally official languages: Catalan, Basque and Galician.

differentiation refers to the language difference within the same dimension, the dubbing. Hence, if the exposure to VoD directly decreases the preference for Catalan versions, it should do so in favour of the Spanish version as long as the user does not switch dimension. Therefore, the preference for Spanish might go either way, depending on which effect is stronger.

To analyse the effect of VoD platforms, I use survey data from 2014 to 2019 containing information about language version preferences and the use of VoD platforms, among other relevant characteristics. By using probability models with different specifications and checks, I find a strong positive effect of VoD on the preference for original versions. The effect on both Spanish and Catalan is negative but only significant for some specifications; the heterogeneous analysis shows how the effect greatly depends on individual characteristics.

The rest of the paper is structured as follows. In the next section, I provide some background relating to the audiovisual market in Catalonia. After this, in Section 3, the data used for the empirical analysis are described. In Section 4, the empirical method is presented. In Section 5 I display the results and discuss them, and finally, in Section 6, I conclude this paper with some final comments.

2. Background

Before the introduction of VoD platforms in Spain, there were two main sources of cinema consumption by most of the population: TV and movie theatres. In TV, following the definitive change from analogue to digital terrestrial television in 2010, although by default the language movies are shown in is the language of the broadcaster, the original version (if different) can be chosen, as well as adding subtitles. The reality is, though, that this is not a common practice. Hence, we can take the percentages of TV audiences by language as a good indicator of the market size of each language in television. In Catalonia, the share of annual time watching TV in Catalan was approximately 20% (and thus approximately 80% for Spanish) from 2017 to 2020 (Statistical Institute of Catalonia [IDESCAT]).

On the other hand, in the movie theatres of Catalonia the Catalan language was less present, only representing approximately 3%–4% of the total number of screenings (IDESCAT). Since in Spain dubbing is the norm, most of the rest of the screenings were in Spanish, either original or dubbed, although in the last years the presence of original versions has increased. There are no official data on the percentage of screenings in the original versions (with subtitles) in past years, but they represented approximately 11.5% of the total number of screenings in Catalonia in 2020 and 2021 (Mallén, 2023).²

As explained in the introduction, the Catalan language has a much lower presence on the globalized VoD platforms than in the other sources of audiovisual consumptions previously described. In May 2020, the Pro-Language Platform (in Catalan, *Plataforma per la Llengua*), a non-governmental organization concerned with the situation of the Catalan language, issued the first report analysing the availability of the Catalan versions of movies on Netflix (the most consumed VoD platform). In this report, they showed that only 0.02%

² This comes from own-collected data using webscraping techniques. I tracked all the screenings in movie theatres in Catalonia from June 2020 to May 2021.

of the movies in the Netflix catalogue had a Catalan-language version available, even though 16% already had a version in Catalan (so Netflix should not need to incur the cost of dubbing or subtitling³). One year later, the Audiovisual Council of Catalonia (*Consell Audiovisual de Catalunya*), an independent regulatory authority, published a similar study showing that the percentage of movies with a Catalan version available on Netflix was only 0.5%. It is important to note that the percentage of screenings in movie theatres and the percentage of films with the version available on VoD platforms are not the same indicator. For instance, during 2020–2021, while screenings in Catalan made up 4.7% of the total, this meant that almost 18% of the movies were also available in Catalan, although this version was exclusive, unlike VoD.

While the availability of Catalan-language versions has decreased in this new source of audiovisual consumption, VoD platforms, the availability of original versions has sharply increased. Actually, it is possible to watch all the movies in their original version (with subtitles, if needed).

To sum up, the introduction of Netflix decreased the exposure to the Catalan language from 20% in TV and 3%–4% of the screenings (or 18% of the films) in movie theatres to 0.2%–0.5% in the online catalogue, while it increased the availability of original versions from 0% in TV and approximately 10% in movie theatres to 100% in the online catalogue. The exposure to Spanish did not experience any considerable change: this language version is still fully available in the online catalogue.

³ The dubbed versions are freely available upon request to Catalan public television, which is the owner.

3. Data and Descriptive Statistics

The data used in this paper come from the Òmnibus surveys that are periodically carried out by the Opinion Studies Center (in Catalan, *Centre d'Estudis d'Opinió*), for the years 2014, 2015, 2018 and 2019. These surveys asked individuals a wide set of questions including their language preferences for different cultural goods (books, cinema, theatre and music). Moreover, they include information on several sociodemographic and economic characteristics such as age, gender, language (mother tongue, language of identification and language of use), education, country of birth, self-reported income, social class, labour status and the size of the municipality of residence, among others.

In addition to that, the waves after the introduction of Netflix (2018 and 2019⁴) asked about the use of VoD platforms. The number of observations per year can be seen in Table 1.

Table 1

Observations Per Year

Years	Observations
2014	4,800
2015	1,050
2018	1,200
2019	1,200
Total	8,250

The sample used for the analysis is restricted to those individuals who did answer all the relevant questions: preference for language in cinema, mother tongue, labour status and education level. Thus, “do not know” or “do not answer” are treated as missing values, and therefore excluded (205 observations dropped). The final sample consists of 8,045 respondents. Table 2 displays the summary statistics of the estimation sample.

Table 2

Summary Statistics Control Variables

⁴ Netflix was introduced in October 2015 and the 2015 survey was carried out in March.

Variable	Obs	Mean	Std. dev.
VoD	8,045	0.060	0.238
Mother tongue			
Catalan	8,045	0.412	0.492
Spanish	8,045	0.499	0.500
Catalan and Spanish	8,045	0.040	0.197
Income			
Up to 1000€	8,045	0.156	0.363
Between 1,000€ and 2,000€	8,045	0.354	0.478
Between 2,000€ and 3,000€	8,045	0.219	0.413
Between 3,000€ and 4,000€	8,045	0.091	0.288
Between 4,000€ and 5,000€	8,045	0.0286	0.167
More than 5,000€	8,045	0.023	0.151
Does not know/does not answer	8,045	0.128	0.334
Class			
Low	8,045	0.106	0.308
Middle-low	8,045	0.288	0.453
Middle	8,045	0.517	0.500
Middle-high	8,045	0.062	0.241
High	8,045	0.005	0.069
Does not know/does not answer	8,045	0.022	0.146
Age	8,045	49.45	17.105
Sex	8,045	0.513	0.500
Place of birth			
Catalonia	8,045	0.680	0.467
Rest of Spain	8,045	0.234	0.424
EU	8,045	0.042	0.201
Non-EU	8,045	0.044	0.205
Municipality size			
Below 50,000 inhabitants	8,045	0.549	0.498
Between 50,000 and 150,000	8,045	0.214	0.410
Between 150,000 and 1M	8,045	0.084	0.278
More than 1M	8,045	0.153	0.360
Labour status			
Works	8,045	0.479	0.500
Does not work	8,045	0.500	0.500
Temporarily on leave	8,045	0.021	0.143
Education			
Primary education or less	8,045	0.118	0.323
Secondary education	8,045	0.227	0.419
Professional course	8,045	0.378	0.485
Tertiary education (university degree, master's or PhD)	8,045	0.277	0.448

It is important to note that the Òmnibus surveys only include people with Spanish nationality; that is why we observe low percentages of individuals born outside of Spain (both

EU and non-EU), relative to official statistics. In addition, due to this restriction the percentages of Catalan-speaking respondents are higher than the average when non-nationals are included. According to the Survey on Language Uses of the Population (*Enquesta d'Usos Lingüístics de la Població*) of 2018, for 31.5% of the population of Catalonia (including non-nationals), Catalan is their mother tongue, while for 52.7% it is Spanish; this is due to non-nationals from Spanish-speaking countries.

Only 6% of individuals in the sample are VoD users. This is normal since all the observations for 2014 and 2015 are necessarily non-users. In order to obtain a better idea of the exposure to VoD, Table 3 reports the proportion of users in 2018 and 2019. Moreover, the dependent variables relating to preferences are reported. In Table 4 the preferences on movies are reported by year. We can observe some changes during this period: a decrease in the preference for Catalan versions and an increase in the preference for Spanish versions. The empirical strategy, explained in the next section, allows for a *ceteris paribus* estimation of these effects.

Table 3

Summary Statistics Dependent and Independent Variables

Variable	Obs	Mean	Std. Dev.
Preference for movies			
Catalan	8,045	0.236	0.425
Spanish	8,045	0.419	0.493
Original version	8,045	0.153	0.360
Indifferent	8,045	0.187	0.390
Preference for books			
Catalan	8,045	0.338	0.473
Spanish	8,045	0.384	0.486
Original version	8,045	0.071	0.257
Indifferent	8,045	0.194	0.395
VoD			
2018	1,175	0.191	0.393
2019	1,175	0.221	0.415

Table 4

Movies Preference by Year

	2014	2015	2018	2019
Catalan	0.283	0.224	0.162	0.134
Spanish	0.376	0.454	0.476	0.501
Original Version	0.134	0.170	0.193	0.174
Indifferent	0.203	0.152	0.163	0.180
VoD	0	0	0.191	0.221
N	4,664	1,031	1,175	1,175

Means reported for the selected sample.

4. Empirical Methodology

This section describes the empirical approach followed to analyse the effect of the exposure to VoD platforms on the changes in the language version preferences. The identification strategy exploits the irruption of Netflix in the late 2015 in Spain as a game changer in the movie market; from this point on, consumers could be exposed to another growing source of movies consumption that had a different provision of versions than the previous ones, which were the cinema and the TV. Thus, while the individuals of the sample for the 2014 and 2015 surveys did not have the chance to be exposed to VoD, those in 2018 and 2019 did.

The main econometric specification is the following:

$$P(CP_i = j) = \alpha + \beta VoD_i + \gamma' X_i + \phi_t + \varepsilon_i \quad (1)$$

Here, cinema preference (CP_i) represents the outcome under investigation of individual i for version j . I use a dummy variable for “Catalan”, “Spanish”, “Original version” and “Indifferent”, respectively, in separate models (one regression for each dependent variable). VoD_i is the independent variable that takes a value of 1 if the individual i is a user, and 0 otherwise (note that this dummy takes a value of 0 for all respondents in the waves of 2014 and 2015). X_i is a vector of control variables: dummies for mother tongue or first language, the polynomial of age, sex, place of birth (Catalonia, rest of Spain, rest of EU and non-EU), size of the municipality, labour status and education level. The term ϕ_t contains year dummies, capturing the evolution of preferences towards the cinema over time. I start with a baseline model with only the exogenous variables and then the other variables are added.

There are two potential issues with the exogeneity of the VoD_i dummy. The first is the omitted variables that were correlated with both VoD and the preference. The year dummies are crucial to avoid a bias due to changes in preferences over time, that could cause the beta to capture some spurious correlation. On the other hand, the vector of control

variables contains all the relevant socioeconomic characteristics, as well as cultural traits, so as to avoid this kind of bias. Hence, the reliability of the beta is based on the assumption that both X_i and ϕ_t make the ε_i uncorrelated with the VoD dummy.

The second potential issue is the reverse causality: that is, the preference in relation to the language version of movies could somehow be driving the decision to be a VoD user. There is no straight way to know whether this occurs, but it is likely to be the case (at least for some individuals).

In order to make sure that the coefficient is capturing the causal effect of the VoD exposure on language preferences, I perform a placebo test with another cultural good that is not related to the cinema: books. If the only relevant causal chain is that the exposure to VoD affects language preferences for that specific good (and it does not capture unobserved factors or reverse causality), exposure to VoD should be unrelated to language preferences towards other goods that have nothing to do with VoD. Conceptually, the placebo test works when the sample used as the placebo is not affected by the intervention, in this case the VoD exposure. Here, I am running a placebo test with the same sample but a different outcome. Hence, the crucial point is that if there was some sort of language preference change that was spuriously correlated with VoD exposure, it should also affect the language preference for books. Therefore, finding a significant relationship between VoD exposure and the language preferences in terms of books would be suggestive that the β coefficient from the main regression is likely to be picking up unobserved factors or reverse causality, invalidating the causal interpretation of the results.

Indeed, the main reason why I pick the language preference for books as the placebo, and not theatre or music preference (for which I also have information), is that while theatre and music are cultural goods that share features with movies, books are a sufficiently different cultural good. At least, it is not evident that VoD exposure could affect such preferences. Moreover, there is an additional reason why I prefer to use books as the placebo.

All the surveys first asked about the language preference for books, followed by cinema, theatre and finally music. Hence, the answer relating to books is less likely to have been conditioned by the answer for cinema, which was asked afterwards. In the case of theatre and music, since these questions were asked immediately after the cinema preference, they could have been conditioned by the previous answer.

As an additional check to understand whether the coefficient of interest can be plausibly interpreted as the causal effect of VoD use on the endogenous change in language preferences, I also perform an instrumental variable (IV) approach exploiting the information about self-reported internet quality at home, on a scale from 0 (no internet) to 10 (excellent connection). Such a question was not asked in the surveys for 2014 and 2015. Therefore, this variable is equal to 0 for observations from the first two waves. However, this is in line with the fact that the VoD dummy varies only for individuals included in the 2018 and 2019 waves, for which the endogeneity issue could be relevant. In order to be a valid instrument, the variable capturing internet quality must satisfy two conditions. On the one hand, it should be a strong determinant of the use of VoD platforms (the relevance condition), which is a testable hypothesis. On the other, it should not directly affect the preferences towards the languages of movies. In order to test this hypothesis, I use dummies for internet quality to run the overidentification test.

The first stage is:

$$VoD_i = \omega + \mu IntQ_i + \delta' X_i + \theta_t + u_i \quad (2)$$

In which $IntQ$ is the internet quality measured from 0 to 10. I expect μ to be positive and significant, since in order to be a VoD user having a good internet connection at home is a must. In many cases, connection quality is determined by the infrastructure, which does not depend on individual decisions or characteristics. If this is the case, the instrument will not only be relevant but also exogenous, satisfying the exogeneity condition.

However, one may argue that the quality of internet at home could reflect other individual characteristics that might directly affect language preferences, such as social status or cultural level. Nonetheless, we control for several demographic and socioeconomic variables that are likely to pick up these unobservable factors, meaning that the variable capturing internet quality can be plausibly considered conditionally exogenous with respect to language preferences, thus satisfying the exogeneity condition.

Additionally, in order to test for the validity of the instrument, I estimate reduced-form equations for the four cultural goods (cinema, books, theatre and music), replacing VoD with the internet quality. Finding no reduced-form effects of internet quality on language preferences for other cultural goods than movies would be a strong argument in favour of the validity of the instrument.

In addition, I estimate a model in which the VoD effect is interacted with time dummies. This model is repeated for each cultural good and preference and shown graphically. The main insight of this alternative way of presenting the results is to show that the trends persist in 2018 and 2019 only for the cinema preference.

4.1. Robustness Checks

In order to further test the stability of the results, I perform several robustness checks. First, I exclude those with a native language that is neither Catalan, nor Spanish (nor both). These individuals are grouped in the category “others”, which is a very mixed category of all other languages spoken in Catalonia, from Occitan to Arabic. This is not a homogeneous group, which makes the interpretation of the coefficient unclear. For this reason, I repeat the main estimation after excluding them from the sample. I also applied another restriction of the sample similar to this one: the inclusion of only those born in Spain.

Another robustness check consists of excluding from the sample the two last waves of surveys made in 2014. There is a concern that 2014 is overweighed in the estimation sample since the survey was carried out three times in that year and therefore, I take more

observations (see Table 1). This is due to the fact that for the subsequent years, 2015, 2018 and 2019, I am using only one wave for each one, while for 2014 there were three waves, 2014-1, 2014-2 and 2014-3; hence, I exclude 2014-2 and 2014-3, obtaining a more balanced sample in terms of years. Finally, I add some subjective controls: the dummified self-reported class and income.⁵ As an additional robustness check, I include in the Appendix the average marginal effects of the main results using non-linear Logit and Probit models.

4.2. Heterogeneous Effects

After presenting the results of the placebo tests and the robustness checks, which provide evidence in favour of the interpretation of the exposure variable as “conditionally exogenous”, I provide additional results regarding the heterogeneous effects of the VoD exposure. Since this paper analyses the endogeneity of the language preferences, the effects may differ depending on the linguistic group (by mother tongue). As reported in Table 2, there are two main linguistic groups in Catalonia, Catalan speakers and Spanish speakers, as well as a third mixed group of individuals who have both languages as their first one, which is smaller in size. There are other minorities such as Occitan speakers or migrants who speak foreign languages, but due to the restriction of the survey to individuals with Spanish nationality, they are quantitatively too few in the sample (and too diverse) to consider them an additional separate group.

Thus, in this last part of the econometric analysis I will look at the effect of the VoD exposure to the language preferences of the three linguistic groups according to their mother tongue or first language. Finally, I also analyse whether the effect of VoD exposure is heterogeneous according to the other sociodemographic characteristics of the respondents, namely gender, education level and age group.

In order to do so, I use interactions of the independent variable *VoD* with the classifying variables first language (four categories), sex (two categories), education level (four

⁵ Including the “Do not know” category as a dummy.

categories) and age group (three categories). To assess the significance of the difference between the coefficients, an equality test is undertaken for each one.

5. Results

The main results are displayed in Table 5, which reports the estimates of the main coefficient of interest from Equation (1) for each of the possible outcomes obtained from different specifications (see the full results in Table A1.2 of the Appendix). In the first row, we can observe the basic regression with no controls (except year dummies). The relationship between the independent variable VoD and the version preferences are significant in all cases: positive for the original version and negative for the others. However, once the individual exogenous characteristics are taken into account, the effect on the preference for Catalan and on the probability of being indifferent with respect to the language versions of movies vanishes, while the effect on the Spanish and original versions remain stable in terms of sign and are highly significant. The introduction of the rest of the controls does not change the significance of the estimates, although they decrease in magnitude, especially in terms of preference for original versions. The exposure to VoD platforms reduces the probability of preferring Spanish versions by 6.26 percentage points (p.p.), while it increases the likelihood of a preference for the original version by 7.86 p.p. The result regarding preferences for original versions is in line with the initial hypothesis. Moreover, the a priori ambiguous effect on the preference for Spanish versions appears to be negative: the increase in variety has led to a switch from the Spanish to the original versions for some users. Nonetheless, the lower provision of Catalan versions has not led to a significant decrease of such a preference. I further discuss this specific result in Section 5.2.

In addition to the main evidence, Table A1.2 in the Appendix shows the estimated coefficients for control variables, which are worth commenting on. The correlation of the first language and the language preferences in columns (1) and (2) is in line with the

expectations; since the reference category is Catalan speakers, we observe that Spanish speakers prefer Spanish versions to Catalan ones to a great extent, and the same happens to a lower extent for individuals who have both Catalan and Spanish as their first languages (mixed use in the family). What is somehow unexpected is that Spanish speakers are less inclined to prefer the original versions than Catalan speakers. One potential reason could be that, since the availability of Catalan versions is low in cinemas and TVs compared to Spanish versions (not as low as on Netflix and VoD platforms), Catalan speakers have become less accustomed to versions shown in their own language and switch more easily to original versions. It is also worth commenting that individuals who have both Catalan and Spanish as their first language are significantly more indifferent in terms of preferences, which is an expected result since the emotional attachment does not incline the balance towards any of the languages.

The effect of age⁶ is quadratic, and, surprisingly, women are less inclined to Catalan versions while they prefer Spanish versions; women are also more indifferent than men. As expected, being born outside Catalonia is negatively correlated with the preference for Catalan versions, but only those born in the rest of Spain have a greater preference for Spanish versions. Foreigners from other EU countries or outside the EU show a greater predilection for original versions. In order to understand the conditional correlation of the municipality size with the dependent variables, it is important to distinguish three main levels with different demolinguistic landscapes: small municipalities are present in inner Catalonia, in which most of the inhabitants are Catalan speakers; medium-sized municipalities, with more than 50,001, are mostly present in the metropolitan area of Barcelona, and most of the inhabitants come from other Spanish-speaking regions, or are second- or third- generation internal migrants; lastly, Barcelona, the capital and largest city of Catalonia (1.6M inhabitants), is a cosmopolitan metropolis in which Catalan and Spanish are present along

⁶ I also attempted the model with a polynomial of three, but the cubic term was not significant.

with other languages. Hence, the results indicate that being in a medium-sized municipality decreases the preference for Catalan while increasing the preference for Spanish, with little or no effect on the preferences for original versions and, if anything, makes an individual less indifferent; living in Barcelona decreases both the predilection for Catalan and Spanish versions while strongly increasing the preference for the original version. People who do not work have a greater preference for Spanish versions at the expense of Catalan ones, while they do not show a difference in terms of the original version.

Finally, as the education level increases, the predilection for the Catalan versions and especially for the original versions increases at the expense of Spanish versions, whose preference is strongly reduced.

Table 5

Main Results

	(1)	(2)	(3)	(4)
Controls	Catalan	Spanish	Original Version	Indifferent
No	-0.0520*** (0.016)	-0.0779*** (0.025)	0.176*** (0.023)	-0.0425** (0.018)
Only exogenous	-0.0228 (0.016)	-0.0905*** (0.024)	0.113*** (0.023)	0.00426 (0.018)
All	-0.0204 (0.016)	-0.0626*** (0.024)	0.0786*** (0.021)	0.00754 (0.018)
N	8045			

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

Year dummies included. Coefficient beta of the variable VoD

As explained in Section 4, a placebo test is performed by looking at the effect of the VoD exposure to the language preferences in books. The results reported in Table 6 indicate that the point estimates of the placebo coefficients are generally very close to zero and are not statistically significant for any outcome, which speaks in favour of the validity of the approach followed.

Table 6

Placebo Test. Preferences for VoD and Books

	(1)	(2)	(3)	(4)
Books	Catalan	Spanish	Original Version	Indifferent
<i>VoD</i>	0.000635 (0.019)	-0.0144 (0.023)	0.0164 (0.015)	0.00439 (0.020)
N	8045			

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01
Dependent variable is the language preference in books

As also explained in Section 4, in order to further investigate whether the main results are potentially affected by any endogeneity bias, I use the connection quality at home to instrument the independent variable *VoD*. Table 7 reports the coefficients from the first stage (lower panel) and for the structural equation for each possible outcome. Note that the N is now 8,043 because I had to exclude two individuals who did not answer the question regarding “internet quality”.

Table 7

IV Using Connection Quality

	(1)	(2)	(3)	(4)
	Catalan	Spanish	Original Version	Indifferent
<i>VoD</i>	-0.128 (0.098)	-0.271** (0.137)	0.215** (0.091)	0.194 (0.120)
First Stage				
<i>IntQ</i>			0.0208*** (0.002)	
N	8043			

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01
IV with full specification

The first stage proves that internet quality is indeed a relevant instrument, since it is strongly correlated with *VoD*. The coefficients in columns (1) to (4) lead us to the same conclusion, that the effect of *VoD* is negative and significant for the preference for Spanish versions, but positive and significant for the preference for original versions. In column (1) we can see that although negative, the coefficient is still non-significant. Nonetheless, the magnitude of the coefficients is much higher in the IV estimation. This tells us that the coefficients found in Table 5 might be biased towards zero and are thus a lower bound of

the actual effect. Another possible explanation for the higher coefficient is that the IV approach captures the effect among those who are induced to use VoD because they have a better quality of internet (the compliers), not among all the treated.

Table 8 reports the estimates from the reduced forms with all the cultural goods, to assess the validity of the instrument. Note that the coefficients are only significant for the cultural good under investigation, Cinema, in the second and third column (which is in line with the main results and with the IV estimates). However, no clear relationship is detected between internet quality and preferences towards other cultural goods.⁷ This evidence is again suggestive that the main results are capturing a causal relationship between VoD exposure and preferences towards movies' language.

Table 8

Test of the Validity of the Instrument "Internet Quality"

	(1)	(2)	(3)	(4)
	Catalan	Spanish	Original Version	Indifferent
<i>IntQ - Cinema</i>	-0.003 (0.002)	-0.006** (0.003)	0.004** (0.002)	0.004 (0.003)
<i>IntQ - Books</i>	0.001 (0.002)	-0.002 (0.003)	0.001 (0.001)	0.002 (0.002)
<i>IntQ - Theatre</i>	-0.000 (0.002)	-0.002 (0.003)	0.002 (0.002)	0.003 (0.003)
<i>IntQ - Music</i>	0.003* (0.002)	-0.004 (0.003)	0.003 (0.003)	-0.002 (0.003)
N	8043			

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

Reduced form IV

In Table 9 I also show an additional estimation in which I use dummies for internet quality as instruments, which provides an overidentified model that enables testing for overidentification. The corresponding null hypothesis is not rejected, which is indicative that the hypothesis of excludability of the instrument is satisfied.

⁷ The coefficient on the preference for music in Catalan, first column and last row, is only weakly significant at the 10% level.

Table 9*Over-Identification Test with Dummified IntQ*

	(1)	(2)	(3)	(4)
	Catalan	Spanish	Original Version	Indifferent
<i>V_{oD}</i>	-0.103 (0.088)	-0.252** (0.120)	0.207*** (0.076)	0.160 (0.107)
Over identification test (p-value)	0.609	0.091	0.455	0.847
N	8043			

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

Over-identified IV with full specification. The instruments are the dummified variable *IntQ*. Wooldridge's score test of overidentifying restrictions, which is robust to heteroscedasticity.

In Figures A1, A2 and A3 in the Appendix, the predictive margins during the period analysed are shown for the four cultural goods (cinema, books, theatre and music), as explained in the empirical methodology section. The graphs are not very conclusive; we only observe a significant difference in both 2018 and 2019 in the case of the preference for the original version in cinema, which is still strong evidence in favour of this particular outcome.

5.1. Robustness Checks

The results of the robustness checks explained in Section 4.1 are reported in Table 10.

Table 10*Robustness Checks*

	(1)	(2)	(3)	(4)
	Catalan	Spanish	Original Version	Indifferent
<hr/> Excluding L1 Others <hr/>				
<i>VoD</i>	-0.0293*	-0.0378	0.0662***	-0.00375
	(0.017)	(0.025)	(0.022)	(0.019)
N	7658			
<hr/> Excluding 2014-2 and 2014-3 <hr/>				
<i>VoD</i>	-0.0151	-0.0495**	0.0674***	0.00225
	(0.016)	(0.024)	(0.021)	(0.018)
N	5130			
<hr/> Excluding Born outside Spain <hr/>				
<i>VoD</i>	-0.0320*	-0.0292	0.0493**	0.0060
	(0.0188)	(0.0267)	(0.0233)	(0.0212)
N	7353			
<hr/> Additional controls: self-reported class and income <hr/>				
<i>VoD</i>	-0.0217	-0.0618***	0.0783***	0.00804
	(0.0156)	(0.0237)	(0.0210)	(0.0183)
N	8045			

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01
Full specification applied.

The exclusion of the speakers of non-official languages slightly changed the coefficients. On the one hand, the effect of *VoD* on the preference for Catalan versions is now weakly significant and negative. On the other hand, the coefficient of the second column is no longer significant, meaning that the exposure to VoD might not have any effect on the preference for Spanish versions. However, the coefficient in column (3), the effect of *VoD* on the preference for original versions, is still very significant and positive. This stresses the importance that the main expected effect on original version is very stable. I provide further explanations and interpretations of these differences in Section 5.2, when analysing the heterogeneous effects by first language.

The second robustness check consisted on an even stricter restriction. As we can see, excluding the speakers of non-official languages led to a sample of 7,658 individuals, and the exclusion of the two last waves of surveys of 2014 (2014-2 and 2014-3) leads to a sample of 5,130 individuals. Nonetheless, the results do not vary much from Table 5 and the

coefficients in columns (2) and (3) are still significant. Hence, the main results are not driven by an overrepresentation of individuals interviewed in 2014.

To further check the results, those born outside Spain were excluded. As explained in Section 4.1, this sample restriction works similarly to the first robustness check and thus, the outcomes are also very similar to the first robustness check. This is not surprising since the individuals excluded in both restrictions might be the same in a large proportion: those with a foreign native language are likely to be born outside of Spain. In addition, we do not observe significant changes compared to the main results in Table 5 when we add additional controls such as self-reported class and income.

Finally, in the Appendix the average marginal effects of the main results using non-linear Logit and Probit models are reported in Table A2. The effects are robust. As in two of the previous checks, the effect of VoD on the Catalan preference is weakly significant. Still, we cannot find a consistent negative effect on Catalan, as predicted in the hypothesis, but this points out that indeed VoD exposure might be affecting such preference somehow. This is clearly seen in the next section.

5.2. Heterogeneous Effects

The next table (Table 11) reports the results of the heterogeneous effects of the VoD exposure by linguistic group, gender, education level and age group. As explained in Section 4.2, I use interactions of the explanatory variable *VoD* with the classifying variables and report the p-value of the equality tests.

Table 11

Heterogeneous Effects by Linguistic Group, Gender, Education Level and Age Group

	(1)	(2)	(3)	(4)
	Catalan	Spanish	Original Version	Indifferent
<i>VoD. L1 Catalan</i>	-0.100*** (0.036)	-0.0171 (0.037)	0.0796** (0.037)	0.0330 (0.031)
<i>VoD. L1 Spanish</i>	0.0166 (0.014)	-0.0453 (0.033)	0.0713*** (0.027)	-0.0393* (0.022)
<i>VoD. L1 Catalan-Spanish</i>	-0.0727** (0.030)	-0.0161 (0.108)	-0.000618 (0.101)	0.0440 (0.094)
<i>VoD. L1 Others</i>	0.0511 (0.036)	-0.264*** (0.061)	0.132** (0.060)	0.118** (0.054)
Equality test (p-value)	0.000	0.004	0.6856	0.0176
<i>VoD. Male</i>	-0.0175 (0.020)	-0.0469 (0.032)	0.0806*** (0.028)	-0.0130 (0.022)
<i>VoD. Female</i>	-0.0240 (0.021)	-0.0813** (0.032)	0.0762** (0.030)	0.0320 (0.027)
Equality test (p-value)	0.814	0.421	0.909	0.170
<i>VoD. Primary education or less</i>	0.0297 (0.038)	-0.00811 (0.070)	-0.0173 (0.043)	0.0129 (0.057)
<i>VoD. Secondary education</i>	-0.0416 (0.030)	-0.00191 (0.057)	0.0884* (0.050)	-0.0366 (0.034)
<i>VoD. Professional course</i>	-0.00962 (0.025)	-0.0491 (0.036)	0.0370 (0.031)	0.0285 (0.029)
<i>VoD. Tertiary education</i>	-0.0388 (0.024)	-0.117*** (0.037)	0.150*** (0.038)	-0.00104 (0.028)
Equality test (p-value)	0.371	0.232	0.018	0.497
<i>VoD. 18-39 years old</i>	-0.00797 (0.018)	-0.102*** (0.032)	0.0994*** (0.029)	0.0142 (0.021)
<i>VoD. 40-64 years old</i>	-0.0354 (0.024)	-0.0115 (0.034)	0.0652** (0.030)	-0.0112 (0.028)
<i>VoD. 65 or older</i>	-0.0333 (0.100)	-0.0608 (0.097)	-0.0703 (0.055)	0.117 (0.118)
Equality test (p-value)	0.621	0.118	0.022	0.477
N	8045			

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

Full specification applied

The p-value of the equality tests show which coefficients are significantly different depending on the classifying variable, that is, that there are heterogeneous effects on this regard. The first language seems to matter when it comes to the effect of the VoD exposure on the preference for Catalan and Spanish versions, columns (1) and (2), as well as the

indifference in column (3) to a lesser extent (the p-value is not as low). The high p-values of the equality tests performed on the impact of VoD by gender prove that there are no differential effects depending on whether the individuals are males or females. On the other side, we observe low p-values and hence significant differences between the coefficients in column (3) for the two last classifying variables: education level and age group.

These results are quite interesting, since they enable distinguishing the two dimensions of differentiation: the horizontal differentiation, between languages (Catalan or Spanish), and the vertical differentiation, between types of versions (dubbed, including both Catalan or Spanish, and original versions). Thus, the exposure to VoD generates a switch from dubbed versions to original ones; such an effect is stronger (or existent) for younger and more educated individuals. Regarding the horizontal dimension, characterized in columns (1) and (2), we can observe how it greatly depends on the first language. VoD does not decrease the preference for Catalan for Spanish speakers and foreign-language speakers, probably because they did not prefer these versions previously. The decrease is greater for Catalan speakers than for individuals with both Catalan and Spanish as first tongues, since they might have a greater preference for these versions previously. However, we do not observe the reverse in the second column; the coefficient of Spanish speakers is negative but non-significant. One potential reason is that, as explained in the hypothesis, Spanish versions are still fully available on VoD platforms, unlike Catalan, so the previous users can maintain their audiovisual consumption habits. However, the effect is very large and significant for foreign-language speakers, who switch largely to the original versions or are indifferent. The interpretation of the fourth column is unclear since we cannot know to which options the individual is indifferent. Nonetheless, we can suppose that the indifference increases in the vertical dimension in that case, as an intermediate step between the dubbed and original versions.

6. Conclusions

In this paper, I investigated the effect of the irruption of VoD platforms in Spain, a major change in the movie market, on the preferences relating to different language versions. The spread of VoD platforms and especially Netflix, which was launched in late 2015, totally changed the landscape in terms of the language versions available for movies. In the specific case of Catalonia, it meant a new source of movie consumption with fewer films in Catalan versions than the other ways of consuming movies: cinema and TV. In turn, the VoD expansion dramatically increased the availability of films in original version (with subtitles), while the availability of films in the state language (in the case of Catalonia, Spanish) remained the same: virtually all were available in that language. Hence, the main hypothesis tested in this paper was that the increased (decreased) exposure to original (Catalan) versions should lead to an increase (reduction) of the preference for such a version. In the case of Spanish versions' preference, there was no clear prediction, since their availability remained the same, but there was an increase in the variety of other versions available, due to their non-exclusivity.

The results clearly confirmed the positive effect of VoD exposure on the preference for original versions. The estimates indicate that being exposed to a VoD platform increased at least by 7.86 percentage points the probability of preferring films in original versions. The general results also showed a negative but non-significant effect on the preference for Catalan versions. The effect on the preference for Spanish versions, that was subject to two counterposed effects, turned out to be negative as well; being a VoD user, *ceteris paribus*, decreases the probability of preferring Spanish versions by 6.26 p.p. The empirical analysis shows that the vertical effect (the acceptance of original versions with subtitles) is intense, while the horizontal effect (Spanish versus Catalan in dubbing) is negligible. Finally, no significant effect was found on the indifference towards language preferences.

In this work I also looked at the heterogeneity of the impact of VoD exposure on language preferences. The results show that it only reduced the preference for Catalan versions for the Catalan speakers, most likely due to the fact that such preference was already low for non-Catalan speakers. Such an effect was actually very high, with a reduction of 10 p.p. for Catalan speakers and 7.27 p.p. for individuals with both Catalan and Spanish as their first language. The other heterogeneous effects demonstrate that more educated and younger individuals have a higher propensity to be affected by the exposure to VoD, indicating that they are more adaptable and/or they have the ability to understand other languages or read fast (to understand original versions with subtitles). In addition, these results highlight the two dimensions of differentiation: the horizontal (between languages) and the vertical (between the dubbed versions and the original versions with subtitles). The VoD expansion greatly affected the vertical dimension, with more individuals switching to original versions from dubbed ones, especially young, educated individuals. Whether VoD reduced the preference for Catalan or Spanish (or both) depends on the linguistic characteristics of the individuals; thus, VoD did not make individuals switch within the horizontal dimension, that is, between languages.

The evidence reported in this paper indicates that individual preferences are endogenous: they adapt to previous experiences. Individuals' tastes accustom or accommodate to market changes. Specifically, the case of VoD shows how the introduction of a new technology in a market (in the case of this paper, the movie market) can rapidly shape the preferences of consumers.

In terms of policy implications, the results show the importance of taking into account the endogeneity of the language preferences for policy design. A specific regulation that could alter the linguistic landscape of versions for films will likely create an adaptive change in consumers habits, since they will adapt their preferences. Thus, the welfare analysis of the implementation of the General Law of Audiovisual Communication of 2022,

explained in the introduction, that established quotes for production in the different official languages of Spain, should take into account that in the mid-run it will generate a change in the preferences for different language versions for some consumers.

The expansion of the VoD that continued after the period of analysis of this paper and was boosted by the Covid lockdown will have predictably increased even more the preference for original versions among the population. This could potentially increase individuals' English language skills (Micola et al., 2019) and therefore lead to a subsequent rise in international trade (Ku & Zussman, 2010) and economic growth (Reksulak et al., 2004).

7. References

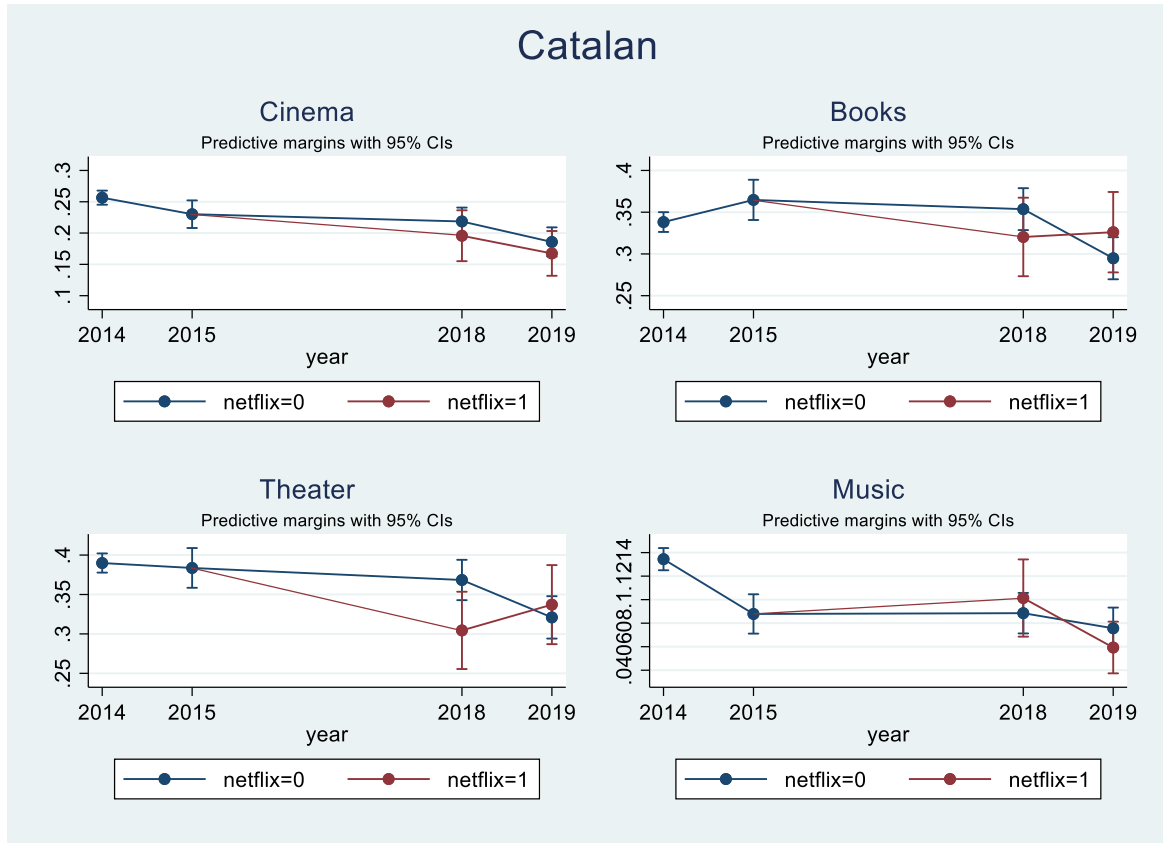
- Anderson, C., & Andersson, M. P. (2013). *Long tail*. Bonnier fakta.
- Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, 36(1), 75–111.
- Butlletí d'informació sobre l'audiovisual a Catalunya* (No. 20). (2022).
- Caminal, R. (2010). Markets and linguistic diversity. *Journal of Economic Behavior and Organization*, 76(3), 774–790
- Dahl, G., & DellaVigna, S. (2009). Does movie violence increase violent crime?. *The Quarterly Journal of Economics*, 124(2), 677–734.
- DellaVigna, S., & Kaplan, E. (2007). The Fox News effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3), 1187–1234.
- Dietrich, F., & List, C. (2013). Where do preferences come from? *International Journal of Game Theory*, 42(3), 613–637.
- Eisensee, T., & Strömberg, D. (2007). News droughts, news floods, and US disaster relief. *The Quarterly Journal of Economics*, 122(2), 693–728.
- Gentzkow, M. (2006). Television and voter turnout. *The Quarterly Journal of Economics*, 121(3), 931–972.
- Institut d'Estadística de Catalunya (IDESCAT). *Televisió i continguts audiovisuals. Audiència per llengua*. <https://www.idescat.cat/pub/?id=eac&n=3.1.2.02>
- Ku, H., & Zussman, A. (2010). Lingua franca: The role of English in international trade. *Journal of Economic Behavior & Organization*, 75(2), 250–260.
- Mallén, B. (2023). *The Effect of Competition on Language Diversity in the Movie-Theatre Industry* (No. 202302). University of Barcelona, Regional Quantitative Analysis Group, revised May 2023

- Micola, A. R., Fenoll, A. A., Banal-Estañol, A., & Bris, A. (2019). TV or not TV? The impact of subtitling on English skills. *Journal of Economic Behavior & Organization*, 158, 487–499.
- Quan, T. W., & Williams, K. R. (2018). Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics*, 49(4), 877–913.
- Reksulak, M., Shughart, W. F., & Tollison, R. D. (2004). Economics and English: Language growth in economic perspective. *Southern Economic Journal*, 71(2), 232–259.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

8. Appendix

Figure A1

Catalan. Margins with Full Specification for Cinema, Books, Theatre and Music



Note: CIs refers to Confidence Intervals.

Figure A2

Spanish. Margins with Full Specification for Cinema, Books, Theatre And Music

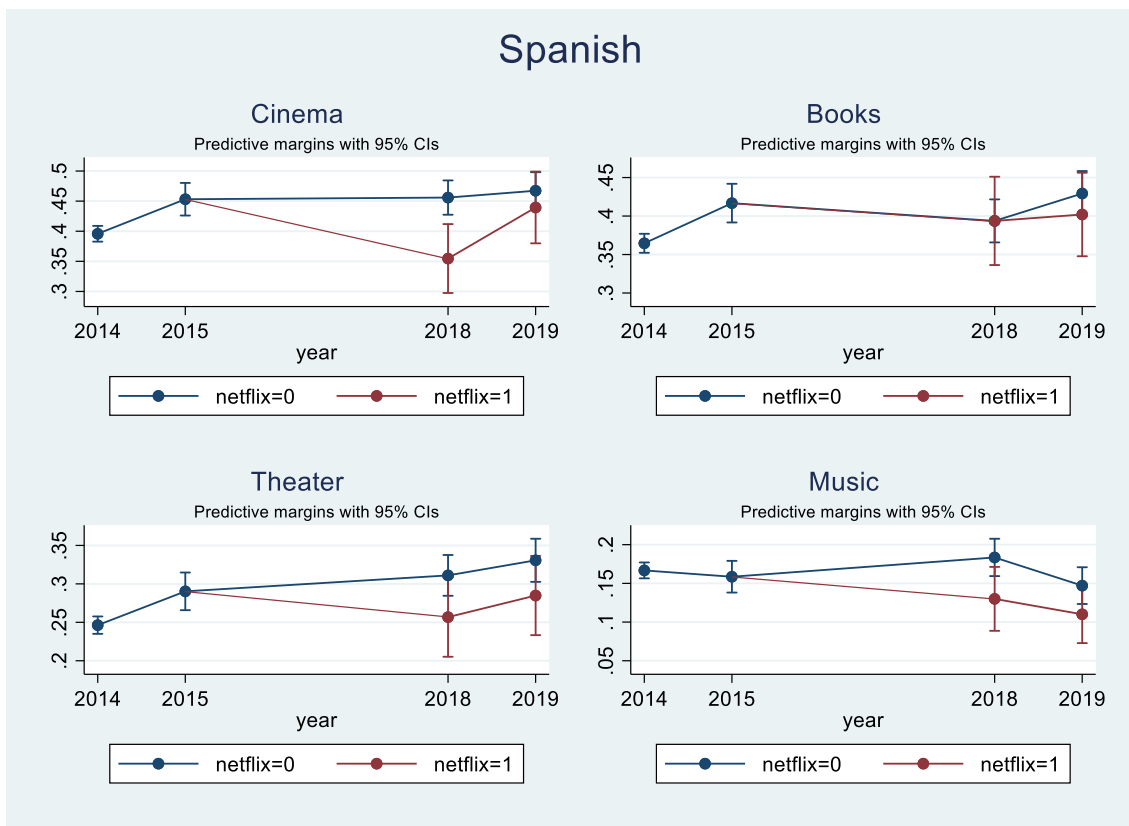


Figure A3

Original Version. Margins with Full Specification for Cinema, Books, Theatre and Music

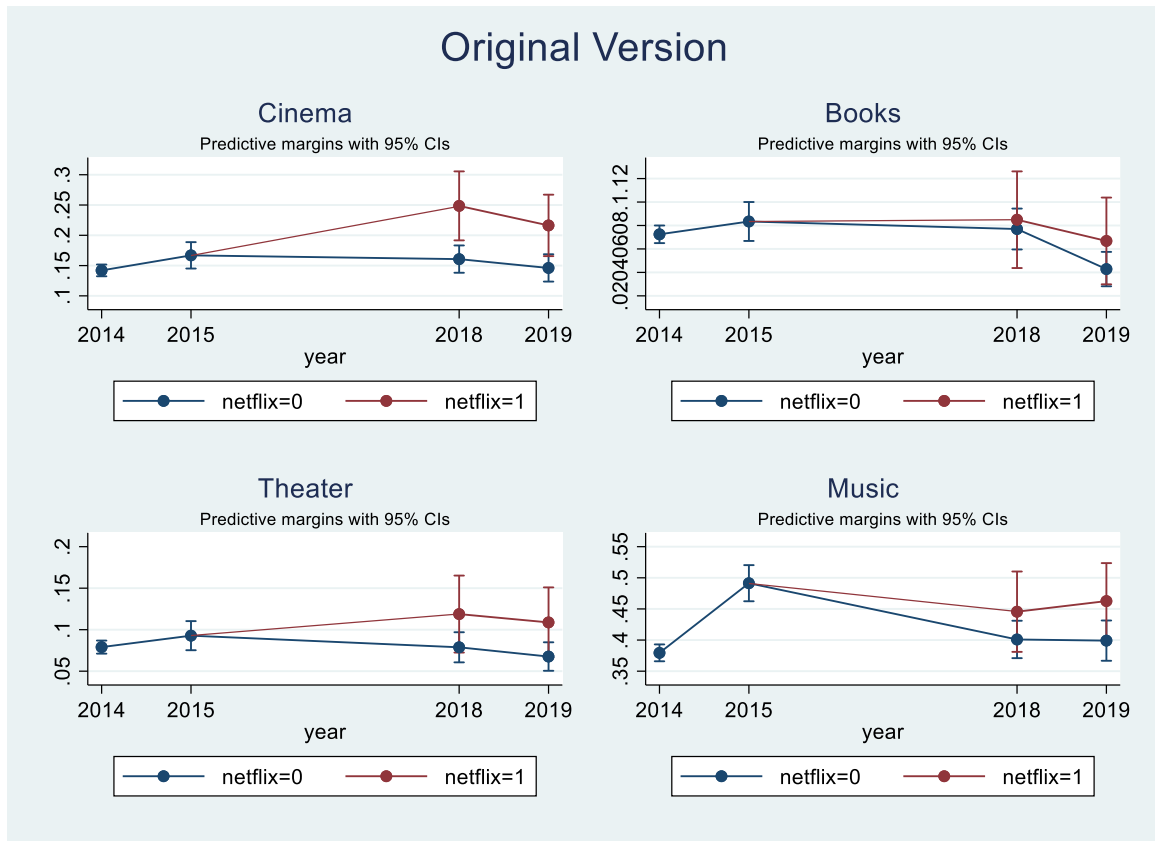


Table A1.1

Basic Regression and Exogenous Characteristics

	(1)	(2)	(3)	(4)
	Catalan	Spanish	Original Version	Indifferent
Basic regression				
<i>Netflix</i>	-0.0520*** (0.016)	-0.0779*** (0.025)	0.176*** (0.023)	-0.0425** (0.018)
ϕ_{2014}	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
ϕ_{2015}	-0.0592*** (0.0146)	0.0783*** (0.0171)	0.0355*** (0.0127)	-0.0503*** (0.0126)
ϕ_{2018}	-0.112*** (0.0132)	0.115*** (0.0169)	0.0255** (0.0126)	-0.0311** (0.0128)
ϕ_{2019}	-0.138*** (0.0128)	0.143*** (0.0172)	0.00141 (0.0123)	-0.0136 (0.0136)
Constant	0.283*** (0.007)	0.376*** (0.007)	0.134*** (0.005)	0.203*** (0.006)
Individual exogenous characteristics as controls				
<i>Netflix</i>	-0.0228 (0.016)	-0.0905*** (0.024)	0.113*** (0.023)	0.00426 (0.018)
L1. Catalan	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
L1. Spanish	-0.380*** (0.010)	0.405*** (0.010)	-0.0349*** (0.008)	0.0125 (0.009)
L1. Catalan-Spanish	-0.270*** (0.022)	0.163*** (0.028)	0.0112 (0.023)	0.0931*** (0.025)
L1. Others	-0.341*** (0.017)	0.273*** (0.027)	0.0207 (0.022)	0.00595 (0.020)
Age	0.0139*** (0.001)	-0.0114*** (0.002)	-0.00944*** (0.001)	0.00728*** (0.001)
Age squared	-0.000107*** (0.000)	0.0000907*** (0.000)	0.0000516*** (0.000)	-0.0000400*** (0.000)
Female	-0.0382*** (0.008)	0.0203** (0.010)	-0.00701 (0.008)	0.0267*** (0.009)
ϕ_{2014}	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
ϕ_{2015}	-0.0323** (0.0127)	0.0479*** (0.0155)	0.0415*** (0.0123)	-0.0531*** (0.0125)
ϕ_{2018}	-0.0472*** (0.0118)	0.0509*** (0.0157)	0.0356*** (0.0124)	-0.0382*** (0.0129)
ϕ_{2019}	-0.0782*** (0.0117)	0.0822*** (0.0162)	0.0139 (0.0123)	-0.0224 (0.0137)
Constant	0.137*** (0.035)	0.462*** (0.044)	0.486*** (0.037)	-0.0989*** (0.032)
N	8045			

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

ϕ are the year dummies

Table A1.2

Full Specification

	(1)	(2)	(3)	(4)
	Catalan	Spanish	Original Version	Indifferent
<i>Netflix</i>	-0.0204 (0.016)	-0.0626*** (0.024)	0.0786*** (0.021)	0.00754 (0.018)
L1. Catalan	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
L1. Spanish	-0.313*** (0.011)	0.330*** (0.013)	-0.0462*** (0.010)	0.0333*** (0.011)
L1. Catalan-Spanish	-0.242*** (0.022)	0.145*** (0.028)	-0.00581 (0.023)	0.100*** (0.025)
L1. Others	-0.274*** (0.020)	0.243*** (0.033)	-0.0647** (0.025)	0.0663*** (0.024)
Age	0.0126*** (0.001)	-0.00826*** (0.002)	-0.0114*** (0.001)	0.00730*** (0.001)
Age squared	-0.0000767*** (0.000)	0.0000339* (0.000)	0.0000792*** (0.000)	-0.0000382** (0.000)
Female	-0.0353*** (0.008)	0.0135 (0.010)	-0.00349 (0.008)	0.0275*** (0.009)
Born. Catalonia	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
Born. Spain	-0.122*** (0.011)	0.110*** (0.016)	0.0362*** (0.011)	-0.0232* (0.013)
Born. European Union	-0.0339* (0.018)	-0.0246 (0.034)	0.121*** (0.027)	-0.0896*** (0.022)
Born. Non-EU	-0.0453*** (0.015)	-0.0235 (0.029)	0.118*** (0.026)	-0.0525** (0.020)
Municipality. Less than 50,001 inhabitants	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
Municipality. Between 50,001 and 150,001	-0.0345*** (0.010)	0.0681*** (0.013)	-0.00133 (0.009)	-0.0326*** (0.011)
Municipality. Between 150,001 and 1M	-0.0520*** (0.014)	0.0339* (0.019)	0.0249* (0.014)	-0.00894 (0.016)
Municipality. More than 1M	-0.0714*** (0.012)	-0.0321** (0.015)	0.137*** (0.013)	-0.0322** (0.013)
Works	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
Does not work	-0.0201** (0.010)	0.0246** (0.012)	-0.00991 (0.009)	0.00432 (0.011)
Primary education or less	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
Secondary education	0.0459*** (0.014)	-0.0457** (0.019)	0.00549 (0.010)	0.00280 (0.017)
Professional course	0.0641*** (0.015)	-0.117*** (0.019)	0.0429*** (0.011)	0.0223 (0.017)
Tertiary education	0.0381** (0.016)	-0.222*** (0.020)	0.183*** (0.013)	0.0120 (0.018)
ϕ_{2014}	Ref. Cat.	Ref. Cat.	Ref. Cat.	Ref. Cat.
ϕ_{2015}	-0.0264** (0.0127)	0.0575*** (0.0153)	0.0248** (0.0121)	-0.0524*** (0.0126)
ϕ_{2018}	-0.0384*** (0.0124)	0.0531*** (0.0157)	0.0203 (0.0125)	-0.0326** (0.0134)
ϕ_{2019}	-0.0702*** (0.0129)	0.0794*** (0.0167)	0.00217 (0.0125)	-0.0101 (0.0145)
Constant	0.0987** (0.0386)	0.574*** (0.0480)	0.417*** (0.0377)	-0.112*** (0.0379)

N

8045

Robust standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

 ϕ are the year dummies

Table A2*Main Results with Non-Linear Models*

	(1)	(2)	(3)	(4)
	Catalan	Spanish	Original Version	Indifferent
<i>V_oD</i> (Logit)	-0.0429* (0.0240)	-0.0598*** (0.0225)	0.0441*** (0.0145)	0.0067 (0.0221)
<i>V_oD</i> (Probit)	-0.0411* (0.0228)	-0.0584*** (0.0225)	0.0480*** (0.0152)	0.0063 (0.0213)
N	8045			

Standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

Full specification. Marginal effects reported.

The logo for UBIREA, featuring the text 'UBIREA' in a bold, white, sans-serif font inside a white rounded rectangle.

UBIREA

Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

WEBSITE: www.ub.edu/irea • **CONTACT:** irea@ub.edu

The logo for AQR, featuring a green circular icon with a white dot inside, followed by the text 'AQR' in a bold, white, sans-serif font inside a white rounded rectangle.

AQR

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu