UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2021

Gravity and Trade in Video on Demand Services

Annette Broocks, European Commission Joint Research Centre, Seville
Zuzanna Studnicka, University College Dublin

WP21/27

December 2021

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Gravity and trade in video on demand services*

Annette Broocks[†] Zuzanna Studnicka[‡]

December 3, 2021

Abstract

We estimate the patterns of catalogue availability (extensive margin) and number of clicks per title (intensive margin) using a novel data set containing the information on Netflix catalogues and viewing across 20 countries. Our results show evidence of the gravity framework explaining both margins of Netflix watching. In particular, we find that Netflix users have a strong preference for domestic productions. Detailed information on film and TV show characteristics gives us a unique opportunity to estimate the importance of quality in determining the patterns of Netflix watching. Independent viewers' ratings and a title's age play a key role in explaining the number of clicks directed at a particular title. Finally, Netflix Original productions attract a disproportionately large number of clicks.

JEL Classification: F10, L82, Z10
Keywords: Netflix; subscription video on demand; gravity equation; services trade

1 Introduction

Rapid advances in digital technology gave rise to the development of new types of trade and new goods and services being traded. Among these advances, an increased penetration of high-speed internet made it possible for consumers to stream audiovisual content and substantially lowered trade frictions associated with shipping physical goods to different markets. As a result of these new opportunities, a wide variety of streaming platforms including Disney+, HBO, Hulu, Netflix, Spotify, YouTube, major sports leagues, and television networks are operating in many countries across the planet. Grouped together video on demand services (VOD) account for more than 60% of global downstream internet traffic (Sandvine, 2019). This share is so substantial that Internal Market Commissioner Thierry Breton asked the large streaming platforms to temporarily reduce quality (either

*We would like to thank Ron Davies, David De Remer, Benjamin Elsner, Martina Lawless, Dorothee Hillrichs, Hylke Vandenbussche, Gonzague Vannoorenberghe and seminar participants at University College Dublin, KU Leuven, and the Joint Research Centre in Seville for helpful comments. We thank participants at the 2021 European Trade Study Group conference and the Trade & Investment in Services Associates 2021 conference. All errors are our own. The views expressed are purely those of the authors and may not under any circumstances be regarded as stating an official position of the European Commission.

[†]European Commission, Joint Research Centre (JRC), Seville, Spain, annette.broocks@ec.europa.eu
[‡]University College Dublin, zuzanna.studnicka@ucd.ie

streaming bit-rates or resolution) in order to avoid a collapse of the internet infrastructure during the first Covid-19 lockdowns in early 2020 (BBC, 2020).

Given the differences in their business models, it is important to define the three main types of VOD: subscription video on demand (SVOD), transactional video on demand (TVOD) and ad-based video on demand (AVOD). SVOD allows users to access an entire catalogue for a recurring fee (e.g., Netflix, Hulu, Disney+, HBO), TVOD allows to buy content on a pay-per-view basis (e.g., Apple TV, Google Play), while AVOD refers to video on demand that is free to its consumers but generating ad revenue (e.g., YouTube and ViacomCBS's Pluto TV).[1]

According to Statista (2020), in 2019 the user penetration rate of SVOD amounted to almost 22% in Central and Western Europe and 28% in Northern Europe, with expected increases of up to 2 percentage points in 2020. The number of subscribers to SVOD in Europe exceeded 100 million for the first time in 2019 and is projected to be close to 200 million by 2025 (European Audiovisual Observatory, 2020; DigitalTV Europe, 2020). Among the large SVOD players, Netflix is the strongest competitor capturing 23% of the global streaming market, followed by YouTube (13%) and Amazon Prime Video (4%), as reported in Sandvine (2019). While accounting for only 0.99% of connections, Netflix alone consumes around 13% of global downstream traffic (Sandvine, 2019). It has risen from delivering DVD's in 2007 to the most widely available SVOD platform and content producer since the launch of "House of Cards" in 2013. It is now present with its streaming service in more than 190 countries (all countries and regions except for China, Crimea, North Korea and Syria). An important feature of Netflix is its independence from TV broadcasters, traditional production studios and tech companies. It is, therefore, different from other video on demand services such as Amazon Prime Video (part of Amazon), Apple TV (part of Apple), BBC iPlayer (part of BBC), or Peacock (part of NBCUniversal).[2] In addition, in contrast to platforms such as TikTok and YouTube, hosting mostly user-created non-professional content, Netflix offers professional scripted and unscripted video on demand services.[3]

SVOD is not only widely-used but also economically relevant. SVOD revenues are the main driver of the EU audiovisual market growth in Europe, representing 83% (EUR 1.55 billion) of the total revenue growth of EUR 1.87 billion in 2018 (European Audiovisual Observatory, 2020). Despite its popularity and a rich literature on the economics of films,[4] research on SVOD is still very limited and most of it is conducted in the fields of media studies and computer science.

The aim of this paper is to study Netflix viewing patterns from the perspective of international trade. The reason behind this is that Netflix acts as an intermediary connecting content producers with content consumers, including viewers in countries other than where content is produced. In this way, a viewer in country i that watches content created in country j imports the services of that content's producer via Netflix. Taking this further, we can consider Netflix's role in this transaction as operating on both an extensive margin (as distributor that manages content availability, that is, whether a given content from j is even available in i) and an intensive margin (as retailer that attracts clicks directed at a given title by viewers in i).[5] Although availability via Netflix removes

[1] Some platforms offer a mix of those services e.g., Amazon Prime Video.

[2] See Burroughs (2019), and Lotz, Lobato and Thomas (2018) for details.

[3] Note that scripted content is produced with a script. Examples include TV series and films. Unscripted content is produced without a script. Examples include talk shows, documentary-series, and game shows.

[4] See McKenzie (2021) for an overview.

[5] Our paper is therefore related to the line or research decomposing international trade into various

many of the barriers between viewers and producers, it is unlikely to remove all of them, especially those related to copyrights or culture (e.g. language). Thus, we anticipate that many of the factors that drive international trade will continue to be important when considering trade in services via SVOD. Our goal is to examine the role of these factors in the provision of audiovisual content via Netflix.

By answering this question, we contribute to a rather limited field in the international trade literature analysing trade in cultural goods and services using the gravity framework. Most papers in this field focus on the role of factors such as cultural proximity, linguistic similarity, past colonial links, or migrants in determining trade in this type of product. In particular, Marvasti and Canterbery (2005) analyse the determinants of US films' exports to 33 countries for the period 1991-1995. They find a positive impact of language, education, and religion on exports. Hanson and Xiang (2011) confirm the importance of market size and language for bilateral US film exports and find that countries with more trade barriers import fewer American films. Alaveras, Gomez-Herrera and Martens (2018) show that success in the home market, and the size of the film budget have a positive effect on trade in films between countries. Their findings also show that relative to their success in the home market, American films have a lower propensity to be exported than EU films. Disdier, Tai, Fontagné and Mayer (2010) analyse the determinants of trade in cultural goods (books, media, visual arts etc.) between a wide range of countries. They find that common language and past colonial links matter for trade, the former being particularly important for trade in books and newspapers. Ferreira and Waldfogel (2013) analyse trade in music among 22 countries and find that it follows similar patterns with distance and common language playing a role. Surprisingly, they find that, despite an increase in availability (due to an increase in local MTV channels and Internet penetration), the degree of home bias has increased since the 1990s. Finally, Hellmanzik and Schmitz (2015) analyse audiovisual services using the gravity framework. They find that virtual proximity (based on bilateral hyperlinks and bilateral website visits between countries) plays an important role in explaining this kind of trade. In summary, the existing literature points to importance of cultural factors and home bias in this kind of trade.

In our paper we use a novel data set that allows us to analyse the patterns of both catalogue availability and viewing of Netflix in twenty countries. Existing papers analysing Netflix focus on the first question. In particular, the paper most closely related to our research, Batikas, Gomez-Herrera and Martens (2015), analyses the content of the Netflix catalogues in eleven EU countries. They find that geographical distance and common border do not significantly affect the availability, while the languages index[6] and the common language dummy are both significant. Overall, their results suggest consumer preferences for home market products and for linguistic proximity. We add to their extensive margin results by also considering the intensive margin.

Furthermore, considering the intensive margin gives us a unique opportunity to estimate the importance of product quality in international trade that goes beyond the traditional measure of quality which is product price or unit value. In general, literature on quality and trade relies, with rare exceptions,[7] predominantly on these two measures of quality and focuses on the determinants of exports or imports of higher quality goods at the country level (Schott, 2004; Hallak, 2006) or at the firm-level (see e.g., Manova and Zhang, 2009).[8] The measure of quality used in our paper are independent show/film

margins (see e.g., Hummels and Klenow, 2005).

[6] Defined as the composite common language index introduced by Melitz and Toubal (2014).

[7] See e.g., Crozet Head and Mayer (2012), Chen and Juvenal (2016) who use more direct measures of quality.

[8] It is also important to note that even more direct measures of product quality used by Crozet et al.,

ratings downloaded from the internet movie database (IMDb). Using this measure is particularly interesting because it allows us to separate price from quality perceived by viewers. This is because, even if a highly-ranked content is usually more costly to produce, this does not translate into a higher price charged by Netflix to the consumer. Our second research question is therefore, what effect does quality have on Netflix viewing patterns?

By analysing the questions of content availability and viewing, our research contributes to the policy debate whether, by making a large amount of US-origin and Netflix-produced content available online, the expansion of Netflix can negatively affect smaller non-US content producers. Empirical evidence studying the first issue shows that even if Netflix distribution favours US-origin content, the degree of this advantage is smaller via Netflix than it would be via traditional theatrical distribution (Aguiar and Waldfogel, 2018; Alaveras et al., 2018). However, the empirical evidence on the distribution of Netflix-produced content as well as on the extent of content viewing is still missing. This is worth knowing as it is a part of a larger trend for streaming platforms including Amazon Prime, Hulu etc. to produce their own content. Our third research question is to what extent is Netflix availability and viewing driven by Netflix-produced content?

Last, our research is closely related to the policy issue of access control - geoblocking - that can be considered a threat to the European Digital Single Market.[9] Despite successful efforts by the European Commission regarding physical trade, the issue of market fragmentation in audiovisual services trade has been exempted from regulations until at least 2022, when the next short-term review of the Regulation (EU) 2018/302 "the geoblocking regulation" is due. Under current legislation, a great portion of content financing comes from distributors with exclusive rights to deliver content to each country (Aguiar and Waldfogel, 2018). It is, therefore, in their interest to keep the audiovisual sector fragmented by keeping users' access restricted based on their geographic location. In order to understand the issue of fragmentation we provide an extensive analysis of Netflix catalogues and viewing across countries. This research question is addressed in the descriptive analysis of this paper.

Based on our research questions, we focus on four main hypotheses. First, because of geoblocking we expect a large variation in the content of Netflix catalogues across countries, reflecting different distribution patterns such as copyrights requirements etc. Second, although distance and culture (traditional gravity variables) do not affect the cost for Netflix when providing a title as an option, they do affect the value of consumption. Therefore, we anticipate a greater effect at the intensive rather than the extensive margin. Third, quality is a major determinant of Netflix viewing. Finally, because of the algorithms used by Netflix, its original content attracts a large proportion of clicks.

Our findings confirm our hypotheses. In particular, we find that Netflix catalogues are very different across countries. Despite the fact that Netflix does not require any physical transport of its content, gravity variables matter for Netflix availability and viewing with distance and common language being the most important among them. Beyond the traditional gravity variables, we find that viewers have a strong preference for domestic productions. Finally, Netflix viewing depends on content's rating and age, and Netflix Original productions attract a large share of clicks. This latter result, for instance, suggests a potential need for competition regulators to continue examining VOD markets.

The remainder of the paper is structured as follows. Section 2 describes the data. Section 3 presents the descriptive analysis focusing on the issue of fragmentation of the

(2012) and by Chen and Juvenal (2016) are reflected in product prices.

[9] A relatively large body of literature focuses on the issue of geoblocking (see e.g., Lobato, Meese, Rugg and Burroughs, 2016, Aguiar and Waldfogel, 2018).

Netflix catalogue. Section 4 contains our methodology. Section 5 presents the results. Section 6 presents some robustness checks. The final section concludes.

2 Data

In this paper we explore the patterns of content availability and viewing using data on Netflix streaming services coming from SimilarWeb, Netflix and Ampere Analysis. Additional content-specific data is obtained from the European Audiovisual Observatory and IMDb.

Based on these sources we construct a data set containing the information on Netflix catalogues available in twenty countries as well as on the number of clicks that go to each specific title. We focus on two types of content: scripted films and TV shows, ignoring unscripted formats such as stand-up comedy, reality-TV, game shows, and documentaries. We treat the country of viewing as the destination country and the main production country as the country of origin. Our destination countries include 15 EU countries and five non-EU countries.[10] Our countries of origin include 80 main production countries across six continents.[11]

In order to construct our final data set we proceed in four steps. In the first step we collect the information on title clicks in the desktop version of Netflix, using the data coming from SimilarWeb (a company that provides data on usage behaviour of internet and mobile phone users).[12] The click data contains the share of clicks going to individual pages of the Netflix domain.[13] We extract a list of pages from the netflix.com domain and their corresponding traffic share per page for desktop traffic.[14] We download the data for the period from January 2018 to December 2019 in intervals of half-years.[15] Based on the structure of each page, we can differentiate between different activities on the platform and even for individual titles. For example, we can tell which share of clicks is spent on browsing and searching activities and which share is spent on watching a title.[16] Unfortunately, our data does not allow to identify whether a user finished watching a given show/film or only started to watch and resumed watching at another point in time, which would count as two watching instances. After extracting the shares, we use an auxiliary data set containing the total clicks to the netflix.com domain per country, which allows us to translate the share allocated to each title into the number of clicks. Another caveat is

[10] EU countries include: Austria, Belgium, Czechia, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, Poland, Portugal, Spain, Sweden and the UK. The non-EU countries include: Australia, Canada, Norway, Switzerland, and the US.

[11] See Table B.1 in Appendix B for a list of production countries.

[12] Read more on SimilarWeb data here: https://support.similarweb.com/hc/en-us/articles/360001631538-SimilarWeb-Data-Methodology

[13] Due to distinct website structures, such detailed data is not available for other internationally operating streaming platforms. For this reason, our paper focuses on one platform only.

[14] For the Netflix app on Android and Mac OS, a decomposition into single titles is not available. We have to assume that Netflix usage is the same when watching in a browser window and when using the app. During the time covered by our sample, the mean ratio of desktop users to Android app users was about 1.12:1. Some users may use both the app and desktop version.

[15] SimilarWeb's popular pages metric summarizes the shares of the 1,000 most clicked pages within a month over each half year interval. Therefore, the number of pages varies between countries, depending on the heterogeneity of watching behaviour. For each country, the shares add up to a figure of around 50%. It means that there is a long right tail with each title receiving a very low number of clicks, not captured by SimilarWeb's data collection method.

[16] On average, 40% of observed clicks are within the "browsing and searching" categories, 5% of clicks go to inspecting a specific title, 51% of clicks are within administration and login, and only 3% of clicks are within the "watch" category.

that, when computing the number of clicks, we are not able to allocate the share of clicks going to individual episodes (or seasons) of TV shows. For this reason, in our analysis we need to consider films and TV shows separately.

In the second step, we merge our click data with more comprehensive information on each TV show/film, scraped directly from Netflix. To match SimilarWeb with Netflix data we use a unique, (usually) 8-digit-long, title id, which allows us to scrape the corresponding title overview page on the Netflix website.[17] We scrape the lists of title pages for each of the twenty countries in our data set using corresponding VPN servers to take local availability into account. This allows us to collect the TV show or film title corresponding to a specific title id, as well as information on its release year (which allows us to compute its age) and genre.

In the third step we match the SimilarWeb and Netflix data to Ampere data. The Ampere Analysis Analytics SVoD App provides a large database that contains detailed monthly information about the full catalogues of Netflix and other streaming platforms worldwide. The information available includes 23 variables such as the primary production country, average episode length, IMDb id number, and an identifier of Netflix Original productions.[18]

In the last step we merge our data with the information on co-productions coming from the European Audiovisual Observatory Lumiere VoD data, with the award and ratings data and film's length data from IMDb,[19] and with the gravity data coming from CEPII. In addition, the last data set on linguistic proximity used in our robustness checks comes from http://faridtoubal.com.

3 Descriptive analysis

We start our analysis by looking at the characteristics of Netflix content in general, and within national catalogues.[20] While Lobato (2018) states that the content of national Netflix catalogues reflects local distribution patterns, copyrights, and licensing rather than the actual demand of national audiences, we believe that it is, at least to some extent, selected to respond to the preference of local demand. This is particularly true in recent years during which Netflix opened a number of offices in Europe (Amsterdam, Berlin, London, Madrid and Paris), and announced that it was going to produce more of its original content for local markets (Netflix 2019; Netflix, 2020). Moreover, Directive 2010/13/EU (The "Audiovisual Media Services Directive") of the European Parliament and of the Council states that "*[On-demand audiovisual media services] should, where practicable, promote the production and distribution of European works and thus contribute actively to the promotion of cultural diversity.*"

Starting with the overall Netflix catalogue for the twenty countries in our data set, Table 1 presents the total number of distinct titles available, and the total number of observed clicks, distinguishing between TV shows and films. As can be seen, the total number of available titles was 14,379 with almost 12,000 titles available each year.[21] During these two years, Netflix scripted content attracted 764.3 million watch clicks. It is important to

[17] Netflix pages are of the form https://www.netflix.com/title/80057281 (Stranger Things), where the number 80057281 corresponds to title id. Title ids are identical across countries.

[18] See https://www.ampereanalysis.com/products/about/analytics-svod for more details.

[19] IMDb data including worldwide ratings is generated by viewers from all sources e.g., cinema, DVD sales, television broadcast and more.

[20] Note that, for the sake of clarity, in parts of our analysis we aggregate our half year periods to years.

[21] The overlap between 2018 and 2019 was 9,039 titles.

note that even if films accounted for 81% of titles, their share of watched scripted content was less than 1%. This is a direct result of the issue mentioned earlier, namely we do not observe the number of clicks allocated to individual episodes of TV shows. Finally, although Netflix Original content accounted for only 3.44% of titles, it attracted more than 65% of clicks, suggesting that Netflix may be pursuing a strategy that diverts views to its own content. In addition, Netflix Exclusives accounted for a similar share of titles but only 8.6% of clicks.[22]

Table 2, displays viewing patterns by genre. It shows that Crime & Thriller was the genre with the largest number of titles available (19.18% of titles), but the most popular genre in terms of the number of clicks was Comedy (44.5% of clicks).

Similar to other platforms such as YouTube, Netflix has superstar titles and titles that hardly anyone watches (also displayed in Figure 3). The most popular titles watched during our sample are listed in Table 3. Amongst them, an American TV show - "The Office" attracted 104 million clicks followed by "Brooklyn Nine-Nine" and "Stranger Things". The most watched film was "Black Mirror: Bandersnatch" followed by "El-Camino: A Breaking Bad Movie". Interestingly, these two most watched films are based on Netflix's TV shows.

Table 1: Number of titles and clicks

	Number	Share	Clicks	Share of clicks
All	14,379		764.3M	
2018	11,730		363M	47%
2019	11,688		401M	53%
Films	11,656	81.06%	4.3M	0.56%
TV Shows	2,723	18.94%	760M	99.44%
Netflix Original	496	3.44%	499.8M	65.40%
Netflix Exclusive	538	3.73%	65.9M	8.60%

Table 2: Number of titles and clicks by content type

Genre	Number	Share	Clicks	Share clicks
Action & Adventure	1,285	8.93%	35.2M	4.6%
Children & Family	2,319	16.10%	19.4M	2.5%
Comedy	2,581	17.93%	340M	44.5%
Crime & Thriller	2,760	19.18%	122M	16.0%
Drama	1,406	9.77%	30M	3.9%
Horror	609	4.23%	21M	2.7%
Romance	2,167	15.05%	21M	2.7%
Sci-Fi & Fantasy	1,252	8.69%	174M	22.8%
Total	14,379	100	764.3M	100%

[22] According to the definitions provided by Ampere Analysis, "Netflix Original" refers to a piece of content funded, commissioned or produced by Netflix. "Netflix Exclusive" refers to a piece of content for which Netflix has exclusive rights in at least one market, and therefore brands as an Original. Note that because for the consumer this difference is not apparent, we combine Netflix Original and Netflix Exclusive content in our regression analysis.

Table 3: Most watched content

	Title	Type	Original	Year	IMDb	Wins	Origin	Clicks (M)
1	The Office (U.S.)	TVShow	No	2005	8.8	48	US	104
2	Brooklyn Nine-Nine	TVShow	No	2013	8.4	12	US	32.5
3	Stranger Things	TVShow	Yes	2016	8.8	67	US	27.9
4	Big Mouth	TVShow	Yes	2017	8	0	US	20
5	13 Reasons Why	TVShow	Yes	2017	7.8	3	US	19.1
6	Money Heist	TVShow	Yes	2017	8.5	23	ES	18.8
7	Disenchantment	TVShow	Yes	2018	7.2	0	US	18.3
8	Love, Death & Robots	TVShow	Yes	2019	8.6	9	US	15.4
9	Lucifer	TVShow	Yes	2015	8.2	0	US	14.6
10	You	TVShow	Yes	2018	7.8	1	US	14.5
11	The Umbrella Academy	TVShow	Yes	2019	8	6	US	14.4
12	The Good Place	TVShow	Excl.	2016	8.2	16	US	14.3
13	Friends	TVShow	No	1994	9	77	US	14.3
14	The End of the F***ing World	TVShow	Excl.	2017	8.1	4	UK	13.7
15	Orange Is The New Black	TVShow	Yes	2013	8.1	50	US	13.3
90	Black Mirror: Bandersnatch	**film**	Yes	2018	7.2	4	US	1.7
122	El Camino: A Breaking Bad Movie	**film**	Yes	2019	7.4	4	US	0.89

In the second step we look at Netflix catalogues by country. We begin our analysis by looking at the number of titles available in each country's individual catalogue. Figure 1 reports the number of titles available in each country in 2018 and in 2019 (sorted by growth rate). It can be seen that Canada had the largest catalogue with 6,443 titles available in 2019, followed by the UK with 5,840 titles available in 2019. Spain, on the other hand, had the smallest catalogue with only 3,429 titles available in 2019. With respect to its growth, the Polish catalogue experienced the largest increase in the number of titles available: from 2,825 in 2018 to 5,236 in 2019, followed by the Danish and Portuguese catalogues (although the size of the latter was only 4,010 titles in 2019). Overall, we observe a substantial heterogeneity in terms of both catalogue size and growth across countries in our sample.

Figure 2 shows the composition of Netflix catalogues by country based on the content's origin. First thing to note is that in most countries, American content was the most prevalent. Quite surprisingly, the US was the country with the lowest share of American content in its catalogue (around 39% of titles), while Finland was the country with the highest share of American content (around 54% of titles). Domestic content (other than American) had the highest share in France (around 9% of titles) followed by the UK (around 8% of titles). Portugal and Switzerland were the two countries with just a handful of domestic titles (the share was close to 0%). The average share of EU-produced content across EU countries was around 19%, much smaller than the 30% envisaged in the 2018 amendment to the Audiovisual Media Services Directive mentioned above.[23] This finding, together with the fact that the overlap between catalogues, especially for films, is rather limited between countries (see Appendix A for a more complete description of catalogue overlap), suggest that even if countries' catalogues are growing and include more and more EU-produced content, there is still room for improvement in terms of content availability for consumers.

In the last step of the descriptive analysis, we look at the distribution of the number of clicks by title. Figure 3 shows the number of clicks going to individual titles by year. It shows that the vast majority of titles were never watched, and only a handful of titles received any clicks at all. Figure 4 shows the percentage of watched TV shows and films by destination. The share of watched TV shows was the lowest in the UK and the highest in Finland, but never reached 10%. The share watched films was even lower than of TV shows (below 1%). Similarly to TV shows, it was the lowest in the UK and the highest in Finland. This figure also indicates that the proportion of watched content was

[23] Directive (EU) 2018/1808 of The European Parliament and of The Council of 14 November 2018 amending Directive 2010/13/EU.

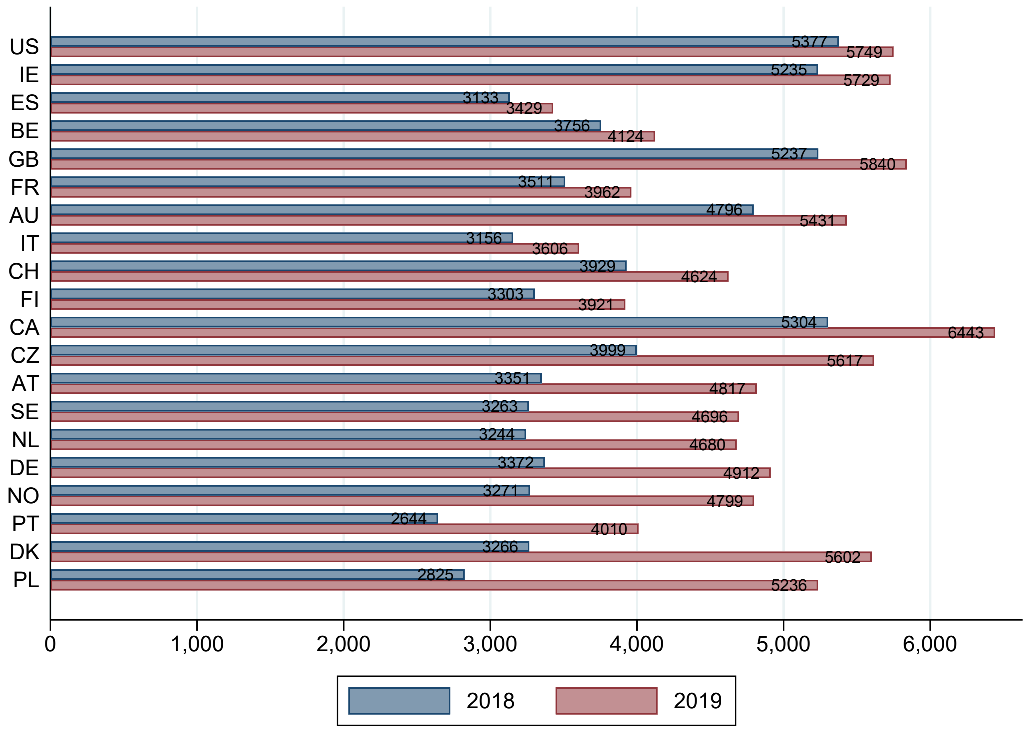

Figure 1: Number of titles by country

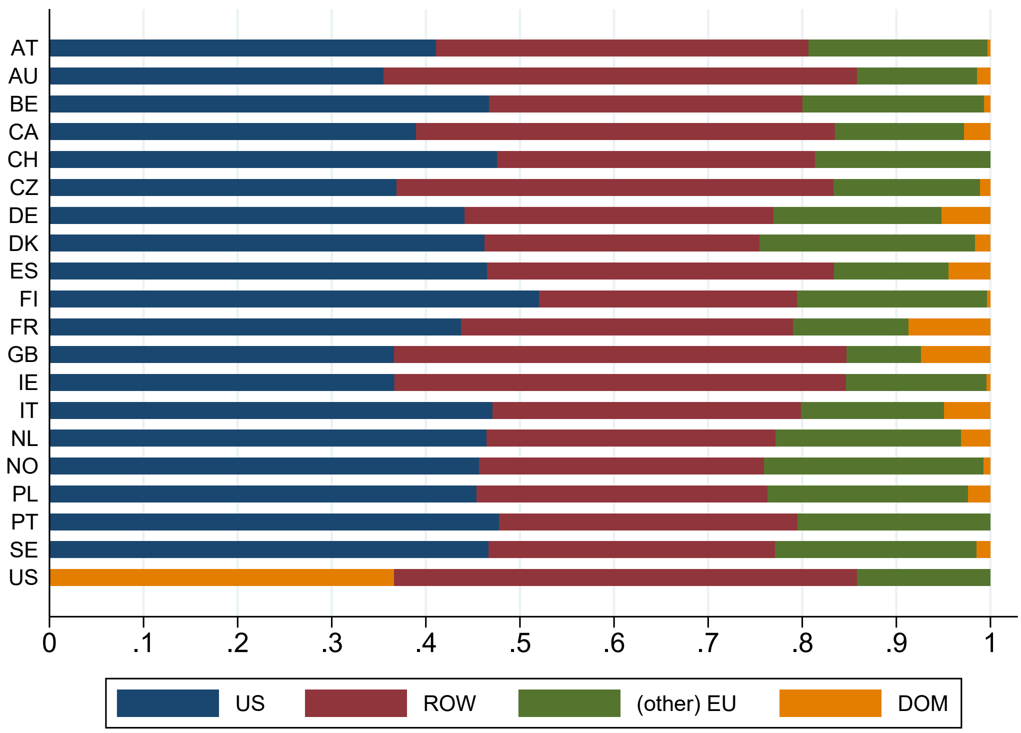

Figure 2: Share of different production countries, 2019

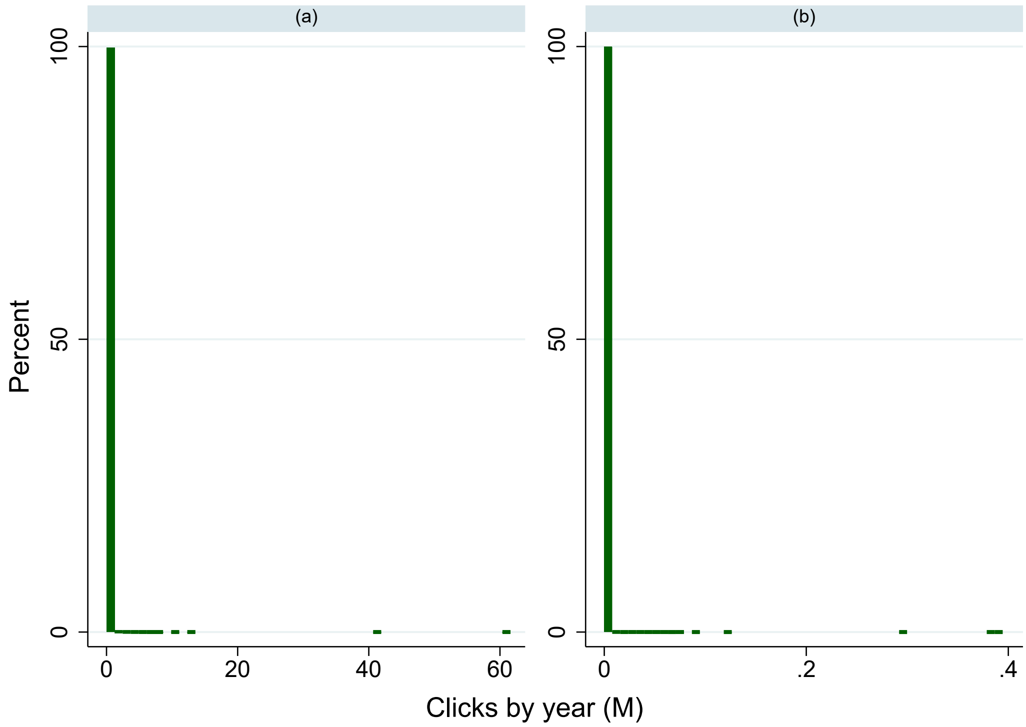

Figure 3: Distribution of clicks going to individual titles, panel (a): TV shows, panel (b): films

predominately driven by Netflix Original productions in case of TV shows.

To sum up, our descriptive analysis uncovers some interesting patterns about both margins of Netflix watching. First of all, content availability varies by country in terms of both catalogue size and its composition. Catalogues overlap to some extent, but the overlap does not necessarily depend on cultural proximity or common language. Second, despite the large size of available catalogues, only a handful of titles attract most of the clicks. TV shows are more popular than films, yet less than 10% of them are watched at all. This is reminiscent to what we know about trade in goods where only a few goods account for the majority of firm's exports (see e.g., Bernard, Jensen, and Schott, 2011) and a handful of firms account for the majority of a country's trade (see e.g., Mayer and Ottaviano, 2008). In the next step we use regression analysis to see which factors drive both margins of Netflix watching.

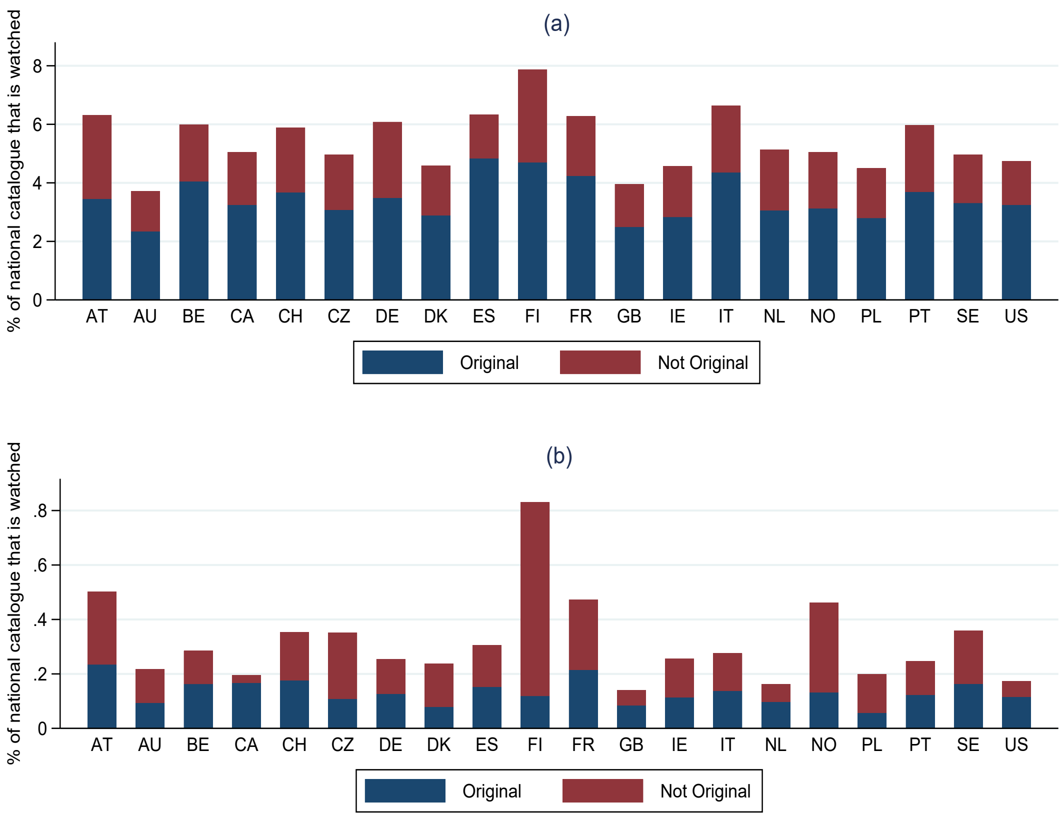

Figure 4: Watched titles as percentage share of full catalogues, 2019 H2
panel (a): TV shows, panel (b): films

4 Estimation strategy

In this section, we estimate an extended gravity equation relating the number of available titles (extensive margin) and the number of clicks per title (intensive margin) to standard country-pair characteristics.

We begin with the extensive margin and estimate the following specification.

$$Y_{i,j,t} = \alpha + fe_{it} + fe_{jt} + \beta_1 x_{i,j} + \varepsilon_{i,j,t} \tag{1}$$

Where $Y_{i,j,t}$ is the number of titles from country j available in country i at time t (we use half-year periods), and $x_{i,j}$ is a set of time-invariant gravity variables, fe are the fixed effects, and $\varepsilon_{i,j,t}$ is the error term.

In the second step we estimate the following intensive margin equation.

$$y^p_{i,j,t} = \alpha + fe_{it} + fe_{jt} + \beta_1 x_{i,j} + \beta_2 x_p + \beta_3 age_{p,t} + \sum_{g=1}^{8} \gamma_g genre_{p,g} + \varepsilon^p_{i,j,t} \tag{2}$$

where $y^p_{i,j,t}$ is the number clicks going to an individual title p at time t, $x_{i,j}$ is a set of time-invariant gravity variables, x_p is a set of time-invariant show/film characteristics (IMDb rating, length and Netflix Original dummy), $age_{p,t}$ is the title's age at time t, and $genre_{p,g}$ is a set of genre dummies.

Among the gravity variables, using distance between countries may seem counter-intuitive when it comes to trade where non-physical transport of goods takes place. However, following the literature, we consider it as a proxy for cultural distance between the two countries. The measure we use is the distance between capitals.[24] Other gravity variables such as common border, common language and colonial links are standard proxies used to measure trade costs or trade facilitators. In our case, we believe that the effect of the common language shared by the destination country and the main production country should be particularly important, as viewers usually prefer content in their mother tongue (see e.g., Marvasti and Canterbery; 2005). We use the common language dummy as our preferred measure of common language and present additional results, using different language proximity measures in the robustness checks section.

Several estimation methods are used in trade literature to estimate the gravity equation. We estimate both the extensive and the intensive margin models using the Poisson pseudo-maximum likelihood (PPML) method suggested by Santos Silva and Tenreyro (2006) for estimating the gravity equation when the number of zeros is particularly large. We introduce origin-, and destination-time (half-year) fixed effects and cluster the standard errors at the same level.[25] We estimate our model using `PPMLHDFE` Stata command developed by Correia, Guimarães and Zylkin (2019). An advantage of using this package is that it automatically drops singletons from the regression analysis, improving the accuracy of standard errors estimation, but reducing the sample size. For this reason, the sample used in our regressions is different than the sample used in our descriptive analysis. For completeness, we present the summary statistics for the sample used in our regression analysis in Table 4 and the summary statistics for the entire sample used in Section 3 in

[24] We have also experimented with alternative distance measures such as the population weighted distance or distance between the most populated cities. Our results had no qualitative and only a small quantitative impact of the results presented here. They are available on request.

[25] Please note that we have also tried different levels of clustering. The significance levels remained unchanged.

Table B.2 in Appendix B.

We estimate our intensive margin regressions for films and TV shows separately, as Table 1 suggests that due to the vast majority of clicks going to TV shows, a pooled estimation would be driven by views of TV shows.[26]

[26] We present the results of the pooled estimations in Table C.1 in Appendix C.

Table 4: Summary statistics

		All			
Variable	Obs.	Mean	Std. Dev.	Min	Max
Extensive margin	5,940	48.41	204.76	0.00	2,464
Clicks	203,688	3,680	137,969	0.00	33,400,000
logDist.	203,688	8.37	1.01	4.23	9.88
Domestic	203,688	0.06	0.23	0.00	1.00
Common language	203,688	0.20	0.40	0.00	1.00
Contiguity	203,688	0.07	0.26	0.00	1.00
Colony	203,688	0.13	0.33	0.00	1.00
Rating	203,688	6.53	1.17	1.40	9.50
logRating	203,688	1.86	0.20	0.34	2.25
Age	203,688	9.30	9.43	1.00	100.00
logAge	203,688	1.85	0.86	0.00	4.61
Average Length	203,688	79.91	36.47	1.00	251.00
logAv. Length	203,688	4.22	0.65	0.00	5.53
Netflix Original	203,688	0.23	0.42	0.00	1.00
		TV Shows			
Variable	Obs.	Mean	Std. Dev.	Min	Max
Clicks	68,432	10,892	237,853	0.00	33,400,000
logDist.	68,432	8.41	1.04	4.23	9.80
Domestic	68,432	0.05	0.21	0.00	1.00
Common language	68,432	0.17	0.37	0.00	1.00
Contiguity	68,432	0.06	0.24	0.00	1.00
Colony	68,432	0.11	0.31	0.00	1.00
Rating	68,432	7.25	1.01	1.70	9.50
logRating	68,432	1.97	0.16	0.53	2.25
Age	68,432	6.85	7.26	1.00	68.00
logAge	68,432	1.58	0.80	0.00	4.22
Average Length	68,432	36.88	17.91	1.00	120.00
logAv. Length	68,432	3.48	0.55	0.00	4.79
Netflix Original	68,432	0.45	0.50	0.00	1.00
		Films			
Variable	Obs.	Mean	Std. Dev.	Min	Max
Clicks	135,256	30.95	1,904	0.00	394,453
logDist.	135,256	8.35	0.99	4.23	9.88
Domestic	135,256	0.06	0.24	0.00	1.00
Common language	135,256	0.22	0.41	0.00	1.00
Contiguity	135,256	0.08	0.27	0.00	1.00
Colony	135,256	0.14	0.34	0.00	1.00
Rating	135,256	6.17	1.08	1.40	9.30
logRating	135,256	1.80	0.20	0.34	2.23
Age	135,256	10.54	10.13	1.00	100.00
logAge	135,256	1.98	0.86	0.00	4.61
Length	135,256	101.68	20.73	2.00	251.00
logLength	135,256	4.60	0.25	0.69	5.53
Netflix Original	135,256	0.12	0.32	0.00	1.00

5 Results

In Table 5 we present our results for the number of available titles. Columns (1) and (2) present the results of estimating the gravity equation for content coming from all countries (including the destination country), the third column presents the results for foreign content only. Columns (4) and (5) keep the same sample of countries as in our intensive margin regressions, columns (6) and (7) exclude the US as country of origin as it has a disproportionately large share of titles in each country's catalogue which may affect our results.

Our results are, to some extent, in line with the gravity literature; a lot of the coefficients are, however, insignificant (in line with our second hypothesis). Among the gravity variables the domestic dummy, the number of migrants, and common language are all significant suggesting that content availability may be chosen to reflect viewers preferences for domestic titles. The magnitude of the domestic dummy varies between 58%[27] and 100% suggesting that the number of domestic titles available is on average between 58% and 100% larger than the number of other titles. The magnitude of the common language coefficient is slightly smaller (between 31% and 100%).

When we exclude the US and look at all content (foreign and domestic), the distance coefficient and colony dummy become significant suggesting that some of our initial results may be driven by American content. For this reason, we exclude it in our robustness checks for the intensive margin regressions. Our results are in line with Batikas et al. (2015) who find that content availability is driven by consumer preferences for home market products.

Table 5: Content availability (extensive margin)

	All countries			The same as in IM		Excl. US	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	all	all	foreign	all	foreign	all	foreign
logDist.	-0.097**	0.009	0.024	0.023	0.033	-0.116***	-0.069**
	(0.038)	(0.042)	(0.036)	(0.037)	(0.033)	(0.031)	(0.030)
Domestic		0.456***		0.529***		0.713***	
		(0.166)		(0.155)		(0.121)	
logMigrants			0.097***		0.102***		0.062***
			(0.021)		(0.023)		(0.014)
Colony	0.046	0.027	0.013	0.019	0.010	-0.129**	-0.226***
	(0.034)	(0.040)	(0.041)	(0.039)	(0.041)	(0.051)	(0.060)
Common language	0.273**	0.318***	0.132	0.354***	0.154	0.694***	0.564***
	(0.115)	(0.118)	(0.102)	(0.122)	(0.104)	(0.098)	(0.101)
Contiguity	-0.156	0.043	0.050	0.071	0.064	-0.039	0.011
	(0.097)	(0.102)	(0.102)	(0.098)	(0.101)	(0.056)	(0.061)
Constant	6.651***	5.726***	4.663***	5.776***	4.699***	5.347***	4.412***
	(0.308)	(0.356)	(0.356)	(0.311)	(0.353)	(0.255)	(0.280)
Observations	5,940	5,940	5,628	2,480	2,412	5,860	5,552
Pseudo R2	0.956	0.956	0.960	0.949	0.955	0.903	0.907
Log likelihood	-26963	-26573	-22721	-20231	-17240	-21396	-19267

PPML estimation results. Dependant variable: number of titles available. All results include origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time and destination-time level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Now we move on to the determinants of the number of clicks (intensive margin) which is the core of our analysis. Table 6 presents the results of our baseline specifications for TV shows.[28] We find that gravity variables have an impact of Netflix viewing (again in line with our second hypothesis). In particular, we find that distance to the production country has a negative effect on the number of clicks, while sharing a common language

[27] exp(0.456)-1 in column 2.

[28] See Table C.1 in Appendix C for TV shows and films combined.

increases it. The elasticity of the distance is initially -0.469 (column 1), but it is mitigated by including a dummy for domestic productions in columns 2-7. Both of these variables together suggest a strong preference for domestic content (home bias). The statistical significance of the negative impact of contiguity and past colonial relationship on the number of clicks vanishes as more explanatory variables are added to the estimation.

Among the remaining gravity variables, the effect of common language is the most important, sharing a common language increases the number of clicks by up to 388%,[29] depending on the specification. In addition, a domestic production receives between 79% and 114% more clicks on average.

Table 6: Baseline results, TV Shows (intensive margin)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
logDist.	-0.469***	-0.262***	-0.324***	-0.299***	-0.336***	-0.305***	-0.306***
	(0.058)	(0.092)	(0.106)	(0.096)	(0.104)	(0.100)	(0.099)
Domestic		0.666***	0.582**	0.755***	0.647**	0.753**	0.756***
		(0.246)	(0.258)	(0.253)	(0.308)	(0.295)	(0.275)
Common language	1.402***	1.498***	1.585***	1.531***	1.524***	1.519***	1.513***
	(0.220)	(0.201)	(0.203)	(0.201)	(0.231)	(0.229)	(0.210)
Contiguity	-0.604*	-0.181	-0.305	-0.183	-0.278	-0.158	-0.163
	(0.362)	(0.368)	(0.439)	(0.395)	(0.438)	(0.408)	(0.405)
Colony	-0.358**	-0.372***	-0.364**	-0.293*	-0.317	-0.296	-0.288
	(0.163)	(0.138)	(0.156)	(0.163)	(0.200)	(0.187)	(0.176)
logRating			10.197***	11.799***	11.931***	11.355***	11.305***
			(0.505)	(0.742)	(0.818)	(0.843)	(0.830)
logAge				-0.789***	-0.791***	-0.501**	-0.493***
				(0.158)	(0.178)	(0.207)	(0.191)
Children & Family					-1.740***	-1.728***	-1.997***
					(0.356)	(0.356)	(0.462)
Comedy					0.640**	0.616**	0.438
					(0.279)	(0.261)	(0.350)
Crime & Thriller					-0.597*	-0.561*	-0.546*
					(0.327)	(0.322)	(0.317)
Drama					-0.896*	-0.834*	-0.767
					(0.502)	(0.500)	(0.490)
Horror					-0.210	-0.192	-0.253
					(0.148)	(0.150)	(0.159)
Romance					-0.335	-0.171	-0.241
					(0.510)	(0.509)	(0.496)
Sci-Fi & Fantasy					0.356	0.383	0.314
					(0.403)	(0.404)	(0.436)
Netflix Original						0.790***	0.821***
						(0.179)	(0.162)
logAv.Length							-0.344
							(0.218)
Constant	14.151***	12.198***	-8.098***	-10.475***	-10.356***	-10.393***	-8.999***
	(0.422)	(0.805)	(1.116)	(1.479)	(1.314)	(1.212)	(1.014)
Observations	68,432	68,432	68,432	68,432	68,432	68,432	68,432
Pseudo R2	0.319	0.319	0.423	0.462	0.510	0.519	0.520
Log likelihood	-2.320e+09	-2.320e+09	-1.970e+09	-1.830e+09	-1.670e+09	-1.640e+09	-1.640e+09

PPML estimation results. Dependant variable: number of clicks. All results include origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time and destination-time level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

When measuring the effects on the intensive margin we can include additional variables specific to a given show (columns 3-7). Among these factors, the IMDb rating has a positive effect while the age of the show has a negative effect on the number of views. Even if the magnitude of the IMDb coefficient may seem large, it is important to remember that most

[29] exp(1.585)-1, in column (3).

titles are not viewed, and only a handful of titles are viewed millions of times. Therefore a 1% increase in ratings translating into an around 12% increase in the number of views does not necessarily imply a big increase in the number of clicks. Overall, these findings are consistent with our third hypothesis that content quality is an important driver of the number of clicks.

In addition, in columns (5)-(7) we include dummies for different genres (our reference category are shows belonging to the Action and Adventure genre). Including these dummies shows that comedies attract the largest number of clicks. Its effect, however, becomes insignificant when we control for the average length of a given show in column (7). In column (6) amd (7) we include a dummy for Netflix Original.[30] It shows that the effect of Netflix Original is very important (in line with our fourth hypothesis), increasing the number of clicks by around 120%. This finding is not surprising given the fact that, as documented by (Lobato, 2018) Netflix's algorithms are designed to promote Netflix Original content. Finally, in column (7) we add the average episode length. Its effect is negative and significant at 15% level only, nevertheless suggesting that viewers prefer shorter TV shows.

Moving on to foreign TV shows (Table 7), we can see that the magnitude of the distance coefficient becomes smaller and less significant in column (1), while the effect of common language remains about the same. Including the number of migrants (from the production country in the country of destination) in column (2) increases the size of the distance effect. The effect of migrants is positive and significant. Overall, all of our findings show the importance of domestic productions in Netflix watching. Viewers have a strong preference for TV shows produced in their home countries, and migrants choose to watch content produced in their country of origin. Colonial linkages are now highly significant and negative, suggesting a bias against content coming from countries sharing a colonial relationship in the past. When watching foreign shows, viewers are even more likely to choose Netflix Original content. The effect of Netflix Original dummy is larger - around 150%. Finally, the average episode length is not significant.

In the next step, we estimate our most complete specifications (columns 6 in Table 6 and Table 7) for films instead of TV shows. We present our results in Table 8. Here again, we estimate our model for all content and foreign content only. In addition, in columns (3) and (6), we include a dummy variable equal to one if a given film was co-produced by the country of destination.[31] We immediately see that the distance and other gravity variables are statistically insignificant. In particular, the fact that the coefficient on common language is insignificant comes as a surprise. A possible explanation is that even if the films catalogue is very large, we only observe 536 observations with positive number of clicks, the initial reason to split the analysis into TV shows and films.

Nevertheless, it is important to note that our results for films confirm our previous findings on the relevance of domestic productions (or co-productions) in determining the extent of Netflix watching. The magnitude of the domestic dummy is initially larger (albeit less significant) for films than for TV shows. Its effect becomes insignificant once we include a co-production dummy. The length of a film does not have a significant effect on views in either specification. The effect of migrant population on viewing patterns appears to be similar to the effect we have found for foreign TV shows, with an elasticity of between 0.48 and 0.53.

Regarding individual film characteristics, the effect of the IMDb rating is of a slightly lower magnitude for films than for TV shows (around 9-10.5%) while the absolute effect

[30] Note that we combine Netflix Original and Netflix Exclusive shows.
[31] Note that we do not have the data on co-productions of TV shows.

Table 7: Foreign TV Shows (intensive margin)

	(1)	(2)	(3)	(4)	(5)	(6)
logDist.	-0.198**	-0.272***	-0.292***	-0.296***	-0.285***	-0.284***
	(0.091)	(0.078)	(0.084)	(0.073)	(0.072)	(0.072)
logMigrants		0.417***	0.409***	0.426***	0.445***	0.441***
		(0.109)	(0.112)	(0.114)	(0.112)	(0.110)
Common language	1.514***	1.467***	1.534***	1.526***	1.562***	1.553***
	(0.293)	(0.246)	(0.247)	(0.255)	(0.255)	(0.246)
Contiguity	0.055	-0.656**	-0.672**	-0.633**	-0.590**	-0.584**
	(0.280)	(0.287)	(0.302)	(0.276)	(0.266)	(0.264)
Colony	-0.473***	-0.396***	-0.425***	-0.363***	-0.327***	-0.325***
	(0.092)	(0.090)	(0.101)	(0.116)	(0.110)	(0.111)
logRating			8.610***	10.129***	9.706***	9.652***
			(0.583)	(0.787)	(0.787)	(0.778)
logAge				-0.976***	-0.648***	-0.639***
				(0.229)	(0.246)	(0.238)
Children & Family				-2.012***	-1.998***	-2.202***
				(0.449)	(0.453)	(0.548)
Comedy				0.258	0.243	0.109
				(0.382)	(0.367)	(0.457)
Crime & Thriller				-0.570*	-0.552*	-0.538*
				(0.337)	(0.332)	(0.326)
Drama				-0.902	-0.862	-0.807
				(0.591)	(0.574)	(0.573)
Horror				-0.364	-0.382	-0.423
				(0.268)	(0.268)	(0.270)
Romance				-0.151	0.011	-0.047
				(0.715)	(0.712)	(0.691)
Sci-Fi & Fantasy				0.307	0.351	0.298
				(0.401)	(0.398)	(0.434)
Netflix Original					0.937***	0.957***
					(0.215)	(0.207)
logAv.Length						-0.272
						(0.279)
Constant	10.848***	6.785***	-10.485***	-12.173***	-12.747***	-11.581***
	(0.806)	(1.231)	(1.787)	(1.842)	(1.805)	(1.684)
Observations	64,809	64,809	64,809	64,809	64,809	64,809
Pseudo R2	0.246	0.248	0.344	0.452	0.465	0.466
Log likelihood	-1.240e+09	-1.230e+09	-1.070e+09	-8.980e+08	-8.770e+08	-8.750e+08

PPML estimation results. Dependant variable: number of clicks. All results include origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time and destination-time level. *** p<0.01, ** p<0.05, * p<0.1.

of age is larger, making older films relatively less attractive than TV shows with the same production year (of the first season). The size of the Netflix Original coefficient for films is much larger than for TV shows (around 400% vs. 150% respectively).

Table 8: Films (intensive margin)

		All			Foreign	
	(1)	(2)	(3)	(4)	(5)	(6)
logDist.	0.224	0.227	0.139	0.215	0.216	0.173
	(0.190)	(0.188)	(0.188)	(0.266)	(0.266)	(0.268)
Domestic	1.801*	1.805*	1.206			
	(1.010)	(1.004)	(1.100)			
logMigrants				0.483***	0.483***	0.529***
				(0.121)	(0.121)	(0.131)
Common language	1.085	1.086	0.719	0.366	0.369	0.034
	(0.858)	(0.854)	(0.896)	(0.301)	(0.299)	(0.342)
Contiguity	1.281	1.290	1.167	0.048	0.049	0.020
	(0.861)	(0.861)	(0.871)	(0.974)	(0.971)	(0.968)
Colony	0.154	0.158	0.164	-0.322	-0.315	-0.330
	(0.456)	(0.456)	(0.408)	(0.252)	(0.254)	(0.243)
logRating	9.030***	9.359***	9.285***	10.162***	10.456***	10.213***
	(1.868)	(2.218)	(2.125)	(2.339)	(2.703)	(2.563)
logAge	-1.790***	-1.791***	-1.783***	-1.852***	-1.846***	-1.780***
	(0.384)	(0.390)	(0.382)	(0.344)	(0.340)	(0.335)
Children & Family	1.251	1.028	0.985	0.354	0.191	0.224
	(1.043)	(0.991)	(0.904)	(0.604)	(0.720)	(0.704)
Comedy	-0.113	-0.248	-0.269	-0.943***	-1.025***	-0.936**
	(0.987)	(0.881)	(0.776)	(0.357)	(0.379)	(0.376)
Crime & Thriller	1.991*	1.956*	1.904*	0.808	0.826	0.897
	(1.133)	(1.096)	(1.031)	(0.844)	(0.834)	(0.827)
Drama	-1.278	-1.322	-1.360	-2.414**	-2.398**	-2.270**
	(1.118)	(1.087)	(1.010)	(1.048)	(1.013)	(0.937)
Horror	-0.534	-0.541	-0.603	-1.777***	-1.739***	-1.706***
	(0.913)	(0.875)	(0.786)	(0.254)	(0.259)	(0.266)
Romance	1.075	1.007	0.958	0.152	0.134	0.208
	(1.223)	(1.147)	(1.040)	(0.623)	(0.623)	(0.650)
Sci-Fi & Fantasy	3.134**	3.076***	2.944***	2.226***	2.217***	2.177***
	(1.264)	(1.183)	(1.067)	(0.699)	(0.694)	(0.692)
Netflix Original	1.655***	1.654***	1.632***	1.544***	1.545***	1.540***
	(0.366)	(0.373)	(0.371)	(0.402)	(0.406)	(0.429)
logLength		-0.329	-0.294		-0.317	-0.273
		(0.452)	(0.435)		(0.480)	(0.463)
Co-production			0.858*			1.067***
			(0.484)			(0.411)
Constant	-14.650***	-13.720***	-12.873***	-20.722***	-19.822***	-19.852***
	(3.853)	(2.697)	(2.436)	(5.219)	(4.095)	(4.038)
Observations	135,256	135,256	135,256	123,207	123,207	123,207
Pseudo R2	0.581	0.582	0.586	0.602	0.602	0.607
Log likelihood	-1.250e+07	-1.250e+07	-1.230e+07	-9.104e+06	-9.092e+06	-8.983e+06

PPML estimation results. Dependant variable: number of clicks. All results include origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time destination-time level. *** p<0.01, ** p<0.05, * p<0.1.

Since one of our research questions is to evaluate the effect of Netflix Original content on watching patterns, in the next step, we estimate our model for Netflix Original content and other content separately. We do this for TV shows only, as the data on films appears less reliable, and estimate our model for all TV shows (columns 1 and 2) and for foreign TV shows separately (columns 3 and 4). We present our results in Table 9.[32]

First thing to note is that excluding Netflix Original renders gravity variables more important. The coefficient on distance for non-original content ranges from -0.805 in column (1) to -1.093 in column (3). Common language and domestic dummy are also of a

[32] Note that, due to the dropping of singletons, our sample sizes are smaller than in Table 6 and Table 7.

Table 9: By Netflix Original, TV Shows (intensive margin)

	All		Foreign	
	(1)	(2)	(3)	(4)
	not nfo	nfo	not nfo	nfo
logDist.	-0.805***	-0.185**	-1.093***	-0.246***
	(0.177)	(0.087)	(0.215)	(0.060)
Domestic	3.342***	1.009***		
	(0.625)	(0.271)		
logMigrants			0.472**	0.533***
			(0.221)	(0.120)
Common language	2.912***	1.258***	3.564***	1.332***
	(0.476)	(0.188)	(0.346)	(0.262)
Contiguity	-1.218	0.326	-1.572	-0.514*
	(0.792)	(0.288)	(1.034)	(0.263)
Colony	0.407	-0.156	0.088	-0.122
	(0.319)	(0.142)	(0.324)	(0.090)
logRating	26.007***	6.909***	14.642***	7.093***
	(1.458)	(0.837)	(1.981)	(0.971)
logAge	0.284**	-0.878***	-0.162	-0.926***
	(0.134)	(0.333)	(0.290)	(0.307)
Children & Family	-0.458	-2.395***	-0.804	-2.676***
	(0.816)	(0.375)	(1.021)	(0.454)
Comedy	2.737***	-0.337	2.241***	-0.396
	(0.314)	(0.524)	(0.482)	(0.561)
Crime & Thriller	-1.232**	-0.769*	-0.402	-0.780**
	(0.586)	(0.400)	(0.507)	(0.386)
Drama	-1.735**	-1.015*	-2.378***	-1.081*
	(0.706)	(0.560)	(0.688)	(0.571)
Horror	-1.274***	-0.469**	-0.556	-0.603**
	(0.410)	(0.212)	(0.379)	(0.304)
Romance	1.293**	-1.947***	2.353***	-2.794***
	(0.576)	(0.458)	(0.642)	(0.490)
Sci-Fi & Fantasy	-1.281***	0.231	-0.574*	0.268
	(0.361)	(0.507)	(0.333)	(0.474)
logAv.Length	-1.128***	0.212	-1.237***	0.243
	(0.257)	(0.199)	(0.264)	(0.257)
Constant	-38.556***	-1.396	-15.662***	-7.932***
	(2.894)	(2.077)	(5.359)	(2.243)
Observations	29,056	30,059	26,561	28,776
Pseudo R2	0.705	0.547	0.618	0.483
Log likelihood	-3.820e+08	-8.740e+08	-1.810e+08	-5.340e+08

PPML estimation results. Dependant variable: number of clicks. All results include origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time and destination-time level. *** p<0.01, ** p<0.05, * p<0.1.

much greater magnitude for content that is not produced by Netflix. These results are not surprising, given our original assumption that Netflix favours its original content. When looking only at the non-original content, older titles seem to be in favour but this effect becomes insignificant once the sample is reduced to foreign productions only. For Netflix Original productions we observe the contrary: viewing decreases with age of the title.

Comparing columns (1) and (2), the size of the IMDb rating coefficient is between 3 and 4 times greater for TV shows that are not made by Netflix. Finally, non-original series attract fewer clicks when the average episode length is long, while this coefficient is insignificant for Netflix Original shows. This may also be explained by the fact that some Netflix Original shows have more of a film feeling and therefore length matters less.

6 Robustness checks

To check the robustness of our results we first exclude American TV shows as they are the most prevalent in Netflix catalogue across countries (Table 10). Columns (1)-(3) exclude the US as country of origin, columns (4)-(5) as destination, finally column (6) excludes the US as both origin and destination. Most of our results hold. In particular, we still observe a preference for domestic content and content in common language. Interestingly, the magnitude of the Netflix Original dummy becomes very large for TV shows produced outside the US.

In the second step, we check if our findings are robust to different measures of linguistic proxmity by using the measures developed by Melitz and Toubal (2014). The measures include common official language (col), common spoken language (csl), common native language (cnl), adjusted linguistic proximity between different native languages (lp), and common language (index).[33]

We run these specifications for foreign TV shows, as these measures are not developed for domestic trade. Our results still hold. In particular, each of the language measures has a positive effect on the number of clicks.

Finally, we estimate additional specifications using title fixed effects in addition to origin-time and destination-time fixed effects. These can be found in Appendix D.

[33] See Melitz and Toubal (2014) for a detailed description of these variables.

Table 10: Robustness check: excluding the US

		Excl. as origin		Excl. as destination		Excl. as both
	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	Foreign	All	Foreign	Foreign
logDist.	-0.528***	-0.168	-0.043	-0.145**	-0.178***	-0.164**
	(0.083)	(0.105)	(0.109)	(0.072)	(0.063)	(0.078)
Domestic		1.411***		0.767***		
		(0.316)		(0.222)		
logMigrants			0.538***		0.181***	0.399***
			(0.098)		(0.070)	(0.078)
Common language	1.577***	1.515***	1.450***	1.573***	1.693***	0.448**
	(0.365)	(0.353)	(0.336)	(0.223)	(0.277)	(0.211)
Contiguity	0.002	0.476*	-0.209	0.093	-0.148	0.016
	(0.286)	(0.262)	(0.228)	(0.268)	(0.180)	(0.158)
Colony	-0.502	-0.864**	-0.865**	-0.140	-0.245**	0.246
	(0.373)	(0.400)	(0.361)	(0.130)	(0.117)	(0.216)
logRating	12.986***	12.949***	12.813***	9.666***	9.589***	13.756***
	(2.004)	(2.001)	(2.076)	(0.801)	(0.749)	(2.763)
logAge	-1.131***	-1.132***	-1.039***	-0.655***	-0.617**	-1.097***
	(0.301)	(0.301)	(0.298)	(0.246)	(0.247)	(0.329)
Children & Family	-3.109***	-3.085***	-2.976***	-2.271***	-2.216***	-3.295***
	(1.028)	(1.013)	(1.152)	(0.568)	(0.574)	(1.083)
Comedy	-0.980	-0.965	-1.058	0.201	0.222	-1.103
	(0.689)	(0.684)	(0.797)	(0.446)	(0.463)	(0.784)
Crime & Thriller	-0.832	-0.825	-1.006*	-0.391	-0.414	-0.681
	(0.576)	(0.581)	(0.565)	(0.327)	(0.324)	(0.549)
Drama	-1.755**	-1.735**	-1.895**	-0.672	-0.666	-1.672**
	(0.771)	(0.775)	(0.749)	(0.580)	(0.593)	(0.707)
Horror	-0.492	-0.486	-0.531	-0.389**	-0.374**	-0.581
	(0.851)	(0.848)	(0.878)	(0.161)	(0.165)	(0.635)
Romance	-2.522***	-2.505***	-4.004***	0.186	0.131	-3.498***
	(0.641)	(0.642)	(0.682)	(0.602)	(0.658)	(0.643)
Sci-Fi & Fantasy	-0.202	-0.193	-0.276	0.366	0.377	-0.317
	(0.769)	(0.771)	(0.797)	(0.434)	(0.446)	(0.689)
Netflix Original	2.367***	2.417***	2.721***	0.824***	0.810***	2.167***
	(0.423)	(0.418)	(0.476)	(0.143)	(0.147)	(0.464)
logAv.Length	-0.879	-0.863	-0.851	-0.281	-0.265	-0.906
	(0.828)	(0.825)	(0.949)	(0.273)	(0.280)	(1.023)
Constant	-9.645*	-12.542**	-19.625***	-7.882***	-9.541***	-18.469***
	(5.039)	(5.023)	(5.257)	(1.391)	(1.259)	(6.760)
Observations	40,954	40,954	38,950	64,435	62,455	36,596
Pseudo R2	0.606	0.611	0.632	0.465	0.467	0.640
Log likelihood	-2.140e+08	-2.120e+08	-1.760e+08	-8.270e+08	-7.920e+08	-1.080e+08

PPML estimation results. Dependant variable: number of clicks. All results include origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time and destination-time level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 11: Robustness check: Melitz and Toubal linguistic proximity measures

	(1)	(2)	(3)	(4)	(5)	(6)
	col	csl	cnl	three	three + lp	index
logDist.	-0.284***	-0.349***	-0.297***	-0.323***	-0.349***	-0.366***
	(0.072)	(0.080)	(0.065)	(0.070)	(0.060)	(0.057)
logMigrants	0.441***	0.464***	0.458***	0.449***	0.467***	0.470***
	(0.110)	(0.115)	(0.111)	(0.112)	(0.106)	(0.100)
Common language	1.553***			0.731**	0.873***	
	(0.246)			(0.337)	(0.307)	
Common spoken language		3.160***		1.227***	-0.375	
		(0.547)		(0.426)	(0.576)	
Common native language			2.410***	0.682	2.267***	
			(0.375)	(0.547)	(0.775)	
Adjusted Linguistic Proximity					0.429***	
					(0.108)	
Common language Index						3.037***
						(0.390)
Contiguity	-0.584**	-0.697**	-0.623**	-0.734***	-0.878***	-0.858***
	(0.264)	(0.279)	(0.248)	(0.275)	(0.239)	(0.207)
Colony	-0.325***	-0.070	-0.460***	-0.296***	-0.366***	-0.286**
	(0.111)	(0.118)	(0.121)	(0.106)	(0.120)	(0.113)
logRating	9.652***	9.638***	9.652***	9.651***	9.657***	9.652***
	(0.778)	(0.773)	(0.776)	(0.777)	(0.777)	(0.774)
logAge	-0.639***	-0.642***	-0.639***	-0.640***	-0.639***	-0.640***
	(0.238)	(0.238)	(0.238)	(0.238)	(0.237)	(0.237)
Children & Family	-2.202***	-2.207***	-2.203***	-2.204***	-2.202***	-2.203***
	(0.548)	(0.549)	(0.548)	(0.548)	(0.548)	(0.549)
Comedy	0.109	0.107	0.109	0.109	0.111	0.111
	(0.457)	(0.457)	(0.457)	(0.457)	(0.457)	(0.456)
Crime & Thriller	-0.538*	-0.533	-0.538*	-0.537	-0.541*	-0.540*
	(0.326)	(0.328)	(0.326)	(0.328)	(0.326)	(0.326)
Drama	-0.807	-0.806	-0.808	-0.807	-0.809	-0.809
	(0.573)	(0.576)	(0.574)	(0.577)	(0.577)	(0.574)
Horror	-0.423	-0.423	-0.423	-0.423	-0.423	-0.422
	(0.270)	(0.272)	(0.269)	(0.270)	(0.271)	(0.272)
Romance	-0.047	-0.037	-0.045	-0.043	-0.045	-0.043
	(0.691)	(0.686)	(0.690)	(0.690)	(0.692)	(0.690)
Sci-Fi & Fantasy	0.298	0.296	0.299	0.298	0.300	0.299
	(0.434)	(0.433)	(0.433)	(0.434)	(0.434)	(0.434)
Netflix Original	0.957***	0.955***	0.959***	0.958***	0.955***	0.953***
	(0.207)	(0.211)	(0.208)	(0.208)	(0.208)	(0.208)
logAv.Length	-0.272	-0.278	-0.273	-0.274	-0.270	-0.271
	(0.279)	(0.280)	(0.279)	(0.279)	(0.279)	(0.279)
Constant	-11.581***	-12.726***	-11.598***	-11.971***	-11.808***	-11.993***
	(1.684)	(1.819)	(1.672)	(1.707)	(1.666)	(1.736)
Observations	64,809	64,809	64,809	64,809	64,809	64,809
Pseudo R2	0.466	0.465	0.466	0.466	0.467	0.467
Log likelihood	-8.750e+08	-8.760e+08	-8.750e+08	-8.740e+08	-8.730e+08	-8.730e+08

PPML estimation results. Dependant variable: number of clicks. All results include origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time and destination-time level. *** p<0.01, ** p<0.05, * p<0.1.

7 Conclusion

Video on demand services play an increasingly important role in audiovisual content's production, distribution and viewing patterns. This paper examines the factors determining both content availability (extensive margin) and viewing (intensive margin) of Netflix from the perspective of international trade using a novel data set covering Netflix catalogues and viewing across twenty countries. In particular, we propose a number of hypotheses that are tested in our empirical analysis. Our first key result is that Netflix catalogues are very different across countries. Our second result is that, despite the fact that Netflix does not require any physical transport of its service, gravity variables matter for Netflix availability and viewing with distance, domestic dummy and common language being the most important among them. Comparing the extensive and intensive margin results, our findings suggest, however, that gravity variables matter more for the intensive margin than for the extensive margin, confirming our initial hypothesis. This finding is not surprising since content viewing can be seen as a two-sided transaction (in which the exporter and importer have agreed to the trade) normally estimated in a trade regression whereas content availability is one sided (i.e. where Netflix is only a potential exporter and viewers have not yet agreed to accept the trade). Beyond the role of gravity variables we find that the Netflix-viewing depends on show's/film's rating, age, genre and length. Finally, Netflix Original productions attract a disproportionately large number of clicks.

Our findings suggest several important policy implications. In particular, since the share of the EU productions in Netflix scripted content is still below the proportion suggested by the European Audiovisual Media Services Directive, and also below some more strict national regulations, and since Netflix users seem to have a strong preference for domestic productions, more effort should be directed by Netflix towards producing and promoting (using specific algorithms) European content in Netflix catalogues across European countries. At the same time, both national and EU regulators should make an effort to enforce local content regulations.

References

[1] Aguiar, L. & Waldfogel, J. (2018). Netflix: global hegemon or facilitator of frictionless digital trade? *Journal of Cultural Economics*, 42(3), 419-445.

[2] Alaveras, G., Gomez-Herrera, E. & Martens, B. (2018). Cross-border circulation of films and cultural diversity in the EU. *Journal of Cultural Economics*, 42(4), 645-676.

[3] Batikas, Gomez-Herrera and Martens (2015). Film Availability in Netflix Country Stores in the EU, JRC Digital Economy working paper nr 2015-11.

[4] BBC (2020). Netflix to cut streaming quality in Europe for 30 days. https://www.bbc.com/news/technology-51968302

[5] Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2012). The empirics of firm heterogeneity and international trade. *Annu. Rev. Econ.*, 4(1), 283-313.

[6] Burroughs, B. (2019). House of Netflix: Streaming media and digital lore. *Popular Communication*, 17(1), 1-17.

[7] Chen, N., & Juvenal, L. (2016). Quality, trade, and exchange rate pass-through. *Journal of International Economics*, 100, 61-80.

[8] Crozet, M., Head, K., & Mayer, T. (2012). Quality sorting and trade: Firm-level evidence for French wine. *The Review of Economic Studies*, 79(2), 609-644.

[9] DigitalTV Europe (2020). Western Europe to approach 200 million SVOD subs by 2025 with Netflix as clear leader.
https://www.digitaltveurope.com/2020/09/21/western-europe-to-approach-200-million-svod-subs-by-2025-with-netflix-as-clear-leader/
(Last retrieved February 22, 2021).

[10] Disdier, A. C., Tai, S. H., Fontagné, L., & Mayer, T. (2010). Bilateral trade of cultural goods. *Review of World Economics*, 145(4), 575-595.

[11] European Audiovisual Observatory (2020). Trends in the VOD market in EU28. https://rm.coe.int/trends-in-the-vod-market-in-eu28-final-version/1680a1511a
(Last retrieved February 22, 2021).

[12] Ferreira, F., & Waldfogel, J. (2013). Pop internationalism: Has half a century of world music trade displaced local culture? *The Economic Journal*, 123(569), 634-664.

[13] Correia, S., Guimarães, P., & Zylkin, T. (2019). PPMLHDFE: Stata module for Poisson pseudo-likelihood regression with multiple levels of fixed effects.

[14] Hallak, J. C. (2006). Product quality and the direction of trade. *Journal of International Economics*, 68(1), 238-265.

[15] Hanson, G., & Xiang, C. (2011). Trade barriers and trade flows with product heterogeneity: An application to US motion picture exports. *Journal of International Economics*, 83(1), 14-26.

[16] Hellmanzik, C., & Schmitz, M. (2015). Virtual proximity and audiovisual services trade. *European Economic Review*, 77, 82-101.

[17] Hummels, D., & Klenow, P. J. (2005). The variety and quality of a nation's exports. *American Economic Review*, 95(3), 704-723.

[18] Lobato, R. (2018). Rethinking international TV flows research in the age of Netflix. *Television & New Media*, 19(3), 241-256.

[19] Lobato, R., Meese, J., Rugg, A., & Burroughs, B. (2016). Geoblocking and Video Culture. Amsterdam: Institute of Network Cultures

[20] Lotz, A., Lobato, R., & Thomas, J. (2018). Internet-distributed television research: a provocation. *Media Industries*, 5(2), 35-47.

[21] Manova, K., & Zhang, Z. (2009). *Quality heterogeneity across firms and export destinations*. National Bureau of Economic Research.

[22] Marvasti, A., & Canterbery, E. R. (2005). Cultural and other barriers to motion pictures trade. *Economic Inquiry*, 43(1), 39-54.

[23] Mayer, T., & Ottaviano, G. I. (2008). The happy few: The internationalisation of european firms. *Intereconomics*, 43(3), 135-148.

[24] McKenzie, J. (2021). The economics of movies (revisited): A decade of literature in review. SSRN. http://ssrn.com/abstract=3924209

[25] Melitz, J., & Farid Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, 93(2), 351-363.

[26] Netflix (2019). Producing for the world: Netflix launches its first European production hub in Madrid. https://about.netflix.com/en/news/producing-for-the-world-netflix-launches-its-first-european-production-hub-in-madrid (Last retrieved March 19, 2021).

[27] Netflix (2020). Netflix opens new Paris office and increases investment in France. https://about.netflix.com/en/news/netflix-opens-new-paris-office-and-increases-investment-in-france (Last retrieved March 19, 2021).

[28] Sandvine (2019). The Global Internet Phenomena Report September 2019.

[29] Santos Silva, J.M.C. & Tenreyro, S. (2006). The log of Gravity. *The Review of Economics and Statistics*, 88(4), 641-658.

[30] Schott, P. K. (2004). Across-product versus within-product specialization in international trade. *The Quarterly Journal of Economics*, 119(2), 647-678.

[31] Statista (2020). Statista Digital Market Outlook - Video Streaming (SVOD).

Appendix: Supporting Tables and Graphs

A Catalogue Overlap

Looking at the country of origin of shows and films is not sufficient to fully understand the similarity of Netflix catalogues between countries. To do that, we plot heat-maps displaying the overlap of catalogues by country and content type (Figure A.1 and Figure A.2). Looking at these figures, we can see that the overlap was higher for TV shows than for films. Spain, Italy and Finland had the most distinct catalogues, while Denmark, Norway and Poland were the countries with the highest degree of overlap (for both TV shows and films). For example the Danish catalogue overlapped the most with the catalogues of Sweden, Norway and France. English speaking countries, in particular the US, had also very distinct catalogues.[34] The US catalogue overlapped with the Canadian one, but the degree of overlap was much lower, than for example, the degree of overlap between the Danish and French catalogues. We observe a similar pattern for German speaking countries. The German catalogue was very similar to the Swiss one but its overlap with the Austrian catalogue was lower than its overlap with the French catalogue. Overall, even though there seems to be an overlap between catalogues in Scandinavian countries, language and cultural proximity do not seem to be the only drivers of title availability. Fragmentation of Netflix catalogues appears to be larger than what we would see by looking at the number of titles available by country.

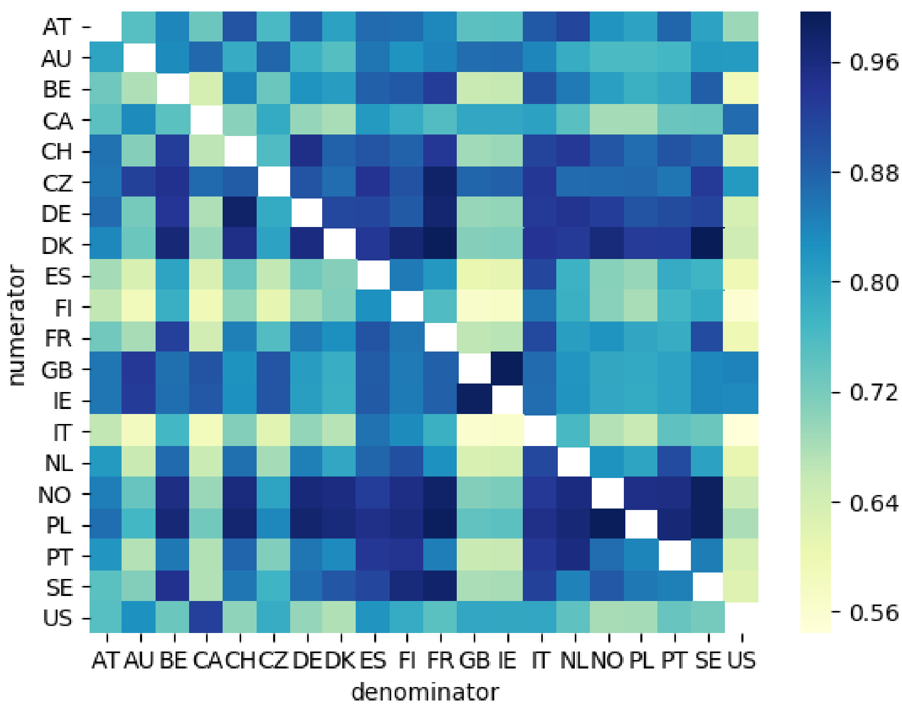

Figure A.1: TV Shows catalogues overlap by country, 2019 H2

[34] Looking at Figure 1 this does not appear to be driven by catalogue size.

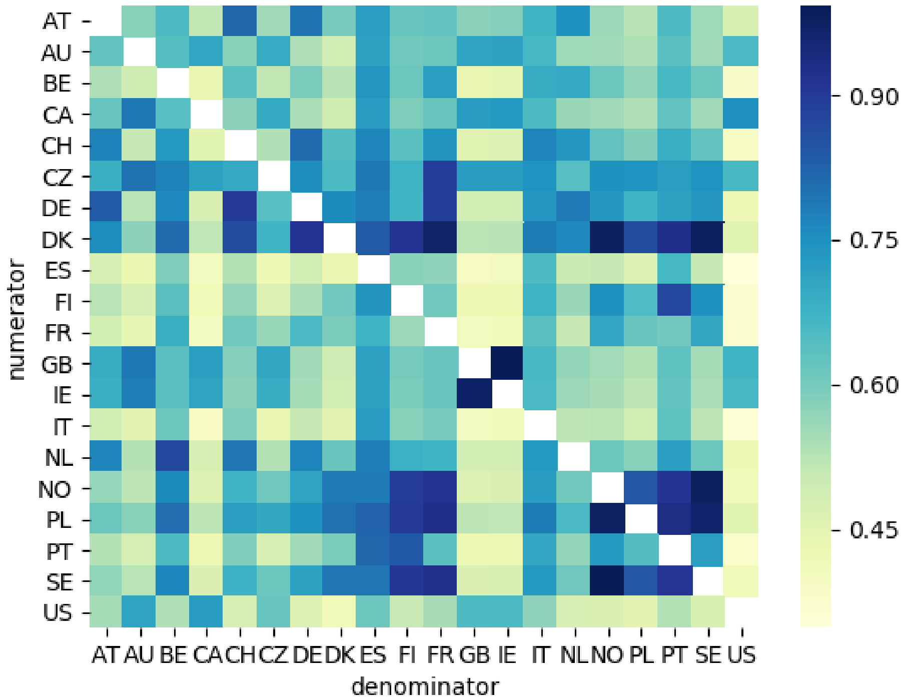

Figure A.2: Film catalogues overlap by country, 2019 H2

B Additional statistics

Table B.1: List of countries of origin

Argentina	Georgia	Luxembourg	Saudi Arabia
Australia	**Germany**	Malaysia	Singapore
Austria	Ghana	Malta	Slovak Republic
Bahrain	Greece	**Mexico**	Slovenia
Bangladesh	Guatemala	Morocco	**South Africa**
Belgium	Hong Kong	**Netherlands**	**South Korea**
Brazil	Hungary	**New Zealand**	**Spain**
Bulgaria	Iceland	Nigeria	Sri Lanka
Cambodia	**India**	Norway	**Sweden**
Canada	**Indonesia**	Pakistan	Switzerland
Chile	Iran	Panama	Taiwan
China	**Ireland**	Paraguay	Thailand
Colombia	**Israel**	Peru	**Turkey**
Croatia	**Italy**	Philippines	UAE
Czechia	**Japan**	**Poland**	**UK**
Denmark	**Jordan**	Portugal	**USA**
Egypt	Kazakhstan	Puerto Rico	Ukraine
Estonia	Kenya	Qatar	Uruguay
Finland	Kuwait	Romania	Venezuela
France	Lebanon	**Russia**	Vietnam

Countries in **bold** are included in the intensive margin regressions.

Table B.2: Summary statistics, full sample

Variable	Obs.	Mean	Std. Dev.	Min	Max
Extensive margin	6,400	44.93	197.66	0.0	2,464
Clicks	287,586	2,657.52	116,509.1	0.0	33,400,000
Domestic	287,586	0.05	0.22	0.00	1.00
Common language	287,586	0.22	0.41	0.00	1.00
logMigrants	271,596	10.36	1.82	1.95	16.26
logDist	287,586	8.4	1.03	4.23	9.88
Contiguity	287,586	0.07	0.25	0.00	1.00
Colony	287,586	0.11	0.31	0.00	1.00
Rating	275,189	6.46	1.19	1.30	9.50
logRating	275,189	1.85	0.21	0.26	2.25
Age	287,533	8.82	9.07	1.00	107
logAge	287,533	1.81	0.83	0.00	4.67
Average Length	286,181	85.47	36.95	1.00	251.00
logAv. Length	286,181	4.30	0.63	0.00	5.53
Netflix Original	287,586	0.19	0.39	0.00	1.00

C Pooled regressions for TV shows and films

In this section we present our results for TV shows and films combined. Table C.1 shows that the coefficients are similar to the coefficients in Table 6 and therefore are mostly driven by TV shows watching. The negative and significant coefficient for length further confirms that viewers prefer shorter content (TV shows) over longer content (films).

Table C.1: TV Shows and films

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
logDist.	-0.456***	-0.249***	-0.348***	-0.295***	-0.325***	-0.283***	-0.296***
	(0.055)	(0.086)	(0.113)	(0.094)	(0.104)	(0.097)	(0.096)
Domestic		0.670***	0.583**	0.804***	0.728**	0.859***	0.803***
		(0.225)	(0.259)	(0.239)	(0.294)	(0.279)	(0.266)
Common language	1.178***	1.277***	1.525***	1.488***	1.476***	1.482***	1.477***
	(0.207)	(0.188)	(0.203)	(0.190)	(0.215)	(0.218)	(0.198)
Contiguity	-0.542	-0.116	-0.323	-0.168	-0.257	-0.082	-0.138
	(0.351)	(0.341)	(0.463)	(0.382)	(0.423)	(0.386)	(0.390)
Colony	-0.254*	-0.268**	-0.272*	-0.237	-0.269	-0.256	-0.253
	(0.153)	(0.128)	(0.159)	(0.154)	(0.185)	(0.179)	(0.169)
logRating			14.408***	15.738***	16.051***	14.815***	13.112***
			(0.479)	(0.738)	(0.853)	(0.924)	(0.820)
logAge				-1.082***	-1.075***	-0.634***	-0.534***
				(0.143)	(0.154)	(0.188)	(0.170)
Children & Family					-1.519***	-1.484***	-2.272***
					(0.389)	(0.373)	(0.427)
Comedy					1.002***	0.878***	0.323
					(0.286)	(0.260)	(0.303)
Crime & Thriller					-0.324	-0.377	-0.370
					(0.336)	(0.318)	(0.319)
Drama					-0.935*	-0.915*	-0.657
					(0.522)	(0.507)	(0.484)
Horror					0.020	-0.038	-0.232
					(0.172)	(0.167)	(0.156)
Romance					-0.227	-0.131	-0.339
					(0.501)	(0.517)	(0.493)
Sci-Fi & Fantasy					0.521	0.519	0.285
					(0.395)	(0.396)	(0.409)
Netflix Original						1.210***	1.072***
						(0.193)	(0.178)
logAv.Length							-0.964***
							(0.137)
Constant	12.651***	10.692***	-17.251***	-18.670***	-19.215***	-18.373***	-10.907***
	(0.408)	(0.752)	(1.171)	(1.507)	(1.440)	(1.319)	(0.997)
Observations	203,688	203,688	203,688	203,688	203,688	203,688	203,688
Pseudo R2	0.197	0.197	0.435	0.509	0.556	0.571	0.589
Log likelihood	-3.400e+09	-3.400e+09	-2.390e+09	-2.080e+09	-1.880e+09	-1.810e+09	-1.740e+09

PPML estimation results. Dependant variable: number of clicks. All results include origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time and destination-time level. *** p<0.01, ** p<0.05, * p<0.1.

D Regressions with title fixed effects

In this section we present results for estimations that include tile fixed effects in addition to the gravity variables. Coefficients for title-specific attributes such as genre, Netflix Original dummy or rating are therefore omitted. For all regressions, the effect of common language is very similar to baseline regressions in Tables 6 and 7. For the full sample, we observe two changes: the distance has a less negative effect, while the domestic dummy coefficient becomes significantly larger. For the sample only including foreign TV shows, the decrease of the distance coefficient is somewhat smaller.

Table D.1: Title fixed effects, TV Shows

	All		Foreign	
	(1)	(2)	(3)	(4)
logDist.	-0.522***	-0.186**	-0.181**	-0.257***
	(0.053)	(0.076)	(0.081)	(0.064)
Domestic		1.124***		
		(0.237)		
logMigrants				0.419***
				(0.093)
Common language	1.308***	1.484***	1.495***	1.464***
	(0.225)	(0.174)	(0.262)	(0.220)
Contiguity	-0.674*	-0.021	0.192	-0.492**
	(0.369)	(0.286)	(0.222)	(0.228)
Colony	-0.179	-0.215*	-0.324***	-0.235**
	(0.168)	(0.123)	(0.098)	(0.096)
Constant	16.913***	13.717***	12.767***	8.689***
	(0.399)	(0.687)	(0.705)	(1.040)
Observations	25,238	25,238	23,850	23,850
Pseudo R2	0.786	0.788	0.710	0.713
Log likelihood	-5.700e+08	-5.660e+08	-3.550e+08	-3.510e+08

PPML estimation results. Dependant variable: number of clicks. All results include title, origin-time and destination-time fixed effects. Standard errors are clustered at the origin-time and destination-time level. *** p<0.01, ** p<0.05, * p<0.1.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP21/02 K Peren Arin, Kevin Devereux and Mieszko Mazur: 'Taxes and Firm Investment' January 2021
WP21/03 Judith M Delaney and Paul J Devereux: 'Gender and Educational Achievement: Stylized Facts and Causal Evidence' January 2021
WP21/04 Pierluigi Conzo, Laura K Taylor, Juan S Morales, Margaret Samahita and Andrea Gallice: 'Can ❤s Change Minds? Social Media Endorsements and Policy Preferences' February 2021
WP21/05 Diane Pelly, Michael Daly, Liam Delaney and Orla Doyle: 'Worker Well-being Before and During the COVID-19 Restrictions: A Longitudinal Study in the UK' February 2021
WP21/06 Margaret Samahita and Leonhard K Lades: 'The Unintended Side Effects of Regulating Charities: Donors Penalise Administrative Burden Almost as Much as Overheads' February 2021
WP21/07 Ellen Ryan and Karl Whelan: 'A Model of QE, Reserve Demand and the Money Multiplier' February 2021
WP21/08 Cormac Ó Gráda and Kevin Hjortshøj O!Rourke: 'The Irish Economy During the Century After Partition' April 2021
WP21/09 Ronald B Davies, Dieter F Kogler and Ryan Hynes: 'Patent Boxes and the Success Rate of Applications' April 2021
WP21/10 Benjamin Elsner, Ingo E Isphording and Ulf Zölitz: 'Achievement Rank Affects Performance and Major Choices in College' April 2021
WP21/11 Vincent Hogan and Patrick Massey: 'Soccer Clubs and Diminishing Returns: The Case of Paris Saint-Germain' April 2021
WP21/12 Demid Getik, Marco Islam and Margaret Samahita: 'The Inelastic Demand for Affirmative Action' May 2021
WP21/13 Emmanuel P de Albuquerque: 'The Creation and Diffusion of Knowledge - an Agent Based Modelling Approach' May 2021
WP21/14 Tyler Anbinder, Dylan Connor, Cormac Ó Gráda and Simone Wegge: 'The Problem of False Positives in Automated Census Linking: Evidence from Nineteenth-Century New York's Irish Immigrants' June 2021
WP21/15 Morgan Kelly: 'Devotion or Deprivation: Did Catholicism Retard French Development?' June 2021
WP21/16 Bénédicte Apouey and David Madden: 'Health Poverty' July 2021
WP21/17 David Madden: 'The Dynamics of Multidimensional Poverty in a Cohort of Irish Children' August 2021
WP21/18 Vessela Daskalova and Nicolaas J Vriend: 'Learning Frames' August 2021
WP21/19 Sanghamitra Chattopadhyay Mukherjee: 'A Framework to Measure Regional Disparities in Battery Electric Vehicle Diffusion in Ireland' August 2021
WP21/20 Karl Whelan: 'Central Banks and Inflation: Where Do We Stand and How Did We Get Here?' August 2021
WP21/21 Tyler Anbinder, Cormac Ó Gráda and Simone Wegge: '"The Best Country in the World": The Surprising Social Mobility of New York's Irish Famine Immigrants' August 2021
WP21/22 Jane Dooley and David Madden: 'Ireland's Post Crisis Recovery, 2012-2019: Was It Pro-Poor?' September 2021
WP21/23 Matthew Shannon: 'The Impact of Victimisation on Subjective Well-Being' September 2021
WP21/24 Morgan Kelly: 'Persistence, Randomization, and Spatial Noise' October 2021 *(For revised version of this paper see WP21/25)*
WP21/25 Morgan Kelly: 'Persistence, Randomization, and Spatial Noise' November 2021
WP21/26 Eliane Badaoui and Frank Walsh: 'Productivity, Non-Compliance and the Minimum Wage' November 2021

UCD Centre for Economic Research
Email economics@ucd.ie