Vertical Integration in Amazon Marketplace

Luca Bennati *

March 18, 2024

Abstract

By vertically integrating, Amazon enters as a seller on its own marketplace, thereby competing with the third-party (3P) sellers it hosts. I investigate how this dual role affects sellers' prices and product offerings in the headphones market. The demand estimates show that demand increases for products carried by Amazon and by 3P-sellers using Amazon's logistics (FBA). Moreover, analyzing the marginal costs and the distribution of market power across products, results suggest that Amazon's entry is driven by both efficiency and profitability reasons. Finally, I find that Amazon's weekly entry cost ranges between 56$ and 297$ for each product it offers.

JEL Classification: L10, L40, C57

Keywords: Competition Policy, Platforms, Entry Games

---

*Toulouse School of Economics. luca.bennati@tse-fr.eu

I am grateful to my advisors Christian Bontemps, Daniel Ershov and Bruno Jullien for their support and guidance. I am also grateful to Michele Bisceglia, Alexandre De Corniere, Olivier De Groote, Isis Durrmeyer, Oscar Jara, Chuqing Jin, Nicolas Martinez, and seminar participants at Toulouse School of Economics and Compass Lexecon for their useful comments and discussion.

1 Introduction

Digital platforms have taken a central stage in the modern economy. From the news we read to the clothes we buy and the booking of a physician appointment, much of this passes through many types of platforms. In e-commerce, digital marketplaces have allowed to remove physical constraints in terms of number of products and sellers they can host. It is no surprise then that many retailers have opened they digital stores to outside independent merchants, also called third-party (3P) sellers. In this way, the marketplace becomes vertically integrated, that is, the owner of the platform operates, and competes, with the 3P-sellers it hosts. Some well-known examples include Apple's and Google's App Stores, Walmart, JD.com, Mercado Libre and, of course, Amazon.

This dual role brings many potential benefits in terms of larger product choice. However, there is a concern that some platforms may have abused their dominant position or employed anti-competitive practices harming 3P-sellers. For instance, both Apple and Google have faced legal challenges over abusive rules and excessive fees in their App store [1]. Other practices which have been under the lens of Antitrust authorities are self-preferencing and use of 3P-sellers' private data [2]

Recent theoretical has suggested that vertical integration could be motivated by different incentives. On the one hand, the marketplace owner may enter for efficiency reasons, in order to take advantage of its cost efficiencies (Etro [2021]), to reduce double marginalization (Etro [2021]), or to regulate 3P-sellers' prices (Jeon and Rey [2022]). On the other, the marketplace owner may enter to foreclose rival 3P-sellers by self-preferencing (Hervas-Drane and Shelegia [2021], Hagiu et al. [2022], de Cornière and Taylor [2019]),

[1] The EU Commission has recently fined Apple over abusive rules in the music apps sectors (https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1161). Epic Games won another antitrust case against Google for abuse of dominant position (https://www.reuters.com/legal/google-epic-games-face-off-app-antitrust-trial-nears-end-2023-12-11/

[2] In 2017 the EU Commission fined Google for giving an unfair advantage to its own shopping comparison service (https://ec.europa.eu/commission/presscorner/detail/en/IP_17_1784. In 2022, Amazon reached an agreement with the European Commission to limit the use of non-public 3P-sellers' data, to limit self-preferencing in the Buy-Box algorithm and in assigning the Prime badge (see European Commission Press Release: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7777

by raising the fees (Anderson and Bedre-Defolie [2021]) or copy innovative 3P-sellers' products (Hagiu et al. [2022], Madsen and Vellodi [2021]). However, even when these last incentives prevail, the effects remain mixed for consumers.

Therefore, whether 3P-sellers should be protected from marketplace owner competition in a vertically integrated platform remains an open empirical question.

In this paper, I investigate this question by developing a structural model which incorporates both sellers' pricing and entry decisions and estimate it on data from Amazon marketplace. In particular, the focus will be on sellers' entry in terms of which products to offer in a market.

I will use the estimated structural model to assess the impact of limiting the owner ability to operate as a seller on its own marketplace. If the benefit of vertical integration for consumers was substantial, limiting vertical integration would be negative for consumers. Yet, if the contribution to consumers surplus was small, Antitrust authorities and regulators could advance policies to limit vertical integration in order to protect 3P-sellers. Hence, accounting for sellers' entry becomes very relevant in order to assess the overall impact on product offerings and prices.

The model has two stages. In the first stage Amazon and 3P-sellers choose their portfolio of products in order to maximize total profit, the difference between variable profit and the entry cost of offering a portfolio of products. Critically, given a set of existing products, sellers' can potentially choose to offer any of these products by operating as downstream retailers. In the second stage, sellers choose prices to maximize variable profits. Both pricing and entry decisions are assumed to be static. For the demand model, I use a Logit specification. The main feature of the supply model is that it includes the referral fees, the percentage of the final price which 3P-sellers pay to Amazon. In the entry part, I model entry costs as depending from a parameter constant across products, but heterogeneous across sellers.

I estimate demand using the headphones market in Amazon.com between March 2023 and September 2023. Then, I estimate the entry cost for Amazon using a moment inequalities approach from Pakes [2010] and inference from Chernozhukov et al. [2019].

In the first results of the paper, I find that both marketplace incentives for vertical integration may hold together within the same market.

When we analyse the distribution of the Lerner Index across different products sold by Amazon and 3P-sellers, we can notice a substantial difference. While the 3P-sellers' Index distribution is more uniform between 1% and 31%, the distribution of Amazon's Index is much more skewed to the right: Amazon will have limited market power on a majority of product, with a Lerner Index below 15%, but it will also enjoy larger market power, with a Lerner Index above 30%, for a small part of the products it offers.
A possible source of Amazon market power could also be consumers preference for Amazon. In fact, demand increases when Amazon is the seller of the product, and, while sellers adopting Amazon logistics (FBA) experience a positive increase in demand too, this is lower than Amazon's.
In addition to this, I investigate whether Amazon may also have some cost efficiencies compared to 3P-sellers. To do so, I compare the costs of products sold in the same week both by Amazon and 3P-sellers. It turns out that Amazon is not systematically more efficient than 3P-sellers, but rather the opposite, as on average 3P-sellers marginal cost is 15$ lower than Amazon's.
Overall, these results suggest that vertical integration could be explained both by efficiency reasons (regulate 3P-sellers' prices, increase product availability) and profitability reasons (take advantage of larger demand, expropriate 3P-sellers' profit).
On the entry side, I find that Amazon entry costs at the weekly level are bounded between 56$ and 297$ for each product. These bounds suggest that there might be heterogeneity in entry costs across products too.
Finally, in the first counterfactual simulations, I evaluate the effect of banning Amazon's products (both private labels and those sold as retailer) when 3P-sellers' entry decisions are fixed. As expected, I find that consumers and Amazon are harmed by such a policy: consumer surplus and Amazon profits decrease by 17.83% and 18.66%, respectively.

Literature. This paper is related to the theoretical literature on two-sided markets started with Caillaud and Jullien [2003], Rochet and Tirole [2003] and Armstrong [2006]. This literature has recently focused on studying platforms decisions to operate as a pure intermediary, as pure seller, or in a dual mode (Condorelli et al. [2018], Etro [2021], Jiang et al. [2011b], Hagiu and Wright [2015], Hagiu et al. [2020]). Different papers on dual mode platforms have looked at the effect of self-preferencing (de Cornière and Taylor [2019], Hagiu et al. [2022], Hervas-Drane and Shelegia [2021]), imitation (Hagiu et al. [2022], Madsen and Vellodi [2021]) and fees (Anderson and Bedre-Defolie [2021]) on 3P-sellers' and consumers, and what are the platform's incentive to regulate 3P-sellers' prices (Jeon and Rey [2022]).

On the empirical side, the literature has mainly focused on pricing in Amazon marketplace. First, there is evidence that the recommendation system tends to give higher visibility to Amazon (Chen and Tsai [2021], Lee and Musolff [2023], Lam [2023]). However, Lee and Musolff [2023] and Lam [2023] find that the recommendation system seems to tighten price competition too. More recent paper have been trying to measure more in detail under which circumstances self-preferencing is harmful for consumers (Farronato et al. [2023], Reimers and Waldfogel [2023]).

Then, Chen and Tsai [2023] show that Amazon holds better information on rivals and exploits it in its pricing, while Gutierrez [2022] finds some evidence of Amazon internalizing platform's network externalities, as Amazon seems to take into account consumer surplus.

I contribute to this literature in two ways. First, my results tend to confirm that 3P-sellers market power is limited, and this is the case for many products offered by Amazon too, consistent with the evidence found in Lee and Musolff [2023], Lam [2023] and Gutierrez [2022]. However, I also find that Amazon holds a large market power on a smaller, but significant, number of products it offers. Moreover, as in Gutierrez [2022], I find that, when entry is exogenous, a ban on Amazon's products is negative for consumers. Secondly, this is the first paper estimating the entry cost of Amazon.

Finally, this paper is related to the empirical literature in entry games studying product variety and using moment moment inequalities (Canay et al. [2023], Fan and Yang [2022], Wollmann [2018]).

Roadmap. In Section 2, I describe the general structure of Amazon marketplace and the data I am using. In Section 3, I will provide some descriptive evidence on the headphones market. Then, in Section 4, I present the structural model and, in Section 5, the estimation strategy. Finally, in Section 6, I will present and discuss the results.

2 Setting and Data

2.1 Setting: the Structure of Amazon Marketplace

Amazon marketplace works as an intermediary between consumers on one side and sellers on the others (Figure 1). Amazon operates different geographical marketplaces; for instance, Amazon.com serves mainly the US market, Amazon.mx the Mexican market, Amazon.fr the French market.

We can distinguish two types of sellers, retailers and producers. In the former case, we refer to downstream retailers buying products from upstream producers and reselling them on the marketplace (e.g. a 3P-seller or Amazon buys Sony WH-1000XM5 headphones and resell them on the marketplace). In the latter case, we refer to upstream producers selling directly their product on the marketplace (e.g. Sony sells directly the Sony WH-1000XM5 headphones on the platform). Amazon may operate either as a retailer selling its own products or as a producer selling its own private brands (e.g. Amazon Basics, Amazon Essentials, Solimo, Wag, Mama Bear). While generally a product could be sold either by a retailer or by its own producer, Amazon's private brands can be sold only by Amazon itself.

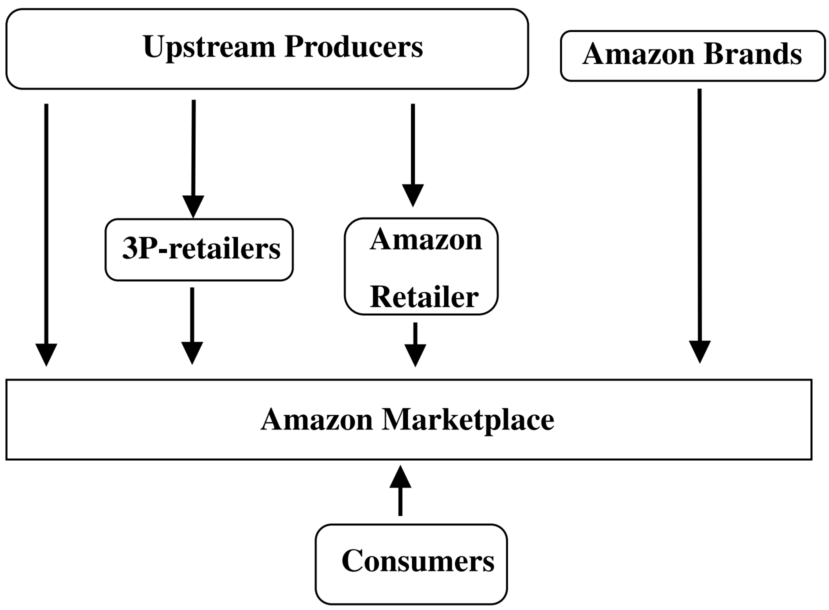


Figure 1: Marketplace Structure

Products can belong to different categories e.g. Consumer Electronics, Electronic

Accessories, Home and Kitchen, Office Products. In each category, we can find several markets. For instance, in the Office Products category a market could be staplers, or in the Electronic Accessories headphones. For every market, a product is assigned a unique barcode, called ASIN, and each barcode can be sold by multiple sellers at the same time, with the only exception of Amazon's brand barcodes (Figure 2).

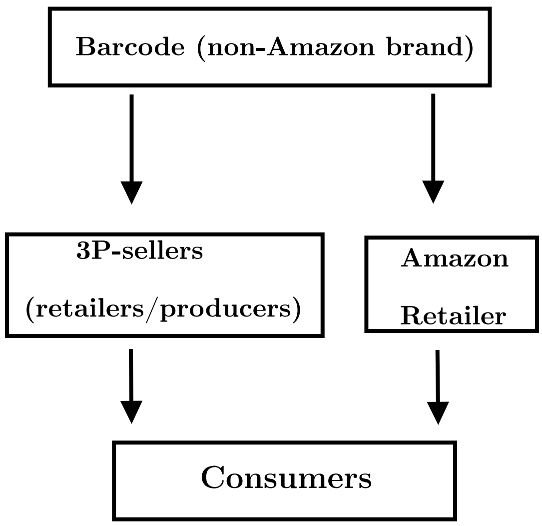


Figure 2: Barcode Sellers

There are three features which are central to the functioning of Amazon Marketplace. The first one is logistics, that is, how the delivery of the product is fulfilled. 3P-sellers can choose whether to deliver the product independently or to use FBA. When they use this logistic service, they will send the items to an Amazon warehouse and, upon purchase, Amazon will be in charge of delivering the item from its warehouse to the consumer. Hence, FBA is a way to outsource the storage and delivery of a product and, while Amazon may not be the only provider of this outsourcing service, it is the most commonly used by 3P-sellers. In order to use this service, 3P-sellers pay to Amazon an FBA fee, which is a unit fee depending on the size and weight of the product, and how long the product has been stored in the warehouse. When it comes to Amazon logistic choice, it will use by default its own service FBA and it will not pay the FBA fees.

The second feature are the referral fees. These are an ad-valorem fee and they amount to a percentage of the final price paid by consumers. These fees are typically 8% or 15% depending on the product category, although there are a few cases of non-linear fees.

In this case, the fee may vary depending on the price (e.g. in Grocery and Gourmet in Amazon.com, the referral fee is 8% for a price below or equal to 15%, and 15% otherwise) or on the portion of the price (e.g. for Electronics Accessories [3] in Amazon.com, the fee is 15% for the portion of the total sales price up to $100.00, and 8% for any portion of the total sales price greater than $100.00.).

The last feature is the recommendation system, which is made up by three main components. The first is the page ranking. When costumers looks for an item, they will type a keyword in the search bar and a list of products will appear, the order depending by the relevance to the keyword, the product rating, the number of reviews and the available offers characteristics (prices, sellers' rating, product). The second is the Buy-Box. As mentioned before, multiple sellers may offer the same barcode simultaneously (Figure 2) and there is not a limit on how many sellers can offer the same barcode. When this happens, all sellers' offers will be grouped in the same product page, but only one of them will be given more visibility to consumers. This offer will be place in a window on the top of the product page called the Buy-Box and its seller will be called the Buy-Box seller. Moreover, the product's price costumers observe in the page ranking will refer to the one of the Buy-Box offer. Finally, other products may be recommended in a product page under a page called Frequently Bought Together.

2.2 Data

I collect data from two sources. The first source of data is Keepa a website scraping Amazon and providing information on product and offers characteristics of all currently listed products. Data from Keepa has also been recently used in Cabral and Xu [2021], Lee and Musolff [2023], Gutierrez [2022] and Chen and Tsai [2023]. Product characteristics include title, brand, manufacturer, product description, in addition to real time changes in sales rank (a measure of aggregate product sales relative to other products in the same category), product rating, number of reviews and Buy-Box seller. Offer characteristics include seller's name, logistic method and real time changes in prices and shipping costs. I complement the data from Keepa with data from AmzScout, a market intelligence company used by 3P-sellers. First, AmzScout sales estimator allows to estimate the

[3] Headphones belong to this category.

aggregate quantity sold for a product in a given period from the sales rank. I provide more information about this in Appendix A.1.1. Secondly, I use data from number of keyword searches for headphones to compute the potential market (Appendix A.1.2). This information will also be useful to extract some relevant product characteristics.
Moreover, while I can estimate the aggregate quantity sold using data from Keepa and AmzScout, I still do not observe the quantity sold by each single seller. However, it is generally reported that most of the sales go through the Buy-Box seller and therefore, as in Gutierrez [2022], I will assumed that observed sales are realized only by the Buy-Box seller [4]. Finally, while in this paper I focus on the headphones market, I have collected data on other markets as well that I plan to use in future extensions.

3 Descriptive Evidence

In this empirical analysis I am going to consider the headphones market in Amazon.com between March 2023 and September 2023. In particular, I will focus on wireless-bluetooth headphones part of the market. The choice of the headphones market is driven by the fact that this is a market where small scale innovation (such as introducing an headphones with a particular design or certain features), and therefore 3P-sellers' entry, are particularly important. Also, I will focus on the wireless-bluetooth section of the market since it contains the most recent and higher demand products.
Although the headphones market is not the largest markets on Amazon, and I am considering only a section of it, we can see in Table 1 that, during the sample period, the market is characterized by a large number of barcodes and sellers, 5,918 and 4,329. Overall, Amazon has offered at least once 8.28% of these barcodes and only a small number of them is an Amazon brand (there are 12 Amazon branded products, all of them Amazon Basics). Yet, Amazon remains the largest seller in the market. On average, Amazon is selling around 363.4 barcodes per week, much higher compared to the 3P-sellers, which offer on average 3.47 barcodes per week.

[4] "Industry experts estimate that about 80% of Amazon sales go through the Buy Box, and the percentage is even higher for mobile purchases" ("Investigation of Competition in Digital Markets" Subcommittee on Antitrust, Commercial and Administrative Law, House of Representative, 2020)

Table 1: Main Summary Statistics

n° of barcodes	5,918
n° of sellers	4,329
% barcodes Amazon offered at least one day	8.28%
n° Amazon brand barcodes	12
weekly mean (median) n° sellers per barcode	2.03 (1)
weekly mean (median) n° of barcodes per non-Amazon merchant	3.47 (1)
weekly mean (median) n° of barcodes sold by Amazon	363.4 (355.5)

Moreover, these statistics suggest that Amazon may play a more important role as a retailer rather than as a producer.
This seems to emerge when we look at the sales distribution in Figure 3. The average weekly sales of products offered by Amazon as a retailer is much larger than the average across all products (44 compared to 10), while Amazon brand weekly sales are just above (around 16). Also, the sales distribution displays the long-tail typical of the e-commerce sector: there will be few products with very large sales, while the majority of products has fewer sales.

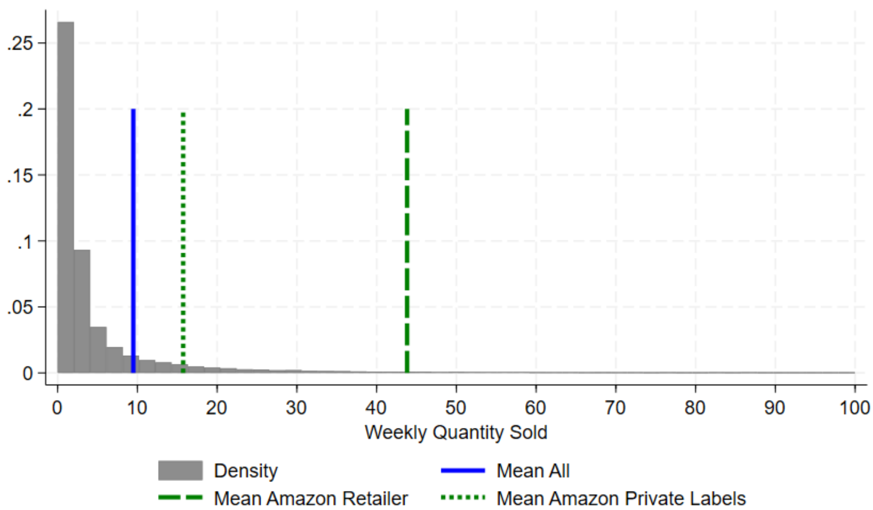


Figure 3: Weekly sales below 99th percentile.

4 Model

4.1 Setup

Before presenting the demand and the entry models, I am going to introduce the general framework of the models.

Consider one differentiated product market in Amazon marketplace. I index sellers by $s \in \mathcal{S}$, products by $j \in \mathcal{J}$ and time period by $t \in \mathcal{T}$, with cardinality $S \equiv |\mathcal{S}|$, $J \equiv |\mathcal{J}|$ and $T \equiv |\mathcal{T}|$. There are two types of sellers s: $3P$, the third party sellers, and A, Amazon. The sellers play a static entry game, where by entry I denote the choice of which products to sell in the market. Every period t, a seller s makes a static decision of whether to offer the portfolio of products $\mathcal{J}_{st}$, that is, it offers $\mathcal{J}_{st}$ if and only if the per-period profit is positive.

The timing of the game is as follows:

- **Stage 1**: seller s chooses the product portfolio $\mathcal{J}_s \subseteq \mathcal{J}$
- **Stage 2**: seller s chooses the price for every $j \in \mathcal{J}_s$
- **Stage 3**: profits are realized

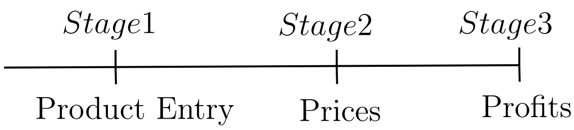


Figure 4: Entry Stages

The per-period profit is denoted by $\Pi_{st}(\mathcal{J}_{st}, \boldsymbol{X}_{st})$ and it is a function of the product portfolio $\mathcal{J}_{st}$ and of the product and seller characteristics $\boldsymbol{X}_{st}$. These characteristics may vary during time (e.g. product rating, number of reviews, prices), or they might be fixed (e.g. identity of the seller, logistic method [5]).

The profit will be equal to the difference between the variable profit and the entry cost of offering $\mathcal{J}_{st}$. Critically, profits will be different between 3P-sellers and Amazon, as the

[5]A seller may use different logistics methods for the same barcode at different time periods or in the same time period. This happens when the seller decides to use FBA for some products and the independent logistics for others. For the moment, I assume these are two separate offers.

latter collects fees from the other sellers. I am going to consider only the referral fees and I am going to assume there is perfect pass-through (that is, the fee is fully passed on to consumers)[6].

The functional form of 3P-sellers' and Amazon's per-period profits is:

$$\Pi_{3Pt}(\mathcal{J}_{3Pt}, \boldsymbol{X}_t) = \sum_{j\in\mathcal{J}_{3Pt}} \underbrace{\mathcal{M}_t \cdot s_{j3Pt}(\boldsymbol{p}_t, \boldsymbol{X}_t)}_{\text{demand}} \cdot \underbrace{\big((1-\phi_{jst})\cdot p_{j3Pt} - c_{j3Pt}\big)}_{\text{markup}} - \underbrace{F_{3P}(\mathcal{J}_{3Pt})}_{\text{entry cost}} \tag{1}$$

$$\begin{aligned}\Pi_{A,t}(\mathcal{J}_{At}, \boldsymbol{X}_t) = &\sum_{j\in\mathcal{J}_{At}} \underbrace{\mathcal{M}_t \cdot s_{jAt}(\boldsymbol{p}_t, \boldsymbol{X}_t)}_{\text{demand}} \cdot \underbrace{(p_{jAt} - c_{jAt})}_{\text{markup}} + \\ &\underbrace{\cdot \sum_{3P} \sum_{j\in\mathcal{J}_{3Pt}} \phi_{jst}\cdot p_{j3Pt}\cdot \mathcal{M}_t \cdot s_{j3Pt}(\boldsymbol{p}_t, \boldsymbol{X}_t)}_{\text{fee revenues}} - \underbrace{F_A(\mathcal{J}_{At})}_{\text{entry cost}}\end{aligned} \tag{2}$$

The 3P-sellers' variable profit function (Eq. 1) will depend on the sum of the sales of each single product in the portfolio, where $\mathcal{M}_t$ denotes the market size in period t, $s_{j3Pt}(\boldsymbol{p}_t, \boldsymbol{X}_t)$ the market share (where $\boldsymbol{p}_t$ indicates the vector of sellers' prices and $\boldsymbol{X}_t$ the vector of sellers' and products' characteristics), p_{jst} the price of product j, c_{j3Pt} the constant marginal cost of product j, and ϕ the referral fee. The entry cost, $F_{3Pt}(\mathcal{J}_{3Pt})$, depends on the portfolio of products borne by the seller in period t.

For Amazon, the variable profit function (Eq. 2) takes a similar functional form, but with two differences. First, Amazon will not pay a referral fee on prices. Secondly, Amazon receives a revenue from referral fees paid by sellers. Finally, like the sellers, it will pay an entry cost $F_{At}(\mathcal{J}_{At})$ depending on its product portfolio in period t.

4.2 Demand

Every period, a mass $\mathcal{M}_t$ of consumers enters the market on Amazon marketplace and decides whether to purchase one of the offered products.

The indirect utility of consumer i for product j sold by seller s at time t is:

$$U_{ijst} = \boldsymbol{\beta X_{jst}} - \alpha p_{jst} + \gamma_{month} + \gamma_j + \xi_{jst} + \varepsilon_{ijst} \tag{3}$$

This is a function of $\boldsymbol{X_{jst}}$, the product and seller characteristics (number of reviews, product's rating, logistics, whether the seller is Amazon, and shipping cost), p_{jst}, the

[6] In later extensions of the paper, I plan to include the FBA fees too.

price, γ_{month}, the month fixed effect, γ_j, the product fixed effect, ξ_{jst}, the product-seller random shock, ε_{ijst}, the consumer specific random shock distributed Type 1 EV [7].

The utility of the outside option is

$$U_{iot} = \varepsilon_{iot} \tag{4}$$

where the outside option is the choice of not buying any of the product offered in the market on Amazon marketplace [8].

Given the distribution of ε_{ijst}, the market share predicted by the demand model is:

$$s_{jst} = \frac{\exp(\delta_{jst} + \xi_{jst})}{1 + \sum_{s=1}^{S} \sum_{j \in \mathcal{J}_s} \exp(\delta_{jst} + \xi_{jst})} \tag{5}$$

where

$$\delta_{jst} = \boldsymbol{\beta X_{jst}} - \alpha p_{jst} + \gamma_{month} + \gamma_j \tag{6}$$

As mentioned before, we will assume that consumers observe only the seller in the Buy-Box. Therefore, the observed market shares is

$$s_{jst}^{obs} = \frac{\%BB_{jst} \cdot Q_{jt}}{\mathcal{M}_t} \tag{7}$$

where $\%BB_{jst}$ is the percentage of time spent by seller s in the Buy-Box of product j during period t, Q_{jt} is the observed quantity sold for product j during period t and $\mathcal{M}_t$ is the potential market during period t.

In the estimation I will then match the observed market shares, s_{jst}^{obs}, with the predicted market shares, s_{jst}.

[7] In this specification, I am modelling demand using Logit. Due to the presence of referral fees in the profit functions, the computation of marginal costs and counterfactual profits is more elaborate compared to standard profit functions. Therefore, I have opted for a simpler specification of demand, although I plan to use Nested Logit or Mixed Logit at later stages.

[8] Therefore, this includes both the choice of not buying any product at all or to buy a product from another store.

4.3 Supply

In Stage 2, Amazon and 3P-sellers play a Bertrand pricing game and choose prices to maximize the per-period variable profits from Eq. 1 and Eq. 2:

$$\max_{p_{j3P}} \pi_{3Pt} = \sum_{j \in \mathcal{J}_{3Pt}} \mathcal{M}_t \cdot s_{j3Pt}(\boldsymbol{p}_t, \boldsymbol{X}_t) \cdot \big((1-\phi_{jst}) \cdot p_{j3Pt} - c_{j3Pt}\big) \tag{8}$$

$$\max_{p_{jA}} \pi_{At} = \sum_{j \in \mathcal{J}_{At}} \mathcal{M}_t \cdot s_{jAt}(\boldsymbol{p}_t, \boldsymbol{X}_t) \cdot (p_{jAt} - c_{jAt}) +$$

$$\sum_{3P} \sum_{j \in \mathcal{J}_{3Pt}} \phi_{jst} \cdot p_{j3Pt} \cdot \mathcal{M}_t \cdot s_{j3Pt}(\boldsymbol{p}_t, \boldsymbol{X}_t) \tag{9}$$

Given the competition assumption, we can derive the sellers' marginal costs.

To better display the marginal cost function, I rewrite the variable profit functions in Eq. 8 and Eq. 9 into one function:

$$\Pi_{jst} = \sum_{j \in \mathcal{J}_{st}} \mathcal{M}_t \cdot s_{jst}(\boldsymbol{p}_t, \boldsymbol{X}_t) \cdot \big((1 - \phi_{jst} \cdot \mathbb{1}_{s=3P}) \cdot p_{jst} - c_{jst}\big) +$$

$$\mathbb{1}_{s=A} \cdot \cdot \sum_{3P} \sum_{j \in \mathcal{J}_{3Pt}} \phi_{jst} \cdot p_{j3Pt} \cdot \mathcal{M}_t \cdot s_{j3Pt}(\boldsymbol{p}_t, \boldsymbol{X}_t) \tag{10}$$

The first order condition is then

$$\text{FOC}_{p_{jst}} : (1 - \mathbb{1}_{s=3P} \cdot \phi_{jst}) \cdot s_{jst} + \sum_{k \in \mathcal{J}_s} [(1 - \mathbb{1}_{s=3P} \cdot \phi_{jst}) \cdot p_{kst} - c_{kst}] \cdot \frac{\partial s_{kst}}{\partial p_{jst}} +$$

$$\mathbb{1}_{s=A} \sum_{3P} \sum_{k \in \mathcal{J}_{3Pt}} \phi_{jst} \cdot p_{k3Pt} \cdot \frac{\partial s_{k3Pt}}{\partial p_{jst}} = 0 \tag{11}$$

Let JSt be the total number of offers in period t. Then, in matrix notation, Eq. 11 is equivalent to

$$(\mathbf{1} - \mathbb{1}_{s=3P} \odot \boldsymbol{\phi}) \odot S + \tilde{\Omega}[(\mathbf{1} - \mathbb{1}_{s=3P} \odot \boldsymbol{\phi}) \odot P - C] + \mathbb{1}_{s=A} \tilde{\Omega}' \boldsymbol{\phi} \odot P = 0 \tag{12}$$

where $\mathbf{1}$, $\mathbb{1}_{s=3P}$, $\mathbb{1}_{s=A}$, $\boldsymbol{\phi}$, P, C are a column vector of length JSt, and $\tilde{\Omega}$ and $\tilde{\Omega}'$ are square matrices of length JSt.

$\tilde{\Omega}$ is equal to $O \odot \Omega$. The first term, O, is the ownership matrix, whose term are equal to 1 when the row-index offer and the column-index offer belong to the same seller, and

0 otherwise. The second term, Ω, is the matrix of market shares derivatives with respect to prices.

Then, $\tilde{\Omega}'$ is equal to $O' \odot \Omega$. The first term, O', is the non-ownership matrix, whose term are equal to 1 when the row-index offer and the column-index offer do not belong to the same seller, and 0 otherwise. The second term, Ω, is the same as defined before [9].

From Eq. 12, I can then derive the vector of marginal costs:

$$C = (1 - \mathbb{1}_{s=3P} \odot \boldsymbol{\phi}) \odot [\tilde{\Omega}^{-1} S + P] + \tilde{\Omega}^{-1} \tilde{\Omega}' \mathbb{1}_{s=A} \odot \boldsymbol{\phi} \odot P \tag{13}$$

Finally, I model the marginal cost as function of seller identity (Amazon or not), product characteristics and month.

$$\log c_{jst} = \beta_1 Amazon + \boldsymbol{\beta}_2 \boldsymbol{X^c} + \gamma_{week} + \omega_{jst} \tag{14}$$

4.4 Entry

As mentioned before, the per-period profits are the difference between variable profits and entry costs.

$$\Pi_{st}(J_{st}, \boldsymbol{X}_t) = \underbrace{\pi_{st}}_{\text{variable profit}} - \underbrace{F_s(J_{st})}_{\text{entry cost}} \tag{15}$$

[9] I provide here an example for illustration:

$$\Omega = \begin{pmatrix} \frac{\partial s_{11t}}{\partial p_{11t}} & \cdots & \frac{\partial s_{J1t}}{\partial p_{11t}} & \cdots & \frac{\partial s_{SJ_t}}{\partial p_{11t}} \\ \vdots & & & & \vdots \\ \frac{\partial s_{11t}}{\partial p_{JSt}} & \cdots & \frac{\partial s_{J1t}}{\partial p_{JSt}} & \cdots & \frac{\partial s_{JSt}}{\partial p_{JSt}} \end{pmatrix}$$

$$O = \begin{pmatrix} 1 & \cdots & 1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

$$O' = \begin{pmatrix} 0 & \cdots & 0 & \cdots & 1 \\ \vdots & & & & \vdots \\ 1 & \cdots & 1 & \cdots & 0 \end{pmatrix}$$

The entry costs depend on the identity of the seller and how many products the seller is offering.

$$F_s(J_{st}) = \sum_{j \in J_{st}} (\theta_s + V_{jt}) \tag{16}$$

where V_{jt} is a random shock.

We assume that in Stage 1 each seller observes V_{jt}, but not ξ_{jst} and ω_{jst}. Therefore, each seller chooses J_{st} to maximize the expected per-period profit with respect to ξ and ω:

$$J^*_{st} : \arg\max_{J_{st}} E_{\xi,\omega}[\Pi_{st}(J_{st}, \boldsymbol{X}_t)] \tag{17}$$

5 Estimation

5.1 Demand

I am going to estimate demand at the weekly level using IV-GMM. The moment conditions are:

$$E[\xi|Z] = 0 \tag{18}$$

where Z are the instruments and ξ is obtained using the market shares inversion:

$$\xi_{jst} = \log(s_{jst}) - \log(s_{ot}) - \delta_{jst} \tag{19}$$

For the instrument Z, I use the other products characteristics as (BLP) instruments. In particular, for this paper, I use the average number of headphones suitable for sport:

$$z_{jst} = \frac{1}{JSt - 1} \sum_{st} \sum_{k \neq j} X^{sport}_{kst} \tag{20}$$

where X^{sport}_{kst} is a dummy equal to 1 if the headphone is suitable for sport, and zero otherwise. Since sport is a characteristic positively values by consumers, the average number of sport headphones can be interpreted as a measure of how much crowded the market is, so that, the higher the average, the higher the competitive pressure on sellers.

5.2 Entry

I am going to estimate entry at the weekly level. Therefore, given the entry cost function in Eq. 16, θ_s is the average entry cost of offering one product for seller s.

In order to estimate θ_s, I am going to employ a moment inequality approach. The choice is driven by the dimension of the problem: since the number of players is very large and the potential number of choices for each player is enormous (sellers may potentially have to make an entry decision for thousands of products), characterizing all the unique and multiple equilibrium regions is unfeasible. Another alternative could be to employ simulated based methods, but again, given the dimension of the problem, it risks being computationally too burdensome.

To derive the moment inequalities, we just need to assume that the observed outcome is an equilibrium of the game: this means that any deviation from this outcome is unprofitable. In practice, this implies that, when a seller offers a product portfolio $\mathcal{J}_{st}$, adding or removing one product from $\mathcal{J}_{st}$ is unprofitable.

This one-product deviations . One-product deviations, which has been used by other papers too, such as Canay et al. [2023], Nosko [2010], Eizenberg [2014], Wollmann [2018] and Fan and Yang [2022], will allow me to derive the moments necessary to estimate θ_s.

5.2.1 Deriving the Moment Inequalities

We introduce some extra notation to derive the moment inequalities using the apporach in Pakes [2010] [10].

I denote by $D_{jst} = \{0, 1\}$ seller s decision on whether to sell product j in period t and by $D_{st} = (D_{1st}, ..., D_{Jst})$: set of product offerings. Given D_{st}, sellers s gains variable profit $\pi_{st}(D_{st})$. Then, $\Delta_j \pi_{st} = \pi_{st}(\partial_j D_{st}) - \pi_{st}(D_{st})$ is the differential variable profit given one product deviation $\partial_j D_{st}$, where $\partial_j D_{st} = (D_{1st}, ..., 1 - D_{jst}, ..., D_{Jst})$.

Then, the estimated variable profit following this deviation is

$$\Delta_j \hat{\pi}_{st} = E[\Delta_j r_{st}] + U_{jt}$$

where U_{jt} is a specification error with $E[U_{jt}|D_{jst}] = 0$. Given the entry cost in Eq. 16, the non-profitable deviation can be rewritten in two possible ways.

If $D_{jst} = 0$, adding product j in $\mathcal{J}_{jst}$ is not profitable:

$$(\Delta_j \hat{\pi}_{st} - U_{jst} - (\theta_s + V_{jst}))(1 - D_{jst}) \leq 0 \tag{21}$$

[10] This notation follows closely Canay et al. [2023]

that is, the expected profit minus the entry cost is not positive.

If $D_{jst} = 1$, removing product j from $\mathcal{J}_{jst}$ is not profitable:

$$(\Delta_j \hat{\pi}_{st} - U_{jst} - (\theta_s + V_{jst}))(1 - D_{jst}) \leq 0 \tag{22}$$

that is, the expected profit plus the saved the entry cost is not positive.

These two inequalities will be the basis of the moments conditions. One potential issue is selection: while we have assumed that sellers do not observe the demand shock ξ before entering, they still observe V.

Therefore, two extra assumptions are introduced.

Assumption 1 The instrumental variable W_{jst} satisfies

$$E[U_{jst}|W_{jst}, D_{jst}] = 0 \text{ and } E[V_{jst}|W_{jst}] = 0$$

Assumption 2 For some known and positive value $\bar{V}$, the conditional expectation satisfies

$$E[V_{jst}|W_{jst}, D_{jst}] \leq \bar{V}$$

Following Canay et al. [2023], we can use these two assumptions to derive moment inequalities not depending on V.

I do this for Eq. 21, as it will be similar for Eq. 22. Given a function of the instrument $h(W_{jst})$, and by multiplying each side of Eq. 21 by $h(W_{jst})$, we obtain:

$$\Delta_j \hat{\pi}_{st} - U_{jst} - (\theta_s + V_{jst}))(1 - D_{jst})h(W_{jst}) \leq 0 \tag{23}$$

Then, since $E[U_{jst}(1 - D_{jst})h(W_{jst})] = 0$ by Ass. 1, by taking the expectation we get:

$$E[(\hat{\pi}_{st} - \theta_s)(1 - D_{jst})h(W_{jst})] + E[V_{jst}(1 - D_{jst})h(W_{jst})] \leq 0 \tag{24}$$

Since by Ass. 1 $E[V_{jst}h(W_{jst})] = 0$ and by Ass. 2 $E[V_{jst}D_{jst}h(W_{jst})] \leq E[\bar{V}D_{jst}h(W_{jst})]$, Eq. 24 is rewritten as:

$$E[((\hat{\pi}_{st} - \theta_s)(1 - D_{jst}) - \bar{V}D_{jst})h(W_{jst})] \leq 0 \tag{25}$$

In this way, we find a moment condition which does not depend neither on U_{jst} nor on V_{jst}.

Using a similar argument for Eq. 22, we can now express the two moment inequalities:

$$m^l_{jst}(\theta_s) \equiv E[((\hat{\pi}_{st} - \theta_s)(1 - D_{jst}) - \bar{V}D_{jst})h(W_{jst})]$$
$$m^u_{jst}(\theta_s) \equiv E[((\hat{\pi}_{st} + \theta_s)D_{jst} - \bar{V}(1 - D_{jst}))h(W_{jst})]$$

Therefore, for each period t and each firm s, we have potentially $k = 2 \cdot J$ moments:

$$m_{st}(\theta_s) = (m^l_{1st}, ..., m^l_{Jst}, m^u_{1st}, ..., m^u_{Jst}) \tag{26}$$

where $E[m_{st}(\theta_s)] \leq 0$.

Finally, an important aspect of the moment inequalities defined above is that they admit partition, that is, the moments for seller s are not overlapping with other sellers'. Thus, inference can be done seperately for each $m_s(\theta_s)$.

5.2.2 Inference

First of all, we set up the inference problem [11]. In a model defined by moment inequalities, for a final dimensional parameter vector $\theta_0 \in \Theta$, we have that:

$$E[m_t(\theta_0)] \leq 0$$

where $m_t \equiv (m_{1t}, ..., m_{St})'$ and the inequality is interpreted component wise.

The identified set for θ_0 is the set of values for θ_0 satisfying the moment inequalities

$$\Theta_0 = \{\theta \in \Theta : E[m_t(\theta)] \leq 0\}$$

Then, the confidence region C_T for the identified set Θ_0 is:

$$\lim_{T\to\infty} \inf \inf_{\theta \in \Theta_0} P\{\theta \in C_T\} \geq 1 - \alpha$$

This is accomplished by exploiting the duality between confidence regions and inverting the tests of each individual null hypotheses.

$$H_\theta : E[m_t(\theta)] \leq 0$$

For each θ, a test of H_θ, $\phi_T(\theta)$, is available that satisfies

$$\lim_{T\to\infty} \sup_{\theta \in \Theta_0} E[\phi_T(\theta)] \leq \alpha$$

Consider a number T of periods. The confidence region is then

$$C_T \equiv \{\theta \in \Theta : \phi_T(\theta) = 0\}$$

[11] Again, I will follow the exposition in Canay et al. [2023].

The test for the null hypothesis is

$$\phi_T(\theta) \equiv I\{\mathcal{T}_T(\theta) > c_T(1-\alpha, \theta)\}$$

where I is an indicator function, $\mathcal{T}_T(\theta)$ is the test statistic and $c_T(1-\alpha,\theta)$ is the critical value.

As mentioned before, the model admits partitioning and we can estimate θ_s separately for each seller. Therefore, for simplicity, I will present here the test for one seller s.

In order to build the test statistic, I use the method in Chernozukov et al. (2019):

$$\mathcal{T}_T(\theta_s) = \max_{1\leq l\leq k} \frac{\sqrt{T}\bar{m}_{lsT}(\theta_s)}{\hat{\sigma}_{lsT}(\theta_s)} \tag{27}$$

where $\bar{m}_{lsT} = \frac{1}{T}\sum_t m_{lst}(\theta_s)$, $\hat{\sigma}_{lsT} = \sqrt{\frac{1}{n}\sum_t (m_{lst}(\theta_s) - \bar{m}_{lsT})^2}$ and k is the total number of moments.

Finally, Chernozukov et al. (2019) derive the critical values in two steps.

1. Let $0 < \beta < \alpha/2$ be a tuning parameter and Φ the distribution function of the standard normal distribution. Let

$$\hat{c}_{ksT}(1-\beta,\theta_s) = \frac{\Phi^{-1}(1-\beta/k)}{\sqrt{1-\Phi^{-1}(1-\beta/k)^2/n}}$$

$$\hat{k}_{sT} = \sum_{l=1}^{k} I\{\frac{\sqrt{T}\bar{m}_{lsT}(\theta_s)}{\hat{\sigma}_{lsT}(\theta_s)} > -2\hat{c}_{ksT}(1-\beta,\theta_s)\}$$

2. Then, the critical value is:

$$\hat{c}_{sT}(1-\alpha,\theta_s) = \left(\frac{\Phi^{-1}(1-(\alpha-2\beta)/\hat{k}_{sT}}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta)/\hat{k}_{sT})^2/T}}\right) I\{\hat{k}_{sT} \geq 0\}$$

6 Results

I present here the results of the demand, supply and entry.

6.1 Demand

I present the demand estimates of $\{\beta, \alpha\}$ in the Table 2 [12]. We can see that demand increases in product rating and number of reviews, which can be interpreted as proxies of quality, and it decreases in shipping cost. Moreover, the Amazon and FBA 3P-sellers receives larger demand compared to independent 3P-sellers, although the demand increase is larger for Amazon.

Variables	**Estimates (Standard Error)**
Prices	-0.1738^{***} (0.0635)
Amazon	1.3102^{***} (0.108)
Fulfilled by Amazon	0.2889^{***} (0.0719)
Shipping Cost^2	-0.003^{**} (0.0012)
Product Rating$_{week-1}$	0.3087^{***} (0.081)
Log Reviews$_{week-1}$	0.1303^{***} (0.0474)
Product Fixed Effects	Yes
Month Fixed Effects	Yes
Instrumental Variable	Yes

Table 2: $^*p < 0.05$, $^{**}p < 0.025$, $^{***}p < 0.01$. Instruments: mean number of sport headphones.

Looking at the prices elasticities (Fig. 5), we can notice that the average own price elasticity during the period is large, between -14.9% and -15.6%. These estimates are just above the elasticities found in other papers estimating demand in Amazon (Gutierrez [2022], Lee and Musolff [2023], Tai Lam, 2023), which are usually situated around 10%.

[12] To estimate demand, I use the package PyBLP from Conlon and Gortmaker [2020]

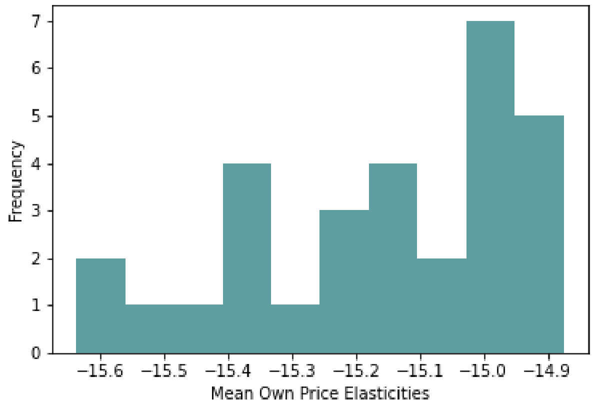
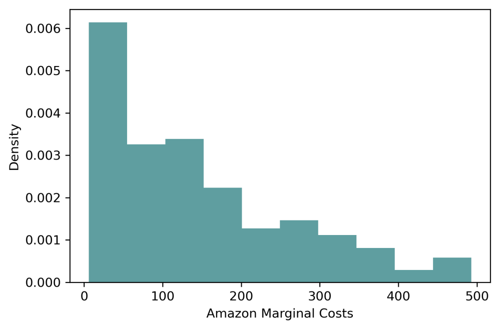


Figure 5: Distribution of the mean of the own price elasticity of demand across weeks

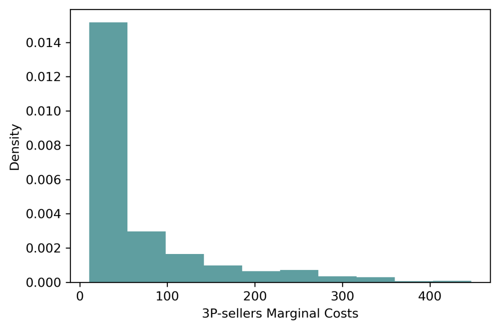
6.2 Supply

6.2.1 Marginal Costs

Analysing the sellers' marginal costs derived from Eq 13, Amazon's and 3P-sellers' appear to have a similar distribution of marginal costs (Fig 6), both displaying a positive skeweness due to the presence of fewer high-quality and high-cost products. However, Amazon seems to have larger marginal costs than the 3P-sellers: the median (mean) marginal cost for Amazon is 121$ (148$), while for 3P-sellers is 30$ (66$). There are two possible explanations for this: Amazon tend to sell higher-quality products, having higher-costs, or Amazon has higher marginal costs.



(a)



(b)

Figure 6: Marginal Costs Distribution

In order to investigate this issue, I compare Amazon's marginal costs to 3P-sellers offering the same product in the same period. To do so, I compute the difference between 3P-sellers' and Amazon's marginal costs ($\Delta_{j3Pt} = c_{j3Pt} - c_{jAt}$) and display the results in Fig. 7. Indeed, it appears that the difference in marginal costs distribution is partly due to 3P-sellers having lower marginal costs (on average, they are 15$ lower than Amazon's). However, this difference in marginal costs for the same products is less pronounced compared to when we consider all products. Therefore, it seems that most of the difference is coming from Amazon concentrating in the high-end of the market, where products have higher quality and costs.

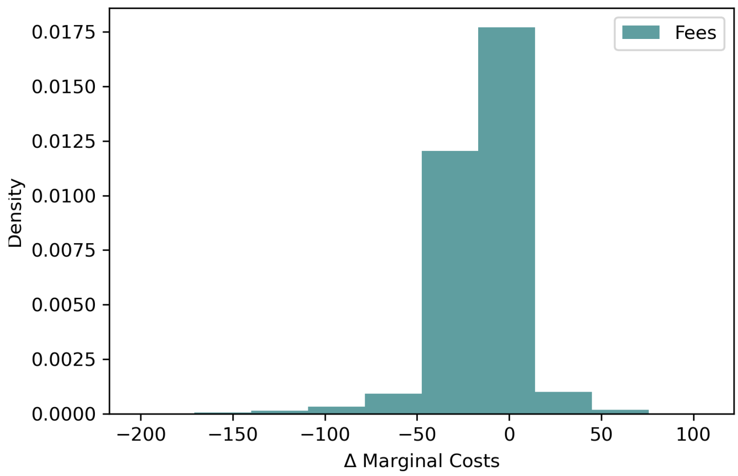


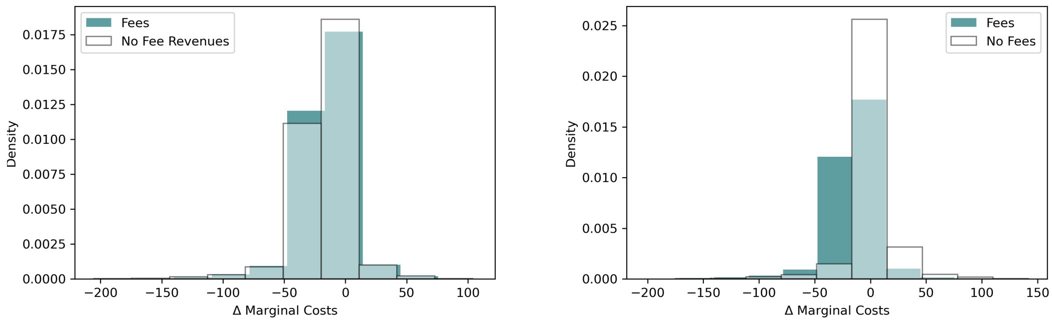
Figure 7: Difference 3P-sellers' and Amazon's marginal costs for the same product and period

Moreover, it turns out that accounting for referral fees play a crucial role in deriving 3P-sellers marginal costs. To understand this, I compute the difference in marginal costs when Amazon does not receive revenues from fees (Fig. 8a), but 3P-sellers keep paying referral fees [13], and the difference in marginal costs when there are not referral fees at all

[13] In this case, the marginal costs equation is changed into:

$$C = (1 - \mathbb{1}_{s=3P} \odot \phi) \odot [\tilde{\Omega}^{-1} S + P]$$

[14] (Fig. 8b). As we can notice, including the referral fees paid by 3P-sellers in the profit function shifts substantially the distribution left-ward.



(a) (b)

Figure 8: Marginal costs distribution under different models.

6.2.2 Lerner Index

Given the computed marginal costs, we can finally compute 3P-sellers' (28) and Amazon's (29) Lerner Index.

$$\text{Lerner Index}_{j3Pt}\ (\%) = \frac{p_{j3Pt} \cdot (1 - \phi_{j3Pt}) - c_{j3Pt}}{p_{j3Pt} \cdot (1 - \phi_{j3Pt})} \cdot 100 \tag{28}$$

$$\text{Lerner Index}_{jAt}\ (\%) = \frac{p_{jAt} - c_{jAt}}{p_{jAt}} \cdot 100 \tag{29}$$

As we can see from Fig. 9, there is a large difference in the distribution of the two Indices. On the one hand, the 3P-sellers' Lerner Index is more uniform between close to 1% and 31%, with an average Index of 14%. On the other, Amazon's Lerner Index is positively skewed, with many product having an Index below 10%, but it also enjoys large market power on other products: while only 1% of the products offered by 3P-sellers have a Index above 30%, 7% of the products offered by Amazon have an Index above 30%.

[14]The marginal costs equation becomes the standard equation in a model without fees:

$$C = \tilde{\Omega}^{-1} S + P$$

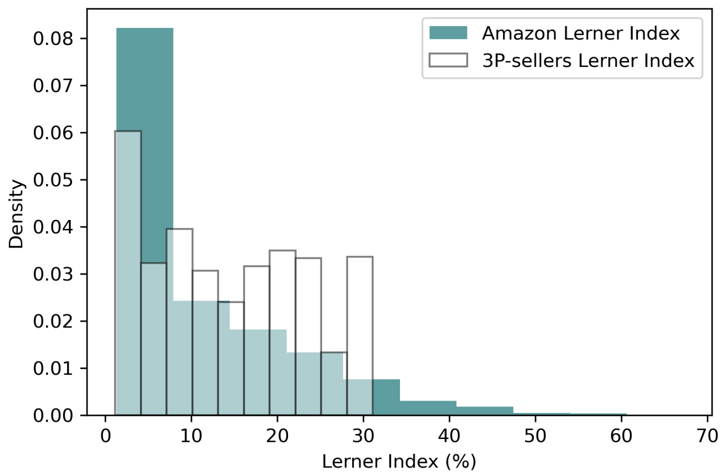


Figure 9: Lerner Index Distribution

6.3 Entry

I start to estimate the weekly entry cost for Amazon using a random sample of 10 products between August 2023 and September 2023. The estimation proceeds in two steps.

First of all, given the estimated demand parameters and marginal costs, we compute the expected variable profit $\hat{\pi}_{At}$ over the week. To do so, I simulate 50 random draws of ξ_{jAt} and compute the optimal price in equilibrium. Then, I take the mean of the realized profit across simulations, $\bar{\hat{\pi}}_{At}$. For simplification, here I will not simulate over ω_{jAt}.

In order to simulate the demand shock, I split ξ_{jst} into a mean component common across all sellers for j and t,ξ_{jt}, and random shock *iid* and following a standard normal distribution, $\tilde{\xi}_{jst}$.

$$\xi_{jst} = \xi_{jt} + \tilde{\xi}_{jst} \tag{30}$$

I will take random draws of $\tilde{\xi}_{jst}$ to simulate ξ_{jst}.

For the marginal costs, I will use the predicted value from the estimation of Eq. 14 [15].

[15] In addition to an Amazon dummy and the month fixed effect, the regression for the marginal cost includes brand dummies and characteristics dummies (noisecancelling, kids, iphone, gym, folding, dj, gaming, mask, headband, sleeping, sport, usb, usbc, workout). This specification works well at predicting the marginal costs, with an $R^2 = 0.87$

Table 3: Amazon's weekly entry costs bounds.

$\hat{\theta}_A$	[56\$, 297\$]

Parameters: $T = 8; k = 11; \bar{V} = 15\$; \alpha = 0.1; \beta = 0.026$

The results suggest that Amazon's entry costs vary between 56$ and 297$ dollar per product every week (Table 3) [16]. Since the estimated bounds are large, this suggests that heterogeneity across products should be accounted for in the entry costs function.

6.4 Counterfactual Results

In the preliminary counterfactual results, we are interested in analysing the impact of banning Amazon's products (both private labels and those sold as retailer) when the entry decision is exogenous, that is, the 3P-sellers do not adjust their portofolio of products. As expected, both Amazon's profit and consumer surplus decrease substantially when Amazon is banned (-18.66 % and -17.83%, respectively). Whereas, on the 3P-sellers side, it appears that prices and markups are not changing substantially.

Table 4: Counterfactual Results from Banning Amazon products

%Δ Consumer Surplus	-17.83%
%Δ 3P-sellers Prices	1.33e-03 %
%Δ 3P-sellers Markups	0.016%
%Δ Amazon Profit	-18.66 %

%Δ Consumer Surplus and %Δ Amazon Profit refer to the entire difference over the sample period. %Δ 3P-sellers Prices and %Δ 3P-sellers Markups refer to the average difference over the sample period.

[16] Note that $k = 11$ since some products will not have all the moments. These are products which are always observed or never observed during the sample period. Since 8 products are never sold by Amazon, 1 is always sold, and 1 product sale is interrupted during the sample period, we have in total 11 moments.

7 Conclusion

This papers investigates sellers' prices and entry strategies in Amazon marketplace.
First, I develop a structural model which includes both the pricing stage and the entry stage, where by entry I refer to the sellers' choice of which products to offer in one market. The supply model incorporates the referral fees paid by the seller, so that Amazon will take into account the fee revenues when choosing its prices. The main component of the entry part is the entry cost parameter. This is paid out by sellers for each product they offer, and it is constant for different products, but heterogeneous across firms. To estimate it, I employ a moment inequality approach which allows me to obtain the confidence region of the parameter. Although I do not have point identification, this approach is very tractable considering this setting with an extremely large number of products and sellers. Both pricing and entry decisions will be static.
I estimate demand and entry at the weekly level. Consistent with previous findings, demand estimates show that demand increases when the product is sold by Amazon or by a seller using FBA. Then, while Amazon's and 3P-sellers' market power is limited on many products, Amazon holds a large market power on a number of products it offer. Moreover, when Amazon and 3P-sellers offer the same product, 3P-sellers seem to have a small cost advantage compared to Amazon, on average. Therefore, vertical integration might be driven both by efficiency and profitability motivations
Regarding the entry costs, I find that Amazon's entry cost for one product ranges between 56\$ and 297\$. Finally, in the first counterfactual simulations, I find that, when entry decisions are fixed, a ban on Amazon's products is negative both for consumers and Amazon.
In future extensions of the paper, I am going to improve on the estimates of Amazon's entry cost and to estimate 3P-sellers' entry costs. Then, I will repeat the counterfactual simulations allowing for 3P-sellers to adjust their offering and prices once Amazon is removed from the market. Moreover, I plan to include the Buy-Box algorithm in the demand model in order to study how the recommendation system impacts entry at the market level. I provide some preliminary evidence on the Buy-Box in Appendix B.

References

S. Anderson and O. Bedre-Defolie. Hybrid platform model. Working Paper, 2021.

M. Armstrong. Competition in two-sided markets. *The RAND Journal of Economics*, 37 (3):668–691, 2006. ISSN 07416261. URL http://www.jstor.org/stable/25046266.

L. Cabral and L. Xu. Seller reputation and price gouging: Evidence from the covid-19 pandemic. *Economic Inquiry*, 59(3):867–879, 2021. doi: https://doi.org/10.1111/ecin.12993. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecin.12993.

B. Caillaud and B. Jullien. Chicken egg: Competition among intermediation service providers. *The RAND Journal of Economics*, 34(2):309–328, 2003. ISSN 07416261. URL http://www.jstor.org/stable/1593720.

I. Canay, G. Illanes, and A. Velez. A user's guide for inference in models defined by moment inequalities. Working Paper, 2023.

N. Chen and H.-T. Tsai. Steering via algorithmic recommendations. Working Paper, 2021.

N. Chen and H.-T. Tsai. Price competition under information (dis)advantage. Working Paper, 2023.

V. Chernozhukov, D. Chetverikov, and K. Kato. Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies*, 86(5 (310)):pp. 1867–1900, 2019. ISSN 00346527, 1467937X. URL https://www.jstor.org/stable/26802782.

D. Condorelli, A. Galeotti, and V. Skreta. Selling through referrals. *Journal of Economics & Management Strategy*, 27(4):669–685, 2018. doi: https://doi.org/10.1111/jems.12251. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jems.12251.

C. Conlon and J. Gortmaker. Best practices for differentiated products demand estimation with pyblp. *The RAND Journal of Economics*, 51(4):1108–1161, 2020. doi: https://doi.org/10.1111/1756-2171.12352. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12352.

G. Crawford, M. Courthoud, R. Seibel, and S. Zuzek. Amazon entry on amazon marketplace. Working Paper, 2022.

C. de Chaisemartin and X. D'Haultfoeuille. Difference-in-differences estimators of intertemporal treatment effects. Working Paper, 2020.

A. de Cornière and G. Taylor. A model of biased intermediation. *The RAND Journal of Economics*, 50(4):854–882, 2019. doi: https://doi.org/10.1111/1756-2171.12298. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12298.

A. Eizenberg. Upstream innovation and product variety in the u.s. home pc market. *The Review of Economic Studies*, 81(3 (288)):1003–1045, 2014. ISSN 00346527, 1467937X. URL http://www.jstor.org/stable/43551618.

F. Etro. Product selection in online marketplaces. *Journal of Economics & Management Strategy*, 30(3):614–637, 2021. doi: https://doi.org/10.1111/jems.12428. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jems.12428.

Y. Fan and C. Yang. Estimating discrete games with many firms and many decisions: An application to merger and product variety. Working Paper, 2022.

C. Farronato, A. Fradkin, and A. MacKay. Self-preferencing at amazon: Evidence from search rankings. *AEA Papers and Proceedings*, 113:239–43, May 2023. doi: 10.1257/pandp.20231068. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20231068.

M. S. Goeree. Limited information and advertising in the u.s. personal computer industry. *Econometrica*, 76(5):1017–1074, 2008. doi: https://doi.org/10.3982/ECTA4158. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA4158.

A. Goolsbee and J. Chevalier. Measuring prices and price competition online: Amazon and barnes and noble. *Quantitative Marketing and Economics*, 1, 08 2002. doi: 10.2139/ssrn.319701.

G. Gutierrez. The welfare consequences of regulating amazon. Working Paper, 2022.

A. Hagiu and J. Wright. Marketplace or reseller? *Management Science*, 61(1):184–203, 2015. ISSN 00251909, 15265501. URL http://www.jstor.org/stable/24551078.

A. Hagiu, B. Jullien, and J. Wright. Creating platforms by hosting rivals. *Management Science*, 66(7):3234–3248, 2020. doi: 10.1287/mnsc.2019.3356. URL https://doi.org/10.1287/mnsc.2019.3356.

A. Hagiu, T.-H. Teh, and J. Wright. Should platforms be allowed to sell on their own marketplaces? *The RAND Journal of Economics*, 53(2):297–327, 2022. doi: https://doi.org/10.1111/1756-2171.12408. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12408.

S. He and B. Hollenbeck. Sales and rank on amazon.com. Working Paper, 2020.

A. Hervas-Drane and S. Shelegia. Retailer-led marketplaces. Mimeo, 2021.

D.-S. Jeon and P. Rey. Platforms, hold up and innovation. mimeo, 2022.

K. Jerath and Z. J. Zhang. Store within a store. *Journal of Marketing Research*, 47(4): 748–763, 2010. ISSN 00222437. URL http://www.jstor.org/stable/20751538.

B. Jiang, K. Jerath, and K. Srinivasan. Firm strategies in the "mid tail" of platform-based retailing. *Marketing Science*, 30(5):757–775, 2011a. ISSN 07322399, 1526548X. URL http://www.jstor.org/stable/23011976.

B. Jiang, K. Jerath, and K. Srinivasan. Firm strategies in the "mid tail" of platform-based retailing. *Marketing Science*, 30(5):757–775, 2011b. ISSN 07322399, 1526548X. URL http://www.jstor.org/stable/23011976.

H. T. Lam. Platform search design and market power. Working Paper, 2023.

K. H. Lee and L. Musolff. Entry into two-sided markets shaped by platform-guided search. Working Paper, 2023.

E. Madsen and N. Vellodi. Insider imitation. Working Paper, 2021.

C. Nosko. Competition and quality choice in the cpu market. Working Paper, 2010.

A. Pakes. Alternative models for moment inequalities. *Econometrica*, 78(6):1783–1822, 2010. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/40928462.

I. Reimers and J. Waldfogel. A framework for detection, measurement, and welfare analysis of platform bias. Working Paper, 2023.

J.-C. Rochet and J. Tirole. Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003. ISSN 15424766, 15424774. URL http://www.jstor.org/stable/40005175.

T. G. Wollmann. Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, 108(6):1364–1406, June 2018. doi: 10.1257/aer.20160863. URL https://www.aeaweb.org/articles?id=10.1257/aer.20160863.

F. Zhu and Q. Liu. Competing with complementors: An empirical look at amazon.com. *Strategic Management Journal*, 39(10):2618–2642, 2018. doi: https://doi.org/10.1002/smj.2932. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.2932.

Appendices

A Data Appendix

In this section, I describe how the dataset is constructed and the variable in the dataset.

A.1 Dataset

I collect data on all listed products in a market from Keepa, where each product is distinguished by a unique number called ASIN.

For each product, I observe the all the real-time changes in sales rank, number of reviews, sellers' logistics, prices and shipping costs since the product was tracked by Keepa. Other information include the changes in Buybox (Buybox seller and price), product's title and description.

I rearrange the data in order to create a weekly panel of offers for the products. An offer is defined as a seller-day-ASIN combination and it contains the price, shipping cost, sales rank, number of reviews and rating for a seller of an ASIN in a certain day. Then, I will add whether the seller was in the Buybox during that day and compute the percentage of time spent in the Buy-Box.

One issue which might occur is that the Buybox seller may have more than one offer for the same ASIN varying only by the logistic method, that is, there will be an offer using FBA and one using FBM. Since I do not know which of the two offers is contained in the Buybox, I will assume that the FBA offer is the Buybox one.

Finally, I will add the product characteristics. Since these are contained in the product description and the title, I use the following strategy. First, I download all the keywords associated with "headphones" from AmzScout, another market intelligence company; this set includes approximately 190 keywords. In addition to the word headphones, these keywords are informative of the most salient characteristics as usually consumers search for products having particular characteristics e.g. noise-cancelling, sleeping, kids, iphone, gaming. Therefore, while these characteristics are not exhaustive (for instance, they do not capture more technical characteristics or aesthetic features), they are useful as a starting cream-skim of product differentiation. To extract these characteristics, I first select the first 100 keywords in terms of average number of monthly searches during the sample period. After some text cleaning and after having removed the brands name, I am left with about 40 words. Then, I add each word the the panel dataset as a dummy equal to 1 if the word is contained in the text or in the title of the ASIN, and 0 otherwise.

A.1.1 Estimated Quantity sold

Different methods have been proposed to approximate the quantity sold from the sales rank. Goolsbee and Chevalier [2002] and He and Hollenbeck [2020] estimate a Pareto distribution model using the category sales rank and actual sales [17]. By taking the logs, the Pareto distribution model can be transformed into a relationship between the log of sales rank and the log of sales.

$$\log(Quantity_t) \approx \alpha - \beta \log(SalesRank_t) \tag{31}$$

In the case of He and Hollenbeck [2020], the model is estimated using the average sales data at the weekly level and the weekly observations of the category sales rank. Chen and Tsai [2021] estimate the same model using the daily sales rank. However, since they do not have actual data on sales, they assume β in the regression is equal to one, while, given their model specification, they do not need to estimate α.

Finally, Gutierrez [2022] collects sales estimates from two leading market intelligence companies for Amazon sellers, AmzScout and JungleScout. These companies use data on actual sales and SalesRank in given period to estimate the relationship between them. Since it appears that AmzScout employs a power test model, Gutierrez [2022] uses a sample of estimated quantities and SalesRank from this website in order to retrieve the estimated parameters; then, he repeat the same procedure for a sample from JungleScout, but this time using a spline, which is the model JungleScout seems to employ. He uses the estimated parameters in order to find estimates of quantities sold in his dataset.

Here, I take a similar approach to Gutierrez [2022] and, using data from AmzScout, I estimate a model which approximates the one used by AmzScout. To collect the sample, I start from a electronics sales rank [18] equal to 1 and then double the sales rank until the

[17] Goolsbee and Chevalier [2002] use data from a seller and own experiments for the book category. He and Hollenbeck [2020] compute sales using changes in inventory reported by Amazon.com

[18] Electronics is the rootcategory in case for headphones

estimated quantity sold remains constant.

As Gutierrez [2022], we do not know the precise model used by AmzScout to estimate the quantity sold. Therefore, I start from the simplest model found in Chevalier and Goolsbee (2003) and He and Hollenback (2020).

$$log(\hat{Q})\_AmzScout = \alpha - \beta log(sales\_rank\_electronics) \tag{32}$$

	(1)
VARIABLES	$log(\hat{Q})\_AmzScout$
$log(sales\_rank\_electronics)$	-0.874***
	(0.0661)
Constant	12.59***
	(0.419)
Observations	24
R-squared	0.888

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Log-log regression estimation

Finally, I use $\hat{\alpha}$ and $\hat{\beta}$ to compute the estimated quantity sold in my data.

A.1.2 Market Size

I use data from the estimated number of keyword search from AmzScout. First, I collect information on all the keywords with two to five words containing the word *headphones*; according to AmzScout, this corresponds to about 700 keywords. Together with the keywords, AmzScout provides also the estimated number of searches in a month.

For the above application, I assume that the headphones market can be divided in different subgroups and I start considering the "wireless bluetooth" subgroup. So, I select all the keywords containing the word *wireless*, *bluetooth*, or both, and I assume that every single search corresponds to a consumer. Therefore, the market size for the "wireless bluetooth" subgroup is built as the sum of the market sizes for each separate keyword.

Wireless Bluetooth Keywords		
Headphones Wireless Bluetooth	Wireless Headphones	Bluetooth Headphones
Beats Headphones Wireless Bluetooth	Headphones Bluetooth	Headphones Wireless Bluetooth Noise Cancelling
Beats Solo3 Wireless On-Ear Headphones	Wireless Headphones Over Ear	Sleep Headphones Wireless Bluetooth
Wireless Headphones Bluetooth	Bluetooth Headphones Wireless	Sennheiser Headphones Wireless
Bluetooth Headphones with Mic	Noise Cancelling Headphones Bluetooth	Dual Wireless TV Headphones
Tribit Xfree Tune Bluetooth Headphones	Bluetooth Headphones Over the Ear	JBL Headphones Wireless Bluetooth
TV Headphones Wireless	Bluetooth Eyemask Sleep Headphones	P47 Wireless Headphones
Kids Bluetooth Headphones	Mpow Headphones Wireless Bluetooth	White Headphones Wireless
Wireless Headphones for TV	White Wireless Headphones	Sony Wireless Headphones
Bose Headphones Wireless Bluetooth	iPhone Headphones Bluetooth	Sony WH-CH500/B Stamina Wireless Headphones
Bluetooth Headphones Wireless Earbuds	Open Ear Headphones Wireless Bluetooth	Wireless Kids Headphones
Sleep Mask with Bluetooth Headphones	Tagry Bluetooth Headphones	Bluetooth Headphones for Kids
Spiderman Kids Volume-Limiting Bluetooth Headphones	Neckband Bluetooth Headphones	Wireless Headphones with Microphone
Headphones Wireless	Wireless Headphones Gaming	Bluetooth Running Headphones
Lenovo TH30 Wireless Bluetooth Headphones	Sony Headphones Wireless Bluetooth	Wireless Gaming Headphones
Beribes Headphones Wireless Bluetooth	Veatool Bluetooth Headphones	Kids Headphones Bluetooth
Sleep Headphones Bluetooth Headband		

Table 6: Keywords containing *headphones* + *wireless* and/or *bluetooth*.

Month	monthly market size
March 2023	246'000
April 2023	319'340
May 2023	1'271'830
June 2023	804'670
July 2023	919'990
August 2023	1'010'870
September 2023	441'970

Table 7: Potential Market by Month for Bluetooth-wireless headphones

B Buy-Box

Given the importance it plays in the recommendation system, we are also interested in understanding how the Buy-Box algorithm works. Following a methodology similar to Lee and Musolff [2023], I approximate the algorithm into a discrete choice model.

Consider a fraction of time τ between $t-1$ and t. The Buy-Box valuation of product j for seller s in period τ is

$$v_{jBBs,\tau} = \alpha^{BB} p_{jst} + \boldsymbol{\beta}^{BB} \boldsymbol{X}_{jst} + \xi_{jBBs} + \varepsilon_{jBBs,\tau} \tag{33}$$

The valuation when no Buy-Box is provided is

$$v_{j0} = \varepsilon_{j0} \tag{34}$$

Assuming that ε_{jBBs} is distributed EV Type I, the probability of seller s being chosen in the Buy-Box is

$$s_{jBBs,t} = \frac{\exp(\delta^{BB}_{jst} + \xi_{jBBs})}{1 + \sum_{sj=1}^{Sj} \exp(\delta^{BB}_{jst} + \xi_{jBBs})} \tag{35}$$

Using the standard market share inversion, I can then estimate the parameters of the model 2SLS, using as instrument the number of available sellers for a product.

Variables	**Estimates (Standard Error)**
Prices	-0.0013 (0.001)
Shipping Cost	0.0095 (0.014)
Amazon	0.5078^{***} (0.075)
Fulfilled by Amazon	0.0986^{*} (0.053)
Product Fixed Effects	Yes

Table 8: Weekly Buy-Box Estimates. $*p < 0.05$, $**p < 0.025$, $***p < 0.01$