Business stealing in a decentralized social network - The case of Mastodon

Mathias Dachert

October 19, 2023

Abstract

We apply the model of Berry & Waldfogel (1999) to a Data collected on the decentralized social network Mastodon. A decentralized social network allows users to create different servers (like email for instance). We find a higher level of business stealing on the platform and lower standard deviation of log fixed cost than the paper of 1999 which show evidence of a form of ineffiency in number of entry. However, regarding the promises of Mastodon, it is a good sign that servers steal business since none of them is supposed to have power over the network.

I would like to specially thank my supervisor Michele Fioretti for his guidance during the master thesis and further. This thesis is also dedicated to Sofia who always supports me in my adventures.

Part I

Introduction

The concentration of some Online Social Networks (OSN) - e.g. Facebook, Twitter, Youtube - has become a problem known also as "The Social Dilemma" according to the 2020 Netflix documentary of Jeff Orlowski. Many activists think the best solution would be for platforms to adopt a decentralized architecture Jonathan DeYoung (2020). A decentralized architecture implies to have different servers owned by users communicating together by a protocol. The email protocol is an example of a decentralized architecture. Under the same protocol companies can create different email services (Gmail, Outlook or protonmail). This question is of a higher importance since decentralized platforms represent also a new business model and a new logic on the way attention is required online. A decentralized network moves away from the two sided tale of Media Anderson & Gabszewicz (2006). Indeed, Facebook like newspaper make their revenue from advertisements and their algorithm is guided to attribute the attention of users to companies' advertisements. From its definition : "Attention is focused mental engagement on a particular item of information. Items come into our awareness, we attend to a particular item, and then we decide whether to act." Beck & Davenport (2001), attention has become a scarce resource since that companies compete for it. Deciding to use a decentralized architecture puts the question of attention back to the hands of users and fight against the so called "Social dilemma". By giving the possibility to host and rule your own server this new type of network creates competition and spread the capacity to guide users through a unique algorithm.

After the buyout concluded in November 2022 of Twitter (now X) by Elon Musk many mentioned another microbbloging platform called Mastodon. Founded by the developer Eugene Rochko in 2018 Mastodon is the main software in the Fediverse. Legally, Mastodon is a non-profit company registered in Germany. This new OSN was designed to be a substitute to Twitter and its user-base rose from 500 thousands to 2.5 millions within the month following Elon Musk's acquisition. Although many people joined the platform, the "migration" was not dramatic and Mastodon is now around 1.5 millions of user-base. There are many reasons why Mastodon is not taking off as traditional OSN did. One obvious reason is that Mastodon is a non-profit and there is no private data collected on the Network. For this reason many previous users of Very Large Online Platforms are not satisfied with the service Mastodon proposes. But the difference in architecture itself, centralized or decentralized, has a more ambiguous effect on the expansion of the Network.

All the servers on Mastodon give access to the same network. When a new user creates an account it has to choose a server. Choosing a server rather than another one will only affect the user's position on the Network. Let's say you decide to join the server econtwitter.net you will have access to two timelines, one general and the other one with only people interested in economics. The entire network is still reachable but you might start to follow people with your server's interest. By using the ActivityPub Protocol, Mastodon is giving the opportunity to online communities like academics in economics to enjoy a service like the one of Very Large Online Platforms while experiencing a full control on the server's policy. The service is intended to be a substitute to traditional OSN like Twitter (now X), nonetheless decentralization creates a market structure where users choose a server to be active on and servers enter the network, like a market with free entry. But how can we make sure free entry is efficient for the network? According to economic literature decentralized markets can be victims of inefficiency in the number of entry Mankiw & Whinston (1986). My master thesis's objective is to give an empirical analysis following Berry & Waldfogel (1999) of the potential inefficiencies on Mastodon's network regarding the number servers.

The contribution of my master thesis is to apply methods of empirical Industrial Organizations to propose a new perspective on the Mastodon case. There is a growing literature on the Decentralized Web in other disciplines and Mastodon has been used as a good example of it. Following the recommendations of Kobbi Nissim (2020), social sciences can help to understand how people adopt new technologies. After understanding the market structure behind Mastodon's network I believe economics becomes more than relevant to the case. Recent work in economics showed that the frontier between for profit and non-profit firms can be thin Fioretti (2022). Although Mastodon is registered as a non-profit its scaling capacity is constrained by economic reasoning of users to spend time on the application and to create servers.

The paper is organized with the following sections: after a literature review of the different topics (II) there will be the description of the Data (III), then the description the Model (IV), the identification (V), there will then be The estimation (VI), a discussion of the results (VIII), succeeded by the proposition of counterfactual (IX), the second last section will be the limitations (X) and finally I will conclude (XI).

Part II

Literature Review

1 Economy of attention

My paper relies on the Economy of attention literature in the sense that Mastodon represents a different paradigm in term of optimization of users time. Since Mastodon is crowdfunded and it is not suppose to expose users to advertisements the way attention is required is different. Mastodon is closer to recreate a real life social interactions, more hazardous but also more meaningful according to the testimonies about Mastodon. The starting point we take in this literature is the work of Simon *et al.* (1971) who seems to be the first to have underpinned the scarcity problem around attention against information. The metaphora he gives is the one of a land with salad and rabbits. Rabbits represent the information and salad the attention. We see that in this story information reproduces quickly where attention stay constant. Even if this work were applied to organization and not to the digital world, recently in this literature attention has been as been used in a model to evaluate the value of these free goods like OSN Brynjolfsson *et al.* (2023). The authors in this work propose a model quantifying the consumers spending in time of digital services since the price of most OSN is zero. They then quantify to contribution of OSN by running a counterfactual experiment. We could say that if all OSN were adopting the business model of Mastodon they would be less addictive. However from another point of view Mastodon is not using modern algorithms and building a network might be at first longer and more painful. The monetary driven algorithm used by traditional OSN might be a way to guide the users to useful information and so be more efficient than the model proposed by Mastodon. But not a great share in this surplus of time spent online is going to the users where they represent the real resource. The economy of attention literature is the reason why we use the Monthly Active Users variable as a criterion of performance for Mastodon's servers in this paper.

2 Industrial Organizations literature

"Economists typically presume that free entry is desirable for social efficiency." Mankiw & Whinston (1986) is the first sentence of Mankiw and Whinston's paper about free entry and social inefficiency and is building block of my paper. From an intuitive point of view Mastodon must be more efficient than a centralized OSN because it enables to split the hosting cost between many individuals. We can think of traditional OSN as monopolies where users pay a higher cost (by giving information on their preferences and attention). By allowing users to self host the platform and fix their own rules we would think about it as socially beneficial. However, according to Mankiw and Whinston, under business stealing effect and fixed set-up cost a free-entry homogeneous market theoretically leads to too many entries. Business stealing and set-up cost are sufficient conditions for free entry to lead to a duplication of fixed cost due to the fact that firms don't internalize their business stealing externality. Following this theoretical result, Berry and Waldfogel proposed an empirical strategy to estimate such inefficiency from market data Berry & Waldfogel (1999). In their paper Berry and Waldfogel introduce the case of radio stations. Radio stations offer a similar service with a difference in variety (Jazz, Rock ...) and they all have to pay an entry cost represented by the frequency rights. In the same way Mastodon creates an homogenized service by giving a standard to servers that propose variety in communities (tech, academics...). The set-up cost is indeed lower on Mastodon since creating a server is easier than creating a radio station. Nonetheless it seems that hosting and administrating even a small server can be time consuming and most of servers call for donations to recover this cost. Berry and Waldfogel use a nested logit formuation for listener's utility functions following Berry (1994). Using US data on different geographical markets they estimate a very high business stealing and the total welfare loss of free entry is 45% of revenue. In the continuation of this work and with the participation of other authors, Berry *et al.* (2016) showed that using the same data but with more heterogeneity between stations in the estimation leads to a lower estimation of the business stealing effect. My paper will focus on the 1999 specification and I keep for further work the possibility to use thiner nests and heterogeneity. The other important feature of Mastodon is that they don't use advertisement as a source of revenue. Economists were looking for the growth of platforms in the indirect network effect that advertisement induces Rysman (2004). Network effects has been raised as the main argument for the concentration of OSN. However Rysman's paper is based on the yellow pages example where advertisement is the information consumed by consumers. Similar studies has been made on digital networks Farronato *et al.* (2020) and they show there is a trade-off in consumer's surplus between high network effects when users are on the same platform and lack of diversity. The protocol proposed by ActivityPub may seem like a reasonable answer to this trade-off and it's not a secret that traditional OSN's firms like meta are interested in this model Meta (2023).

3 Mastodon literature

My paper connects also to the existing work on the OSN Mastodon and Decentralized OSN. Powered by the W3C, decentralized networks gained in interest among computer scientists. Projects like the one of the MIT Solid (2023) or the one of Kobbi Nissim from Georgetown University Kobbi Nissim (2020) started to spread the idea that decentralized architecture can guarantee a better protection of privacy. Despite that the application to decentralized architecture to social media happened before Mastodon and the Fediverse it appeared as being the first large scaled experiment of it. Zignani *et al.* (2018) was the first paper published on the topic after Mastodon started its activity. They used network theory to study the evolution of the social media, and according to their analysis "Mastodon is set to become a valid alternative to established platforms like Twitter." Zignani *et al.* (2018). One year later another study on similar data showed that users on Mastodon concentrate on few servers having then a central position in the network "10% of instances host almost half of the users" Raman *et al.* (2019). The authors mention it as a challenge for the decentralized web since the network becomes dependent on the well being of these few servers. Although it can be a problem for technical reasons, from an economic point of view, concentration is not necessarily a challenge. If some communities are bigger than others it can still be efficient to provide the service in a decentralized architecture. On the other hand if servers don't actually bring their community when they enter the network but they mostly steal the business to other severs, that can be a challenge that decentralized web didn't mention yet.

Part III

Data description

The Data set I use was scrapped from the 29th of January 2023 to the 1st of June 2023 at a rhythm of once every two days. My main data on the Mastodon servers come from the api instances.social and the other data studying the patreon accounts of Mastodon servers come from scrapping data. While making my query for the api I specified from the very beginning to the end of the data collection that servers must have at least 100 users. Here I would like to introduce some vocabulary in my data: a user is an account created by someone, at some point, that is not necessary active ; an active user is an account that connected to the server at least once in the last month (Monthly Active Users) ; and an instance is the specific name servers are given on Mastodon (for simplicity I will stick to "server" in my master thesis) ;

finally a market in my Data is a date and each observation is identified by a market (date) and a server's name. After a little data cleaning the summary statistic of the main Data set collected on instances.social are in the next figure.

Table 1: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
users	79,091	5,864.656	41,259.770	100	1,150,343
active_users	65,000	654.518	3,760.519	1	128,788
statuses	79,091	523,236.200	3,781,624.000	−2	70,552,342
connections	79,091	16,367.900	37,630.610	0	1,414,322

From this procedure we see that servers differ a lot in the intensity of their activity. With a mean of active users around 654 active users and a maximum 128,788 (for the main server mastodon.social). We note that many servers have a NA value for their active users in our Data set from the number of observations in this row. The only unexplained statistic is the negative minimum number of statuses that is probably due to an error of data collection that I couldn't correct in the data cleaning. According to the literature on the Economy of Attention we can consider the Monthly Active Users (MAU) as a consumer (spending scarce time against accessing the network). Then the total number of MAU represents the size of total market. We would like to know in a reduced form regression how this total size of the market evolves when we change the total number of servers from market to market. Our objective there is to get a sense of the market stealing in our Data.

This graph shows a clear evidence of business stealing in our Data set since as the number of servers grow the total size of the market seems to diminishes.

Part IV

Model

Choice model for users

Our model is a simultaneous choice model where users decide on which nest $g/0$, which server j to be active and servers owners decide whether to enter the market at each period t. The first part of the model, the choice users make, is a classical multinomial nested logit McFadden (1977),Berry (1994). In this model there is one nest for the main server

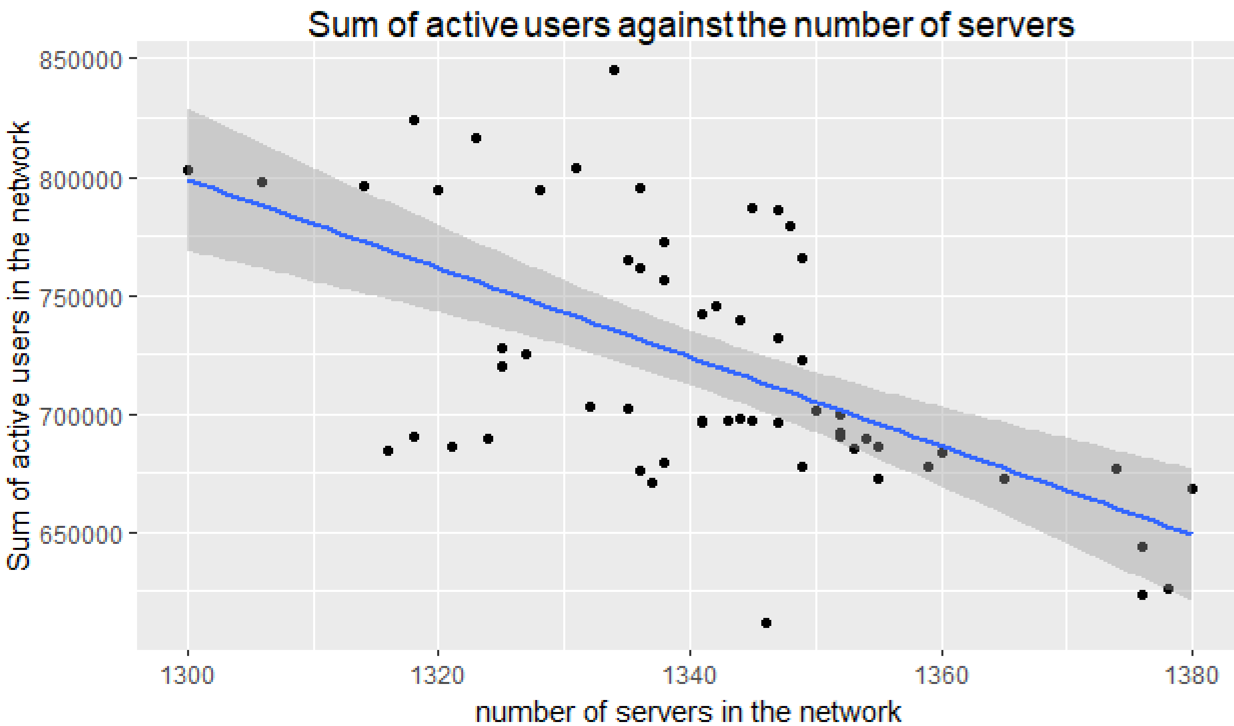


Figure 1: Sum of active users in Mastodon against the number of servers in each market

(mastodon.social) considered as the outside option (denoted by 0 as its utility is normalized) and one nest for the other servers. To explain our choice of nest I add a screenshot of the main phone application proposing to join either the main server or all the other servers (annex 2).

Users derive utility from the following utility function, to alleviate the notation we don't write the time subscript, but this equation holds for each market t:

$$u_{ij} = \delta_j + \nu_i(\sigma) + (1-\sigma)\epsilon_{ij} \tag{1}$$

Note that the second nest is normalized so we simply write the utility for the nest with all the other servers. The identification of this equation is based on the distribution of the term $\nu_i(\sigma)$ that follows the one specified in Cardell (1997) and is the same error term for all the servers. For the identification to be correct we have to understand that as σ goes to 0, $\nu_i(\sigma)$ goes to zero and let us back with a classical multinomial logit where all the servers have a different utility for user i. And as σ goes to 1, the multinomial logit error term disappear and all servers have the same deviation in the utility of user i. For the entry model to be tractable, we will have to impose δ_j to be constant across servers, but for the identification of σ we can use differentiation in mean utility across servers.

Entry model for servers

We make the assumption that revenue per active user is the result of a Cournot game that only depends on number of entry. The role of the identification will be to find a functional form to this theoretical formula. We start from the formula of revenue per listener in the case of radio station Berry & Waldfogel (1999):

$$p\left(N\right)=p\left(Ns\left(N\right)\right) \tag{2}$$

For the entry model to be tractable we assume all servers are ex-ante identical. So the market share becomes the size of the inside nest divided by the number servers at that period. To find an empirical application of this formula we will rely on the data collected on patreon.com. Even if there is only a subset of servers in this data we have in it an idea of how much donations can a server gather at each period.

From there we can build the revenue function for the servers in each market:

$$\pi\left(N\right)=Mp\left(N\right)s\left(N\right)-F \tag{3}$$

Where M is the market size so in our case the total number of active users at each period. We assume the existence of a fixed cost that doesn't vary with the number of active users. Which is what we empirically see when checking the prices of the hosting platforms. The prices of these services is quickly decreasing when we take the per active user price as shown in the following figure.

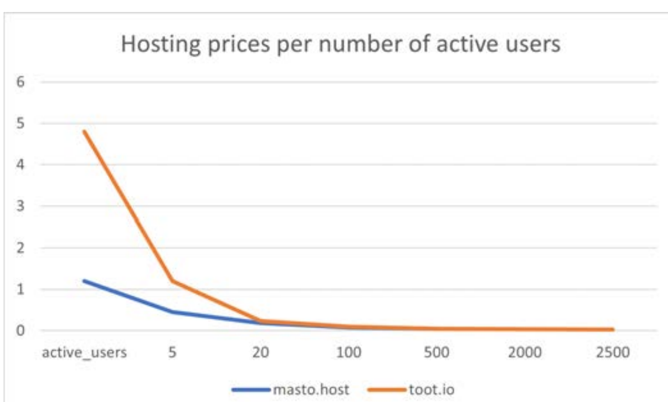


Figure 2: Price per active users of the two main hosting services for Mastodon's servers

Then according to Mankiw & Whinston (1986) the number of entry in each market must respect the two conditions

$$\begin{aligned}\pi\left(N_e\right) \geq & 0 \\ \pi\left(N_e+1\right) < & 0\end{aligned}$$

Which states that the number of servers we observe in each market is the result of an equilibrium in a Cournot competition model of entry.

Part V

Identification

Identification of the choice model for users

First we parameterize the mean utility to be dependent on server's characteristics:

$$\delta_j = x_j\beta + \xi_j$$

In our specification we use the activity on each server, which means in our data the number of connections and the number of statuses.

According to the derivation presented in Berry (1994)the model is identified by an instrumental variable regression solving the following equality:

$$\log\left(s_j\right) - \log\left(s_0\right) = x_j\beta + \sigma\log\left(s_{j/g}\right) + \xi_j$$

Where $\log\left(s_{j/g}\right)$ is the log of the share of server j in the nest where there are all the servers except mastodon.social. This function of the data is endogenous since s_j is present on both side of the equation. For identification we then need an instrument that shifts the share among the nest with all servers but do not change s_j. The best instruments I found are: the share of open servers in each market and a dummy for the extension ".social". We have information on whether servers are open to registrations or closed as a dummy. Taking the share of this excluded from the station j we can consider that a new user is more likely to be active on the server j the more the other servers are closed (by substitution). The second instrument is based on extension of the domain of servers. When servers need to set their domain name they need to choose an extension (.net, .com, .uk etc...). After some research it appears that certain domain or more expensive and/or more trendy than others. There is an extension that is very trendy among servers and it is the ".social". Because this extension makes the server in general more trendy I decided to base my second instrument on this parameter. The second instrument, based on extension, is a dummy variable equal to one if the server's domain has the extension ".social".

Identification of the entry model for servers

We adopt the same constant elasticity specification as Berry & Waldfogel (1999) for the function p:

$$p(N) = \alpha(S(N))^{-\eta}$$

Where α is a parameter that shifts the demand and $S(N)$is the share of the nest with all the servers except the main one.

We assume that α is a function observed and unobserved demand shifters. In our estimation it will be the share of servers that call for donations in their welcome page. We can then rewrite the log in this way:

$$\log(\alpha) = x_t\gamma + \omega_t$$

We can then identify our parameters by running the following regression:

$$\log(p_t) = x_t\gamma - \eta\log(S_t) + \omega_t$$

Where p_t is the observed money collected through patreon.com per active users and S_t is the share of the nest with all the servers except the main one at each period (the markets in our data are periods). To run this regression we will instrument $\log(S_t)$ by the share of extension .social in the period and the share of open severs in the period.

We finally assume a parametric form for the log of the fix cost:

$$\log\left(F_t\right) = x_t\mu + \lambda\nu_t$$

Where ν_t is distributed following a normal standard between zero and one. We are stating that the fixed cost is determined by an observed demand shifter (in our case the share of server calling for donations in their welcome page) and a deviation normally distributed (some period the fixed is higher for unobserved reasons). To identify these two new parameters we use the conditions we wrote on the revenue function, for each period to be an equilibrium. We can derive from there a moment condition. To alleviate the notation we don't denote the time/market subscript.

$$\begin{aligned}
Mp\left(N\right)s\left(N\right) - F \geq & 0 > M_t p\left(N+1\right)s\left(N+1\right) - F_t \\
Mp\left(N\right)s\left(N\right) \geq & F > Mp\left(N+1\right)s\left(N+1\right) \\
\log\left(Mp\left(N\right)s\left(N\right)\right) \geq & \log(F) > \log\left(Mp\left(N+1\right)s\left(N+1\right)\right) \\
\log\left(Mp\left(N\right)s\left(N\right)\right) \geq & x\mu + \lambda\nu > \log\left(Mp\left(N+1\right)s\left(N+1\right)\right) \\
\log\left(Mp\left(N\right)s\left(N\right)\right) - x\mu \geq & \lambda\nu > \log\left(Mp\left(N+1\right)s\left(N+1\right)\right) - x\mu \\
\frac{\log\left(Mp\left(N+1\right)s\left(N+1\right)\right) - x\mu}{\lambda} \geq & \nu > \frac{\log\left(Mp\left(N\right)s\left(N\right)\right) - x\mu}{\lambda}
\end{aligned}$$

So we know that assuming we observe an equilibrium this inequality is probably holding. However since we know the distribution of ν we can write the likelihood function corresponding to this inequality.

$$L\left(\theta\right) = \Phi\left(\frac{\log\left(Mp\left(N\right)s\left(N\right)\right) - x\mu}{\lambda}\right) - \Phi\left(\frac{\log\left(Mp\left(N+1\right)s\left(N+1\right)\right) - x\mu}{\lambda}\right)$$

Finally we can use the partial derivative of this likelihood function as a set of two moment conditions and our parameters are identified.

Part VI

Estimation

The estimation of the business stealing parameter is summarized in the following table:

Table 2: Estimation of the business stealing

	Model 1	Model 2	Model 3	First stage
(Intercept)	1.260 13***	1.714 10***		
	(0.307 56)	(0.213 24)		
σ	0.941 568***	0.993 606***	0.985 067***	
	(0.035 447)	(0.024 427)	(0.010 542)	
connections		$-1.4561e{-8}$	$-6.2769e{-8}$	
		(9.1297e−8)	(8.2100e−8)	
statuses		$-5.8764e{-9}$	5.1577e−8*	
		(3.9143e−9)	(2.3516e−8)	
share of open servers				−42.678
				(62.969)
share of open servers $\times$ .social extension				−3.7525**
				(1.3160)
Num.Obs.	65 000	65 000	65 000	65 000
R2	0.984	0.987	1.000	0.983
R2 Adj.	0.984	0.987	1.000	0.982
F	74.011	98.616	65.291	
Std.Errors	by: name	by: name	by: name	by: name
FE: Date			X	X
FE: Name			X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The first model takes mean utility of servers to be a constant. In the second model we add heterogeneity in mean utilities by adding the covariates statuses and connections. In the third model we add the time and server fixed effects to the regression. All three models are IV and their F-stat is reported. The instruments we use are the one tested in the first stage (last column): the share of open servers at each date, the dummy of the extension ".social" and their interaction. Note that in the first stage the coefficient of the extension instrument doesn't appear since it is co-linear with the server fixed effect.

For the entry model we used the same moment conditions as Berry & Waldfogel (1999). I found a point of convergence around starting values of 0.5 to 1 which sounds fair starting values compared to the lambda of 0.22 found in BW 99. The parameters I find by minmizing the GMM objective function in R are $\lambda = 0.05386161$ and $\mu = 4.09643206$. We can plot these results in a levelplot but the precision of the plot is not good enough to see a unique point. The plot only show that these coordinates indeed lower the objective function.

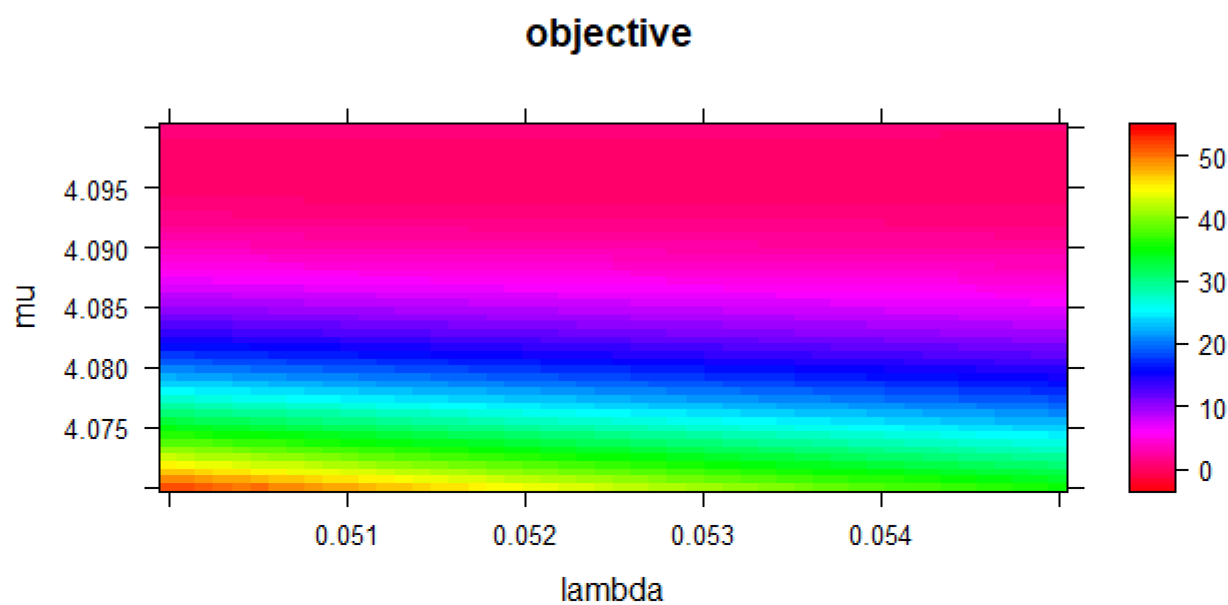


Figure 3: Levelplot Entry model

Part VII

Discussion of the results

I found it puzzling on how to interpret the estimate I find on business stealing. It is even higher than the one found for radio stations in Berry & Waldfogel (1999)Berry *et al.* (2016) which we can think it is a real problem for Mastodon growth. Our finding suggests that servers steal active users from each other without creating new active users in the meanwhile. But looking at the decentralized architecture of Mastodon we can think this result is comforting the promises of institutions like Mastodon. Indeed, the revealed objective of Mastodon is not to reproduce Twitter but to create a space where all servers have freedom but no power over the network. Where past studies showed a risk of concentration of the servers among few big providers Raman *et al.* (2019) our results shows an opposite movement. It looks like even though there are big servers concentrating a large part of the network all the other servers exchange active users constantly. And it is important to know because it means that servers can't deviate too much from the service they are supposed to propose (free speech, control of spam ...) otherwise active users switch easily to another server. Business stealing for Mastodon is both a hurdle for growth but also the guarantee for the network to work correctly. Imagine a network where servers capture their audience and can prevent competition that would mean servers that have power over the network what would reproduce the Twitter problem. To nuance our result it's important to mention that we would have to test on a different time period to see if the results hold in the same way. During the time of the Data collection Mastodon user-base was mostly decreasing which is not the case in every time period since Mastodon was created. It may probably be the case that business stealing is varying over in the same network.

Regarding the results of the entry model, it looks like I am finding a lower standard deviation of log fixed cost rather than Berry & Waldfogel (1999) which finds 0.22 against around 0.05 in our model. This goes in the direction that fixed cost on digital platform are very low and not very volatile. Indeed, a mastodon server doesn't cost much to settle in and its price doesn't vary neither. We didn't do the last paper of BW 1999 to see the impact on welfare, however this second result might contradict the first one. Inefficiencies might be lower than the ones of radio stations even though there is a higher business stealing. Since mastodon servers don't pay exploding fix cost we can think that having many servers is a less important problem.

Part VIII

Counterfactual

The main counterfactual I would like to do is to use another time period to compare the results on the business stealing parameter. Indeed I know from the Data of other studies Zignani *et al.* (2018) that the market of Mastodon was expanding at other time period. It would be interesting to compute the business stealing parameter in a time of market expansion.

Another counterfactual and good exercise would be to include more or different covariates in the mean utility of servers. There is in my data a variable that I couldn't make sense called "http score" and taking a letter from "A+" to "E". I think there is an idea of the time response of the server but since the API was giving no other information I preferred not using it. Finding variables closer to describe the user experience than the number of connections or the number of statuses would be a good exercise.

Finally, a very good improvement of my work would be to find thinner nest than the one we used following Berry & Waldfogel (1999). Indeed I have in the data the topic of each servers. Going from the empty set for generalist servers to "tech" or "lgbt" for specific servers these topic could be used to create thinner nests where the Independence of Irrelevant Alternatives (IIA) may hold with more certainty. It would also probably change our estimate of business stealing maybe substantially according the 2016 follow up of Berry and Waldfogel work Berry *et al.* (2016).

Part IX

Limitations

There are many limitations to the piece of work I propose as my master thesis. The first one is the fact that my data was collected every two days and so considering each time period as an independent market in my identification can be criticized. I think it is necessary to correct for autocorrelation in my estimation which is something that I only do thanks to fixed effects. The second limitation is probably my estimation procedure on the coefficients of the entry model. Indeed the optimizer is not converging for starting points higher and lower than the interval $[0.5, 1]$. Another method than the one I usee (Nelder Mead) would be better to put positivity constraints on the parameters (for example L-BFGS-B). The last limitation and not the least is that the two data-sets don't included the same number of servers. The Data-set scrapped on patreon only includes around 50 servers where the instances.social Data-set has 1300 approximatically per date. It looks obvious to say that there many other ways to gather money than patreon.com and that this data is a narrow subset of the money collected by servers.

Part X

Conclusion

My research contribution was to show that decentralized OSN induce a market structure that looks to be inefficient in the number of servers. However we underline that even though it might be a problem for the growth of the platform, business stealing looks to be an important feature to guarantee Mastodon's promises. There is still a lot to be done on decentralized platforms in economics. My work has also to be taken as a research project that open the path to a new object in economics. How do people control each other in decentralized platforms? What about the production of information and the scarcity of attention? And of course the business stealing effect can be better identified as the counterfactual part and limitations part suggest. Social networks have a great impact on the way people communicate, and as we know in economics the mechanism is link to its outcomes, the experiment of decentralized networks is an economic experiment that may lead to a different equilibrium than the situation described in The Social Dilemma.

References

Anderson, Simon P, & Gabszewicz, Jean J. 2006. The Media and Advertising: A Tale of Two-Sided Markets. *Handbook of the Economics of Art and Culture*, **1**, 567–614.

Beck, John C, & Davenport, Thomas H. 2001. The Attention Economy: Understanding the New Currency of Business. *Boston, MA, Harvard Business School Pres.*

Berry, Steven, Eizenberg, Alon, & Waldfogel, Joel. 2016. Optimal Product Variety in Radio Markets.

Berry, Steven T. 1994. Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics*, **25**(2), 242.

Berry, Steven T, & Waldfogel, Joel. 1999. Free Entry and Social Inefficiency in Radio Broadcasting.

Brynjolfsson, Erik, Kim, Seon Tae, & Oh, Joo Hee. 2023. The Attention Economy: Measuring the Value of Free Goods on the Internet. *Information Systems Research*, Aug.

Cardell, N. Scott. 1997. Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity. *Econometric Theory*, **13**(2), 185–213.

Farronato, Chiara, Fong, Jessica, & Fradkin, Andrey. 2020. *Dog Eat Dog: Measuring Network Effects Using a Digital Platform Merger*. Tech. rept. National Bureau of Economic Research.

Fioretti, Michele. 2022. Caring or Pretending to Care? Social Impact, Firms' Objectives, and Welfare. *Journal of Political Economy*, Sept., 000–000.

Jonathan DeYoung. 2020 (June). *Decentralized Tech Can Protect Activists From Social Media Crackdowns*. https://cointelegraph.com/news/decentralized-tech-can-protect-activists-from-social-media-crackdowns.

Kobbi Nissim. 2020. *(167) Data Cooperatives in the Real World: Progress and Challenges - Ligett, Nissim, Ruhaak - RxC 2020 - YouTube*. https://www.youtube.com/watch?v=vUbuOiyosjI.

Mankiw, N. Gregory, & Whinston, Michael D. 1986. Free Entry and Social Inefficiency. *The RAND Journal of Economics*, **17**(1), 48.

McFadden, Daniel. 1977. Modelling the Choice of Residential Location.

Meta. 2023. *About Threads and the Fediverse | Instagram Help Center*. https://help.instagram.com/169559812696339.

Raman, Aravindh, Joglekar, Sagar, Cristofaro, Emiliano De, Sastry, Nishanth, & Tyson, Gareth. 2019. Challenges in the Decentralised Web: The Mastodon Case. *Pages 217–229 of: Proceedings of the Internet Measurement Conference*. Amsterdam Netherlands: ACM.

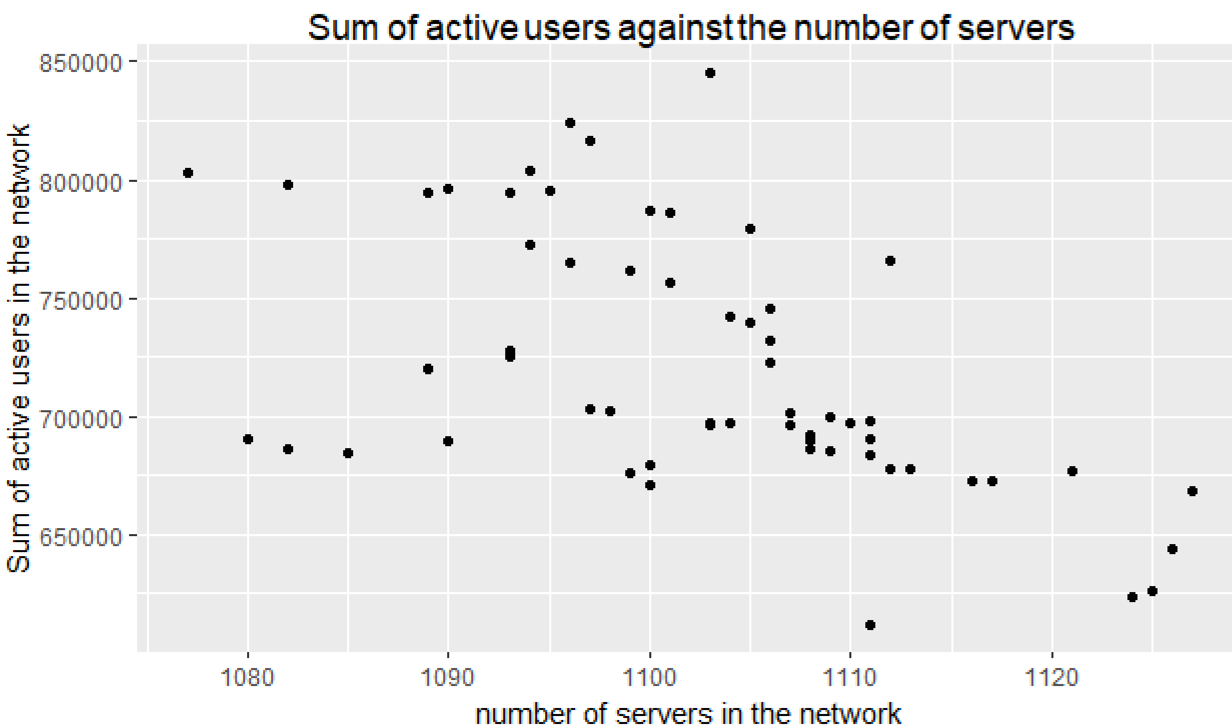
Rysman, Marc. 2004. Competition between Networks: A Study of the Market for Yellow Pages. *The Review of Economic Studies*, **71**(2), 483–512.

Simon, Herbert A, *et al.* 1971. Designing Organizations for an Information-Rich World. *Computers, communications, and the public interest*, **72**, 37.

Solid. 2023. *Solid*. https://solid.mit.edu/.

Zignani, Matteo, Gaito, Sabrina, & Rossi, Gian Paolo. 2018. Follow the "Mastodon": Structure and Evolution of a Decentralized Online Social Network. *Proceedings of the International AAAI Conference on Web and Social Media*, **12**(1), 541–550.

Annex 1 : The reduced form regression when removing the servers with NA in their active user column.



Annex 2 : Screenshot of the main application (motivation for the nests)