The effect of broadband Internet access on the number of firms in Brazilian municipalities.

Rafael Neves

Economics

Defended March 22, 2023

Committee:

Dr. Scott Savage, Department of Economics, Thesis Advisor

Dr. Brian Cadena, Department of Economics, Honors Council Representative

Alan Paradise, Department of Computer Science, Outside Reader

Abstract

This paper studies the relationship between Information Communication Technologies and market efficiency. A double-log regression model is estimated on panel data from Brazilian municipalities that relates the number of firms to Internet penetration and quality, and other market controls. The results show that although the total number of firms in a given municipality decreases with increase in broadband Internet access, the number of new firms and closing firms both increase with access. The findings suggest that while broadband Internet access reduce the overall number of firms of the country, at the same time, it increases market efficiency and allows for new firms to enter the market.

1. Introduction

The Internet and related technologies caused a dramatic technological change, which in the beginning were small and incremental, but more recently have been more drastic and restructured many ways in which businesses and activities were carried out (Borenstein & Saloner, 2001). They allow for more ways to communicate, connecting people from all over the globe, both synchronously and asynchronously, increasing collaboration and the amount of information available to anyone that can access it. The Internet's impact on society is still hard to measure and is constantly evolving. Many markets, however, were fast to adapt and thrived with this innovative technology. Large companies like Google, Meta and Amazon rapidly grew and influenced the way information and trade are done both on a micro and macro level. Many other businesses also started to emerge in local economies providing services that were previously only available in person, such as goods, entertainment, and finances. The focus of this research is on businesses and their relation to the Internet. Thus, a question is formed: what is the effect of broadband Internet access on the number of firms?

The wider adoption of the Internet allowed many new companies and industries to emerge. The technology increases consumers' information on the products and services available, thus decreasing search costs. It reduces inefficiencies from slow communication and allows for more complex supply chains and gives people tools to learn new skills and provide new services. It is important to note that the infrastructure required to provide broadband Internet has high upfront fixed costs, but comparatively low maintenance and expansion costs. The cabling and machinery necessary to operate this information and communication technology (ICT) requires previous infrastructure to exist. Namely, a robust electrical system, as the machines need constant and reliable power to work properly. Also, adequate road infrastructure is needed, not only to allow maintenance of the equipment, but also because a lot of cabling is done under or by roads. Additionally, relay stations, computers, servers, physical security of the equipment all can be considered part of the upfront costs. However, once the initial construction is concluded, maintaining, and expanding such systems has comparatively lower costs. So once a region starts to have these basic requirements, it is easier to roll out services to other nearby areas and improve upon what is there. Factors such as the quality of the service, which technologies are used, and the legal system supporting it impact people's use of the Internet and the scope of its access. For example, if a municipality gets access to a fiber optic connection when it only had xDSL connections before, it can use its existing cabling infrastructure to provide households with better quality Internet connection which has lower costs than building everything new.

This paper assumes that there is a relation between information and communication technologies and market effects. As ICT infrastructure improves it increases the available information for consumers and producers, which in turn decreases search and matching costs. This means that more goods can be traded in the market, but also means that inefficient firms will be pushed out of the market. Meaning that ICTs can lead to frictionless markets, or at least approximate the market to that assumption. Additionally, the Internet allows firms to operate over larger areas and with lower costs, as orchestration of their services can be automated and facilitated by the ICT. This increases competition as consumer preferences can now be better matched, but at the same time creates new opportunities for new entrepreneurs as it decreases the barrier of entry. For example, a small local bakery before would have to have a storefront to sell their goods, now can own just a kitchen and deliver to their customers on a need basis. This reduces fixed costs, while better matching production to consumers demands. This paper also assumes that all actors in the market are digitally literate. Implications of this would be an increase in the total number of firms as access increases, or, perhaps, a change in the number of firms entering the market or leaving it.

Previous literature on this topic focused on mobile phone infrastructure to measure the relation between ICTs and market efficiencies (Jensen, 2007; Aker & Mbiti, 2010). For regions such as South India and Central Africa in the early 2000s the technological leap to mobile phones caused a substantial change in the communication possibility and had noticeable implications for how the market operates. In the case of South India these changes were attributed to increases in efficiency, as the communication possibilities for the individuals in the study increased. There were positive impacts measured on prices and a decrease in waste and non-productive time (Jensen, 2007). This was observed for cellular phone communication, not Internet access, which is a major distinction between previous work and this paper. Although both cellphones and the Internet are ICTs, the effects of Internet access on the market are expected to have unique distinctions from the effect from telephone communication.

For the places that adopted the Internet earlier, such as the United States or the United Kingdom, these changes were slow as the technology was still in its infancy. Even if a household gained Internet access it wouldn't have as many options of what to do online as a user has today. Therefore, using them as the basis for this study would require data over long time periods which could limit the analysis without an incredibly robust dataset. The solution was to choose a place that shifted rapidly from not having anything, in most places, to sufficiently providing its citizens Internet connection. The country of Brazil fitted these needs. Outside of large metropolitan areas, vast regions of the country had no reliable Internet connection during the first decade of the 21st century. In 2009, the president of the National Telecommunications Agency (ANATEL), Ronaldo Sardenberg, said that the country only 15 million people had broadband access and that the goal was to reach over 165 million people by 2018 by heavily investing in the sector. This would mean that roughly 80% of the population would have Internet access.

In 2010 the federal government launched the National Program of Broadband (PNBL) with the goal of offering broadband access to all municipalities by the year of 2016. After the end of the PNBL in 2016, the project Intelligent Brazil (PBI) was launched in the same year with the goal of expanding the fiber optic infrastructure throughout the country. Later, in 2018 the PBI was revoked, and it became one of the competencies of ANATEL to, in general, expand the telecommunications infrastructure of the country. The period observed will be for the years of 2010 until 2019, as it covers the years for which these programs were in effect and one year after. Noting that Internet access continued to grow after the conclusion of the projects. To avoid noise in the data due to the COVID-19 pandemic, the years of 2020 and 2021 will not be used, even though the reports for those years from ANATEL and the Ministry of Finance (MF) could provide useful insights for further research.

During the period observed by this research there were also policy changes that facilitate the process of opening, but especially closing businesses, as well as an increase in the number of digital methods to operate a business. In the early 2000s it could take well over 150 business days to get a firm's closure approved, while in 2019 that number dropped to an average of 44 hours (Ministry of Finance Map of Firms report, 2020). Notably, the National Network for the Simplification of Registration and Legalization of Companies and Business (REDESIM) created in 2007 had the goal of integrating and facilitating the entire process of opening, operating, and closing a firm, with a focus on smaller businesses. Not only the efforts of REDESIM, but also the Law of Access to Information (Lei nº 12.527) of 2011 forced the federal government to provide citizens with public information through the Internet. This in turn incentivizes an increasing number of services to be made available online, which facilitates the process of opening and closing businesses. As even more evidence of this, note the Way Simpler program (BMS) in 2015, or the Brazil Efficient program (BE) in 2017. Both programs provide new tools for entrepreneurs to open, operate and close their businesses in line with REDESIM protocols. Even more so when recalling the PNBL and PBI programs, which increased the infrastructure for Internet access through the country.

The analysis of the panel data indicates that Internet access has an important positive impact on the rate at which new firms are created or close. It shows that independent of overall market trends, as municipalities gain more access, their markets become more competitive, although it does force more firms out of the market than allows for new firms to be created. It indicates that increasing Internet access has a comparatively smaller, but negative, effect on the total number of firms for a given. The findings also go against initial the intuition as they indicate that Internet access actually decreases markets.

This paper is organized as follows. Section 2 will go over where economic thought is on this topic and where it is possible to add to it by stating the hypothesis. Section 3 presents the model that will be used and section 4 which introduces the data. Then section 5 will go over analyzing the results and providing some interpretation of the results. Concluding with section 6 and a discussion of the results and future work. Additionally, there are two appendixes, the first on the construction of the dataset and the second on some robustness checks for some of the regressions.

2. Literature Review

The literature for this thesis can be broken down into two large groups of work. The first one being how information and communication technologies (ICTs) impact markets and consumers (Jensen, 2007; Aker & Mbiti, 2010; Forman et al., 2012; Brown & Goolsbee, 2002; Kim & Orazem, 2017). The second one is on how market efficiency impacts the formation of new companies (Barbosa, 2016; Phillips & Kirchhoff, 1989; Manjón-Antolín & Arauzo-Carod, 2008). There is also some literature on the specificities of these phenomena in Brazil. Giving us some great insight into some nuances that appear in the data. Such as, Internet access being considered a luxury good especially for areas with widespread access (Silva et al., 2019). Also, how e-commerce allows national businesses to participate in international value chains, instead of simply importing goods and services (Tigre, 2006).

On the relation between ICTs and markets most of the research either focuses on producers, or consumers, getting better access. This would present itself as either measuring access, as how people can use the ICT, quality of access, as what is the condition of the access, or use, as how people utilize the technology. Then researchers would measure something that the theory would assume to be closely related to the desired metric, such as the number of cellphones per individual, or coverage, or penetration rate, or a combination of them (Jensen, 2007; Forman et al., 2012). The principle behind why ICT improvements lead to more competitive markets is that as communication capacity increases, it reduces the friction of supply and demand, as day-to-day fluctuations would be more easily absorbed by the market. It approximates the market to the perfect information assumption by reducing search and transaction costs. In markets such as the American insurance market for 1995 to 1997, Internet access seems to have reduced term life prices by $8-15\%$ (Brown & Goolsbee, 2002). They conclude that, at least for financial products, the reduction in search costs created by the Internet may lead to larger consumer welfare. This could mean that the benefit from the Internet not only comes from gaining access to the product, but also the ability to know all other products and consumers available.

Notably, previous literature would consider the fluctuations on supply were endogenously controlled by the data of the research and assume that demand was exogenous, or vice-versa. For example, Aker and Mbiti (2010) studied this by measuring cellphone use in sub-Saharan Africa and the socio-economic impacts of if. They assumed that people were able to get cellphone connection when they purchased their devices. This is a fair assumption to make, and it helps to control for issues that might arise with omitted variable bias from the data available. Most research on this topic is based on developing nations and so it can be hard sometimes to get some desired metrics and control for things such as differentiating access to devices and access to connections. This is important to consider because broadband Internet access requires, not only the access, but also a router and a computer to be used. Since this paper measures access, it will be assumed that people have the machines required to use such access.

Regardless of what the researchers focused on, they all concluded that ICT improvements have immediate and noticeable impacts on people's lives as they can access more services and with lower costs (Jensen, 2007). It also shows that firms benefited from these improvements as they can reach larger areas, reduce externalities and benefit from agglomeration economics (Kim & Orazem, 2017). Kim & Orazem's (2007) study suggests that the effect of broadband Internet access on firm creation on rural municipalities has a higher impact on rural municipalities closer to urban centers. Which indicates that when dividing the dataset between rural and urban municipalities controlling for distance to urban centers can be valuable.

The second grouping of the literature is to support the claim that even if ICT development increases market efficiency, would this relate to new companies emerging. Most research focuses on metrics such as survival rate or number of companies. The studies then restrict their analysis for datasets in which they have the required level of data to run models such as Cox Proportional Hazard model (Barbosa, 2016; Phillips & Kirchhoff, 1989). My research diverges in scope, since the question requires data on a nationwide level in Brazil a metric such as survival rate is not as interesting as these would vary wildly even in the same region with similar Internet access (SEBRAE, 2008). Even though their research focuses on survival rates, their results still indicate that, in general, Internet access does attract new firms to the market. Although the results don't explicitly measures Internet access, it does indicate that being able to offer services to a wider audience increases the number of firms (Manjón-Antolín & Arauzo-Carod, 2008).

Possibly the most influential paper on this topic is Jensen's (2007) as he studies cellphone adoption and the market effects on market performance and wellbeing in South Indian fisheries. In this study the author focuses on both producer and consumer metrics and shows strong correlations with how ICTs improve the market and people's conditions. My thesis is remarkably similar to his work. Both investigate developing BRICS nations and their metrics around the time of the 2000s commodity boom. Measuring the use of new ICT access that shows a fast adoption by the population.

The main divergence between my thesis and Jensen's work is twofold. One is that he used cellphones as a measure of communication efficiency, which is a significant metric for the research he has done. However, it can be argued that the Internet has a different effect on how people communicate that would produce different results. As the Internet not only allows for instant communication, but also collective storage of information and automation of services. This means that it also has some specificities that are different from that of phones, mainly the importance of measuring quality of the connection. Which leads to the second divergence as Jensen restricts his research to one market. He uses mostly data and analyzes how the fishing market adopted and behaved with the introduction of cellphones. This restriction allows him to produce very precise results, but it is not clear if his results hold true to a more general nation level. Since this research uses the Internet as the observed ICT, it looks upon the impact it has on firms across the country. Also, it's not a major difference but it is important to note that he tries to find a connection between ICT adoption and both consumer and producer dependent variables, while my research is on exclusively on a producer dependent variable with the use of the number of firms.

In relation to the presented literature my thesis is mostly an application of previous work on a different location and with a unique dataset. The scope of the dataset is unusual, as most researchers focus on smaller areas or markets, while this research uses country wide data. This thesis ponders if the impact of ICTs is general enough to be able to be sensed at a country level.

The hypothesis is that Internet access should have a relatively small positive impact on the number of firms. As the literature on ICT improvements and market efficiency suggests that efficient markets could increase the number of firms that operate in each market. However, it should have a comparatively higher positive effect on the number of firms opening and closing for municipalities, from the evidence from how market efficiency allows for new firms to enter the market and how increase in information can increase competition and push firms out of business.

3. Methodology

Understanding the relation between the number of firms and Internet access requires some way to understand how these changes manifest over time and location. Thus, a panel data that examines 5556 Brazilian municipalities over the years of 2010 to 2019 was constructed.

The form of the models uses a log-log specification to alleviate potential heteroskedasticity of the data. An implication of this specification is that estimated coefficients can be conveniently interpreted as elasticities. Thus, measuring the effect can describe how the number of firms varies for each observation through time. To analyze the constructed database the following model will be used:

$$\ln(firmTotal_{it}) = \beta_1 \ln(access_{it}) + \beta_2 ln(gdppc_{it}) + \beta_3 \ln(pop_{it}) + \alpha_i + \phi_t + \varepsilon_{it} \quad (1)$$

The variables are over municipality and year, as i and t, respectively. The main variables firm total, and Internet access will be explained in section 4. The control variables are GDP per capita of the municipality, as it approximates average income of the unit of observation, and population, which should account for population effects. Since the number of firms is also related to people's income, production levels, infrastructure development and factors of consumption and production, these control variables are representative of these metrics.

Since the panel data covers many years and very different municipalities the models control for fixed effects. Using α_i to control for time invariant, municipality specific effects and using ϕ_t to control for municipality invariant, time specific effects. The error term ϵ_{it} controls for time and municipality varying effects. It is assumed that the error term is not correlated to the independent variables to account for omitted variable bias. The important coefficient for this regression will be β_1, as the coefficient for Internet access on the total number of firms and it should be interpreted as percent change between the variables. Meaning that a one percent increase in Internet access corresponds to a β_1 percent change in the number of firms. The hypothesis would be confirmed by a positive β_1 as it is expected that Internet access and the number of firms to be positively related.

There are two variations for the base model, one for number of firms opening and one for firms closing as the dependent variable. They indicate the rate at which Internet access impacts the number of firms entering or exiting a market. These models allow us to observe if the change in the total number of firms comes from Internet access changing the number of firms entering and exiting the market, or if it is some other unobserved reason.

$$\ln(firmOpen_{it}) = \beta_1 \ln(access_{it}) + \beta_2 ln(gdppc_{it}) + \beta_3 \ln(pop_{it}) + \alpha_i + \phi_t + \varepsilon_{it} \quad (2)$$

$$\ln(firmClose_{it}) = \beta_1 \ln(access_{it}) + \beta_2 ln(gdppc_{it}) + \beta_3 \ln(pop_{it}) + \alpha_i + \phi_t + \varepsilon_{it} \quad (3)$$

The interpretation for these models is the same as for the first model. However, with these models it is more relevant to compare the coefficients for Internet access between the two models, which can provide insights on how the market is adapting to the change in technology. The hypothesis assumes that both coefficients are positive. However, the coefficient for firms opening is greater than the coefficient for firms closing, as it is expected that the market will grow with technological improvements.

A unique contribution of this paper is not only being able to control for the quantity of access, but also the average quality of it. A variable that represents the average download speed was constructed to avoid omitted variable bias on the access variable. Not considering the quality of the service could lead to positive bias on the Internet access coefficient, since it expected that more access would also mean better quality service. This is due to the relatively lower cost of upgrading the Internet infrastructure after a municipality gains access. Additionally, higher quality service is expected to attract more customers, thus increasing access. Also, this variable controls for different average speeds between different observation groups, if the data were to be restricted. The model will be used with all three dependent variables for number of firms and be of the form:

$$\ln(firm_{it}) = \beta_1 \ln(access_{it}) + \beta_2 ln(gdppc_{it}) + \beta_3 \ln(pop_{it}) + \beta_4 \ln\,(A.speed_{it}) + \alpha_i + \phi_t + \varepsilon_{it} \quad (4)$$

Note that for this equation $firm_{it}$ corresponds to all three dependent variables, total number of firms (equation 4.1), number of firms opening (equation 4.2) and number of firms closing (equation 4.3). Then β_4 should show the effect that the average quality of the Internet has on the coefficient for number of firms, with a negative value showing an inverse relation, a positive value a direct relation and a zero-value showing no relation. Also, if comparing two groups or between the dependent variables, the magnitude of the coefficient indicates the sensitivity of the group to the quality of access. This is relevant, because even if a location has low quality access, there could be no demand for better access by that market. So, the coefficient is proportional to their preferences for better quality Internet.

Lastly a model to understand the growth rate and shrinking rate of the market shall be used which looks at the percent change on the percentile of which the market grew or shrank:

$$\ln\left(\frac{firm_{it}}{firmTotal_{it}}\right) = \beta_1 \ln(access_{it}) + \beta_2 ln(gdppc_{it}) + \beta_3 \ln(pop_{it}) + \beta_4 \ln\,(\text{A}.Speed) + \alpha_i + \phi_t + \varepsilon_{it} \quad (5)$$

It is important to note two things about this model. Firstly that $firm_{it}$ corresponds only to the number of firms opening and the number of firms closing. Secondly, only for this model, the firm total variable will not be leaped, thus the variable is constructed such as the ratio between number of firms that opened, or closed, by December of year t and the total number of firms that a municipality had in January of year t.

These models will be used on the constructed dataset to study the relation. It is pertinent to point out that these models assume that the fixed effects are indeed time or municipality invariant and that the error term is uncorrelated to the variables in the model. Also, the interpretations for β_1 imply correlation and not causation, so they can indicate some possible cause and effect, but do not prove that there is any. To better understand the results is important to first understand the main variables. Section 4 focuses on the data summary of variables used by the models.

4. Data

The panel data built for the purposes of this work is the aggregation of multiple publicly available federal datasets joined on the Brazilian Institute of Geography and Statistics (IBGE) unique municipal identifier and year. All variables used in this model come from distinct databases and had to be combined into the final dataset.

Table 1: Variable Description

Variable	Description
FirmTotal	The number of firms that are registered to a municipality at the beginning of the fiscal year by their registration to the CNPJ* registry
FirmOpen	The sum of number of firms that registered to a municipality by the end of the fiscal year by their registration to the CNPJ* registry
FirmClose	The sum of number of firms that were approved to close by a municipality by the end of the fiscal year by their registration to the CNPJ* registry
Access	The number of broadband Internet access endpoints registered to ANATEL, reported by Internet Service Providers
A. Speed	Average download speed in Mbps over many different technologies for the measured broadband Internet access endpoint by ANATEL
Gdppc	The GDP per Capita at market price of a municipality in reference to prices of 2010 as reported by IBGE in Reais
Pop	The number of people that live in a municipality. Measured by the Census of 2010 and estimated for future years as reported by IBGE

Note: * National Registry of Legal Entity (CNPJ). See appendix A for the data source of the dataset used for this table.

The data for the number of firms that closed and firms that opened came from the Ministry of Finance's Map of Firms. This data was collected monthly and for the purposes of this research aggregated up to each year. This means that these values correspond for the total at the end of the year. This was done for two reasons. Mainly because it was unlikely that monthly changes due to Internet access would be detectable or have much economic significance, as a percentage change of a dependent variable between two months would be very small. Additionally, for especially the smaller municipalities, there would be a lot of missing data which would make the panel data very unbalanced. It's important to make it explicit that for all observations these values are positive integers, meaning that every municipality had some fluctuations for each year. The data for the total number of firms comes from IBGE. It measures the number of firms operating in that municipality at the beginning of the fiscal year. Therefore, in the regressions it shall be used lagged by a year to match the time of the other variables.

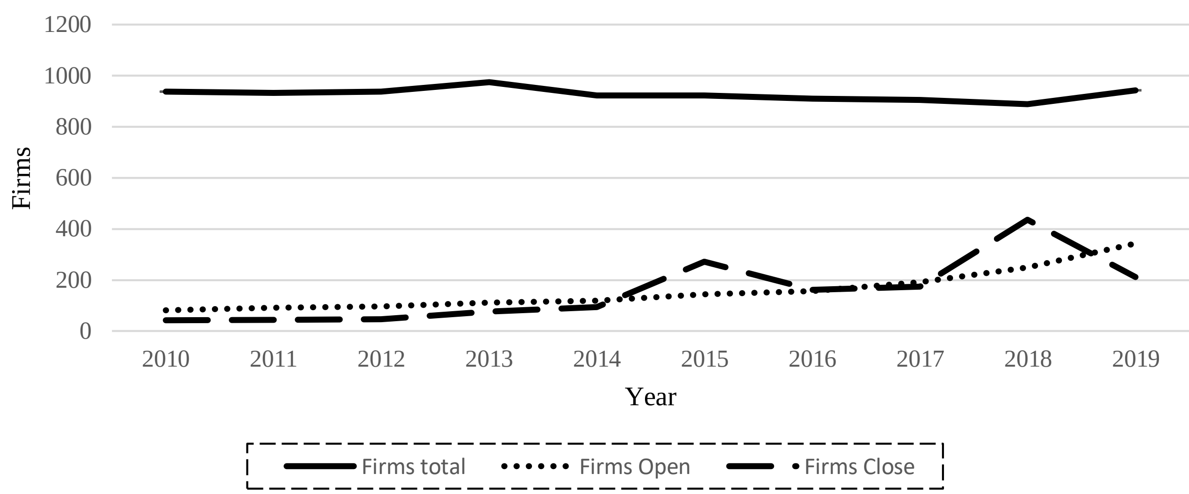
Table 2 below describes the summary statistics of the variables for the total number of firms, number of firms that opened and number firms that closed. Figure 1 shows how the mean of each variable changes through the years of the observation.

Table 2: Summary statistics of dependent variables for all years

	Mean	Std. Dev.	Median	Q_1	Q_3	Min	Max
FirmTotal	927.69	8628.90	166	71	457	3	588673
FirmOpen	158.15	1517.16	26	11	72	1	180740
FirmClose	156.14	1698.62	21	7	65	0	219425

Note: Std. Dev. Is the standard deviation of the variable. Q_1 is the upper bound of the first quartile. Q_3 is the lower bound of the third quartile. Min is the minimum value observed for the variable. Max is the maximum value observed for the variable. See appendix A for the data source of the dataset used for this table.

Figure 1: Average number of firms of municipalities over time



Note: See appendix A for data source of the dataset used for this figure

It is important to note that the averages of firms opening and closing are very similar. However, firms closing have some outliers that firms opening do not have, which is consistent with the trends on the total number of firms as can be seen in figure 1, above.

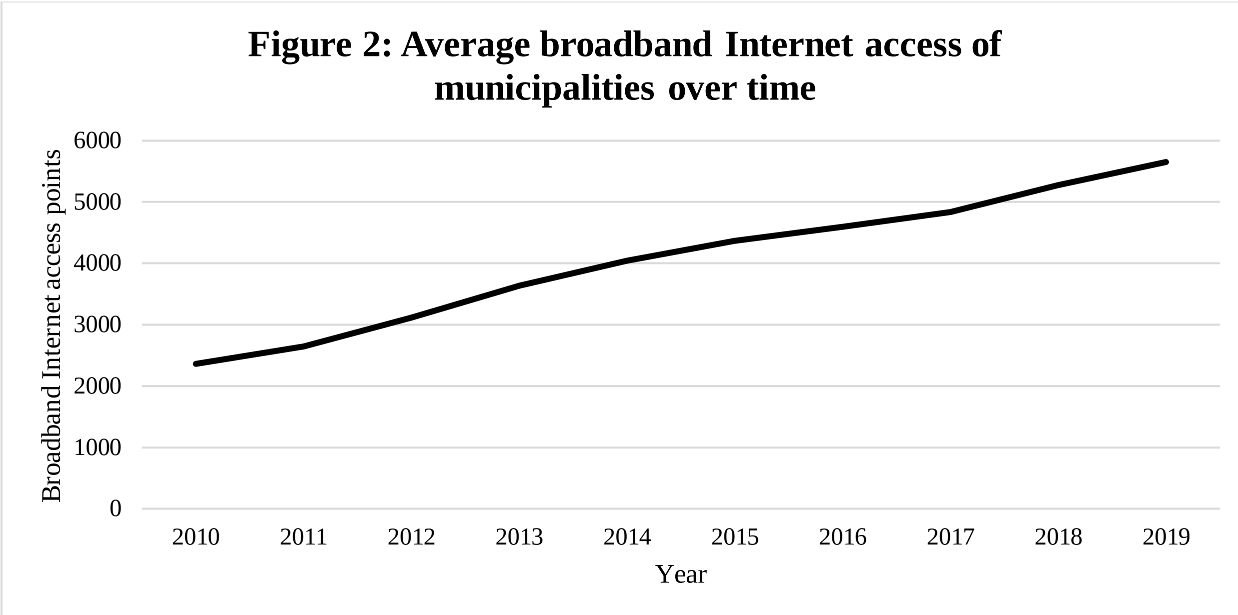
The data for broadband Internet access comes from the National Telecommunications Agency (ANATEL). It was also collected monthly and aggregated to the year level for the same reasons as explained above. This means that these values correspond for the total at the end of the year. This variable measures the number of broadband Internet access points registered to ANATEL. Mostly these are xDSL and cable modem connections for home consumption. This means that the data is a good enough metric for the purposes of approximating consumer access to the Internet. Also, it can be expected that most consumers will make purchases using Wi-fi or their home computers, meaning that the variable should be proportional with consumption.

Table 3 below describes the summary statistics of the variables for Internet access and average Internet speed. Figure 2 and figure 3 show how the mean of each variable changes through the years of the observation, respectively.

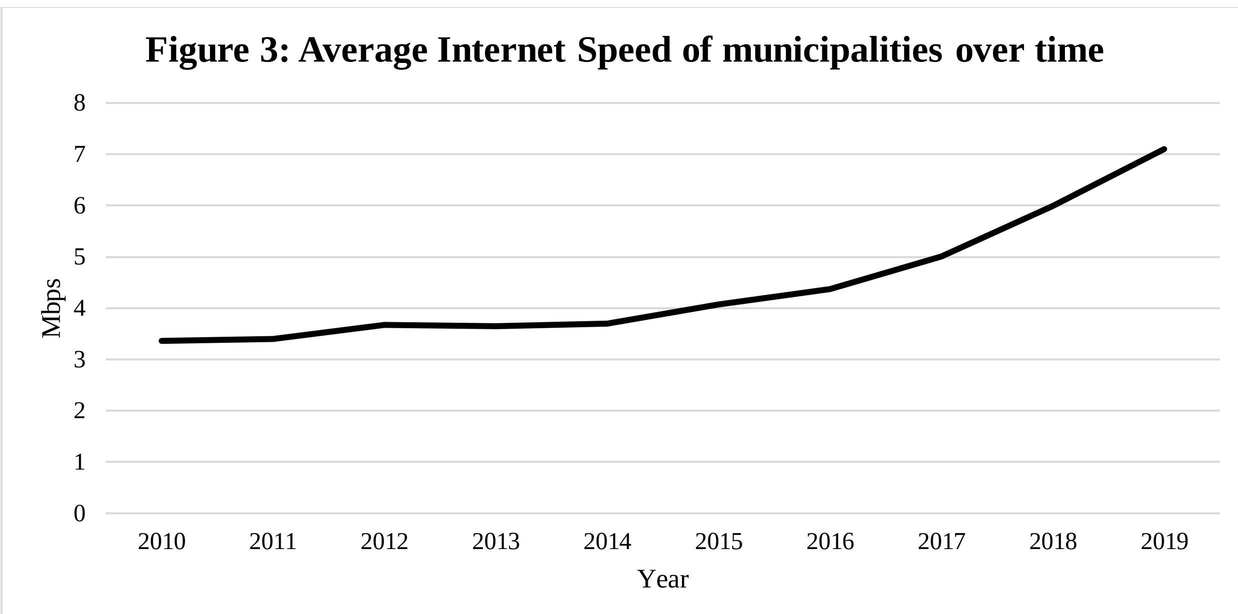
Table 3: Summary statistics of independent variables for all years

	Mean	Std. Dev	Median	Q_1	Q_3	Min	Max
Access	4058.95	49052.98	223	71	893	1	3438564
A. Speed	4.44	2.38	4.09	2.75	5.84	0.25	34

Note: Std. Dev. Is the standard deviation of the variable. Q_1 is the upper bound of the first quartile. Q_3 is the lower bound of the third quartile. Min is the minimum value observed for the variable. Max is the maximum value observed for the variable. See appendix A for the data source of the dataset used for this table.



Note: See appendix A for data source of the dataset used for this figure



Note: See appendix A for data source of the dataset used for this figure

As it can be observed broadband Internet access and average Internet speed both grow over time. It is interesting to observe that the curve for average Internet speed has a nonlinear slope, which is consistent with the rate of technological development for Internet speeds (Federal Communications Commission, 2013). Also, it agrees with the intuition that expanding and upgrading the Internet service grid is easier than first installing due to the relatively lower maintenance costs in relation to upfront fixed costs.

For the control variables, GDP per Capita and population come from the IBGE database, Sidra. Since in 2010 Brazil performed a Census, the data for that year comes from the nationwide survey. For the later years, data from the 2010 Census and other smaller surveys are used to estimate these variables. This estimation is done according to IBGE standards and procedures and are consistent across the data, as these standards are also used by all other Brazilian federal institutions.

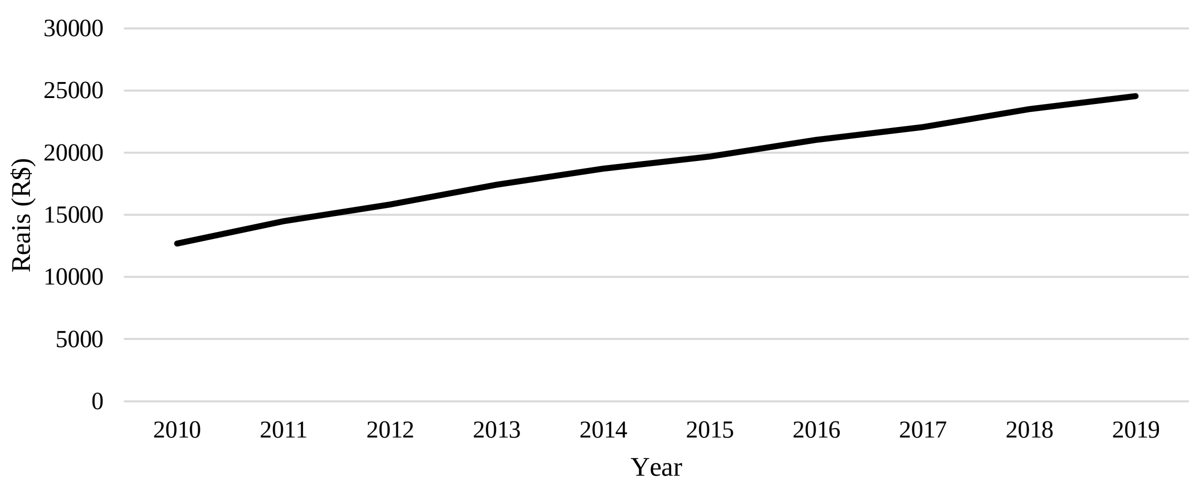
Table 4 below describes the summary statistics of the variables for GDP per capita and population. Figure 4 and figure 5 show how the mean of each variable changes through the years of the observation, respectively.

Table 4: Summary statistics of control variables for all years

	Mean	Std. Dev	Median	Q_1	Q_3	Min	Max
Gdppc	19011.21	21077.23	13647.05	8021.97	23243.1	301.6	815698
Pop	36708.38	215513.5	11546	5474	24942	781	$1.23E07$

Note: Std. Dev. Is the standard deviation of the variable. Q_1 is the upper bound of the first quartile. Q_3 is the lower bound of the third quartile. Min is the minimum value observed for the variable. Max is the maximum value observed for the variable. See appendix A for the data source of the dataset used for this table.

Figure 4: Average GDP per capita of municipalities over time



Note: See appendix A for data source of the dataset used for this figure

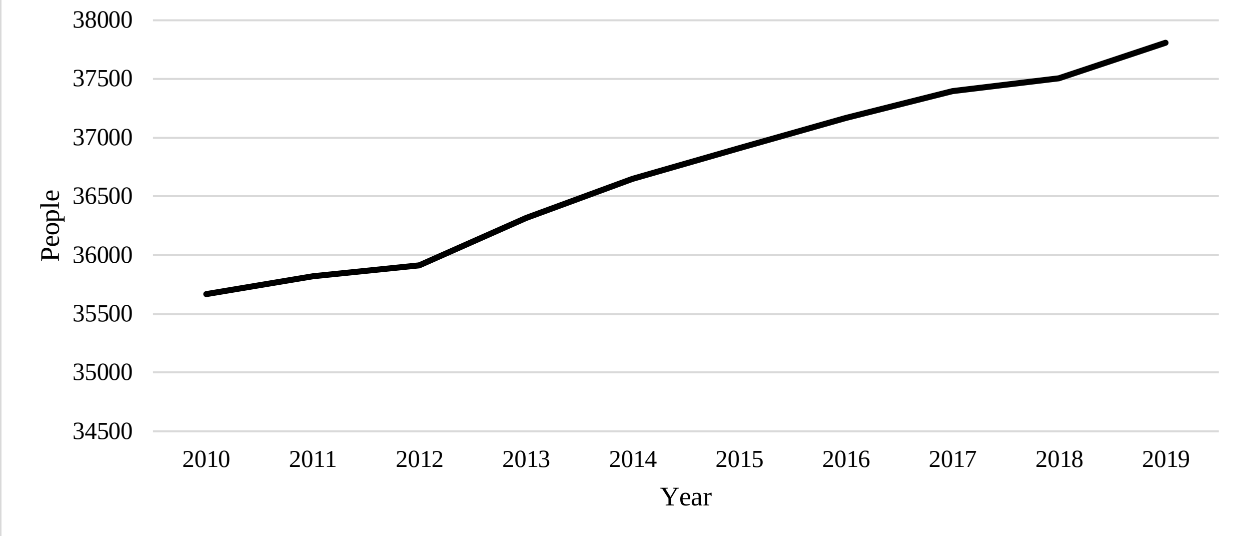


Figure 5: Average population of municipalities over time

Note: See appendix A for data source of the dataset used for this figure

Note that the averages for GDP per Capita and population increase over time as expected from the period observed. It is important to make it explicit that these metrics are year averages and thus not all municipalities experience this growth, and some might even show negative growth overtime for any of the variables presented.

It is important to understand that while some municipalities might have changed names or were divided up into smaller municipalities or were subject to specific political conditions, if their IBGE identification code is the same, they should not be impacted in the analysis. Also, some municipalities are missing a few years of data. All these conditions make the panel not perfectly balanced. Out of the 5556 municipalities in the dataset, 5354 (96.36%) have data for all years, 153 (2.76%) municipalities are missing just one year, with 65 (1.17%) of those missing 2010 specifically. Since most of the data is reasonably balanced, no further modifications to the dataset will be made for the general analysis.

Table 5 describes the summary statistics differentiating between the municipalities which have observations for all years as "full" and those who don't as "partial". From the table it is possible to observe that these municipalities have clearly lower averages then the rest of the sample. This suggests that they should be kept in the panel data to avoid outlier bias.

Table 5: Data summary of variables on full and partial municipalities

	Mean	Std. Dev.	Obs.	Median	Q_1	Q_3	Min	Max
FirmTotal								
Full	956.80	8767.36	53,530	175	76	477	4	588673
Partial	38.27	40.99	1,752	28	18	44	3	447
FirmOpen								
Full	163.12	1541.53	53,530	27	12	75	1	180740
Partial	6.23	8.31	1,752	4	2	8	1	185
FirmClose								
Full	161.06	1725.97	53,530	22	8	68	0	219425
Partial	5.60	9.03	1,752	3	1	7	0	165
Access								
Full	4189.85	49843.84	53,530	238	78	945	1	3438564
Partial	59.25	102.16	1,752	33	14	65	1	1524
A.Speed								
Full	4.44	2.37	53,530	4.09	2.75	5.85	0.25	34
Partial	4.33	2.65	1,752	4.06	2.48	5.65	0.25	20.5
Gdppc								
Full	19207.0	21213.65	53,530	13885.75	8120.75	23464.8	2256.42	815698
Partial	13029.05	15204.07	1,752	8618.37	6410.67	13819.1	301.6	203973
Pop								
Full	37761.35	218929.2	53,530	12014	5786	25637	781	$1.23E07$
Partial	4536.25	3501.64	1,752	3352	2451	5168	1034	21368

Note: Std. Dev. Is the standard deviation of the variable. Q_1 is the upper bound of the first quartile. Q_3 is the lower bound of the third quartile. Min is the minimum value observed for the variable. Max is the maximum value observed for the variable. See appendix A for the data source of the dataset used for this table.

In the analysis the dataset will be divided into rural and urban municipalities. This division uses the IBGE 2017 standard and considers to be rural (61.26 %) the following categories: remote rural (5.79 %), rural adjacent (54.42 %) and intermediate remote (1.08 %). It considers urban (38.74 %) the following categories: urban (26.33 %) and intermediate adjacent (12.39 %). In parenthesis is the percentage of municipalities that fall into each category. It is important to understand how the variables differ for rural and urban municipalities as they are also an approximate qualitative measure of distance from urban centers.

Dividing the sample into rural and urban municipalities makes it evident that in all metrics rural municipalities have lower average values. This is expected from how population and resources are distributed throughout the country (Wagner & Ward, 1980). With 85% of the population living in urban areas in 2015 (IBGE, 2015). Observe that there is an important distinction between urban areas and urban municipalities. The definition for urban area is a fiscal one that is intra-municipal, while urban municipalities is a generalization based on socio-political-geographical characteristics of each municipality. This will be important for interpreting results in the analysis section when using the models on the different groups as all municipalities have urban and rural areas. It is mostly that rural municipalities have a higher proportion of rural areas than urban municipalities. Table 6 provides the summary statistics of all variables for each group, rural and urban municipalities.

Table 6: Data summary of variables on rural and urban municipalities for all years

	Mean	Std. Dev.	Observations	Median	Q_1	Q_3	Min	Max
FirmTotal								
Urban	2166.01	13770.92	21,419	579	246	1391	13	588673
Rural	144.43	155.01	33,863	93	51	182	3	2268
FirmOpen								
Urban	371.29	2421.90	21,419	89	38	221	1	180740
Rural	23.33	28.01	33,863	1	7	29	1	709
FirmClose								
Urban	369.01	2715.07	21.419	75	26	609	0	219425
Rural	21.48	31.09	33,863	11	4	26	0	664
Access								
Urban	10103.14	78425.82	21,419	1227	358	4070	1	3438564
Rural	235.87	410.80	33,863	111	45	267	1	18072

A.Speed								
Urban	5.22	2.53	21,419	4.95	3.30	6.84	0.25	17.83
Rural	3.94	2.14	33,863	3.67	2.49	5.19	0.25	34
Gdppc								
Urban	23523.60	22726.25	21,419	22726.25	11131	28437.6	2486.75	428020
Rural	16157.03	19430.62	33,863	10949.90	7003.7	19368.1	301.6	815698
Pop								
Urban	78439.88	341918.5	21,419	26952	14465	58614	1487	1.23E07
Rural	10312.4	9008.59	33,863	7500	4171	13454	781	82094

Note: Std. Dev. Is the standard deviation of the variable. Q_1 is the upper bound of the first quartile. Q_3 is the lower bound of the third quartile. Min is the minimum value observed for the variable. Max is the maximum value observed for the variable. See appendix A for the data source of the dataset used for this table.

Now having a deeper understanding of the data and how the panel data was created, section 5 uses the models from section 3 to study the dataset. The next session also provides some initial interpretations that will be further explored in section 6. For further information on the technical aspect of creating the dataset, appendix A goes over it with greater detail.

5. Analysis

Running the first model on the effect of Internet access on the number of firms and controlling for fixed effects, produces the results seen in Table 1. As expected, Internet access has a positive effect on the number of firms that open in a municipality. Interestingly, at the same time having a positive effect on the number of firms closing in a municipality. The curious effect is on the total number of firms, as it has a negative value. The reasons for these values are what will guide this analysis.

Table 7: Log of independent variables on number of firms controlling for fixed effects

	(1)	(2)	(3)
ln(access)	-0.0354^{***} (0.0013)	0.1580^{***} (0.0030)	0.3424^{***} (0.0050)
ln(gdppc)	-0.0520^{***} (0.0035)	0.8240^{***} (0.0084)	1.5023^{***} (0.0134)
ln(pop)	0.0930^{***} (0.0228)	1.5750^{***} (0.0499)	4.1496^{***} (0.0787)
Nº Obs.	49,581	55,282	53,702
R^2	0.0195	0.8507	0.6880

Notes. Model for three distinct dependent variables, total number of firms, firms that were opened, firms that close. ***significant at the 0.001 level. **significant at the 0.05 level. *significant at the 0.1 level. Standard error in parenthesis. All regressions control for time and municipal fixed effects.

A technical interpretation of the coefficients for equation (1) is that for a 1% increase in Internet access the total number of firms is expected to decrease by 0.0354%, all else constant. This means that as Internet access grows a municipality is expected to have fewer firms operating in it. A similar interpretation can be done for GDP per capita, as for a 1% increase in GDP per capita represents a 0.052% decrease in the total number of firms in a given municipality, all else constant. Lastly, population shows a positive relation, so if population increases by 1% the total number of firms is also expected to increase by 0.093%, all else constant. This indicates that municipalities are more likely to have an increase in their markets from a growing population, while an increase in Internet access or wealth decreases these markets.

A possible interpretation for this is that an increase in population leads to an increase in demand as people will have to consume regardless of wealth or other factors. Which goods they are consuming or how they are consuming can differ, but bigger populations mean more consumers, which means more demand. Also, a bigger population means, on average, a bigger labor supply, which all attract firms. A similar interpretation can be used to explain the coefficient for GDP per capita. If we interpret this variable as a measure of wealth, an increase in wealth leads to an increase in demand. However, since most people are workers, an increase in wealth must mean an increase in wages, which would push firms out of the market. Also, more wealth could mean that people are buying better products or spending that extra money outside of the municipality, which would not affect local demand and explain the negative value of the coefficient. The interpretation for why the coefficient for Internet access is negative requires us to understand how this variable interacts with the firms that wish to enter the market and firms that wish to leave the market.

Observing equation 2, we see that a 1% increase in Internet access the number of firms opening is expected to increase by 0.1580%, all else constant. This means that as Internet access grows a municipality is expected to have more firms entering its market each year. A similar interpretation can be done for GDP per capita, as for a 1% increase in GDP per capita represents a 0.8240% increase in the number of new firms in a given municipality, all else constant. Lastly, population also shows a positive relation, so if population increases by 1% the number of firms opening is expected to increase by 1.575%, all else constant. This indicates that as these variables grow more and more firms are expected to enter the market each year.

Observing equation 3, a very similar reading can be done. As the variables increase by 1%, the number of firms that close increases by the coefficient as a percentage. This means that at the same time as Internet access grows, more firms enter the market, but also more firms leave.

A possible interpretation for the difference between the β_1 could mean two things, either Internet access forces more firms out of the market than allows for new ones, or it makes it easier for non-efficient firms to leave the market. It's important to recall that the variable for firms closing is recorded when the firm is approved by the government to close. Thus, this variable can be seen as proportional to the barrier to exit the market, with the variable being inversely proportional to the barrier to exit.

Since many of the procedures for closing a business can be done online since 2001, and even more were available with the REDESIM, BMS and BE programs, increasing Internet access reduces this barrier. It can be argued that Internet access also reduces the barrier of entry, since many of the procedures for opening a business can also be done online. While this explains the β_1 for firms opening, the results indicate that the effect on firms closing is more significant. This difference could explain numerically why the coefficient for Internet access on total number of firms is negative.

The data does not discriminate on firms or sectors; therefore, nothing can be said about which firms are entering or exiting the market. As perhaps, different industries, or sectors, observe different effects from Internet access. It is expected that the gaming industry, for example, will benefit from the Internet, while it is unclear how food establishments would be impacted. However, it could be that the reason why the rate of firms closing is higher than opening is because Internet access increases competition across all industries. This is consistent with the intuition that

Internet access allows for businesses to expand their customer base by reaching new consumers more cheaply and consistently.

Another interpretation for why the β_1 values for opening and closing firms are positive is the idea that Internet access allows consumers to reduce search costs. This can be understood as reducing the friction from shifts in supply and demand. As the market gets closer to having perfect information two things happen, one related to firms entering the market and another to firms exiting.

When consumers have better information on a market, they can now choose the product that best suits their needs with the reduction of search costs. This means that before they could be consuming in a non-optimal way and thus switch their purchasing habits. This means that non-efficient firms have a higher pressure to improve, or they will be pushed out of the market. This explains the positive coefficient between Internet access and firms closing. By reducing this inefficiency, consumers might have more purchasing power and thus incentivize new firms to enter the market, explaining the positive, but smaller, coefficient for firms opening.

Due to the symmetry of information provided by the Internet, the same way consumers have more information on producers, producers have more information on consumers. This allows for firms to better decide when to enter or exit the market, which might also contribute to the values observed.

Using this interpretation, the explanation for why the value for firms closing is bigger than that of opening is that markets with more Internet access are more efficient. Thus, there are more firms that are forced out of the market than those that are incentivized to enter. Additionally, it could be that due to the Internet, firms can operate in neighboring municipalities without having a location there. Thus, businesses can count for the number of firms in one area, while increasing the competition of neighboring areas and making more firms leave the market.

It is most likely that the results found are explained by a combination of both interpretations. As with lower barriers to entry, or exit, the market is more responsive to demands to attract or repel firms and therefore the higher observed coefficients. While the reduction of the friction of the supply and demand curves makes it so that there is a force for firms to enter or leave the market, which again increases the coefficients.

However, it could be that this behavior of more firms leaving then entering is not consistent throughout our data. As seen in Jensen's (2007) results, it is expected that areas with lower levels of access to show more significant results by increasing access, then ones with higher levels of access (Jensen, 2007). It is expected that such areas will mainly be rural areas as they show overall lower Internet access. This expectation is consistent with the data, as shown before.

The same model as before is used, but the dataset is divided into rural and urban municipalities. To observe the distinctions in the base model due to the different types of municipalities the groups are put side to side. The important row is still Internet access.

Table 8: Log of independent variables on number of firms by rural and urban municipalities

	(1) (Urban)	(1) (Rural)	(2) (Urban)	(2) (Rural)	(3) (Urban)	(3) (Rural)
ln(access)	-0.0443^{***}	-0.0321^{***}	0.1855^{***}	0.1449^{***}	0.4580^{***}	0.2955^{***}
	(0.0183)	(0.0017)	(0.0047)	(0.0039)	(0.0084)	(0.0060)
ln(gdppc)	-0.0360^{***}	-0.0605^{***}	0.6784^{***}	0.8355^{***}	1.2569^{***}	1.5494^{***}
	(0.0044)	(0.0050)	(0.0114)	(0.0115)	(0.0203)	(0.0177)
ln(pop)	0.1300^{***}	0.0761^{**}	3.7555^{***}	0.5030^{***}	6.4301^{***}	2.7458^{***}
	(0.0303)	(0.0321)	(0.0730)	(0.0677)	(0.1299)	(0.1047)
Nº Obs.	19,274	30,307	21,419	33,863	21,326	32,376
R^2	0.2114	0.0044	0.8419	0.5296	0.6726	0.4468

Notes. Model for three distinct dependent variables, total number of firms, firms that were opened, firms that close. ***significant at the 0.001 level. **significant at the 0.05 level. *significant at the 0.1 level. Standard error in parenthesis. All regressions control for time and municipal fixed effects.

The results of this regression show that the effect of Internet access on the total number of firms is proportionally the same for rural and urban municipalities. The interpretations for all the results are the same as before, as a 1% increase in the variable means an increase of the dependent variable by the coefficient as a percentage. The interesting result from this model is that urban areas are more sensitive to improvements in Internet access than rural areas.

An interpretation for why rural areas have lower coefficients could be that they also have slower and more unreliable Internet connection. This would explain lower coefficient between urban and rural for the number of firms opening and closing as the effect of access in rural municipalities is still positive but less than urban areas, which are expected to have better Internet connection. This can be assessed by using a variable that represents the average Internet speed for a given unit observation. Then by running equation 4 on the divided dataset should control for differences in the quality of the Internet.

Table 9: Log of independent variables controlling for average Internet speed on number of firms by rural and urban municipalities

	(4.1) (Urban)	(4.1) (Rural)	(4.2) (Urban)	(4.2) (Rural)	(4.3) (Urban)	(4.3) (Rural)
ln(access)	-0.0406^{***} (0.0019)	-0.0312^{***} (0.0017)	0.1290^{***} (0.0045)	0.1174^{***} (0.0039)	0.4181^{***} (0.0086)	0.2676^{***} (0.0061)
ln(A.Speed)	-0.0274^{***} (0.0026)	-0.0078^{**} (0.0026)	0.3203^{***} (0.0062)	0.1815^{***} (0.0058)	0.2259^{***} (0.0116)	0.1775^{***} (0.0091)
ln(gdppc)	-0.0277^{***} (0.0045)	-0.0567^{***} (0.0052)	0.5563^{***} (0.0109)	0.7330^{***} (0.0118)	1.1716^{***} (0.0206)	1.4440^{***} (0.0184)
ln(pop)	0.1882^{***} (0.0308)	0.0801^{**} (0.0321)	3.0061^{***} (0.0699)	0.4436^{***} (0.0667)	5.8978^{***} (0.1315)	2.6925^{***} (0.1040)
Nº Obs.	19,274	30,307	21,419	33,863	21,326	32,376
R^2	0.4580	0.0010	0.8508	0.5460	0.6779	0.4445

Notes. Model 4 for three distinct dependent variables, total number of firms (4.1), firms that were opened (4.2), firms that close (4.3). ***significant at the 0.001 level. **significant at the 0.05 level. *significant at the 0.1 level. Standard error in parenthesis. All regressions control for time and municipal fixed effects.

From the results in table 9, the interpretations from table 8 still hold. However, these results disprove the assumption that the reason why rural areas are less affected by Internet access was the quality of the service available. This is due to the model now controlling for Internet speed as a proxy for quality and the comparison of the two groups indicating that rural municipalities are less sensitive to changes in quality than urban. Meaning that urban municipalities not only improve more than rural ones when access increases, but also when quality increases.

As shown by Kim & Orazem (2017) the effect of Internet access on the rates of new firms is dependent on proximity to urban centers due to agglomeration economies. Meaning that although Internet access increases the rate in which firms enter the market, the effects are higher in urban areas compared to rural areas. As firms now can reduce costs by being closer to their main services in urban centers and still provide for the demand of the region. This interpretation is consistent with the results observed by the model.

The last piece of this analysis is to observe if the number of firms in a municipality affects the rates for opening and closing as shown in equation 5. This model should control for the case that the size of the market has different effects on the rates of opening and closing.

Table 10: Log of independent variables on open or close ratio

	$\ln\left(\frac{\# open}{\# total}\right)$	$\ln\left(\frac{\# close}{\# total}\right)$	$\ln\left(\frac{\# open}{\# total}\right)$	$\ln\left(\frac{\# close}{\# total}\right)$
ln(access)	0.1906^{***} (0.0033)	0.3760^{***} (0.0051)	0.1491^{***} (0.0033)	0.3369^{***} (0.0052)
ln(gdppc)	0.8823^{***} (0.0092)	1.5619^{***} (0.0142)	0.7475^{***} (0.093)	1.4354^{***} (0.0146)
ln(pop)	1.4014^{***} (0.0546)	3.9617^{***} (0.0844)	1.1659^{***} (0.0533)	3.7505^{***} (0.0839)
ln(A. Speed)			0.2563^{***} (0.0048)	0.2341^{***} (0.0076)
Nº Obs.	55,282	53,702	55,282	53,702
R^2	0.0195	0.0113	0.0271	0.0126

Notes. Same model for two distinct dependent variables, open ratio, and close ratio. ***significant at the 0.001 level. **significant at the 0.05 level. *significant at the 0.1 level. Standard error in parenthesis. All regressions control for time and municipal fixed effects.

The results are consistent with previous models. Note that the coefficient for Internet access has a different meaning than previous models. Now β_1 indicates how the rate in which a market is growing or shrinking is affected by the rate of growth of Internet access. In other words if Internet access increases by 1%, then the rate in which the market grows or shrinks increases by β_1%, all else constant. These results show a higher sensitivity of the rate in which the markets shrink to Internet access when compared to that of growing. This reinforces the idea that although Internet access allows for new firms, overall, it reduces the number of firms operating for a given municipality as inefficient firms are pushed out and firms can now service larger areas.

This interpretation is consistent with the results found for the coefficient of average Internet speed, as the rate in which markets grow is more sensitive to the quality of the service. Meaning that for a market to grow not only more access is required, but also better-quality access.

6. Discussion

The results observed by the work done in this paper are strongly in line with classical economic theory. One of the assumptions made by economists about perfectly competitive markets is perfect information. This assumption says that if consumers and producers know everything about the market, by removing search costs the market can reach perfect equilibrium. Theory also points to in the long run, a perfect competitive market has no barrier to entry or exit. However, with these assumptions, theory does not make any statements about the overall number of firms in this market.

As shown in the models done by this paper, an increase in broadband Internet access results in an increase in the number of firms entering or leaving a given municipality. Which is consistent with the interpretation that Internet access increases the amount of information each individual has. In other words, it approaches perfect information. Then Internet access is understood as proportional to the amount of information in the market, with perfect information being the asymptotic behavior of Internet access. It must be noted that there is probably a limit to how much Internet access can reduce search costs. As the Internet can also overload individuals with information and thus increase search costs.

The results for the effect of Internet access on the total number of firms under this interpretation are also in accordance with theory. Since the perfect information assumption does not imply anything about the size of the market and the coefficients for Internet access on the total number of firms were comparatively small. Not only that, but also it is consistent with the results of firms closing and opening as the rate in which firms close is higher than open. This suggests that for some municipalities, as access grows, they probably lose local firms to new ones that are not local. As the rural comparison model suggests this is likely because of firms now having a bigger reach on where they can provide services and moving to urban centers to benefit from aggregation economies.

Further work that can be done from here comes from two areas in which this paper does not consider. Firstly, analyzing the data for different industries and seeing if these results differ across them. It would be expected that the service sector would feel a bigger impact than agribusiness or industrial sectors. However, it would be very useful to understand by how much, or even if at all. It is important to recall that Brazil has a substantial agribusiness sector, which recently has pushed harder for automation and digitalization of many procedures. Also, looking at the productivity of firms could better explain the difference between the rate of opening and closing. Perhaps, the impact of international firms that provide e-commerce to Brazilians should also be considered. Additionally in this topic, understanding the effect of e-banking and the PIX payment method and how these are related to broadband Internet access.

Secondly, broadband Internet access has limitations. Since it only counts for home consumption and many people would purchase or use services using mobile Internet access. Doing some of this analysis and expanding on it for mobile Internet could prove useful. Especially because remote areas of Brazil mostly have mobile Internet access of 3G or maybe 4G technologies. This research would be extremely useful together with the e-banking analysis, as many people access their banks through their phones.

Lastly, the results of this paper could further be improved by having more accurate controls for people's consumption budgets, urban development metrics and other population statistics. The data from the 2022 Brazilian census should prove particularly useful. Due to the COVID-19 pandemic the market was forced to digitalize more quickly, and this might affect the results found in this work. This could have caused a shift from the Internet being a luxury good, to a required good. People would not be able to perform many activities without some form of Internet connection.

In conclusion, broadband Internet access has a noticeable effect on the rate in which firms enter and leave the market, but not in the quantity of firms operating in it. This is consistent with the literature and classic economic theory. It is supported by the understanding that ICTs allow for better information and reduce search costs and information related inefficiencies. It also shows that rural municipalities benefit from better access, but it is important to further study how they are affected and if intervention is required to control for negative externalities.

Appendix A

The construction of the dataset has some design decisions that must be made explicit. Due to the size of the raw datasets from which each variable is constructed, MySQL version 8.0.30 was used. Since each variable comes from a unique dataset, the methodology used was to first create a table containing just the desired variable, the year and the IBGE municipal code. Using the year and municipal code as a composite key to join the tables together and output the final dataset that will then be exported to Stata version 17.0 for the analysis portion of this work.

Three of the variables required very little work to be added to the dataset. The firm total variable comes from IBGE and already was in such a way that was organized by year and municipal code. The same was true for GDP per capita as it also came from IBGE. For population, each year was organized by municipality, but the years were in distinct files, so they were appended to their table and easily constructed.

For broadband Internet access, the raw dataset provided by ANATEL's data panel was divided into six files of 2 years each and a unit of observation in each file was at a year, month, municipality, technology, speed and Internet service provider level. Thus, one municipality has multiple rows for any given month with different kinds of access. To aggregate to the month level the rows were grouped by year, month and municipality and then summed. Then to aggregate to the year level, the average of the available months was taken. This is because the month aggregated data was the total access for that municipality, but those numbers were not constant or constantly growing and some municipalities only had a few months of each year. It would be unreliable to only get data for one specific month and by taking the average, the trend over the years is still maintained and reduces noise that would appear in between months.

From the same dataset as broadband Internet access and following a similar construction, the variable for average Internet speed was constructed. The way it was recorded by the dataset was in groups of speed. Then each observation was made to correspond to the average speed of that group in Mbps and then when aggregating to the month, municipality level a weighted average based on the number of accesses for each speed was taken. Then again, a weighted average was taken when aggregating up to the year, municipality level. For example, if a municipality had for a given month 3 accesses that fell into the 0 Mbps to 0.5 Mbps group and 5 accesses that fell into the 2 Mbps to 12 Mbps group, the monthly average Internet speed would have been $\frac{3\times 0.25+5\times 7}{8} = 4.46875$ Mbps and 8 accesses for that month. Then using these averages, it would take another weighted average when grouping up to the year level.

For the number of opening firms and number of closing firms, the data was available through a Tableau endpoint hosted by the Ministry of Finance’s Map of Firms. However, the password for the endpoint was not public. As a work around, it was possible to download the raw binaries and then using the student client version of Tableau rebuild the databases. Then it was possible to extract the data, but the structure of the database was unknown. This procedure produced two tables. For firms open it gave monthly data on the number of firms that opened in that given month and the municipal code was available as well as the years required. For firms close the rows were the same as firms open, however the column for municipal code was missing. Which required using the municipality name and state to match the strings with the municipal codes. With the two tables, the last step was to aggregate them to the year level by summing the number of firms for each month.

With all the variables in their respective tables, it was the case of joining them on year and municipal code to generate the final dataset.

Appendix B

As noted on section 4, some of the values for the number of firms closed were 0. Therefore, when taking the natural logarithm of that variable, those values were excluded. Although it is not expected for it to skew the data too much, below are the results for running all the models on two constructed variables. One which is the number of firms close plus one and another which is the number of firms close plus a half. This adds those missing data points into the regressions and the results can be seen below.

Table 11: Log of independent variables on number of closed firms plus one

	ln(# firms close+1)	ln(# firms close+1) (Urban)	ln(# firms close+1) (Rural)	$\ln\left(\frac{\# close+1}{\# total}\right)$	$\ln\left(\frac{\# close+1}{\# total}\right)$
ln(access)	0.3126*** (0.0044)	0.3978*** (0.0082)	0.2474*** (0.0054)	0.3451*** (0.0047)	0.3110*** (0.0048)
ln(gdppc)	1.4208*** (0.0122)	1.1322*** (0.0198)	1.3540*** (0.0163)	1.4791*** (0.0131)	1.3680*** (0.0135)
ln(pop)	4.1146*** (0.0727)	5.8854*** (0.1264)	2.6750*** (0.0919)	3.9411*** (0.0777)	3.7471*** (0.0772)
ln(A. Speed)		0.2158*** (0.0111)	0.1535*** (0.0079)		0.2110*** (0.0069)
Nº Obs.	55,282	21,419	33,863	55,282	55,282
R^2	0.6927	0.6839	0.4512	0.0053	0.0061

Notes. Model for firms that close plus one. ***significant at the 0.001 level. **significant at the 0.05 level. *significant at the 0.1 level. Standard error in parenthesis. All regressions control for time and municipal fixed effects.

Table 12: Log of independent variables on number of closed firms plus a half

	ln(# firms close+.5)	ln(# firms close+.5) (Urban)	ln(# firms close+.5) (Rural)	$\ln\left(\frac{\# close + .5}{\# total}\right)$	$\ln\left(\frac{\# close + .5}{\# total}\right)$
ln(access)	0.3492^{***} (0.0047)	0.4178^{***} (0.0084)	0.2869^{***} (0.0060)	0.3818^{***} (0.0050)	0.3488^{***} (0.0048)
ln(gdppc)	1.5028^{***} (0.0133)	1.1544^{***} (0.0115)	1.4826^{***} (0.0182)	1.5610^{***} (0.0141)	1.4537^{***} (0.0145)
ln(pop)	4.1392^{***} (0.0790)	5.8820^{***} (0.1305)	2.8897^{***} (0.1030)	3.9656^{***} (0.0837)	3.7782^{***} (0.0834)
ln(Avg. Speed)		0.2183^{***} (0.0084)	0.1434^{***} (0.0089)		0.2040^{***} (0.0075)
Nº Obs.	55,282	21,419	33,863	55,282	55,282
R^2	0.6809	0.6800	0.4404	0.0151	0.0163

Notes. Model for firms that close plus one half. ***significant at the 0.001 level. **significant at the 0.05 level. *significant at the 0.1 level. Standard error in parenthesis. All regressions control for time and municipal fixed effects.

As the results show, the coefficients don't significantly change when considering the missing values due to taking the natural logarithm. The interpretations and conclusions presented by this research still hold for the non-altered variable.

Literature Cited

Aker, Jenny C., and Isaac M. Mbiti. 2010. "Mobile Phones and Economic Development in Africa." Journal of Economic Perspectives, 24 (3): 207-32. DOI: 10.1257/jep.24.3.207

Barbosa, Evaldo. 2016. "Determinants of Small Business Survival: The Case of Very Small Enterprises of the Traditional Manufacturing Sectors in Brazil." Available at SSRN: https://ssrn.com/abstract=2802706 or http://dx.doi.org/10.2139/ssrn.2802706

Borenstein, Severin, and Garth Saloner. 2001. "Economics and Electronic Commerce." The Journal of Economic Perspectives 15 (1): 3–12. http://www.jstor.org/stable/2696535.

Brown, Jeffrey R., and Austan Goolsbee. 2002. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." Journal of Political Economy 110 (3): 481–507. https://doi.org/10.1086/339714.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2012. "The Internet and Local Wages: A Puzzle." American Economic Review, 102 (1): 556-75. DOI: 10.1257/aer.102.1.556

Jensen, Robert. 2007. "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector." The Quarterly Journal of Economics 122 (3): 879–924. http://www.jstor.org/stable/25098864.

Kim, Younjun, and Peter F. Orazem. 2017. "Broadband Internet and New Firm Location Decisions in Rural Areas." American Journal of Agricultural Economics 99 (1): 285–302. http://www.jstor.org/stable/44132477.

Manjón-Antolín, M.C., and JM. Arauzo-Carod. 2008. "Firm Survival: Methods and Evidence." Empirica 35 (1): 1–24. https://doi.org/10.1007/s10663-007-9048-x.

Phillips, Bruce D., and Bruce A. Kirchhoff. 1989. “Formation, Growth and Survival; Small Firm Dynamics in the U.S. Economy.” Small Business Economics 1 (1): 65–74. http://www.jstor.org/stable/40228494.

Paulo Bastos Tigre. 2003. "Brazil in the Age of Electronic Commerce." The Information Society 19 (1): 33-43. DOI: 10.1080/01972240309476.

Silva, T. C., Coelho, F. C., Ehrl, P., and Tabak, B. M. 2020. "Internet Access in Recessionary Periods: The Case of Brazil." Phys. A Stat. Mech. Appl. 537:122777. Doi: 10.1016/j.physa.2019.122777.

Wagner, F. E., and John O. Ward. 1980. “Urbanization and Migration in Brazil.” The American Journal of Economics and Sociology 39 (3): 249–59. http://www.jstor.org/stable/3486104.