Politecnico di Torino

Master Degree in Engineering and Management

Hedonic Price Model on real estate: influence of Covid-19 virus on house prices

Academic Year 2020/2021
December 2021

Supervisor:
Luigi Benfratello

Candidate:
Rossella Ragusa

Summary

Index of figures and tables

Introduction

The housing sector is very widely associated with the economic wealth of a nation. This study focuses on the consumption aspect of the housing market and therefore it is crucial to outline and understand the attributes and characteristics contributing to the value of a unit. Literature has shown that the market price of a housing unit can be determined by the buyer's evaluations of the housing unit's bundle of inherent attributes, such as locational, structural or neighbourhood attributes. The most adopted method scholars have been using to model real estate market price is the hedonic price model.

Nowadays, since the Covid-19 outbreak the general economic context suffered a dramatic reduction in consumption, resulting in a further drop in prices and a decrease in workers' income per capita and the real estate market will hardly move independently. On 9th March 2020, Italian Government imposed a lockdown all over the Country and people started to experience homeworking at a worldwide scale. I believe that this would positively affects the marginal utility of a good internet connection available at the house.

Surprisingly, it appears there is a lack in literature on this matter, with only a few studies carried out many years ago in US and UK, and with even less taking into account the infrastructure for mobile connectivity.

Leveraging a dataset provided by Italian Revenue Agency of 7.499 transactions that took place in the province of Turin (Italy) between 2018 and 2021, the purpose of this thesis is to show that both broadband connection and the quality of mobile connectivity provides a price premium that consumers are willing to pay when buying a house, even before the pandemic. Furthermore, I will show how the role played by internet connection gets more important in the last few years, mitigating the negative effect of the Covid-19 on prices.

The thesis is structured in the following way:

1. In Chapter 1 I started from the housing sector analysis followed by a rigorous review of the literature available on hedonic model from a general point of view down to domestic sector findings.
2. In Chapter 2 I developed the main hypothesis meant to be investigated by the thesis itself as well as the model used to perform such examinations.
3. Chapter 3 is split into two main subparts: the first one concern the sample composition as well as high level analysis of the observation collected. The second one is dedicated to a detailed description of the dependent, explanatory and control variables included in the model and their operationalization.
4. In Chapter 4 descriptive statistics over the sample was performed, followed by the specification of the process used for the identification of potential outliers. Subsequentially, empirical results were examined in deep, supported by sensitivity analysis and robustness checks. Finally, the chapter was completed by introducing objective remarks on outcomes that could give birth to future further study.

CHAPTER 1

Literature review

1.1 The real estate market

1.1.1 Real estate market definition

The housing sector is very widely associated with the economic wealth of a nation, considering it usually triggers growth in other economic sectors. From an economic theory point of view, it is agreeable that buying a house represents an investment and consumption decision at the same time, as buyers may purchase a house both for their own personal consumption but also as assets to generate income. Therefore, to study the impact of variables to a property price becomes of extreme importance [1]. Boone, L. and N. Girouard have shown, indeed, how the housing market has a significant influence on the overall consumer behaviour and consumption choices [2].

This study focuses on the consumption aspect of the housing market and therefore it is crucial to outline and understand the attributes and characteristics that contribute to the value of a good in the real estate market.

One main characteristic which makes a difference when comparing it to other goods is that the housing market manifests durability, heterogeneity, and spatial fixity [1], that is indeed a unique feature. Literature has shown that the market price of a housing unit can be determined by the buyers' evaluations of the housing unit's bundle of inherent attributes, such as locational, structural, or neighbourhood attributes [3]. Many studies concerned the impact of the most classic attributes (such as the school quality and proximity, studied by

Jud and Watts in Charlotte, North Carolina [4]). Researchers have gone even further, studying the effect of a particular attribute in a specific region (e.g. Li *et al* studied the effect of landfill proximity in Hong Kong [5], Ibeas *et al* demonstrated a reduction of 1.1% in prices for each additional minute in travelling time to the Central Business District in Santander [6], whilst Bottero *et al* valued buildings energy efficiency in 2018 in Piedmont region of Italy [7]).

To model the underlying differentiation effectively, countless scholars have adopted the hedonic price model.

1.1.2 Main attributes influencing real estate prices

As stated in chapter 1.1.1., an effective method for deriving a house market price is to treat it as a bundle and study its attributes.

A more economic explanation for it lays in the notion of implicit market, that denotes the process of exchange and consumption as well as production of commodities that are almost exclusively traded in bundles, and whose transactions are thereby unobservable [8]. The related explicit market, on the other hand, with observed prices and transactions, is for the bundles themselves, and it might as well be thought of as constituting several implicit markets for the components themselves. This is of particular importance when the bundles are not homogeneous (like formerly stated, a house is an heterogenous good) but vary in some way depending on the quantity of different components they contain [8].

Ultimately, the existence of product differentiation implies that a wide variety of alternative packages are available and, therefore, transactions in products are equivalent to *tied sales* when thought of as bundles of characteristics [9].

In this sense, it is of extreme importance to explore and select all possible attributes that may have an influence.

At a high level, they can be grouped in four main categories (even though this grouping is just merely representative), listed below with some examples [1]:

- *Locational*:
    - *Fixed locational*: measured in some form of accessibility (travelling time, cost of travel, convenience), it has been shown that consumers are willing to pay a premium for higher accessibility to public transportation [10].
    - *Relative locational*: evaluated by surrogate measures, it has been shown a correlation between price and relative locational factors. For instance, consumers are willing to pay more for ocean view and/or waterfront [11], whilst avoiding landfill sites and cemetery views.
- *Structural*:
    - *Number of rooms*: positive effect, buyers pay more for more space.
    - *Floor area*: obviously positively correlated.
    - *Age*: negative effect because the age of a home means less security and therefore higher maintenance and repair costs [12]. Though, it is not rare to find positive coefficients on age in some submarkets. Clapp and Giaccotto developed a rational expectations framework for interpreting the coefficient on age in a standard hedonic model, stating that when a negative sign emerges, it is reasonable to expect that depreciation is the dominant factor being measured. On the other hand, when we observe a positive effect, then a "vintage" effect dominates the obsolescence in that particular submarket and point in time, due to the customer preferences [12].
- *Neighbourhood services*:
    - *Local governments and municipal services*: especially for families, school and hospital proximity and quality have a positive impact [4].

- *Neighbourhood quality*:
    - *Socio-economic variables*: like neighbour average social class [13].
    - *Other externalities*: think about traffic noise, neighbour pollution, crime rate, and other disturbances [14].

The table below summarizes it all.

Attribute		**Effect on house price**
Locational	Distance from Central Business District	-
	View of hills/valley/sea/ocean/lake/rivers	+
	Obstructed view	-
	Length of land lease	+
Structural	Number of rooms	+
	Floor area	+
	Basement/garage	+
	Floor level	+
	Structural quality (e.g., design. Materials and fixtures)	+
	Facilities (e.g., swimming pool, gymnasium)	+
	Age of the building	*varies from place to place*
	Building services (e.g., lift, air conditional system)	+
Neighbourhood services	Proximity to shopping center	*varies from place to place*
	Proximity to schools	+
	Proximity to hospitals	*varies from place to place*
	Proximity to places of worship	+
Neighbourhood quality	Crime rate	-
	Traffic noise	-
	Income of residents	+
	Environmental quality (e.g., landscape, garden)	+

Figure 1 – List of Commonly Used Housing Attributes in Hedonic Price Models

1.1.3 The impact of COVID-19 pandemic on real estate market

COVID-19 is an ongoing pandemic of coronavirus disease 2019, caused by severe acute respiratory syndrome coronavirus 2. This pandemic (also called "SARS-CoV-2") is surely causing a dramatic reduction in consumption, resulting in a further drop in prices and a decrease in workers' per capita income like a domino effect. In turn, there is also a decrease in employment rates, which will further depress consumption.

As stated, real estate market is tightly connected to the overall economy of a country, hence, at least in the short and mid-term, it will not tend to move independently from the general economic context [15]. In addition, it's not arguable that these negative effects will be long lasting because companies will require several months to recover from the losses suffered and customers will require much time in order to adapt to the new economic environment.

Moreover, the coronavirus disease possesses unique features with respect to other pandemics because such an event is happening in a globalised and interconnected world, like it never was during previous ones [16]. The scale of reduction in final wealth generation has yet to be assessed.

With regards to the real estate market, the effect of health emergencies in general is poorly explored in international literature. Some studies though were being conducted; for instance, in 2008 Wong has shown the effect of the SARS epidemic in Hong Kong and founded a 1.6% general decrease in property prices, highlighting that the absence of price overreaction is likely to be related to housing market characteristics [17].

On 9th March 2020, Italian government imposed a "lockdown" all over the country in the attempt to stop the spread of the virus. After two months of lockdown more or less, and after further two months of gradual recovery of production and work activities, the pandemic's economic effects begin to appear in all their intensity. Something that is undeniable and well documented is a drop in the number of real estate transactions, still the corresponding drop in prices that one may expect is not a certainty.

Homeowners experience a dramatic loss of real estate capital, a significant economic shock to the local and regional economy, negative health impacts, and in many cases, forced evacuation and relocation during periods of decontamination and disinfection [18]. A major consequence of the aforementioned restrictive measures is a change in the perception of the efficient use of space that may have an impact on the demand for existing and new real estate assets [19]. In addition, people started to experience homeworking at a worldwide scale [20]. This inevitably led to a different conception of the house spaces, with lots of families suffering for the lack of space necessary for working from home. Especially the households expecting to work from home even for a few days per week started looking for property outside downtown, having incentives in buying bigger houses at lower prices per square meter.

As a matter of fact, Liu and Su have shown how housing demand shifted away from high population density areas [21]. The decreased demand for density is mostly driven by the dwindling need for living close to work. Another factor is the diminished attraction of local amenities, gym for instance. The following graphs extracted from Liu and Su paper in 2021 pictures the diminishing rates over time for gym and other point of interest overall.

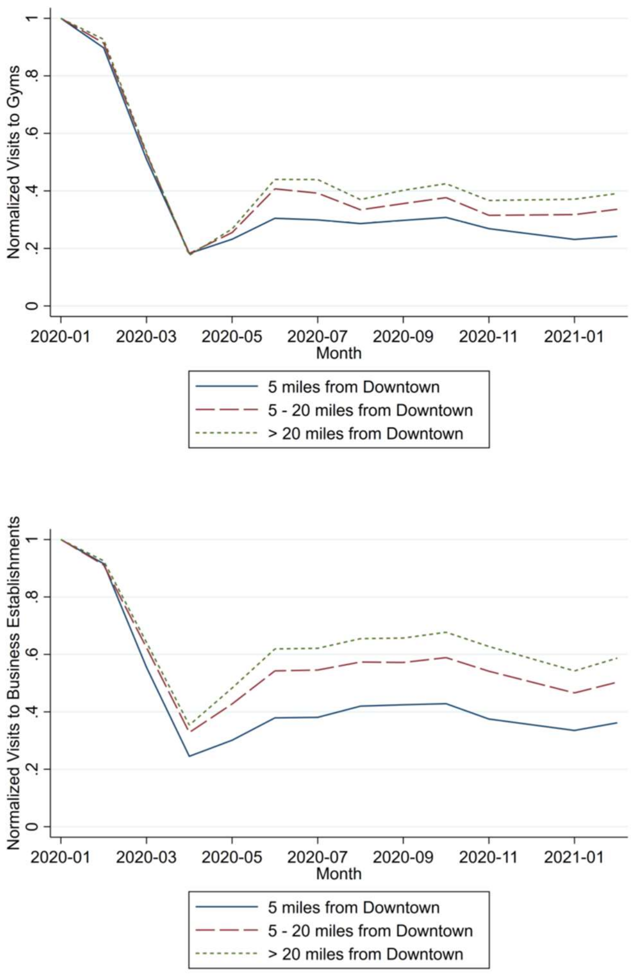

Figure 2 – Visitor traffic to Point-of-Interest by Type and distance to downtown [21]

In Italy, for instance, the housing market data collected by the Real Estate Market Observatory of the Italian Revenue Agency ("*Agenzia delle Entrate"*) shows, for the year 2019, as regards the entire regional territory, a decrease in the average property prices for main cities (0.6%) and stationary data for the rest of the provinces [22].

1.1.4 The impact of internet broadband on real estate market

Nowadays, broadband technology has revolutionized the way services are delivered and how people go about their everyday lives. High-speed Internet, or broadband, provides users with highest quality internet services, such as videoconferencing, and performing a variety of everyday tasks from the comfort of their own homes. Worldwide, policy makers discuss many proposals that would increase the deployment of faster household Internet speed, being aware of its increasing demand.

For instance, the current Italian strategy "*Verso la Gigabit Society*", approved on 25th May 2021 by the *Comitato Internministeriale per la Transizione Digitale* (CITD), tempts to achieve the coverage of the entire national territory with 1 Gbit/s connectivity by 2026 end of year [23].

Numerous academics tried to study the household benefits from broadband as an influencing factor on the value of residential real estate.

Conley *et al.* investigated in 2020 the impact of broadband availability on rural households by carrying out a hedonic price model in two US rural counties, finding out no measurable premium exists for broadband access across these counties. Nevertheless, they highlighted that a premium may exist in other rural areas depending on individual county settings as well as a lack of inclusion of mobile broadband connections that could play a role in the lack of impact seen for traditional wireline access [24].

Conversely, Molnar *et al.* showed that homes in neighbourhoods with access to relatively higher speeds will have higher transaction prices. However, it looks like rural residents are unwilling to pay a larger premium than urban residents for faster Internet access [25]. Nevertheless, they used data from 2011 to 2014 and even though it may be reasonable to think that people choosing to live in rural areas are less interested in staying connected to the world, one may expect that this finding is no longer valid in more recent times, especially after the spread of Covid-19. The latter expectation might be justified by the changes that society is experiencing in the way of living (labour and homeworking, education and e-learning, shopping and e-commerce and so on).

1.2 Hedonic price model: a review

1.2.1 Theoretic definition and main assumptions

The term "hedonic" derives from the Greek word *hedonikos {εδονικοσ}* ("pleasure"), and this model posits that goods are typically sold as a bundle of inherent attributes and their price is a summation of all marginal implicit prices [9]. Originally introduced by the pioneer Court in 1939 [26], and further developed by Rosen and Griliches soon after, it was widely adopted throughout the XX and XXI centuries for capturing quality changes using multiple regression, being defined by Hulten (2003) as the "[…] *most intellectually satisfying of the various quality-adjustment methods because it appeals to an underlying economic structure rather than opportunistic proxies*" [27].

Hedonic price models have spread to study quality changes of inherent attributes in goods such as automobiles [28] and other multicomponent products, like smartphones [29]. In this sense, real estate market is no different.

Recall what has been said about the way to analyse critically the attributes constituting the final worth of a house, seen as a bundle of the inherent constituents. For such a reason, a wide number of scholars have chosen the hedonic price model to deal with it.

Hedonic price models allow researchers to estimate the individual effect of each housing factor, *ceteris paribus* [1], as market price is seen as a function of each tangible and intangible building characteristic and other outside influencing factors [30]. As stated, hedonic analysis is performed to uncover the value of a particular attribute and such analysis is successful to the extent that it isolates the magnitude of this suspected cause and effect relationship. To the extent that undesired variation in houses value can be controlled, a more accurate measure of this relation is possible [31].

However, when one is to adopt a model, there are main assumptions that must hold true for exploiting its benefit with statistical significance, and the hedonic model is no exception. The main assumptions are:

- *Perfect competition*: this underpinning assumption is justified considering there are many buyers and many sellers, and practically nobody can actually influence the price, given that a single unit constitutes a quite negligible portion of the market [1].
- *Perfect information*: although it is conceivable that perfect knowledge is unachievable in reality, in real estate market it seems at least quite reasonable, as consumers and sellers can acquire most of the relevant information from brokers, newspapers (media in general) and, above all, real estate agents [1].
- *No market segmentation*: this assumption can be considered satisfied if and only if the geographical definition is properly defined. A too broad one is surely an improper aggregation, whilst a too narrow one may lead to imprecise and biased estimates [1]. One must therefore deal with it depending on the geographical location of empirical data. This work has its foundations in the Piedmont region, and the way in which it faces the market segmentation problem is widely explained in chapter 1.3.
- *Market equilibrium*: although to have a price vector adjusting instantaneously to changes in either supply or demand is unachievable in reality, real estate market is among the most stable markets existing, with theoretical demand and supply functions that only rarely change significantly over time [9].
- *Homogenous good*: to say housing market is homogeneous is a very strong assumption. As a matter of fact, in chapter 1.1.1 the exact opposite was asserted, that is properties are heterogenous goods. What can be thought as homogenous are the inherent attributes instead. Given that technology is not much different in property joint production, hedonic prices of housing characteristics vary with the quantity of characteristics.
- *No interrelationship between attributes*: this assumption cannot be met, generally speaking. One can only address the problem by putting so much effort in the attribute

selection and paying a lot of attention to their correlation, something that is a must anyway when dealing with statistical experimentation [32].

In the light of the previous assumptions, literature agrees now that hedonic price model is not as generally applicable as it was thought, rather you need to use a lot of caution in applying such a model [33].

1.2.2 Choice of functional form and variables in hedonic models

When deriving the model of a hedonic price function, two main aspects are to be considered: the list of explanatory variables and the choice of the functional form.

Regarding functional form, researchers have found little basis for choosing one form over the other. Since the early adoptions of the model, scholars have chosen the functional form simply empirically. In short, though you can find theoretical support for a functional form that incorporates at least some interactions among the various characteristics of housing, empirical experience suggests that most of the feasible approximations to the correct form are close substitutes [34]. Taking Rose, for instance, he does not define *a priori* a functional form for the relationship between good/commodity and its attributes, he rather uses the goodness-of-fit criterion.

The functional form of the hedonic regression equation can either be linear, semi-log, or log-log form. Most common is the semi-logarithmic form which has the advantage that the coefficient estimates are proportions of the price that are directly attributable to the respective characteristic [35]. On the other hand, the main advantage of the log-log form is that the regression equation estimates elasticities with respect to each characteristic. In addition, using logs of the dependent variable you also incorporate the prices non-negativity, remembering to pay attention to the normality assumption.

In practice, in real estate sector the linear model is often preferred, given the obvious advantage of directly estimating monetary values of housing characteristics [36].

With respect to the attribute's selection, Richard Butler well expressed how a misspecification is practically unavoidable, still significant estimates can be produced nonetheless:

"*Since price depends on its attributes, hedonic model generally suffers a lot in misspecification of variables (over-specification when including something irrelevant, under-specification when not including something worthy). However, to some extent (large or small as it may be), all hedonic models are misspecified, so it is just sufficient to include a small number of key variables, choosing among the ones that are costly to produce and yield the highest utility* […]" [34]

More importantly, Butler works suggests that the biases deriving from independent variable exclusion is small in practice. Therefore, biased as they may be, estimates are generally not invalidated by the inevitable misspecification of independent variable set.

It logically follows that the list of independent variables should be limited to housing characteristics, to be chosen among the four aforementioned groups (locational, structural and neighbourhood services and quality) or a mathematical transformation thereof. Including, for example, demander characteristics such as income is a clear misspecification and the same stands for buyer's [37].

Furthermore, the intrinsic clustering of characteristic combinations into a relatively low number of configurations leads to multicollinearity in estimates when including a generous number of relevant variables, meaning that any estimate of the hedonic relationship must omit some of the relevant variables.

To conclude, urban economic theory suggests that the function should include some land values that depend on location, to make sure to properly incorporate the spatial aspect.

1.3 Hedonic price method in Italian market

1.3.1 Relevant findings in Italian real estate market

Hedonic prices methods have been widely adopted by Italian researchers to study the effect of housing characteristics in the Italian real estate market. Different scholars have studied the price premium of energy efficient buildings through hedonic models, like Bottero *et al* [7], already mentioned in chapter 1.1.1, and Bisello *et al* in 2019 [38], as well as the urban improvement impact on housing values by Rosato *et al* in Venice [39].

The most important features emerging from the researcher's methodologies can be summarized in:

1. Papers whose area of study is limited to a region, city province or even a single city, are preferred and more reliable. Expanding to a larger geographical zone is sometimes accepted only when trying to overcome the little dynamism in terms of number of trades characterizing the Italian market.
2. The effect of location has a major influence (this is valid everywhere) on housing price. Intuitively, along with the square meters, it is the only housing characteristic that is never going to change over time (remember the spatial fixity) [40]. Hence, one must evaluate it carefully since location cannot be change after the deal is concluded and it will always remain so.

With point 1 above, I aim to justify my choice of restricting my analysis to a very specific region.

Regarding point 2, scholars have adopted very sophisticated statistical and econometric models, the so-called "spatial models". However, if a country is properly segmented and divided in zones so that each one may well correspond to a sub-market, then a way simpler adaptation of the model can be considered: the use of binary variables to identify the various real estate sub-markets that exist in a given territory [40]. Indeed, Bourassa *et al* show that the gain in terms of correctness and accuracy that is derived from the inclusion in a standard hedonic model of binary variables that refer to the various sub-markets is not inferior to the gain which is derived from the use of spatial methods [41].

Luckily, in Italy there are the OMI zones.

1.3.2 The OMI zones in Italy

The OMI of the Italian Revenue Agency ("*Agenzia delle Entrate"*), standing for *Osservatorio del Mercato Immobiliare*, is a database that allows to consult data on the real estate quotations and leases throughout the national territory.

The entire country is segmented in the so-called *OMI zones*, defining homogeneous sectors of the local real estate market of a particular city, in which there is a substantial uniformity of appreciation for economic and socio-environmental conditions.

The key process of the OMI is the definition of the OMI zones, which are meant to be homogeneous market areas.

It segments the market based on positional characteristics, assuming that they prevail clearly in the definition of real estate values. This criterion mainly derives from the scarce availability of complete information on all the property characteristics.

The main groups are:

- Central (code B): identifies that portion of the municipal territory that coincides with the urban center of the municipality itself.
- Semi-Central (code C): identifies that portion of the municipal territory that is in a position immediately adjacent to the urban center.
- Outskirt (code D): identifies that portion of the municipal territory which is contiguous to the central or semi-central sections.
- Suburban (code E): identifies that portion of the municipal territory that contains the urbanized areas that are separated from the urban agglomeration of the municipality by an undeveloped area, by a natural or artificial barrier.

- Extra urban (code R): identifies the remaining part of the municipal territory not included in the previous sections and it is delimited by the administrative boundary of the municipality; it is therefore a residual section.

Each section is then divided into numbered homogeneous zones that present uniformity in qualitative characteristics. Below the map of Turin as example.

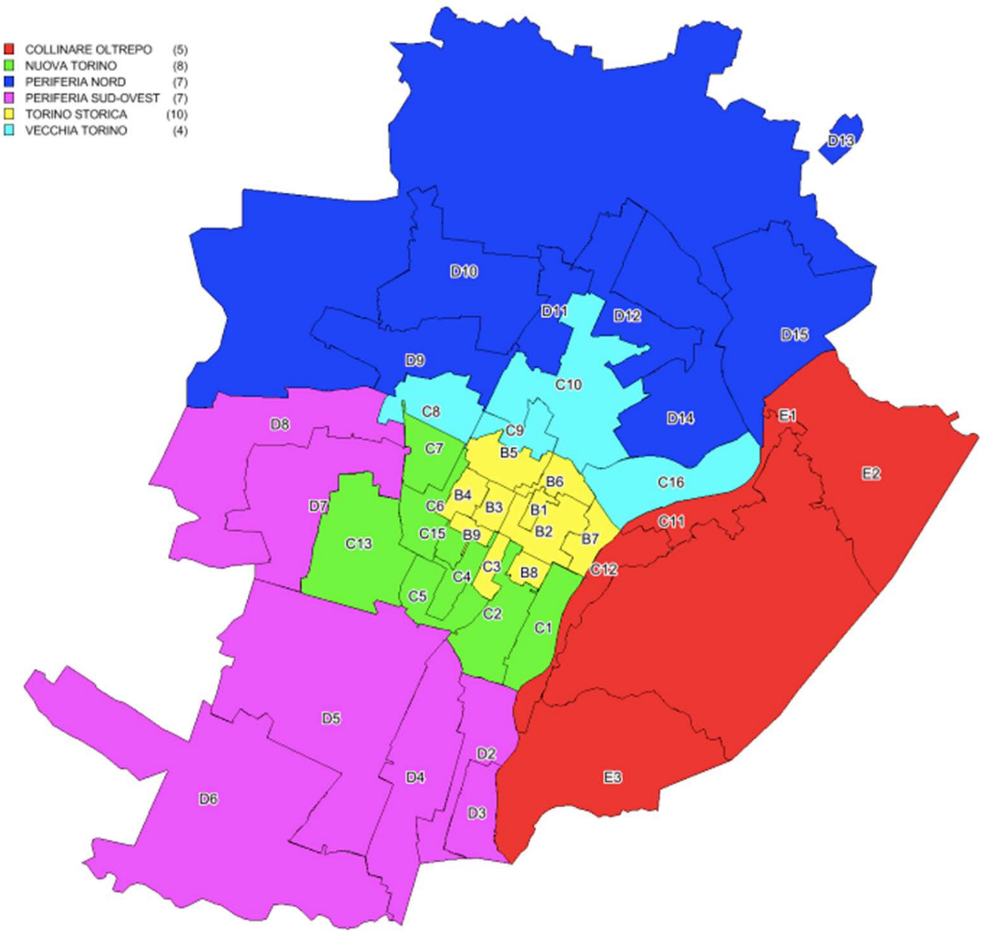

Figure 3 – OMI zones in Turin

The OMI zone, therefore, identifies a real estate submarket, and it can distinguish between different locations in the sense that two similar dwellings, with respect to characteristics, could have very different prices due to the fact that they belong to two different OMI zones. Consequently, considering binary variables, each one referring to a specific OMI zones of the geographical region of interest, it grasps the effect that the location of the property in a particular real estate submarket (that is, an OMI zone) has on its overall house price, simply by including such binary variables in a standard hedonic model, as if they were further housing characteristics [40].

If an OMI zone identifies a particular sub-market, then by aggregating more OMI zones, an extended real estate market is built, with higher dynamism (a considerable increase in the number of observations) and, therefore, a net increase in the degrees of freedom of the regression model, with a consequent increase in the economic reliability and statistical significance.

An important attention point, however: obviously, the dummy variables that identify more than one category must be interpreted with respect to the category chosen as a reference. It is therefore necessary to subjectively define the reference OMI zone (it does not affect the results at all) and, once the regression model is estimated, the interpretation of the results obtained for an OMI zone, must be viewed with respect to the reference zone. Lisi *et al* explained this point very well in 2020, saying that when interpreting the result of one of these binary variables, you must answer the following question: "*How does the price change when considering a home location in an OMI zone that is different from the reference zone?*" [40]. Of course, the difference in the model accuracy and significance when including such kind of variables strongly relies in the quality of the clusters offered by the OMI zones.

CHAPTER 2

Hypothesis and related models

2.1 Hypothesis 1

The thesis investigates the impact of the internet connectivity on a real estate unit. As discussed in Chapter 1, there is no evidence in Italy about the effect of a good internet infrastructure on house prices. In addition, there is even less evidence about such an effect when it comes to the mobile connectivity.

My first hypothesis **H1** is:

"*The price of a real estate unit is positively influenced by both the quality of the broadband technology coverage and the quality of the mobile connectivity, holding all the other attributes constant*"

If a scholar builds an index representing the internet quality, taking into account both homes broadband and mobile connectivity, while controlling for all the relevant attributes, then such an effect can be measured. The hedonic price model is perfect for doing such a job; therefore, I adopted the following model for testing **H1**:

$$Value = \beta_0 + \beta_1 Internet_{score} + \sum_{k=1}^{K} \beta_{S,k} S_k + \sum_{i=1}^{I} \beta_{NQ,i} NQ_i + \sum_{j=1}^{J} \beta_{NS,j} NS_j + \sum_{m=1}^{M} \beta_{L,m} L_m \quad \textbf{(2. 1)}$$

where S_k is the *k*-th structural attribute, NQ_i is the *i*-th neighbourhood quality attribute, NS_j is the *j*-th neighbourhood services attribute and L_m is the *m*-th locational attribute. In equation (2. 1), I expressed the control variables in different terms in the attempt of emphasizing the importance of including factors for each main attribute category, for a framework meant to model the real estate market would probably not be capable of isolating the effect of a single factor otherwise.

2.2 Hypothesis 2

The reviewed literature is suggesting that real estate market should have suffered a shock owing to the pandemic, same as the general economy. But not only general economy was impacted: people are changing their way of living. Given that the number of tasks being performed from home with a good connection is increasing and that the time spent at home is increasing as well, having a good internet connection should have even more relevance nowadays. For instance, a survey conducted in the city of Naples in 2021 states that, among the respondent *"*[…] *35.1% would like a better internet connection (an element that proved even more fundamental during isolation)* […]" [42].

My second hypothesis **H2** is therefore:

"*The price of a real estate unit suffered a negative shock due to the spread of Covid-19 disease, that is positively moderated by both the quality of the broadband technology coverage and the quality of the mobile connectivity, holding all the other attributes constant*"

To test **H2**, one should firstly prove **H1** and then use the same index for the internet quality along with its interaction with variable modelling the Covid-19 effect. The model for proving **H2** true is therefore the following:

$$Value = \beta_0 + \beta_1 Internet_{score} + \beta_2 Covid\ effect + \beta_3 Internet_{score} Covid\ effect +$$
$$+ \sum_{k=1}^{K} \beta_{S,k} S_k + \sum_{i=1}^{I} \beta_{NQ,i} NQ_i + \sum_{j=1}^{J} \beta_{NS,j} NS_j + \sum_{m=1}^{M} \beta_{L,m} L_m$$

(2. 2)

CHAPTER 3

Research sample

3.1 Data collection and statistics

3.1.1 Composition of the sample and sources

The research sample is mostly based on a dataset gently provided by Italian Revenue Agency of real estate transactions that took place between the beginning of 2018 and the first half of 2021 in the province of Turin [43], reporting the transaction value along with a wide set of attributes of the building, including locational attributes, neighbourhood's, and structural characteristics. Using the reference year and month of the transactions, it is possible to virtually split the dataset in transaction pre-Covid and post-Covid.

Moreover, I looked up for some sort of proxy for the quality of the internet connection. Using the open-source data available at AGCOM website ("*Agenzia per le Garanzie nelle Comunicazioni"*), I was able to find the internet coverage for each city in the province of Turin present in the aforementioned dataset [44].

Specifically, for deriving an internet quality score for each city I used the number of homes covered by ADSL, the number of homes covered by FTTC (*Fibre To The Cabinet*), and the number of homes covered by FTTH (*Fibre To The Home*), with respect to the total number of homes inspected PF1 (more on that on chapter 3.2.1). Moreover, in the attempt of including the mobile broadband connection, I collected data about average download/upload speed, number of providers and average latency for each city, using the OPENSIGNAL mobile application [45].

3.1.2 Sample statistics

The dataset is composed by 7.499 transactions, ranging from the 1st semester of 2018 to the 1st semester of 2021, across 45 cities (Turin included). The stack chart below shows the number of transactions per city pre- and post-Covid.

Overall, there are 6.021 transactions pre-Covid (taking July 2020 as watershed) and 1.478 post-Covid. The sample presents some missing values for the variables of interest and, hence, the number of transactions taken in consideration will be lower.

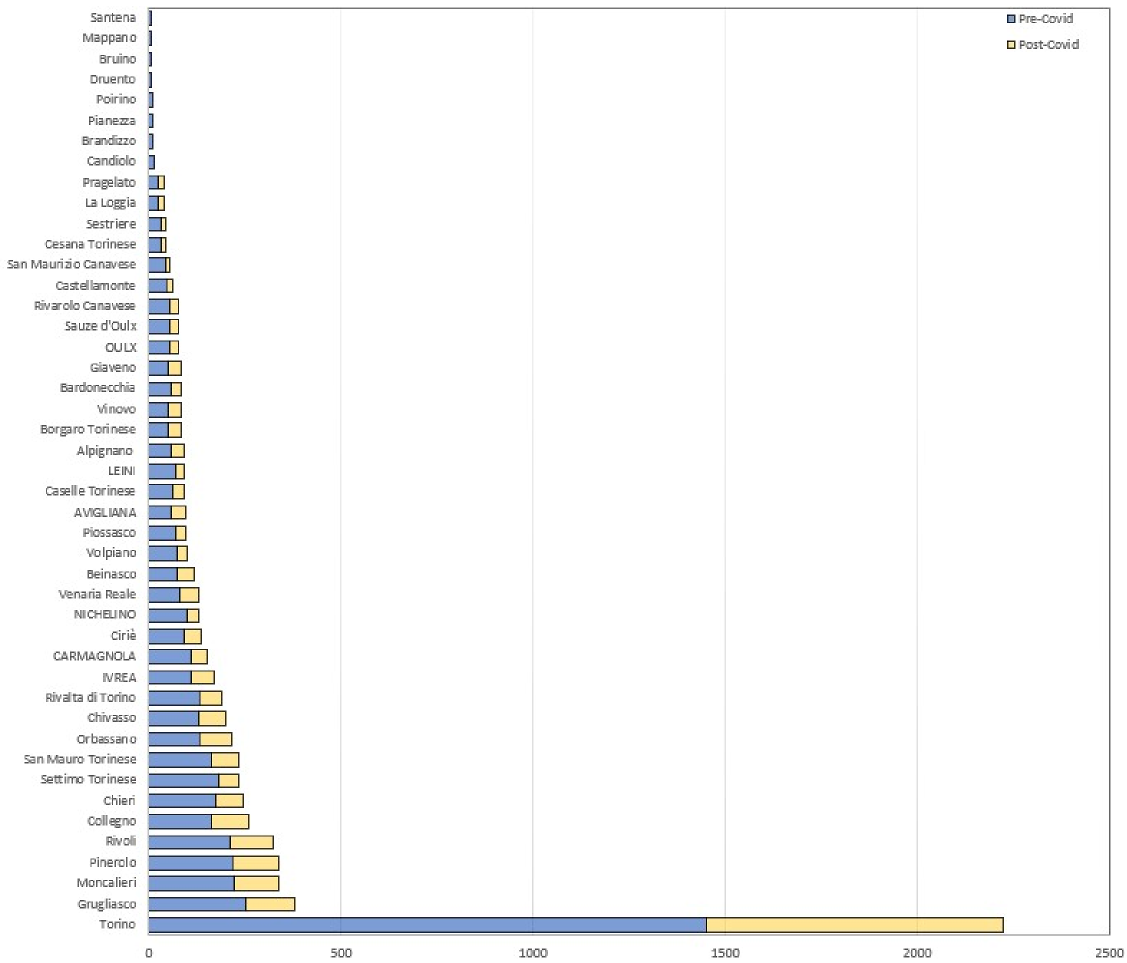

Figure 4 – Sample stat: Number of transactions per city

In the figure above, you can observe there are indeed cities with a dramatically low number of transactions, both pre- and (especially) post-Covid.

These cities (from Castellamonte upwards) have been excluded from the analysis, for they are not statistically significant.

With respect to the internet quality, the map below shows the cities coloured within a scale from light yellow to red, with increasing internet score (more on score computation on paragraph 3.2.1).

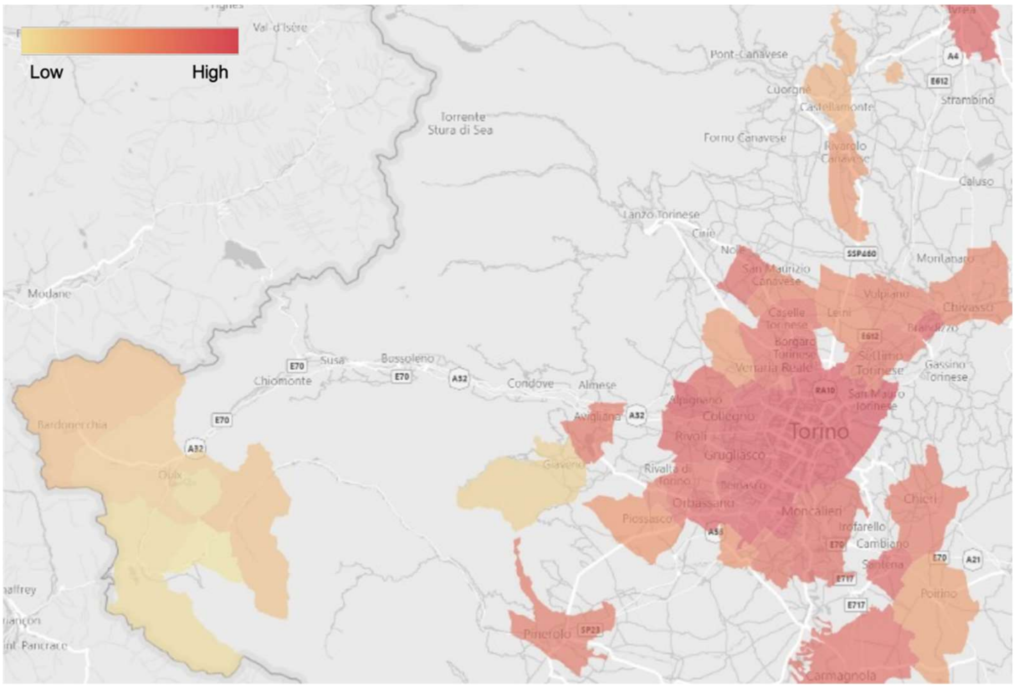

Figure 5 – Sample stat: Internet score map

With respect to the transaction value magnitude, the line chart below shows the average and the minimum value over time (by month). The vertical orange line (right on March 2020) is meant to divide pre- and post-Covid trends (still, remember I will use July as watershed).

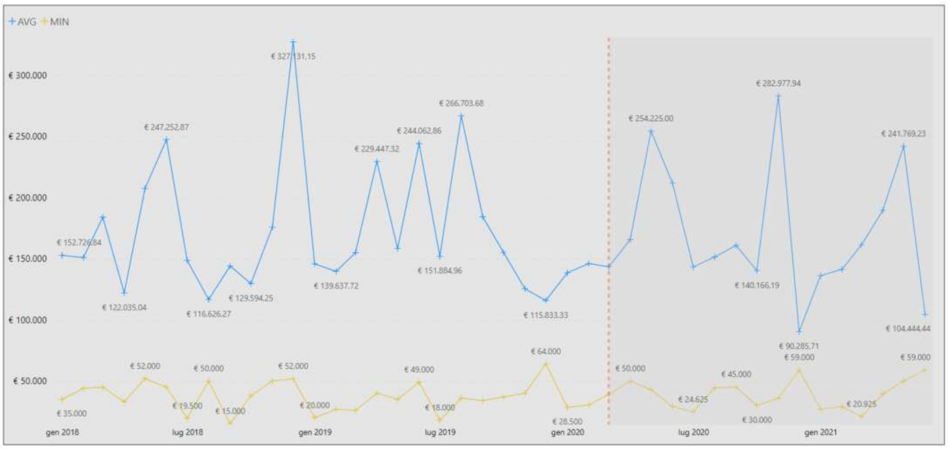

Figure 6 – Sample stat: Transaction value over time

As observable, there is no obvious downward trend. Data is telling us that, at least graphically, neither the average nor the minimum transaction values suffered a dramatic decrease after the spread of the Covid-19 disease.

In conclusion, experts from Italian Revenue Agency confirmed that the sample provided is complete with exclusion of only commercial, productive and services typologies. Hence, it is acceptable to assume that the sample for the real estate market segment of interest is random and complete.

In addition, they highlighted that each transaction refers to economic and/or notary agreements stipulated averagely a few months ago before the actual transaction date, thus confirming that selecting July 2020 as transition date for pre- and post-Covid is reasonable.

3.2 Variables description

3.2.1 Dependent and explanatory variables

As dependent variable I selected the transaction value, which represents the transaction price incurred when purchasing the property. Since the sample distribution was skewed and too far from normality, I transformed the value using the natural logarithm, and, as a matter of fact, it worked better statistically wise. Moreover, to adjust for inflation, I considered the yearly average Italian inflation indexes FOI[1] computed by ISTAT for the years 2021, 2020 and 2019 and considered 2018 as reference year for prices. In conclusion:

$$Value = \ln\left(Price_{adjusted}\right)$$

$$\text{where } Price_{adjusted} = \begin{cases} Price\ (1 + FOI_{2021})(1 + FOI_{2020})(1 + FOI_{2019}) & \text{if occurred in 2021} \\ Price\ (1 + FOI_{2020})(1 + FOI_{2019}) & \text{if occurred in 2020} \\ Price\ (1 + FOI_{2019}) & \text{if occurred in 2019} \\ Price & \text{otherwise} \end{cases}$$

(3. 1)

Empirically, it does not change a lot, but the results acquire more economic relevance.

In literature, there is a huge adoption of this variable since the underlying goal is to determine the correlation between building characteristics and the transaction price and, ultimately, to measure the influencing of those attributes on the overall transaction price.

As explanatory variables, I used *Internet score*, *Covid effect* and their interaction, called *Moderator*. *Internet score* is a time-independent index empirically built that should reflect the average goodness of the city broadband and mobile connectivity infrastructures, and for the city *c* it has been computed as follow:

$$Internet_{score,c} = \alpha\, Home_{score,c} + (1 - \alpha)\ Mob_{score,c}$$

$$\text{where } Home_{score,c} = \frac{ADSL_c + FTTC_c + FTTH_c}{PF1_c} \text{ and } Mob_{score,c} = \sum_{i=1}^{3} X_{i,c} \quad where \quad X_{i,c} = \frac{x_{i,c} - X_{min,i}}{X_{max,i} - X_{min,i}}$$

(3. 2)

[1] The FOI index is known by the economists as the Consumer Price Index, and it is based on a basket of goods and services representing the consumption of the average Italian family and it is widely adopted for prices adjustments, such as house rental prices.

where $ADSL_c, FTTC_c, FTTH_c$ are the number of homes covered by ASDL, Fiber To The Cabinet and Fiber To The Home technologies in the city c respectively and $PF1_c$ is the number of total homes inspected in the city c, whilst $X_{1,c}$, $X_{2,c}$ and $X_{3,c}$ are the normalized average download speed, upload speed and latency of the mobile connectivity of city c. Intuitively, both $Home_{score,c}$ and $Mob_{score,c}$ are continuous variables within the domain [0,3], and so it is their weighted average $Internet_{score,c}$. The weight α has been put equal to 0.5.

The *Covid effect* is meant to measure the effect (if any), *ceteris paribus*, on the house price as a shock suffered by the real estate market after the spread of the pandemic. For this sake, it was computed as follow:

$$Covid\ effect = \begin{cases} 1 & \text{if transaction occured after July 2020} \\ 0 & \text{otherwise} \end{cases} \tag{3. 3}$$

I have chosen July 2020 as watershed because it is reasonable to assume that the shock did not begin to manifest before a couple of months after the first lockdown imposed by the Italian government.

Regarding the interaction, it was simply computed as the *Internet score* times the *Covid effect*.

3.2.2 Control variables

Among the ones available in the dataset, I selected the following measures as control variables:

- Structural*: Surface[m^2]*, *Age*, *Lift_dummy*, *Bath_no* and *Maintenance*;
- Neighbourhood quality: *Neighbourhood*;
- Neighbourhood services: *Services*;
- Locational: *Transports* and *Distance* (from city centre).

Including the surface area is pretty much obvious considering the choice of the dependent variable (value instead of value by square meter). I used the natural logarithm transformation because it exhibits a distribution closer to the normality. In appendix (Figure 25 and Figure 26), density plots about both surface and its logarithm are presented. Moreover, I also tested whether the relationship between surface and value could be quadratic (assuming that the same percentage increase at a high level of surface values would imply a higher percentage increase on value compared to the same increase for low level of surface). This provided better results (along with logarithmic transformation), hence:

$$Surface = (\ln(Total\ surface))^2 \tag{3. 4}$$

I have chosen *Age* as a control variable since a correlation with property value has been shown practically in all hedonic work in literature. I calculated *Age* as the difference between the transaction and construction year (data available in the dataset), then I operationalized *Age* as its natural logarithm, because it resulted a little bit skewed and it provided better results (same as surface). Again, density plots are shown in appendix (Figure 27 and Figure 28).

In the dataset the number of lifts of the fabricate was available, still I have included it in the model as a dummy control variable, *Lift_dummy*, calculated as follow:

$$Lift\_dummy = \begin{cases} 1 & \text{if at least one lift is available for the fabricate} \\ 0 & \text{otherwise} \end{cases} \tag{3. 5}$$

Bathroom number *Bath_no* was included as control variable for the same reason as *Age*, and it is simply the number of bathrooms in the house. It is important to include also this variable as proxy of the building size since, even though it usually correlated with the surface, one may expect that it may have a positive impact on the consumer preference (holding constant the house surface).

The last structural attribute is *Maintenance,* and it is a qualitative score from 1 to 3, serving as a proxy of the overall health level of the building, where:

$$\begin{cases} 1 \rightarrow \text{poor} \\ 2 \rightarrow \text{normal} \\ 3 \rightarrow \text{optimal} \end{cases} \tag{3. 6}$$

The assessment of the maintenance level of the building considers the state of the ceilings, the internal fixtures, the walls and floors coatings, the electrical, heating, sewage and plumbing systems and the burglar alarm (if any). Quite intuitively, this has been treated as a categorical variable (and not as an integer).

The categorical variable *Neighbourhood* accounts for the neighbourhood quality of the house in terms of several factors such as crime rate and pollution. The score is again on scale of equation (3. 6) and it is not house-specific: instead, the Italian Revenue Agency calculated this index for each OMI zone and then extended the score to each transaction accordingly. If one believes in the sectorization done by the agency and considering that the neighbourhood quality is not the main core of this thesis, then this variable alone is enough as neighbourhood quality attribute.

Regarding the services, I used *Services* categorical variable available in the dataset, accounting for proximity to local amenities (schools, banks, pharmacies, post offices, hospitals…). The raw data was in a descending scale. For an easier interpretation, I re-arranged data on the following scale[2]:

$$\begin{cases} 1 \rightarrow \text{close} \\ 2 \rightarrow \text{afar} \\ 3 \rightarrow \text{absent} \end{cases} \tag{3. 7}$$

No further transformation was performed on the variable.

I used the categorical variable *Transports* to control for the locational attributes (fixed).

[2] Please note that the indication of close/afar/absent is measured according to the local reality. For instance, for an urban area with relatively high dimensions, a radius of 300-400 meters is considered as close.

For assigning the score, Italian Revenue Agency detected the presence of bus, tram, and subway stops nearby. In this case as well, the raw data related to the variable *Transports* was in a qualitative descending scale, and therefore I re-arranged data following the scale (3. 7) with no additional transformation on the variable.

Finally, using the OMI zone classification I computed the variable *Distance* which accounts for the distance from the city centre as follow:

$$Distance = \begin{cases} 1 & \text{if OMI Zone is not in } \mathrm{B}^* \text{ or } \mathrm{C}^* \\ 0 & \text{otherwise} \end{cases}$$

(3. 8)

that is a dummy variable activating for outskirts, suburban and extra-urban areas. This one falls too under the fixed locational attributes.

Unfortunately, other locational variables present in the dataset that may accounts for relative locational attributes such as view quality were nonetheless excluded from the model since they exhibited to many missing values and, therefore, they would have compromised the regression analysis.

CHAPTER 4

Analysis and results

4.1 Preliminary considerations

4.1.1 Descriptive statistics

The correlations of the variables included in the models are presented in the correlation matrix in appendix (Figure 29). The only relevant correlation is the one between the number of bathrooms and surface of the real estate unit. With a correlation of 66%, which was quite expectable, it is still worth to include both as they could capture size-related consumer preferences in a different fashion. No other correlations reached warning levels. As for other descriptive statistics, you can find below the frequencies of the involved dummy variables (Figure 7, Figure 8 and Figure 9) and a table with means and standard deviations for the continuous and for categorical ones (Figure 10 and Figure 11).

```
Lift_dummy

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/7,242

            tabulation:  Freq.  Value
                         2,514  0
                         4,728  1
```

Figure 7 – Lift availability dummy variable frequencies

```
Distance

                  type:  numeric (float)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/7,242

            tabulation:  Freq.  Value
                         4,138  0
                         3,104  1
```

Figure 8 – Distance from city center dummy variable frequencies

As observable, dataset is well distributed between transactions for homes in the city centers (and semi-central) and transactions for homes in suburban and extra-urban areas.

Unfortunately, the number of post-covid transactions is quite less than the pre-covid transactions. Notwithstanding, 1.448 observations are enough for providing statistically relevant results.

```
Covid_effect

                  type:  numeric (float)

                 range:  [0,1]                       units:  1
         unique values:  2                       missing .:  0/7,242

            tabulation:  Freq.  Value
                         5,794  0
                         1,448  1
```

Figure 9 – Covid dummy variable frequencies

The statistics for the *Internet Score* show a mean close to the maximum value observed. Still there is a relevant variability and therefore the index is not compromised.

The average *Surface* is quite reasonable if you think of the average home size in Turin and its province. The same stands for the *Age* and *Bath_no*. Nonetheless, their maximum values suggest there is room for outliers identification.

Variable	Obs	Mean	Std. Dev.	Min	Max
Internet_s~e	7,493	2.000466	.4501673	.3849701	2.364344
Surface_m2	7,494	102.6797	48.97979	19	625
Age	7,369	48.19758	34.26618	1	982
Bath_no	7,035	1.343994	.5777032	1	5

Figure 10 – Descriptive statistics for continuous variables

Categorical variables descriptive statistics are shown below.

Variable	Obs	Mean	Std. Dev.	Min	Max
Maintenance	7,242	2.01657	.1735164	1	3
Neighbourh~d	7,242	2.021403	.19155	1	3
Services	6,051	1.337961	.492568	1	3
Transports	6,108	2.853635	.3662415	1	3

Figure 11 – Descriptive statistics for categorical variables

The dependent variable *Value* shows a reasonable mean with a relatively large standard deviation. The Figure 12 also presents the percentiles, providing a clear picture of the sample distribution. In this case as well descriptive statistics suggest that an outlier identification process should be executed.

Transaction value (Euros)

	Percentiles	Smallest		
1%	39000	15000		
5%	54000	18000		
10%	64000	19500	Obs	7,494
25%	85000	20000	Sum of Wgt.	7,494
50%	120000		Mean	157524.2
		Largest	Std. Dev.	142262.9
75%	173000	1470000		
90%	274000	1500000	Variance	2.02e+10
95%	398000	1730000	Skewness	4.255581
99%	800000	2150000	Kurtosis	30.0539

Figure 12 – Descriptive statistics of Value

4.1.2 Outliers identification process

Before performing the regression analysis, I have looked for potential outliers, as it is well known in literature that OLS is dramatically sensible to outliers. Starting from the insights provided by descriptive statistics, I used Box Plot instrument to get a detailed view for *Surface* and *Age*. Instead of analysing *Value* independently, I found it more reasonable to analyse the joint Value by square meters.

Box plots for *Surface* and *Age* confirmed the presence of actual outliers. To avoid compromising the randomness of the sample, I cautiously removed only observations presenting at the same time a surface greater than 350 m^2 and aging over 100 years. Residential units falling in this range can be thought as "big historical buildings" and they could therefore be considered as belonging to a different market segment.

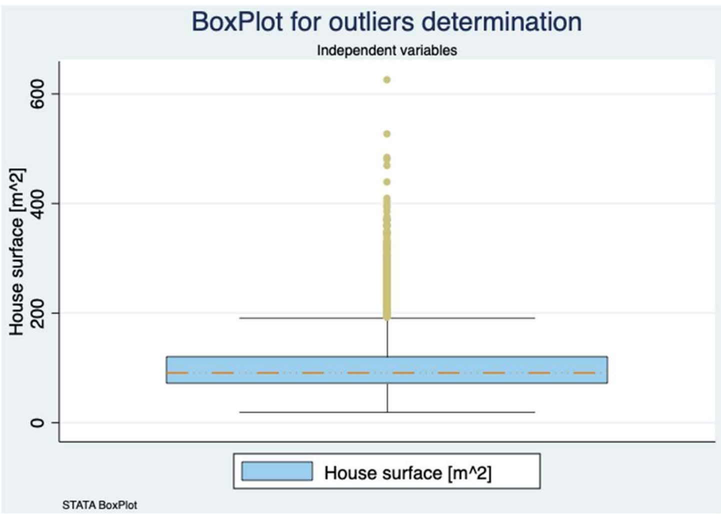

Figure 13 - Box Plot for Surface

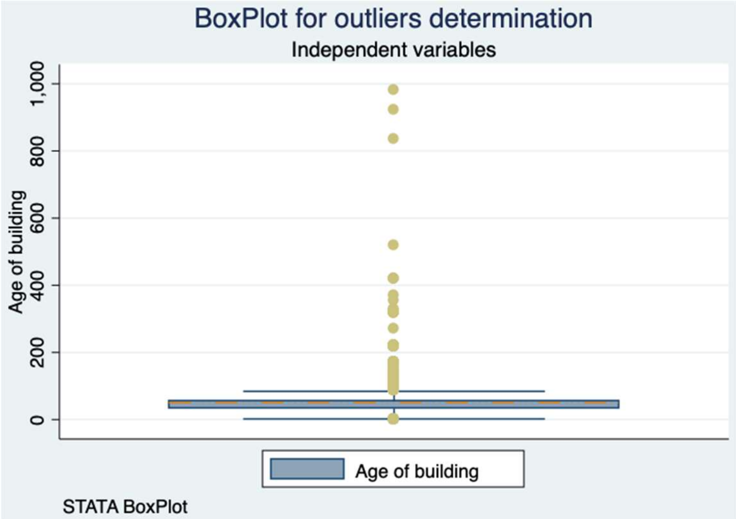

Figure 14 – Box Plot for Age

As for the value by square meter, I dropped the following: $Value/m^2 < 520€/m^2$ or $Value/m^2 \geq 3.450€/m^2$.

I considered observations with a value under the lower thresholds as outliers, classifying them as extraordinary cheap transactions, even though they are not highlighted statistically by the box plot (this is due to the high skewness caused by the huge number of observations in the right tail).

Moreover, looking at web data available at Immobiliare.it for real estate average quotations [46], this look quite reasonable since no city in our dataset is showing an average price below 900 €/m^2 (Figure 16).

For similar reasons, observations above the high threshold have been dropped as extremely expensive transactions. This time the box plot in Figure 15 supports the chosen upper bound; moreover, if you look at real estate quotations of Italian Revenue Agency, whatever the year, only historical buildings in Turin centre with optimum status of conservations and other peculiarities reached that price level.

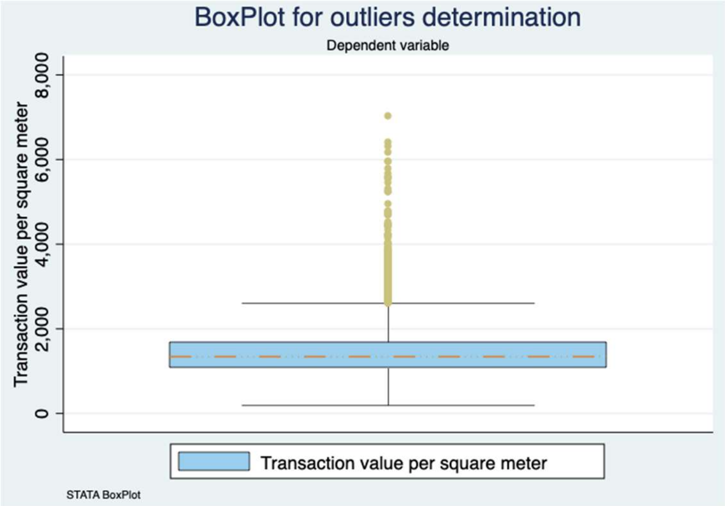

Figure 15 – Box Plot for Value by square meter

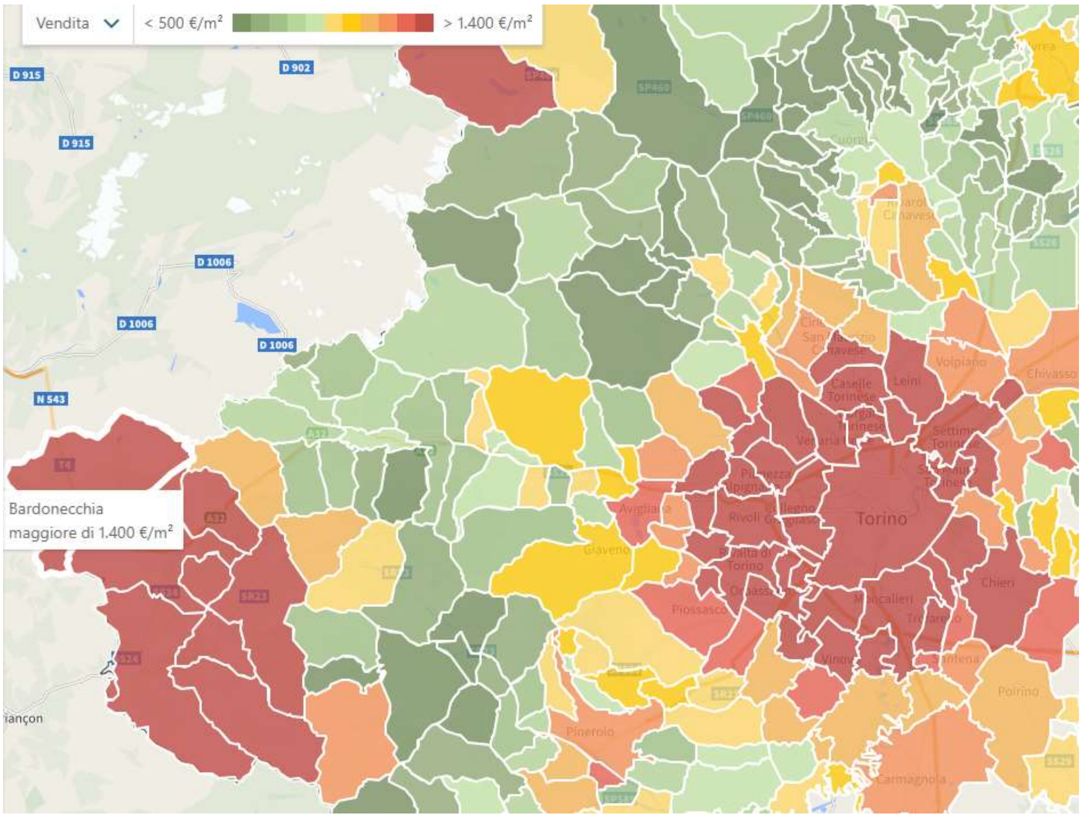

Figure 16 - Quotation map (Province of Turin) from Immobiliare.it

4.2 Results and comments

4.2.1 Empiric results and robustness checks

Table below summarizes regression results on the real estate unit value (in logarithm).

	Value (log)	Model 1 Base Model	Model 2 Full Model
Structural	Surface (log^2)	0.109***	0.109***
		(0.00)	(0.00)
	Age (log)	-0.058***	-0.058***
		(0.01)	(0.01)
	Lift (dummy)	0.168***	0.167***
		(0.01)	(0.01)
	#Bath	0.147***	0.147***
		(0.01)	(0.01)
	Maintenance - 1st	0.170***	0.169***
		(0.05)	(0.05)
	Maintenance - 2nd	0.264***	0.264***
		(0.06)	(0.06)
Neigh.hood quality	Neighbourhood (OMI) - 1st	0.192***	0.190***
		(0.04)	(0.04)
	Neighbourhood (OMI) - 2nd	0.440***	0.437***
		(0.05)	(0.05)
Neigh.hood services	Services - 1st	0.001	0.001
		(0.01)	(0.01)
	Services - 2nd	0.149***	0.145**
		(0.05)	(0.05)
Locational	Transports - 1st	0.197**	0.203**
		(0.06)	(0.06)
	Transports - 2nd	0.172**	0.177**
		(0.06)	(0.06)
	OMI Distance (dummy)	-0.085***	-0.085***
		(0.01)	(0.01)
Explanatory variables	**Internet Score**	**0.064*****	**0.058*****
		(0.01)	(0.01)
	Covid (dummy)		**-0.155***
			(0.08)
	Moderator		**0.093***
			(0.04)
	constant	8.737***	8.747***
		(0.10)	(0.10)
	R-sqr	**0.714**	**0.714**
	dfres	5519	5517

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

Figure 17 – Regression results

Model 0 – Base Model is meant to test **H1**, while Model 1 – Full Model is meant to test **H2**.

From Figure 17, it is evident that all estimates share a high significance level. The only estimate not reaching at least the 95% significance is the 1st level of *Services*. Moreover, all regression coefficients' signs are coherent with literature (see Figure 1).

However, before jumping to conclusions, I performed several sensitivity and robustness checks. First, I regressed the Base and the Full model on the same set of variables, changing *Internet_score* computation. Specifically, I computed the weighted average with α=0.55 (Model 3 and Model 4) and with α=0.6 (Model 5 and Model 6) to put a bit more importance on the home score, for testing how it would change the results. Specifically, I aimed to detect whether control variables remain stable and whether the explanatory variables maintain their sign and significance level. Results are present in appendix (Figure 30) and they confirmed that both the Base and Full models are stable, at least when varying *Internet_score* computation. As a matter of fact, all signs and significance levels are the same and all regression coefficients show values very close to the ones of Figure 17.

The major robustness check performed is the Jarque-Bera normality test on residuals [47]. This is a χ^2 test meant to assess whether residuals are distributed as $\backsim \mathcal{N}(0,1)$. In Figure 18, the results of the tests conducted over all the six models are shown.

Jarque-Bera residual normality test	**χ2**	***p*-value**
Model 1	5,489	**0,0643**
Model 2	5,049	**0,0801**
Model 3	6,076	**0,0479**
Model 4	5,912	**0,0520**
Model 5	7,494	**0,0236**
Model 6	7,624	**0,0221**
Reject normality if *p*-value < 0,05		

Figure 18 – Jarque-Bera residual normality test

Model 3, Model 5 and Model 6 failed to pass the test. For the others, the hypothesis of residuals being distributed as $\backsim \mathcal{N}(0,1)$ cannot be rejected. Hence, I kept Model 1 and Model 2 as reference models for checking **H1** and **H2** and, therefore, the comments will be based on their regression coefficients. The residual density plots in Figure 19 and Figure 20 provide a graphical idea of how good the two chosen models predict the dataset.

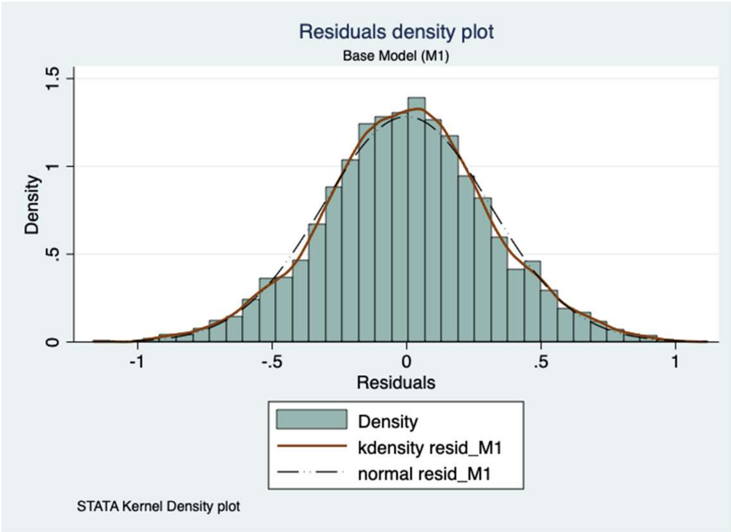

Figure 19 – Residual density plot of Base Model

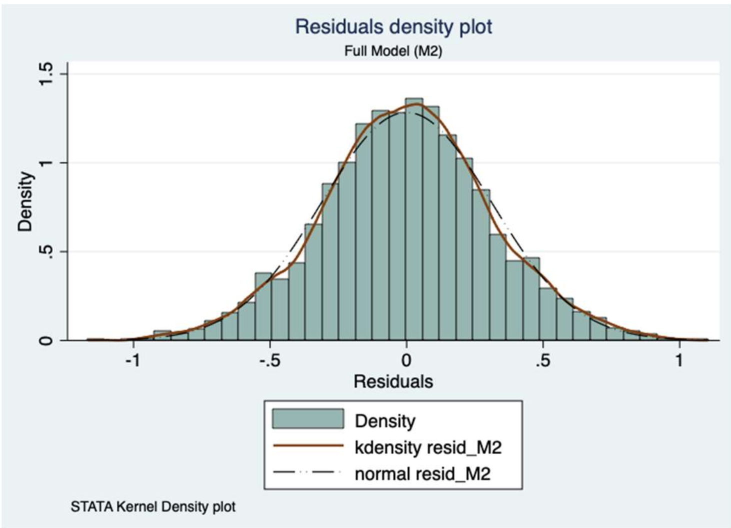

Figure 20 – Residual density plot of Full Model

4.2.2 Comments

As stated, comments are based on Model 0 and Model 1. Considering regression results, sensitivity analysis and the conducted robustness check, I consider both **H1** and **H2** proven. Looking in deep at estimation coefficients, they exhibit reasonable values. Figure 21 contains all the general comments.

Attribute Category	Variable	Impact	Significant	Comments
Structural	Surface	+	Yes	Positive impact on price. A 10.9% increase on value for a quadratic percentage unit increase of the floor area
Structural	Age	-	Yes	As the building gets 1% older, the value decreases by 5.8%
Structural	Lift_dummy	+	Yes	The presence of a lift generally increases value by 16.7%
Structural	Bath_no	+	Yes	An additional bathroom in the unit (holding constant the floor area) increases its value by 14.7%
Structural	Maintenance	+	Yes	Having a normal maintenance level increases property value by 16.9%. The same increase does not stand when passing from normal to optimal, obtaining an extra increase by only 26.4 - 16.9 = 9.5%
Neighbourhood quality	Neighbourhood	+	Yes	A qualitative normal neighborhood makes a 19% increase in value. Moreover, consumers are willing to pay an extra-premium (more than linear) for superior quality, with a 43.7% increase from poor to optimal
Neighbourhood services	Services	+	Partially	The coefficient for the absent→afar transition did not prove itself significant, whilst passing from afar to close significantly increase the value by 14.5%. This could be due to consumers perceiving the positive effect of local amenities proximity only when they are very close
Locational	Transports	+	Yes	When transport services are present, consumers are willing to pay a 20% premium. The premium is very similar (surprisingly even a bit inferior) when passing from absent to very close
Locational	OMI distance	-	Yes	Being located away from the city center decreases property value by 8.5%
Explanatory	Internet Score	+	Yes	Each 1 point of score of the Internet index accounts for a 5.8% increase in the property value
Explanatory	Covid_dummy	-	Yes	A real estate unit without internet connection and placed in a zone without mobile connectivity (Internet Score = 0) would ideally suffer a 15.5% decrease in price
Explanatory	Moderator	+	Yes	The negative Covid effect is mitigated by the internet connection quality

Figure 21 – Comments on estimation coefficients

The results for the explanatory variables reveal what I have postulated. The internet connection quality has indeed a positive impact on the property value. Furthermore, even though real estate market is suffering a decline in prices due to the pandemic, such an effect is mitigated if households can benefit from a good internet connection. Figure 22 shows it graphically.

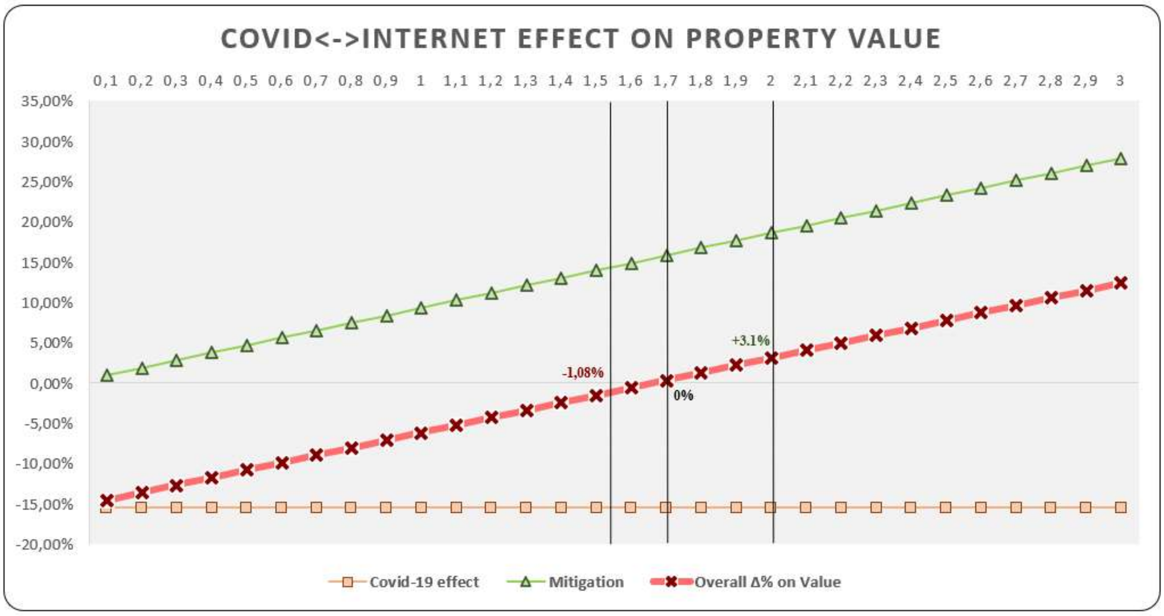

Figure 22 - Post-Covid effect on value by Internet score

The breakeven point is reached for $Internet_{score} \approx 1.7$. As highlighted in the graph, for a house with the average score ($Internet_{score} \approx 2$), the price is already above the breakeven point, with a net increase of 3.1%, whilst houses with a relatively bad internet connection are the ones actually suffering the pandemic effect (for instance, if you go 1 standard deviation point below the mean, it is $Internet_{score} \approx 1.55$ and $\Delta\%(Value) = -1.08\%$).

An additional fundamental comment is the following: the fact that **H2** has been proven true gives credibility to the evidence for **H1**. In fact, one could argue that considering the internet index is time invariant and city-related, it may account for the variance of other hidden attributes not included in the model which behave like the internet index and are city-related too. Nevertheless, when you add the evidence for **H2**, these alleged hidden factors should also interact with the Covid-19 effect in the same way the internet does, thus making their existence much less likely.

4.2.3 Results excluding Turin

To assess model validity even further I performed the same regressions (Base Model and Full Model) removing Turin from the dataset. Such an analysis would both prove whether the results are robust even removing a large portion of the dataset (Turin accounts for roughly 29% of the observations) and whether the evidence is true for the rest of the Province when removing its Central Business District.

Estimations of Model 7 and Model 8 are reported in appendix on Figure 31. Below the evidence:

- Estimations are in general slightly less significant, but this is very likely to be just due to the considerable reduction in the number of observations (and therefore a reduction in the degrees of freedom).
- Almost all control variables show pretty much the same value, meaning their results remain valid. The only relevant exception is the *Distance*, being not significant and positive. That could be because in most rural areas the "peripherical negative effect" of being distant from the center is outweighed, since rural outskirts do not exhibit the common feature of a suburban areas: instead, they are usually places with beautiful natural amenities (waterfront, lake, green landscapes, mountains for example), closeness to ski centers, quiet surroundings that are indeed positive factors not explicitly included in the model.
- Moderator and the Covid effect show consistent values close to the original ones, only with a slightly reduction in significance (they are significant at 10%). On the other hand, the Internet Score is not statistically significant, showing a value negative and very close to 0. An immediate interpretation of this result could be that in rural areas people had not been interested in the quality of their internet connection since the Covid-19 has changed their routines. In a practical sense, this actually supports the original hypothesis **H2**. Indeed, if one considers the whole sample, there is a

positive *a priori* effect of the internet quality, increased by the synergy with the pandemic; on the other hand, when the investigation is restricted to the rural areas this effect is absent before the pandemic, and appears only afterwards.

4.2.4 Limitations and further studies

From a statistical standpoint, estimates appear to be solid and backed up by reliable robustness checks. Nevertheless, there are some limitations to the investigation mainly linked to the sample and the explanatory variable.

- *Sample*: first, the sample does not cover all cities in the province. If one would like to extend the results to an entire geographical segment, data from all entities of that segment should be included, to eliminate the risk of mis-extending results to sub-segments which could potentially have different behaviours instead. In addition, considering the chosen watershed for classifying post-Covid data, only eleven months are available, compared to the pre-Covid transactions which are covered by thirty-one months. Hence, to be sure that the evidence describes a long-lasting behaviour instead of just a "flash in the pan", one could argue that a larger time span for post-Covid transaction would be required.
- *Explanatory variable*: as stated, the index for the internet connection quality is time invariant. It means I was forced to make the strong assumption that the infrastructure did not improve significantly over the time window. Even though to extend and to strengthen such kind of infrastructures requires considerable time, this is still a limitation. Moreover, the index is city-related, and that leads to an even stronger assumption. While this may sound reasonable for a very small city, that's quite pretentious for a big one like Turin. Furthermore, shrinking the entity specification of the internet index (ideally down to a single house unit) would reduce the risk this attribute is actually explaining the variance of other hidden factors.

Based on the above considerations, a direct extension of this thesis would be to repeat a similar investigation when more post-Covid data is available and to capture the internet quality with a more reliable index (built over time, shrank to smaller zones).

Moreover, it would be interesting to carry out the same type of investigation in other regions to prove that the evidence does not reveal a geographical limited phenomenon but rather an actual change in consumers preferences.

Conclusions

The two hypotheses of the thesis cannot be statistically rejected, as estimates have been proven to be significant and robust.

I have shown that, at least in Turin province, the price of a real estate unit is positively influenced by both the quality of the broadband technology coverage and the quality of the mobile connectivity, holding all the other attributes constant. Although the sample cannot be considered representative for the entire national territory, let alone greater geographical regions, I believe that this result could be proven true at least in most of the industrialized societies. This can be material for future studies.

I have also shown that in the Turin province, real estate market suffered an apparent decline in prices, that is mitigated by both the quality of the broadband technology coverage and the quality of the mobile connectivity. After the pandemic, the average house unit in terms of internet quality registered an increase in price of a roughly 3%, whilst the ones below the average have been left behind, drastically suffering the negative effects due to physical isolation and, in general, of spending much more time at home.

Moreover, I investigated whether this effect is the same when restricting investigation to rural areas. The evidence is that if one considers the whole sample there is a positive *a priori* effect of the internet quality, increased by the synergy with the pandemic; on the other hand, when the investigation is restricted to the rural areas this effect is absent before the pandemic, and appears only afterwards.

This result has significant economic implication, since it further supports the on-going digitalization strategy pursued by the government, which acquires even more relevance for a fast economic recovery of the Country.

References

[1] K. W. Chau and T. L. Chin, “A Critical Review of Literature on the Hedonic Price Model,” *International Journal for Housing Science and Its Applications,* vol. 27, no. 2, pp. 145-165, 2003.

[2] L. Boone and N. Girouard, “The Stock Market, the Housing Market and Consumer Behaviour,” *OECD Economic Studies,* vol. 2002, no. 2, 2003.

[3] A. M. Freeman, “Hedonic prices, property values and measuring environmental benefits: A survey of the issues,” *Scandinavian Journal of Economics,* vol. 81, pp. 154-171, 1979.

[4] D. G. Jud and J. M. Watts, “Real estate values, school quality, and the pattern of urban development in Charlotte, North Carolina,” *Economics of Education Review,* vol. 1, no. 1, pp. 87-97, 1981.

[5] R. Y. M. Li and H. C. Y. Li, “Have Housing Prices Gone with the Smelly Wind? Big Data Analysis on Landfill in Hong Kong,” *Sustainability,* vol. 10, 2018.

[6] A. Ibeas, R. Cordera, L. Dell'Olio, P. Coppola and A. Dominguez, “Modelling transport and real-estate values interactions in urban systems,” *Journal of Transport Geography,* vol. 24, pp. 370-382, 2012.

[7] M. Bottero, M. Bravi, F. Dell'Anna and G. Mondini, “Valuing buildings energy effiency through Hedonic Prices Method: are spatial effects relevant?,” *Journal Valori e Valutazioni,* vol. 21, 2018.

[8] S. Sheppard, “Hedonic analysis of housing markets,” *Handbook of Regional and Urban Economics,* vol. 3, pp. 1595-1635, 1999.

[9] S. Rosen, “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,” *Journal of Political Economy,* vol. 82, no. 1, pp. 34-55, 1974.

[10] M. L. Martinez and J. M. Viegas, “Effects of Transportation Accessibility on Residential Property Values: Hedonic Price Model in the Lisbon, Portugal, Metropolitan Area,” *Transportation Research Record,* vol. 2115, no. 1, pp. 127-137, 2009.

[11] N. H. Lansford Jr. and L. L. Jones, “Recreational and Aesthetic Value of Water Using Hedonic Price Analysis,” *Journal of Agricultural and Resource Economics,* vol. 20, no. 2, pp. 341-355, 1995.

[12] J. M. Clapp and C. Giaccotto, “Residential Hedonic Models: A Rational Expectations Approach to Age Effects,” *Journal of Urban Economics,* vol. 44, no. 3, pp. 415-437, 1998.

[13] H. W. Richardson, J. Vipond and R. A. Furley, “Determinants of Urban House Prices,” *Urban Studies,* vol. 11, pp. 189-199, 1974.

[14] D. Thaler, “A note on the value of crime control: Evidence from the property market,” *Journal of Urban Economics,* vol. 5, pp. 137-145, 1978.

[15] V. Del Giudice, F. De Paola and F. P. Del Giudice, “COVID-19 Infects Real Estate Markets: Short and Mid-Run Effects on Housing Prices in Campania Region (Italy),” *Social Sciences - Department of Industrial Engineering, University of Naples “Federico II”,* vol. 9, pp. 114-132, 2020.

[16] M. Billio and S. Varotto, “A New World Post COVID-19: Lessons for Business, the Finance Industry and Policy Makers,” *Edizioni Ca' Foscari - Digital Publishing,* 2020.

[17] G. Wong, “Has SARS infected the property market? Evidence from Hong Kong,” *Journal of Urban Economics,* vol. 63, pp. 74-95, 2008.

[18] J. Giesecke, W. Burns, A. Barrett, E. Bayrak, A. Rose, P. Slovic and M. Suher, “Assessment of the regional economic impacts of catastrophic events: A CGE analysis of resource loss and behavioral effects of a radiological dispersion device attack scenario,” *Risk Analysis,* vol. 32, pp. 583-600, 2012.

[19] P. Taltavull, “COVID 19 and real estate: a major challenge,” *Journal of European Real Estate Research,* vol. 13, no. 1, pp. 1-3, 2020.

[20] A. Bartik, Z. Cullen, E. Glaeser, E. Luca and C. Stanton, “What jobs are being done at home during the COVID-19 crisis? Evidence from firm-level surveys,” *Harvard Business School Working Paper,* vol. 20, 2020.

[21] S. Liu and Y. Su, “The impact of the COVID-19 pandemic on the demand for density: Evidence from the U.S. housing market,” *Economics Letters,* vol. 207, 2021.

[22] Agenzia delle Entrate, [Online]. Available: www.agenziaentrate.gov.it. [Accessed October 2021].

[23] Ministero dello Sviluppo Economico, “Strategia Italiana per la Banda Ultralarga "Verso la Gigabit Society",” MITD - Ministero per l'innovazione tecnologica e la transizione digitale, 2021.

[24] K. L. Conley and B. E. Whitacre, “Home is where the internet is? High-speed Internet's impact on rural housing values,” *International Regional Science Review,* pp. 1-30, 2020.

[25] G. Molnar, S. J. Savage and D. C. Sicker, “High-speed internet access and housing values,” *Applied Economics,* 2019.

[26] A. T. Court, “Hedonic price indexes with automotive examples,” *The Dynamics of Automotive Demand - General Motors,* pp. 98-119, 1939.

[27] C. Hultern, “Price Hedonics: A Critical Review,” *Economic Policy Review,* vol. 9, pp. 5-15, 2003.

[28] Z. Griliches, “Hedonic price indexes for automobiles: an econometric analysis of quality change,” in *The Price Statistics of the Federal Goverment*, vol. 31, NBER and University of Chicago, 1961, pp. 173-196.

[29] G. Sidak and J. Skog, “Hedonic Prices for Multicomponent Products,” *The Criterion Journal on Innovation,* vol. 4, pp. 301-379, 2019.

[30] M. Monson, “Valuation Using Hedonic Pricing Models,” *Cornell Real Estate Review,* vol. 7, 2009.

[31] E. Cassel and R. Mendelsohn, “The Choice of Functional Forms for Hedonic Price Equations: Comment,” *Journal of Urban Economics,* vol. 18, no. 2, pp. 135-142, 1985.

[32] N. Dusse and C. Jones, “A hedonic price model of office rent,” *Journal of Property Valuations and Investment,* vol. 16, no. 3, pp. 297-312, 1998.

[33] M. P. Murray, “Hedonic prices and composite commodities,” *Journal of Urban Economics,* vol. 5, no. 2, pp. 188-197, 1978.

[34] R. V. Butler, "The Specification of Hedonic Indexes for Urban Housing," *Land Economics,* vol. 58, pp. 94-108, 1982.

[35] S. Herath and G. Maier, "The hedonic price method in real estate and housing market research: a review of the literature," *Institute for Regional Development and Environment,* pp. 1-21, 2010.

[36] R. Maurer, M. Pitzer and S. Sebastian, "Hedonic price indices for the Paris housing market," *Journal of the German Statistical Society,* vol. 88, no. 3, pp. 303-326, 2004.

[37] M. R. Straszheim, "Estimation of the Demand for Urban Housing Services from Household Interview Data," *The Review of Economics and Statistics,* vol. 55, no. 1, pp. 1-8, 1973.

[38] A. Bisello, V. Antoniucci and G. Marella, "Measuring the price premium of energy efficiency: a two step analysis in the Italian housing market," *Energy and buildings,* 2019.

[39] P. Rosato, M. Breil, C. Giupponi and R. Berto, "Assessing the impact of urban improvement on housing values: A hedonic pricing and multi-attribute analysis model for the historic centre of Venice," *Buildings,* vol. 7, 2017.

[40] G. Lisi and M. Iacobini, "Measuring the Effect of Location on House Prices in Italy," *Regional Economic Development Research,* 2020.

[41] S. C. Bourassa, E. Cantoni and M. Hoesli, "Spatial Dependence, Housing Submarkets, and House Price Prediction," *The Journal of Real Estate Finance and Economics,* vol. 35, no. 2, pp. 143-160, 2007.

[42] P. De Toro, F. Nocca and F. Buglione, "Real Estate Market responses to COVID-19 crisis: which prospects for the metropolitan area of Naples (Italy)?," *Urban Science,* vol. 5, no. 23, 2021.

[43] "Schede Province Torino (Excel file)," Immobiliare, Agenzia delle Entrate - Territorio - Osservatorio del Mercato, Rome, 2021.

[44] "AGCOM - Autorità per le Garanzie nelle Comunicazioni," [Online]. Available: https://www.agcom.it. [Accessed 10 2021].

[45] "OPENSIGNAL," © 2021 Opensignal Limited - Mobile app, 2021.

[46] "Immobiliare.it - Mercato immobiliare - Quotazioni (Italia/Piemonte/Provincia di Torino)," Immobiliare.it, [Online]. Available: immobiliare.it/mercato-immobiliare/piemonte/torino-provincia/. [Accessed 11 2021].

[47] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review,* vol. 55, pp. 163-172, 1987.

[48] P. Pavese, "Hedonic Housing Price Indices: The Turinese Experience," *RIVISTA DI POLITICA ECONOMICA,* 2007.

[49] R. Halvorsen and H. Pollakowski, "Choice of Functional Form for Hedonic Price Equations," *Journal of Urban Economics,* vol. 10, pp. 37-49, 1981.

[50] A. Grybauskas, V. Pilinkienė and A. Stundžienė, "Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic," *Journal of Big Data,* vol. 8, pp. 105-125, 2021.

[51] M. J. Bailey, R. F. Muth and H. O. Nourse, "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association,* vol. 58, no. 304, pp. 933-942, 1963.

[52] G. R. Parson, "Hedonic Prices and Public Goods: An Argument for Weighting Locational Attributes in Hedonic Regressions by Lot Size," *Journal of Urban Economics,* vol. 27, no. 3, pp. 308-321, 1990.

[53] Idealista, 2021. [Online]. Available: www.idealista.it. [Accessed September 2021].

[54] "Italian National Institute of Statistics (ISTAT) website," ISTAT, 2021. [Online]. Available: www.istat.it. [Accessed September 2021].

Appendix

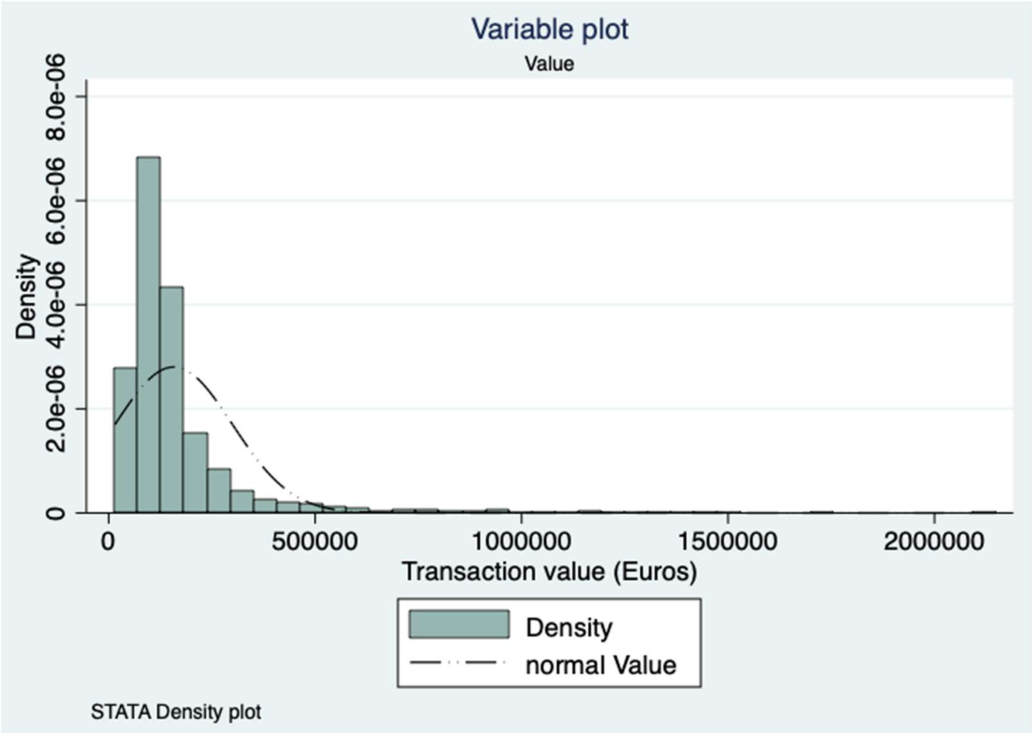

Figure 23 – Variable plot (Value)

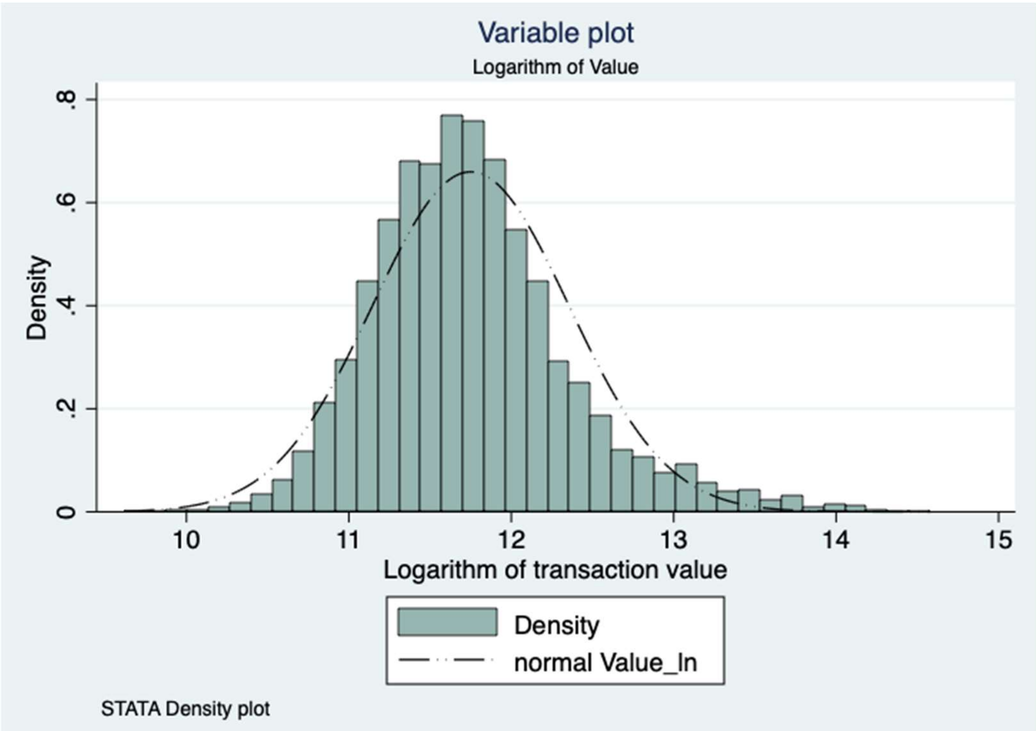

Figure 24 – Variable plot (Log of Value)

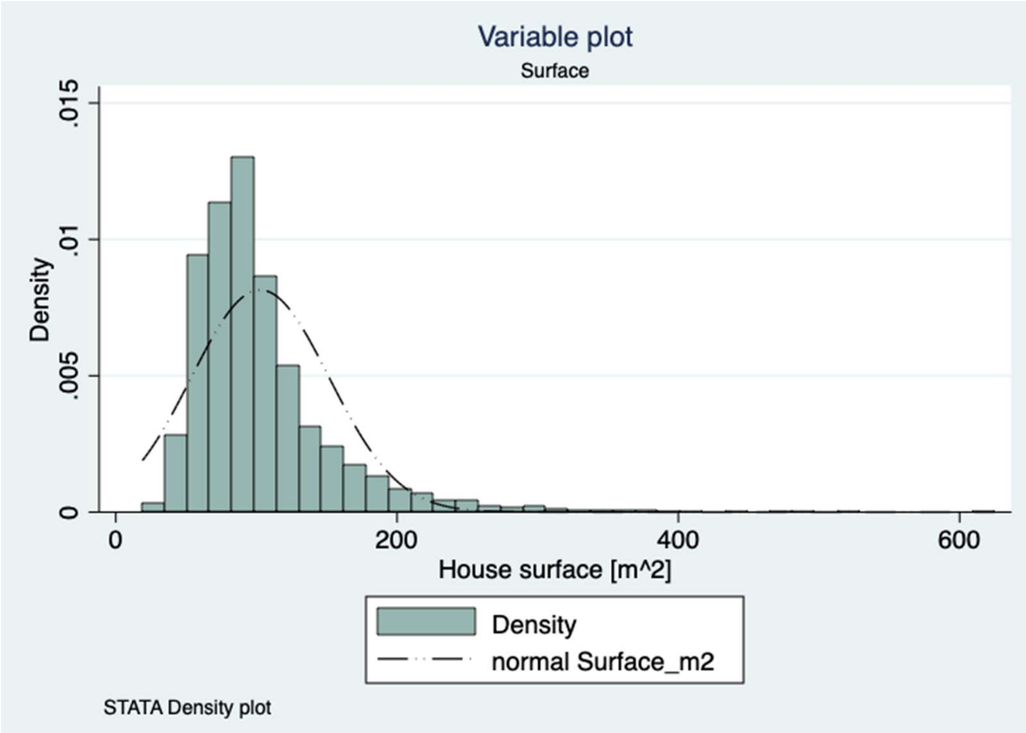

Figure 25 – Variable plot (Surface)

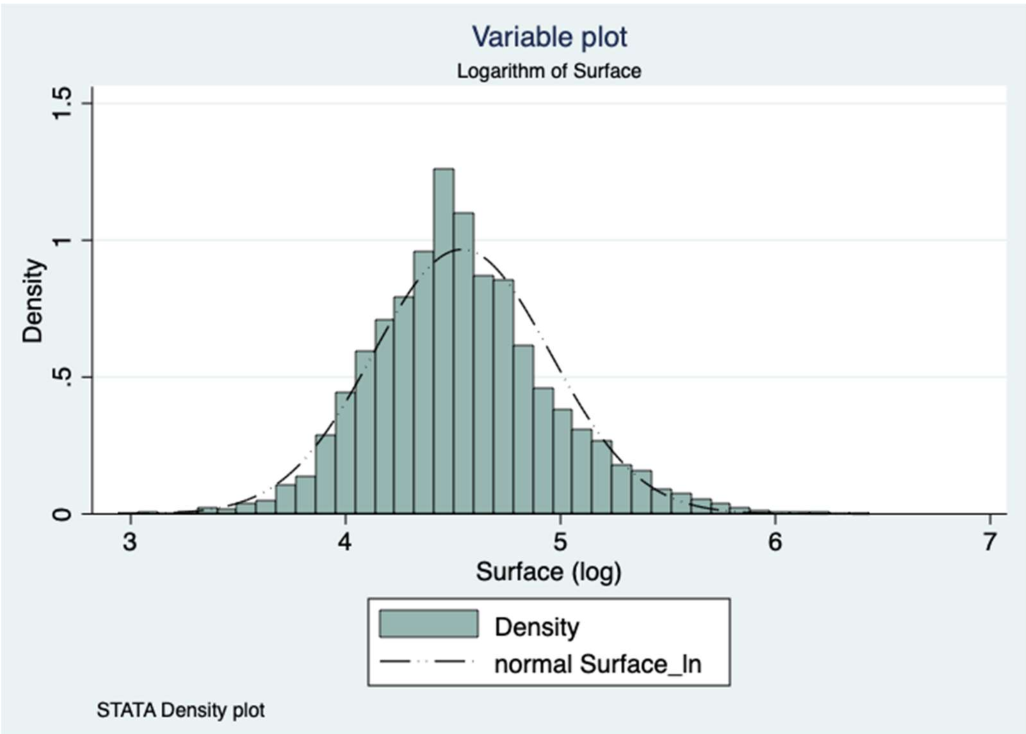

Figure 26 – Variable plot (Log of Surface)

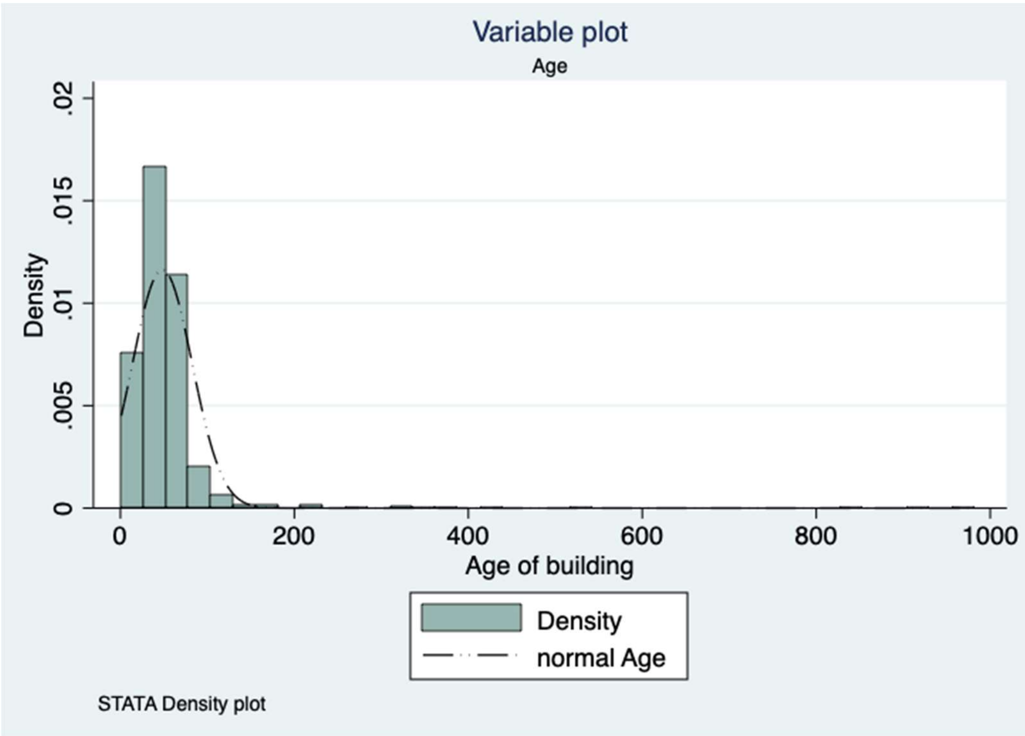

Figure 27 – Variable plot (Age)

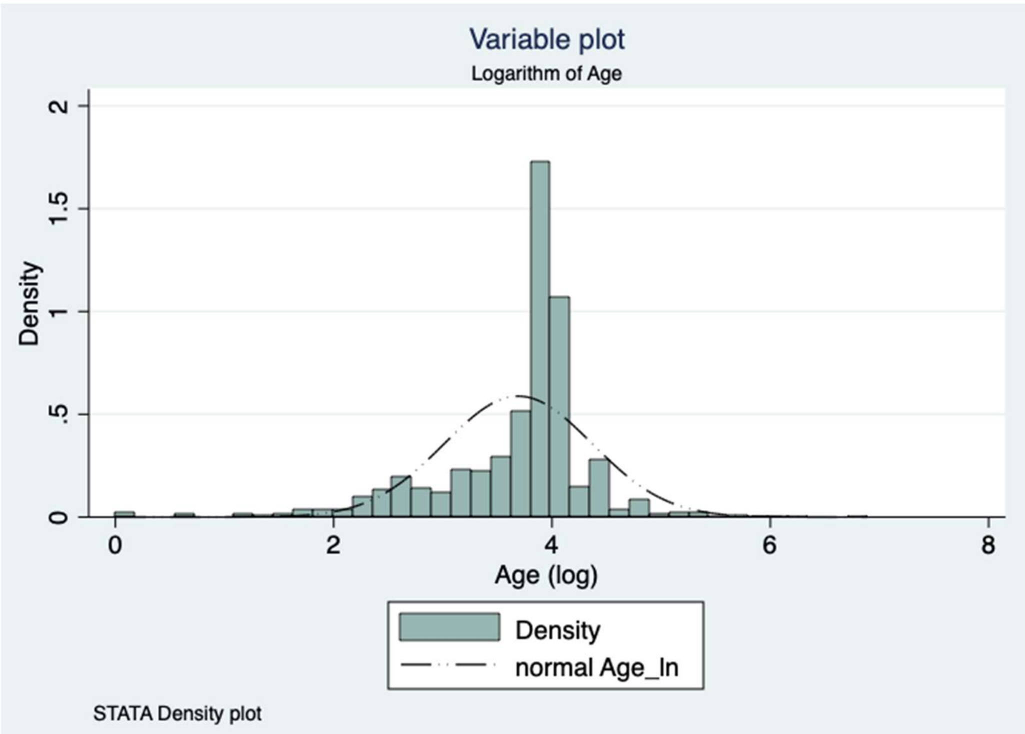

Figure 28 – Variable plot (Log of Age)

Correlation Matrix	*Internet score*	*Surfac (log^2)*	*Age (log)*	*Lift dummy*	*Bath no*	*Services*	*Transports*	*Neighborhood*	*Maintenance*	*Distance*
Internet score	1.000									
Surfac (log^2)	0.1734	1.000								
Age (log)	0.1754	0.0189	1.000							
Lift dummy	0.2221	-0.0142	-0.0478	1.000						
Bath no	0.0981	0.6644	-0.0586	0.0366	1.000					
Services	-0.2234	0.0758	-0.2170	-0.1574	0.0839	1.000				
Transports	0.2645	-0.0322	0.1500	0.1984	-0.0164	-0.3389	1.000			
Neighborhood	-0.1043	-0.0117	0.0570	0.0294	0.0340	-0.0715	0.0299	1.000		
Maintenance	-0.0046	0.0622	-0.2719	0.0397	0.0717	0.0475	-0.0328	0.0165	1.000	
Distance	0.0472	-0.0477	-0.1431	0.0440	-0.0220	0.1963	-0.0566	-0.0562	0.0204	1.000

Figure 29 – Correlation Matrix

	Value (log)	Model 3 Base Model (α=0.55)	Model 4 Full Model (α=0.55)	Model 5 Base Model (α=0.6)	Model 6 Full Model (α=0.6)
Structural	Surface ($\log^2$)	0.110***	0.110***	0.110***	0.110***
		(0.00)	(0.00)	(0.00)	(0.00)
	Age (log)	-0.056***	-0.056***	-0.054***	-0.055***
		(0.01)	(0.01)	(0.01)	(0.01)
	Lift (dummy)	0.171***	0.170***	0.173***	0.173***
		(0.01)	(0.01)	(0.01)	(0.01)
	#Bath	0.147***	0.147***	0.147***	0.147***
		(0.01)	(0.01)	(0.01)	(0.01)
	Maintenance - 1st	0.169***	0.168***	0.168***	0.167***
		(0.05)	(0.05)	(0.05)	(0.05)
	Maintenance - 2nd	0.266***	0.266***	0.267***	0.267***
		(0.06)	(0.06)	(0.06)	(0.06)
Neigh.hood quality	Neighbourhood (OMI) - 1st	0.190***	0.188***	0.189***	0.186***
		(0.04)	(0.04)	(0.04)	(0.04)
	Neighbourhood (OMI) - 2nd	0.433***	0.431***	0.426***	0.424***
		(0.05)	(0.05)	(0.05)	(0.05)
Neigh.hood services	Services - 1st	-0.001	-0.001	-0.003	-0.003
		(0.01)	(0.01)	(0.01)	(0.01)
	Services - 2nd	0.141**	0.137**	0.133**	0.128**
		(0.05)	(0.05)	(0.05)	(0.05)
Locational	Transports - 1st	0.194**	0.200**	0.191**	0.197**
		(0.06)	(0.06)	(0.06)	(0.06)
	Transports - 2nd	0.172**	0.178**	0.172**	0.178**
		(0.06)	(0.06)	(0.06)	(0.06)
	OMI Distance (dummy)	-0.083***	-0.083***	-0.082***	-0.081***
		(0.01)	(0.01)	(0.01)	(0.01)
Explanatory variables	**Internet Score**	**0.048*****	**0.042*****	**0.032****	**0.025***
		(0.01)	(0.01)	(0.01)	(0.01)
	Covid (dummy)		**-0.169***		**-0.198***
			(0.08)		(0.08)
	Moderator		**0.096***		**0.109****
			(0.04)		(0.04)
	constant	8.751***	8.762***	8.770***	8.784***
		(0.10)	(0.10)	(0.10)	(0.10)
	R-sqr	**0.713**	**0.713**	**0.712**	**0.713**
	dfres	5519	5517	5519	5517

* p<0.05, ** p<0.01, *** p<0.001

Figure 30 – Regression results for sensitivity analysis

Value (log)	Model 1 Base Model	Model 2 Full Model	Model 7 Base Model No Turin	Model 8 Full Model No Turin
Surface ($\log^2$)	0.109***	0.109***	0.093***	0.093***
	(0.00)	(0.00)	(0.00)	(0.00)
Age (log)	-0.058***	-0.058***	-0.147***	-0.147***
	(0.01)	(0.01)	(0.01)	(0.01)
Lift (dummy)	0.168***	0.167***	0.093***	0.093***
	(0.01)	(0.01)	(0.01)	(0.01)
#Bath	0.147***	0.147***	0.108***	0.108***
	(0.01)	(0.01)	(0.01)	(0.01)
Neighbourhood (OMI) - 1^{st}	0.192***	0.190***	0.195***	0.194***
	(0.04)	(0.04)	(0.05)	(0.05)
Neighbourhood (OMI) - 2^{nd}	0.440***	0.437***	0.398***	0.399***
	(0.05)	(0.05)	(0.05)	(0.05)
Maintenance - 1^{st}	0.170***	0.169***	0.155**	0.155**
	(0.05)	(0.05)	(0.05)	(0.05)
Maintenance - 2^{nd}	0.264***	0.264***	0.110	0.109^{Δ}
	(0.06)	(0.06)	(0.06)	(0.06)
Services - 1^{st}	0.001	0.001	-0.007	-0.007
	(0.01)	(0.01)	(0.01)	(0.01)
Services - 2^{nd}	0.149***	0.145**	0.097*	0.095*
	(0.05)	(0.05)	(0.04)	(0.04)
Transports - 1^{st}	0.197**	0.203**	0.123*	0.125*
	(0.06)	(0.06)	(0.06)	(0.06)
Transports - 2^{nd}	0.172**	0.177**	0.090	0.092
	(0.06)	(0.06)	(0.06)	(0.06)
OMI Distance (dummy)	-0.085***	-0.085***	0.012	0.012
	(0.01)	(0.01)	(0.01)	(0.01)
Internet Score	**0.064*****	**0.058*****	**-0.012**	**-0.017**
	(0.01)	(0.01)	(0.01)	(0.01)
Covid (dummy)		**-0.155***		**-0.141^{Δ}**
		(0.08)		(0.08)
Moderator		**0.093***		**0.074^{Δ}**
		(0.04)		(0.04)
constant	8.737***	8.747***	9.630***	9.638***
	(0.10)	(0.10)	(0.10)	(0.10)
R-sqr	**0.714**	**0.714**	**0.642**	**0.642**
dfres	5519	5517	3909	3907

Δ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Figure 31 - Regression results removing Turin from sample