A novel research strategy of measuring housing disadvantages of vulnerable populations for all income levels: the Propensity Score Matching approach.

Heijs, Joost and Cruz-Calderón, Selene Cruz

Universidad Complutense de Madrid

27 February 2023

A novel research strategy of measuring housing disadvantages of vulnerable populations for all income levels: the Propensity Score Matching approach.

JEL: J15; R21

Authors: Joost Heijs y Selene Fabiola Cruz-Calderón

1.- Introduction

1.1. Background and Main objective of the study

As will be explained in depth, this paper offers a new strategy—based on the Propensity Score Matching (PSM) technique and a factor analysis approach—to quantify the size of the housing quality gap (HQG) between individuals of the vulnerable and non-vulnerable segments of the population, considering the structural differences between both groups. The PSM method, in combination with a quantitative Housing Quality Index (HQI) based on a continuous numerical scale, makes it possible to isolate the part of the level of inequality of vulnerable groups caused by their ethnic or racial origin and the part caused by structural differences (income, education, among others). The proposed methodological approach also allows to improve several aspects of the existing empirical literature broadening its scope—as will be explained in detail—solving one of the main methodological problems: the selection bias. This, to infer the causal relationship between the ethnic or racial origin and the worse housing quality, comparing individuals with the same sociodemographic and socioeconomic characteristics.

The empirical studies show that there is a link between inequality and the cultural and ascriptive conditions of vulnerable individuals. Inequality and social exclusion go hand in hand leading to unequal distributions of human and physical assets and differential access to markets and services (Hopenhayn, 2008; Gandelman, 2011). For the selection of the vulnerable population assessed in this paper, we will follow Castellino's definition or classification of vulnerable groups: *"A group, numerically inferior to the rest of the population of a State, in a non-dominant position, whose members—being nationals of the State—possess ethnic, religious or linguistic characteristics differing from those of the rest of the population and show, if only implicitly, a sense of solidarity, directed towards preserving their culture, traditions, religion or language"* (2009; P. 11). In fact, Indigenous people, such as those from American countries, Australia or New Zealand, immigrants, people of African descent, women, people with disabilities and other groups defined by culture or adscription are considered vulnerable groups because they have often been the most disadvantaged in terms of social citizenship and public voice. In that sense, the measurement of social exclusion has turned out to be a challenge in social research (Hopenhayn, 2008).

Even though there has been progress in terms of political institutions, legislation and justice, and social policies, the social and power gaps that condemn these groups to greater poverty and exclusion or that impose greater social integration barriers on them than other groups in society

remain strong (Hopenhayn, 2008; Allman, 2013). However, over the last decades the social demands of vulnerable groups defined by their identity, culture or affiliation are acquiring unprecedented influence in politics of a wide range of countries (Dalton, 2017), posing growing challenges in the fields of political fairness and social justice. For all these reasons an in-depth analysis of the worse living conditions and the identification of the most affected group within the vulnerable population would be a relevant issue for policymakers.

In order to explain the usefulness of the research strategy we propose, we selected, for the Mexican case, two different vulnerable groups in terms of ethnic and racial origin: indigenous people and afro-Mexicans residing in the Metropolitan Area of Mexico City (MAMC). Analysing the case of the MAMC allows us to compare indigenous and Afro-Mexican population that share the same urban areas and the same public services with mestizos and white Mexicans (including areas with similar conditions of marginality and disadvantages)[1].

The availability of data has largely driven the selection of which dimensions of exclusion to measure such as: financial situation, ownership of durable goods, housing quality, perception of the neighbourhood, personal social relationships, physical health, and psychological wellbeing (Silver, 2007).In this study we analyse the social exclusion to which indigenous and Afro-Mexicans are exposed in terms of their access to good quality housing. . Housing, as a means to integration, reflects the way of accessing various social benefits, improving one's standing and reducing vulnerability (Valero, Coca and Miranda, 2010). Thus, housing and the organisation of the residential space reflect the life experience of vulnerable groups and can be considered as an important determinant of how each society addresses their integration and discrimination problems[2]. In this context the importance of dwelling units in the household economy has led to the development of many socio-demographic studies (Among others, Osili 2004; Navarro and Ayala, 2008; Palvarini and Pavolini, 2010; Bradley and Putnick, 2012; Roy *et al*., 2020), and there are many that analyse the marginalisation of vulnerable social groups in the housing market (Arias and De Vos, 1996; Johnston *et al*., 2007; Delgadillo, 2011; Reibel and Regelson, 2011; Saporito, 2011; CEPAL, 2021 .).

The selection of indicators is based on the multidimensionality of housing quality. The empirical literature uses large number of quiet different indicators and each of them has a particular importance in different scenarios of geographic location such as national context[3] or according to different income levels. Therefore, it would be a particularly interesting exercise to test whether the methodological strategy we propose works in such different settings and for different vulnerable groups. As mentioned, for the Mexican case, we compared the housing quality of

[1] In the regions of Mexico (Oaxaca, Chiapas, Guerrero, Chihuahua) where these groups of the population are more represented they are less likely to cohabit with mestizos or white Mexicans.

[2] For example, recent studies have found that one in every three households in Latin America and the Caribbean (LAC) struggles with inadequate housing: 5 million households rely on another family for shelter, 3 million live in houses that are beyond repair, and 34 million lack one or more of the following: running water, sewage, adequate flooring and sufficient space (Bouillon 2012; McTarnaghan *et al*., 2016).

[3] For example, the datasets used for this paper show that 98% of the population in Spain has direct access to all basic services (water, electricity and sewerage) while in Mexico the percentage of population without access to these services is 26.7%.

indigenous and afro-Mexican population versus that of the mestizos and white Mexicans. However, the proposed research strategy can also be used for other vulnerable groups that suffer different forms of marginalization or discrimination, such as sexual preferences or gender differences or for other indicators of inequality (like the level of wages or over education).

The paper has the following structure. In the next subsection (1.1) the novelties of our methodological approach and the research questions analysed are discussed followed by a discussion of the main results (1.2). Section 2 offers a literature review on the situation of the indigenous and afro-Mexican population versus that of the mestizos and white Mexicans in the MAMC and the following section describes the database used and specifies the development process of the first stage of our research: the creation of a numerical housing quality indicator with a continuous range of values, using the factor analysis approach.

Section 4 and 5 reflect the second stage of our research in which we used the PSM technique to estimate the individual and average Housing Quality Gap (HQG). Section 4 specifies the equations that the PSM models and discusses the selection of matching variables and specifies the technical adjustments of the models. While section 5 presents the main results obtained from the PSM models and their econometric suitability. Section 6 offers the profile of the persons with a lower or higher HQG—based on a regression model on the HQG.

Section seven offer a brief analysis using the same methodology to analyse the Housing quality gap of immigrants in Spain. While section 7 offers a final discussion and the approach to the methodological problems solved during the process of preparing this paper.

1.2. Research strategy and novelty of the methodological approach

One of the main contributions of the methodology used is that the Propensity Score Matching (PSM) technique considers structural differences. In fact, when it comes to the development of techniques that make it possible to understand the phenomena inequalities caused by ethnic and racial disparities, there is fully awareness that the main aspect to take in consideration is that the characteristics of those who are in a situation of vulnerability are very different from the rest of the population. People's multiple exchanges in their daily social lives occur within a certain set of social, economic, and political institutions (formal and informal) that provide the opportunities and services needed, social exclusion takes place when those rules negatively constrain the performance of some groups or agents (Gandelman, 2011). Such process engendered structural social economic differences between vulnerable and non-vulnerable groups. In fact, the aforementioned vulnerable groups (indigenous and Afro-Mexican population) have on average, lower wages and educational levels and are more likely to work in the economy's informal sector without social protection compared to the rest of the population (Hopenhayn, 2008). These structural differences—resulting from the accumulation of a long process of marginalization and cultural traditions and habits—make it difficult to detangle the causes of the disadvantages in their social-economic situation as a direct product of discrimination based on their origin.

This means that if we simply compare the average housing quality (HQ) between vulnerable groups and the rest of the population, a substantial part of the differences observed might be the consequence of structural differences in their level of income and education or cultural preferences and traditions rather than discrimination against their ethnic or racial origin. More precisely, if we were to measure the gap in housing quality caused by discrimination against these vulnerable groups using that simple mean, we would incur in selection bias. In this paper we propose the Propensity Score Matching technique which—as a semi-experimental approach—focusses on the creation of an unbiased control group for the vulnerable population. Such unbiased group makes it possible to isolate the part of the worst housing conditions generated by the structural differences between vulnerable and non-vulnerable individuals and that part of the housing quality gap (HQG) generated by discrimination against their ethnic-racial condition, solving the aforementioned selection bias.

For this purpose, the proposed methodological approach intends to analyse a broader set of aspects than that of the existing empirical literature in order to broaden the scope of empirical evidence. A first way to broaden the scope of existing studies is to move beyond their focus on the disadvantage and/or segregation of only the poorest sectors of the population and their focus on only on those in low-paying jobs, often in informal labour markets. Our method complements the use of mainly a binary criterion of disadvantage by establishing a "level" of disadvantage. Likewise, most studies establish a minimum level of conditions required to ensure a socially acceptable standard of living, considering that (vulnerable) people below that level are considered deprived. In that way, in most cases the studies only analyse individuals based on the identification of their deprivation. Although some studies compare the level of deprivation between the vulnerable and non-vulnerable groups, showing that the level of deprivation is lower for the non-vulnerable, they continue to focus on the poorest sectors of the population.

All these studies are very useful for policy makers because they identify important problems and offer interesting policy recommendations. However, to the best of our knowledge, they do not directly compare vulnerable individuals with clones of the non-vulnerable population in the same socio-economic situation. Descriptive studies show the differences but do not assure that they can be attributed—causally—to the condition of vulnerability. Which brings us to our first research question:

1. Do vulnerable people live worse than their non-vulnerable clones with the same socio-demographic and socio-economic characteristics? —and therefore, this situation can be causally attributed to their condition as vulnerable individuals?

The idea is to analyse the intersection of the condition of vulnerability by ethnicity and race mainly with the class condition, that is, with different income levels. Likewise, structural differences would intersect, such as educational level, age as proxy for the life cycle, gender or household structure. The PSM technique—as proposed in this paper—makes it possible to identify for each vulnerable individual a nonvulnerable clone or counterfactual with exactly the same socio-economic characteristics except for their ethnic and racial condition and their housing quality (HQ). Once the clone has been identified, the differences in HQ between the two individuals

establish the so-called housing quality gap (HQG), which can be interpreted as an unbiased estimation of the level of disadvantage directly caused by ethnic or racial condition. This would reflect the HQ the vulnerable persons would have had if they were in a non-vulnerable position based on race or ethnicity.

Another way to broaden the scope of existing studies is to complement the binary bottom-line approach in two manners. On the one hand, a profile of the people below that line can be developed (using a logistic regression model). On the other hand, it is also interesting to analyse the profile of those disadvantage people who have the worst HQ conditions. Such profile could identify those disadvantage population groups that require greater attention from the government and, therefore, would make it possible to fine-tune public policy measures. Therefore, instead of a bottom-line level approach we propose a multidimensional complex housing quality indicator (HQI) that implies a continuous numerical measure. Again, the PSM method would be useful to analyse our second research question:

2.- Are there persons within the same vulnerable and disadvantage population group and with a similar income level in housing quality who show a wider HQ gap than others?

To answer this question, we use only the sample of vulnerable persons and analyse the characteristics of those with the largest HQG. Especially, we identify within the vulnerable group with a low-income level those subsections with specific characteristics that suffer more disadvantages (a wider HQ gap) in order to recommend specific adjustments for policymakers.

The bottom-line approach also implies that the effects on the living condition of vulnerable people are basically analysed for the poorest echelons of the population. It would be interesting to analyse whether individuals having a higher income level, although belonging to a vulnerable group (not being "deprived" at least in economic terms) also suffer worse living conditions than their non-vulnerable counterfactuals with the same socio-economic situation. Therefore, we included in our HQ indicator not only housing conditions related to basic services and amenities but also secondary and even luxury aspects. By creating such a synthetic index, we are able to capture the gap for individuals of all kinds of income levels and we are able to identify those who are more disadvantaged than others for each of the different income levels in society. Which bring us to two additional research questions:

3.- Do vulnerable people from the middle- and high-income strata also suffer worse living conditions (housing) than their non-vulnerable clones? —and therefore, this situation can be causally attributed to their condition as vulnerable individuals?

4.- Is the profile of the persons with a larger/lower Housing Quality Gap (HQG) different depending on their income level?

As mentioned before, to address these questions the rigorous statistical analysis that the PSM technique offers makes it possible to differentiate between the part of the housing quality gap caused by structural differences and the part caused by discrimination against racial-ethnic condition. We designed three stages research strategy using data from the Population and Housing Census (PHC) of Mexico.

Stage 1: Creation of a synthetic HQ index using the factor analysis approach.

Stage 2: Calculation of the HQG for individuals of different income intervals using the PSM method.

Stage 3: Elaboration of the profile of those people with a vulnerable condition who suffer a greater or smaller HQG.

In the first stage we create a complex numerical synthetic housing quality indicator (HQI) based on a broad set of 28 variables that characterise the dwellings, that include, among others information on the size in terms of the number of persons by square meter or by room (level of overcrowding), the technical quality of the building, and the availability of a broad range of amenities and services. This last group consist of binary variables that indicate the availability of certain services and amenities. To obtain several quantitative indicators with a continuous range of values instead of binary ones, we created combined indicators summing up the number of different types of services or amenities (being the number of basic services; the number of secondary amenities and the number luxury non-essential equipment). Such numerical approach is important, because this is what allows—as will be explained in detail—to quantify (estimate) the quantitative size of the housing quality gap (HQG) for each vulnerable person compared to their non-vulnerable counterfactual observation.

Once the different quantitative indicators are created, we applied a factor analysis to reduce them into a single concise synthetic indicator. The factor analysis (FA) throws up three factors—that will be used as partial indices—and for each individual the average value of these three factor scores will be used as our unique synthetic Housing Quality Indicator (HQI). An important advantage of the use of the FA is that, implicitly, it generates a standardisation of the data measured in very different scales. The obtained factors or partial indexes reflect standardized values (mean of zero and standard deviation of 1), therefore their values and the size of the later estimated partial and global gap will be directly comparable to each other. Like a final remark it can be stated that the developed index contemplates a continuous measurement of characteristics that represent not only the housing quality of the most disadvantaged sectors of the population, but the entire spectrum and, therefore, can be used to calculate the HQG for different types of income levels. This way of analysis has an important implication because most existing literature explore basically the most disadvantage or poorest sectors of society.

Once established the qualitative (HQI) with a continuous range of values we use—in a second stage—the Propensity Score Matching (PSM) technique to address the so-called selection problem aforementioned. Despite being originally developed to unbiasedly estimate the effect of an

intervention (medical treatments or public policies), this method is nowadays used to analyse discrimination against vulnerable population.

The outcome of the PSM model offers the average of the differences in the housing quality of all matched couples: the average housing quality gap (AHQG). In addition, repeating the models for different income levels we find out whether such gap also exists for the higher income groups. This way, we answer our first three research questions. In the third stage of our study, we answer the question 2 and 4. Which profile of vulnerable individuals suffer a larger HQG than others? This individual HQG will be used as the dependent variable in a regression model to characterise the people most affected by a wider or smaller HQG. While as (independent) explanatory variables of such gap we include their socio-cultural and economic situation. The definition of this profile is the second main contribution to the existing literature on this subject.

Concluding, in order to answer the four proposed research questions, we have to solve the so-called selection bias problem, creating unbiased control groups of non-vulnerable individuals before trying to directly compare both types of individuals (vulnerable versus non-vulnerable). Comparing the average HQ of both groups based on the whole sample would not reflect either the real level of discrimination against vulnerable groups or a situation of causality in the gap. In this sense, the selection bias correction is the main novelty of the proposed research strategy. Although this study does not propose a definitive solution to solve this methodological problem, our approach—based on Propensity Score Matching—at least offers the possibility of analysing with greater precision the explanatory factors of inequality in terms of housing that affects the vulnerable population of our interest.

In the following lines we briefly explain the database used and the main results of our analysis to compare the HQ between indigenous or Afro-Mexicans and the rest of the Mexican population. We used data from the Mexican "Population and Housing Census" [4]. The sample includes 692,444 observations covering the whole population between 16 and 75 years of age, of which 7.9% self-identify as indigenous people, 1.3% as Afro-Mexican, and 1.4% are indigenous that speak an indigenous language. The factor analysis applied to the combined 28 basic variables that reflect the HQ of each individual to synthesised them, lead us to obtained three partial indices: (F1) Services, supply and amenities of the dwelling; (F2) Housing density; (F3) Construction materials. We synthesised these partial indices in a unique HQ indicator by calculating the simple average.

The results of the PSM models show that indigenous people who speak an indigenous language suffer from a clearly lower HQ than their paired clones (mestizos or white Mexicans). This difference in socio-economic level also exists, albeit less pronounced, for self-identified indigenous people (non-indigenous language speakers) compared with other Mexicans. Afro-Mexicans seem to present, on average, similar housing conditions to their mestizo counterparts. However, this result does not imply the absence of housing problems for this population.

Hence, it could be possible to consider that the housing conditions of indigenous people in MAMC can be explained by causes related to their ethnic identification. Our results could help

[4] Conducted by the National Institute for Statistics and Geography (Spanish acronym: INEGI)

policymakers identify specific aspects to redefine existing policies at different governmental levels resolving existing social inequalities and discrimination processes, improving the living conditions of indigenous people and Afro-Mexicans.

2.- Housing quality and disadvantages of the vulnerable populations in Mexico

In Mexico the recognition of otherness was occupied almost exclusively by the indigenous peoples leaving aside Afro-Mexicans who had been invisible from the national consciousness, however, the stigmatization of both populations is strongly linked to nationalist "mestizaje", Mexico's dominant racial paradigm (Vaughn, 2013). Even though in Mexico, the racism experience has been always linked to European and American cultures, the marginalization and inequalities between Indigenous groups and Afro-Mexican indicate that racism is a problem that must be addressed (López, 2019). Separating classism from racism and attributing all inequities to class disparities has never been a right path to achieve equity (López, 2019). To us, it is important to acknowledge that vulnerable groups are affected by multiple aspects that intersect making them less or more susceptible to exclusion or poverty.

The marginalisation of the indigenous population living in Mexico, and in Mexico City, is a well-known problem generated by a historical process of segregation and supported by their negative and prejudicial colonial representation (Nieto, 2018). On the other hand, Afro-descendants in Mexico have remained an invisible population whose situation is rarely studied, showing their exclusion from the notion of "Mexican," while suffering from prejudice and negative stereotypes (Torre and Sánchez, 2019). However, the role of ethnic or racial origin has recently gained more attention in the academic literature of Latin America and the Caribbean (Gandelman *et al*., 2011). The few studies (Barbary, 2015; Torre and Sánchez, 2019; CEPAL, 2021) on the subject have revealed that indigenous and Afro-descendants live in worse conditions than mestizos and whites in terms of income, education, health, physical infrastructure, access to the labour markets, formal jobs and housing (Gandelman *et al*., 2011). In terms of housing, this author highlights that a substantial part of this population suffers from a lack of access to basic services (water supply, heating and drainage), adequate dwelling size and the materials used for construction. Yet, their exclusion, especially that of Afro-Mexicans, has been rarely studied in the Mexican case, basically due to the lack of systematic and representative statistical data. However, since 2015, Mexico included a question about Afro-descendant self-identification in the Intercensal Survey that enhanced the visibility of this segment of the population who lacked an official count and statistical recognition (Torre and Sánchez, 2019).

Most studies analyse the bad situation or deprivation of those groups but do not compare their situation directly with very poor mestizos or white Mexicans. Although the conclusions of such descriptive studies are probably correct, the lack of rigorous econometric analysis in the available literature makes it hard to statistically accept or reject the hypothesis of ethnic discrimination. Such a situation increases the need for research to go beyond paying special attention to causal relationships.

Several authors highlight that the lack of systematic and representative statistical data has made it difficult to analyse this subject using sound econometric models that statistically test whether socio-economic disparities among Afro-Mexicans and indigenous Mexicans are a result of human capital composition or of an ethno-racial discrimination process (Ñopo *et al.,* 2010; Torre and Sánchez, 2019). Moreover, existing studies basically compare the two populations (indigenous or

Afro-Mexican versus other Mexicans) without considering the self-selection problem: the ethnic population is quite different in most characteristics from the rest of the inhabitants. Another limit is that they focus mainly on the poorest levels of incomes, whereas in this study we try to analyse whether ethnicity also affects the housing conditions for persons on high and medium-high incomes.

In this section, we give an overview of the situation of the living conditions of the indigenous population and of the Afro-Mexicans, following a review of the scant literature available. Following Nieto (2018), indigenous marginality is increasingly becoming an urban matter, in spite of decades of intense migration from rural areas to metropolitan Mexico City. Native groups settled in the city are still considered "outsiders" and "deterritorialised". This dynamic determines the consolidation of prejudicial portrayals of natives, identifying them not only as 'outsiders', but also as 'rural', 'poor', 'ignorant', 'dirty' and 'drunk' people (Oehmichen, 2001; Saporito, 2011). In addition to this snapshot, indigenous Mexican people are often located in informal settlements referred to as areas for "the excluded society"[5], settling in self-built popular colonies or squatting in the crumbling dwellings of the city centre (Saporito, 2011).

Native communities in Mexico City are mainly concentrated in the in the *vecinades*[6], of the urban centre and in the popular colonies of the east, most of them sharing the same urban areas with similar conditions of marginality and deprivation as the non-indigenous lower classes or under-classes (Saporito, 2011). Looking at their standard of living, Saporito (2011) highlighted that the indigenous population showed a very low level of quality of life, in particular due to the reduced accessibility to social services, such as healthcare, housing and education.

Putting the focus on Afro-Mexicans, the existing evidence is significantly limited, especially due to the lack of (official) statistical data. From a legal point of view, the Mexican Political Constitution has recognised the multicultural composition of the nation since 2001, originally based on indigenous populations (Lara, 2011). However, under federal legislation, Afro-descendants were not mentioned as part of this cultural mosaic[7] until 2019. When the new Article 2 of the Political Constitution of the United Mexican States, recognised the Afro-Mexican population and communities, whatever their self-denomination, as part of the multicultural composition of the nation (DOF, 2019). This lack of official recognition posed several disadvantages for this segment of the population with important effects on their quality of life, being excluded from the support from governmental programs aimed at improving poor living conditions (Torre and Sánchez, 2019).

[5] Working as hawkers or domestic workers, shop workers, men enlisted in the military, seasonal workers or as craftsmen selling their products on the streets.

[6] A type of multi-family housing that normally consists of small houses located on the sides of a corridor or central patio. Families living in a *vecindad* sometimes share "some other services". There are two types: 1) old houses or buildings that were abandoned 2) buildings expressly build for affordable housing to lower-class sectors of the population.

[7] For a long period of time, only a regional law of 1998 recognised the so-called "Afro-Mexican" ethnic group, in the state of Oaxaca (Lara, 2011).

In fact, following our sample, 7.3% of the Afro-Mexican population in MAMC live in the city centre, while 36% and 31.4% live in peripheral areas (contours number 3 and 4, respectively)[8]. In these two contours, especially in contour 4, despite some of their rural characteristics, there is intense demographic pressure and land speculation (Toscana and Pimienta, 2018), a fact that affects the cost of housing.

According to the scant literature on the subject, Afro-descendants are more visible seems to experience higher rates of economic and social marginalisation, reflected in their limited access to basic infrastructure, education and health services, as well as other political, social and cultural rights (Velázquez and Iturralde, 2012). However, Torre and Sánchez (2019) unexpectedly found that Afro-descendant self-identification is associated with higher levels of education and better occupational status. In fact, they are similar to those of the rest of the population. Such findings contrast with recent information reported by the Economic Commission for Latin America and the Caribbean[9] in 2021, which points out their overrepresentation in situations of poverty and vulnerability to poverty in Latin America. The percentage of Afro-descendants who live in overcrowded housing and who suffer severe or moderate deprivation of access to basic services is higher than that of non-Afro-descendants, in both urban and rural areas, in most of the countries for which information is available. The most recent census data indicate that, in urban areas in 11 out of 15 countries of L.A., the percentage of Afro-descendants deprived of access to water, ranges from 3.1% in Costa Rica to 29.9% in Mexico (CEPAL, 2021).

As mentioned, in contrast to the living conditions of the indigenous population, the situation of Afro-Mexicans in Mexico is scarcely documented. Therefore, it is very important to analyse how ethnicity plays a role in determining indigenous and Afro-descendant exclusion to access of good quality housing.

[8] For details of contours conformation see Annex I.
[9] Spanish acronym: CEPAL.

3.- Creation of the numerical housing quality indicator with a continuous range of values: the factor analysis approach

3.1. The use of housing quality indicators in the existing empirical studies

The assessment of housing quality is a long-standing research area and also includes the adequacy of the neighbourhood, these being aspects that have been studied since at least the 1950s (Nelbach, 1950; Kain *et al.,* 1970). Housing quality is a multifaceted concept involving multiple themes and ideas (Clapham, 2009; Hatuka and Bar, 2017). Throughout this section, a taxonomy[10] of the indicators used in the evaluation of housing quality will be developed. The foregoing, from the review of the literature on the subject in the international context, particularly the European and Latin-American. Despite the long-standing tradition, it is difficult to define an internationally accepted standardised housing quality index due to differences between countries in terms of the level of income, culture and habits, expectations and minimum requirements that broadly differ between them. Especially because of the need to adapt them to the circumstances of each country, region, income level of the inhabitants, etc. This means that the usefulness of the results for certain policy recommendation for a country or region is not always suitable for other social or national contexts. Furthermore, they are often presented in formats inaccessible to a practitioner audience, limiting the degree to which research has informed housing policies and programs (McTarnaghan et al., 2016).

The potential set of indicators can be limited to the most technical-physical aspects of the housing condition: size, quality, deterioration of the dwelling and the existing amenities and public services being directly part of the living conditions. However, the literature also offers a broader view on housing quality that includes several socio-cultural and institutional aspects.

The United Nations Human Settlement Program (UN-HABITAT) established in 1996 a comprehensive set of criteria or parameters, designed at an international level, that should be taken into account (Habitat II Conference, 1996; McTarnaghan et al., 2016). The categories include not only the physical-technical aspects though also the social aspects and security.

The next four categories proposed by UN-HABITAT include parameters that measure the technical quality of the dwellings and the costs of housing, considering:

1. **Habitability**: physical safety or provide adequate space, as well as protection against the cold, damp, heat, rain, wind, other threats to health, and structural hazards.
2. **Availability of services, materials, facilities, and infrastructure**: safe drinking water, adequate sanitation, energy for cooking, heating, lighting, food storage, or refuse disposal.
3. **Affordability**: costs should not threaten or compromise the occupants' enjoyment of other human rights.

[10] This classification is partially based on the comprehensive overviews of the existing literature carried out by Sinha *et al* (2017) and Brkanić (2017) and their papers were broadly used for this section.

The following four categories include—at least implicitly—social aspects and security circumstances:

4. **Location**: employment opportunities, access to health care services, schools, childcare centres, and other social facilities, and the level of pollution or criminality.
5. **Security of tenure**: legal protection against forced evictions, harassment, and other threats.
6. **Cultural adequacy**: take into account the expression of cultural identity.
7. **Accessibility**: specific needs of disadvantaged and marginalized groups are taken into account.

Although several studies have proposed a measurement model for HQ trying to follow these parameters, the empirical literature still uses several ways to measure security aspects and socio-cultural circumstances. Mainly due to the lack of the appropriate information. To get an overview, we studied a broad number of additional papers on housing quality and pay especial attention to the comprehensive overviews of the existing literature carried out by Sinha et al (2017) and Brkanić (2017).

The literature review made it possible to identify and classify the most used "objective" or "technical" indicators without taking into account other relevant subjective aspects related to the perception, priorities or available income of the residents of the dwellings, being an aspect outside of the scope of empirical analysis of the thesis.
The identified indicators were the following:

1. **Technical quality of the housing unit** (available space and its functional distribution; physical quality of the construction materials; wear and tear and lack of maintenance).
2. **Technical design of the unit** in terms of basic services, non-essential amenities and functional distribution.
3. **Aesthetic design, ambience and comfort of the dwelling.**
4. **Features and adequacy of the geographic and social environment of the neighbourhood.**
5. **The direct costs of housing (affordability)** (rent, mortgage, maintenance, costs of basic services).

The first group of indicators—the technical quality of the housing unit—reflects, in fact, the basic intrinsic characteristics of the dwelling. An initial aspect would be **the space available,** such as size, the number and type of rooms and their adequate distribution. Another aspect is **the physical quality of the construction materials,** which is not only related to the satisfaction of the immediate user of the unit but also the foreseeable cost of maintenance. Part of the physical quality of the housing unit (level of insulation for damp, cold or noise) is intrinsic to the materials used in construction, though a complementary set of indicators measures **the wear and tear and lack of maintenance** (damp, leaks and/or rotting walls, floors, foundations or window frames, etc.) (Streimikiene, 2015; Borg, 2012, 2015). According to Sinha (2017), it is also important to consider

the adequacy of the housing unit according to the needs or desires of the residents. In some cases, it is possible to reform the house in an easy and cheap way or to refurbish and resolve the wear and tear of the unit, although in other cases this could be very expensive or even impossible. Consequently, the potential costs and flexibility to adjust such distribution and restore the wear and tear would be an element of the housing quality (Rohe and Stegman, 1994).

A second group of indicators that reflects the quality of a dwelling is the **design of the unit in terms of basic services, non-essential amenities and functional distribution.** The availability of basic facilities (toilet, shower, heater, storage space, etc.) is broadly analysed as a way to measure housing deprivation for poor families or in studies in poor countries. Furthermore, the available space, basic services and also the functional distribution of the spaces should be considered (Mustapha *et al.,* 1995; Sinha *et al.,* 2017). Furthermore, aspects related to the privacy of residents, accessibility and circulation between various spaces (kitchen, dining room, storage, etc.) or the availability and utility of areas of common use must be included. Some of them are directly related or correlated to the space or size of the housing unit (like the presence of a living room, entrance hall or a large kitchen). While other aspects that can be considered are non-essential amenities or facilities (like a garden, parking lot or a swimming pool).

The different elements of these first two main groups can be measured on a more or less "objective" way, regardless of the level of the resident's satisfaction, although the real value that a resident perceives of such apparently "objective" aspects is also influenced by his/her subjective priorities and judgement (Onibokun, 1975; Sinha *et al.,* 2017), especially important for the third group of characteristics of the dwelling, which is "**aesthetic design, ambience and comfort of the dwelling**". This group of aspects includes, among others, the type of house, its orientation to the sun (light and brightness of the house), the views from the windows and its aesthetic design.

Many authors[11] stress that the quality of a specific housing unit also depends on, **the features and adequacy of the geographic and social environment of the neighbourhood.** The variables that measure the level of satisfaction with the qualities of the neighbourhood can be classified in five groups[12].

- Basic collective public services (transport, education, and health, pedestrian walkways).
- Social-cultural situation (cost of living, ageing structure of the population, security and criminality, social status, cultural integration, etc.).
- Non-basic or secondary physical amenities and facilities (parking arrangements, shopping facilities, public parks, playgrounds, and sports facilities, etc.).
- Urban services for housing (public supply of water and electricity and waste and drainage systems)[13].

[11] Like Nelbach, 1950; Kain et al., 1970; Uehara, 1994; Marr and McCready, 1997; Kutty, 1999; Palvarini et al., 2010; Sinha et al., 2017; Brkanić. 2017

[12] Sinha *et al.,* (2017) include two main groups of location characteristics: the environmental aspects of the neighborhood and the institutional facilities and services of the neighborhood. While Sirgy and Cornwell (2002) base their description of the "neighborhood environment" on physical, social and economic features. In this case, we have combined those two classifications.

[13] This group of aspects (the urban services) is in fact already part of the features that define the quality of the dwelling.

- Pollution and environmental problems (emission of polluted air or stenches, noise, traffic jams and overcrowded streets).

Another main group of variables that affects the quality of the housing unit—as mentioned by Sinha *et al.,* (2017)—is "**The direct costs of housing (affordability)".** This includes the expense on rent for tenants and the cost of a mortgage in the case of owners; the monthly costs of basic services (water supply, heating or waste); the payment or monthly costs of the "community" and taxes. In fact, choosing a dwelling is often a trade-off between almost intrinsic characteristics related to the technical quality and design of the dwelling unit, its aesthetic design and comfort and the features of the geographic and social environment of the neighbourhood. This trade-off is influenced by the level of income of the resident and the cost (affordability) of the housing.

This trade-off between the desired qualities of the housing unit, the neighbourhood and housing costs is highly influenced by personal aspirations and needs. Moreover, as mentioned by Sengupta and Tipple (2007), Zhu and Shelton (1996), and Sinha *et al.,* (2017), a subjective perception of different user groups exists regarding the quality of a house with the same conditions. In this regard, it can be inferred that people from developing versus developed countries will have quite a different opinion about a house with the same "objective conditions". And, within each country, the perceived quality is directly related to the personal opinion of the person living in the dwelling regarding this objective situation (Sinha *et al.,* 2017).

In order to understand better the empirical approach of the few studies that- -as done in this Paper- analyse the lower quality housing of certain (vulnerable) groups of the population on the basis of sex, and racial or ethnic origin we offer a short review of the methodologies and indicators used. The empirical results and conclusions will be discussed in the corresponding sections.

The main aspects of these studies are synthesised in Table 2.1 and 2.2. to get a fast overall overview and optimize the information on the indicators and models used. As can be seen, most empirical studies use a limited number of technical-physical indicators on housing quality. Seven studies used 5 or less, and three of them used 12/13 indicators. Despite the fact that most studies use few variables; a total number of 37 different indicators were detected in the 13 studies reviewed. This reflects the fact that there is a wide heterogeneity in the indicators used. Only 4 variables were used in 5 or more studies, being the overcrowding indicator the most used (8 studies). And the presence of toilet; deterioration of the building and renter/owner status are used by 5 different studies.

Several studies use the variables of the market value as estimated by the owner and being the owner (tenure) of the house or the price paid for the housing as an implicit measure of Housing Quality (like Marr et al, 1997; Spain, 1990). In our opinion these should not be considered as such because it depends on a broad range of determinants like the localisation of the dwelling (periphery-centre) or the public services and infrastructures or the average price of the demand-supply relationship.

Table 2.1.- Empirical evidence on the HQ deprivation of vulnerable population: Objectives of and methods used in the studies.

	Roy et al. (2020)	Castillo et al. (2020)	Borg (2012, 2015)	Navarro et al. (2008)	Schill et al. (1998)	Brandolini et al. (1998)	Uehara (1994)	Mack et al. (1985)	Spain (1990)	Marr et al. (1997)	Kutty (1999)	Palvarini et al. (2010)
Country	Mexico	Mexico	Europe	Spain	United States	Italy	United States	Britain	United States	Canada	United States	4 clusters Europe
Objective of study Overall life quality index that include HQ (OLQ); HQ-index (HQI) HQ deprivation yardstick (HQ-DEP)	HQ-DEP	OLQ	HQ-DEP	HQ-DEP	Disadvan-taged HQ	OLQ	Disadvan-taged HQ	OLQ	HQI	HQI	HQ-DEP	HQ-DEP
Analyse directly or indirectly the discrimination for vulnerable groups	No	No	Immigrants-gender	No	Immigrants - Race - Ethnecity	Gender	Race and gender	Age (children)	Race - Ethnecity	Gender	Race	No
Include (implicitly) a profile of the characteristics the individuals with a higher/lower HQI (Like the level of income or education, type of family)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Spatial analisis	Yes
Dependent variable(s)	Combined continuous variable	Crowding ration and solid con-struction rate	Combined binary variable of deprivation	Multidimen-sional housing deprivation index	Over-crowding Lack of maintenance soundness	Self-evaluation HQ and living area; Crowding ratio and disposal of heating	Low quality of living[14]	Seven variables on HQ as part of OLQ	Crowding ratio	Crowding Ration	Inadequate HQ[15]	Affordability Crowding (space availability) Physical quality Quality of the neighbourhood
Types of dependent variables on HQ Continuous, binary or both	Cont.	Both	Binary	Con.	Binary	Both	Binary	Binary	Cont.	Cont.	Binary	Binary – descriptive tables
Yardstick of deprivation (binary) Continuous indicator HQ-Level	HQ-level	Yardstick	Yardstick	HQ-level	Yardstick	HQ level	Yardstick	Yardstick	HQ-level	HQ-level	Yardstick	HQ-level
Type of indicators on HQ: Combined: Arithmetic (average or sum) or factor analysis (FA) versus Individual indicators	FA	Arith-metic	Arith-metic + individual	Combined	Arith-metic	Arith-metic	Arith-metic	Arith-metic	Individual	Arith-metic	Factor analysis	Arith-metic
Total number of used indicators on housing quality (as dependent or independent variable)	**6**	**12**	**7**	**12**	**5**	**2**	**2**	**7**	**2**	**5**	**7**	**13**

Source: Own elaboration based on the papers included.

[14]The authors consider the living arrangement as "low quality" if the "case manager" observe at least one of the following situation: (1) a Single Room Occupancy unit rented by the day or week; (2) a physically "unsafe" unit; (3) a living unit located in a physically unsafe block; or (4) a short-term or time-limited arrangement.

[15] A dwelling is inadequate if it has at least one of the following features: 1. Incomplete plumbing 2. Three or more heat breakdowns during the previous winter that lasted more than six hours 3. Fuse or breakers blew three or more times in a period of 90 days and wiring is not concealed and every room does not have an electric outlet 4. Three or more problems with the upkeep, such as water leaks, holes in the floor, cracks in wall, broken plaster or paint in an area larger than one square foot, and sign of rats or mice in the past 90 days 5. Three or more problems with hallways and common spaces (lights broken or missing in public hallways, hazardous steps on common stairways, stair railings missing or not firmly attached, or lack of an elevator in buildings with four or more stories) 6. Three or more flush toilet breakdowns of six hours or more in the past 90 days 7. Main type of heat equipment used is a room heater (using oil or gas) without vent 8. Incomplete kitchen facilities.

Table 2.2a. Empirical evidence on the HQ deprivation of vulnerable population: variables used.

Housing quality criteria			Total	Roy et al. (2020)	Castillo et al. (2020)	Borg (2012, 2015)	Navarro et al. (2008)	Schill et al. (1998)	Brandolini et al. (1998)	Uehara (1994)	Mack et al. (1985)	Spain (1990)	Marr et al. (1997)	Kutty (1999)	Palvarini et al. (2010)
		Total number of indicators used	**81**	**6**	**12**	**7**	**12**	**5**	**2**	**2**	**7**	**2**	**5**	**7**	**13**
Spaces of the dwelling	Essential living spaces	Size of the dwelling	1			1									
		Number of persons per room (overcrowding)	8	1	1			1	1			1	1	1	1
		Enough beds	1								1				
	No essential living spaces	Garden	1				1								
		Garage/Parking lot	0												
		Provision of balcony / Place to sit outside	1												1
		Existence of space for children to play	0												
		Separate study room	0												
Equipment and features of the unit dwelling	Basic equipment	Kitchen (indoor/outdoor)	1				1								
		Toilet (indoor/outdoor)	5	1		1	1				1				1
		Type of toilet (flushing, latrine, pit or hole)	3			1	1								1
		Bath/Shower	4			1	1				1				1
		Refrigerator	1								1				
		Washing machine	1								1				
		Heating	4				1		1			1			1
		Incomplete plumbing	1											1	
		Incomplete kitchen facilities	1											1	
	Others	Accessibility (lift, accessibility for disabled)	0												
		Carpets	1								1				

Source: Own elaboration based on the papers included.

Table 2.2b. Empirical evidence on the HQ deprivation of vulnerable population: variables used.

Housing quality criteria			Total	Roy et al. (2020)	Castillo et al. (2020)	Borg (2012, 2015)	Navarro et al. (2008)	Schill et al. (1998)	Brandolini et al. (1998)	Uehara (1994)	Mack et al. (1985)	Spain (1990)	Marr et al. (1997)	Kutty (1999)	Palvarini et al. (2010)
Construction quality	Structural	Type of private home	0												
		Walls quality	0												
		Floor quality	1	1											
		Ceiling quality	0												
		Temporary structures	1	1											
	Physical	Age	2										1	1[a]	
		Overall condition of the building	1		1										
		Wrong functioning of basic facilities	1											1	
		Specific condition of the building (damp, rot in window frames and floor, darkness)	5			1	1	1			1				1
Household Services	Basic	Sewerage	2	1	1										
		Piped and public water	2	1	1										
		Hot running water	3		1		1								1
		Light	2		1		1								
		Telephone line	1		1										
	Secon	Internet	1		1										
		Waste Management	1		1										
Neighbourhood environment	Polluti	Noise	2				1								1
		Pollution	3		1		1								1
	Locati	Distance to work	0												
		Nearness to the neighbourhood facilities	0												
	Social aspects	Crowding	0												
		Social integration with the neighbours	1		1										
		Own perception of the neighbourhood quality	1											1[a]	
		Crime and vandalism	4		1		1			1					1
Tenure, renting and Affordability		Yearly costs of housing / affordability	3					1					1		1
		Percentage of income used for housing	3					1					1		1
		Tenure/renting	5			1		1					1	1[a]	
		Type of tenure	1			1									
		Unstable short term renting contracts	1							1					

Source: Own elaboration based on the papers included.

Regarding housing quality in developed and developing countries, it is important to note that given the high number of housing and precarious settlements that exist in Latin American countries, researchers have defined habitability, in order to address aspects not only related to the absence of infrastructure but also to the housing conditions necessary to good health and psychological aspects of the inhabitants (Hastings, 2011; Mercado (2011). For example, Mercado (2011) stressed that conventionally, the minimum level of quality of the dwellings can be valued through three concepts that reflect some Un-Habitat's parameters: overcrowding (number of people per bedroom), precariousness (dwellings with walls, ceilings and floors without the basic appropriated material attributes) and deterioration (aging of the housing stock that is close to completion or has already exceeded its useful life). As a fourth one the presence of basic services or facilities: drinking water[16], drainage [17] and connection to electricity. While Rindfüss proposed the establishment of housing quality by measuring elements of the structure, such as electricity, the components used for the roof, and indoor plumbing (Rindfuss et al., 2007).

Thus, the history of European and Latin American housing conditions and organization are very different. In the Latin American case, the accretion of informal and squatter settlements often accompanies the growing and pauperized and shrinking neighbourhoods (Soja, 2000). The lack of suitable housing in this part of the world leads to the evaluation of a qualitative deficit that usually refers to the percentage of homes in the country or region that would require major improvements to be at minimum standards for habitability as those that are not connected to the water or sewage system (Saiz, 2022). In addition, informal or illegal housing—those structures that are built without an official permit, perhaps on a tract of land to which the dweller does not have official title—has unfortunately been a necessary resort for the poor to get access to urban shelter in most Latin American countries (Saiz, 2022).

Furthermore, the percentage of households facing shortages of basic infrastructure, in Mexico for example, is 26.7%, considering the most recent data (2020). This situation contrasts with what happens for example in Spain, where more than 99% of the population have primary housing services covered[18], and where the deficit is observed in aspects such as housing density, general accessibility, status of the building and secondary facilities. Therefore, any housing solution or improvement to informal housing in the Latin-American scenario would require master planning approaches at the neighbourhood scale (Saiz, 2022). This means that the development of a housing quality indicator in such a different social and national context is a necessary challenge to make it accessible for the development of housing policies and programs.

[16] Private homes that have piped water inside the home or outside the home but within the land, as well as a public tap or hydrant.

[17] Private homes that have drainage connected from the street or connected to a septic tank, as well as those that drain into the ground, into a river or lake.

[18] The data for Spain and Mexico are based on our own calculation with data from the Spanish census in 2011 and the Mexican census in 2020.

3.2 Measuring Housing Quality: the development of a synthetic indicator.

Having presented the taxonomy of the type of indicators used in the literature, it is worth analysing how they are used in studies. Particularly, how they are combined in a workable set of limited indicators. As mentioned, the measurement of housing quality is a multi-layered and complex activity (Brkanić, 2017; Dinc *et al.,* 2013); nevertheless, the presence of a broad number of diverse indicators is an advantage. Several studies underpin the existence of groups of bundled indicators showing similar though slightly different aspects (for example the presence rotten windows, rotten floors, bad working or broken sanitary amenities). The correlation between them means that their simultaneous inclusion in econometric models generates some major methodological problems, especially in terms of collinearity. Furthermore, not only the slightly different indicators can be highly correlated, but also very different indicators could be correlated (such as the level of overcrowding and building deterioration).

One of the main problems is how to reduce a large number of complementary indicators to a manageable well-defined composite indicator that synthesises the quality of housing without losing information. The papers reviewed employ several ways to do this. One of the most comprehensive ones is the use of a factor analysis that combines variables based on their correlation. Several studies highlight the fact that the bundles of variables for each type of aspect are highly correlated, and this fact makes them especially suitable for a factor analysis such as the one used by Kain and Quigley (1970), Kutty (1999) and Roy *et al.,* (2020). Some define, for each indicator, a minimum quality threshold and if one of them is below that threshold, housing deprivation would exist.

Such a combination implies the creation of a binary variable often used in studies on deprivation (see Brandolini *et al.,* 1998 and Mack and Lansley, 1985), but it means the loss of information that could be used to have a more qualitative perspective. For example, by summing up the number of variables in which deprivation exists or employing another form of pure arithmetic addition, as done by Navarro *et al.,* (2008). Other studies that only use a few indicators did analyse them individually (Módenes *et al.,* 2013; Borg, 2012), seeking whether the level of deprivation differs depending on the indicator used.

However, only a few of these studies actually use housing quality to analyse discrimination on the basis of sex, and racial or ethnic origin. Aspect that would be address throughout this research.

Given that the objective of this paper is the creation of a strategy able to measure the gap in housing quality between people from vulnerable groups of different socioeconomic levels and the rest of the population, a quantitative indicator of housing quality with a continuous range of values is required. Such a quantitative approach would be much more suitable for this purpose rather than using binary methods as has often been done in previous research. Therefore, these binary variables—indicating the existence of certain services, features or equipment—are added into a single variable, adding the number of services for each of the aforementioned groups, reflecting the number of amenities or services available in the dwelling.

The factor analysis technic will be applied to reduce the resulting variables to a few factors, with a synthesis based on the correlation between individual variables. Therefore, each factor would reflect

the same aspect of the dwelling or combine aspects that empirically co-exist even if they are conceptually or theoretically different. The details of this methodological application would be explored in the corresponding Chapter for each country context.

3.3.- Elaboration Housing Quality Index: the Mexican case.

The Population and Housing Census (PHC) of 2020 of the INEGI of Mexico provide 28 variables that reflect the housing quality of the individual (see figure 3.1). It include also broad information on demographic, cultural, economic and social characteristics and identify the individuals by its racial/ethnic "self-identificated" background: Indigenous and Afro-descendant. Moreover, distinguishes whether people who self-identify as indigenous speak an indigenous language or not, which is considered an important factor behind continued social exclusion (Gandelman *et al*., 2011; Torre and Sánchez, 2019). Our study focuses on people between 16 and 75 years old (the potential working age) who live in the MAMC, which includes Mexico City and 59 agglomerated municipalities[19]. The sample consists of 692,444 observations of which 7.9% self-identify as indigenous people and 1.3% as Afro-Mexican. Only 1.4% of the population speak an indigenous language. In practical terms, we will henceforth refer to those who self-identify as indigenous as indigenous non-native speakers, and those who speak an indigenous language as indigenous native speakers.

We classified the 28 variables from the PHC that capture the housing characteristics by six types of aspects (see Figure 3.1.) The first group—**basic services and supply**—includes six variables on the existence of basic services and five on methods of supply. The availability of basic facilities (toilet, shower, heater, storage space, etc.) is broadly analysed as a way to measure housing deprivation for poor families or in studies for developing countries (Boarini and d'Ercole, 2006; Castillo *et al*., 2020).

The second group are two indicators based on three variables that reflect **"habitability"**: number of rooms, number of bedrooms and total number of household members, which are used to measure overcrowding (number of persons per room). We include a variable that somehow reflects the level of privacy based on the number of persons per bedroom. In fact, the overcrowding variable is one of the most universally used indicator in the empirical studies (Palvarini and Pavolini, 2010; Roy *et al*., 2020). A third group is the **"quality of the construction materials"** of the ceiling and the walls, which not only relates to the satisfaction of the immediate user of the unit but also to the wear and tear and the foreseeable cost of maintenance.

We identified three groups of indicators related to housing amenities or facilities, distinguishing between **"basic equipment"** (water, refrigerator and TV); **"secondary services"** (such as Internet and telephone); and **"non-essential equipment"** (such as washing machine and air conditioning).

Especially the last group is important because we aim to compare the level of discrimination against indigenous people and Afro-Mexicans from different socio-economic strata. For the same reason we require a quantitative indicator of HQ with a continuous range of values. Adding the number of services for each of the aforesaid groups, would reflect the number of amenities or services available

[19] One of them located in the state of Hidalgo and the rest in the State of Mexico.

in the dwelling. Such a quantitative approach would be much more suitable for this purpose rather than using binary methods as often done in most of the previous research.

As a methodological remark, it can be stated that despite having information on the quality of flooring materials and the type of dwelling, our preliminary analysis showed that there is no clear and well-defined correlation between these two and the other housing quality variables. In other words, for these two variables a classification from bad to good or high quality cannot be established because most people in the MAMC live in dwellings with concrete floors regardless of the quality of the other materials used for the walls and roof. Furthermore, the different types of dwelling cannot be ranked from worst to best in terms of quality. The same "type" of dwellings also exists in high- and low-income neighbourhoods. Therefore, you may have a beautiful house with all the features and services coexisting with an illegal building, built by the owner but lacking several basic amenities. The lack of a clear relationship between housing quality and floor type or dwelling type could generate noise in the Housing Quality (HQ) values, so they were removed from the initially proposed HQ indicator.

The combination and reclassification of the variables provided lead us to nine indicators to which we applied an Exploratory Factor Analysis (EFA) in order to create a Housing Quality Index (see Table 3.1). In which each factor combines the most correlated variables supposing that they somehow reflect the same housing aspect or combine different conceptual and theoretical aspects that empirically co-exist.

The EFA grouped the variables in three factors[20], which correspond to distinguishable aspects of housing conditions in line with the theoretical and conceptual framework on the subject: (F1) Services, supply and amenities; (F2) Housing density; (F3) Construction materials (see Table 3.1).

The first factor captures five indicators on the number of essential services, the convenience and security of the means by which they are supplied, and the number of secondary services, or amenities. Factor 2 synthesises the two variables that reflect overcrowding, related to the need of the household members to satisfy their physiological needs for privacy and living space. The third factor includes the quality of materials used to build the walls and roof of the dwelling, ranging from essentially disposable materials like mud, cardboard, plastic, trash and palms, to non-disposable materials like bricks. The three factors were considered as partial indices, their factor scores were then used to calculate a solely synthetic HQI based on their simple average. These partial and overall indices—that have a continuous scale—are used to calculate the Housing Quality Gap in the following section.

In terms of the statistical adequacy of the factor analysis, it can be stated that the statistical test[21] achieve the minimum levels required to consider the model adequate (Farrell and Rudd, 2009).

[20] The values of each of the final factor scores are also determined, albeit only in a marginal way, by the variables of the other factors. For our purpose, we wanted independent sub-indices whose values are exclusively defined by the principal variables of each of the factors obtained above. Therefore we applied additional EFA separately for the variables that determine each of the factors.

.

[21] Kaiser–Meyer–Olkin tests the adequacy of the inclusion of each variable in the model and the complete model. It reflects the proportion of variance among variables that might have common variances. You can improve the KMO value by removing items (variables) which have a low factor loading (less than 0.05). The Bartlett test of sphericity

Figure 3.3.1.- Variables of the PHC survey housing chapter and their transformation

Basic services and their supply	Habitability
1. Number of available basic services (one variable) Enjoyment of: Electricity, Piped water, Sewage system, Toilet, Kitchen, Shower. (Minimum 0, maximum 6). **2. Index of the number of basic services and convenience of their supply (one variable)** 1. Sewage system (0 to 3) 2. Piped water supply (1 to 5) 3. Water intake (1 to 3) 4. Place to cook (0 to 3) 5. Use of the toilet (binary) (shared versus individual) 6. Fuel (0 to 3) (Minimum 0, maximum 18)	**3. Bedroom overcrowding** (Continuous variable) $\frac{1}{\left(\frac{Number\ of\ household\ members}{Number\ of\ bedrooms}\right)}$ **4. Dwelling overcrowding** (Continuous variable) $\frac{1}{\left(\frac{Number\ of\ household\ members}{Number\ of\ rooms}\right)}$

Construction materials	Basic equipment
5. Quality of the roof (1 to 5) **6. Quality of the walls (1 to 4)**	**7. Number of items of basic equipment (one variable)** Enjoyment of: Water heater, Refrigerator, TV. (Minimum 0, maximum 3).

Secondary services and common living spaces	Non-essential equipment
8. Number of secondary services (1 variable) Enjoyment of: Telephone, Garbage removal, Internet. (Minimum 0, maximum 3).	**9. Number of items of non-essential equipment** Enjoyment of: Washing machine, Solar water heater, Air conditioning, Solar panels, Energy-saving light bulbs. (Minimum 0, maximum 5).

Source: author with data from the Population and Housing Census (INEGI, 2020).

suggests there is a substantial and sufficient level of correlation between the variables (the correlation matrix is not an identity).

Table 3.3.1. Housing Quality Index: results of the factor analysis

Indicators	Exploratory Factor Analysis Principal Components					
	Factor 1 Services supply and Amenities	Factor 2 Housing density	Factor 3 Construction materials	Factor 1 Services supply and Amenities	Factor 2 Housing density	Factor 3 Construction materials
Number of items of basic equipment	0.832	0.129	0.138	0.862		
Number of items of basic services and infrastructure	0.777	0.152	0.205	0.823		
Number of secondary services	0.662	0.058	0.050	0.661		
Method to supply services	0.643	0.015	-0.026	0.629		
Number of items of non-essential equipment	0.588	0.120	0.150	0.610		
Overcrowding per bedroom	0.111	0.935	0.038		0.943	
Overcrowding per room	0.147	0.928	0.059		0.943	
Quality of the walls	-0.010	0.000	0.891			0.626
Quality of the roof	0.370	0.111	0.635			0.626
Total explained variance (EV)	29.59%	20.10%	14.32%	52.49%	88.97%	63.87%
	63.98%					
KMO-Test	0.739			0.758	0.500	0.758
Bartlett's sphericity test	1929699.743			1007488.946	660278.311	1007488.946
gl.	36			10	1	10
Significant	0			0	0	0

Source: author with data from the Population and Housing Census (INEGI, 2020)

4.- Estimating the Housing Quality Gap: The PSM technique for biased samples.

4.1.- PSM specification: Equations of the models.

The Propensity Score Matching method as a novel means to evaluate racial and ethnic discrimination as a causal factor of worse housing conditions among vulnerable populations. As is well-known in the empirical literature, the characteristics of the vulnerable population in a society—such as immigrants, indigenous people, or minorities in terms of race or religion and their interaction with gender—are very different from the structural characteristics of the majority of population in a given country[22] (Bayona, 2007; Basabe and Bobowik, 2013). Structural social, economic and cultural differences imply that the use of the complete sample of the Mexican population to compare and estimate difference in housing between vulnerable groups and the rest of the population, using the simple average of housing quality (HQ), would not be correct. Mainly, because these differences are partially caused by the disparities between these groups in terms of the cultural and social traditions, habits, customs and preferences. In addition, these disparities are related to differences in income, level of education or health conditions.

Using the differences between the averages for the whole sample would generate what in econometrics is called a selection bias or sample-bias and would not reflect the real level of discrimination of vulnerable groups nor a situation of causality on the gap in HQ. In fact, the sample of the non-vulnerable group as a whole is not representative as a control group and the average differences in HQ would not reflect the real differences caused by their origin or ethnicity, though partially would simply reflect the structural differences between both types of inhabitants.

A frequently employed approach to overcome this so-called selection problem is the Propensity Score Matching (PSM) method. This method allows us to calculate—or rather to estimate—the counterfactual situation of the HQ of a person of a vulnerable group as if he/she would belong to the non-vulnerable population. This method identifies and matches two clones or pairs (a vulnerable person with a non-vulnerable person), both with the exact same characteristics (age, level of studies, type of job, income, etc.) except for the variable of interest (HQ). In other words, the PSM allows us to create an unbiased sample in which both groups only differ in their status of vulnerability and housing quality. Thus, the focus is on inferring causal relationship between being indigenous or Afro-Mexican and having worse housing conditions than the rest of the population.

In fact, we use the PSM to obtain a sample with an exact balance on all the relevant contextual characteristics that differentiate both groups of the population and that explain their HQ regardless of their vulnerability. In this matched sample the average difference in HQ between both individuals of each couple would reflect unbiasedly the Housing Quality Gap (HQG). Concluding, in this thesis the average differences of the HQ of all paired couples would estimate the HQG caused by discrimination

[22] In the case of this study, the indigenous and afro-Mexicans versus the mestizos and white Mexicans.

or traditions and habits as—*ceteris paribus*—the difference in HQ caused by all aspects of ethnicity, race or being an immigrant was isolated.

The PSM was originally developed to unbiasedly estimate the effect of an intervention (such as medical treatment or public employment policies). However, the method is nowadays broadly applied in order to create an unbiased control group for people differentiated by one binary characteristic (race, immigrants or gender: See Pais*,* 2011; Jennings *et al.,* 2014; Stringer and Holland, 2016; Deng *et al.,* 2020). Although this approach is different from the common use of PSM (pairing ‘‘treatment” and ‘‘control” units) the central aim is—following Pais*.,* (2011) —the same. In this way, the PSM is applied to obtain an exact balance on all the relevant contextual characteristics that differentiate vulnerable from non-vulnerable population which could influence their housing quality. The average difference in housing quality between both groups of individuals paired as a couple (a vulnerable person and their non-vulnerable counterpart) unbiasedly reflects the HQG.

When applying the PSM for our purpose, we have two possible states for each individual:

F= 1 if the individual belongs to a vulnerable group, and *F*= 0 otherwise, and the Individual Housing Quality Gap (IHQG) of an individual *i* can be written as:

$$IHQG_i = HQI_{1i} - HQI_{0i} \qquad (1.1)$$

where HQI_{1i} denotes that the individual is a person that belongs to a vulnerable group and HQI_{0i} if the individual does not belong to a vulnerable group.

To evaluate the impact of belonging to a vulnerable group on individuals, we are interested in the estimation of the Average Housing Quality Gap (AHQG)[23] between the two groups:

$$\text{AHQG: } E[IHQI_1\_IHQI_0 \mid F = 1] = E[IHQI_1 \mid F = 1] - E[IHQI_0 \mid F = 0] \qquad (1.2)$$

Calculating the level of discrimination implies a great risk of falling into selection bias since the conditions of vulnerable population are structurally different from those non-vulnerable. Knowing this, we need to estimate the so-called counterfactual, constructing a control sample of non-vulnerable individuals with similar characteristics to those considered vulnerable. Following Rosenbaum and Rubin (1983), the PSM represents a useful alternative, because the application of this non-parametric approach condenses the information of all characteristics (*X*) in just one—the propensity score or estimated likelihood of being foreign—born conditioned by the matrix of *X*.

[23] Those used to handle the theoretical literature of the PSM employ the concept of Average Treatment Effect on the Treated (*ATET*) here, because this method is normally used to evaluate the impact of the policy intervention (treatment) on a certain variable. Here we have adapted this concept to the Average Housing Quality Gap (AHQG).

The Conditional Independence Assumption (CIA) on which the methodology is based stresses that given an adequate set of observable covariates (X), the potential outcomes (housing quality) are independent of the fact of belonging to a vulnerable population (Rubin, 1977).

$$(HQI_{1i}, HQI_{0i}) \perp F \mid X \qquad (1.3)$$

If the CIA holds, the estimated housing quality gap at an individual level - $\hat{\tau}_i$ - can be obtained by substituting the non-observed IHQ_{0i} for the outcome of a vulnerable individual with that of a nonvulnerable with a similar propensity score ($\widehat{IHQI_{0i}}$):

$$\widehat{IHQG_i} = IHQI_{1i} - \widehat{IHQI_{0i}} \qquad (1.4)$$

Consequently, we can compute the *AHQG* as the mean of estimated individual treatment effects:

$$\hat{A}HQG = \frac{1}{N}\sum_{i=1}^{N} \widehat{IHQG_i} \qquad (1.5)$$

This last equation is used to estimate the AHQG based on the unweighted means of the individual differences in the housing quality indicator for each matched couple.

Regarding the interpretation of the HQG, it must be pointed out that the differences in housing conditions of vulnerable versus non-vulnerable population reflect a combination of several aspects of affordability, discrimination and personal decisions. However, indirectly, the differences in the housing quality of people with similar level of studies, responsibilities by type of job and age also reflect the discrimination in economic opportunities (based, among others on their salaries and in terms of over-qualification) to a certain extent.

The PSM method tries to define a kind of *ceteris paribus* situation, comparing housing quality while maintaining everything else equal except the status of vulnerability. The observations for both groups are matched when they have the same Propensity Score and each couple can only be considered as exact clones if all the relevant information is included; therefore, the selection of the correct variables is essential (Rosenbaum and Rubin, 1983). Moreover, although an equal PS should reflect a perfect clone, an exact match is recommended to control some very important variables that could affect the preferences, capability (affordability) and limitations for certain types of dwelling or their level of discrimination. In other words, the observations are not only matched because of having the same propensity score, but they are also exactly equal in some of the very relevant characteristics.

In that sense, as will be further seen, we require an exact match for three characteristics of each individual (potential income dummies, sex and age) and for two aspects of their geographic location (living in the same region and in a municipality of a similar size). However, the inclusion of a very large number of often similar or highly correlated variables makes it difficult to find exact matches.

This fact will be explained in the chapters where the models, for which a large set of variables are used, are developed.

Several variables reflect the socio-economic situation of the persons: the level of education, the potential income of the observed individuals or on family level and the fact of being the owner or tenant of the house. Another set of variables reflect the configuration and size of the family nucleus. In addition to the aforementioned characteristics of a person (sex and age), we also include some variables on residential location (central versus more peripheral region or neighbourhood and the size of municipality).

Finally, once the PSM model was estimated its suitability was tested to ensure that the unbiasedness of the sample obtained is satisfactory. Therefore, we use standard tests such as the t-test for the means of each of the variables used for averages between the group of vulnerable persons and the control group. If the model fits well all the means between both groups should be statistically non-significant. Another control test is the estimation of the PROBIT models using only the new reduced samples—using as the dependent variable the binary variable of each vulnerable versus non-vulnerable group. In this case none of the variables should have any predictable power to correctly classify both groups in the reduced sample to reflect their unbiasedness. The receiver operating characteristic curve (ROC) was also analysed, providing a statistical method for evaluating the diagnostic accuracy of these tests, being used to determine the cut-off value with the highest sensitivity and specificity and assess the discriminative capacity of the model, that is, their ability to differentiate individuals from a vulnerable versus non-vulnerable population.

4.2.- PSM specification: Selection of matching variables and adjustments of the models

As just discussed the PSM process all the individuals' relevant information should be included to assure that those with the same Propensity Score (PS) can be considered as exact clones. Therefore, the selection of the right set of variables for the matching procedure is essential (Rosenbaum and Rubin, 1983) and should include all the characteristics of the people that influence their preferences in housing and their capability (affordability) to access a certain quality level. Therefore, a broad set of variables were used not just in order to calculate the PS. Those couples were not only match using the native control observation with the same PS but it was also requested that they were, regardless of their PS, equal for some of the very relevant characteristics. The foregoing, forcing an exact match for several characteristics of each individual (income, sex, age and being owner versus renter of the dwelling). Moreover also we applied an E-match for the geographic location: living in the same region and in a municipality with a similar size. The simultaneous inclusion implies that somehow we measure if they live in the centre versus the more peripheral locations.

As shown in Figure 4.2.1, our models for Mexico include 10 variables to characterise the differences between indigenous people and Afro-Mexicans. Four variables reflect their socio-economic situation: level of education, employment benefits, retirement income, household income and their status as owner or renter. One variable reflects the type of cohabitation at home; two indicate the structural

demographic characteristics of a person (sex and age); and two variables indicate the geographic location (urban area and town size).

Table 2.4.1 Variables used in the PSM: The Mexican case

Outcome variables	Global Index of Housing Quality Synthetic composite variable
Sub-samples by vulnerable group	*Self-identified indigenous people *Indigenous people that speak an indigenous language *Afro-Mexican
Individual characteristics	
Sex (PSM and E-match)	Dummy variable, 1 if Man, 0 if woman
Age (PSM and E-match)	1: between 16 and 25 years old 2: between 26 and 35 years old 3: between 36 and 55 years old 4: between 56 and 75 years old
Level of studies completed (PSM)	1: No education 2: Basic education 3: Baccalaureate or medium grade or technical education 4: Higher education degree 5: Post higher education 6: Not specified
Retirement income (PSM and E-match)	Dummy variable, 1 yes, 0 no
Number of employment benefits (PSM)	Sum of seven dummy variables (indicating 0 to 7 benefits) Christmas bonus, Annual leave, Medical service, Utilities, Paid disability, Retirement savings, Housing loan
Home structure	
Type of cohabitation (Structure of the household) (PSM)	1: Couples without children 2: Couples with children 3: Single parents 4: Single
Household monthly income (HMI) (American dollars - PSM and E-match)	Using the 2020 average exchange rate of 21.64 Mexican pesos per American dollar 1: From 0 to 171 dollars 2: From 172 to 245 dollars 3: From 246 to 337 dollars 4: From 338 to 536 dollars 5: From 537 to 832 dollars 6: From 833 to 1,086 dollars 7: From 1,087 to 1,617 dollars 8: More than 1,617 dollars
Owner/Renter (PSM and E-match)	Binary variable differentiating between 0: Renters or other situations 1: Home-owners
Geographical location of the vulnerable groups	
Contours (binary) (PSM and E-match)	City centre and 4 contours
Town Size (By number of inhabitants) (PSM and E-match)	1: Below 2,500 2: Between 2,500 and 14,999 3: Between 15,000 and 49,999 4: Between 50,000 and 99,999 5: Over 100,000

Source: author with data from the Population and Housing Census (INEGI, 2020).

The most important characteristic that, *ceteris paribus*, determines the housing quality is family income, because people with a higher income can clearly choose a better dwelling (Schill *et al*., 1998). The **level of income** is not only important in financial terms, but the occupational status of the current job is perceived as a factor that has an important effect on socio-cultural integration (Mazza, 2010).

To analyse this aspect, we separated the samples into eight levels or intervals (see table 4.1)[24] of the total household income from work from all the inhabitants of the dwelling. We focused on people of a potential working age (16-75 years old), being the effective labour market retirement in Mexico (75 years old)[25]. For a large group of people the effective age to retire is clearly lower, therefore, we included the binary variable on receiving a pension or retirement income, or not, and use it as an exact match. Considering the importance of income as an explanatory variable of the HQ, an exact match was applied for this variable ensuring that the native clone does not differ in this specific aspect.

Some studies (Evans *et al.,* 2000; Coley and Leventhal, 2012; Viljoen *et al*., 2020) stress that women have different opinions and preferences when they select their type of housing, especially when they have children. Gender roles influence integration in the ideal labour market, access to resources and housing affordability. In fact, Mexican national gender roles cause gender-specific disadvantages; the result of cultural norms and practices may be shared by all Mexicans but are operationalised differently across ethnicity, region and class (Newdick, 2005). For all these reasons, we introduced the variable sex as an exact match.

In traditional housing studies, economic and demographic factors, especially the lifecycle of the individuals, are referred to as the most important determinants of residential choice (Painter *et al.,* 2001; Grigolon *et al*., 2014). The search for a dwelling is often a trade-off between affordability (the level of financial resources) and individual preferences or needs, based, among others, on the size of the family and number and age of the children (Schill *et al*., 1998; Karsten, 2010) and each of these aspects—as will be discussed below—is closely related with the lifecycle of a person. In other words, housing preferences and consumption changes throughout the life cycle, due to changing household needs related with the ageing process (Clark and Onaka, 1983; Bayoh and Haab, 2006). The life-cycle is correlated with having a partner or having children and their age, supposing that more and older children mean higher living costs. It could also mean obtaining better paid jobs due to work experience, thus overcoming the lack of formal education. Therefor another important variable used to calculate the Propensity Score (PS) is the **age of the individual**. Its inclusion denotes several implicit adjustments **in terms of the lifecycle of a person**, as well as social aspects and their labour experience. In relation to this last aspect, a person's income often increases—even within the same job—as they get older and/or the number of years worked (experience gained). Moreover, older people may have a better dwelling not only because they earn higher salaries within the same occupation but also because they could have accumulated savings in the past or inherited or received continuous support from their parents or even their adult children. On the other hand, young people often accept a lower housing quality for very dissimilar reasons: their greater flexibility as regards

[24] The definition of income intervals is based on the Decision Tree Technique. Our data present the income variable in pesos; here we have used dollars, applying the 2020 exchange rate in Mexican pesos per American dollars.

[25]Following the statistics of the "Effective age of labour market exit" a significant proportion of older adults remain in the labour market even after reaching the legal age to retire from work (65 years old) (CEPAL/OIT, 2018).

overcrowding; to save money to send part of it to their family abroad; to buy a house in the near future; to spend more on consumption or to purchase necessary durable consumer goods. Finally, the age—or the lifecycle—is correlated to the cost of having children, not only because of the possibility of having them, but also because of their age. At any rate, such costs have a non-linear relationship, being lower when children are very young, increasing when they are around 11 to 21 years old and decreasing or completely disappearing when they start working. An exact match was applied for this variable, to ensure that the mestizos and white Mexicans native clone does not differ in certain aspects related to the lifecycle of a human not included in the survey.

As a fifth indicator, considered for an exact match, we included the status as **home-owner versus renter.** In the broader literature this has been the most frequently used indicator to highlight differences in the degree of security, wealth and overall quality related to the housing experience (Borg, 2015). Particularly if the owner lives in the house, whereby upgrading the condition of the dwelling or simply refurbishing due to wear and tear is an investment from which they benefit. While such upgrading and maintenance are costs for a renter, and its benefits are for the renter. Moreover, this defines the formal position of residents in their capacity as owner and user of their dwelling (Borg, 2015).

We required an exact match for the variables that contemplate the individual's **residential location** considering **the size of the locality** and its "contour", differentiating between the city centre and the different more outlying municipalities, each of them with its own characteristics. We consider that the five contours in which the metropolitan area has been divided capture a broad part of the dynamics of the MAMC. The city centre (Contour 0) attracts, due to higher costs of living, population segments with a higher level of education and higher income, followed by contours 2 and 3. While contours 3 and 4, which contain the most recent and more outlying municipalities, tend to concentrate the population with the lowest level of education and income. These are contours with less accessibility to employment centres, education and other services than the population of the city centre and first two contours (Toscana and Pimienta, 2018).

In fact, the size of the municipalities accounts for geographic and rural/urban differences, among others, that could, together with the contours, potentially explain crime levels along with the quality and accessibility of the public infrastructure. However, even when poor people live in urban zones with services that are unavailable in rural localities, their access to these resources is socially determined as they lack the funds to pay for such goods as education, health services and transportation and suffer from social exclusion (Gutiérrez and Atienzo, 2011). It is important to note that several of the aforesaid variables are interrelated. For example, age is directly related to the level of studies (very young people have not yet finished their studies); the level of income often increases as people get older due to their proven work experience (years worked); and age is also correlated with having children or not. This reinforces the quality of the PSM technique because those with the same Propensity Score will also be more similar in the non-observed aspects that are highly correlated with specific settings or configurations of the variables used.

Regarding the rest of the variables included in the PSM, the first variable is the **level of studies** completed, as this provides information on the social and cultural capital of the human beings. Frequently, a higher educational level implies a broader cultural view or capital than those persons with a lower level of education, and this affects preferences regarding where to live (Toscana and Pimienta, 2018) and certainly influences preferences regarding certain aspects of the dwelling. In Mexico, even within highly marginalised rural areas, indigenous children have, on average, clearly worse educational outcomes (years of schooling and educational performance) than non-indigenous children. Such disparities affect adulthood occupational pathways, as low levels of education prevent indigenous people from entering higher-paid skilled jobs (Canedo, 2019).

Another aspect determining housing affordability is the intrinsic cost of living that depends on the **size and structure of the household**, especially the number of people relying on their household income. Therefore, we included four types of cohabitation, distinguishing between couples with and without children, single parents and single households ((bachelors, widows or others). Each form of household structure implies a different panorama of costs and socio-cultural requirements for the dwelling. This type of cohabitation is clearly related to the life cycle (age) of the individuals. Both indicators—together with the level of education—reflect, to some extent, the non-observed social or cultural preferences. Additionally, we include an aspect that reflects the specific working conditions of each observation: the number of additional employment benefits (paid annual leave, Christmas bonus, etc.) to capture the level of job security of the household members.

It is important to note that several of the aforesaid variables are interrelated. For example, age is directly related to the level of studies (very young people have not yet finished their studies); the level of income often increases as people get older due to their proven work experience (years worked); and age is also correlated with having children or not. The use of the E-match of six out of ten variables reinforces the quality of the PSM technique because those with the same Propensity Score will also be more similar in the non-observed aspects that are highly correlated with specific settings or configurations of the variables used.

We should also highlight the specific parameters of the models and the way we treated the methodological requirement—not mentioned before—of the so-called Consistency Hypothesis. The right PSM models require that all the units treated receive the same treatment, which generates the same potential effect for each observation [26] (Imbens and Rubin, 2015: 10; Keele, 2015b: 5). Therefore, we repeat the PSM models separately for the three different groups. Firstly, we estimate a specific model for the Afro-Mexican population. Moreover, we also divided the indigenous population into two groups: those that speak their native indigenous language[27] (indigenous native speakers)—often the part of the population with a lower economic level and more marginalised—and those that do not speak their native language (indigenous non-native speakers). This is important as indigenous native speakers perform at a higher economic level almost equivalent to their non-

[26] A requirement being part of the Stable Unit Treatment Assumption SUTVA.
[27] Primarily Otomí, Mixteco, Zapoteco, Mazateco and Mazahua (Domínguez and Rodríguez, 2017)

indigenous counterparts in Mexico (Gandelman *et al*., 2011). In fact, our sample shows that the average salaries of the three groups are clearly different: the highest salaries for Afro-Mexicans and the lowest for indigenous native speakers. Therefore, we estimate the AHQG for each group of those traditionally excluded or vulnerable populations, also considering, at least implicitly, the ethnic differences, language and other aspects related to the community form of organisation, requirements and housing expectations.

Two last methodological details of the model that we implemented are the matching procedure without replacement, so that one mestizo or white Mexican cannot be matched more than once, and we applied a maximum distance (calliper) of the PS to ensure more similarity between the coupled pairs of observations.

5.- Estimating the Average Housing Quality Gap (AHQG): the results of the PSM

Once we detailed our housing quality index, the specification of the PSM model and the variables used in the matching process we dispose of all aspects required to estimate the average Housing Quality Gap. Being unbiased average difference in the housing quality of each pair of matched observations. In other words, the PSM method calculate for each individual of the vulnerable group an individual gap compared to his non-vulnerable match or counterfactual. And the simple average of all individuals would reflect *Average Housing Quality Gap (AHQG)*[28] *attributable to their status of vulnerability.* Where a negative HQG average value for a certain traditionally excluded group implies that this ethnic minority has dwellings with worse housing conditions, while a positive score implies, in a broad sense, having a better dwelling than their mestizo or white Mexican counterparts.

5.1.- Average Housing Quality Gap (AHQG): Mexico

In the case of Mexico we calculated, as just mentioned, the gap for three different groups: indigenous non-native speakers, indigenous native speakers and Afro-Mexicans. And negative HQG average value for a certain traditionally excluded group implies that this ethnic minority has dwellings with worse housing conditions, while a positive score implies, in a broad sense, having a better dwelling than their mestizo or white Mexican counterparts. Our results (see Table 4.1) suggest that the indigenous population in MAMC live in worse housing conditions than the rest of the Mexican population. Moreover, indigenous native speakers suffer an even larger negative gap than indigenous non-native speakers, meaning that they have a higher probability of living in worse housing conditions than the rest of the population. A fact observed for each of the three partial indexes.

Surprisingly, for Afro-Mexicans, we did not find statistically significant differences. Certainly, despite the documented historical discrimination faced by Afro-descendants (see Losilla and de Indias, 2020), we found no evidence that Afro-Mexicans in MAMC suffer from worse housing conditions than mestizos or white Mexicans. Such results coincide with what was set out by Torre and Sánchez (2019). However, observing the AHQG bus the partial indexes a negative gap was found in terms of number of services and amenities while regarding to overcrowding a positive gap seems to exist. In other words, the divergent results of the literature sometimes could be explained by the used indicators, although also by the different social-cultural settings of particular countries.

Looking in detail to the partial indexes an interesting conclusion is that the larger negative HQG for each of the three groups is basically caused by the partial index that represents the number of amenities of the dwelling. This sub-index reflects a much larger HQG than the index of the quality of the construction materials. As mentioned, the overcrowding gap is even positive for Afro-Mexicans, reducing, on average, the size of the overall gap. In conclusion, the three vulnerable groups, compared with mestizos or white Mexicans, clearly have worse basic services, inferior forms of supply and

[28] In the statistical programs like STATA this effect is known as the Average Treatment Effect on the Treated (ATET)

fewer secondary services and non-essential equipment. In fact, for Afro-Mexicans this is the only partial index that is negatively statistically significant. For this population, no overall gap exists because the negative gap by number of services is neutralised by the positive gap for overcrowding. Hence the contradictory results found for this part of the population may depend on the indicator used.

One of the reasons to use the PSM and to create a continuous indicator of housing quality—instead of establishing a minimum yardstick of housing deprivation[29]—was to analyse whether this gap exists for people on different income levels. We not only analyse whether the poorest people with indigenous and Afro-Mexican roots have worse living conditions than the poorest mestizos or white Mexicans. Instead, we analyse whether people in these vulnerable groups that reached a high level of income suffer such gaps. Therefore, we repeated the PSM models for separated samples into five levels[30] of total household income from work[31]. The results, show that the negative AHQG subsists for all five income levels in the case of each indigenous group, although again the gap in each model is more pronounced for indigenous native speakers than for non-native speakers. This was particularly the case for low-, medium- and high-income levels, being around 4-5 times higher. While for the lowest and the highest income levels, the gap in the differences is smaller; for the lowest it is twice as high and for the highest incomes it is over 50% higher.

The main conclusion stemming from these results is that even those on a high or a very high level of income still suffer from worse living conditions than mestizos and white Mexican clones. For the non-native speaking indigenous populations, the gap seems to be higher for those on "medium"-level incomes, a bit lower for the poorest people and the lowest for the two intervals of the highest incomes. For indigenous native speakers, the panorama is different. The gap is larger for persons from the three sub-samples on low, medium and high incomes, the highest for the poorest sector while a clearly smaller gap exists for the interval of those on a very high income.

In the case of indigenous non-native speakers, we identified a wider negative HQG for the sub-sample on the lowest levels of income (less than $245). For the other four intervals, the size of the gap is nearly half that of the poorest people. In contrast, for indigenous native speakers, the higher gaps are found in the low-, medium-, and high-income levels (between $246 and 1,617). While the gap for the poorest sector is a bit lower and for the wealthiest people it is very low - although it still exists. In the case of self-identified Afro-Mexicans, we do not find any significant positive or negative gap at any income level.

[29] Only useful for the poorest echelon.

[30] The definition of the income intervals is based on the Decision Tree Technique. Our data present the income variable in pesos; here we used dollars applying the 2020 exchange rate for Mexican pesos per American dollar.

[31] The intervals used and the distribution of the sample by those intervals is presented in Table 4.2

Table 5.1. Average Housing Quality Gap (AHQG) by vulnerable ethnic-race group: The case of Mexico

	Average Housing Quality Gap (AHQG)		Services, supply and amenities of the dwelling (API1)		Housing density (API2)		Construction materials (API3)		Details on the sample and the requirements of the caliper			
Groups treated	AHQG	S.E.	API1	S.E.	API2	S.E.	API3	S.E.	Observations			Comm Cal(.)
									Matched	Not matched	Total	
Indigenous non-native speakers	-0.068***	0.005	-0.124***	0.007	-0.032***	0.006	-0.048***	0.007	53,872	3,668	57,540	0.0002
Indigenous native speakers	-0.228***	0.011	-0.443***	0.016	-0.126***	0.013	-0.115***	0.019	10,027	759	10,786	0.0001
Afro-Mexican	-0.007	0.010	-0.044***	0.014	0.035**	0.015	-0.012	0.015	9,441	204	9,645	0.0002

Source: author with data from the Population and Housing Census (INEGI, 2020). Notes: ***: p-value<0.01, **: p-value<0.05, *: p-value<0.1. NNM(n) = Nearest Neighbour Matching with one individual. Comm = common support. Cal(.) = the maximum distance allowed between treated and control individuals.

The test for adequacy of our PSM models and the unbiased results of the samples obtained are satisfactory[32]. The t-test for means of each of the variables used for averages between traditionally excluded groups versus mestizos or white Mexicans does not show any statistically significant differences. The PROBIT models for the new reduced samples—using the binary variable of each vulnerable group versus mestizos or white Mexican as the dependent variable show that none of the 35 variables have any predictable power to correctly classify both groups in our reduced sample reflecting its unbiased nature.

[32] The results of all the tests mentioned will be available online in the journal and by request from the authors. PROBIT models of the PSM; marginal effects; test of means, all of them before and after matching (these annexes are available for the revisors in a separate file).

Table 5.2. Average Housing Quality Gap (AHQG) by five intervals of the monthly total income from work by household: The case of Mexico

	Average Housing Quality Gap (AHQG)		Details on the sample and the requirements of the caliper			
Groups treated	AHQG	S.E.	Observations			Comm Cal(.)
			Matched	Not matched	%	
Indigenous non-native speakers	-0.068***	0.005	55,976	1,564	100	0.0002
Very low income	-0.102***	0.010	15,101	312	26.8	0.0009
Low income	-0.059***	0.008	17,435	410	31.0	0.0009
Medium income	-0.053***	0.007	16,058	373	28.6	0.0009
High income	-0.037***	0.012	4,344	265	8.0	0.0009
Very high income	-0.045***	0.013	3,038	204	5.6	0.0009
Indigenous native speakers	**-0.228***	0.011	10,371	415	100	0.0001
Very low income	-0.195***	0.024	2,809	83	26.8	0.0009
Low income	-0.250***	0.018	3,700	123	35.4	0.0009
Medium income	-0.259***	0.020	2,798	97	26.8	0.0009
High income	-0.256***	0.036	659	67	6.7	0.0009
Very high income	-0.059	0.040	405	45	4.2	0.0009
Afro-Mexican	-0.007	0.010	9,544	101	100	0.0002
Very low income	-0.004	0.026	2,130	16	22.2	0.0009
Low income	-0.001	0.021	2,615	43	27.6	0.0001
Medium income	-0.011	0.017	2,876	20	30.0	0.0009
High income	-0.018	0.023	1,054	12	11.1	0.0009
Very high income	-0.017	0.023	869	10	9.1	0.0009

Source: author with data from the Population and Housing Census (INEGI, 2020). Notes: ***: p-value<0.01, **: p-value<0.05, *: p-value<0.1. NNM(n) = Nearest Neighbour Matching with one individual. Comm = common support. Cal(.) = the maximum distance allowed between treated and control individuals.

Table 5.3.- Test of average difference between the AHQG; The case of Mexico

	Very low income			Low income			Medium income		
Vulnerable groups	Indigenous non-native speakers	Indigenous native speakers	Afro-Mexicans	Indigenous non-native speakers	Indigenous native speakers	Afro-Mexicans	Indigenous non-native speakers	Indigenous native speakers	Afro-Mexicans
Indigenous non-native speakers		Larger	Smaller		Larger	Smaller		Larger	Smaller
Indigenous native speakers	Smaller		Smaller	N.S.		Smaller	Smaller		Smaller
Afro-Mexicans	N.S.	Larger		N.S.	N.S.		Larger	Larger	
	High- Income			Upper High-Income			Entire Sample		

Source: author with data from the Population and Housing Census (INEGI, 2020). Note: The results reflect the t-test of means where Larger HQG or Smaller HQG implies that the vulnerable group mentioned in the columns has a statistically significant Smaller HQG or Larger HQG than the group indicated in the rows. N.S means that the difference in the HQG is statistically non-significant.

The test that should confirm the adequacy of our PSM models and the unbiased results of the samples obtained are satisfactory[33]. A first measure is the t-test for means of each of the variables of the PSM model applied for averages between traditionally excluded groups versus mestizos or white Mexicans in the reduced unbiased sample. Once we only include the matched couples none of the means between vulnerable versus non vulnerable groups should not show -as it is in our results- any statistically significant differences. Another control test is the estimation of the PROBIT models, applied again for only the new reduced samples. Using the binary variable of each vulnerable group versus mestizos or white Mexican as the dependent variable the PROBOT models show that none of the 41 variables have any predictable power to correctly classify both groups in our reduced sample confirming its unbiased nature.A last test -that also is bases on the predictive capacity- is the Receiver Operating Characteristic Curve (ROC) that provide the diagnostic accuracy of these tests assessing the discriminative capacity of the model before and after matching. After matching, the ROC value indicates that it is not possible to differentiate individuals from a vulnerable versus non-vulnerable population.

An important discussion would be the interpretation of the differences in housing detected by the PSM method. It is without doubt very difficult to define which part of the HQG is caused by direct discrimination of the vulnerable groups in the housing market, which part is related with discrimination in terms of labour opportunities and salary and what part is the consequences of personal decisions based on social-cultural values and preferences.

The differences in housing conditions of immigrants versus non-immigrants reflect a combination of several aspects of affordability, discrimination and personal decisions. As discussed choosing a certain level of HQ and/or personal satisfaction (subjective perception) often depends on household characteristics, such as the age of the residents, the number of persons in the dwelling and their economic situation. Besides, it can be inferred that poor people versus rich ones will have quite a different opinion regarding a house with the same “objective conditions”. Such a fact cannot be ignored when comparing the housing conditions of vulnerable versus non vulnerable persons. The observed HQ gap for indigenous people could be related to the acceptance of worse housing conditions due to socio-cultural differences in combination with their personal decision to spend less money in order to support their family, to save money for the future (pension funds or saving to buy a house or durable goods in the future) or simply to maintain a higher immediate level of consumption.

Another part of the interpretation is that the differences in the HQ of people with a similar level of studies, responsibilities by type of job and age, somehow reflect the discrimination in economic opportunities (based, among others, on their salary and high-qualification). In particular, the unstable job for most indigenous often makes the access to better housing very difficult, tenants often ask for higher rents to compensate possible losses if they lose their job. However, without any doubt, the HQG partially reflects the social aspect of discrimination.

[33] The results of all the mentioned tests can be consulted in Annexes IV, V and VI. PROBIT models of the PSM; marginal effects; test of means, ROC curve, all of them before and after matching.

6.- The profile of the vulnerable persons in Mexico most affected by a HQG

Once the way in which the housing quality index is created and the definition and basic specification of the PSM model used have been defined, it is explained how the profile of those vulnerable persons who suffer worse living conditions than the non-vulnerable clone or counterfactual was created. The characterisation of the type of persons from different vulnerable groups with a wider HQG is the main novelty of this thesis within the empirical literature. This profile considers variables that delve into aspects that may intersect: race, gender, class, disability, family structure, place of residence, age (lifecycle) and other axes of inequality that may constitute systems of disadvantage. Such systems produce very different life experiences that can either enhance or challenge adulthood development pathways and the ability to ensure good housing access.

As mentioned, we estimate a normal regression model (with bootstrapping)[34] using the estimated Individual Housing Quality Gaps (IHQGs) as the dependent variable estimating the following equation:

$$\widehat{IHQG} = \beta_0 + \beta_j Z_i + \varepsilon \qquad (1.6)$$

Where $\widehat{IHQG}$ reflects the estimated individual HQG and Z represents the matrix of a broad set of independent variables reflecting different aspects in terms of socio-demographic and socio-economic characteristics of the people, the structure of the families and their origins and geographic localization and town size. It can be stressed that the estimation of the model for this third stage is based—as logical—on the sample that only includes the vulnerable population.

As will be observed in the following sections, the models offer a broad profile and provide new information about the vulnerable groups of interest stressing that some members with specific conditions or characteristics have a higher negative HQG. In order to facilitate the interpretation, it can be stated that a negative regression coefficients or relationship implies a wider HQ gap and positive value a lower gap.

For the Mexican case, a global model and five models for sub-samples by income levels will be estimated to see if the profile of the most affected vulnerable groups differs by income level. For most of the variables, the global models show the same or very similar results. Therefore, the global model will be used as basis to present the results and the other models will only be mentioned if they tinge them. Moreover, the results are presented by groups of variables; the basic determinants; the social aspects of individuals; the existence of additional sources of income (public subsidies, income transfers and pension payments) and the aspects of the geographical location and type of the dwelling.

[34] The bootstrap method is applied because the IHQG is an estimation and therefore requires a more demanding method to assure the correct results (Beck et al., 2016).

Table 6.1a.- A regression model of the effect of specific characteristics of vulnerable groups on their HQG: Intrinsic individual characteristics.

Variables	(Very low income)	(Low income)	(Medium income)	(High income)	(Very high – income)	Global
Ln_Household monthly income	-0.012	0.168***	0.175***	0.303***	-0.009	0.060***
	(0.009)	(0.028)	(0.027)	(0.074)	(0.016)	(0.004)
Age						
Between 16 and 25 years old	0.121***	0.087***	0.061***	0.062**	0.019	0.068***
	(0.030)	(0.019)	(0.017)	(0.029)	(0.030)	(0.011)
Between 26 and 35 years old	0.049*	0.034**	0.041***	0.055**	0.045**	0.033***
	(0.025)	(0.016)	(0.015)	(0.024)	(0.023)	(0.009)
Between 36 and 55 years old (reference)						
Between 56 and 75 years old	-0.127***	-0.016	0.014	-0.021	-0.092***	-0.053***
	(0.025)	(0.017)	(0.016)	(0.027)	(0.026)	(0.010)
Level of studies completed						
No education	0.200***	0.143***	0.207***	-0.029	0.081	0.188***
	(0.051)	(0.039)	(0.045)	(0.113)	(0.109)	(0.026)
Basic education	0.098***	0.091***	0.099***	0.135***	0.072***	0.098***
	(0.022)	(0.014)	(0.013)	(0.022)	(0.025)	(0.008)
High School/Technical studies (reference)						
Higher education degree	-0.041	-0.046**	-0.039***	-0.074***	-0.142***	-0.063***
	(0.031)	(0.018)	(0.014)	(0.020)	(0.021)	(0.009)
Post higher education	-0.232*	-0.098	-0.196***	-0.217***	-0.171***	-0.159***
	(0.133)	(0.070)	(0.048)	(0.053)	(0.039)	(0.027)
Not specified	0.066	-0.329**	-0.045	0.487	0.684***	0.041
	(0.193)	(0.166)	(0.141)	(0.315)	(0.087)	(0.096)
Sex	-0.003	-0.008	-0.003	0.001	-0.014	-0.007
	(0.020)	(0.013)	(0.012)	(0.019)	(0.019)	(0.007)
House rental (versus owner)	-0.058***	-0.066***	-0.037***	-0.098***	-0.123***	-0.066***
	(0.021)	(0.014)	(0.014)	(0.022)	(0.023)	(0.008)
Vulnerable group						
Self-identified indigenous (reference)						
Speak an indigenous language	-0.075***	-0.145***	-0.154***	-0.106***	0.007	-0.118***
	(0.026)	(0.017)	(0.019)	(0.032)	(0.033)	(0.011)
Afro-Mexican	0.057**	0.033*	0.011	-0.024	-0.000	0.033***
	(0.028)	(0.018)	(0.015)	(0.023)	(0.021)	(0.010)
Type of cohabitation						
Couples without children	0.006	0.038*	0.014	-0.007	-0.002	0.021*
	(0.030)	(0.020)	(0.018)	(0.028)	(0.027)	(0.011)
Couples with children (reference)						
Single parents	-0.053	-0.054	-0.003	0.066	-0.012	-0.027
	(0.050)	(0.040)	(0.038)	(0.056)	(0.060)	(0.023)
Single	-0.171***	-0.050***	-0.036***	-0.037	-0.035	-0.060***
	(0.023)	(0.015)	(0.014)	(0.023)	(0.022)	(0.008)
Number of members in the household						
Live alone	0.507***	0.550***	0.650***	0.749***	0.695***	0.516***
	(0.046)	(0.040)	(0.054)	(0.089)	(0.146)	(0.027)
Two people	0.208***	0.204***	0.220***	0.180***	0.215***	0.198***
	(0.024)	(0.018)	(0.021)	(0.033)	(0.033)	(0.011)
3 to 4 people (reference)						
5 to 10 people	-0.152***	-0.174***	-0.183***	-0.230***	-0.229***	-0.167***
	(0.029)	(0.016)	(0.013)	(0.021)	(0.022)	(0.008)
Over 10 people	0.236	-0.480***	-0.370***	-0.466***	-0.423***	-0.363***
	(0.167)	(0.107)	(0.038)	(0.043)	(0.034)	(0.021)
Some disability	0.002	-0.042*	-0.028	-0.081	0.016	-0.032**
	(0.033)	(0.024)	(0.026)	(0.052)	(0.045)	(0.015)
Constant	0.291***	-1.296***	-1.457***	-2.905***	0.421**	-0.331***
	(0.081)	(0.250)	(0.265)	(0.756)	(0.187)	(0.041)
Observations	13,056	23,445	21,265	5,900	4,197	66,392
R-squared	0.113	0.105	0.141	0.125	0.156	0.112

Note: Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1 Source: author with data from the Population and Housing Census (INE, 2011). Standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

Table 6.1b.- Regression model of the effect of specific characteristics of vulnerable groups on their HQG: housing loan, monetary income from non-work sources and spatial location aspects

Variables	(Very low income)	(Low income)	(Medium income)	(High income)	(Very high income)	Global
Number of remittances						
1 Other regions or foreign countries	0.118***	0.088***	0.078***	0.032	0.050	0.044***
	(0.033)	(0.025)	(0.025)	(0.049)	(0.035)	(0.015)
2 Other regions and foreign countries	0.268	0.163*	0.238**	-0.166	-0.006	0.098*
	(0.165)	(0.096)	(0.093)	(0.104)	(0.090)	(0.053)
Receive no remittances (reference)						
Receive retirement income	-0.012	0.042**	0.045***	0.062**	0.098***	0.046***
	(0.024)	(0.017)	(0.015)	(0.026)	(0.023)	(0.009)
Social programme income	0.005	-0.012	0.034***	-0.032	-0.040*	0.002
	(0.021)	(0.014)	(0.013)	(0.023)	(0.023)	(0.008)
Received a cheap housing loan						
for poor people (FONHAPO)	-0.214**	0.188**	-0.035	0.172	-0.126	-0.014
	(0.107)	(0.086)	(0.072)	(0.113)	(0.165)	(0.047)
For workers on a medium income level (INFONAVIT or FOVISSSTE)	0.050	0.048**	0.004	-0.014	-0.139***	0.026**
	(0.033)	(0.019)	(0.015)	(0.026)	(0.027)	(0.010)
For all PEMEX workers	0.370***	-0.075**	-0.243	-0.002	-0.036	0.127
	(0.038)	(0.033)	(0.193)	(0.145)	(0.091)	(0.141)
No support (reference)						
Type of dwelling and floor						
Roof top/Cement or mosaic floor	-0.400	-0.288	-0.189**	-1.012***	-0.077	-0.362***
	(0.372)	(0.192)	(0.079)	(0.037)	(0.457)	(0.123)
Vecindad/Cement floor	-0.468***	-0.450***	-0.492***	-0.956***	-0.292	-0.509***
	(0.075)	(0.044)	(0.054)	(0.289)	(0.285)	(0.033)
Vecindad/Mosaic floor	-0.325**	-0.252***	-0.130**	-0.180***	-0.282**	-0.165***
	(0.127)	(0.072)	(0.063)	(0.067)	(0.134)	(0.040)
House/Dirt floor	-1.727***	-1.449***	-1.745***	-1.189***	-1.314***	-1.579***
	(0.074)	(0.058)	(0.077)	(0.243)	(0.138)	(0.038)
House/Cement floor	-0.451***	-0.355***	-0.335***	-0.233***	-0.180***	-0.358***
	(0.021)	(0.013)	(0.012)	(0.019)	(0.022)	(0.007)
House/Mosaic floor (reference)						
Apartment/Cement floor	-0.039	0.047	0.018	-0.067	-0.183***	-0.011
	(0.062)	(0.035)	(0.032)	(0.051)	(0.068)	(0.021)
Apartment/Mosaic floor	0.111***	0.092***	0.014	-0.046*	-0.045*	0.019*
	(0.038)	(0.023)	(0.018)	(0.026)	(0.025)	(0.011)
Centre - outskirts						
City centre	-0.319***	-0.219***	-0.189***	-0.080**	-0.142***	-0.162***
	(0.047)	(0.032)	(0.025)	(0.033)	(0.032)	(0.015)
Contour 1	0.025	-0.033*	-0.037**	0.062**	-0.051*	-0.009
	(0.032)	(0.020)	(0.017)	(0.027)	(0.026)	(0.011)
Contour 2	-0.013	0.006	0.001	0.089***	-0.122***	0.008
Contour 3 (reference)						
	(0.030)	(0.018)	(0.017)	(0.026)	(0.029)	(0.011)
Contour 4	0.052*	0.008	0.010	0.050	-0.070**	0.008
	(0.027)	(0.017)	(0.018)	(0.034)	(0.035)	(0.011)
Town Size (inhabitants)						
Below 2,500	0.091***	0.033	0.004	0.020	-0.061	0.058***
	(0.034)	(0.023)	(0.025)	(0.050)	(0.058)	(0.015)
Between 2,501 and 15,000 (reference)						
Between 15,000 and 49,999	-0.068**	-0.041**	-0.053***	-0.028	-0.005	-0.033***
	(0.031)	(0.020)	(0.019)	(0.036)	(0.041)	(0.012)
Between 50,000 and 99,999	-0.182***	-0.056**	-0.108***	0.016	-0.013	-0.051***
	(0.042)	(0.026)	(0.026)	(0.049)	(0.056)	(0.016)
Over 100,000	-0.110***	-0.141***	-0.140***	-0.092**	-0.036	-0.124***
	(0.032)	(0.021)	(0.021)	(0.039)	(0.041)	(0.013)
Constant	0.291***	-1.296***	-1.457***	-2.905***	0.421**	-0.331***
	(0.081)	(0.250)	(0.265)	(0.756)	(0.187)	(0.041)
Observations	13,056	23,445	21,265	5,900	4,197	66,392
R-squared	0.113	0.105	0.141	0.125	0.156	0.112

Note: Robust standard errors in brackets*** p<0.01, ** p<0.05, * p<0.1 Source: author with data from the Population and Housing Census (INE, 2011). Standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

6.1. Basic determinants of HQ and the magnitude of the gap.

In this sub-section, we start with what we considered to be the basic determinants of the size of the Housing Quality Gap (HQG) within the sample of vulnerable people. These basic aspects were included in the PSM as an exact match process. The most important aspect is undoubtedly the level of income. The global model for the entire sample reflects a statistically significant relationship, a fact confirmed for the three intervals of medium income levels. While for the sub-samples of households belonging to vulnerable groups on a very low income and for the richest, there is no significant relationship.

The interpretation of these results could be that the poorest groups of people use their slightly higher income to improve their quality of nutrition or other basic needs, rather than their dwelling conditions. This is similar to the observation made by Espinosa *et al*. (2015), who detected that support for social housing for the deprived population is often used for better nutrition. Another likely explanation could be that despite the increase in their income within this low level, they continue to live in the same dwelling, for example, for cultural reasons or because their wage increase is not stable. The non-significant relationship for the richest group is possibly related to the fact that once they reach a certain income level, extra payments are not used for better housing.

The second basic variable is age which, as mentioned, reflects aspects of the costs of living and income related to a person's life cycle (cost of having children and their age, being retired and/or receiving a pension). Moreover, as Wong & Espinoza (2003) mention, as age increases and retirement from economic activity approaches, the flow of income loses value as an indicator of economic wellbeing. This increases the probability that they make use of family networks (obtaining financial resources not included in our income indicator) and make use of assets accumulated in the past. Finally, it is important to remark that age is a significant aspect to consider in the dynamic of indigenous people who present higher child labour rates whose contribution to family income is significant and increases with age, while educational attainment, in contrast, reduces their contribution (Bando *et al*., 2005).

The model reveals a negative linear relationship, where the value of the regression coefficient, to some extent, decreases with increasing age. In fact, a negative effect (larger gap) is found for older adults (between 56 and 75 years old) for the global model and the poorest and richest intervals. While for the three sub-samples of intermediate incomes, the regression coefficient of this interval of age is not significant.

For the two intervals of the youngest vulnerable population (between 16 and 35 years old), the HQG is smaller than that of persons between 36 and 55 years old. The smallest gap is the one for those between 16 and 25 years old. Implicitly, the results show that the reference group (those between 36 and 55 years old) have a larger gap than the youngest but smaller than the oldest individuals. This reflects a non-linearity that seems to exist, to some extent, for all sub-samples by income level.

The larger gap for older people could be related to the fact that pension payments are quite low in Mexico and that people in vulnerable groups have lower pensions than the rest of Mexicans,[35] and therefore have less opportunities to save and accumulate capital. The lower, although negative, gap of the youngest could be a result of the fact that they do not often take on the costs of having offspring or the increasing cost of children as they grow. Moreover, considering the exact match by intervals of income levels, the smaller negative gap of younger vulnerable populations could be related to the fact that those on an income similar to that of white and mestizo Mexicans spend more on direct consumption, including better housing, and save less money for the future[36]. A final explanation may be that there is a socio-cultural change where the level of discrimination is reduced over time. Although in theory this should also affect the elderly population, these people do not often want to change their dwelling, and implicitly their neighbourhood either, due to their social relationship with their neighbours. However, as for most of the potential interpretations in this section, we cannot contrast these ideas with empirical data.

The next basic variable would be the level of formal education. This variable is important because better education offers the possibility to access more and better resources and often a high level of education implies different preferences more focused on good housing (Vera and Ateca, 2008; Clark, 2009). Hence, authors like Schmelkes (2009) remark that the very limited representation of indigenous groups in higher education in Mexico has been the result of the poor quality of education received at earlier educational levels or a consequence of cultural customs, racism and discrimination that limit options and opportunities, among others.

The models shows that vulnerable individuals who lack any formal education[37] or who only have preschool or basic education have a smaller negative HQG. The lack of formal schooling could explain the smaller gap for these people. A substantial number of individuals with such a low educational level are actually talented enough that they might have had the necessary skills to apply them to their jobs, regardless of their educational level, becoming more efficient and productive. Therefore, they would have had a higher income than those with higher education in similar jobs.

On the other hand, for that segment of the vulnerable population with education at the highest levels (university degree and higher), a greater negative gap is observed in the global model. In the case of university and postdoctoral degrees, this is confirmed for all the sub-samples, except for those on the lowest incomes. There are several likely explanations. Firstly, for socio-cultural reasons, indigenous and/or Afro-Mexicans with a high level of income and education may demand fewer luxuries, being already satisfied with a certain minimum level of HQ. In addition, they may spend part of the money they earn to support their family and/or their community. The relationship found for very low-income levels refers to well-trained persons, who do not see their formal education reflected in good jobs with a higher salary, suffering to some extent from the problem of over-qualification.

The fourth basic variable is sex. This indicator reflects several aspects that influence preferences and vulnerability to access better housing conditions. As discussed, gender roles influence integration into the ideal labour market, access to resources and housing affordability. Gender-specific disadvantages

[35] Often because they worked a substantial part of their life in a non-official sector and hence were unable to build up a pension fund or savings.

[36] In order to buy a house in the future or have money for unforeseen costs.

[37] For this educational level, this variable is only statistically significant for the lowest three income groups. While in the case of basic schooling, the negative sign exists for all sub-samples.

as a result of cultural norms and practices may be shared by all Mexicans but operate differently based on ethnicity, region and class. Interestingly, this variable does not present a statistically significant relationship in terms of the size of the gap for any of the models. This does not mean that women could not have better housing conditions than men, as reflected by Viljoen *et al*., (2020) who showed that women have different preferences regarding aspects of the quality of a house, especially when they have children. However, the results indicate that the discrimination of man and woman compared with mestizos and white Mexicans seems to be equal, reflected in a similar HQ gap.

The last basic explanatory indicator reflects the condition of ownership. Several studies have analysed the relationship between this variable and some HQ indicators in the case of indigenous people and Afro-descendants in Peru, Ecuador and Honduras (see Benavides *et al*., 2006; Ponce, 2006; González, 2006, respectively). Their studies show that when ownership levels are taken into account, HQ problems (overcrowding, basic needs that are not satisfied) seem to be more present when tenants are indigenous or Afro-descendants. Our results, confirmed for all income levels, and despite the fact that we used an exact match for this variable, show that renters have a wider HQ gap than owners. Bearing in mind that homeowners can decide who they rent their house to, perhaps discrimination is more patent in this variable to explain part of the gap observed in our analysis. However, this interpretation is difficult to prove with the data available.

6.2. Social aspects and the magnitude of the gap (status as vulnerable, type of cohabitation and physical disability).

In this section, we present the characteristics of the profile of the most affected persons within the vulnerable population considering several social aspects: type of cohabitation at home (couples with or without children, single parents and singles), number of people living in the home: vulnerable group of origin and the presence of any physical disability.

The regression models point to the fact that speaking an indigenous language has a statistically significant negative effect on the individual HQG, compared with the reference category: indigenous non-native speakers. This wider negative gap is confirmed for all income levels except for the sub-sample on the highest income. In contrast, being Afro-Mexican implies—at least in the global model—having a smaller HQG, only statistically significant for the sub-samples of the two lowest income levels.

As regards the type of cohabitation, we took couples with children as the reference group. The global model shows that couples without children have a smaller gap, a fact only confirmed for one of the sub-samples by income level. Single-parent families do not suffer more discrimination than couples with children. In addition, singles have a worse HQG, in this case confirmed for the three lowest income levels. This condition is directly related to the variable number of inhabitants of the dwelling. Households with only one inhabitant—singles—show a smaller gap for all income levels, indicating that singles living alone have better living conditions in terms of the HQG. Furthermore, households where two people live show a smaller HQG. Households with more than five inhabitants show a wider gap, particularly for those residences with more than ten people, who are often highly marginalised as poor and/or families that for traditional/cultural reasons all live altogether.

We also address the analysis of whether physical disabilities exacerbate the HQG. There is some scattered evidence that their vulnerability implies less access to education than the rest of the population

(Lémez, 2005). In the main results of the national housing survey published in 2020, it was noted that there is a need to adapt spaces for people with disabilities in homes in the country and it was found that 12.9% need to install handrails, 12.8% ramps, 12.3% adapt bathrooms and 9.9% widen doors. Given that people with disabilities are much less likely to be engaged economically than the general population, it is essential to try to identify how this affects and interacts with existing vulnerabilities. Hence, the results of the models show that a disability is a disadvantage that negatively affects the individual gap in HQ. Such a negative effect is only observed for one of the sub-samples. The small number of disabled persons in the sample could diminish the precision of the test of significance.

6.3. Housing loans and monetary income from non-work sources (retirement payments and remittances) and the magnitude of the gap.

In this sub-section, we focus on the additional sources of income of individuals other than wages. The PHC offers a set of binary variables that refer to whether people receive such additional income or not, but do not reflect the amount of income earned. We distinguish between two types. Firstly, those additional revenues from different agents or persons that can be freely used for all kinds of purposes (pension or retirement payments, and private remittances from other individuals), or subsidies from social support programmes obtained for specific purposes unrelated to housing.

The second group of three variables that to some degree express extra income directly focuses on housing. One variable reflects government support (subsidies) to the poorest people. The second reflects the mortgage support that the labour institution offers medium income level workers to buy or self-build a house. The third reflects the mortgage support from public firms to workers for the expenses of buying their own house. Below we begin with models in relation to the first type of additional income.

As mentioned, payments from pension funds would increase the income level because that variable only includes salaries. At the same time, those households where the pension is their main source of income may be considered more vulnerable because pensions are quite low in Mexico and often imply the loss of several additional rights such as good health insurance, in the case of public servants who have been affected by the changing retirement schemes that began to be introduced in the 90s.The results of our models suggest that receiving a pension implies a smaller HQG than those dwellings where the head of household does not receive a pension. This positive relationship is confirmed for the four sub-samples on the highest income levels. This is probably due to the fact that part of the population has often worked in the formal sector and has generated more stability to obtain a good pension. In other words, this is the group with the highest pension payments. While the poorest mainly worked in the informal sector with no or few accumulated rights, receiving basic pension payments instead of a higher pension based on their former level of wages and the years they contributed to the pension system.

Those people who receive extra income through remittances suffer a lower HQG, a fact confirmed for the three sub-samples of the lowest income levels, at least for one of the two variables used. For those on a high-income level, no effect was detected, as they probably do not need or do not use the extra income to improve their housing conditions.

Another source of extra income includes all types of social support for the poorest families (e.g., subsidies for nutrition or public transport, education). In our global model, these types of benefits do not affect the housing conditions in terms of the HQG. However, for those on medium incomes the effect is positive while for those on the highest incomes the effect is negative. First of all, most of the subsidies included in this variable do not focus on housing and some of them are not even limited to the poorest sectors of the population. Finally, there is no positive effect on the HQG in the sub-samples of the poorest, which does not really come as a surprise. Often, support for really poor families is used to overcome the deprivation of other basic needs. Families on the lowest income level, instead of using the money for better housing, use it to ensure better nutrition or for other basic needs. This could be called the *substitution effect*, where extra support for housing does not increase total spending on housing but rather frees up money for the poorest families—initially allocated to housing—to fulfil other basic needs.

The effect of the four additional sources of income identified above is very heterogeneous and difficult to interpret. Receiving government support for general reasons widens the HQG for the highest income levels. Income reflects the total sum of wages of all people in the household. Therefore, the negative effect could be related to overcrowded households with several wages and who, despite the support, have a wider negative gap than those who do not have access to support due to their income level. In fact, these results should not be interpreted as if Social Programmes do not have a positive impact on certain aspects of the lives of the vulnerable groups that we are studying, mainly because it depends on the purpose that the Social Programme seeks to support and how it is used by the beneficiaries (to improve nutrition, education and health).

As mentioned above, the second type of sources of extra income are the low interest loans directly focused on buying a house or building it oneself. In Mexico, two institutions[38] offer these services. Although these institutions have specific support programmes for vulnerable groups such as indigenous people and people with disabilities, they support the middle-classes with a minimum income level to ensure that the beneficiaries will pay the loan. However, this support is limited to those above a certain income level.

This type of support shows a significant positive effect for the overall dataset and for the low-income level sub-sample. While for the lowest income levels, no significant relationship exists, which is logical because most of them did not have access to these programmes or used the support for other basic needs. Finally, for the very high-income level sub-sample, we detected a negative effect, probably because those people who access these programmes do not need the money to improve their dwelling.

Some public firms offer specific support for workers to buy their own house. This variable not only represents an extra income but is also a support that is offered mainly to those workers who enjoy a more stable and better paid job. They are considered more reliable when they apply for private bank loans, allowing them higher mortgages and subsequently they can purchase better houses. In this case, the effects on the size of the HQG are unclear. It is non-existent for the global model and for the three

[38] The Institute of the National Housing Fund for Workers (INFONAVIT) and the Housing Fund of the Institute for Social Security and Services of State Workers (FOVISSSTE).

highest income levels. The poorest sub-sample shows a positive effect (reflecting a smaller gap) and for those on a low income, the regression coefficient is negative, suggesting that PEMEX support would lead to a higher HQG.

The only support programme focused on the poor population in order for them to build, expand or improve their house is the one offered by the National Popular Housing Trust Fund (FONHAPO) of the National Housing Organisation (ONAVIS). It only subsidises individuals or family groups registered on one of the social programmes that seek to reduce extreme poverty and promoted by the National Government. In other words, the family must be in a situation of displacement. Interestingly, our results for this variable show a statistically significant negative effect for the lowest income group, indicating a wider gap for the beneficiaries of this type of fund. This is more than likely because they use the benefits for better nutrition or other basic needs, such as schooling. Even when they use the public money for housing, they do not increase the total amount they had initially envisaged for housing. In other words, the income from the public subsidies replaces their initially envisaged private or own funds that are freed up for other basic family needs. For the sub-sample of people on a low income, the effect is positive showing that they use the money to improve their dwellings.

As detected for the first type of extra income (not focused on housing), the different types of loans or support directly focused on housing do not show a clear pattern that is easy to interpret either. However, these types of additional income play an important role as a control variable. For example, FONHAPO controls for those very poor people that are entitled to participate in social programmes. The two institutes that support medium-income workers and the support offered by PEMEX control for the stability of their jobs. In fact, all the variables of this sub-section play this control role vis-à-vis extra income not included as salaries.

At any event, if we exclude this group of variables in the model, all the rest of the results are very similar, basically maintaining the same trends. We keep them in order to control income in terms of wages by other financial resources not quantified by their amount although they affect affordability for better housing.

6.4. Type of dwelling, the geographical residence location (contour) and town size versus the size of the gap.

In this last group of variables used to characterise the type of people from the vulnerable groups that suffer a wider HQG, the location of the residence (contour and town size) is taken into account, as well as the type of dwelling combined with the quality of the flooring materials. This last aspect should apparently be related to the HQ, although—as discussed— the relationship between the indicator and the HQ is irregular and, due to its complexity could not be included as a HQ indicator. However, it definitely fits as a conditioner of the size of the gap and will probably work better for sub-samples of income levels.

The characteristics of this variable are the type of dwelling (*Vecindad*, House, **"Condominium" or Rooftop**) and the characteristics of the floor (whether or not it has some type of coating such as pavement/concrete, wood, mosaic, or another coating). The combination of both offers eight types of

dwellings used in our model. Table 6.1 shows the following relationship with the HQG. Those who live in a "Condominium" with a cement floor have a HQG similar to those in the reference group[39], while if the floor is coated, they show a smaller gap. Those who live in a house or *vecindad* and/or have a dirt or a cement floor have a wider gap than those living in the reference dwellings (a more luxury house with a coated floor). The fact that living in "*vecindades*" or in a house with a dirt floor negatively affecting the HQG is not surprising. In the city centre, there are numerous, abandoned or with frozen rent "*vecindades*" that continue to house the indigenous population with the lowest resources[40]. Many of the buildings are on the verge of collapse.

Most of these relationships are confirmed for almost all the income levels. Except for the sub-sample of people on very high incomes and those living on a rooftop and/or a house with a cement floor. Those who live on the rooftop show—in the global model—a higher HQG, although this fact is confirmed for medium and high-income levels and is not significant for the sub-samples of the highest and poorest income intervals. Those living in a "Condominium" with a coated floor that show a clear opposite effect were detected for the lowest and highest income levels. This is logical because for low-income levels having a coated floor is a luxury, a fact that explains the positive effect of the HQG. It is more difficult to explain why richer families living in such a dwelling have a higher HQ gap. In all likelihood the fact that they do not live in a detached house implies that all the spaces are possibly smaller.

Since much of the variation impacting housing conditions can be explained by differences in the local labour market due to urban/rural dynamics, we also include the town size and the contours where they live. These variables also correct the model for differences in prices and housing opportunities. In fact, analysing the residential location to detect in which part of the MAMC a better or worse situation in the HQ exists for vulnerable groups, we found that living in the city centre compared with living in the third contour (reference) represents a wider gap in HQ. These results could confirm that the city centre is one of the most vivid examples of the situation of vulnerability experienced by indigenous people in Mexico City (Oehmichen, 2001). The sub-sample models confirm this wider gap for all income levels.

Living near the centre—in Contour 1—does not seem to affect the HQG, although the results by the income sub-sample refine our conclusion. The wider gap found for the centre also exists for those vulnerable persons living in Contour 1 on a low- or medium-income level. However, a positive effect exists for those on a high income living in Contours 1 and 2, showing a smaller HQG. Moreover, those vulnerable indigenous or Afro-Mexican people on a very high income show a wider housing quality gap for each of the four contours mentioned. Finally, those on a very low income show a smaller gap when they live in Contour 4 and a higher gap when they live in the city centre.

The last aspect included in the models is the size of the municipality. The regression models show negative and statistically significant coefficients (a wider HQG) for those living in towns with more than 15,000 inhabitants. This fact is confirmed, in general, for all income levels except the highest. In contrast, the global model indicates a smaller HQG for the smallest towns (less than 2,500 inhabitants), although the additional sub-samples only confirm this fact for the very low-income level.

[39] Except for those people with a very high income, where a negative effect was detected.

[40] For example, the Mazahuas of the State of Mexico live in a dozen old *vecindades*, while the Triquis of San Juan Copala, Otomi of Querétaro and the Mazahuas of the State of Mexico live crowded together in public buildings and vacant lots that are federally or privately owned (Oehmichen, 2001).

7.- Housing quality gap of immigrants in Spain

7.1.- Data set and creation of the housing quality index

The strategy developed in the former sections can be widely used for all kinds of measures of disadvantage or inequality in several settings. In fact, we analyzed the same questions using the Spanish database comparing immigrants and native Spaniards and found robust results. Here we will present the main findings and tables for al details see Cruz-Calderón, 2023[41].

For Spain we use the latest available data from the Spanish Population and Housing Census (SPHC-2011) which reflects a broad number of demographic, cultural, economic and social characteristics, and also a wide range of indicators on housing conditions. The initial sample consist of 4,107,465 observations; however, we focus on people between 16 and 65 years old that are at a potential working age. Of the 2,684,259 between 16 and 65 years of age, we exclude people with permanent work disability, retirees, early retirees, pensioners or rentiers or in another situation. We exclude young children under the minimum working age because their HQ index is conditioned by their parents and retired people since there is less probability of finding retired immigrants in the sample. Thus, the sample consists of 659,154 observations, of which approximately 9% (56,769) are considered immigrants: people born abroad. However, we consider only those individuals that were born in Spain as native Spaniards with both native parents (600,720), and those born abroad and without any native Spaniard parent as immigrants (45,655).

The data from the 2011 SPHC are very interesting as they reflect the crossroads of several important phenomena in terms of immigration and the housing market. The immigration boom at the former century created a new perception of immigrants (Valero, Coca and Miranda, 2010) in which native workers fear for their jobs, a tendency reinforced during the 2008 crisis. Moreover, the data implicitly considers the accumulated mid-term effects of the deep economic crisis of 2007-2009, which particularly affected the Spanish housing market and was established as a period in which many immigrants and natives became unemployed and struggled paying their rent or mortgages.

The SPHC survey—offers 19 different indicators. In terms of the **"Technical quality of the housing unit"** we have specific data on the habitable space (overcrowding and privacy) and two on the status or deterioration of the building. The age of the building and an evaluation possible status : ruinous, bad, deficient or in a good condition. The aspect of **"Technical design"** is considered by the survey by adding up the four binary indicators on the availability of primary services or facilities (toilet, bath or shower, water supply system and a wastewater evacuation structure). Some additional variables reflect the accessibility of the building or housing unit (lift in the building, number of floors above ground and accessibility of the unit for disabled people). As regards the "**Aesthetic design, ambience and comfort of the dwelling",** there were used 6 binary variables reflecting the presence of a set of secondary services (garage or parking lot, central heating, central hot water, telephone line, Internet access and central gas supply). After aggregating and transforming the 19 initial variables, eight basic indicators were obtained (see **Table 7.1**); all of them are defined so that the higher its value the better the housing

[41] For all details on the Spanish case see Cruz, S. (2023) heterogeneous living conditions for vulnerable groups: segregation and housing quality in Spain in Mexico PhD Thesis, Universidad Complutense Madrid.

quality. However, as more than 99% of the population in Spain have primary housing services covered, it was considered the focus should be on those aspects that allow to identify more differences, such as housing density, general accessibility, status of the building and secondary facilities. Their inclusion would reduce the variability of our indicator of the HQ were bad scores in other aspects would be offset by this specific variable. Therefore, the last 15 variables of the SPHC survey were restructured in seven combined variables.

Our factor analysis (see Table 7.2) identified three factors that reflect conceptually distinguishable aspects of housing conditions that are in line with the theoretical framework: Factor 1 reflects the Housing density (overcrowding and level of privacy: number of persons per room and number of families living together in the dwelling). Factor 2 reflects Secondary services. This factor also includes the two variables that reflect the Physical soundness of the dwelling, an aspect, although conceptually different, highly correlated with the presence of more secondary services. The last factor reflects the Accessibility of the dwelling (the combined variable by number of floors and lift availability which indirectly reflects the accessibility for disabled people to a certain extent) (Factor 3). Afterwards, we combine the values of each of the three factors in one single synthetic indicator with a continuous scale calculating the simple average.

Table 7.1 - Combination and transformation of 19 variables in eight basic indicators

<table>
<tr><th colspan="3">Technical quality of the housing unit</th></tr>
<tr><td colspan="2">Housing density
(3 variables)</td><td>Status of the building
(2 variables)</td></tr>
<tr><td>1. Space (square meters per person)
(Continuous variable)</td><td>$\frac{Useful\ surface\ area\ of\ the\ dwelling}{Number\ of\ household\ members}$</td><td rowspan="3">4. Building status
Ruinous= 1, Bad= 2, Deficient= 3, Goo4

5. Age of the building.
Before 1900=2; 1900-1920=3; 1921-1940=4; 1941-1950=5; 1951-1960=6; 1961-1970=7; 1971-1980=8; 1981-1990=9; 1991-2001=10; from 2002 to 2011, the numbering is successive, up to 20</td></tr>
<tr><td>2. Number of families</td><td>0= 3 or more, 1= 2 families, 2= 1 family, 3= no family</td></tr>
<tr><td>3. Overcrowding
(Continuous variable)</td><td>$\frac{1}{\left(\frac{Number\ of\ household\ members}{Number\ of\ rooms}\right)}$</td></tr>
</table>

<table>
<tr><td>Technical design (2 combined variables)

6. General accessibility (0 to 6): combination of number of floors and lift availability. 1 floor and lift= 6, 2 floors= 5, 3 floors= 4, 4 floors= 3, 5 floors= 2, 6 floors=1, 7 or more floors= 0

7. Enjoy basic facilities: water supply, wastewater evacuation, bath or shower, toilet. (4 variables - Minimum 0, maximum 4)</td><td>Aesthetic design, ambience and comfort of the dwelling
(1 combined variable)

8.- Number of secondary facilities
Enjoyment of: Internet, gas, garage, telephone line, central hot water, central heating. (6 variables: Minimum 0, maximum 5).</td></tr>
</table>

Source: authors with data from the Population and Housing Census (INE, 2011)

Table 7.2. Housing Quality Index: The results of the factor analysis applied to the seven basic indicators.

Indicators	Factor 1 Housing density	Factor 2 Physical soundness of the dwelling and secondary services	Factor 3 Accessibility of the dwelling
Overcrowding (no. persons per room)	0.927	-0.010	0.009
Number of square meters per person	0.871	0.037	0.090
Number of families (0-3)	0.825	-0.054	-0.051
Secondary services (0-6)	-0.040	0.723	-0.216
Age of the building	0.008	0.709	0.319
Building status	0.004	0.620	-0.003
General accessibility	0.022	-0.006	0.952
Total explained variance (EV)	68.29%	KMO-Test	0,647
EV Factor 1	32.88%	Bartlett's sphericity test	1059997.388
EV Factor 2	20.19%	gl.	21
EV Factor 3	15.22%	Significant	0

Source: authors with data from the Population and Housing Census (INE, 2011)

7.2.- HQG of immigrants in Spain: The application of the PSM technic

Besides the variable used in the PSM application (see Table 7. 3). Once the variables that should be used in the PSM were selected, some other specifications should be mentioned. The so-called Consistency Hypothesis of the PSM method requires that all units treated receive the same treatment and such treatment generates the same potential effect[42] for each unit (Imbens and Rubin, 2015: 10; Keele, 2015b: 5). As will be argued, the consistency of the hypothesis is not fulfilled if the models are estimated using the entire group of immigrants. Thus, the AHQG is estimated for similar groups of countries, guaranteeing in this way more homogeneous groups of immigrants.

- African (e.g., Morocco, Algeria, Senegal)
- Asian (e.g., China, the Philippines, Pakistan, India)
- Caribbean countries
- Latin-American countries
- Eastern European countries (e.g., Poland, Romania, Hungary, Russia)
- Southern European countries (basically Italy, Portugal and the Balkan countries)
- Western European[43] countries (basically Germany, France and the Benelux countries)
- Western offshoots (mainly the USA, Canada, New Zealand and Australia)
- The United Kingdom
- Scandinavian and Baltic countries

This disaggregation is necessary because not only the group of immigrants as a whole is different from inhabitants born in Spain, but also the immigrants from different continents or regions can be quite different. For example, a high number of immigrants from African countries often have a lower level of education, are looking for non-qualified jobs and are, on average, younger, while immigrants from Northern Europe are made up of expats or well-qualified workers. Although our sample is limited to

[42]A requirement that is part of the Stable Unit Treatment Assumption (SUTVA).
[43]The grouping of European countries is based on the United Nations "geoscheme" for Europe. https://unstats.un.org/unsd/methodology/m49/

people younger than 66 years old, a large number of the immigrants from Northern European countries like British and Germans are mainly elderly and from a social class with a high purchasing power. They prefer areas with better climate conditions, long hours of sunshine, mild temperatures throughout the year and close to the Mediterranean coast or the islands (Sabater and Marti, 2012).

Table 7.3.- *Variables used in the PSM: The Spanish case .*

Sex	Dummy variable, 1 if man, 0 if woman
Age	1: between 16 and 25 years old 2: between 26 and 35 years old 3: between 35 and 50 years old 4: between 50 and 65 years old
Level of studies completed	1 = incomplete basic education 2 = completed primary education 3 = baccalaureate or middle grade education 4 = degree or industrial master's degree 5 = higher education
Care chores	Dummy variable, 1 if the individual performs the activity
Housework	Dummy variable, 1 if the individual performs the activity
Potential income based on the type of occupation	The average salary in euros of men and women by economic activity and occupation group is based on the official estimations of the National Statistics Office (INE). We considered full-time (1) or part-time work (0.5) and applied it to the average salary. 1 <= 5,800 2 > 5,800 & <=11,600 3 > 11,600 & <=17,400 4 > 17,400 & <=23,200 5 > 23,200 & <=29,000 6 > 29,000 & <=34,800 7 > 34,800 & <=40,600 8 > 40,600 & <=52,200 9 > 52,200 & <=58,000
Number of family incomes (couple)	The occupation and working hours of the individual and their partner are considered simultaneously.
	1 There is no family income (0 incomes)
	2 If the individual observed or their partner works part time (0.5 incomes)
	3 One is employed full time or both partners work part time (1 “full” income)
	4 One is employed full time and one part time (1.5 incomes)
	5 Both are employed full time (2 full incomes)
Home structure	
Type of family nucleus	1 Single (does not belong to any nucleus) 2 Couple without children 3 Couple with children 4 Father or mother with children
Geographical location of the immigrants	
Regions (binary)	1: Galicia, Asturias or Cantabria 2: Basque Country, Navarre or Rioja 3: Aragon or Castile and Leon 4: Catalonia 5: Extremadura or Castile-La Mancha 6: Madrid 7: Valencia or Murcia 8: Andalusia 9: Balearic or Canary Islands, Ceuta or Melilla
Size of municipality	1 <=10,000 inhabitants 2 >10,000 & <=50,000 inhabitants 3 > 50,000 & <=250,000 inhabitants 4 >250,000 inhabitants

Source: authors with data from the Population and Housing Census (INE, 2011)

In addition, some immigrants from low-income countries of the European Union (Romania and Poland) do not need a visa or permission to work in Spain, having access to better jobs than Africans working in the submerged economy without official permission. Therefore, we estimated a PSM for each of the groups of immigrants by type of country[44]. This method of estimation not only controls some of the

[44] The exact allocation of countries for each aggregate region is shown in Annex section (Annex I).

differences just mentioned but also implicitly considers the differences in the homeland culture, language, religion and other non-observed aspects related to their country of origin. Because immigrants from different countries or with different reasons to immigrate have different requirements and housing expectations (Coates *et al.*, 2013).

7.3.- The results of the PSM model and the overall profile of the HQG of vulnerable persons in Spain

As can be observed in Table 7.4, the size and the sign of the HQG for immigrants from different regions is heterogeneous and strongly depends on where they are born. Those from countries that belong to less developed regions—especially Africa—have a negative gap, in other words, their immigrants have a higher probability of living in worse housing conditions. While for those from more developed regions, such "discrimination" does not exist, and on average those immigrants even have a higher HQ level than their Spanish counterparts.

One of the conclusions of the analyses carried out in the second stage for the Spanish case is that immigrants from developing countries have a negative HQG while those from developed countries have a positive gap (living better than their Spanish counterparts). Therefore, the profiles are estimated separately for two groups of countries. This procedure also facilitates the interpretation, because, in the case of developing countries, a wider gap means having a worse Housing Quality (HQ) than the native clones. While in the case of developed countries, a larger gap means better living circumstances than their native counterparts. To facilitate the presentation, the results of the regression model were divided in two.

The data in Table 7.5a reflects the variables or explanatory factors regarding the basic characteristics of the individual (age, sex) and two very relevant determinants of the HQ, the potential income and the condition of being an owner or renter. This part of the table also includes variables that reflect socio-cultural situations like the structure of the family and involvement in care-giving activities. Finally, it includes the specific conditions or status as an immigrant, such as years of residence in Spain, the native status (or not) of the partner and the region of origin of the home-country.

Table 7.4 *Average Housing Quality Gap (AHQG) by region of origin: The case of Spain*

Groups treated	Average Housing Quality Gap (AHQG)		Housing Density (API1)		Physical Soundness and Secondary services (API2)		Accessibility (API3)		Details on the sample and the requirements of the caliper			
									Observations			Common support (Caliper)
Developing countries	AHQG	S.E.	AHQG	S.E.	AHQG	S.E.	AHQG	S.E.	Matched	Not matched	Total	
Africa	-0.328***	0.011	-0.387***	0.018	-0.574***	0.020	-0.024	0.017	6,170	585	6,755	0.0009
Eastern/Oriental Europe	-0.251***	0.009	-0.365***	0.014	-0.302***	0.016	-0.087***	0.014	9,148	360	9,508	0.0009
Asia	-0.229***	0.024	-0.354***	0.037	-0.205***	0.038	-0.129***	0.042	1,482	75	1,557	0.0009
Latin America	-0.226***	0.006	-0.361***	0.010	-0.210***	0.011	-0.107***	0.011	18,091	1,486	19,577	0.0009
The Caribbean	-0.201***	0.018	-0.314***	0.028	-0.199***	0.030	-0.092***	0.032	2,314	131	2,445	0.0002
Developed countries												
United Kingdom	0.071**	0.071	0.128**	0.055	-0.018	0.042	0.103**	0.041	1,079	53	1,132	0.0009
Scandinavian and Baltic countries	0.002	0.038	-0.013	0.070	-0.024	0.066	0.045	0.068	438	4	442	0.0009
Western Offshoots	-0.027	0.060	-0.027	0.102	-0.179*	0.098	0.124	0.105	248	20	268	0.0002
Western Europe	0.014	0.020	0.023	0.037	-0.064**	0.033	0.084**	0.034	1,920	68	1,988	0.0009
Southern Europe	-0.051**	0.020	-0.088**	0.034	-0.091***	0.034	0.026	0.033	1,923	60	1,983	0.0002

Source: authors with data from the Population and Housing Census (INE, 2011). Notes: ***: p-value<0.01, **: p-value<0.05, *: p-value<0.1. NNM(n) = Nearest Neighbour Matching with 1 individual. Common support Caliper = the maximum distance of the PS allowed between treated and control individuals.

***Table 7.5a**.- Regression model of the effect of specific characteristics of vulnerable groups on their HQG: Case of Spain*

Spain	Spain Developed countries	Spain Developing countries
Age		
Between 16 and 25	-0.020	**0.039****
	(0.054)	(0.019)
Between 26 and 35	**-0.079*****	**-0.018****
	(0.030)	(0.009)
Between 36 and 50 (Reference)		
Between 51 and 65	**0.079*****	**0.025****
	(0.026)	(0.012)
Sex (1 man, 0 woman)	-0.027	**-0.026*****
	(0.022)	(0.008)
Potential income level by type of occupation Lowest	-0.035	**-0.036****
	(0.048)	(0.018)
Low-medium	0.003	-0.000
	(0.039)	(0.015)
Medium-high (Reference)		
Highest income level	0.010	**0.068*****
	(0.031)	(0.019)
Number of incomes		
0 (No job)	0.044	**0.068*****
	(0.037)	(0.012)
0.5 (One part-time job)	-0.025	**0.039*****
	(0.042)	(0.013)
1 (Reference) (One full-time or 2 part-time jobs)		
1.5 (One full-time and one part-time job)	-0.011	-0.012
	(0.036)	(0.015)
2 (Two full-time jobs)	-0.008	**-0.035*****
	(0.031)	(0.013)
Type of family nucleus		
Single	0.012	**-0.179*****
	(0.043)	(0.017)
Couple without children	0.146	-0.027
	(0.095)	(0.034)
Couple with children	**0.176***	**0.052***
	(0.093)	(0.030)
Single parent family (reference group)		
Care-giving activities	-0.007	**-0.030*****
	(0.023)	(0.010)
Occupation status: renting (0) versus ownership (1)	**0.092*****	**0.076*****
	(0.025)	(0.009)
Years living in Spain		
1 to 5	0.035	-0.044***
	(0.029)	(0.012)
6 to 10	0.009	-0.030***
	(0.026)	(0.010)
More than 10 years (reference category)		
Spanish partner (yes = 1, no = 0)	-0.022	0.138***
	(0.024)	(0.013)
Region of origin	(A) Developed Countries	(B) Developing Countries
(A) Western Offshoots / (B) The Caribbean	-0.063	-0.015
	(0.059)	(0.017)
(A) Western Europe / (B) Latin America	-0.017	-0.007
	(0.033)	(0.011)
(A) Southern Europe (B) Africa	**-0.058***	**-0.088*****
	(0.032)	(0.015)
(A) Scandinavian and Baltic countries / (B) Asia	-0.033	-0.003
	(0.042)	(0.022)

***Table 7.5b**.- Regression model of the effect of specific characteristics of vulnerable groups on their HQG: Case of Spain*

Education	Spain Developed countries	Spain Developing countries
Incomplete-basic	-0.020	-0.006
	(0.034)	(0.013)
Completed primary education (reference)		
Baccalaureate-middle	-0.001	**-0.027****
	(0.037)	(0.010)
Higher-Industrial master	-0.008	**0.072*****
	(0.054)	(0.026)
Higher	**-0.079***	**0.071*****
	(0.046)	(0.021)
Workers with over-qualification	0.067	**-0.080*****
	(0.050)	(0.024)
Partner's education		
Illiterate	-0.231	**-0.123*****
	(0.168)	(0.032)
No studies	**-0.325*****	**-0.041****
	(0.075)	(0.021)
Second grade (reference)		
First grade	0.005	**-0.043*****
	(0.037)	(0.014)
Third grade	**0.059****	**0.076*****
	(0.029)	(0.013)
Not living with a partner	**0.190****	**0.057***
	(0.090)	(0.030)
Parents' education		
Illiterate	**-0.398*****	**-0.150****
	(0.130)	(0.070)
No studies	-0.281	0.037
	(0.231)	(0.059)
Second grade (reference)		
First grade	-0.157	-0.032
	(0.145)	(0.047)
Third grade	0.038	**0.098****
	(0.109)	(0.041)
Unknown		**0.074****
		(0.031)
Region of residence		
Cantabrian coast	**0.106***	**0.040***
	(0.057)	(0.023)
Madrid	**0.090***	**-0.039****
	(0.052)	(0.017)
The "Castiles"	**0.094****	0.005
	(0.048)	(0.017)
Northeast of Spain (reference)		
Mediterranean Axis	0.041	**-0.072*****
	(0.043)	(0.014)
South of Spain	**0.075***	0.026
	(0.041)	(0.017)
Canary Islands	**0.097***	**0.070*****
	(0.058)	(0.023)
Municipality size		
Until 2,000 inhabitants	**-0.064***	**-0.098*****
	(0.036)	(0.014)
2001 to 10,000 inhabitants	0.039	0.010
	(0.029)	(0.012)
10,001 to 50,000 inhabitants (reference)		
50,001 to 100,000 inhabitants	**0.091****	0.014
	(0.036)	(0.015)
100,001 to 350,000 inhabitants	0.010	**-0.113*****
	(0.041)	(0.014)
More than 350,000 inhabitants	-0.023	**-0.044*****
	(0.030)	(0.012)
Constant	-0.279	**-0.261*****
	(0.190)	(0.059)
Observations	5,590	36,938
R-squared	0.027	0.036

Source: authors with data from the Population and Housing Census (INE, 2011). Standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

In Table 7.5b, we present the part of the regression that reflects the educational background of the persons observed and their family and geographic location (region and municipality size) where the immigrants reside. As will be observed, the model for immigrants from developing countries offers a broad profile highlighting that some specific types of immigrants from those countries have a wider HQG. In contrast, the model for developed countries is very concise, only a few explanatory characteristics of the immigrants affect the size of the HQG. Those variables that are significant basically reflect a similar trend or form of relationship as observed in the profile for immigrants from developing countries.

The results are remarkable and offer novel information about differentiated HQGs for different groups of immigrants according to their place of birth and the income level of their home countries. The models show that immigrants from regions that basically consist of less developed low or medium-income countries have a negative HQG, especially those from Africa. While immigrants from the United Kingdom enjoy a higher housing quality than native Spaniards (except southern Europeans) and those from Western Europe, Scandinavian and Baltic Countries and the so-called Western Offshoots have a similar housing quality.

In the final stage of the analysis —the main purpose of this study— a more comprehensive view of the heterogeneity of the HQG between the different groups of immigrants was offered. We studied the characteristics of those immigrants particularly affected by worse housing opportunities (reflected by a larger negative HQG). In fact, this analysis is the main new feature of this study in terms of international empirical evidence. Regression models using individual HQGs as a dependent variable were estimated. The explanatory variables are basic determinants: aspects of their status as an immigrant; educational background; socio-cultural aspects and geographical differences. Taking into account the second stage results (the negative gaps for developing countries and the absence of a gap or its positive nature for developed countries) these models were separately estimated for two subsamples of immigrants: those from developing countries and those from developed countries. The results suggest that, for immigrants from developing countries, there exists, on average, a negative HQG, especially wider for men, Africans, and those who are high-qualified and who perform care-giving activities. Although the gap is still negative for those with a high level of studies, who have a Spanish partner, and who are home-owners, it is less pronounced. Moreover, the HQG decreases as the number of years living in Spain increase, and as the income level grows. In terms of geographic aspects, the gap is wider in the Mediterranean Axis and those from the Region of Madrid, and lower in the Canary Islands and on the Cantabrian coast. Immigrants residing in the smallest and biggest municipalities are the most affected. For immigrants from developed countries, the profile is very concise. The average gap is positive —they have better housing conditions— and only a few variables influence the size of the HQG. The positive gap is lower for those that are middle aged, with their partner or parents with the lowest levels of education and is higher for those who are older. In terms of the geographical aspects, those living on the “Cantabrian coast” show a higher positive gap. Immigrants residing in the smallest municipalities see their gap negatively affected.

The results stress the importance of an inclusive society that goes beyond only offering immigrants jobs and, instead, that also tries to close the physical, social and economic gaps that keep people separated. As Farber and Allard (2012) set out, the immigration problem might better be framed in terms of vulnerability, to highlight the risk of social exclusion that immigrants may face, which can become a threat to social cohesion and the economic prosperity of the country as a whole.

Although the study shows interesting evidence regarding marginalisation in terms of immigrant housing in Spain, some limitations should be mentioned. The most important could be the lack of information regarding exact household income, and the absence of data on the quality of the neighbourhood atmosphere in terms of services, traffic and public transport or criminality and pollution. Other pending aspects for future research are the impact on the HQG of the settlement patterns of immigrants, or according to gender or ethnic differences. Another interesting avenue for future research could be the analysis of the generational effect: the conditions of second-generation immigrants seem to be fundamental for real comprehension of their integration process.

8.- Conclusions and final remarks

The main objective of this working paper was to explain an alternative and novel methodological research strategy to analyse the Housing Quality Gap (HQG) of vulnerable sectors of the population in terms of ethnic and racial condition. The Mexican case was mainly presented, considering the indigenous and Afro-Mexicans of the Metropolitan Area of Mexico City (MAMC) as a vulnerable population. Therefore, we begin this section with a brief description of the main results and some policy implications. We end with some comments on the methodological implications, limitations and observations.

8.1.- Main results and tendencies

Given the emerging interest in the living conditions of Afro-descendants and the barely explored housing quality of the indigenous population in Mexico, the objective of this study was to analyse the existence of a possible Housing Quality Gap (HQG) between these groups of population and mestizos and with Mexicans. The intention is to detangle the causes of the unequal housing conditions between the vulnerable and non-vulnerable population regarding those aspects directly related with their condition of vulnerability (race or ethnicity) versus those caused by other structural differences (like the level of income, or studies or age). Especially, the inclusion of the Afro-Mexican inhabitants in our analysis might be considered novel, because there are only a few studies in Mexico that analyse their specific situation, showing contrasting results. We acknowledge that besides classism, racism is a problem that must be addressed to better understand how inequalities and discrimination operate affecting vulnerable groups in Mexico through multiple aspects that intersect making them less or more susceptible to exclusion or poverty.

The main results show that, in general, indigenous people suffer from worse housing conditions than the rest of the population in the MAMC, a gap being especially wide for the native speakers. While Afro-Mexicans seem to enjoy similar housing quality to mestizos and white Mexicans. The additional PSM models analyse which of the three partial indices (factors) have the highest impact on the overall HQG. In our model, the factor with the main weight in the overall gap is the number of services, supplies and amenities, being therefore the main features in which this population finds their most deprived situation. In terms of housing density, indigenous native speakers are the most affected, in contrast with Afro-Mexicans, who present, in fact, a positive significant gap. The weight of the third partial index of HQ—condition of the materials used for roofs and walls—within the overall HQG also has a negative effect on the gap, although less marked.

Repeating the PSM models for five sub-samples by levels of income, the wider negative gap detected in the case of indigenous native speakers was confirmed for four of them. Only no statistically significant gap was found at the very high-income levels—probably because of the few observations this interval of income. Those five subsamples also show that the gap seems to be larger for people on a low- to medium-income level and a bit lower for the interval of very low-income. Like for the overall model, in none of the sub-samples for self-identified Afro-Mexicans a negative gap was detected.

Regarding the characteristics of those particularly affected by worse housing opportunities (reflected by a larger negative HQG), we created a profile to characterise the most negative effected persons. The models were estimated for the global sample joining the three vulnerable groups as a whole and

including their ethnicity as an independent variable. The results suggest that, on average, a negative HQG is particularly wide for indigenous people who continue to speak their native language, for households belonging to the three income intervals in the middle, men, older people, renters, couples without children and single parents, those with physical or mental disabilities, those living in a *vecindad* or house with no floor covering, and those residing in the city centre or localities with more than 100,000 inhabitants. The gap is less pronounced—although still negative—for the following types of people: young people, those with low levels of studies, for the models of the same income interval, for those who receive a pension or remittances, or a housing loan from a public institution. From a geographical location point of view, those residing in localities of less than 2,500 inhabitants also have a smaller gap, as do those in the sub-sample of very poor people living in the most peripheral contour. In other words, those poor vulnerable people in those places have a lower negative gap than the rest of the vulnerable population.

8.2.- Policy recommendations.

In terms of policy recommendations, one of the arguments for starting this study was the lack of good research data on some aspects of the housing conditions of vulnerable groups. Especially in relation to the identification of those vulnerable people who suffer the most from the worst housing quality due to their status as immigrant, indigenous or Afro-descendant. Therefore, a call for a systematic effort to produce more and improved data and better additional research should be made. This is an absolutely necessary first step towards the production of analytical knowledge to inform policy-makers about the most effective ways to improve the living standards of these underprivileged groups in our societies. In this regard, the development of the profile that identifies the most affected people can give some light for a more focussed policy approach.

One of the main problems of policy making is that each measure must focus on a specific group of the population that can be easily identified by means of strict measurable criteria (such as age, type of job, or level of education) in order to assure clearly who has the clearance of getting support -or not- and to overcome a discretionally applications of the norms. In this regard, the development of the profile that identifies the most affected people can shed some light for a more focussed policy approach. Several conclusions from the analysis offer more background information on possible bottlenecks that guide useful criteria for adjusting support to specific highly vulnerable groups.

One of the few direct policy recommendations for Mexico is the need for specific attention to the population that speaks their native language and the younger part of the indigenous population. Moreover, it justifies the existing policy based on the construction of dwellings specifically for those indigenous people living in the city centre. The other aspects of the profile are less appropriate to be used in formal legal norms and criteria, though offer interesting background information to the policymakers.

The difficulty to eradicate racial discrimination and its structural effects of social exclusion of the native communities in the MAMC is clear. Therefore, it is urgent to analyse the current policies aimed at native communities to promote and guarantee spaces for cultural miscegenation that respect and protect their identity and diversity. Studies such as the one presented in this paper seek to provide valuable information that helps policy makers to identify the clear elements that intersect in the lives of urban indigenous people and that complicate their possibilities of social mobility. Thus, national and local policy can be formulated to make investment in long-term and efficient policies possible.

Therefore, a call for a systematic effort to produce more and improved data and better additional research should be made.

8.3.- Methodological limitations and remarks.

Particularly, from a methodological point of view, this research poses various important contributions to the literature on this matter. First of all, most existing studies on deprivation (including in terms of housing) offer descriptive statistics focussed on the poorest individuals. While the chosen methodologic approach permits to analyse the level of "discrimination" that exist also for vulnerable people in middle- and high-income groups. The focus on the poorest echelons of the society is related to the fact that most of the existing literature uses a binary criterion when establishing the conditions required to ensure a minimum level of socially acceptable standard of living. Those living below that minimum standard are considered as deprived. Such bottom-base-line approach does not study the heterogeneity between vulnerable persons, not being able to analyse the profile of the are the more or less disadvantaged.

In order to address both shortcomings, we proposed, instead of such bottom-line approach, the creation of a multidimensional complex housing quality indicator (HQI) with a continuous range of values, so that we can the estimate the level of the HQG of each individual in every income level and analyse some specific socio-economic characteristics correlated with the size of the gap. In this paper we used 28 different variables that measures different aspects of housing conditions, and that information was synthesised by applying the factor analysis technique, creating our unique combined HQI. An important advantage FA is that it, implicitly, standardises the data of variables with very different scales.

This unique indicator was used in our Propensity Score Matching (PSM) Models firstly to separate the part of different living conditions caused by race or ethnicity and those caused by other structural aspects (age, income, education level, etc.). In this way, we solved the so-called selection bias problem and obtained an unbiased HQG size caused by discrimination against the racial or ethnic condition of the individuals. The size of the gap is based on the comparison of vulnerable individuals with their non-vulnerable counterfactual clone (with the same socio-demographic and socio-economic characteristics) and implicitly reflects what the individual's housing quality would have been if they had been mestizo or white Mexican.

This research is not exempt from limitations, most of which are mentioned in each of the corresponding sections. However, the most important ones are again highlighted here. First, although we apply the PSM technique as a proxy for the evaluation of causality of discrimination against vulnerable population, regarding the interpretation of the HQG, it cannot be assured that the housing quality gap observed for the vulnerable population is caused solely by discrimination, as it may be due to a combination of several aspects of affordability, socio-cultural traditions, and personal decisions. However, indirectly, differences in the housing quality among people with similar level of education, responsibilities by type of job and age do reflect discrimination in economic opportunities (based, among others on their wages and on terms of over-qualification) to some extent. In any case, the results of the research clearly reflect the existing vulnerability of the population analysed since, having the same socio-economic and socio-demographic characteristics, they choose or are pushed to houses in worse conditions.

Secondly, from a methodological point of view, in recent articles, some aspects of the PSM method have been criticised (see Guo et al, 2020; Choi et all, 2019; Desai and Franklin, 2019), especially in the case of a low number of observations and variables available for the matching process. Considering that in this research there were more than 600,000 observations and that the number of variables used was quite large, it is considered that this would be a minor problem. Furthermore, the alternative solution suggested by the literature is to use the difference in difference method. However, in that method a panel data set is needed, while the Population and Housing Census is only carried out every ten years.

There are important methodological findings of our analysis with implications for future research, as shown by the results obtained for afro-Mexicans. As mentioned, no HQG was detected, however, looking at the gaps for the different partial indices the conclusions are quite different and even contrasting. The absence of a gap was confirmed for the indicator of the quality of the construction materials, although there is a positive gap in terms of overcrowding and a negative gap in basic and secondary services. This means that the choice of one or the other indicator is not impartial since it may imply reaching opposite research results. Hence, the heterogeneity in the results of the empirical studies for Afro-Latin Americans, apart from different contextual particularities in every country, may also depend on the indicator used. Therefore, using several individual indicators together with a synthetic composed overall index could shine better light on the complex reality of the housing conditions. Where, the analysis of individual indicators or partial indices makes it possible to identify which of them explains the global gap.

Finally, a wide area of new research options using the PSM method was detected. Based on the results of our profile of the most affected people, a more in-depth analysis could be considered for certain people's profiles. For example, one could start with a PSM model that compares the housing quality (HQ) for people with multiple vulnerability characteristics (indigenous people, women, people with a relevant physical or mental disability, and homosexuals).

From a geographic location point of view, a novel approach would be the analysis of the HQG between people who live in the areas where indigenous and Afro-Mexican populations are more represented. Another interesting avenue for future research could be a more in-depth analysis of how monetary income from non-work resources affects housing conditions.

As shown the strategy used can be widely used for all kinds of measures of disadvantage or inequality in several settings. In fact, we analysed the same questions using the Spanish database comparing immigrants and native Spaniards and found robust results (See annex VII and also Cruz-Calderón, 2023[45]). Our main conclusions is that For immigrants from developing countries, the HQG is larger for those with lower incomes, with over-education, those that are single and/or male and for those performing care-giving activities. In contrast, the HQG decreases when the years of residence in Spain increase and for higher income levels.

[45] For all details on the Spanish case see Cruz, S. (2023) heterogeneous living conditions for vulnerable groups: segregation and housing quality in Spain in Mexico PhD Thesis, Universidad Complutense Madrid.

In terms of geographical aspects, the gap is wider for those from Africa and those living in the Mediterranean Axis or in the Region of Madrid. Moreover, the HQG is higher in the smallest and the biggest municipalities compared with the small/medium-sized areas. Finally, for immigrants from developed countries, the profile is very concise.

Annexes

Annex I. List of municipalities that make up the Metropolitan Area of Mexico City (MAMC)

Contour	Municipality	Contour	Municipality	Contour	Municipality
Central City	Benito Juárez	Contour 3	Tecámac	Contour 4	Isidro Fabela
	Cuauhtémoc		Valle de Chalco solidaridad		Jilotzingo
	Miguel Hidalgo		Chalco		Juchitepec
	Venustiano Carranza		La Paz		Melchor Ocampo
Contour 1	Álvaro Obregón		Coacalco de Berriozábal		Nextlalpan
	Azcapotzalco		Huixquilucan		Nopaltepec
	Coyoacán		Chicoloapan		Otumba
	Gustavo A. Madero		Cuautitlán		Ozumba
	Iztacalco		Cuautitlán Izcalli		Papalotla
	Iztapalapa		Jaltenco		San Martín de las Pirámides
Contour 2	Tlalpan	Contour 4	Tizayuca		Temamatla
	Ecatepec de Morelos		Acolman		Temascalapa
	Magdalena Contreras		Amecameca		Tenango del Aire
	Naucalpan de Juárez		Apaxco		Teoloyucan
	Nezahualcóyotl		Atenco		Teotihuacán
	Tlalnepantla de Baz		Atlautla		Tepetlaoxtoc
	Xochimilco		Axapusco		Tepetlixpa
Contour 3	Tláhuac		Ayapango		Tepotzotlán
	Cuajimalpa de Morelos		Chiautla		Tequixquiac
	Milpa Alta		Chiconcuac		Texcoco
	Chimalhuacán		Cocotitlán		Tezoyuca
	Atizapán de Zaragoza		Coyotepec		Tlalmanalco
	Tultitlán		Ecatzingo		Tultepec
	Ixtapaluca		Huehuetoca		Villa del Carbón
	Nicolás Romero		Hueypoxtla		Zumpango

Source: Toscana and Pimienta (2018) classification

Annex II. Probability of belonging to a vulnerable group. Probit model before and after the Match

VARIABLES	Non-native speaker indigenous		Native speaker indigenous		Afro-Mexican		Vulnerable groups	
	Before	After	Before	After	Before	After	Before	After
Level of studies completed								
No education	0.145***	0.002	1.041***	0.003	0.025	0.052	0.364***	-0.003
	(0.015)	(0.026)	(0.019)	(0.043)	(0.030)	(0.071)	(0.012)	(0.020)
Basic education	0.123***	-0.001	0.496***	0.002	0.009	-0.005	0.176***	-0.004
	(0.005)	(0.009)	(0.011)	(0.028)	(0.010)	(0.023)	(0.005)	(0.008)
Superior degree	-0.083***	0.003	-0.153***	-0.010	0.008	-0.002	-0.074***	-0.002
	(0.007)	(0.013)	(0.019)	(0.048)	(0.012)	(0.026)	(0.006)	(0.011)
Post superior education	-0.197***	0.001	-0.058	0.010	0.022	0.046	-0.148***	0.003
	(0.020)	(0.045)	(0.045)	(0.124)	(0.029)	(0.068)	(0.018)	(0.037)
Type of Cohabitation								
Couples without children	0.002	0.003	-0.016	-0.002	-0.068***	0.018	-0.015**	-0.000
	(0.008)	(0.014)	(0.015)	(0.032)	(0.015)	(0.035)	(0.007)	(0.012)
Single parents	0.004	-0.004	-0.160***	0.010	-0.004	0.035	-0.024*	0.002
	(0.014)	(0.026)	(0.027)	(0.068)	(0.025)	(0.058)	(0.013)	(0.023)
Single	0.018***	0.000	-0.224***	0.001	-0.046***	0.007	-0.029***	-0.000
	(0.005)	(0.009)	(0.010)	(0.023)	(0.010)	(0.021)	(0.005)	(0.008)
Number of employment benefits								
1	-0.076***	0.006	-0.109***	0.001	-0.001	-0.216**	-0.076***	-0.002
	(0.019)	(0.040)	(0.034)	(0.090)	(0.036)	(0.088)	(0.017)	(0.035)
2	0.043*	0.029	0.075*	0.009	0.032	-0.159	0.057***	0.001
	(0.024)	(0.048)	(0.040)	(0.103)	(0.045)	(0.108)	(0.021)	(0.042)
3	0.031	0.018	0.102**	0.017	0.080*	-0.176	0.058**	-0.004
	(0.026)	(0.052)	(0.044)	(0.112)	(0.047)	(0.114)	(0.023)	(0.046)
5	0.033	0.002	-0.035	0.033	0.175***	-0.207*	0.049**	0.046
	(0.026)	(0.052)	(0.048)	(0.127)	(0.046)	(0.112)	(0.023)	(0.046)
6	0.032	0.009	-0.108**	0.111	0.045	-0.165	0.014	-0.000
	(0.024)	(0.049)	(0.046)	(0.115)	(0.045)	(0.107)	(0.022)	(0.044)
7	-0.040*	0.002	-0.175***	0.012	0.014	-0.218**	-0.058***	-0.002
	(0.021)	(0.044)	(0.040)	(0.103)	(0.039)	(0.095)	(0.019)	(0.038)
8	-0.126***	0.006	-0.262***	0.010	0.049	-0.212**	-0.127***	-0.009
	(0.020)	(0.042)	(0.037)	(0.096)	(0.037)	(0.090)	(0.018)	(0.036)
Town Size								
Below 2,500	0.253***	0.000	0.250***	0.001	-0.196***	0.003	0.213***	0.000
	(0.009)	(0.015)	(0.016)	(0.035)	(0.019)	(0.044)	(0.008)	(0.013)
Between 2,500 and 14,999	0.150***	0.000	0.109***	0.000	-0.067***	0.001	0.121***	0.000
	(0.008)	(0.013)	(0.016)	(0.034)	(0.014)	(0.031)	(0.007)	(0.011)
Between 50,000 and 99,999	0.005	0.000	0.110***	-0.002	-0.097***	-0.002	0.002	0.000
	(0.012)	(0.022)	(0.023)	(0.053)	(0.021)	(0.049)	(0.011)	(0.019)
Over 100,000	-0.107***	-0.000	-0.042***	-0.001	-0.058***	0.001	-0.104***	-0.000
	(0.008)	(0.014)	(0.014)	(0.031)	(0.014)	(0.033)	(0.007)	(0.012)
Observations	661,226	107,744	614,472	20,054	613,331	18,882	681,657	144,214
LR-test: Chi2	6273	1.144	8332	2.453	442.2	9.435	8782	3.129
ROC	0.5978	0.5011	0.7700	0.5026	0.5594	0.5062	0.6012	0.5010
Log of likelihood	-192286	-74682	-48938	-13899	-49405	-13083	-237858	-99960

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 *-Source: authors with data from the Population and Housing Census (INEGI, 2020)

Annex III. Cont. Probability of belonging to a vulnerable group. Probit model before and after the Match

Variables	Non-native speaker indigenous		Native speaker indigenous		Afro-Mexican		Vulnerable groups	
	Before	After	Before	After	Before	After	Before	After
Contours								
Central City	-0.048***	-0.000	0.056***	0.000	0.048**	-0.000	-0.021**	-0.000
	(0.011)	(0.020)	(0.020)	(0.045)	(0.019)	(0.043)	(0.010)	(0.017)
1	-0.065***	-0.000	-0.083***	-0.001	-0.013	0.001	-0.068***	-0.000
	(0.009)	(0.016)	(0.016)	(0.036)	(0.016)	(0.037)	(0.008)	(0.014)
3	-0.031***	-0.000	-0.028**	0.000	0.045***	0.002	-0.022***	-0.000
	(0.007)	(0.012)	(0.012)	(0.026)	(0.013)	(0.031)	(0.007)	(0.011)
4	-0.221***	-0.000	-0.438***	-0.000	0.002	0.004	-0.244***	-0.000
	(0.009)	(0.015)	(0.016)	(0.033)	(0.017)	(0.039)	(0.008)	(0.013)
Household monthly income								
From 0 to 171 dollars	-0.060***	0.000	-0.112***	0.002	-0.015	0.000	-0.070***	0.000
	(0.009)	(0.015)	(0.016)	(0.035)	(0.018)	(0.042)	(0.008)	(0.013)
From 172 to 245 dollars	-0.062***	0.000	-0.023	0.001	0.030	-0.000	-0.049***	0.000
	(0.010)	(0.017)	(0.017)	(0.036)	(0.019)	(0.045)	(0.009)	(0.015)
From 338 to 536 dollars	-0.019**	-0.000	0.003	0.000	0.055***	-0.001	-0.007	-0.000
	(0.009)	(0.015)	(0.015)	(0.032)	(0.017)	(0.038)	(0.008)	(0.013)
From 537 to 832 dollars	-0.034***	-0.000	-0.051***	-0.000	0.074***	-0.001	-0.025***	-0.000
	(0.009)	(0.015)	(0.016)	(0.033)	(0.017)	(0.038)	(0.008)	(0.013)
From 833 to 1,086 dollars	-0.044***	-0.000	-0.053***	-0.000	0.032	-0.001	-0.041***	0.000
	(0.010)	(0.018)	(0.019)	(0.042)	(0.020)	(0.045)	(0.009)	(0.016)
From 1,087 to 1,617 dollars	-0.083***	-0.001	-0.076***	-0.000	0.099***	-0.002	-0.061***	0.000
	(0.010)	(0.018)	(0.020)	(0.044)	(0.019)	(0.043)	(0.010)	(0.016)
More than 1,617 dollars	-0.147***	-0.001	-0.132***	-0.000	0.074***	-0.004	-0.123***	0.000
	(0.012)	(0.021)	(0.023)	(0.054)	(0.020)	(0.046)	(0.010)	(0.018)
Retirement income	0.047***	-0.000	-0.186***	0.003	-0.003	0.001	0.013**	-0.000
	(0.007)	(0.012)	(0.015)	(0.039)	(0.012)	(0.030)	(0.006)	(0.011)
Owner/Renter	0.003	-0.000	-0.165***	0.000	-0.027***	-0.002	-0.032***	-0.000
	(0.005)	(0.009)	(0.009)	(0.020)	(0.009)	(0.021)	(0.005)	(0.008)
Age								
between 16 and 25 years old	-0.072***	-0.000	-0.150***	0.001	-0.044***	0.004	-0.082***	-0.001
	(0.007)	(0.012)	(0.014)	(0.032)	(0.012)	(0.028)	(0.006)	(0.010)
between 26 and 35 years old	-0.036***	-0.000	-0.013	0.000	-0.017	0.002	-0.032***	-0.000
	(0.006)	(0.011)	(0.011)	(0.025)	(0.011)	(0.025)	(0.006)	(0.009)
between 56 and 75 years old	0.108***	0.000	-0.018	0.001	-0.021*	0.002	0.073***	0.000
	(0.006)	(0.011)	(0.012)	(0.025)	(0.012)	(0.029)	(0.006)	(0.010)
Sex	-0.000	-0.001	0.023**	-0.002	0.025***	-0.004	0.009**	0.000
	(0.005)	(0.008)	(0.009)	(0.020)	(0.009)	(0.020)	(0.004)	(0.007)
Constant	-1.205***	-0.005	-1.922***	-0.006	-2.124***	0.205**	-1.026***	0.005
	(0.023)	(0.046)	(0.040)	(0.104)	(0.043)	(0.102)	(0.021)	(0.040)
Observations	661,226	107,744	614,472	20,054	613,331	18,882	681,657	144,214
LR-test: Chi2	6273	1.144	8332	2.453	442.2	9.435	8782	3.129
ROC	0.5978	0.5011	0.7700	0.5026	0.5594	0.5062	0.6012	0.5010
Log of likelihood	-192286	-74682	-48938	-13899	-49405	-13083	-237858	-99960
Pseudo R2	0.0162	7.66e-06	0.0986	8.83e-05	0.00432	0.000360	0.0187	1.57e-05

Source: authors with data from the Population and Housing Census (INEGI, 2020) Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Annex IV. Probability of belonging to a vulnerable group by income level. Probit model before and after the Match

	Very-Low		Low		Middle		High		Upper-High	
Variables	Before	After	Before	After	Before	After	Before	After	Before	After
Level of studies completed										
No education	0.142***	0.031	0.156***	-0.003	0.138***	0.026	0.161**	0.097	0.153*	0.042
	(0.023)	(0.038)	(0.027)	(0.045)	(0.032)	(0.053)	(0.066)	(0.126)	(0.083)	(0.170)
Basic education	0.145***	0.004	0.116***	-0.014	0.113***	-0.003	0.109***	0.011	0.141***	-0.012
	(0.011)	(0.019)	(0.010)	(0.016)	(0.010)	(0.017)	(0.019)	(0.034)	(0.024)	(0.044)
Superior degree	-0.085***	0.016	-0.067***	-0.009	-0.073***	0.008	-0.070***	0.017	-0.150***	-0.008
	(0.015)	(0.027)	(0.014)	(0.025)	(0.012)	(0.022)	(0.020)	(0.037)	(0.023)	(0.042)
Post superior education	-0.289***	0.006	-0.170***	0.081	-0.159***	0.079	-0.207***	0.079	-0.205***	0.007
	(0.048)	(0.105)	(0.057)	(0.118)	(0.040)	(0.082)	(0.049)	(0.102)	(0.040)	(0.081)
Type of Cohabitation										
Couples without children	0.004	0.010	0.007	-0.001	0.006	0.007	0.006	-0.011	-0.042	0.014
	(0.016)	(0.027)	(0.015)	(0.025)	(0.015)	(0.026)	(0.027)	(0.048)	(0.030)	(0.056)
Single parents	0.035	0.006	-0.024	0.047	0.003	0.016	-0.043	-0.007	0.052	0.018
	(0.024)	(0.042)	(0.026)	(0.046)	(0.027)	(0.047)	(0.053)	(0.112)	(0.062)	(0.127)
Single	0.025**	-0.002	0.018*	0.001	0.023**	-0.011	-0.004	0.000	-0.000	0.013
	(0.010)	(0.017)	(0.010)	(0.016)	(0.010)	(0.017)	(0.018)	(0.033)	(0.021)	(0.039)
Number of employment benefits										
1	-0.086*	-0.168*	-0.139***	-0.069	-0.042	-0.045	-0.030	0.020	0.017	0.012
	(0.050)	(0.091)	(0.033)	(0.055)	(0.033)	(0.058)	(0.059)	(0.128)	(0.075)	(0.168)
2	0.006	-0.183*	-0.019	-0.041	0.074*	-0.023	0.145**	0.047	0.125	0.065
	(0.061)	(0.111)	(0.041)	(0.068)	(0.040)	(0.070)	(0.073)	(0.151)	(0.096)	(0.212)
3	-0.021	-0.177	0.003	-0.072	0.042	-0.023	0.043	-0.025	0.197**	0.077
	(0.069)	(0.127)	(0.044)	(0.074)	(0.044)	(0.077)	(0.079)	(0.159)	(0.097)	(0.210)
5	-0.004	-0.139	-0.040	-0.026	0.089**	-0.006	0.044	0.125	0.153	0.097
	(0.071)	(0.128)	(0.045)	(0.077)	(0.043)	(0.077)	(0.078)	(0.161)	(0.097)	(0.209)
6	-0.017	-0.106	-0.009	-0.038	0.037	-0.018	0.124*	0.047	0.162*	0.114
	(0.066)	(0.122)	(0.042)	(0.072)	(0.041)	(0.072)	(0.072)	(0.150)	(0.089)	(0.194)
7	-0.066	-0.111	-0.112***	-0.071	-0.018	-0.038	0.028	0.025	0.091	0.005
	(0.058)	(0.108)	(0.038)	(0.064)	(0.036)	(0.064)	(0.063)	(0.135)	(0.079)	(0.176)
8	-0.149***	-0.171*	-0.206***	-0.074	-0.092***	-0.053	-0.084	0.020	0.018	0.031
	(0.054)	(0.098)	(0.035)	(0.059)	(0.034)	(0.061)	(0.060)	(0.131)	(0.076)	(0.171)
Town Size										
Below 2,500	0.248***	-0.000	0.237***	0.001	0.245***	0.001	0.298***	-0.000	0.346***	0.005
	(0.017)	(0.027)	(0.015)	(0.025)	(0.016)	(0.027)	(0.033)	(0.058)	(0.045)	(0.083)
Between 2,500 and 14,999	0.114***	0.000	0.155***	0.001	0.149***	0.001	0.175***	-0.001	0.264***	0.002
	(0.015)	(0.025)	(0.013)	(0.022)	(0.014)	(0.024)	(0.028)	(0.050)	(0.036)	(0.066)
Between 50,000 and 99,999	-0.034	-0.001	0.002	0.000	0.022	0.001	0.070*	0.000	-0.009	0.003
	(0.024)	(0.041)	(0.021)	(0.037)	(0.021)	(0.038)	(0.041)	(0.077)	(0.057)	(0.112)
Over 100,000	-0.136***	-0.001	-0.092***	-0.002	-0.121***	-0.000	-0.034	-0.002	-0.098***	-0.001
	(0.016)	(0.026)	(0.014)	(0.023)	(0.014)	(0.024)	(0.027)	(0.049)	(0.035)	(0.065)

Source: authors with data from the Population and Housing Census (INEGI, 2020) Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Annex V. Cont. Probability of belonging to a vulnerable group by income level. Probit model before and after the Match

	Very-Low		Low		Middle		High		Upper-High	
Variables	Before	After	Before	After	Before	After	Before	After	Before	After
Contours										
Central City	-0.016	-0.001	-0.008	-0.003	-0.112***	-0.001	-0.068**	-0.004	-0.050	-0.009
	(0.022)	(0.039)	(0.023)	(0.040)	(0.022)	(0.039)	(0.033)	(0.059)	(0.031)	(0.058)
1	-0.020	0.000	-0.049***	-0.002	-0.070***	-0.001	-0.122***	-0.002	-0.147***	-0.003
	(0.018)	(0.031)	(0.017)	(0.030)	(0.016)	(0.028)	(0.027)	(0.049)	(0.029)	(0.054)
3	0.019	0.001	-0.029**	0.000	-0.052***	-0.001	-0.038	-0.001	-0.135***	-0.002
	(0.015)	(0.024)	(0.013)	(0.022)	(0.013)	(0.022)	(0.024)	(0.042)	(0.029)	(0.051)
4	-0.159***	0.001	-0.221***	-0.001	-0.256***	-0.001	-0.192***	-0.001	-0.354***	-0.002
	(0.018)	(0.029)	(0.016)	(0.025)	(0.016)	(0.026)	(0.031)	(0.055)	(0.040)	(0.070)
Retirement income	0.080***	0.001	0.045***	-0.000	-0.003	0.001	0.011	-0.001	0.092***	0.003
	(0.011)	(0.020)	(0.014)	(0.023)	(0.014)	(0.024)	(0.023)	(0.043)	(0.024)	(0.046)
Owner/Renter	-0.018*	-0.000	0.023***	-0.001	-0.006	0.000	-0.026	-0.000	0.077***	-0.002
	(0.010)	(0.016)	(0.009)	(0.015)	(0.010)	(0.016)	(0.018)	(0.033)	(0.022)	(0.042)
Age										
between 16 and 25 years old	-0.081***	0.005	-0.078***	-0.002	-0.068***	0.008	-0.082***	0.004	-0.042	-0.004
	(0.014)	(0.024)	(0.012)	(0.021)	(0.012)	(0.021)	(0.023)	(0.042)	(0.028)	(0.050)
between 26 and 35 years old	-0.034**	0.001	-0.014	-0.001	-0.052***	0.001	-0.063***	-0.000	-0.043*	-0.004
	(0.013)	(0.023)	(0.011)	(0.018)	(0.011)	(0.020)	(0.021)	(0.038)	(0.024)	(0.045)
between 56 and 75 years old	0.144***	-0.000	0.093***	0.003	0.090***	0.001	0.064***	-0.000	0.058**	0.004
	(0.011)	(0.019)	(0.012)	(0.020)	(0.013)	(0.021)	(0.023)	(0.041)	(0.026)	(0.047)
Sex	0.005	-0.004	-0.007	0.000	-0.002	-0.001	0.014	0.001	0.014	-0.004
	(0.010)	(0.016)	(0.009)	(0.015)	(0.009)	(0.015)	(0.016)	(0.029)	(0.019)	(0.035)
Constant	-1.297***	0.161	-1.173***	0.075	-1.226***	0.044	-1.337***	-0.032	-1.447***	-0.016
	(0.054)	(0.098)	(0.038)	(0.064)	(0.038)	(0.066)	(0.069)	(0.143)	(0.087)	(0.190)
Observations	169,928	30,202	187,537	34,870	189,219	32,116	61,693	8,688	52,849	6,076
LR-test: Chi2	1813	6.131	1474	5.629	1512	4.406	454.1	2.886	527.7	2.328
ROC	0.6033	0.5040	0.5849	0.5039	0.5882	0.5038	0.5921	0.5056	0.6150	0.5071
Log of likelihood	-50764	-20931	-58199	-24167	-55095	-22259	-16163	-6021	-11929	-4210
Pseudo R2	0.0178	0.000147	0.0126	0.000116	0.0135	9.88e-05	0.0138	0.000239	0.0214	0.000276

Source: authors with data from the Population and Housing Census (INEGI, 2020) Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Annex V. Quality of the matching procedure: average values for the variables after matching

Variables	3 Vulnerable groups			Indigenous			Speak an Indigenous language			Afro-Mexican		
	Treated	Control	p>|t|	Treated	Control	p>|t|	Treated	Control	p>|t|	Treated	Control	p>|t|
No education	0,031	0,031	0,962	0,026	0,026	0,890	0,083	0,083	1,000	0,017	0,017	0,906
Basic education	0,546	0,546	0,961	0,532	0,534	0,657	0,717	0,717	0,960	0,424	0,425	0,842
Baccalaureate or middle grade education	0,243	0,243	0,746	0,254	0,254	0,960	0,128	0,128	0,893	0,288	0,288	0,960
Technical education with primary	0,015	0,015	0,947	0,019	0,018	0,711	0,006	0,005	0,838	0,017	0,016	0,680
Technical education with baccalaureate	0,012	0,011	0,183	0,013	0,012	0,496	0,004	0,004	0,908	0,016	0,015	0,580
Superior degree	0,144	0,144	0,981	0,148	0,148	0,826	0,056	0,055	0,974	0,219	0,221	0,798
Post superior education	0,009	0,009	0,861	0,008	0,008	0,889	0,007	0,006	0,709	0,018	0,017	0,488
Couples without children	0,106	0,105	0,770	0,107	0,107	0,864	0,113	0,114	0,962	0,099	0,099	0,980
Couples with children	0,503	0,504	0,556	0,481	0,481	0,935	0,631	0,631	0,950	0,487	0,488	0,832
Single parents	0,023	0,022	0,332	0,024	0,024	0,526	0,019	0,019	0,912	0,026	0,025	0,738
Single	0,369	0,368	0,901	0,388	0,388	0,995	0,236	0,237	0,944	0,388	0,387	0,901
1 & <=34 hours worked	0,002	0,002	0,726	0,003	0,002	0,803	0,003	0,002	0,647	0,003	0,002	0,445
35 & <=48 hours worked	0,128	0,128	0,954	0,132	0,131	0,658	0,111	0,109	0,648	0,127	0,128	0,946
> 48 hours worked	0,244	0,245	0,648	0,243	0,244	0,907	0,235	0,238	0,645	0,279	0,281	0,749
Employment Benefits	1,193	1,191	0,888	1,205	1,204	0,940	0,903	0,900	0,921	1,657	1,660	0,935

Notes: ***: p-value<0.01, **: p-value<0.05, *: p-value<0.1. Results obtained using the nearest neighbor procedure with only 1 neighbor, common support and caliper (0.00009) for total Vulnerable groups, (0.0002) for Indigenous non-native speakers and Afro-Mexicans, (0.0001) for Indigenous native speakers. LR-test: Likelihood-ratio test of the joint significance of all regressors. Exact matching is required on Contours, Town Size, Sex, Age, Owner/Renter condition, Retirement income and Household monthly income.

Source: authors with data from the Population and Housing Census (INEGI, 2020)

Annex VI a. Descriptive statistics of the Housing Quality Gap: The case of Mexico

Self-identified indigenous						Variables	Indigenous Native Speakers					
mean	sd	p50	min	max	N		mean	sd	p50	min	max	N
-0.07	**0.9**	**-0.05**	**-6.87**	**6.48**	**53,872**	**Total**	**-0.23**	**0.99**	**-0.2**	**-5.87**	**4.97**	**10,027**
mean	sd	p50	min	max	N	***Age***	mean	sd	p50	min	max	N
-0.1	0.91	-0.08	-5.04	4.97	10777	Between 16 and 25 years old	-0.33	1.08	-0.31	-5.21	4.97	1184
-0.08	0.88	-0.06	-4.98	6.48	10336	Between 26 and 35 years old	-0.25	1	-0.22	-5.84	4.67	2052
-0.06	0.9	-0.04	-6.87	5.53	20318	Between 36 and 55 years old	-0.22	0.97	-0.21	-5.45	4.74	4555
-0.04	0.92	-0.03	-6.75	5.62	12441	Between 56 and 75 years old	-0.18	0.95	-0.12	-5.87	4.59	2236
mean	sd	p50	min	max	N	***Level of studies completed***	mean	sd	p50	min	max	N
-0.09	1.16	-0.05	-5.09	4.89	1392	No education	-0.13	1.15	-0.1	-5.19	3.98	882
-0.07	0.99	-0.05	-6.75	6.48	28949	Basic education	-0.23	0.99	-0.22	-5.87	4.97	7316
-0.08	0.8	-0.06	-5.04	4.71	15410	High School/Technical studies	-0.3	0.91	-0.26	-5.56	4.59	1289
-0.05	0.68	-0.04	-6.87	4.78	7635	Higher education degree	-0.12	0.74	-0.07	-2.89	4.18	478
-0.04	0.65	-0.02	-3.06	2.83	418	Post higher education	-0.06	0.67	-0.06	-1.94	2.6	55
-0.04	0.85	-0.09	-1.86	2.58	68	Not specified	-0.06	1.27	-0.11	-1.9	2.32	7
mean	sd	p50	min	max	N	***Sex***	mean	sd	p50	min	max	N
-0.07	0.9	-0.06	-6.75	5.62	28297	Woman	-0.23	0.99	-0.2	-5.84	4.74	5226
-0.06	0.9	-0.04	-6.87	6.48	25575	Man	-0.23	0.99	-0.21	-5.87	4.97	4801
mean	sd	p50	min	max	N	***Occupation status***	mean	sd	p50	min	max	N
-0.1	0.92	-0.07	-6.87	5.62	15761	Renting	-0.28	0.97	-0.27	-5.79	4.97	4215
-0.06	0.89	-0.04	-6.75	6.48	38111	Ownership	-0.19	1	-0.14	-5.87	4.74	5812
mean	sd	p50	min	max	N	***Vulnerable group***	mean	sd	p50	min	max	N
-0.07	0.9	-0.05	-6.87	6.48	53872	Self-identified indigenous	.	.	.	.	.	0
.	.	.	.	.	0	Speak an indigenous language	-0.23	0.99	-0.2	-5.87	4.97	10027
.	.	.	.	.	0	Afro-Mexican	.	.	.	.	.	0
mean	sd	p50	min	max	N	***Type of cohabitation***	mean	sd	p50	min	max	N
-0.02	0.73	-0.02	-4.17	4.2	5672	Couples without children	-0.19	0.85	-0.15	-5.04	4.29	1082
-0.08	0.89	-0.06	-5.35	6.48	26276	Couples with children	-0.23	0.98	-0.2	-5.87	4.74	6448
-0.09	0.92	-0.07	-4.56	3.47	1219	Single parents	-0.09	0.99	-0.15	-3.2	4.4	179
-0.07	0.96	-0.05	-6.87	5.62	20705	Single	-0.25	1.05	-0.25	-5.79	4.97	2318
mean	sd	p50	min	max	N	***Number of members in the household***	mean	sd	p50	min	max	N
0.34	1.28	0.4	-6.87	6.48	2427	Live alone	0.08	1.16	0.08	-4.6	3.74	429
0.08	0.91	0.12	-4.96	5.53	7172	Two people	-0.05	1.03	-0.02	-5.87	4.67	1202
-0.06	0.85	-0.04	-5.95	5.09	31325	3 to 4 people	-0.24	0.95	-0.21	-5.84	4.61	5800
-0.22	0.88	-0.18	-6.75	4.99	11742	5 to 10 people	-0.33	0.97	-0.29	-5.09	4.74	2391
-0.41	0.85	-0.35	-4.99	2.64	1206	Over 10 people	-0.45	1.17	-0.38	-5.45	4.97	205
mean	sd	p50	min	max	N	***Some disability***	mean	sd	p50	min	max	N
-0.06	0.9	-0.05	-6.87	6.48	50342	No	-0.23	0.99	-0.2	-5.87	4.97	9444
-0.12	0.98	-0.1	-5.04	5.33	3530	Yes	-0.24	1.02	-0.18	-4.3	4.29	583

Source: authors with data from the Population and Housing Census (INEGI, 2020)

Annex VI b. Descriptive statistics of the Housing Quality Gap: The case of Mexico

Self-identified indigenous						Variables	Indigenous Native Speakers					
mean	sd	p50	min	max	N		mean	sd	p50	min	max	N
-0.07	**0.9**	**-0.05**	**-6.87**	**6.48**	**53,872**	**Total**	**-0.23**	**0.99**	**-0.2**	**-5.87**	**4.97**	**10,027**
mean	sd	p50	min	max	N	***Receive retirement income***	mean	sd	p50	min	max	N
-0.08	0.93	-0.06	-6.87	6.48	46529	No	-0.24	1	-0.22	-5.84	4.97	9370
-0.02	0.72	-0.01	-4.44	4.46	7343	Yes	-0.12	0.82	-0.05	-5.87	4.29	657
mean	sd	p50	min	max	N	***Number of remittances***	mean	sd	p50	min	max	N
-0.07	0.9	-0.05	-6.87	6.48	50719	Receive no remittances	-0.24	0.99	-0.21	-5.87	4.97	9580
0	0.91	0	-4.44	4.98	3003	1 Other regions or foreign countries	-0.06	0.95	-0.1	-3.31	4.61	418
-0.02	0.82	-0.03	-2.97	2.58	150	2 Other regions and foreign countries	0.09	0.8	0.04	-1.49	1.71	29
mean	sd	p50	min	max	N	***Social program income***	mean	sd	p50	min	max	N
-0.06	0.9	-0.04	-6.87	6.48	40656	No	-0.24	1	-0.21	-5.87	4.97	8089
-0.10	0.89	-0.09	-4.99	5.62	13210	Yes	-0.19	0.93	-0.17	-4.39	4.41	1938
mean	sd	p50	min	max	N	***Received a cheap housing loan***	mean	sd	p50	min	max	N
0.13	0.63	0.06	-3.16	5.29	4577	For workers on a medium income level (INFONAVIT or FOVISSSTE)	0.14	0.69	0.04	-2.7	4.67	430
0.23	0.55	0.18	-0.73	1.65	26	For all PEMEX workers	0.27	.	0.27	0.27	0.27	1
-0.05	0.61	-0.07	-1.93	1.63	124	For poor people (FONHAPO)	-0.14	0.8	-0.18	-1.68	2.38	21
-0.09	0.92	-0.07	-6.87	6.48	49145	No support (reference)	-0.24	1	-0.22	-5.87	4.97	9575
mean	sd	p50	min	max	N	***Type of dwelling and floor***	mean	sd	p50	min	max	N
.	.	.	.	.	0	Vecindad/Dirt floor	.	.	.	.	.	0
-0.14	0.88	-0.23	-1.46	2.38	32	Roof top/Cement or mosaic floor	-0.66	0.28	-0.66	-1.14	-0.25	14
-0.33	0.86	-0.33	-3.58	4.08	407	Vecindad/Cement floor	-0.47	0.81	-0.56	-3.15	1.95	261
-0.05	0.76	-0.11	-3.16	3.15	241	Vecindad/Mosaic floor	-0.02	0.74	-0.15	-1.26	3.45	81
-1.33	1.42	-1.22	-5.86	3.61	1020	House/Dirt floor	-1.49	1.52	-1.32	-5.79	2.41	431
0.11	0.69	0.03	-2.66	3.58	900	Apartment/Cement floor	0.08	0.81	-0.01	-2.7	4.59	240
0.16	0.64	0.1	-3.52	5.53	3608	Apartment/Mosaic floor	0.05	0.66	0	-1.91	4.67	459
-0.16	0.94	-0.16	-6.87	6.48	32663	House/Cement floor	-0.27	0.96	-0.27	-5.87	4.61	6445
0.17	0.72	0.11	-4.44	5.62	14788	House/Mosaic floor	0.07	0.76	0.03	-3.53	4.74	1982
mean	sd	p50	min	max	N	***Centre - outskirts***	mean	sd	p50	min	max	N
-0.06	0.7	-0.04	-3.84	4.38	2697	City centre	-0.22	0.79	-0.22	-3.98	4.18	537
-0.05	0.72	-0.03	-6.87	4.78	5801	Contour 1	-0.13	0.76	-0.13	-4.54	3.76	1007
-0.09	0.89	-0.05	-5.35	5.59	7606	Contour 2	-0.22	0.99	-0.22	-4.95	4.97	1704
-0.09	0.91	-0.07	-6.75	5.53	19155	Contour 3	-0.28	1.03	-0.23	-5.84	4.74	4467
-0.05	0.97	-0.03	-5.64	6.48	18613	Contour 4	-0.17	1.02	-0.14	-5.87	4.67	2312
mean	sd	p50	min	max	N	***Town Size***	mean	sd	p50	min	max	N
-0.04	1.14	-0.04	-4.98	5.59	6975	Below 2,500	-0.25	1.31	-0.23	-5.84	4.97	1411
-0.04	0.99	-0.04	-5.09	6.48	11070	Between 2,501 and 15,000	-0.21	1.08	-0.17	-5.87	4.67	1647
-0.06	0.89	-0.05	-6.75	5.29	8114	Between 15,000 and 49,999	-0.23	0.95	-0.2	-5.56	4.41	1241
-0.04	0.73	-0.02	-3.44	4.46	2110	Between 50,000 and 99,999	-0.21	0.71	-0.19	-2.24	2.17	383
-0.09	0.8	-0.06	-6.87	5.53	25603	Over 100,000	-0.23	0.88	-0.2	-5.12	4.61	5345
-0.07	**0.9**	**-0.05**	**-6.87**	**6.48**	**53,872**	**Total**	**-0.23**	**0.99**	**-0.20**	**-5.87**	**4.97**	**10,027**

Source: authors with data from the Population and Housing Census (INEGI, 2020)

Annex VI c. Descriptive statistics of the Housing Quality Gap: The case of Mexico

Afro-Mexican						Variables	All three vulnerable groups					
mean	sd	p50	min	max	N		mean	sd	p50	min	max	N
-0.01	**0.86**	**0**	**-4.88**	**7.65**	**9,441**	**Total**	**-0.08**	**0.91**	**-0.07**	**-7.50**	**6.53**	**72,107**
mean	sd	p50	min	max	N	***Age***	mean	sd	p50	min	max	N
-0.03	0.87	0	-4.81	4.11	1973	Between 16 and 25 years old	-0.10	0.93	-0.09	-5.32	5.51	13759
-0.02	0.82	-0.01	-4.69	4.58	2141	Between 26 and 35 years old	-0.09	0.89	-0.08	-4.99	6.53	14197
0.01	0.85	0	-4.88	4.87	3743	Between 36 and 55 years old	-0.08	0.9	-0.06	-7.5	5.34	28145
0	0.92	0.01	-4.41	7.65	1584	Between 56 and 75 years old	-0.06	0.92	-0.05	-6.23	5.38	16006
mean	sd	p50	min	max	N	***Level of studies completed***	mean	sd	p50	min	max	N
-0.13	1.17	0.02	-4.69	3.03	182	No education	-0.11	1.18	-0.1	-5.11	5.34	2351
-0.03	0.96	-0.03	-4.81	7.65	4024	Basic education	-0.10	0.98	-0.08	-7.5	6.53	39846
0	0.81	0	-4.29	4.55	3006	High School/Technical studies	-0.07	0.82	-0.06	-5.73	5.51	19418
0.03	0.67	0.02	-4.88	4.58	2015	Higher education degree	-0.04	0.67	-0.03	-6.23	4.43	9799
0.04	0.65	0.04	-2.39	2.72	197	Post higher education	0	0.69	-0.01	-2.83	4.08	602
-0.25	1.09	-0.09	-1.98	2.19	17	Not specified	-0.27	0.86	-0.17	-2.72	2.22	91
mean	sd	p50	min	max	N	***Sex***	mean	sd	p50	min	max	N
-0.01	0.84	-0.01	-4.77	4.86	4822	Woman	-0.09	0.9	-0.07	-6.14	5.51	37757
0	0.87	0	-4.88	7.65	4619	Man	-0.07	0.91	-0.06	-7.5	6.53	34350
mean	sd	p50	min	max	N	***Occupation status***	mean	sd	p50	min	max	N
-0.01	0.91	-0.01	-4.81	4.87	3091	Renting	-0.11	0.93	-0.1	-7.5	5.59	22628
-0.01	0.83	0	-4.88	7.65	6350	Ownership	-0.07	0.89	-0.05	-6.23	6.53	49479
mean	sd	p50	min	max	N	***Vulnerable group***	mean	sd	p50	min	max	N
.	.	.	.	.	0	Self-identified indigenous	-0.07	0.9	-0.05	-7.5	6.53	53230
.	.	.	.	.	0	Speak an indigenous language	-0.23	0.98	-0.21	-5.32	5.34	9882
-0.01	0.86	0	-4.88	7.65	9441	Afro-Mexican	-0.02	0.82	-0.01	-5.11	5.51	8995
mean	sd	p50	min	max	N	***Type of cohabitation***	mean	sd	p50	min	max	N
-0.02	0.65	-0.01	-3.52	3.16	915	Couples without children	-0.05	0.74	-0.04	-4.81	4.53	7435
0	0.84	0	-4.81	4.87	4664	Couples with children	-0.1	0.9	-0.07	-5.87	6.53	36975
0	0.82	-0.04	-3.05	2.52	261	Single parents	-0.1	0.92	-0.08	-4.25	3.58	1553
-0.01	0.93	0	-4.88	7.65	3601	Single	-0.07	0.96	-0.06	-7.5	5.59	26144
mean	sd	p50	min	max	N	***Number of members in the household***	mean	sd	p50	min	max	N
0.57	1.23	0.64	-4.32	7.65	415	Live alone	0.33	1.26	0.39	-7.5	6.53	3197
0.18	0.86	0.19	-4.69	4.22	1248	Two people	0.08	0.93	0.11	-6.23	5.51	9445
0	0.8	0.01	-4.88	4.87	5659	3 to 4 people	-0.08	0.86	-0.05	-6.14	5.38	42078
-0.22	0.84	-0.19	-4.2	4.85	1865	5 to 10 people	-0.24	0.88	-0.22	-5.32	4.94	15758
-0.39	0.9	-0.33	-3.34	2.52	254	Over 10 people	-0.42	0.9	-0.36	-5.21	5.03	1629
mean	sd	p50	min	max	N	***Some disability***	mean	sd	p50	min	max	N
0	0.86	0	-4.88	7.65	8990	No	-0.08	0.9	-0.06	-7.5	6.53	67620
-0.06	0.84	0	-3.52	3.25	451	Yes	-0.12	0.97	-0.1	-6.63	5.38	4487

Source: authors with data from the Population and Housing Census (INEGI, 2020)

Annex VI d. Descriptive statistics of the Housing Quality Gap: The case of Mexico

Afro-Mexican						Variables	All three vulnerable groups					
mean	sd	p50	min	max	N		mean	sd	p50	min	max	N
-0.01	**0.86**	**0**	**-4.88**	**7.65**	**9,441**	**Total**	**-0.08**	**0.91**	**-0.07**	**-7.50**	**6.53**	**72,107**
mean	sd	p50	min	max	N	***Receive retirement income***	mean	sd	p50	min	max	N
-0.01	0.88	0	-4.88	7.65	8239	No	-0.09	0.93	-0.07	-7.5	6.53	63127
0.01	0.69	0	-4.32	4.52	1202	Yes	-0.03	0.73	-0.02	-5.73	5.38	8980
mean	sd	p50	min	max	N	***Number of remittances***	mean	sd	p50	min	max	N
-0.01	0.86	0	-4.88	4.87	8810	Receive no remittances	-0.09	0.91	-0.07	-7.5	6.53	67955
0.03	0.9	0.01	-3.16	7.65	590	1 Other regions or foreign countries	0	0.88	0	-4.67	4.52	3934
0.18	0.82	0.18	-1.25	2.61	41	2 Other regions and foreign countries	0.05	0.77	0.07	-2.45	2.62	218
mean	sd	p50	min	max	N	***Social programme income***	mean	sd	p50	min	max	N
0.01	0.87	0.01	-4.81	7.65	7527	No	-0.08	0.91	-0.06	-7.5	6.53	55306
-0.09	0.82	-0.07	-4.88	4.87	1914	Yes	-0.1	0.88	-0.08	-6.63	5.38	16795
mean	sd	p50	min	max	N	***Received a cheap housing loan***	mean	sd	p50	min	max	N
0.13	0.66	0.05	-4.88	3.98	1201	For workers on a medium income level (INFONAVIT or FOVISSSTE)	0.13	0.65	0.07	-5.73	5.51	6107
0.31	0.33	0.2	0.07	0.77	4	For all PEMEX workers	0.29	0.49	0.3	-1.19	1.32	31
-0.14	0.45	-0.09	-1.17	0.83	37	for poor people (FONHAPO)	-0.02	0.64	-0.11	-1.52	3.21	183
-0.03	0.88	-0.01	-4.81	7.65	8199	No support (reference)	-0.1	0.92	-0.08	-7.5	6.53	65786
mean	sd	p50	min	max	N	***Type of dwelling and floor***	mean	sd	p50	min	max	N
.	.	.	.	.	0	Vecindad/Dirt floor	.	.	.	.	.	0
-0.04	2.01	-0.04	-1.46	1.38	2	Roof top/Cement or mosaic floor	-0.28	0.68	-0.38	-1.69	2.38	47
-0.13	0.84	-0.05	-2.11	2.21	54	Vecindad/Cement floor	-0.34	0.84	-0.36	-3.54	2.71	711
0.01	0.84	-0.07	-3.05	1.81	47	Vecindad/Mosaic floor	-0.09	0.77	-0.13	-2.98	4.92	359
-1.19	1.44	-1.06	-4.81	2.39	115	House/Dirt floor	-1.35	1.43	-1.24	-7.5	3.87	1551
0.13	0.73	0.09	-1.63	3.88	238	Apartment/Cement floor	0.1	0.69	0.01	-2.66	4.59	1344
0.11	0.61	0.05	-2.81	3.36	1077	Apartment/Mosaic floor	0.14	0.64	0.08	-3.52	4.52	5025
-0.13	0.93	-0.12	-4.32	4.87	4819	House/Cement floor	-0.17	0.93	-0.18	-6.23	6.53	43264
0.17	0.72	0.1	-4.88	7.65	3056	House/Mosaic floor	0.16	0.73	0.09	-5	5.38	19450
mean	sd	p50	min	max	N	***Centre - outskirts***	mean	sd	p50	min	max	N
-0.06	0.74	-0.02	-4.3	2.52	687	City centre	-0.08	0.72	-0.06	-4.7	4.25	3809
0	0.67	-0.01	-4.32	3.36	1164	Contour 1	-0.04	0.73	-0.04	-6.23	5.38	7834
-0.03	0.8	-0.02	-3.81	4.85	1255	Contour 2	-0.1	0.88	-0.07	-5.51	5.59	10314
0	0.87	0.01	-4.88	4.71	3348	Contour 3	-0.11	0.93	-0.09	-6.63	5.38	26513
0	0.95	0.01	-4.29	7.65	2987	Contour 4	-0.05	0.97	-0.05	-7.5	6.53	23637
mean	sd	p50	min	max	N	***Town Size***	mean	sd	p50	min	max	N
-0.05	1.23	-0.04	-4.2	7.65	558	Below 2,500	-0.08	1.16	-0.09	-7.5	5.59	8813
-0.01	1.01	-0.01	-4.81	4.87	1577	Between 2,501 and 15,000	-0.06	1	-0.05	-5.49	6.53	14061
0.01	0.91	0.01	-4.29	4.86	1775	Between 15,000 and 49,999	-0.07	0.9	-0.06	-5.25	5.51	10937
0.05	0.74	0.05	-2.94	2.97	400	Between 50,000 and 99,999	-0.06	0.73	-0.04	-3.76	3.21	2809
-0.01	0.74	-0.01	-4.88	4.85	5131	Over 100,000	-0.1	0.8	-0.07	-6.63	5.38	35487
-0.01	**0.86**	**0**	**-4.88**	**7.65**	**9,441**	**Total**	**-0.08**	**0.91**	**-0.07**	**-7.50**	**6.53**	**72,107**

Source: authors with data from the Population and Housing Census (INEGI, 2020)