Volume 13, Number 1, 2026

REGION

The Journal of ERSA
Powered by WU

Table of Contents

Articles

Funded by

REGION, The journal of ERSA / Powered by WU
ISSN: 2409-5370

ERSA: http://www.ersa.org
WU: http://www.wu.ac.at

All material published under the Creative Commons Attribution License. See http://creativecommons.org/licenses/by/4.0/ for the full conditions of the license.
The copyright of all the material published in REGION is with the authors

Articles

REGION
The Journal of ERSA
Powered by WU

Volume 13, Number 1, 2026, 1–21
DOI: 10.18335/region.v13i1.484

journal homepage: region.ersa.org
ISSN: 2409-5370

A systematic literature review on the relations between "firm-region nexus" and firm productivity

Pierre-François Wilmotte[1], Didier Van Caillie[1], Isabelle Reginster[2], Marcus Dejardin[3] Jean-Marie Halleux[1]

[1] Université de Liège, Liège, Belgium
[2] IWEPS, Namur, Belgium
[3] University of Namur, Namur, Belgium; UCLouvain, Louvain-la-Neuve, Belgium

Received: 1 March 2023/Accepted: 29 January 2026

Abstract. An increasing number of studies focus on the impact of firm location on firm productivity. Here, *location* refers to localised resources or externalities available to firms. These studies have become possible due to advancements in firm-level data availability and productivity estimation methods. There is a growing need to better identify the effects of localised externalities – "the firm-region nexus" – on firms, which are the primary targets of many regional development policies. However, relying solely on administrative regions risks committing the ecological fallacy: attributing to all firms in a region the effects observed for a region as a whole.

This article presents the state of the art as of 2025, highlighting key issues in this area of research. A total of 165 articles were selected, mainly published in the past decade. From a methodological perspective, there is no real consensus on the models utilised, whether to estimate productivity or to investigate the relationships between the "firm-region nexus" and firm productivity. Spatial integration is still not adequately taken into account: (a) Total Factor Productivity estimations do not account for spatial dependence among methodological concerns; (b) 70% of the articles do not discuss methodological issues linked to the use of firm locations. A deeper grasp of firm location, particularly the "head office bias", emerges as critical to improving the robustness of analyses.

Quantifying the "firm-region nexus" remains heterogeneous across national and local contexts (diverging points of interest, data availability, etc.). Comparing the effects of various types of externalities across countries therefore appears ambitious. Some articles focus on the effect of one category of localised externalities, aiming not only to identify a relationship but also its type: spatial or temporal effect, linear or non-linear relationships, and threshold effects.

JEL classification: D24, O47, R58, R32, R11

Key words: firm-level, firm productivity, total factor productivity, firm location

1 Introduction

Understanding the potential contribution of the "firm-region nexus" to the productivity of existing firms is not only a common research concern in regional sciences but also a central political issue for public authorities. The “firm-region nexus” refers to localised externalities that provide firms with a local advantage.

Firms take advantage of these localised externalities. Consequently, firms are targeted by public policies that enhance the “firm-region nexus” or localised externalities. However, most existing research about the effects of this “firm-region nexus” is carried out at the aggregate statistical unit level, such as the NUTS classification in Europe. Drawing conclusions regarding regional development strategies in these studies exposes decision-makers to the ecological fallacy (Openshaw 1984, Fotheringham, Wong 1991, Wrigley 1995). Since relationships seen at higher levels may not always mirror realities at the individual level (in this case, firm level), the ecological fallacy is a mistake that can result in misunderstandings and poor decisions. Public policies can also benefit from individual-level research (Duranton 2011). Indeed, firms are the target group of most public policies in regional development in order to boost socio-economic development of local communities as the ultimate beneficiaries.

The development of research on relationships between territorial resources and productivity measured at the firm level depends on four conditions: the availability of reliable accounting data, reliable information on the location of activities, robust indicators to quantify productivity, and quantification of the “firm-region nexus”.

First, access to business databases varies from country to country. The quality of, and access to, data depends on the institutions, public and private, that collect the information, the issue of data protection, and the amount of information requested from businesses. These influence the distribution of case studies around the world.

Second, it remains difficult to link the creation of added value by a firm to a specific location: accounting data usually are available at the legal-unit level, implying that added value is located at headquarters. However, headquarters are not necessarily the main place where value is produced or where a firm benefits most from localised externalities.

Third, the productivity of a firm can be measured using various indicators. In the 1950s, the “Solow residual” was conceptualised to compare the productivity of firms regardless of differences in activities (Abramowitz 1956, Griliches 1996, 1998, Hulten 2001, Solow 1957, Van Beveren 2012, Van Ark 2014, Nadiri 1970). The operationalisation of the concept, better known as Total Factor Productivity (TFP), is now widespread. The various techniques for estimating TFP seek to deal with classical biases in econometrics, including price omission bias, simultaneity bias, endogeneity of explanatory variables (capital and labour), and selection bias linked to the use of panel data (Van Beveren 2012).

Fourth, quantifying “firm-region nexus” remains a methodological challenge due to data availability or the level of details to capture spatial variability due to, for instance, agglomeration economies. Some authors quantify agglomeration economies by proxies such as population or firm density, whereas other authors quantify agglomeration economies through sophisticated variables designed to capture different forms such as localisation, urbanisation, or competition economies.

In the 2010s, a growing number of studies sought to identify the interactions within the “firm-region nexus” and their effects on productivity measured at the firm level, seeking to integrate the aforementioned issues. This article reports on the Systematic Literature Review (SLR) implemented to identify 165 articles constituting the most exhaustive literature review possible. The SLR exploits the content of the 165 identified articles to highlight the methodological issues and to present the state of knowledge on the interactions between territorial resources and firm productivity at firm level.

The article is organised as follows. Section 2 discusses the methodology used to select and analyse the 165 articles. Section 3 analyses the metadata of 165 selected articles (date of publication, location of case study, authors…). Section 4 discusses three important factors in the methods used to analyse the 165 articles: (i) the sample or statistical population studied (size and types of firms); (ii) the methods to estimate productivity and the impact of “firm-region nexus”; and (iii) the variables quantifying this “firm-region nexus”. Section 5 presents the state of knowledge on the relationship between the “firm-region nexus” and firm productivity. Finally, Section 6 concludes.

Table 1: Query formulated in Scopus for the systematic literature review

```
ALL ( "plant-level" OR "firm-level" ) AND TITLE-ABS-KEY ( ( "firm" OR
"enterprise" OR "plant" ) W/5 ( "productivity" OR "TFP" OR "Total factor
productivity" ) ) AND TITLE-ABS-KEY ( "regional" OR "location" OR
"territorial" OR "localization" OR "agglomeration" OR "city" OR "spatial"
OR "place-based" OR "proximity" OR "spatial clustering" OR "local" OR
"distance" ) AND NOT ALL ( "FDI" AND "foreign direct investment" ) AND (
LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) )
```

2 Method and Data

This section presents the method and data collected to conduct the Systematic Literature Review (SLR). The first subsection defines the SLR and details how it was applied. The second subsection details the data collected in each article considered.

2.1 SLRs: definition and application

Initially, SLRs were developed and used in the life sciences. The main advantage of SLR is that it provides a comprehensive inventory of the literature. It must meet five qualities: be reproducible, transparent, objective, unbiased, and rigorous. The selection of keywords, the chosen databases, and the interpretations are all choices that must be justified by the researchers (Boell, Cecez-Kecmanovic 2015).

The list of keywords comprises three categories that were to be represented in the selected articles (see Table 1). The first category corresponds to a clear mention (anywhere in the article) to an analysis conducted at a firm-level (and not at regional level). The second category refers to some words describing a clear mention of spatial and/or geographical dimensions. To select papers on the spatial dimension of firm productivity, these words must be in the title, abstract or keywords of each article. In the third category, an association must be made between a firm (or two main synonyms) and its productivity. This ensures that the selected articles focus on the firm-region (productivity) nexus issue. No specific period has been included in the query.

The query was conducted on December 12, 2023, using the *Scopus* search engine (see Table 1). *Scopus* is preferred over other search engines based on the two arguments made by Mongeon, Paul-Hus (2016). First, *Scopus* inventories a larger number of scientific journals, particularly in the social sciences. Second, the diversity of origins of scientific journals is broader in *Scopus* compared to other search engines like the *Web of Science*. It should be noted that *Google Scholar* was not used because it does not distinguish between peer review articles and the 'grey' literature.

The initial query identifies 920 relevant documents. We introduced two specific codes to limit the selection of articles:

- The first limits the selection to articles (to ensure the same quality standards) published in English as scientific *lingua franca*, reducing the number of documents to 839; this may impact the location of case studies.
- The second one excludes articles mentioning Foreign Direct Investment (or its acronym FDI) because it appeared that these articles do not use the spatial or geographic dimension to analyse the "firm-region nexus": either the relevant scale is national in these papers or the subnational (regional or local) scale is analysed as control variable. It reduces the number of documents to 574.

The 574 abstracts were reviewed to select relevant articles, analysing the relations between territorial resources and firm productivity. To analyse the "firm-region nexus", selected articles must fulfil the following conditions:

- Be an empirical analysis, excluding purely state-of-the-art or theoretical analysis.
- Analyse primarily the territorial resources as defined above.

- Demonstrate an analysis that includes a spatial or geographic dimension (for instance, the spatial impact of an infrastructure but not the economic impact of a national or regional infrastructure program).
- Be focused on manufacturing or service firms, excluding (thanks to Scopus research engine) articles focused on agriculture, retail or tourism firms. The location of these firms is linked to other specific location factors (high streets in city, tourist attractions, geographic conditions).
- Be focused on externalities, and not on internal characteristics of firms such as type of ownership or firm level of R&D investment.
- Not have as main subject the effect of national or regional regulations in the environmental or fiscal fields.
- Not have as main subject international trade, outsourcing, and/or export ability, because the spatial dimension refers, in these cases, to national or international scales.

The reading of each abstract resulted in a selection of 216 articles. Subsequently each article was read in its entirety: a selection of 165 articles resulted[1]. These articles contain some models explaining the firm productivity by territorial resources with or without internal characteristics. Based on the different criteria, the systematic literature review is based on these 165 articles; at https://orbi.uliege.be/bitstream/2268/340548/-2/Annex_SLR.xlsx is the complete list.

2.2 *Collection of data from articles included in the SLR*

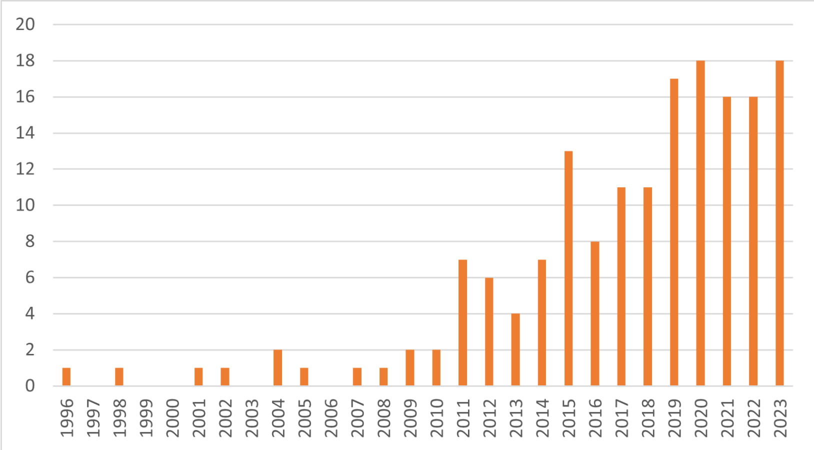
Various information was collected for each of the 165 identified articles. The metadata collected via *Scopus* include author(s), year of publication, and journal name. Data about articles was processed based on four categories.

The first category of data refers to sample details. These details include the type of firms (for instance, manufacturing or services firms), the sample size (i.e., the number of firms analysed), and the years of observation.

The second category of data refers to location information. These details include area where firms are located, the level of accuracy of firm locations, and the inclusion of information on single or multiple establishments of firms. The level of accuracy of firm location is derived from the European NUTS typology. We define seven scales. Four scales are derived from NUTS typology for regional administrative units, from NUTS 0 (country) to NUTS 3 according to the population size of study area. Eurostat, the European Office for Statistics, offers the Local Administrative Units (LAU) to analyse local level data. It corresponds to local authorities, communes, or municipalities. We add two complementary scales: local labour market, usually defined on a scientific basis, and city without any other information provided by the paper authors.

The third category refers to the analysis of relations between variables quantifying the “firm-region nexus” and firm productivity. First, indicators of firm productivity are collected (including methods to estimate productivity). Second, models quantifying the “firm-region nexus” (including the direction of the relationship with the indicators of productivity) are analysed. Most articles include several models, between one to 15 models published per article. It is important to avoid the over-representation of results published in articles testing similar models. That is why, in these cases, models considered are either the baseline model with the most variables or several models in cases where a single model did not include all the variable categories. Based on this selection, data are collected about the statistical methods, variables quantifying the relations between “firm-region nexus” and firm productivity including the basic results (positive, negative, both, or non-significant).

[1] One article was retracted on 9 Oct 2025 during the review process of this article, the retracted article was removed from the final selection.



Source: Scopus

Figure 1: Number of articles identified by SLR by year of publication

Explanatory variables are collected and then classified according to an adaptation of a typology by Tsvetkova et al. (2020). These authors have realised a large state-of-the-art of the "firm-region nexus" based on workshops and literature review within the OECD Spatial Lab. They identify twelve categories of territorial resources with a spatial dimension on firm productivity at subnational scales. These twelve categories are grouped as follows: first, agglomeration economies, including five components (localisation economies, urbanisation economies, competition economies, size and efficiency of local labour markets and global index of agglomeration economies); second, knowledge and ecosystems, including R&D, knowledge diffusion, entrepreneurship and human capital; third, finance and governance, including finance, institutions, governance and policies and regulations; fourth, infrastructure and settlement, including demography, geography and infrastructures; and fifth, local economic structure, including the effect of specific categories of firms on productivity (e.g. local share of large firms).

3 An emerging field of research in a growing number of countries

The question of the relationship between "firm-region nexus" and firm productivity is an emerging theme, as shown by the publication dates of the 165 analysed articles. Almost all the articles were published since the 2010s (see Figure 1). Since 2019, each year has seen more than 15 articles published. Two elements support the observation that this is an emerging topic. On the one hand, no single journal has established itself as a reference for publication on the research theme. No less than 91 different journals have published articles. These journals are either specialised in specific economic themes (productivity, transport geography) or specialised in regional sciences. Six journals published more than five articles in the sample: *Regional Studies*, *Annals of Regional Science*, *Regional Science and Urban Economics*, *Journal of Urban Economics* and *Sustainability*.

Among the 391 different authors of 165 articles, 9% have published more than one article included in this SLR. These authors have published about the same study case: sometimes about different samples of firms but always of the same country. Company databases comprise balance sheets of legal units. However, legal units' definition differs across national laws. Moreover, a firm is sometimes an association of several legal units, united by a distinctive organisational logic. The concept of firm is broader than legal unit: it remains ill-defined entity both in terms of its boundaries and its content (Maskell 2001), due to different viewpoints.

Moreover, there are major differences across countries about available firm data: private, semi-private, or public producers of databases alongside differences between national accounting systems and various methodological choices. The lack of harmonised data across countries means it remains difficult to carry out analyses on several countries

Table 2: Location of study cases of the 165 articles of the SLR

Continents	Countries with more than 3 articles
Europe – 82 articles	Italy – 27 articles United Kingdom – 9 articles France – 6 articles Spain – 5 articles Germany and Sweden – 4 articles Belgium, Portugal, and studies on several European countries – 3 articles
Asia – 63 articles	China – 39 articles South Korea – 4 articles India and Vietnam – 3 articles
America – 9 articles	United States of America – 5 articles
Africa – 8 articles	
Oceania – 3 articles	
Intercontinental – 1 article	

Source: Own elaboration

or for authors to publish on different countries. Researchers face specificities related to the accounting data collected and made available (see Siepel, Dejardin 2020).

The Bureau Van Dijk database, including the Europe's largest firms, or some surveys conducted among firms in several African countries, allows simultaneous analyses of several countries. These two databases are used in 7 of the 165 articles. Further, 142 articles analyse a sample or a population of firms from one country, while 15 articles focus on firms from a part of a country, typically China or Italy.

The study cases are mainly located in China (39 articles) and Italy (27 articles) (see Table 2). Apart from these two special cases, 10 countries are the case studies of more than 3 articles of this SLR (7 in Europe, 2 in Asia and 1 in America). There are several hypotheses explaining this inequal repartition of study cases across the world: first, uneven regional development is a common issue in these 12 countries; second, data availability and robustness is a crucial issue to allow research; third, the spread of English as a scientific lingua franca is not the same throughout the World. For instance, 24 articles out of 39 articles about Chinese firms are based on a firm census conducted during the 2000s.

4 How has research been conducted?

4.1 Selecting firm sample or population to quantify productivity

Most research evaluating the impact of "firm-region nexus" on firm productivity focuses on manufacturing firms (see Table 3). Most exclude some firms for several reasons: first, firm data quality remains a real issue. There are few data quality checks for firms' databases e.g., errors, omissions, omitted relations between legal units. Second, data accuracy for the smallest firms can be weaker because they do not need to report all accounting data (with variations in what must be published depending on the country). Third, some authors focus on market-oriented firms to avoid firms with specific strategies such as public owned firms. Fourth, and finally, some authors focus their research on a specific sector.

Three indicators emerge from the multitude of indicators that measure firm productivity (OECD 2020): labour productivity, wages, outputs and TFP. TFP has the advantage of allowing for comparison *all other things being equal*, but also the disadvantage of being more complex to estimate due to numerous methodological difficulties (Gal 2013, Gordon et al. 2015, OECD 2020). TFP is the most common productivity indicator among the SLR (see Table 4).

A diversity of estimation methods is observed due to the absence of a method that simultaneously deals with all methodological issues (see Van Beveren 2012). Some authors

Table 3: Firm sample or population analysed by the 165 articles of the SLR

Sectoral focus	Without exclusion	With exclusion based on accounting data (e.g., more than x full-time equivalent, amount minimum of sales. . .)	With exclusion based on economic sector (e.g., without financial firms with specific accounts. . .)	Focus on specific group(s) of firms (e.g., sector or technology level. . .)	Total
None (all firms)	30	21	2	12	66
Manufacturing	32	32	17	10	91
Services			1	6	7
Total	63	53	20	28	163

Source: Own elaboration
Notes: No data: 2 articles

Table 4: Frequency of use of each productivity indicator in SLR

Indicator	Number of articles in which the indicator is used
Total Factor Productivity	104
Added value	53
Outputs	13
Wages	2
No clear information	1

Source: Own elaboration

(19% of the 165 articles) favour statistical methods like ordinary least squares estimators (OLS), fixed effects, models including instruments, or generalised moment models (GMM) (see Table 4). OLS is used despite the lack of robustness because it remains easy to use and to interpret (Ehrl 2013).

Other authors (55% of the 165 articles) use *ad hoc* estimation methods discussed by Van Beveren (2012). These *ad hoc* methods are the methods of Olley, Pakes (1996), Levinsohn, Petrin (2003), Ackerberg et al. (2015), and Wooldridge (2009). These two last methods are the published version or the last version of working papers used by Van Beveren. These *ad hoc* methods account for about half of the 140 different TFP estimates identified among the 104 articles (see Table 5).

The lack of consideration in econometric research for phenomena with a spatial dimension (Anselin 2010, Tsvetkova et al. 2020) is evident in the methods used to estimate TFP. However, there is a large body of work demonstrating the existence of spatial dependence between firms. In Belgium, Dhyne, Duprez (2016) show that the probability

Table 5: TFP estimation methods across the 165 articles

TFP estimation method		Number of concerned articles
Ad hoc methods identified by Van Beveren (2012)	Olley, Pakes (1996)	23
	Levinsohn, Petrin (2003)	47
	Ackerberg et al. (2015)	14
	Wooldridge (2009)	6
Estimation of TFP by generic methods	OLS	14
	Fixed effects	6
	Use of instruments	9
	GMM	3
	Specific method (with development in article)	2
	Other	10
	No information	6

Source: Personal elaboration

of a transaction between two firms decreases as the distance between them increases. Surprisingly, the discussion of Van Beveren (2012) on the methodological issues, does not mention spatial autocorrelation as a potential bias. This element testifies to the existence of a gap between the research conducted in general econometrics and in spatial econometrics (Anselin 2010).

This gap is filled by three articles in our sample. First, Micucci, Di Giacinto (2009) have included a spatial lag of their explanatory variables. Their results are robust in terms of spatial autocorrelation. Second, Owoo (2016) identified spatial dependency between firm's TFP in African countries. Third, Harris et al. (2019) develop a model dealing with all the methodological issues, including spatial dependence.

4.2 *Taking into account the spatial dimension in data and methods*

Spatial dependence can be integrated into distinct parts of the statistical analysis: on the dependent variable, on the measure of the firm productivity (*spatial lag model*), on the explanatory variables, namely territorial resources (*spatial cross-regressive model*), or on the error term (*spatial error-model*). A community of researchers is interested in the Generalized Moment Method (GMM) to both estimate TFP and to integrate spatial dependence in the models. Indeed, GMM allows for dealing with endogeneity, whether spatial (Anselin, Rey 2014) or production factors-related (Wooldridge 2009). Thus, Harris et al. (2019) use an integrated methodology quantifying TFP, without spatial dependence bias. They use a GMM integrating spatial dependence to quantifying TFP in a one-step way. This model appears to fill the gap identified in the previous section.

More generally, spatial analysis methods are poorly employed among the identified articles. Out of 165 articles, 73 (44%) include at least one map and/or an application of spatial analysis tools. Maps are used more frequently than spatial analysis tools. Specifically, only 8 out of 165 articles (5%) contain both at least one map and an application of spatial analysis tools.

Firms may have several establishments. The spatial association of a firm with its headquarters is a strong simplifying hypothesis, leading to major biases (Autant-Bernard et al. 2011). These different establishments can all be places where value is produced, potentially benefiting from specific localised externalities. More generally, the observation made by Autant-Bernard et al. (2011) testifies to an insufficient consideration of the geographical complexity and sub-national variability of productivity in the scientific literature, as highlighted by, among others, Anselin (2010) and Tsvetkova et al. (2020).

The SLR confirms that the spatial association of a firm with its headquarters is generally the rule. It is striking and challenging that there is no methodological discussion about the quality of firm location data in 115 articles (70%). In these cases, it is assumed that the firm is located at its headquarters. Among the 50 remaining articles (30%), 28 articles (17%) include a methodological discussion about the data quality of firm location, assuming the limits of these data. As for the 22 other articles, 9 articles have created a dummy variable to distinguish firms with one-single location corresponding to the headquarters and firms with several establishments or plants. Harris, Moffat (2012) identify that multi-establishments firms have higher TFP than firms with one location because multi-establishments firms may have capabilities to benefit from localised externalities from several locations. Ten other articles select only single plant firms (9 cases) or multi-plants firms (1 specific case to analyse this population, see Okubo, Tomiura 2012). Finally, the 3 remaining articles assume a split of firms between different plants according to the weight of each plant (by, e.g., number of employees as weight) (Boschma et al. 2009, Gibbons et al. 2019, Mitze, Makkonen 2020).

5 What is the state of knowledge on the relationship between "firm-region nexus" and firm productivity?

In this section, the variables assessed by the different models are analysed following the methodology and the layout of Beaudry, Schiffauerova (2009). Results are also presented by article to avoid the over-representation of results from articles including several models (see Table 4).

5.1 Between 0.4% and 11% of firm TFP depends on the firm location

Some articles identify the proportion of TFP variability resulting from firm location. The spatial variability of firm TFP does not exceed 11% and is generally lower than the sectoral variability. It depends also on the study cases, the models, and the variables. According to Campisi et al. (2022), 0.4 % of firm TFP variability is only explained by the regional location, but 44% of firm TFP variability is related to the interactions between location (at regional level) and sector among Italian knowledge intensive business services. By comparison, firm TFP variability depends also on the country: intranational variability is 3.6% in France, 9.9% in Spain, and international variability is up to 15% for firms from seven European countries according to Aiello, Ricotta (2016). Between these extreme values, the magnitude of spatial variability is around 5% for firm TFP.

5.2 Few articles focus on all categories of localised externalities

Most articles (91 out of 165) include a dummy variable identifying a region and/or a control variable of spatial variability. Indeed, authors do not necessarily try to understand all the spatial variability but the effect of one specific category of localised externalities. However, these variables, by capturing part of the spatial variability, are also "catch-all", thus preventing their interpretation. The use of fixed effect to capture spatial variability is frequent. By way of illustration, some articles use this type of variable to distinguish territories with remarkably high disparities, as in Italy (Aiello et al. 2014, 2015, Antonioli et al. 2016, Cardamone et al. 2016, Lasagni et al. 2015), but also elsewhere in the world (for instance, in Africa, see Owoo, Naudé 2017; in Great Britain, see Harris, Moffat 2015; in Canada, see Baldwin et al. 2008).

5.3 Analysis of variables quantifying the "firm-region nexus"

5.3.1 Agglomeration economies and their components

Agglomeration economies are the most studied type of localised externalities, with 49% of tested variables among the 165 articles. Variables quantifying agglomeration economies are present in 57% of articles of this SLR. Simple indicators quantifying agglomeration economies are firm or population density, or local GDP (per capita). These global indicators are used to consider the size effect of cities or urban areas on firm productivity. These indicators allow authors to focus on other types of localised externalities. Indeed, these global indicators do not aim to understand the different categories of agglomeration economies.

Relevant articles are focused on countries with cities concentrating high productivity. In dense areas of France, the productivity advantage ranges from 4.8% for the least productive firms to 13.9% for the most productive firms (Combes et al. 2012). In Belgium, a firm located in Brussels, the largest Belgian city, is 4% more productive once firm sorting and agglomeration economies integrate (Kampelmann et al. 2018). In China, doubling city size implies a productivity premium for firms between 3 and 4% (Hashiguchi, Tanaka 2015, Pan, Zhang 2002). In United Kingdom, the intranational productivity gap between the most and least productive places is broken into two roughly equal parts: on the one hand, between London, the most productive place, and the second most productive place and, on the other hand, between this second most productive place and the least productive places (Harris, Moffat 2022). This illustrates the difference between localised resources in London and in the rest of the country.

Among the four components of agglomeration economies, the competition dimension is the least tested, but the positive effect is the most unanimous among articles of SLR. The competition dimension is the most important source of agglomeration economies of manufacturing firms in Italy according to Cainelli et al. (2015), but not for the case of German firms (Brunow, Blien 2015).

The localisation and urbanisation economies are more frequently studied. The positive effect of localisation economies on firm productivity is more frequently observed than the positive effect of urbanisation economies on firm productivity. Beaudry, Schiffauerova (2009) suggest that detection of localisation and/or urbanisation economies depends on scale and sectoral aggregation. For these authors, it remains difficult to detect the right scale of aggregation and/or the right level of sectoral aggregation. Regarding the relevant scale of agglomeration economies, a threshold of 1 km has been identified for services located in cities in Spain (Coll-Martinez 2019) and in Sweden (Andersson et al. 2019): the effect of localisation economies is stronger at neighbourhood scale (< 1km) and urbanisation economies have an effect at city level (> 1km). Nevertheless, localisation economies have a positive effect beyond 1km as observed in Canada (Baldwin et al. 2010) or in Italy (Cainelli, Ganau 2018).

The positive effect of localisation and urbanisation economies reach consensus (see e.g. Beaudry, Schiffauerova 2009), but this is not reflected by results of SLR. Non-significant or negative effects reflect methodological designs aiming to understand the effect of agglomeration economies according to firm age, firm size, technological or knowledge intensity, or other internal characteristics of firms. Each firm has a specific equilibrium between advantages and disadvantages of agglomeration economies (Lall et al. 2004, Badr et al. 2019, Bellofatto et al. 2023, Cainelli et al. 2016, Harris et al. 2019, Martin et al. 2011) depending on the internal capabilities to capture agglomeration economies (Spanos 2019), the firm size (Knoben et al. 2016, Ramachandran et al. 2020), and the cumulative effects (with other categories of localised externalities) for some firms (Fafchamps, Hamine 2017, Di Giacinto et al. 2020). Thus, relations between agglomeration economies and firm productivity are non-linear (Carreira, Lopes 2018), including the threshold effect (Dai et al. 2017) or different temporal effects (Fazio, Maltese 2015, Martin et al. 2011); all this exerting a selection effect on firms (Hasan et al. 2018).

The quality of local labour markets is poorly addressed among the 165 articles as it remains difficult to quantify this resource. However, the local level of skilled workers (quality and quantity) is related to the firm productivity to explain difficulties of certain regions in the United Kingdom (Morris et al. 2020). The quality of local matching of these skilled workers is a component of related variety advantages as observed in Denmark by Timmermans, Boschma (2014). In China, social mobility is positively linked to firm productivity (Wang, Luo 2022).

5.3.2 Local economic structure

The local importance of certain types of firms influences nearby firm productivity. By contrast with agglomeration economies, this category refers to the effect of certain types on firms as a whole. Here, 27% of the articles include a variable related to this category into the tested models. A high local share of frontier firms influences TFP for all firms: a frontier firm leads to higher productivity for nearby firms (Serafinelli 2019, Du, Vanino 2021). A high local level of foreign direct investments influences firm productivity positively in China (Lin et al. 2011) and in Germany (Mitze, Makkonen 2020). In Italy, risk aversion of familial owned firms negatively influences their productivity and, as corollary, the productivity of neighbouring firms (Amato et al. 2022). According to the cases, the local importance of services improves productivity of manufacturing firms (Yang et al. 2018, Chen et al. 2023). The industrial specialisation also positively influences firm productivity in China (Tian et al. 2023) and in Europe (Aiello, Ricotta 2016).

5.3.3 Knowledge, innovation, human capital, and R&D

Variables in this category represent 12% of tested variables and are tested in 28% of articles. Human capital is frequently quantified by the local share of the most highly qualified (for instance, local share of university degree). This variable positively influences firm productivity according to the SLR: more than 50% positive association between human capital and firm productivity are observed among the 165 articles. In China, increasing by one year the average number of years of study of city's population increase by 14% firm productivity (Liu 2014). Training of workers is also a way to take the most of localisation economies as observed in France by Morin, Védrine (2022).

Beside human capital, R&D is frequently tested because data are available. Also, they usually positively influence firm productivity, but there are sometimes unsignificant relations. Also, the effect of universities on neighbouring firms is not yet clear. Beneficiaries of academic activities are not well identified (Raspe, van Oort 2011).

Knowledge diffusion is poorly tested by comparison with the two previous subcategories (R&D and human capital) of "firm-region nexus". There are many ways to quantify knowledge diffusion from institutions like universities: between firms and between (certain types of) workers, among others. Knoben et al. (2016) studied the effect of face-to-face contacts in combination with internal characteristics of firms and agglomeration economies. Face-to-face contacts of a firm influence the firm capability to take benefit from agglomeration economies. Antonelli, Scellato (2013, p. 92) observed different relevant scales for contacts and benefits from proximity in terms of productivity according to sectoral proximity: "The Jacobian inter-industrial -within region- vertical knowledge interactions, the intra-industrial nationwide horizontal knowledge feedbacks, and the localised intra-industrial knowledge interactions that take place within geographical clusters".

Finally, a high local entrepreneurship dynamism, quantified by the local share of new firms, is not necessarily linked to a high firm productivity: influence of new firms is negative at the time of creation but it becomes positive several years after the initial entry (Andersson et al. 2012). However, younger firms adopt likely faster more advanced technologies than older firms, which improves their productivity (Howell 2020).

5.3.4 Quality of governance and local institutions

Our analysis shows that variables quantifying the quality of local governance and institutions positively influence firm productivity. Several international institutions have developed regional index quantifying the quality of local governance or institutions such as the European Quality of Government Index (EQI) (Charron et al. 2014, 2021). These indicators are positively associated with firm productivity even if the indicators are constructed differently in respective global regions (e.g. Manzocchi et al. 2017, Rodriguez-Pose et al. 2021, Abeberese et al. 2023, Hussen, Çokgezen 2022). Besides these indexes, low levels of corruption, rapid delivery of building permits, and low local authority debts tend to positively influence firm productivity (Agostino et al. 2020, Giordano et al. 2020, Lasagni et al. 2015, Demir et al. 2022, Qi et al. 2022).

Access to finance also positively influences firm productivity. A high local level of private fundings (by banks) or a high number of local banks positively influence firm productivity, as observed in Italy (Moretti 2014, Ganau 2016, Manzocchi et al. 2017). A high quality of local institutions moderates the negative effect of the lack of funding by financial institutions (Rodriguez-Pose et al. 2021). The results are not similar regarding public fundings (by comparison with fundings from banks), especially subsidies, leading to a sorting effect of firms (Zhu et al. 2022).

5.3.5 Infrastructure and geography

Within this category, where numerous variables have been tested (235 variables across 73 articles), most authors seek to understand the effect of geography, infrastructure, or spatio-temporal phenomena on firm productivity.

First, some authors identified the influence of human or physical geography on firm productivity: a location in less developed regions of Mexico, United Kingdom, Belgium, or Italy implies a lower firm productivity (Deichmann et al. 2004, Harris, Moffat 2012, Kampelmann et al. 2018, Di Giacinto et al. 2020). Beyond regional disparities, the effects of spatial phenomena are also studied: (a) a firm located less than 100 miles from a site hit by a terrorist attack is less productive (Mun et al. 2021); (b) the proximity to a border affects the productivity of Italian firms (Fantechi, Fratesi 2023); (c) in China, firm productivity is higher in warmer areas (Xue et al. 2021) but the projected impact of climate change, including an increasing of temperature, could negatively impact the future firm productivity until -12% of output (without adaptation) (Zhang et al. 2018); and (d) finally, air pollution is linked to a lower firm productivity in China (Liu et al. 2023).

Second, local demography affects firm productivity in several ways. Firm productivity could be improved in places (a) where immigration increases (for firms with TFP under median in France) (Mitaritonna et al. 2017); (b) with demographic diversity (e.g. nationalities) at the regional level (but not at firm level) for German firms (Trax et al. 2015); and (c) with a younger population (Di Giacinto et al. 2020). The proximity between the firm and the entrepreneur's home influences the firm's capabilities to benefit from agglomeration economies (especially Marshall-Arrow-Romer spillovers in case of proximity between firm and residence) and the lock-in situation in Italian manufacturing firms (Stefano et al. 2023).

Third, the effects of infrastructure on firm productivity are an important point of interest. New highways are linked to an increase of manufacturing firm productivity in the United Kingdom (Gibbons et al. 2019), Spain (especially in more developed regions) (Holl 2016, Holl, Mariotti 2018), and South Korea (Lee 2021). However, the effect of new highways on firm productivity has a decreasing marginality in China (Li et al. 2017, Tian et al. 2023). Indeed, a larger number of economic agents is accessible within the same time thanks to new highways. These infrastructures improve the agglomeration economies. The accessibility benefit decreases the more the region is already connected. Moreover, a negative effect on firm productivity is identified if the better accessibility is linked to a too small city (Lin, Huang 2021). The same reasoning applies to high-speed lines for accessibility to cities: a decrease of time-distance by train of 1% is linked to an increase of 1.5% firm TFP (Yué, Nan 2019), which strengthens exchanges between cities (Zhao et al. 2021).

Fourth, IT infrastructure is crucial to ICT access and ultimately to firm productivity. This is observed by a survey in New Zealand (Grimes et al. 2012) and is linked to the quality of preexisting networks (time effect) in Italy (Cambini et al. 2023). Zhou, Wang (2023) identify a stronger positive effect of proximity to main train stations or airports for labour-intensive, non-exporting, private and small firms.

Fifth, Special Economic Zones (SEZ) have been developed in many countries to host Foreign Direct Investments or exporting firms. These zones offer facilities in terms of land, taxes, institutions, and infrastructure depending on countries. Given that the real effect depends on national or local cases, it remains difficult to generalise about SEZs. For instance, many studies identify a selection effect (Koster et al. 2019, Li et al. 2021), border effect (in or out the SEZ) (Zhou, Liu 2023), and heterogeneous effects by firm categories (Howell 2019, Hasan et al. 2020).

Sixth, regional path dependency impacts firm productivity. The quantification of regional path dependency depends on the scope of articles. The variables are sometimes used as instrumental variables. Many articles refer to phenomena many centuries ago. For instance, middle age climate is used as a variable to identify vulnerable regions that have set up institutions at an early stage. The early development of institutions explains current firm productivity of manufacturing in Europe (Rodriguez-Pose et al. 2021). The road network of Antiquity and the Middle Ages is an instrumental variable that identifies the positive effect of highways on firm productivity (Holl 2016, Lee 2021). Closer in time, local literacy levels, historical population density, and/or the existence of a cadastre at the end of the 19^{th} century are positively linked to current firm productivity (Békés, Harasztosi 2013, Agostino et al. 2020).

6 Concluding remarks

The objective of this SLR is to identify and analyse articles that quantify the relations "firm-region nexus" and productivity at the firm level. Based on a query conducted using the *Scopus* search engine, 165 articles were analysed, revealing an emerging research field.

This literature review highlights the fact that too little research is conducted on the impact of territorial resources on productivity as measured at the firm level by comparison with the abundant research conducted at regional level. From this viewpoint, there is a need to better identify the effect of territorial resources on firms, which are the target group for a large part of regional development policies. Staying on the level of administrative regions exposes researchers and decision makers to the ecological fallacy of attributing an observed effect for the community to everyone in that community (Openshaw 1984, Wrigley 1995). Moreover, "head office" bias could lead to overestimating the real magnitude of agglomeration economies and localised externalities linked to largest cities.

The elements discussed here not only directly impact the results of the studies about the relations between the "firm-region nexus" and firm productivity, but consequently also the conclusions regarding regional development, a major economic policy objective. That is why policymakers should pay greater attention to data and method issues. That said, new research may lead to increasingly robust insights about the impact of territorial resources on firm productivity.

The availability of reliable data on firms and the use of robust estimates of TFP as a comparable indicator of firm productivity seem to jointly condition the development of field research. The analysed articles usually focus on national contexts because data, typically produced at the national level, are not standardised across countries. This continues to make international comparisons challenging (Schuh et al. 2017). Although authors normally work within a single national context, ongoing exchanges between scientists across national boundaries are making it possible to achieve a consensus regarding appropriate statistical analysis methods and the effect of "firm-region nexus" on firm productivity. For example, some innovative variables could be assessed in multiple national contexts, notably in terms of access to financial services or the quality of supply-demand matching in local labour markets.

Few articles take seriously into account the firm location issue: only 30% of 165 articles contain a specific methodological discussion about it. Some of these articles detail the potential bias in their methodological section. Therefore, we can conclude that there is a lack of consideration for the geographical complexity and subnational variability of productivity as highlighted by Anselin (2010), Autant-Bernard et al. (2011), and Tsvetkova et al. (2020). Neglecting the "head office bias", the real advantage of agglomeration probably tends to be overestimated (Gaubert 2018, Bouba-Olga, Grossetti 2020).

Classical methods to estimate firm TFP neglect the bias induced by spatial dependence between firms. From this perspective, it is striking to observe that, for instance, Van Beveren (2012) does not include it in his list of methodological challenges. This potential bias is another methodological issue for researchers in regional science that reflects an additional lack of consideration for the spatial dimension. In the future, studies should address the remaining gap in the integration of the spatial dimension, notably by developing ad hoc methodologies and tools. First, it is observed that cartography and spatial analysis are still underused tools for describing fundamentally spatial phenomena. In parallel, the development of tools in spatial data analysis, such as *Geoda Space* (Anselin, Rey 2014), is likely to facilitate wide usage of these techniques. We should also mention Harris et al. (2019), which develops an innovative methodology combining, within the same model, a robust estimation of the TFP, spatial dependence, and an explanatory model of territorial resources using GMM.

The quantification of "firm-region nexus" remains very varied, depending on national or local cases (diverging points of interest, data availability, different ways to quantify "firm-region nexus" ...). A critical perspective is more developed in articles focusing on one category of externalities. Thus, a "general" effect (identified by a classical regression) could hide specific effects or relations (threshold effect, decreasing marginality...).

These more detailed analyses explain a part of the non-significant, heterogeneous, or negative effects on firm productivity among the 165 articles. There are some tests of specific relations between firm productivity and variables. Many articles challenge the real effect of agglomeration economies on firm productivity: spatial and/or temporal effect, linear or non-linear relations, and the threshold effect. Regarding entrepreneurial dynamics, some articles identify the key importance of firm age on TFP of nearby firms with, from a temporal perspective, a negative and then a positive effect.

Among the 165 articles, it appears that the "firm-region nexus" influences firm TFP. Regarding articles including the variety of variables quantifying this "firm-region nexus", firm location potentially influences TFP within a range of 0.4% to 11%. Therefore, location does influence firm productivity, but it is a minor factor compared to internal company organisation.

Among the different categories of localised externalities, scopes diverge. Agglomeration economies increase firm productivity. Beyond this stylised fact, the scientific debate focuses on the spatial and/or temporal effect, the magnitude of each type of agglomeration economies, the cumulative advantage, the selection effect on firms with varying degrees of interest for agglomeration economies, and the local magnitude of diseconomies of agglomeration. The quality of matching on the local labour market is another aspect of agglomeration economies. A better understanding of the "head office bias" could improve analysis robustness. In addition to agglomeration economies, some firms have a specific effect on firm productivity nearby due to their nature (e.g., size, frontier firms, financial services...).

The local quality of governance, institutions and human capital are less frequently tested but are more frequently positively linked to firm productivity. This is clearer than knowledge diffusion or local level of innovation due to difficulties in quantifying these categories.

Finally, geography, demography, history, and infrastructure could influence firm productivity, but these variables are strongly related to agglomeration economies, institutions, or human capital. For instance, infrastructure improves accessibility to economic agents and therefore to agglomeration economies. SEZ are developed to improve agglomeration economies in locations open to the world. Historical variables, quantifying past levels of development (sometimes even from the Roman era for European countries, see Holl 2016), seem to explain the variability across space of firm TFP in the modern era.

References

Abeberese AB, Barnwal P, Chaurey R, Mukherjee P (2023) Democraty and firm productivity: Evidence from Indonesia. *Review of Economics and Statistics* 105: 713–722. CrossRef

Abramowitz M (1956) Resources and outputs trends in the US since 1870. *American Economic Review* 46: 5–23

Ackerberg DA, Caves K, Frazer G (2015) Identification properties of recent production function estimators. *Econometrica* 83: 2411–2451. CrossRef

Agostino M, Di Tommaso MR, Nifo A, Rubini L, Trivieri F (2020) Institutional quality and firms' productivity in European regions. *Regional Studies* 54: 1275–1288. CrossRef

Aiello F, Pupo V, Ricotta F (2014) Firm heterogeneity in TFP, sectoral innovation and location. Evidence from Italy. *Spatial Economic Analysis* 9: 51–70. CrossRef

Aiello F, Pupo V, Ricotta F (2015) Explaining total factor productivity at firm level in Italy. Does location matter? *International Review of Applied Economics* 29: 579–607. CrossRef

Aiello F, Ricotta F (2016) Firm heterogeneity in productivity across Europe: Evidence from multilevel models. *Economics of Innovation and New Technology* 25: 57–89. CrossRef

Amato S, Pieroni V, Lattanzi N, Vitali G (2022) Are family firms good neighbours? A spatial analysis of Italian technology-intensive firms. *Journal of Small Business and Enterprise Development* 29: 663–693. CrossRef

Andersson M, Braunerhjelm P, Thulin P (2012) Creative destruction and productivity: Entrepreneurship by type, sector and sequence. *Journal of Entrepreneurship and Public Policy* 1: 125–146. CrossRef

Andersson M, Larsson JP, Wernberg J (2019) The economic microgeography of diversity and specialization externalities - firm-level evidence from Swedish cities. *Research Policy* 48: 1385–1398. CrossRef

Anselin L (2010) Thirty years of spatial econometrics. *Papers in Regional Science* 89: 3–25. CrossRef

Anselin L, Rey SJ (2014) *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.* GeoDa Press LLC, Chicago, USA

Antonelli C, Scellato G (2013) Complexity and technological change: Knowledge interactions and firm level total factor productivity. *Journal of Evolutionary Economics* 23: 77–96. CrossRef

Antonioli D, Borghesi S, Mazzanti M (2016) Are regional systems greening the economy? Local spillovers, green innovations and firms' economic performances. *Economics of Innovation and New Technology* 25: 692–713. CrossRef

Autant-Bernard C, Guironnet JP, Massard N (2011) Agglomeration and social return to r&d: Evidence from French plant productivity changes. *International Journal of Production Economics* 132: 34–42. CrossRef

Badr K, Rizk R, Zaki C (2019) Firm productivity and agglomeration economies: Evidence from Egyptian data. *Applied Economics* 51: 5528–5544. CrossRef

Baldwin JR, Beckstead D, Brown MW, Rigby DL (2008) Agglomeration and the geography of localization economies in Canada. *Regional Studies* 42: 117–132. CrossRef

Baldwin JR, Brown WM, Rigby D (2010) Agglomeration economies: Microdata panel estimates from Canadian manufacturing. *Journal of Regional Science* 50: 915–934. CrossRef

Beaudry C, Schiffauerova A (2009) Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research policy* 38: 318–337. CrossRef

Bellofatto AA, Domínguez B, Dwyer EC (2023) Uncovering urban advantages: Evidence from Australian firm-level data. *Economic Record* 99: 13–34. CrossRef

Boell SK, Cecez-Kecmanovic D (2015) On being 'systematic' in literature reviews in IS. *Journal of Information Technology* 30: 161–173. CrossRef

Boschma R, Eriksson R, Lindgren U (2009) How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity. *Journal of Economic Geography* 9: 169–190. CrossRef

Bouba-Olga O, Grossetti M (2020) The French version of the CAGE mythology (competitiveness – attractiveness – globalization – excellence) and some ideas on how to escape from it. In: Pumain D (ed), *Theories and models of urbanization.* Springer, Cham, Switzerland, 133–150. CrossRef

Brunow S, Blien U (2015) Agglomeration effects on labor productivity: An assessment with microdata. *REGION* 2: 33–53. CrossRef

Békés G, Harasztosi P (2013) Agglomeration premium and trading activity of firms. *Regional Science and Urban Economics* 43: 51–64. CrossRef

Cainelli G, Fracasso A, Marzetti G (2015) Spatial agglomeration and productivity in Italy: A panel smooth transition regression approach. *Papers in Regional Science* 94: S39–S67. CrossRef

Cainelli G, Ganau R (2018) Distance-based agglomeration externalities and neighbouring firms' characteristics. *Regional Studies* 52: 922–933. CrossRef

Cainelli G, Ganau R, Iacobucci D (2016) Do geographic concentration and vertically related variety foster firm productivity? Micro-evidence from Italy. *Growth and Change* 47: 197–217. CrossRef

Cambini C, Grinza E, Sabatino L (2023) Ultra-fast broadband access and productivity: Evidence from Italian firms. *International Journal of Industrial Organization* 86: 102901. CrossRef

Campisi D, Mancuso P, Mastrodonato SL, Morea D (2022) Productivity dispersion in the Italian knowledge-intensive business services (kibs) industry: A multilevel analysis. *Management Decision* 60: 940–952. CrossRef

Cardamone P, Pupo V, Ricotta F (2016) Do firms benefit from university research? Evidence from Italy. *Italian Economic Journal* 2: 445–471. CrossRef

Carreira C, Lopes L (2018) Regional knowledge spillovers: A firm-based analysis of non-linear effects. *Regional Studies* 52: 948–958. CrossRef

Charron N, Dijkstra L, Lapuente V (2014) Regional governance matters: Quality of government within European Union member states. *Regional Studies* 48: 68–90. CrossRef

Charron N, Lapuente V, Bauhr M (2021) Sub-national quality of government in EU member states: Presenting the 2021 European Quality of Government Index and its relationship with Covid-19 indicators. Working paper series of the quality of government institute 2021 (4)

Chen G, Liu Y, Gao Q, Zhang J (2023) Does regional services development enhance manufacturing firm productivity? A manufacturing servitization perspective. *International Review of Economics and Finance* 86: 451–466. CrossRef

Coll-Martinez E (2019) Creativity and the city: Testing the attenuation of agglomeration economies in Barcelona. *Journal of Cultural Economics* 43: 365–395. CrossRef

Combes PP, Duranton G, Gobillon L, Puga D, Roux S (2012) The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica* 80: 2543–2594. CrossRef

Dai M, Li X, Lu Y (2017) How urbanization economies impact TFP of R&D performers: Evidence from China. *Sustainability* 9: 1766. CrossRef

Deichmann U, Fay M, Koo J, Lall SV (2004) Economic structure, productivity, and infrastructure quality in Southern Mexico. *Annals of Regional Science* 38: 361–385. CrossRef

Demir F, Hu C, Liu J, Shen H (2022) Local corruption, total factor productivity and firm heterogeneity: Empirical evidence from Chinese manufacturing firms. *World Development* 151: 105770. CrossRef

Dhyne E, Duprez C (2016) Three regions, three economies? Economic Review 12/2016, National Bank of Belgium, Brussels, Belgium, https://www.nbb.be/doc/ts/publications/economicreview/2016/revecoiii2016_h4.pdf

Di Giacinto V, Micucci G, Tosoni A (2020) The agglomeration of knowledge-intensive business services firms. *Annals of Regional Science* 65: 557–590. CrossRef

Du J, Vanino E (2021) Agglomeration externalities of fast-growth firms. *Regional Studies* 55: 167–181. CrossRef

Duranton G (2011) California dreamin': The feeble case for cluster policies. *Review of Economic Analysis* 3: 3–45. CrossRef

Ehrl P (2013) Agglomeration economies with consistent productivity estimates. *Regional Science and Urban Economics* 43: 751–763. CrossRef

Fafchamps M, Hamine SE (2017) Firm productivity, wages, and agglomeration externalities. *Research in Economics* 71: 291–305. CrossRef

Fantechi F, Fratesi U (2023) Border effects on firm's productivity: The role of peripherality and territorial capital. *Papers in Regional Science* 102: 483–506. CrossRef

Fazio G, Maltese E (2015) Agglomeration externalities and the productivity of Italian firms. *Growth and Change* 46: 354–378. CrossRef

Fotheringham AS, Wong DWS (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23: 1025–1044. CrossRef

Gal P (2013) Measuring total factor productivity at the firm level using OECD-ORBIS. OECD Economics Department Working Papers, 1049, Paris, France, OECD Editions. CrossRef

Ganau R (2016) Productivity, credit constraints and the role of short-run localization economies: Micro-evidence from Italy. *Regional Studies* 50: 1834–1848. CrossRef

Gaubert C (2018) Firm sorting and agglomeration. *American Economic Review* 108: 3117–3153. CrossRef

Gibbons S, Lyytikäinen T, Overman HG, Sanchis-Guarner R (2019) New road infrastructure: The effects on firms. *Journal of Urban Economics* 110: 35–50. CrossRef

Giordano R, Lanau S, Tommasino P, Topalova P (2020) Does public sector inefficiency constrain firm productivity? Evidence from Italian provinces. *International Tax and Public Finance* 27: 1019–1049. CrossRef

Gordon J, Zhao S, Gretton P (2015) On productivity: concepts and measurement. Productivity Commission Staff Research Note, Australian Productivity Commission, Canberra, Australia

Griliches Z (1996) The discovery of the residual: A historical note. *Journal of Economic Literature* 34: 1324–1330

Griliches Z (1998) *R&D and productivity: The Econometric Evidence*. University of Chicago Press, Chicago, USA

Grimes A, Ren C, Stevens P (2012) The need for speed: Impacts of internet connectivity on firm productivity. *Journal of Productivity Analysis* 37: 187–201. CrossRef

Harris R, Moffat J (2012) Is productivity higher in British cities? *Journal of Regional Science* 52: 762–786. CrossRef

Harris R, Moffat J (2015) Plant-level determinants of total factor productivity in Great Britain 1997-2008. *Journal of Productivity Analysis* 44: 1–20. CrossRef

Harris R, Moffat J (2022) The geographical dimension of productivity in Great Britain, 2011-18: The sources of the London productivity advantage. *Regional Studies* 56: 1713–1728. CrossRef

Harris R, Moffat J, Evenhuis E, Martin R, Pike A, Sunley P (2019) Does spatial proximity raise firm productivity? Evidence from British manufacturing. *Cambridge Journal of Regions, Economy and Society* 12: 467–487. CrossRef

Hasan S, Faggian A, Klaiber HA, Sheldon I (2018) Agglomeration economies or selection? An analysis of Taiwanese science parks. *International Regional Science Review* 41: 335–363. CrossRef

Hasan S, Klaiber HA, Sheldon I (2020) The impact of science parks on small- and medium-sized enterprises' productivity distributions: The case of Taiwan and South Korea. *Small Business Economics* 54: 135–153. CrossRef

Hashiguchi Y, Tanaka K (2015) Agglomeration and firm-level productivity: A Bayesian spatial approach. *Papers in Regional Science* 94: S95–S114. CrossRef

Holl A (2016) Highways and productivity in manufacturing firms. *Journal of Urban Economics* 93: 131–151. CrossRef

Holl A, Mariotti I (2018) Highways and firm performance in the logistics industry. *Journal of Transport Geography* 72: 139–150. CrossRef

Howell A (2019) Heterogeneous impacts of China's economic and development zone program. *Journal of Regional Science* 59: 797–818. CrossRef

Howell A (2020) Picking 'winners' in space: Impact of spatial targeting on firm performance in China. *Journal of Regional Science* 60: 1025–1046. CrossRef

Hulten CR (2001) Total factor productivity: A short biography. In: Hulten CR, Dean ER, Harper MJ (eds), *New Developments in Productivity Analysis*, NBER Studies in income and wealth No. 63. University of Chicago Press, Chicago, 1–47

Hussen MS, Çokgezen M (2022) Relationship between innovation, regional institutions and firm performance: Micro-evidence from Africa. *African Journal of Science, Technology, Innovation and Development* 14: 316–332. CrossRef

Kampelmann S, Rycx F, Saks Y, Tojerow I (2018) Misalignment of productivity and wages across regions: Evidence from Belgium. *Regional Studies* 52: 1695–1707. CrossRef

Knoben J, Arikan AT, van Oort F, Raspe O (2016) Agglomeration and firm performance: One firm's medicine is another firm's poison. *Environment and Planning A* 48: 132–153. CrossRef

Koster HRA, Cheng FF, Gerritse M, van Oort FG (2019) Place-based policies, firm productivity, and displacement effects: Evidence from Shenzhen, China. *Journal of Regional Science* 59: 187–213. CrossRef

Lall SV, Shalizi Z, Deichmann U (2004) Agglomeration economies and productivity in Indian industry. *Journal of Development Economics* 73: 643–673. CrossRef

Lasagni A, Nifo A, Vecchione G (2015) Firm productivity and institutional quality: Evidence from Italian industry. *Journal of Regional Science* 55: 774–800. CrossRef

Lee JK (2021) Transport infrastructure investment, accessibility change and firm productivity: Evidence from the Seoul region. *Journal of Transport Geography* 96: 103182. CrossRef

Levinsohn J, Petrin A (2003) Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies* 70: 317–341. CrossRef

Li X, Wu X, Tan Y (2021) Impact of special economic zones on firm performance. *Research in International Business and Finance* 58: 101463. CrossRef

Li Z, Wu M, Chen BR (2017) Is road infrastructure investment in China excessive? Evidence from productivity of firms. *Regional Science and Urban Economics* 65: 116–126. CrossRef

Lin HL, Li HY, Yang CH (2011) Agglomeration and productivity: Firm-level evidence from China's textile industry. *China Economic Review* 22: 313–329. CrossRef

Lin Y, Huang J (2021) Can highway networks promote productivity? Evidence from China. *Journal of Advanced Transportation* 6860979. CrossRef

Liu S, Yang Y, Cai L (2023) Impact of air quality on enterprise productivity: Evidence from Chinese listed companies. *Frontiers in Environmental Science* 10. CrossRef

Liu Z (2014) Human capital externalities in cities: Evidence from Chinese manufacturing firms. *Journal of Economic Geography* 14: 621–649. CrossRef

Manzocchi S, Quintieri B, Santoni G (2017) Local manufacturing productivity markers: An empirical study of the Italian counties. *Annals of Regional Science* 59: 255–279. CrossRef

Martin P, Mayer T, Mayneris F (2011) Spatial concentration and plant-level productivity in France. *Journal of Urban Economics* 69: 182–195. CrossRef

Maskell P (2001) The firm in economic geography. *Economic Geography* 77: 329–344. CrossRef

Micucci G, Di Giacinto V (2009) The producer service sector in Italy: Long-term growth and its local determinants. *Spatial Economic Analysis* 4: 391–425. CrossRef

Mitaritonna C, Orefice G, Peri G (2017) Immigrants and firms' outcomes: Evidence from France. *European Economic Review* 96: 62–82. CrossRef

Mitze T, Makkonen T (2020) When interaction matters: The contingent effects of spatial knowledge spillovers and internal R&I on firm productivity. *Journal of Technology Transfer* 45: 1088–1120. CrossRef

Mongeon P, Paul-Hus A (2016) The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics* 106: 213–228. CrossRef

Moretti L (2014) Local financial development, socio-institutional environment, and firm productivity: Evidence from Italy. *European Journal of Political Economy* 35: 38–51. CrossRef

Morin Y, Védrine L (2022) Do agglomeration economies affect firms' returns to training? Evidence based on French industrial firms. *Papers in Regional Science* 101: 1135–1156. CrossRef

Morris D, Vanino E, Corradini C (2020) Effects of regional skill gaps and skill shortages on firm productivity. *Environment and Planning A* 52: 933–952. CrossRef

Mun S, Han SH, Kim HJ (2021) Terrorist attacks and total factor productivity. *Economics Letters* 202: 109808. CrossRef

Nadiri MI (1970) Some approaches to the theory and measurement of total factor productivity: A survey. *Journal of Economic Literature* 8: 1137–1177. https://www.jstor.org/stable/pdf/2720722.pdf

OECD (2020) *Enhancing Productivity in UK Core Cities: Connecting Local and Regional Growth*. OECD Urban Policy Reviews. OECD Publishing, Paris, France. CrossRef

Okubo T, Tomiura E (2012) Industrial relocation policy, productivity and heterogeneous plants: Evidence from Japan. *Regional Science and Urban Economics* 42: 230–239. CrossRef

Olley S, Pakes A (1996) The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64: 1263–1297. CrossRef

Openshaw S (1984) Ecological fallacies and the analysis of areal census data. *Environment and planning A* 16: 17–31. CrossRef

Owoo N, Naudé W (2017) Spatial proximity and firm performance: Evidence from non-farm rural enterprises in Ethiopia and Nigeria. *Regional Studies* 51: 688–700. CrossRef

Owoo NS (2016) A spatio-temporal analysis of non-farm enterprise performance in Uganda: 2010-2012. *African Journal of Economics and Management Studies* 7: 535–546. CrossRef

Pan Z, Zhang F (2002) Urban productivity in China. *Urban Studies* 39: 2267–2281. CrossRef

Qi Y, Yu J, Yang S, Xie X (2022) Local government consumption and firm performance. Evidence from the "TPCs" in China. *Journal of Asian Economics* 80: 101477. CrossRef

Ramachandran R, Reddy K, Sasidharan S (2020) Agglomeration and productivity: Evidence from Indian manufacturing. *Studies in Microeconomics* 8: 75–94. CrossRef

Raspe O, van Oort FG (2011) Firm heterogeneity, productivity and spatially bounded knowledge externalities. *Socio-Economic Planning Sciences* 45: 38–47. CrossRef

Rodriguez-Pose A, Ganau R, Maslauskaite K, Brezzi M (2021) Credit constraints, labor productivity, and the role of regional institutions: Evidence from manufacturing firms in Europe. *Journal of Regional Science* 61: 299–328. CrossRef

Schuh B, Dallhammer E, Derszniak-Noirjean M, Kaucic J, Kirchmayr-Novak E, Tordy J, Oberholzner T, Enichlmair C, Kaufmann P (2017) Small and medium-sized enterprises in European regions and cities. Scientific annex. ESPON, Luxembourg, Grand Duchy of Luxembourg

Serafinelli M (2019) "Good" firms, worker flows, and local productivity. *Journal of Labor Economics* 37: 747–792. CrossRef

Siepel J, Dejardin M (2020) How do we measure firm performance? A review of issues facing entrepreneurship researchers. In: Saridakis G, Cowling M (eds), *Handbook of Quantitative Research Methods in Entrepreneurship*. Edward Elgar Publishing, Cheltenham, 4–20

Solow RM (1957) Technical change and the aggregate production function. *The review of Economics and Statistics* 39: 312–320. CrossRef

Spanos G (2019) Firm organization and productivity across locations. *Journal of Urban Economics* 112: 152–168. CrossRef

Stefano A, Zare S, Nicola L (2023) Is too small always bad? The role of place attachment in harnessing location advantages. *International Entrepreneurship and Management Journal* 20: 265–308. CrossRef

Tian Z, Hu A, Lin Y (2023) Does proximity to expressways improve manufacturing productivity? Evidence from Chinese firms. *Journal of Development Studies* 59: 1867–1884. CrossRef

Timmermans B, Boschma R (2014) The effect of intra- and inter-regional labour mobility on plant performance in Denmark: The significance of related labour inflows. *Journal of Economic Geography* 14: 289–311. CrossRef

Trax M, Brunow S, Suedekum J (2015) Cultural diversity and plant-level productivity. *Regional Science and Urban Economics* 53: 85–96. CrossRef

Tsvetkova A, Ahrend R, Oliveira Martins J, Lembcke AC, Knutsson P, Jong D, Terzidis N (2020) The spatial dimension of productivity: Connecting the dots across industries, firms and places. *OECD Regional Development Working Papers* 2020: 1–65. CrossRef

Van Ark R (2014) Total factor productivity: Lessons from the past and directions for the future. NBB working paper research, 271, National Bank of Belgium, Brussels, Belgium, https://www.nbb.be/doc/ts/publications/wp/wp271en.pdf

Van Beveren I (2012) Total factor productivity estimation: A practical review. *Journal of Economic Surveys* 26: 98–128. CrossRef

Wang J, Luo C (2022) Social mobility and firms' total factor productivity: Evidence from China. *Sustainability* 14. CrossRef

Wooldridge JM (2009) On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters* 104: 112–114. CrossRef

Wrigley N (1995) Revisiting the modifiable areal unit problem and the ecological fallacy. In: Cliff AD, Gould PR, Hoare AG, Thrift N (eds), *Diffusing Geography*. Wiley, 49–71

Xue S, Zhang B, Zhao X (2021) Brain drain: The impact of air pollution on firm performance. *Journal of Environmental Economics and Management* 110: 102546. CrossRef

Yang FF, Yeh AGO, Wang J (2018) Regional effects of producer services on manufacturing productivity in China. *Applied Geography* 97: 263–274. CrossRef

Yué S, Nan S (2019) Railway speedup and firms' total factor productivity. *Economics Letters* 178: 21–23. CrossRef

Zhang P, Deschenes O, Meng K, Zhang J (2018) Temperature effects on productivity and factor reallocation: Evidence from a half million Chinese manufacturing plants. *Journal of Environmental Economics and Management* 88: 1–17. CrossRef

Zhao T, Xiao X, Dai Q (2021) Transportation infrastructure construction and high-quality development of enterprises: Evidence from the quasi-natural experiment of high-speed railway opening in China. *Sustainability* 13: 13316. CrossRef

Zhou D, Wang W (2023) Can internet adoption promote the productivity of firms in China: Evidence from a longitudinal study. *China Agricultural Economic Review* 15: 795–813. CrossRef

Zhou Z, Liu Z (2023) The impact of place-based policies in firm performance: Evidence from China. *Sustainability* 15: 6623. CrossRef

Zhu J, Xu H, Zhang Y (2022) Local government debt and firm productivity: Evidence from China. *Research in International Business and Finance* 63: 101798. CrossRef

© 2026 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Louvain-la-Neuve, Belgium. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

REGION
The Journal of ERSA
Powered by WU

Volume 13, Number 1, 2026, 23–42
DOI: 10.18335/region.v13i1.616
journal homepage: region.ersa.org
ISSN: 2409-5370

Regional Disparities in Italy: Developing a Composite Indicator of Physical Activity, Health, and Mobility

Giorgio Cecchi[1]

[1] Università Telematica degli Studi – IUL, Florence, Italy

Received: 1 July 2025/Accepted: 23 February 2026

Abstract. Quality of Life in urban settings is intricately tied to active transportation, public health, and physical activity, capturing the growing societal emphasis on sustainable mobility and well-being. This study develops a composite indicator to evaluate regional Quality of Life across Italy, utilizing data from the 2021 ISTAT (Italian National Institute of Statistics) survey "Aspects of Daily Life." Employing Principal Component Analysis for weighting, the indicator identifies key dimensions of urban well-being, including active transportation, physical activity, and general health. The results reveal pronounced disparities between central-northern and southern Italy, driven by differences in infrastructure, access to health resources, and socio-economic conditions. Northern regions consistently outperform southern counterparts, where structural challenges hinder improvements in Quality of Life. These findings underscore the urgent need for targeted, equity-focused policies to address regional disparities and enhance urban environments, with a particular focus on improving accessibility and public health in underperforming regions. The study provides valuable insights for policymakers aiming to promote sustainable and inclusive urban development.

Key words: Composite Indicator; Quality of Life; Active Transport; Health; Physical Activity

1 Introduction

Active transportation and general health are critical determinants of Quality of Life (QoL) in urban and regional contexts (Pazhuhan et al. 2020). In Italy, the extent to which citizens engage in active modes of transportation, such as walking and cycling, varies significantly across regions, as does the accessibility of public transport systems (Beria et al. 2017). These factors play a crucial role in shaping mobility patterns (Huan et al. 2023), reducing car dependency (Wiersma 2020), alleviating traffic congestion, and improving air quality (Cheshmehzangi, Thomas 2016). Simultaneously, engaging in regular physical activity, including cycling and walking, has been shown to improve overall health outcomes, reducing the risk of chronic diseases such as cardiovascular conditions and obesity (Miles 2007).

In addition to transportation, QoL is heavily influenced by the health status of the population, which reflects broader socio-economic and environmental factors (Bowling, Windsor 2001). Regions with higher levels of participation in regular sports activities tend to report better health outcomes (Lera-López, Marco 2018) as consistent physical activity is essential for both physical and mental well-being (Kim et al. 2017). However, regional disparities in access to opportunities for active living, such as infrastructure for cycling and walking or sports facilities, contribute to differences in citizens' health across Italy (Costa et al. 2020).

Traditional indicators of well-being, such as GDP per capita, as well as multidimensional QoL frameworks developed by national and international institutions – including the Human Development Index (Anand, Sen 1994), the OECD Better Life Index, the Italian BES (Benessere Equo e Sostenibile) framework, and the Environmental Performance Index (Wendling et al. 2020) – typically rely on a broad combination of socio-economic, environmental, and subjective dimensions. While these approaches provide comprehensive assessments of societal well-being, many of the included variables are only indirectly related to everyday mobility behaviours and physical activity.

GDP per capita is widely used as a general proxy for well-being (Bechtel 2019) but is deliberately excluded from the present composite indicator. Although GDP is correlated with overall QoL, it also captures expenditures that do not necessarily enhance well-being – such as those related to tobacco consumption or healthcare costs – thereby limiting its suitability for assessing health- and mobility-related dimensions of urban well-being (Balestra et al. 2018, Shrotryia, Singh 2020). Conversely, indices focusing on transportation or environmental performance – such as the EU-framework for Sustainable Urban Mobility Indicators (Finger, Serafimova 2020) primarily assess infrastructural or environmental conditions and seldom incorporate direct measures of population health, such as physical activity levels or self-reported well-being.

This fragmentation leaves a gap for metrics that specifically evaluate how urban mobility patterns and health behaviours jointly shape regional well-being.

The contribution of this study lies precisely in addressing this gap. The index proposed here is innovative because it isolates two strictly interdependent and mutually reinforcing domains – (1) Physical Health and Active Transportation, and (2) Public and Pedestrian Commuting – capturing a specific dimension of urban well-being that is both health-related and environmentally grounded. Unlike existing QoL measures, the proposed indicator deliberately excludes economic and subjective dimensions that may obscure the direct effects of mobility patterns and physical activity on well-being. By prioritising variables that simultaneously promote personal health and environmentally sustainable behaviours, the index provides a concise, policy-relevant, and domain-specific tool particularly suited to regional analysis within the Italian context. Moreover, the indicator allows for a clearer interpretation of cross-regional disparities, which are known to follow a North–South divide in Italy in terms of health, socio-economic conditions, and mobility infrastructures (Leydesdorff 2021).

The first dimension, Physical Health, and Active Transportation examines the percentage of people who cycle to work, the proportion of the population reporting good health, and those who engage in regular sports activities. These variables capture the broader influence of lifestyle choices and regional infrastructure on health outcomes and overall well-being.

The second dimension, Public and Pedestrian Commuting, focuses on the percentage of population using public transport or walking to commute to work. These indicators provide insights into the accessibility and sustainability of transportation systems across regions, highlighting areas that support more active and environmentally friendly commuting practices.

Using data from the Italian National Institute of Statistics, ISTAT (https://www.istat.it), the study constructs a composite indicator to facilitate cross-regional comparisons and track changes over time.

The remainder of the paper is structured as follows. Section 2 reviews the literature on the mobility–health nexus and on QoL measurement, highlighting gaps in existing approaches. Section 3 outlines the theoretical framework and motivates the choice of the two dimensions. Section 4 describes the data and methodology used to construct the composite indicator. Section 5 presents and discusses the empirical results, while Section 6 concludes with policy implications, limitations, and directions for future research.

2 Literature review

2.1 The Mobility–Health Nexus in Regional Studies

Research at the intersection of regional studies, mobility behaviours, and public health has increasingly demonstrated that the spatial organisation of cities and regions plays a decisive role in shaping QoL. A large body of evidence shows that features of the built environment – including walkability, street connectivity, land-use mix, green and recreational spaces, and active mobility infrastructure – significantly influence both physical activity levels and health outcomes (Baobeid et al. 2021, Pontin et al. 2022). Moreover, mobility is not only a matter of physical movement but also a social and relational practice: everyday travel shapes opportunities for interaction, connectedness, and inclusion, influencing how individuals experience urban and regional environments (Te Brömmelstroet et al. 2017). This reinforces the idea that mobility conditions affect well-being through both health-related pathways and broader social mechanisms.

Walkable and transit-oriented environments are associated with higher rates of active commuting (Saelens, Handy 2008), lower incidence of obesity and cardiovascular diseases (Sallis et al. 2012) and improved mental well-being (Nakamura 2022). This literature provides a strong conceptual basis for considering active transportation and health behaviours as key determinants of urban and regional well-being.

The regional mobility literature further demonstrates that accessibility – defined as the ease with which individuals can reach jobs, services, and opportunities – constitutes a core dimension of regional equity and QoL (Geurs, van Wee 2004). High levels of public transport accessibility not only reduce car dependency and emissions but also enhance social inclusion and labour market participation (Papa, Bertolini 2015). Conversely, regions with weak transit systems often face compounded disadvantages, including limited mobility choices, higher exposure to pollution, and reduced participation in social and economic life (Lucas 2012). These mechanisms justify the inclusion of *public and pedestrian commuting* as a structural component of well-being.

At the same time, health outcomes vary significantly across regions due to differences in lifestyles, socio-economic structures, and access to facilities for sports, recreation, and active living (Rey 2025). Regular physical activity, whether achieved through leisure sports or active commuting, is strongly associated with lower morbidity and higher perceived health (Warburton, Bredin 2017). The regional distribution of sports participation, cycling rates, and self-reported health therefore captures critical variations in lifestyle-driven health that are relevant for QoL assessments. Integrating these indicators into the measurement of well-being is consistent with both public health research and the broader agenda of sustainable regional development.

Within regional studies, inequalities in mobility and health intersect with long-standing territorial imbalances shaped by socio-economic structures, governance capacity, and institutional performance (Rodríguez-Pose 2020). The Italian context exemplifies this pattern: the persistent North–South divide manifests not only in income and productivity gaps but also in disparities in transport infrastructure, urban accessibility, and public health indicators (De Muro et al. 2011). Regions in Southern Italy exhibit lower public transport supply, reduced walkability, and more limited opportunities for active living (Asso 2023), all of which contribute to weaker health outcomes and lower regional QoL. These territorial disparities make Italy a pertinent empirical setting for analysing the interaction between mobility, health, and well-being (Ferrara, Nisticò 2015).

While concepts such as territorial cohesion, governance quality, and regional resilience remain central to understanding regional disparities, an excessively broad review risks diluting the argument. What is most relevant for the present study is that resilient and inclusive regions are those that offer equitable access to mobility options and promote healthy behaviours (Medeiros, Rauhut 2020, Simmie, Martin 2010). Improvements in active transportation infrastructure, public transit, and pedestrian environments have been shown to simultaneously strengthen environmental sustainability, social inclusion, and population health – key components of resilient regional development.

2.2 Quality of Life Measurements

QoL measurement has become an essential focus in social science and policy due to its comprehensive implications for health, well-being, and societal satisfaction. Traditional unidimensional indicators, such as income or employment rates, have proven inadequate for capturing the complexity of QoL, which involves multidimensional factors including health, social relationships, environment, and individual satisfaction (Maggino 2023). QoL is a critical societal indicator, impacting policies and personal well-being (Verdugo et al. 2005), and has therefore garnered attention from various stakeholders, including governments, researchers, and national and international statistical institutes. Given its complexity, recent efforts have emphasized the use of composite indicators to address the limitations of single metrics. This section examines several prominent projects and composite indices developed to provide a more comprehensive assessment of QoL.

One of the pioneering efforts to measure equitable and sustainable well-being on a national scale is Italy's BES project (https://www.istat.it/statistiche-per-temi/focus/benessere-e-sostenibilita/la-misurazione-del-benessere-bes/gli-indicatori-del-bes/). It was launched in 2013 by ISTAT in collaboration with the National Council for Economics and Labour (CNEL). BES integrates multiple dimensions of QoL, including health, education, environment, economic well-being, and civic engagement, with the goal of assessing well-being at both national and regional levels (De Rosa 2018). By incorporating both objective indicators (such as income and employment) and subjective aspects (such as perceived safety and satisfaction), BES stands out as one of the earliest frameworks to combine objective data with self-reported measures. This innovative approach enables policymakers to address regional disparities, allowing for targeted interventions that consider both economic and social dimensions of well-being. By 2024, BES is composed of 152 indicators organized into 12 domains, covering a broad range of QoL aspects. Although this extensive coverage provides a comprehensive view of well-being, the large number of indicators and domains makes the results complex to interpret and synthesize for policy decisions or comparisons across regions.

At the European level, the EU Statistics on Income and Living Conditions (EU-SILC; https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions/) launched in 2004, serves as a pan-European initiative to assess material and social conditions in households across EU member states. While EU-SILC primarily targets income distribution, poverty, and social exclusion, it also provides insights into broader QoL indicators, such as housing quality, health, and education. The initiative's approach represents a shift in European social policy, focusing on social inclusion, housing, and healthcare to enhance living conditions (Engsted 2013). By providing comparable data across countries, EU-SILC enables researchers and policymakers to identify and address disparities across the EU. The data collected through EU-SILC informs policies on social inclusion, housing, and healthcare, demonstrating the program's importance in shaping interventions aimed at improving the QoL across various societal dimensions (Wolff et al. 2010). A limitation of EU-SILC is that it emphasizes household-level indicators and may not fully capture individual subjective experiences or urban-specific aspects of QoL.

On a more focused scale, the Safe Cities Index (https://impact.economist.com/projects/safe-cities/), first developed by the Economist Intelligence Unit in 2015, concentrates on measuring QoL within urban environments, specifically addressing the essential factors of safety and security. This index ranks cities worldwide on a range of parameters, including digital security, health security, infrastructure, and personal safety. Given that safety is an indispensable aspect of QoL, this index provides insights into how cities can enhance residents' QoL by prioritizing security. The Safe Cities Index is particularly useful for comparing urban safety across various socio-economic contexts, illustrating how investments in infrastructure and public safety initiatives can shape both perceived and actual QoL in cities. However, the index's emphasis on safety underscores the need to integrate additional dimensions, such as environmental quality and access to public services, within broader QoL frameworks. While it provides detailed insights on urban safety, it does not integrate broader dimensions like environmental quality, social inclusion, or access to services.

In addition to these targeted initiatives, the Better Life Index by the Organisation for Economic Co-operation and Development (OECD https://www.oecd.org/), introduced in 2011, offers a comprehensive and customizable approach to QoL assessment. This index includes key indicators like housing, income, jobs, education, environment, civic engagement, and work-life balance. A unique feature of the Better Life Index is that it allows individuals to assign weights to each dimension based on their personal priorities, offering valuable insights into how QoL perceptions vary across cultural and individual contexts (Mizobuchi 2014). By allowing for customized prioritization, the index acknowledges the inherent subjectivity of QoL while providing standardized data for cross-country comparability. The index has notably influenced policy decisions in OECD countries, particularly in areas related to housing, employment, and environmental protection, highlighting its impact on shaping well-being policy in diverse socio-economic settings. Despite its flexibility and individual weighting of dimensions, the index depends on self-reported priorities and may be influenced by cultural norms, complicating cross-country comparability.

An innovative and increasingly influential model in urban planning that impacts QoL is the 15-Minute City concept, first proposed in 2016 and popularized by urban planner Carlos Moreno. The 15-Minute City framework promotes urban design principles that ensure residents can access essential services, such as healthcare, education, leisure, and commerce, within a 15-minute walk or bike ride from their homes (Papadopoulos et al. 2023). By reducing dependency on cars and fostering local connectivity, this model directly supports goals related to environmental sustainability, social cohesion, and overall well-being (Moreno 2024). Although not a statistical measure, the concept complements QoL objectives by emphasizing urban mobility, environmental sustainability, and the importance of local access to services. Cities like Paris have embraced the 15-Minute City model, reshaping urban policies to reduce traffic congestion, promote green spaces, and enhance local service accessibility (Allam et al. 2022). While primarily implemented in urban areas, the 15-Minute City could serve as a guiding framework for future QoL measures by illustrating the value of proximity and accessibility. Although it emphasizes accessibility and urban sustainability, it is primarily conceptual and not yet operationalized as a quantitative measure, limiting its direct applicability in large-scale QoL assessments.

Several additional composite indices contribute to the assessment of urban QoL. The Mercer Quality of Living Survey (https://mobilityexchange.mercer.com/quality-of-living-reports), for example, evaluates cities on aspects like political stability, healthcare, education, and environmental quality. Since its inception in 1994, this survey has helped multinational organizations and expatriates evaluate liveability in various cities, influencing economic and labour migration decisions (Okulicz-Kozaryn, Valente 2019). Its main limitation is that it is oriented toward expatriates and multinational organizations, which may not reflect the experiences of the local population. Another relevant indicator, the Global Liveability Index, also developed by the Economist Intelligence Unit, ranks cities based on healthcare, culture, environment, and infrastructure (Giap et al. 2014). It provides useful comparisons but may overlook local social inequalities and subjective well-being factors.

Taken as a whole, these frameworks illustrate the diverse approaches to QoL measurement, recognizing well-being as a multidimensional construct. While each provides valuable insights, they also present limitations, including emphasis on specific domains, unweighted or overly complex sets of indicators, potential overemphasis on economic factors, or reliance on subjective assessments. These considerations highlight the potential value of a well-balanced composite QoL measure based on reliable official statistics, carefully structured to integrate multiple relevant dimensions, providing a clear, interpretable, and comparable assessment of QoL across regions and populations.

3 Theoretical Framework

The development of a composite indicator for assessing QoL requires the selection of dimensions that capture key aspects of well-being while remaining measurable, policy-relevant, and comparable across contexts (Maggino, Zumbo 2011). In this study, the two selected dimensions – Physical Health and Active Transportation, and Public and Pedestrian Commuting – are grounded in both theoretical and practical considerations and address central mechanisms through which social, environmental, and health-related factors shape regional disparities in QoL. These domains are also closely aligned with areas of public policy intervention aimed at improving citizens' well-being, making them particularly suitable for cross-regional analysis.

Focusing on only two dimensions reflects a deliberate emphasis on simplicity and transparency. The composite indicator is designed to allow a clear and immediate interpretation as a measure of sustainable health-related behaviours and sustainable urban mobility. Limiting the number of dimensions is therefore intended to enhance interpretability and communicability, rather than to provide an exhaustive representation of all QoL domains (Bruzzi et al. 2022).

These two dimensions capture key aspects of daily life that are measurable, policy-relevant, and comparable across regions. They are also areas in which public interventions are particularly prominent, making them especially suitable for cross-regional analysis. Importantly, the indicator adopts an explicitly individual-centred perspective: it focuses on behaviours and conditions experienced directly by individuals, which simultaneously generate broader social and environmental effects. Consequently, the selected indicators refer to the proportion of people engaging in specific practices – such as using public transport, walking or cycling to work, engaging in physical activity, or reporting good health – rather than to system-level characteristics. This choice reflects a conceptual distinction between environment-centric indicators (e.g. pollution levels or transport infrastructure supply) and individual-centric indicators, where sustainability emerges through aggregated individual behaviours.

Similarly, environmental indicators, such as air pollution or emissions, are excluded following the same rationale. Although highly relevant for QoL, their inclusion would shift the focus of the composite indicator towards environmental outcomes rather than individual behaviours, thereby altering its conceptual scope and complicating interpretation. Likewise, transport-related indicators referring to the quantity or quality of available services are not considered, as they describe infrastructural conditions rather than actual mobility practices. By contrast, commuting behaviours – particularly the use of public transport and walking – capture how individuals interact with urban mobility systems in their everyday lives, while also addressing key issues of accessibility, sustainability, and social inclusion (Cottrill et al. 2020).

Finally, relying exclusively on quantitative, non-subjective indicators further supports the clarity and comparability of the composite measure. Subjective indicators, while valuable in QoL research, are more sensitive to cultural norms, individual expectations, and response styles, which may reduce cross-regional comparability (Maggino 2017, Diener et al. 2018). All variables included in the indicator are therefore expressed as percentages of individuals who report a given condition or behaviour, enhancing statistical robustness, facilitating aggregation, and ensuring a more straightforward interpretation of the resulting index. The first dimension, Physical Health, and Active Transportation, underscores the importance of active lifestyles and their direct impact on health outcomes. Regular cycling to work and participation in physical activities are well-known for their positive effects on both mental and physical health, contributing to reduced risks of chronic diseases and healthcare costs. However, significant disparities exist in physical activity levels across Italy. ISTAT data from 2021 reveals that only 23.6% of Italians aged 3 and older engage in regular physical activity, with much higher rates in northern regions like Trentino-Alto Adige (39.8%) and lower rates in southern regions such as Campania (14.3%). The inclusion of cycling in this dimension reflects the importance of regional infrastructure, such as bike lanes, in shaping commuting patterns and promoting active lifestyles. By incorporating both cycling to work and general health indicators (e.g., the percentage of people reporting good health), this dimension offers a holistic view of how lifestyle choices and access to physical activity opportunities shape public health.

The second dimension, Public and Pedestrian Commuting, is critical in assessing how accessible and sustainable mobility is within different regions. Public transport plays a dual role, reducing car dependency and lowering emissions, while also promoting social inclusion by improving access to work and services. Walking to work, similarly, enhances both physical well-being and urban sustainability. According to ISTAT data, in 2021, 10.5% of the Italian population regularly used public transport to commute, with significant regional variation: in regions like Lazio, where Rome is located, usage is much higher (26.0%), whereas in southern regions, reliance on private cars is more common. Walking also represents an active form of commuting, albeit with variation depending on the urban context and available infrastructure. The integration of public transport use and walking into this composite indicator highlights their importance in promoting both environmental sustainability and personal health.

In summary, the two dimensions identified for the composite indicator – commuting via public transport and walking, and physical activity, and general health status – provide a comprehensive framework for evaluating urban QoL across regions. Each dimension reflects essential aspects of daily life that collectively influence both individual well-being and broader societal health. The decision to focus on active transportation and physical health aligns with this principle, as these domains are both measurable and critical determinants of QoL at the regional level (Maggino 2017). By integrating these interconnected areas, the composite indicator offers a nuanced perspective on how factors such as transportation infrastructure, active commuting, and health outcomes interact to shape regional disparities in QoL.

4 Methodology

4.1 Data

The study uses data from the ISTAT "Aspetti della Vita Quotidiana" (Aspects of Daily Life) survey. The sample survey "Aspects of Daily Life" is part of the integrated system of Multipurpose Surveys on families, which was launched in 1993 to gather information on individuals and households. Conducted annually, the survey provides insights into the daily habits of citizens and the challenges they encounter. The survey, based on a sample, is conducted annually, typically in March. It involves approximately 20,000 households and 50,000 individuals. Since 2018, the survey has been administered using a mixed technique of sequential CAWI (Computer Assisted Web Interview)/PAPI (Paper And Pencil Interview) data collection. The survey design generally involves proposing a web interview to all sample families first, and subsequently sending an interviewer to non-responding families.

For the 2021 survey, the sample was integrated with the sampling design used for the Master Sample of the Permanent Population Census[1]. Specifically, the sample municipalities for this survey were selected as a sub-sample from the 2,850 municipalities of the Master Sample used in 2018. To this end, the classic sampling scheme for family surveys was applied to the sub-universe of municipalities surveyed during the Permanent Census in October 2018. Within each area, defined by the intersection of regions and six types of municipalities, the municipalities were divided into two subsets: the larger municipalities were classified as Self-Representative (AR), while the remaining ones were classified as Non-Self-Representative (NAR). NAR municipalities were further divided into strata of equal size based on population, and two sample municipalities from each stratum were selected with probabilities proportional to their size. For each municipality (both AR and NAR), a cluster sampling method was used: clusters (families) were randomly selected from the population registry, and all members of each selected family were surveyed. The minimum number of sample families per municipality was set at 24.

[1]The Permanent Population Census in Italy is based on a combination of sample surveys and statistically processed administrative data. It is conducted annually and is part of the Integrated System of Statistical Registers managed by the Italian National Institute of Statistics (ISTAT). This system was established in 2012 (art. 3 of d.lgs 179/2012, converted with modifications into law 221/2012). The strategy of permanent censuses is extended to all thematic areas, including agriculture, starting from 2021.

Families were chosen from the theoretical sample selected for the Master Sample, and for each family included, data were collected on all household members.

Within this framework, it is important to note that the age ranges used in the indicators are determined by the design of the ISTAT survey and cannot be modified. For instance, the commuting indicator ("percentage of individuals aged 15 and older who commute to work by foot, public transport, or bicycle") and the physical activity indicator ("percentage of individuals aged 3 and older who engage in regular physical activity") follow the age specifications provided by ISTAT in the publicly available datasets. While these ranges include children or younger adolescents who may participate in physical activity as part of mandatory school programs, the composite indicator is intended to capture the prevalence of behaviours associated with health and sustainable mobility, regardless of the voluntariness or motivation behind them. Similarly, in the case of commuting, individuals may use public transport or walk out of necessity rather than preference, yet these behaviours remain relevant for assessing sustainable mobility. These constraints reflect the design of the survey rather than arbitrary choices by the authors and are explicitly acknowledged as a limitation of the study.

The distributions of the simple indicators used in the study are reported in Table 1. The table shows notable differences in commuting patterns, health status, and physical activity across regions, metropolitan zones, and municipality sizes. Walking and cycling rates are generally higher in smaller municipalities and central metropolitan areas, while public transport use is concentrated in central metropolitan areas. Northern regions tend to have higher rates of cycling and better reported health, whereas southern regions show lower cycling rates and slightly lower health and physical activity levels.

4.2 Analytical Approach

To construct the composite, an approach based on PCA is adopted to determine the weights of the variables within the index. This methodology is particularly suitable for multidimensional constructs, as it allows for the consideration of intercorrelations among variables while reducing the dimensionality of the data (Jolliffe 2002, Mazziotta, Pareto 2019). Unlike more subjective weighting methods, PCA provides an objective, data-driven procedure for assigning weights based on each variable's contribution to the overall variance in the dataset (Abdi, Williams 2010). This choice aligns with tha argument of Maggino (2017) that multidimensionality must be effectively captured in composite indicators to reflect the complexity of phenomena like QoL.

As shown in Table 1, the individual indicators all display relatively small ranges. To address this issue and ensure that each indicator had an appropriate influence on the composite index, the rescaling technique in (1) was applied.

$$I_{ic} = \frac{x_{ic} - (\min(x_i) - 0.05)}{\max(x_i) - (\min(x_i) - 0.05)} \tag{1}$$

Where

I_{ic} is the value of indicator i after the transformation for case c, and
x_{ic} is the value of indicator i before the transformation for case c.

A 0.05 adjustment was introduced to ensure that minimum values do not become zero during the rescaling process, as this would prevent the use of the geometric mean in the aggregation of the elementary indices.

This normalization process increased the effect of indicators with limited variability by expanding their range, thus ensuring that even those with small initial dispersion contribute meaningfully to the composite score. The same rescaling technique was also applied to the final composite indicator, solely to expand its range and ensure better differentiation between regions, while maintaining the original rankings unchanged (Maggino 2006).

In this context, the aggregation of the indicators and dimensions in the composite indicator was performed using a weighted geometric mean. This method was chosen because it reduces compensability, ensuring that poor performance in some indicators

Table 1: Distribution of simple indicators by Italian regions, metropolitan area zones, and municipality size, year 2021

Region	Percentage of individuals aged 15 and older who commute to work on foot	Percentage of individuals aged 15 and older who commute to work using public transportation	Percentage of individuals aged 15 and older who commute to work by bicycle	Percentage of individuals aged 15 and older in good health	Percentage of individuals aged 3 and older who engage in regular physical activity
Piemonte	15.1	9.3	2.6	70.3	26.0
Valle d'Aosta	20.4	7.7	2.3	74.3	32.5
Liguria	15.8	16.5	1.3	71.0	23.1
Lombardia	12.1	16.1	4.8	71.9	28.0
Trentino-Alto Adige	14.2	10.1	9.4	79.6	39.8
Veneto	7.3	5.7	6.7	70.6	27.8
Friuli-Venezia Giulia	8.2	4.8	5.2	70.3	24.2
Emilia-Romagna	10.7	6.1	5.4	70.9	28.0
Toscana	9.8	6.0	3.3	72.3	26.5
Umbria	10.9	4.8	1.4	70.2	23.7
Marche	9.4	2.2	2.7	69.3	25.5
Lazio	11.5	26.0	0.9	72.4	25.8
Abruzzo	13.8	5.5	1.9	70.5	23.5
Molise	13.7	8.8	0.6	67.3	15.4
Campania	17.2	10.7	1.4	73.3	14.3
Puglia	17.0	5.4	1.1	70.3	17.9
Basilicata	15.5	6.2	0.3	65.8	16.0
Calabria	14.0	7.8	0.5	64.0	15.8
Sicilia	11.6	5.1	1.9	70.7	15.6
Sardegna	13.1	4.0	2.3	66.5	22.3
Metropolitan Area (MA) zone					
Central MA	15.9	29.0	4.3	70.7	25.3
Peripheral MA	9.3	14.4	2.2	72.1	24.7
Municipality size					
$\leq$ 2.000 in.	12.6	3.0	0.8	67.7	20.3
2.001 - 10.000 in.	10.6	5.0	2.8	70.2	22.2
10.001 - 50.000 in.	12.3	6.4	3.2	72.1	24.0
>50.000 in.	14.0	7.7	5.2	71.2	23.4

Source: ISTAT Survey "Aspetti Della Vita Quotidiana", year 2021 (https://www.istat.it/microdati/-multiscopo-sulle-famiglie-aspetti-della-vita-quotidiana/)

Notes: According to the ISTAT statistical classification, central municipalities of metropolitan areas are the main municipalities of Italy's largest metropolitan areas (Turin, Milan, Venice, Genoa, Bologna, Florence, Rome, Naples, Bari, Palermo, Catania, and Cagliari). Peripheral municipalities of metropolitan areas are the satellite municipalities surrounding the central municipality of a metropolitan area, defined based on functional relationships such as commuting flows and socio-economic links within the metropolitan area.

The indicator "Percentage of individuals aged 15 and older who commute to work using public transportation" is obtained as the sum of the following indicators: Percentage of individuals aged 15 and older who commute to work using I. Train II. Tram, Bus III. Metro IV Pullman cannot be fully offset by high performance in others, as would occur with an arithmetic mean. This approach is particularly effective when dealing with percentages, as it emphasizes balanced improvements across all dimensions, penalizing those regions with extreme variability in performance (El Gibari et al. 2019). By leveraging the geometric mean, the composite indicator better captures the interdependencies among dimensions, such as active transportation and physical health, and ensures that the most vulnerable areas are not masked by stronger performances in other areas.

The aggregation process involves two main steps:

1. *Aggregation of indicators within each dimension*: For each dimension, the indicators are aggregated using a weighted geometric mean, where the contribution of each indicator is determined by its respective weight. The formula is as follows:

$$D_j = \prod_{i=1}^{n_j} x_i^{w_{ij}} \tag{2}$$

   Where

   D_j is the value of the dimension j
   x_i is the value of indicator i within dimension j
   w_{ij} is the standardized weight of indicator i in dimension j
   n_j is the number of indicators in dimension j

2. *Aggregation of dimensions*: Once the indicators are aggregated within each dimension, the dimensions themselves are aggregated using another weighted geometric mean, with weights assigned to each dimension based on their relative importance. The formula is:

$$CI = \prod_{j=1}^{m} D_j^{w_j} \tag{3}$$

   Where

   D_j is the aggregated value of the dimension j
   w_j is the standardized weight of dimension j
   m is the number of dimensions

The decision to utilize PCA as the primary method for weight selection is supported by existing literature on the development of composite indices for complex phenomena, such as QoL and socio-economic development (Nardo et al. 2005). In this context, PCA not only minimizes the risk of collinearity among variables but also ensures that the composite index reflects the multidimensional nature of the phenomenon under investigation.

In terms of aggregation, the PCA-derived weights are used to compute a weighted sum of the selected indicators for each region, resulting in a composite score that reflects QoL in terms of active transportation and physical health. The analysis was conducted using IBM SPSS Statistics version 26.0 (https://www.ibm.com/it-it/spss).

Table 2 presents the standardized weights assigned to each indicator and dimension in the construction of the composite indicator, which follows a hierarchical model. This model is structured in two levels: the first level consists of the primary dimensions, while the second level includes the specific indicators within each dimension. This hierarchical approach reflects the multidimensional nature of the phenomenon being measured.

The standardized weights for each variable, derived from PCA loadings, reflect their contribution to the composite index, while dimension weights correspond to the percentage of variance explained by each component, ensuring balanced emphasis on both transportation and health outcomes. This approach highlights regional strengths and deficits, with higher composite scores for regions performing well across dimensions and lower scores where active transportation and health outcomes are limited, guiding policy priorities.

Table 2: Standardized weights assigned to variables and dimensions based on principal component analysis

Dimension	Indicator	Standardized weight of the indicator	Standardized weight of the dimension
Physical Health, and Active Transportation	Percentage of individuals aged 15 and older who commute to work by bicycle	0.324	0.640
	Percentage of individuals aged 15 and older in good health	0.322	
	Percentage of individuals aged 3 and older who engage in regular physical activity	0.354	
Public and Pedestrian Commuting	Percentage of individuals aged 15 and older who commute to work on foot	0.526	0.360
	Percentage of individuals aged 15 and older who commute to work using public transportation	0.474	

The PCA revealed that walking to work and using public transport load on the same component, whereas cycling to work loads with indicators of general health and regular physical activity. This grouping is entirely data-driven, reflecting correlations between the indicators, while the dimension names (Physical Health and Active Transportation and Public and Pedestrian Commuting) were chosen by the author to convey their conceptual meaning. These results suggest that walking to work is more strongly shaped by spatial proximity and urban structure, whereas cycling is more closely associated with additional physical activity. This empirical grouping is also consistent with the literature showing that cycling is more strongly associated with intentional physical activity and lifestyle choices, whereas walking to work often reflects spatial proximity and urban structure rather than health-oriented behaviour (Sallis et al. 2004). Figure 1 illustrates the rotated component space and the relationships between individual indicators and the two dimensions.

A key reason for employing PCA is to ensure that the indicator captures the multidimensionality of QoL (Maggino 2015). By leveraging the eigenvalue criteria (typically retaining components with eigenvalues > 1), PCA allows for the identification of the latent structure of the dataset and focuses on the dimensions that account for the most variance. This step ensures that the final composite indicator robustly reflects the core dimensions of active transportation and health.

After confirming the comparability of the variables, the weights derived from the PCA loadings are applied. These loadings are proportional to the degree of variance that each variable contributes to the principal components, ensuring that the most influential indicators have a greater impact on the final composite score. The weighted variables are aggregated using a linear additive model, resulting in a composite score for each region.

The technical specification applied during the PCA process is the Kaiser-normalized varimax rotation, which enhances the interpretability of the principal components by maximizing the variance of the squared loadings across components (Kaiser 1958) Kaiser–Meyer–Olkin -KMO- Test = 0.558 indicating a moderate level of adequacy for conducting factor analysis, and Bartlett's Test: $\chi^2 = 39.1475$; df =10; P-value < 0.001 strongly rejecting the null hypothesis of variable independence. This step ensures that each variable

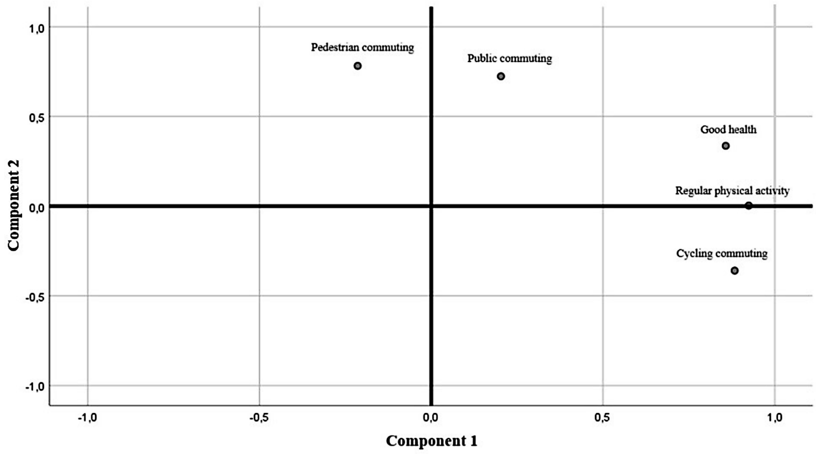


Figure 1: Graph of the rotated components

loads highly on one component, facilitating the interpretation of each dimension's contribution to the overall index. It also mitigates the issue of cross-loadings that could obscure the true influence of individual variables.

Sensitivity and robustness checks are crucial to this study to ensure the stability and reliability of the composite indicator (Greco et al. 2019). A total of 1000 Monte Carlo simulations were conducted using R version 4.1.2 (https://www.r-project.org/), randomly assigning weights from a Uniform distribution (0,1). A mean Pearson's correlation coefficient of 0.90 was calculated, which is considered good for assessing the robustness of the composite indicator. These simulations allowed for the evaluation of how small changes in methodological choices could impact the final rankings of the regions (Nardo et al. 2005).

In conclusion, the internal consistency of the composite indicator is supported by a Cronbach's alpha of 0.782 for the first dimension (Physical Health and Active Transportation), indicating that the variables within this dimension effectively measure a coherent underlying construct (Tavakol, Dennick 2011). The second dimension (Public and Pedestrian Commuting) has a lower alpha of 0.255; however, this is not a concern, as the component captures two complementary aspects of urban mobility – walking and public transport – that reflect distinct but equally relevant behaviours for assessing sustainable commuting patterns. Overall, the use of PCA for determining variable weights and the aggregation method provides a rigorous, statistically grounded framework for assessing regional QoL in Italy. This approach ensures that the composite indicator captures the multidimensional nature of the phenomenon while enabling a nuanced interpretation of the factors contributing to regional disparities, offering meaningful insights for policy and practice.

5 Results

In Table 3, the values are reported in ascending order by region, metropolitan area zone and Municipality size for individual dimensions, and for the composite indicator.

Table 3 reveals pronounced territorial disparities across Italian regions, metropolitan zones and municipality size, confirming a persistent North–Centre versus South divide. This pattern emerges consistently across both dimensions – Physical Health and Active Transportation and Public and Pedestrian Commuting – and is further reinforced in the Composite Indicator.

At the regional level, Southern regions (Calabria, Basilicata, Campania, Molise, Sicily and Puglia) systematically occupy the lower positions in the Composite Indicator ranking. This result is consistent with evidence from the Italian BES framework, according to which the share of individuals reporting good or very good health is, on average, around 10–15 percentage points higher in Northern regions compared to Southern ones.

Table 3: Values sorted in ascending order by region, metropolitan area zone and Municipality size of the two Dimensions and of the Composite Indicator

Region	Physical Health and Active Transportation	Region	Public and Pedestrian Commuting	Region	Composite Indicator
Calabria	0.08	Marche	0.07	Calabria	0.07
Basilicata	0.12	Veneto	0.08	Basilicata	0.11
Campania	0.14	Friuli-Venezia Giulia	0.34	Campania	0.15
Molise	0.18	Toscana	0.52	Marche	0.17
Sicilia	0.30	Umbria	0.52	Molise	0.17
Puglia	0.33	Sardegna	0.55	Veneto	0.23
Sardegna	0.39	Sicilia	0.57	Sicilia	0.23
Lazio	0.43	Emilia-Romagna	0.57	Sardegna	0.29
Liguria	0.43	Abruzzo	0.66	Puglia	0.30
Umbria	0.43	Puglia	0.73	Umbria	0.31
Abruzzo	0.47	Basilicata	0.74	Friuli-Venezia Giulia	0.33
Marche	0.51	Calabria	0.76	Abruzzo	0.39
Piemonte	0.53	Molise	0.78	Toscana	0.43
Friuli-Venezia Giulia	0.59	Trentino-Alto Adige	0.83	Lazio	0.47
Toscana	0.59	Piemonte	0.83	Liguria	0.49
Valle d'Aosta	0.62	Lombardia	0.86	Emilia-Romagna	0.51
Emilia-Romagna	0.65	Valle d'Aosta	0.90	Piemonte	0.53
Lombardia	0.65	Campania	0.92	Valle d'Aosta	0.66
Veneto	0.68	Lazio	0.94	Lombardia	0.67
Trentino-Alto Adige	1.00	Liguria	1.00	Trentino-Alto Adige	1.00
		Metropolitan Area (MA) zone			
Peripheral MA	0.17	Peripheral MA	0.06	Peripheral MA	0.10
Central MA	1.00	Central MA	1.00	Central MA	1.00
		Municipality size			
$\leq$ 2.000 in.	0.06	2.001 - 10.000 in.	0.32	$\leq$ 2.000 in.	0.05
2.001 - 10.000 in.	0.71	$\leq$ 2.000 in.	0.35	2.001 - 10.000 in.	0.30
10.001 - 50.000 in.	0.96	10.001 - 50.000 in.	0.81	10.001 - 50.000 in.	0.81
>50.000 in.	1.00	>50.000 in.	1.00	>50.000 in.	1.00

Similarly, BES indicators on physical activity show that regular participation in sport or physical exercise involves more than half of the population in several Northern and Central regions, while remaining well below 40% in most Southern regions. These gaps are coherently reflected in the lower values of the Physical Health and Active Transportation dimension observed for Southern Italy, where unfavourable socio-economic conditions and less supportive built environments tend to limit active lifestyles (Pirani, Salvini 2012, Costa et al. 2020).

The Public and Pedestrian Commuting dimension further mirrors well-documented regional disparities in transport infrastructure and service provision. BES data on mobility indicate that the availability and use of public transport are substantially higher in Northern regions, where satisfaction with local public transport services exceeds 60% in several cases, compared to values often below 40% in Southern regions (ISTAT 2022). These differences translate into lower effective commuting costs and higher labour market accessibility in the North and Centre, fostering greater reliance on public and pedestrian modes (Vendemmia, Beria 2023, Babapour et al. 2025). The high scores recorded by regions such as Trentino-Alto Adige, Lombardy and Emilia-Romagna in this dimension are therefore fully consistent with BES evidence on transport efficiency and service quality.

Disparities are even more pronounced when considering metropolitan area zones. Central metropolitan areas reach the maximum score across all dimensions, while peripheral metropolitan areas record substantially lower values. In many large cities, such as Rome, reliance on public transport is often driven by the inefficiency of alternatives; private ve-

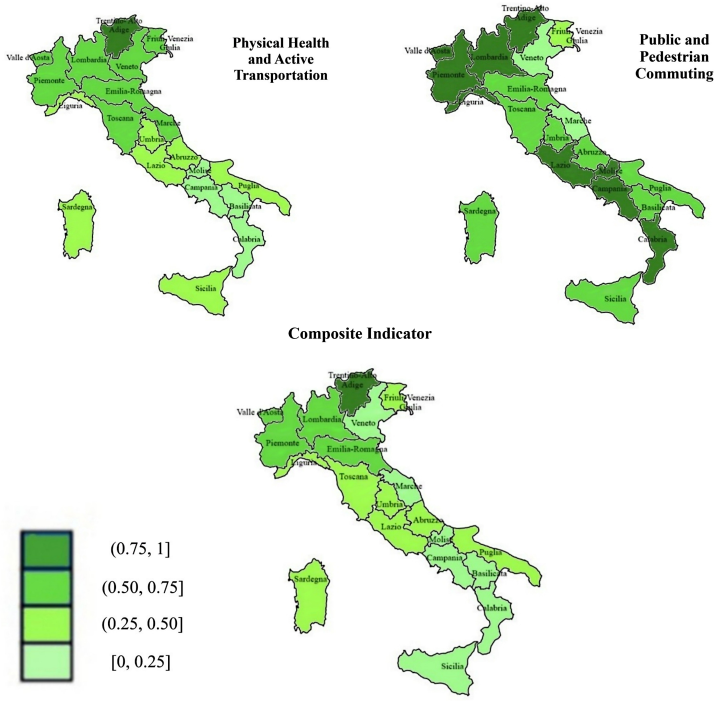


Figure 2: Geographical distribution of the values by region for the two Dimensions and the Composite Indicator

hicle use is often less viable due to congestion and parking challenges (Piccioni et al. 2019, Corazza, Carassiti 2021). This insight underscores that, while public transport is vital for sustainable mobility, its effectiveness and accessibility remain inconsistent across regions. This finding closely aligns with BES indicators on Accessibility to Services, which show that residents of central urban areas enjoy markedly shorter travel times to essential services and higher public transport availability compared to metropolitan peripheries (ISTAT 2022). Urban economics literature has long emphasised how density, mixed land use and functional integration in central areas reduce commuting costs and promote both active mobility and healthier lifestyles (Mariotti et al. 2021).

A similar gradient emerges across municipality size. Very small municipalities ($\leq$ 2,000 inhabitants) consistently score lowest on the Composite Indicator, while municipalities with more than 50,000 inhabitants achieve the highest values. BES data confirm that satisfaction with transport services and accessibility increases monotonically with municipality size, with gaps exceeding 20 percentage points between small rural municipalities and large urban centres (ISTAT 2022). While small municipalities may benefit from lower environmental stress, the scarcity of public transport and limited walkable infrastructure significantly constrain both sustainable commuting and daily physical activity (Giuffrida et al. 2021).

Figure 2 presents maps illustrating the two dimensions alongside the composite indicator, providing a visual understanding of the geographical distribution of values across Italian regions. This mapping enables a clearer analysis of regional patterns and disparities, offering insights into spatial variations in health and transportation metrics and emphasizing areas where interventions might enhance QoL outcomes.

The maps accentuate the regional contrasts between central-northern and southern Italian regions in both individual dimensions and the composite indicator. Veneto stands out as an exception: despite its overall central-northern location, it scores relatively low in the composite indicator due to a particularly low value in the Public and Pedestrian Transportation dimension. In contrast, while some southern regions, such as Molise, Campania, and Calabria, exhibit high scores in the Public and Pedestrian Transportation dimension, their composite indicator values remain lower. This is primarily because they score less favourably in the Physical Health and Active Transportation dimension, and the use of a geometric mean for aggregation results in limited compensation from higher values in one dimension when paired with lower scores in the other.

6 Conclusions

This study set out to develop and test a composite indicator capturing a specific, policy-relevant dimension of regional well-being in Italy, centred on the interaction between physical health, active lifestyles and sustainable mobility. The empirical results confirm pronounced territorial disparities, closely aligned with the long-standing North–Centre versus South divide, but the main contribution of the paper lies not merely in documenting these differences. Rather, it lies in demonstrating how health-related behaviours and everyday mobility practices jointly structure a distinct dimension of regional inequality that is only partially visible in existing multidimensional frameworks.

A first contribution of the study concerns conceptual clarification. While many QoL indices aim at broad coverage by combining economic, social, environmental, and subjective dimensions, the indicator proposed here deliberately adopts a narrower scope. The results show that even when attention is restricted to only two dimensions – Physical Health and Active Transportation, and Public and Pedestrian Commuting – substantial and systematic territorial gradients emerge. This suggests that health-related mobility behaviours constitute a core mechanism through which regional inequalities are reproduced. For this reason, the indicator should not be interpreted as a comprehensive measure of QoL. A more accurate interpretation is that of an Index of Health and Mobility, capturing a specific but structurally important component of regional well-being.

This positioning responds directly to the risk of conceptual overreach identified by the reviewers. Rather than competing with comprehensive frameworks such as the BES, the OECD Better Life Index or other multidimensional QoL measures, the proposed index should be understood as complementary to them. Its value lies precisely in isolating a behavioural and practice-based dimension – how people move and how this movement relates to health – that is often diluted within broader composite indicators. The strong empirical convergence with BES outcomes in the domains of health, environment and services supports the external validity of the index, while observed divergences across regions (e.g. intermediate performances of Lazio or Sardinia) highlight the additional interpretative leverage gained by focusing explicitly on health–mobility interactions.

A second contribution concerns policy interpretability. Unlike many regional statistics that separately report health outcomes, transport supply or infrastructure endowments, the proposed indicator integrates individual behaviours into a single framework. This allows policymakers to distinguish between different underlying configurations of disadvantage. For instance, low composite scores may result from weak transport accessibility, low engagement in physical activity, or both. Conversely, some regions display relatively good performance in one dimension but not in the other, revealing limited compensability between health and mobility outcomes. This feature is reinforced using a geometric mean, which prevents high performance in one domain from masking structural deficits in the other.

From a policy perspective, this has direct implications. Regions with low scores in both dimensions require integrated interventions – combining investments in public transport, walkability and cycling infrastructure with health promotion policies aimed at increasing physical activity. Regions where transport-related scores are relatively high but health-related scores lag behind may instead benefit more from targeted public health initiatives, sports infrastructure and programmes encouraging active lifestyles. In this sense, the indicator provides actionable insight beyond existing regional statistics, helping to prioritise policy levers rather than merely ranking territories.

The results for metropolitan areas and municipality size further strengthen this interpretation. The sharp contrast between central and peripheral metropolitan areas underscores the role of accessibility, density, and functional integration in shaping both commuting patterns and health behaviours. Similarly, the monotonic increase in scores with municipality size suggests that small municipalities face structural barriers not only in transport provision but also in opportunities for active living. These findings suggest that place-based policies should be differentiated not only by region but also by settlement structure, reinforcing the relevance of the index for multi-level governance.

At the same time, the study has clear limitations that must be acknowledged. The restricted number of dimensions, while enhancing interpretability, necessarily excludes other important aspects of well-being, such as income, housing conditions, environmental quality and subjective life satisfaction. Moreover, the indicator relies on cross-sectional data and on behaviours that may be shaped by necessity rather than choice, particularly in peripheral or disadvantaged areas. These limitations do not undermine the validity of the index but define its scope: it is not intended as a synthetic measure of overall QoL, but as a focused analytical tool addressing the health–mobility nexus.

Future research could build on this framework in several ways. One direction would be to integrate the Index of Health and Mobility into broader dashboards alongside BES indicators, explicitly analysing complementarities and trade-offs. Another would be to extend the analysis over time, assessing whether improvements in transport accessibility or active mobility infrastructure translate into measurable gains in health-related behaviours. Finally, comparative applications to other national contexts could test the generalisability of the approach and its relevance for European cohesion policy.

In conclusion, this study contributes to regional well-being research by proposing a theoretically grounded, empirically robust, and policy-oriented indicator that captures a specific but critical dimension of regional inequality. By focusing on health-related mobility behaviours, the Index of Health and Mobility offers a clear and actionable lens through which policymakers can interpret territorial disparities and design targeted, place-sensitive interventions, without claiming to replace more comprehensive QoL measures.

References

Abdi H, Williams LJ (2010) Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2: 433–459. CrossRef

Allam Z, Bibri SE, Chabaud D, Moreno C (2022) The '15-minute city' concept can shape a net-zero urban future. *Humanities and Social Sciences Communications* 9: 1–5. CrossRef

Anand S, Sen AK (1994) Human development index: Methodology and measurement. 5: 1433–1434. New York: Human Development Report Office Occasional Paper 12

Asso PF (2023) New perspectives on old inequalities: Italy's north–south divide. In: della Porta D, Keating M, Pianta M (eds), *Inequalities, Territorial Politics, Nationalism.* Routledge, 22–40

Babapour M, Corazza MV, Gentile G (2025) Urban commuting preferences in Italy: Employees' perceptions of public transport and willingness to adopt active transport based on k-modes cluster analysis. *Sustainability* 17: 5149. CrossRef

Balestra C, Boarini R, Ruiz N (2018) Going beyond GDP: Empirical findings. In: *Handbook of research on economic and social well-being.* Edward Elgar Publishing, 52–103. CrossRef

Baobeid A, Koç M, Al-Ghamdi SG (2021) Walkability and its relationships with health, sustainability, and livability: Elements of physical environment and evaluation frameworks. *Frontiers in built environment* 7: 721218. CrossRef

Bechtel G (2019) GDP is well-being! Results in the USA and China. *Open Journal of Social Sciences* 7: 189–204. CrossRef

Beria P, Debernardi A, Ferrara E (2017) Measuring the long-distance accessibility of Italian cities. *Journal of Transport Geography* 62: 66–79. CrossRef

Bircu C, Cavallaro F, Pozzer G, Nocera S (2024) Exploring the prospects and challenges of sustainable urban mobility: Potential and limits of cycling in Venice. *Journal of Transport and Land Use* 17: 401–422. CrossRef

Bowling A, Windsor J (2001) Towards the good life: A population survey of dimensions of QoL. *Journal of Happiness Studies* 2: 55–82. CrossRef

Bruzzi S, Ivaldi E, Santagata M (2022) Measuring regional performance in the Italian NHS: Are disparities decreasing? *Social Indicators Research* 159: 1057–1084. CrossRef

Cheshmehzangi A, Thomas SM (2016) Prioritizing accessible transit systems for sustainable urban development: Understanding and evaluating the parameters of a transportation system in Mumbai. *Journal of Urban Planning and Development* 142: 4. CrossRef

Corazza MV, Carassiti G (2021) Investigating maturity requirements to operate mobility as a service: The Rome case. *Sustainability* 13: 15. CrossRef

Costa G, Caiazzo A, Marinacci C, Spadea T (2020) Individual and contextual determinants of inequalities in health: The Italian case. In: *The political and social contexts of health*. Routledge, 115–146. CrossRef

Cottrill CD, Brooke S, Mulley C, Nelson JD, Wright S (2020) Can multi-modal integration provide enhanced public transport service provision to address the needs of vulnerable populations? *Research in Transportation Economics* 83: 100954. CrossRef

De Muro P, Monni S, Tridico P (2011) Knowledge-based economy and social exclusion: Shadow and light in the Roman socio-economic model. *International Journal of Urban and Regional Research* 35: 1212–1238. CrossRef

De Rosa D (2018) Capability approach and multidimensional well-being: The Italian case of BES. *Social Indicators Research* 140: 125–155. CrossRef

Diener E, Oishi S, Tay L (2018) Advances in subjective well-being research. *Nature Human Behaviour* 2: 253–260. CrossRef

El Gibari S, Gómez T, Ruiz F (2019) Building composite indicators using multicriteria methods: A review. *Journal of Business Economics* 89: 1–24. CrossRef

Engsted IM (2013) The European context: Measuring social inclusion in the European Union. In: Ruggeri Laderchi C, Savastano S (eds), *Poverty and exclusion in the Western Balkans: New directions in measurement and policy*. 13–27. CrossRef

Ferrara AR, Nisticò R (2015) Regional well-being indicators and dispersion from a multidimensional perspective: Evidence from Italy. *The Annals of Regional Science* 55: 373–420. CrossRef

Finger M, Serafimova T (2020) Towards a common European framework for sustainable urban mobility indicators

Geurs KT, van Wee B (2004) Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport geography* 12: 127–140. CrossRef

Giap TK, Thye WW, Aw G (2014) A new approach to measuring the liveability of cities: The Global Liveable Cities Index. *World Review of Science, Technology and Sustainable Development* 11: 176–196. CrossRef

Giuffrida N, Le Pira M, Inturri G, Ignaccolo M (2021) Addressing the public transport ridership/coverage dilemma in small cities: A spatial approach. *Case studies on transport policy* 9: 12–21. CrossRef

Greco S, Ishizaka A, Tasiou M, Torrisi G (2019) On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social indicators research* 141: 61–94. CrossRef

Huan N, Yao E, Xiao Y (2023) Roles of accessibility and air-rail intermodality in shaping mobility patterns in mega-city regions: Behavioural insights from China. *Cities* 143: 104591. CrossRef

ISTAT (2021) Aspetti della vita quotidiana 2021. https://www.istat.it/microdati/multi-scopo-sulle-famiglie-aspetti-della-vita-quotidiana/

ISTAT (2022) *Benessere equo e sostenibile (BES): Rapporto annuale*. Istituto Nazionale di Statistica. https://www.istat.it/tag/rapporto-bes/

Jolliffe IT (2002) Principal component analysis for special types of data. In: Jolliffe IT (ed), *Principal Component Analysis*. Springer, New York, 338–372. CrossRef

Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23: 187–200. CrossRef

Kim ES, Kubzansky LD, Soo J, Boehm JK (2017) Maintaining healthy behavior: A prospective study of psychological well-being and physical activity. *Annals of Behavioral Medicine* 51: 337–347. CrossRef

Lera-López F, Marco R (2018) Sports participation, physical activity, and health in the European regions. *Journal of sports sciences* 36: 1784–1791. CrossRef

Leydesdorff L (2021) Regions, innovations, and the north–south divide in Italy. In: Leydesdorff L (ed), *The Evolutionary Dynamics of Discursive Knowledge*. Springer, Cham, 115–134. CrossRef

Lucas K (2012) Transport and social exclusion: Where are we now? *Transport policy* 20: 105–113. CrossRef

Maggino F (2006) *Gli indicatori statistici: concetti, metodi e applicazioni*. Firenze University Press, Archivio E-Prints

Maggino F (2015) Assessing the subjective wellbeing of nations. In: Glatzer W, Camfield L, Møller V, Rojas M (eds), *Global handbook of QoL*. Springer, Dordrecht, 803–822. CrossRef

Maggino F (2017) Developing indicators and managing the complexity. In: Maggino F (ed), *Complexity in society*. Springer, Cham, 87–114. CrossRef

Maggino F (2023) *Encyclopedia of quality of life and well-being research*. Springer International Publishing, Cham. CrossRef

Maggino F, Zumbo BD (2011) Measuring the Quality of Life and the construction of social indicators. In: Land KC, Michalos AC, Sirgy MJ (eds), *Handbook of social indicators and Quality of Life research*. Springer, Dordrecht, 201–238. CrossRef

Mariotti I, Burlando C, Landi S (2021) Is local public transport unsuitable for elderly? exploring the cases of two italian cities. *Research in Transportation Business & Management* 40: 100643. CrossRef

Mazziotta M, Pareto A (2019) Use and misuse of PCA for measuring well-being. *Social Indicators Research* 142: 451–476. CrossRef

Medeiros E, Rauhut R (2020) Territorial cohesion cities: A policy recipe for achieving territorial cohesion? *Regional Studies* 54: 120–128. CrossRef

Miles L (2007) Physical activity and health. *Nutrition bulletin* 32: 314–363. CrossRef

Mizobuchi H (2014) Measuring world better life frontier: A composite indicator for OECD better life index. *Social Indicators Research* 118: 987–1007. CrossRef

Moreno C (2024) "the 15-minute city": Redesigning urban life with proximity to services. *Barcelona Societat; Journal on Social Knowledge and Analysis* 30

Nakamura K (2022) The relationship between walkability and QOL outcomes in residential evaluation. *Cities* 131: 104008. CrossRef

Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman S, Giovannini E (2005) Handbook on constructing composite indicators: Methodology and user guide. OECD statistics working papers, no. 2005/03, OECD publishing, Paris. CrossRef

Okulicz-Kozaryn A, Valente RR (2019) Livability and subjective well-being across European cities. *Applied research in quality of life* 14: 197–220. CrossRef

Papa E, Bertolini L (2015) Accessibility and transit-oriented development in European metropolitan areas. *Journal of transport geography* 47: 70–83. CrossRef

Papadopoulos E, Sdoukopoulos A, Politis I (2023) Measuring compliance with the 15-minute city concept: State-of-the-art, major components and further requirements. *Sustainable Cities and Society* 104875. CrossRef

Pazhuhan M, Shahraki SZ, Kaveerad N, Cividino S, Clemente M, Salvati L (2020) Factors underlying life quality in urban contexts: Evidence from an industrial city (Arak, Iran). *Sustainability* 12: 6. CrossRef

Piccioni C, Valtorta M, Musso A (2019) Investigating effectiveness of on-street parking pricing schemes in urban areas: An empirical study in Rome. *Transport Policy* 80: 136–147. CrossRef

Pirani E, Salvini S (2012) Place of living and health inequality: a study for elderly Italians. *Statistical Methods & Applications* 21: 211–226. CrossRef

Pontin FL, Jenneson VL, Morris MA, Clarke GP, Lomax NM (2022) Objectively measuring the association between the built environment and physical activity: A systematic review and reporting framework. *International Journal of Behavioral Nutrition and Physical Activity* 19: 119. CrossRef

Rey S (2025) Spatial inequality. *REGION* 12: 19–45. CrossRef

Rodríguez-Pose A (2020) Institutions and the fortunes of territories. *Regional Science Policy & Practice* 12: 371–386. CrossRef

Saelens BE, Handy SL (2008) Built environment correlates of walking: A review. *Medicine and science in sports and exercise* 40. CrossRef

Sallis JF, Floyd MF, Rodríguez DA, Saelens BE (2012) Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation* 125: 729–737. CrossRef

Sallis JF, Frank LD, Saelens BE, Kraft MK (2004) Active transportation and physical activity: opportunities for collaboration on transportation and public health research. *Transportation research part A: policy and Practice* 38: 249–268. CrossRef

Shrotryia VK, Singh SVP (2020) Measuring progress beyond GDP: A theoretical perspective. *Emerging Economy Studies* 6: 143–165. CrossRef

Simmie J, Martin R (2010) The economic resilience of regions: Towards an evolutionary approach. *Cambridge journal of regions, economy and society* 3: 27–43. CrossRef

Tavakol M, Dennick R (2011) Making sense of Cronbach's alpha. *International journal of medical education* 2: 53. CrossRef

Te Brömmelstroet M, Nikolaeva A, Glaser M, Nicolaisen MS, Chan C (2017) Travelling together alone and alone together: Mobility and potential exposure to diversity. *Applied Mobilities* 2: 1–15. CrossRef

Vendemmia B, Beria P (2023) When commuting is not enough: Towards a measure of territorial marginality based on job mobility. *Geographical Review* 113: 409–432. CrossRef

Verdugo MA, Schalock RL, Keith KD, Stancliffe RJ (2005) Quality of life and its measurement: Important principles and guidelines. *Journal of intellectual disability research* 49: 707–717. CrossRef

Warburton DE, Bredin SS (2017) Health benefits of physical activity: A systematic review of current systematic reviews. *Current opinion in cardiology* 32: 541–556. CrossRef

Wendling ZA, Emerson JW, De Sherbinin A, Esty DC (2020) *2020 Environmental performance index*. Yale Center for Environmental Law & Policy, New Haven, CT. CrossRef

Wiersma JK (2020) Commuting patterns and car dependency in urban regions. *Journal of Transport Geography* 84: 102700. CrossRef

Wolff P, Montaigne F, González GR (2010) Investing in statistics: EU-SILC. *Income and living conditions in Europe* 37

© 2026 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Louvain-la-Neuve, Belgium. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

REGION
The Journal of ERSA
Powered by WU

Volume 13, Number 1, 2026, 43–60
DOI: 10.18335/region.v13i1.575
journal homepage: region.ersa.org
ISSN: 2409-5370

Street View Imagery Analysis in a Computational Notebook

Yuhao Kang[1]

[1] The University of Texas at Austin, Austin TX, USA

Received: November 4, 2024/Accepted: January 6, 2026

Abstract. Street view imagery, capturing detailed streetscapes at human eye-level, has received significant attention in the past decade. Street view images can be leveraged to observe the built environment at both the element and scene levels. This chapter provides an introduction to methodologies for street view image-based analytics, including downloading street view images using Google Street View API, and employing advanced computer vision techniques such as Deep Convolutional Neural Networks to detect and quantify urban elements (trees, cars, buildings) and scenes (park, street). Particularly, this chapter introduces the use of image semantic segmentation to identify distinct urban elements, and image classification techniques to categorize and predict urban scene types. Further, an example of the calculation of the Green View Index (GVI) is provided to demonstrate how street view imagery analysis could contribute to urban data analytics. Through these methods, street view imagery not only helps model the digital environment and enhances our understanding of urban environments, but also offers a variety of insights into geography and urban studies.

Keywords: Street view imagery, urban environment, deep learning, AI, urban data science

1 Introduction

Street view imagery has emerged as a crucial data source for auditing built environments across various domains, including geography, urban planning, environmental science, criminology, and public health (Rzotkiewicz et al. 2018, Kang et al. 2020, Ibrahim et al. 2020, Biljecki, Ito 2021) . Street view images are typically collected as omni-directional panoramas, capturing detailed urban streetscapes, and providing visual information that closely reflects human eye-level perceptions(Anguelov et al. 2010). Because these images are captured at specific points in time, they offer temporal "snapshots" of the built environment. To protect privacy, faces, license plates, and in some cases façade details are automatically blurred. Street view images thus offer comprehensive perspectives for examining various aspects of the urban built environment and have benefited many downstream urban applications.

Beyond allowing users to navigate urban landscapes remotely through panoramic images for an immersive experience, street view images have several significant advantages over traditional data collection methods such as surveys, questionnaires, and mental mapping. These benefits include high coverage and volume, high data quality, low cost, and eye-level scenery (Kang et al. 2020). First, multiple global mapping service providers, including commercial companies like Google, Bing, Baidu, Tencent, and Yandex, as well as crowdsourced platforms such as Mapillary, offer extensive volumes of street view images covering global cities. Despite that street view imagery might be limited in Global South, this extensive coverage enables a comprehensive depiction of the built environment and allows researchers to perform cross-national comparisons of urban landscapes.

Second, street view images typically exhibit relatively higher quality. Images obtained from mapping service providers generally adhere to uniform standards and processes, having a consistent format that ensures a reliable image quality. Even crowdsourced platforms require specific format standards so that images are preprocessed to ensure their quality standards. Third, compared with traditional methods, street view images are effective and easily accessible. Users can conveniently and efficiently download these images to meet their demand at relatively low costs or no cost via APIs (Application Programming Interfaces) provided by mapping services. Moreover, users can customize various parameters of APIs such as location, camera angle, and image size, to get the data to meet specific research objectives. This accessibility facilitates the study of built environments using street view images in diverse cities and regions.

Last, and most crucially, street view images provide an objective observation of the built environment from a human perspective. Unlike aerial photographs and remote sensing images that offer bird's-eye views, street view imagery captures the urban environment from ground level, aligning closely with human perceptions. This feature makes street view imagery particularly effective for observing urban spaces and everyday urban life. Collectively, these benefits make street view images a vital data source for in-depth analysis of urban-related topics and for making informed decisions in urban planning and policy development.

More importantly, the rapid development of Artificial Intelligence (AI) offers various opportunities for quantifying urban features. Researchers are increasingly utilizing advanced computer vision techniques, such as deep learning, to identify and measure various elements (e.g., trees, vehicles) and scenes (urban functions) within the built environment (Li et al. 2015, Zhang et al. 2018). These efforts enable a comprehensive modeling of urban environmental settings, further enabling studies on both physical and socio-economic environments. For instance, researchers have quantified the proportion of urban greenery (Li et al. 2018), assessed the walkability and cyclability of environments (Yin, Wang 2016, Lu 2019), and examined factors contributing to obesogenic environments (Feuillet et al. 2016). These studies leveraged AI to enhance our understanding of urban dynamics, offering critical insights into the complex interplay between urban design and societal outcomes. Additionally, researchers have also assessed the socio-economic dimensions of cities, such as measuring urban gentrification (Ilic et al. 2019), human mobility (Zhang et al. 2019), urban functions (Ye et al. 2021), and housing prices (Kang et al. 2021). Moreover, beyond analyzing objective elements and scenes, researchers are utilizing street view images to delve into residents' subjective perceptions of urban environments. Such studies have provided insights into feelings of safety and depression (Zhang et al. 2018, Kang et al. 2023), thereby enhancing our understanding of the interplay between the built environment and human subjective experiences.

Given these advancements in large-scale street view images and AI methods for processing images, this chapter provides an overview of methods leveraged to utilize this valuable dataset. The following sections will provide detailed instructions about data collection and analysis at the object and scene levels. Several applications will be demonstrated to showcase how street view images might be leveraged to support urban and geospatial data analytics.

2 Computational Environment

In this section, we present the packages and computational environment used in our analysis. The *requests* and *ssl* libraries are used to communicate with web APIs (such as the Google Street View API) over secure connections. *NumPy* and *pandas* provide core numerical and tabular data structures, which we use to organize data. We utilize the *GeoPandas* package to handle and process geospatial data efficiently. We use the *OSMnx* package for downloading network data from OpenStreetMap. This tool helps in generating points along road networks. For visualization, we rely on *matplotlib*, which allows us to create maps and plots. The *PIL* package is utilized to help process imagery data and display street view images. The following codes has been tested on Python 3.9.

[1]:
```
import io
import json
import os
import requests
import ssl
import numpy as np
import pandas as pd
import osmnx as ox
import geopandas as gpd
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
from PIL import Image
```

To effectively implement deep learning methods, it is essential to install relevant packages and establish suitable computational environments. Several tools such as *pip* or *conda* has been widely utilized. *Conda* is a package and environment manager that installs libraries and keeps track of their versions. These packages can help manage the packages and environment, supporting debugging of the environment.

We primarily utilize the *PyTorch* framework, developed by the *Fundamental AI Research (FAIR)* team at Meta. PyTorch has become a popular platform in deep learning due to its comprehensive functionality and ecosystem. This framework provides a wide range of easily accessible pre-trained models, enabling users to apply advanced deep-learning techniques in a user-friendly way.

[2]:
```
# Load packages of PyTorch
import torch
from torchvision import transforms
print(torch.__version__) # we want 2.1.2
```

[2]:
```
2.1.2+cu121
```

PyTorch can be installed in two main ways: a CPU-only version and a GPU-enabled version. The CPU (Central Processing Unit) is the main processor in your computer and can run all of the code in this notebook, but deep learning models may run slowly on large datasets. A GPU (Graphics Processing Unit) is a specialized processor originally designed for graphics, but it is now widely used to accelerate deep learning computations.

For optimal performance, the use of NVIDIA GPUs coupled with the installation of CUDA is recommended to ensure the environment runs efficiently. If you are unsure whether your machine has a suitable GPU or CUDA installed, you may need to check your system specifications or consult your IT support. Unlike traditional methods that are performed on CPUs, utilizing GPUs for model training could significantly accelerate computational tasks. It is crucial for running such computationally intensive experiments on GPUs to analyze place scenes and extract objects from large-scale street view image dataset. To verify if the system is properly configured for GPU utilization, the following command can be used:

[3]:
```
print(torch.cuda.is_available())
```

[3]:
```
True
```

In addition to PyTorch, we incorporate *OpenMMLab*, an extensive toolkit that offers a variety of pre-trained models and tools for deep learning applications. This toolkit provides valuable resources for users so that they do not need to develop and train models from the ground up. By leveraging these advanced computational tools, our research can efficiently identify objects in the physical environment and understand place scenes from street view images. The OpenMMLab provides its own tools mmengine to help install a variety of deep learning models. This enables the researchers to download model packages efficiently.

[4]:
```
# !pip install -U openmim
# !mim install mmengine
# !mim install "mmcv>=2.0.0"
# !pip install "mmsegmentation>=1.0.0"
```

After the installation commands above, the subsequent code loads an OpenMMLab model.

[5]:
```
# Load packages of OpenMMLab
import mmcv
import mmseg
from mmseg.apis import MMSegInferencer
```

3 Street View Images Collection

To demonstrate the practical applications of street view images in urban environment observation, we will use the Mueller neighborhood in Austin, TX, United States as a case study. This example will guide you through the process of acquiring street view images via the Google Street View API. Users must first obtain an API Key from Google Maps, which is utilized to verify authorization and manage data usage. To obtain this API Key, readers can visit the *Google Cloud Platform*, register or log into the account, and follow the steps to generate a key. This key will further be utilized in their code to download street view images.

To download street view images, a prerequisite is to generate a set of geographic points from which street view images will be requested. To accomplish this, we use OSMnx to download nodes from OpenStreetMap that represent intersections and endpoints of streets and roads within the Mueller neighborhood. As illustrated below, nodes are downloaded within the Mueller neighborhood. Each node contains the unique OSM id and coordinates. In practice, beyond utilizing nodes downloaded from OpenStreetMap, users could generate their own geographic points for certain places manually. It should be noted that not all locations have street view images.

3.1 Download node data

[6]:
```
# Download nodes from OpenStreetMap using the OSMnx
G = ox.graph_from_place('Mueller, Austin, TX')
ox.save_graph_geopackage(G, filepath='data/mueller.gpkg')
mueller = gpd.read_file("data/mueller.gpkg")
mueller.head(5)
```

[6]:
```
       osmid          y          x  street_count          highway ref  \
0  152680140  30.288202 -97.700334             3
1  354665612  30.300224 -97.706409             4
2  354665831  30.305895 -97.707555             5  traffic_signals
3  354667206  30.307465 -97.711078             4         crossing
4  354668466  30.302508 -97.708340             3

                     geometry
0  POINT (-97.70033 30.28820)
1  POINT (-97.70641 30.30022)
2  POINT (-97.70755 30.30589)
3  POINT (-97.71108 30.30746)
4  POINT (-97.70834 30.30251)
```

[7]:
```
# Plot all nodes downloaded
mueller.plot(markersize=2)
```

3.2 Download street view images based on latitude and longitude

Once these geographic points are defined, two methods can be employed to download street view images. First, the *Google Street View API* allows users to retrieve street view images by providing latitude and longitude coordinates. This method is straightforward and convenient, returning the nearest street view image to the specific location. However,

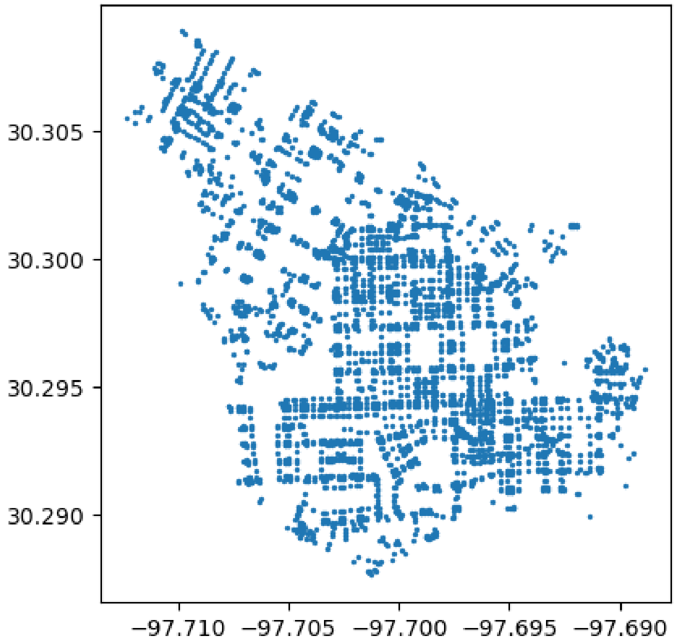


Figure 1: Example street view imagery points downloaded in Mueller, Austin. (Output of code chunk 7)

it should be noted that the image retrieved may not always precisely represent the given locations as it is the closest available image. More details of the parameters can be found in the *Google Street View API*.

[8]:
```python
# The function to download the street view image
def download_sv_latlng(lat, lng, img_path):
    API_key = "--------------------" # Put your API key here
    sv_params = {
        "size": "400x400",
        "location": f"{lat}, {lng}",
        "heading": "150",
        "pitch": "0",
        "key": API_key
    }
    response = requests.get(url="https://maps.googleapis.com/maps/api/streetview",␣
↪params=sv_params)
    print("Download status: ", response.status_code)
    img = Image.open(io.BytesIO(response.content))
    img.save(f"{img_path}")
    return img
```

After running the function, the notebook prints an HTTP response status code (e.g., Download status: 200). This code indicates whether the request to the Google Street View API was successful. A status code of 200 means "OK," indicating that the image was retrieved correctly. Other codes may indicate problems. For example, 400-series codes (such as 400 or 403) usually indicate client-side issues (e.g., invalid parameters, missing or incorrect API key, exceeding quota), whereas 500-series codes reflect server-side errors.

[9]:
```python
# Download a single street view image in the DataFrame mueller based on latitude and␣
↪longitude
sv_img = download_sv_latlng(30.288202, -97.700334, "node_152680140.jpg")
```

[9]:
```
Download status:  200
```

Figure 2 shows an example street view image downloaded given a specific geographic coordinate. It is the first node collected in the *mueller* dataframe.

Figure 2: An example street view imagery. (Output of code chunk 10)

[10]:
```
# Download a single street view image
plt.imshow(sv_img)
plt.axis('off')  # Turn off axis numbers and ticks
plt.title('Street View Image')
plt.show()
```

3.3 Download street view images based on panoids

The second method involves two key steps. Instead of directly retrieving street view images, users first obtain the *panoid*, which is a unique identifier for each street view image, along with its precise latitude and longitude. Once the panoid of the nearest street view image of the location is acquired, users can use it to request the specific street view image. This step does not involve downloading the images themselves but rather gathering information to locate the images requested. After identifying the necessary metadata and removing duplicates to avoid redundancy, the street view image for each unique node is then downloaded.

[11]:
```
# The function to get panoid of the street view image, given a specific coordinate
def get_panoid(lat, lng):
    API_key = "--------------------" # Put your API key here
    sv_params = {
        "size": "400x400",
        "location": f"{lat}, {lng}",
        "heading": "150",
        "pitch": "0",
        "key": API_key
    }
    response = requests.get(url="https://maps.googleapis.com/maps/api/streetview/
↪metadata", params=sv_params)
    panoid = json.loads(response.text)["pano_id"]
    return panoid
```

[12]:
```
# Get panoid of the street view image, given a specific coordinate
panoid = get_panoid(30.288202, -97.700334)
print("panoid: ", panoid)
```

[12]:
```
panoid:  igiDU33L1L1_vYg9CMyZZw
```

[13]:
```
# Get panoid for each osmid point
mueller["panoid"] = mueller.head(5).apply(lambda x: get_panoid(x["y"], x["x"]), axis=1)␣
↪# Just get panoids for the first five nodes
mueller.head(1)
```

[13]:
```
       osmid          y          x  street_count highway ref  \
0  152680140  30.288202 -97.700334             3

                  geometry                  panoid
0  POINT (-97.70033 30.28820)  igiDU33L1L1_vYg9CMyZZw
```

[14]:
```
# Drop duplicates as different places may have the same street view image
panoids = mueller["panoid"].dropna().unique()
print(panoids)
```

[14]:
```
['igiDU33L1L1_vYg9CMyZZw' 'UpIFSkcTE-KYAepxqtXWjw'
 'mzvCkPGNcD08bqMTam6pwA' 'Z7fSMVeWojQcJQG6N4nrGg'
 'wiErl1I_CT64If0GL42BoQ']
```

[15]:
```
# The function to download the street view image
def download_sv_panoid(panoid, img_path):
    API_key = "--------------------"
    sv_params = {
        "size": "400x400",
        "pano": f"{panoid}",
        "heading": "150",
        "pitch": "0",
        "key": API_key
    }
    response = requests.get(url="https://maps.googleapis.com/maps/api/streetview",␣
↪params=sv_params)
    print("Download status: ", response.status_code)
    img = Image.open(io.BytesIO(response.content))
    img.save(f"{img_path}")
    return img
```

[16]:
```
# Download a single street view image
sv_img = download_sv_panoid("igiDU33L1L1_vYg9CMyZZw", "igiDU33L1L1_vYg9CMyZZw.jpg")
```

[16]:
```
Download status:  200
```

Figure 3 shows an example street view image given a specific panoid. Both methods are valuable for accessing Google Street View images, with the choice of method depending on the user's specific needs regarding accuracy and the level of detail in imagery and metadata. By following this workflow, users can effectively leverage street view images to perform analysis of the urban environment.

[17]:
```
# Display the downloaded street view image
plt.imshow(sv_img)
plt.axis('off')  # Turn off axis numbers and ticks
plt.title('Street View Image')
plt.show()
```

This chapter's example focuses on Google Street View imagery. To collect crowd-sourced street view imagery from Mapillary, readers can refer to its website *Mapillary API*.

4 Urban Elements Extraction with Image Segmentation

With street view images retrieved, researchers could model the built environment at both the element- and scene-levels. Element-level observation refers to detecting and extracting individual objects, elements, or items from the urban environment, as well as calculating specific indices for these elements. One notable example refers to the calculation of urban greenery (Li et al. 2015). Other elements that have been analyzed include,

Figure 3: An example street view imagery. (Output of code chunk 17)

but are not limited to, features such as buildings (Gong et al. 2018) and traffic signs (Li et al. 2022). By detecting and measuring urban elements in streetscapes, researchers could further understand the composition and characteristics of urban spaces.

Image segmentation techniques have been widely used for element-level observation. Image segmentation refers to the process of partitioning an image into multiple segments, where each pixel is assigned a meaningful label or annotation. Several commonly detected objects include buildings, trees, roads, and vehicles. The emergence of deep learning methods, such as deep convolutional neural networks (DCNNs), has enabled researchers to extract high-dimensional visual features, facilitating a deeper understanding of images, and achieved high accuracy in image segmentation. Several notable methods include SegNet (Badrinarayanan et al. 2017) and ResNet (He et al. 2016).

In this chapter, we provide an example of using DeepLabV3 (Chen et al. 2018), a specific DCNN, for semantic segmentation, utilizing the ADE20K dataset. The ADE20K dataset is a large-scale annotated image dataset with over 27K images (Zhou et al. 2017). It depicts a variety of place scenes and provides comprehensive labels for numerous objects, such as buildings, vegetation, and vehicles, making it well-suited for urban element-level observation. Based on this dataset, a DeepLabV3 model has been trained to detect 150 objects from images. DeepLabV3 is a state-of-the-art DCNN model for semantic segmentation. It combines atrous (dilated) convolutions with an Atrous Spatial Pyramid Pooling (ASPP) module to capture multiscale contextual information, allowing the model to distinguish objects of different sizes in complex street scenes. It is highly effective for analyzing the urban environment by distinguishing different objects in place scenes, and has achieved competitive accuracy on the ADE20K dataset. We also acknowledge a few limitations of the current deep learning-based methods. The segmentation quality may vary across object classes and image conditions. For example, certain objects might be missing in the training data. Some images might be blurred, which will result in inaccurate results.

To perform image segmentation, the pretrained DeepLabV3 model is first downloaded using the MMSegmentation toolkit, which provides access to a variety of models for semantic segmentation with a simple command line (Chen et al. 2019). It should be noted that, the readers could download other types of deep learning methods based on the needs. In this chapter, we take the following model as an example to show case ho wo perform image semantic segmentation.

4.1 Image Semantic Segmentation based on ADE20K using DeepLabV3

[18]:
```
# Create the model and output directory
# !mkdir ./chec
# !mkdir ./segmentation_results
```

[19]:
```
# Download the deep learning model
# !mim download mmsegmentation --config deeplabv3plus_r101-d8_4xb4-160k_ade20k-512x512␣
↪--dest ./chec
```

[20]:
```
# Set image file, model file and checkpoint path
img= 'node_152680140.jpg'
model = "chec/deeplabv3plus_r101-d8_4xb4-160k_ade20k-512x512.py"
cpt = "chec/deeplabv3plus_r101-d8_512x512_160k_ade20k_20200615_123232-38ed86bb.pth"
# Set output directory
outdir = "segmentation_results/"
```

Rather than training the model from scratch, we load these pretrained weights and apply the model to our street view images to obtain pixel-level labels. This step is often called inference, i.e., running the trained model on new images to generate predictions (segmentation maps).

[21]:
```
# Perform image semantic segmentation with the downloaded DeepLabV3 model
mmseg_inferencer = MMSegInferencer(model, cpt, dataset_name= "ade20k")
result = mmseg_inferencer(img, out_dir = outdir, opacity = 1, with_labels = False)
```

[21]:
```
Inference ----------------------------------------

/miniconda3/envs/geo/lib/python3.11/site-
packages/mmengine/visualization/visualizer.py:196: UserWarning: Failed to add
<class 'mmengine.visualization.vis_backend.LocalVisBackend'>, please provide the
`save_dir` argument.
  warnings.warn(f'Failed to add {vis_backend.__class__}, '
```

4.2 Visualization of Results

After inferencing the image using the selected model, the results are presented as follows. The output is a matrix that matches the dimensions of the input image, with each pixel assigned a value that corresponds to a specific urban object. The segmented output is then visualized by rendering each element in a distinct color within the streetscape. The class names from the ADE20K dataset and their corresponding colors are predefined in the *mmsegmentation* tool, and details can be accessed through the provided links for object names and color codings. This allows researchers to vividly understand and analyze the composition of the built environment.

[22]:
```
# Display results
print(result)
print("Image size: ", result['predictions'].shape)
```

[22]:
```
{'predictions': array([[2, 2, 2, ..., 2, 2, 2],
       [2, 2, 2, ..., 2, 2, 2],
       [2, 2, 2, ..., 2, 2, 2],
       ...,
       [6, 6, 6, ..., 6, 6, 6],
       [6, 6, 6, ..., 6, 6, 6],
       [6, 6, 6, ..., 6, 6, 6]]), 'visualization': []}
Image size:  (400, 400)
```

[23]:
```
# mmsegmentation predefined object names, color schemes and palettes
# objects: https://github.com/open-mmlab/mmsegmentation/blob/main/mmseg/utils/
↪class_names.py#L15-L42
# color schemes and palettes: https://github.com/open-mmlab/mmsegmentation/blob/main/
↪mmseg/utils/class_names.py#L273-L312

obj_classes=['wall', 'building', 'sky', 'floor', 'tree', 'ceiling', 'road', 'bed ',␣
↪'windowpane', 'grass', 'cabinet', 'sidewalk', 'person', 'earth', 'door',
         'table', 'mountain', 'plant', 'curtain', 'chair', 'car', 'water', 'painting',␣
↪'sofa', 'shelf', 'house', 'sea', 'mirror', 'rug', 'field', 'armchair',
         'seat', 'fence', 'desk', 'rock', 'wardrobe', 'lamp', 'bathtub', 'railing',␣
↪'cushion', 'base', 'box', 'column', 'signboard', 'chest of drawers',
         'counter', 'sand', 'sink', 'skyscraper', 'fireplace', 'refrigerator',␣
↪'grandstand', 'path', 'stairs', 'runway', 'case', 'pool table', 'pillow',
         'screen door', 'stairway', 'river', 'bridge', 'bookcase', 'blind', 'coffee␣
↪table', 'toilet', 'flower', 'book', 'hill', 'bench', 'countertop',
         'stove', 'palm', 'kitchen island', 'computer', 'swivel chair', 'boat', 'bar',␣
↪'arcade machine', 'hovel', 'bus', 'towel', 'light', 'truck',
         'tower', 'chandelier', 'awning', 'streetlight', 'booth', 'television␣
↪receiver', 'airplane', 'dirt track', 'apparel', 'pole', 'land', 'bannister',
         'escalator', 'ottoman', 'bottle', 'buffet', 'poster', 'stage', 'van', 'ship',␣
↪'fountain', 'conveyer belt', 'canopy', 'washer', 'plaything', 'swimming pool',
         'stool', 'barrel', 'basket', 'waterfall', 'tent', 'bag', 'minibike', 'cradle',␣
↪'oven', 'ball', 'food', 'step', 'tank', 'trade name', 'microwave', 'pot', 'animal',
         'bicycle', 'lake', 'dishwasher', 'screen', 'blanket', 'sculpture', 'hood',␣
↪'sconce', 'vase', 'traffic light', 'tray', 'ashcan', 'fan', 'pier', 'crt screen',
         'plate', 'monitor', 'bulletin board', 'shower', 'radiator', 'glass', 'clock',␣
↪'flag']

palette= [[120, 120, 120], [180, 120, 120], [6, 230, 230], [80, 50, 50], [4, 200, 3],␣
↪[120, 120, 80], [140, 140, 140], [204, 5, 255], [230, 230, 230], [4, 250, 7],
          [224, 5, 255], [235, 255, 7], [150, 5, 61], [120, 120, 70], [8, 255, 51],␣
↪[255, 6, 82], [143, 255, 140], [204, 255, 4], [255, 51, 7], [204, 70, 3],
          [0, 102, 200], [61, 230, 250], [255, 6, 51], [11, 102, 255], [255, 7, 71],␣
↪[255, 9, 224], [9, 7, 230], [220, 220, 220], [255, 9, 92], [112, 9, 255],
          [8, 255, 214], [7, 255, 224], [255, 184, 6], [10, 255, 71], [255, 41, 10],␣
↪[7, 255, 255], [224, 255, 8], [102, 8, 255], [255, 61, 6], [255, 194, 7],
          [255, 122, 8], [0, 255, 20], [255, 8, 41], [255, 5, 153], [6, 51, 255], [235,␣
↪12, 255], [160, 150, 20], [0, 163, 255], [140, 140, 140], [250, 10, 15],
          [20, 255, 0], [31, 255, 0], [255, 31, 0], [255, 224, 0], [153, 255, 0], [0,␣
↪0, 255], [255, 71, 0], [0, 235, 255], [0, 173, 255], [31, 0, 255], [11, 200, 200],
          [255, 82, 0], [0, 255, 245], [0, 61, 255], [0, 255, 112], [0, 255, 133],␣
↪[255, 0, 0], [255, 163, 0], [255, 102, 0], [194, 255, 0], [0, 143, 255], [51, 255, 0],
          [0, 82, 255], [0, 255, 41], [0, 255, 173], [10, 0, 255], [173, 255, 0], [0,␣
↪255, 153], [255, 92, 0], [255, 0, 255], [255, 0, 245], [255, 0, 102], [255, 173, 0],
          [255, 0, 20], [255, 184, 184], [0, 31, 255], [0, 255, 61], [0, 71, 255],␣
↪[255, 0, 204], [0, 255, 194], [0, 255, 82], [0, 10, 255], [0, 112, 255], [51, 0, 255],
          [0, 194, 255], [0, 122, 255], [0, 255, 163], [255, 153, 0], [0, 255, 10],␣
↪[255, 112, 0], [143, 255, 0], [82, 0, 255], [163, 255, 0], [255, 235, 0], [8, 184,␣
↪170],
          [133, 0, 255], [0, 255, 92], [184, 0, 255], [255, 0, 31], [0, 184, 255], [0,␣
↪214, 255], [255, 0, 112], [92, 255, 0], [0, 224, 255], [112, 224, 255], [70, 184,␣
↪160],
          [163, 0, 255], [153, 0, 255], [71, 255, 0], [255, 0, 163], [255, 204, 0],␣
↪[255, 0, 143], [0, 255, 235], [133, 255, 0], [255, 0, 235], [245, 0, 255], [255, 0,␣
↪122],
          [255, 245, 0], [10, 190, 212], [214, 255, 0], [0, 204, 255], [20, 0, 255],␣
↪[255, 255, 0], [0, 153, 255], [0, 41, 255], [0, 255, 204], [41, 0, 255], [41, 255, 0],
          [173, 0, 255], [0, 245, 255], [71, 0, 255], [122, 0, 255], [0, 255, 184], [0,␣
↪92, 255], [184, 255, 0], [0, 133, 255], [255, 214, 0], [25, 194, 194], [102, 255, 0],
          [92, 0, 255]]
```

[24]:
```
# Match objects and colors
unique_colors = np.unique(result['predictions'].reshape(-1, 1), axis=0)
exist_classes = [obj_classes[int(index)] for index in unique_colors]
exist_palette = [palette[int(index)] for index in unique_colors]
print("The following objects are detected in this image: ", exist_classes)
print("Their palettes are: ", exist_palette)
```

Figure 4: Example of a street view imagery (left) and its corresponding semantic segmentation map (right), where pixels are classified into built-environment elements (e.g., wall, buildings, and sky). (Output of code chunk 25)

[24]:
```
The following objects are detected in this image:  ['wall', 'building', 'sky',
'tree', 'road', 'grass', 'sidewalk', 'car', 'house', 'fence', 'signboard',
'streetlight', 'pole']
Their palettes are:  [[120, 120, 120], [180, 120, 120], [6, 230, 230], [4, 200,
3], [140, 140, 140], [4, 250, 7], [235, 255, 7], [0, 102, 200], [255, 9, 224],
[255, 184, 6], [255, 5, 153], [0, 71, 255], [51, 0, 255]]

/tmp/ipykernel_59716/2720650808.py:3: DeprecationWarning: Conversion of an array
with ndim > 0 to a scalar is deprecated, and will error in future. Ensure you
extract a single element from your array before performing this operation.
(Deprecated NumPy 1.25.)
  exist_classes = [obj_classes[int(index)] for index in unique_colors]
/tmp/ipykernel_59716/2720650808.py:4: DeprecationWarning: Conversion of an array
with ndim > 0 to a scalar is deprecated, and will error in future. Ensure you
extract a single element from your array before performing this operation.
(Deprecated NumPy 1.25.)
  exist_palette = [palette[int(index)] for index in unique_colors]
```

[25]:
```python
# Plot the original and segmented images
out_img_path = os.path.join(outdir, 'vis', img.split('/')[-1])
out_img = mmcv.imread(out_img_path)
orig_img = mmcv.imread(img)

# Create a plot to compare the original and segmented images
fig, ax = plt.subplots(1, 2, figsize=(14,6))
ax[0].imshow(mmcv.bgr2rgb(orig_img))
ax[0].axis('off')
ax[1].imshow(mmcv.bgr2rgb(out_img))
ax[1].axis('off')

# Create a legend for every color
patches = [mpatches.Patch(color=np.array(exist_palette[i])/255.,
                          label=exist_classes[i]) for i in range(len(exist_classes))]
ax[1].legend(handles=patches, bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.,
           fontsize='large')

# Plot the result
plt.tight_layout()
plt.show()
```

5 Urban Scenes Evaluation using Image Classification

With street view images retrieved, researchers can also evaluate the entire urban scene to have an overall assessment of the environment and understand the quality of urban environments. Beyond individual elements, scene-level observation could offer insights into the overall characteristics of the urban landscape, such as place type classification, assessment of the neighborhood quality and functionality. To perform scene-level observation, researchers often leverage DCNNs to perform image classification (He et al. 2016). Image classification refers to assigning a label or category to an entire image based on its visual content. Here, we utilize ResNet for scene-level observation, which is an advanced deep learning model that contains residual connections for effective understanding of complex images. ResNet has been widely leveraged for image classification in various image datasets such as ImageNet (Deng et al. 2009) and Places365 (Zhou et al. 2017). The ResNet model is leveraged to classify place scenes based on the Places365 dataset. Places365 dataset contains millions of labeled images representing a wide variety of scenes, including residential neighborhoods, parks, and hospitals. In total, 365 place scenes are defined. By leveraging Places365 dataset and ResNet, researchers can efficiently categorize urban scenes based on their visual characteristics, enabling scene-level observation. We first downloaded the model and the categories of place scenes.

[26]:
```
# !wget https://raw.githubusercontent.com/CSAILVision/places365/refs/heads/master/
↪categories_places365.txt -P ./chec
# !wget http://places2.csail.mit.edu/models_places365/resnet50_places365.pth.tar -P ./
↪chec
```

5.1 Identify urban scenes from street view images

[27]:
```
from torch.autograd import Variable as V
import torchvision.models as models
from torchvision import transforms as trn
from torch.nn import functional as F
```

[28]:
```
# Load the pre-trained weights
arch = 'resnet50'
model = models.__dict__[arch](num_classes=365)
checkpoint = torch.load("chec/resnet50_places365.pth.tar", map_location=lambda storage,␣
↪loc: storage)
state_dict = {str.replace(k,'module.',''): v for k,v in checkpoint['state_dict'].
↪items()}
model.load_state_dict(state_dict)
model.eval()

# Load the image transformer
centre_crop = trn.Compose([
        trn.Resize((256,256)),
        trn.CenterCrop(224),
        trn.ToTensor(),
        trn.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])
])

# Load the class label
classes = list()
with open("chec/categories_places365.txt") as class_file:
    for line in class_file:
        classes.append(line.strip().split(' ')[0][3:])
classes = tuple(classes)
```

[29]:
```
# Read and process the input street view image
img_array = Image.open(img)
input_img = V(centre_crop(img_array).unsqueeze(0))

# Forward pass
logit = model.forward(input_img)
h_x = F.softmax(logit, 1).data.squeeze()
probs, idx = h_x.sort(0, True)

# Display the input street view image
plt.imshow(img_array)
plt.axis('off')  # Turn off axis numbers and ticks
plt.title('Street View Image')
plt.show()
print('Input image', img)
# Output the prediction
print(f"Scene type: {classes[idx[0]]}")
print(f"Top 5 scene type:")
for i in range(0, 5):
    print(f'{classes[idx[i]]}: {probs[i]}')
```

Street View Image

Figure 5: Example of a street view imagery and its corresponding semantic scene. (Output of code chunk 29)

[29]:
```
Input image node_152680140.jpg
Scene type: industrial_area
Top 5 scene type:
industrial_area: 0.4464263319969177
residential_neighborhood: 0.2901725769042969
highway: 0.15815606713294983
field/wild: 0.025090044364333153
parking_lot: 0.018267285078763962
```

As shown in Figure 5, for a given street view image, the model generates an array of scene categories along with their corresponding probabilities, showing the confidence level of each classification. For instance, the top 5 predicted scenes of the image include an industrial area, residential neighborhood, highway, field, and parking lot. In particular, the top two categories of scenes have a combined probability of approximately 73%. This detailed output helps researchers effectively assess the place type of the contexts of the urban environments.

6 Urban Greenery Calculation

After performing element-level and scene-level observations of the urban environment, these valuable insights could be incorporated into urban data analytics to better understand and model the built environments. Here, we take the measurement of urban greenery as an example to demonstrate how greenery can be quantified from street view images. Urban greenery refers to the presence of vegetation, such as trees, shrubs, and grass, in urban areas. It has been widely studied in various urban research literature (Li et al. 2015, Yang et al. 2021, Lu et al. 2023). Researchers have developed the Green View Index (GVI) to measure the degree of physical greenness within the built environment. In the following paragraphs, we compute the average Green View Index (GVI) for the Mueller neighborhood using street view images. The equation for calculating the Green View Index is defined as follows:

$$GVI = \frac{N_{pixels_g}}{N_{pixels_a}}$$

whereN_{pixels_g} refers to the number of pixels that have been identified as urban greenery (e.g., trees grass, plant, flower, palm, and canopy) of the given image, andN_{pixels_a} indicates the total number of pixels of the street view image

Below, we present an example of how to calculate the Green View Index (GVI) for a given street view image. Specifically, the GVI calculation includes the following categories: trees, grass, plants, flowers, palms, and canopies. The proportions of each category are first computed individually, and then aggregated to determine the overall GVI percentage, providing a comprehensive measure of greenery in the image..

6.1 Compute the Green View Index

[30]:
```
# Display the input street view image
plt.imshow(img_array)
plt.axis('off')  # Turn off axis numbers and ticks
plt.title('Street View Image')
plt.show()

# Compute the Green View Index of a given street view image
total_pixel = result['predictions'].shape[0]*result['predictions'].shape[1]

tree_percentage = (np.sum(result['predictions'] == 4) / total_pixel)*100
print(f"{obj_classes[4]} percentage: {tree_percentage:.2f}%")

grass_percentage = (np.sum(result['predictions'] == 9) / total_pixel)*100
print(f"{obj_classes[9]} percentage: {grass_percentage:.2f}%")

plant_percentage = (np.sum(result['predictions'] == 17) / total_pixel)*100
print(f"{obj_classes[17]} percentage: {plant_percentage:.2f}%")

flower_percentage = (np.sum(result['predictions'] == 66) / total_pixel)*100
print(f"{obj_classes[66]} percentage: {flower_percentage:.2f}%")

palm_percentage = (np.sum(result['predictions'] == 72) / total_pixel)*100
print(f"{obj_classes[72]} percentage: {palm_percentage:.2f}%")

canopy_percentage = (np.sum(result['predictions'] == 106) / total_pixel)*100
print(f"{obj_classes[106]} percentage: {canopy_percentage:.2f}%")

print(f"Total Green:␣
↪{tree_percentage+grass_percentage+plant_percentage+flower_percentage+
palm_percentage+canopy_percentage:.2f}%")
```

[30]:
```
tree percentage: 8.44%
grass percentage: 4.52%
plant percentage: 0.00%
flower percentage: 0.00%
palm percentage: 0.00%
canopy percentage: 0.00%
Total Green: 12.96%
```

For instance, as illustrated in Figure 6, the proportion of trees in the street view image is 8.44%, and grass accounts for 4.5%, resulting in an overall GVI of approximately 13% when all types of urban greenery are considered. Researchers can then iterate over all GVI values within the study area and aggregate them to the neighborhood level for future spatial analysis. These GVI calculations can then be integrated into broader urban research to provide insights into the relationships between urban greenery and various environmental and social aspects of urban environments.

Additionally, the following cell provides codes that combines image semantic segmentation with the calculations of the Green View Index (GVI). This allows for the efficient computation of the GVI from any provided street view image, enabling researchers to assess the extent of urban greenery present in street view images.

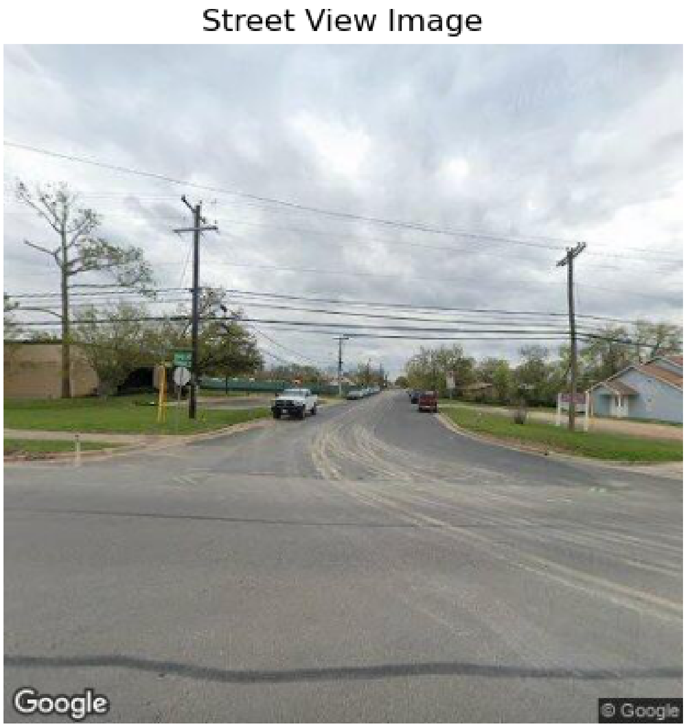


Figure 6: Example of a street view imagery and its corresponding element percentages. (Output of code chunk 30)

We acknowledge potential limitations for the GVI. When extending this calculation to many images along a road network, it is important to recognize that overlapping scenes may occur. Images captured at nearby points can contain very similar or even nearly identical views. If a particularly green segment of a street is sampled many times, this overlap may inflate neighborhood-level GVI estimates.

[31]:
```
# The function to calculate the Green View Index of a given street view image
def cal_gvi(result):
    total_pixel = result['predictions'].shape[0]*result['predictions'].shape[1]
    tree_percentage = (np.sum(result['predictions'] == 4) / total_pixel)*100
    grass_percentage = (np.sum(result['predictions'] == 9) / total_pixel)*100
    plant_percentage = (np.sum(result['predictions'] == 17) / total_pixel)*100
    flower_percentage = (np.sum(result['predictions'] == 66) / total_pixel)*100
    palm_percentage = (np.sum(result['predictions'] == 72) / total_pixel)*100
    canopy_percentage = (np.sum(result['predictions'] == 106) / total_pixel)*100
    gvi = tree_percentage+grass_percentage+plant_percentage+flower_percentage+
    palm_percentage+canopy_percentage
    return gvi
```

[32]:
```
# Given an image, perform image semantic segmentation and calculates the GVI
img= 'node_152680140.jpg'
print("The GVI is: ", cal_gvi(mmseg_inferencer(img, out_dir = outdir, opacity = 1,␣
↪with_labels = False)), "%")
```

[32]:
```
Inference ----------------------------------------

The GVI is:  12.959375 %
```

7 Conclusion

The emergence of street view images provides a valuable data source for auditing the built environment. Compared to traditional methods such as field surveys and secondary data sources, street view images capture the urban physical environment from an eye-level perspective, significantly reducing costs and increasing efficiency. In this chapter, we provided detailed guidance on retrieving street view images using the Google Street View API. Then, we demonstrated how deep learning techniques can be effectively leveraged for both element-level and scene-level observations of the built environment. At the element level, individual urban features, such as trees, buildings, and vehicles, can be identified and detected through image segmentation. We showcased the use of the DeepLabV3 model for image segmentation, where up to 150 categories of objects can be identified from street-view images. At the scene level, we employed image classification techniques using the ResNet model, trained on the Places365 dataset, to categorize urban scenes based on their place types. This classification enables researchers to analyze different urban settings. After illustrating element- and scene-level observations, we calculated the Green View Index (GVI) of a given street view image to demonstrate how street view image analysis can benefit urban studies to advance our understanding of the built environment. The GVI computation allows for quantifying the presence of greenery within the urban environment, supporting a comprehensive understanding of streetscapes, and may benefit a wide range of potential urban applications. Overall, this chapter highlights the significant potential of street view imagery and deep learning in urban environment observation. By integrating advanced methods such as image segmentation and image classification, street view imagery ultimately contributes to more informed geospatial and urban studies.

References

Anguelov D, Dulong C, Filip D, Frueh C, Lafon S, Lyon R, Ogale A, Vincent L, Weaver J (2010) Google street view: Capturing the world at street level. *Computer* 43: 32–38. CrossRef

Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39: 2481–2495. CrossRef

Biljecki F, Ito K (2021) Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning* 215: 104217. CrossRef

Chen K, Wang J, Pang J, Cao Y, Xiong Y, Li X, Sun S, Feng W, Liu Z, Xu J, Zhang Z, Cheng D, Zhu C, Cheng T, Zhao Q, Li B, Lu X, Zhu R, Wu Y, Dai J, Wang J, Shi J, Ouyang W, Loy CC, Lin D (2019) Mmdetection: Open mmlab detection toolbox and benchmark. arXiv preprint, 190607155

Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the european conference on computer vision (eccv) 2018. CrossRef

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. 2009 ieee conference on computer vision and pattern recognition, ieee. CrossRef

Feuillet T, Charreire H, Roda C, Ben Rebah M, Mackenbach J, Compernolle S, Glonti K, Bárdos H, Rutter H, De Bourdeaudhuij I, McKee M, Brug J, Lakerveld J, Oppert JM (2016) Neighbourhood typology based on virtual audit of environmental obesogenic characteristics. *Obesity Reviews* 17: 19–30. CrossRef

Gong FY, Zeng ZC, Zhang F, Li X, Ng E, Norford LK (2018) Mapping sky, tree, and building view factors of street canyons in a highdensity urban environment. *Building and Environment* 134: 155–167. CrossRef

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Proceedings of the ieee conference on computer vision and pattern recognition, 2016. CrossRef

Ibrahim MR, Haworth J, Cheng T (2020) Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities* 96: 102481. CrossRef

Ilic L, Sawada M, Zarzelli A (2019) Deep mapping gentrification in a large canadian city using deep learning and Google Street View. *PloS one* 14: e0212814. CrossRef

Kang Y, Abraham J, Ceccato V, Duarte F, Gao, S. and LL, Zhang F, Näsman P, Ratti C (2023) Assessing differences in safety perceptions using geoai and survey across neighbourhoods in Stockholm, Sweden. *Landscape and Urban Planning* 236: 104768. CrossRef

Kang Y, Zhang F, Gao S, Lin H, Liu Y (2020) A review of urban physical environment sensing using street view imagery in public health studies. *Annals of GIS* 26: 261–275. CrossRef

Kang Y, Zhang F, Gao S, Peng W, Ratti C (2021) Human settlement value assessment from a place perspective: Considering human dynamics and perceptions in house price modeling. *Cities* 118: 103. CrossRef

Li X, Ning H, Huang X, Dadashova B, Kang Y, Ma A (2022) Urban infrastructure audit: An effective protocol to digitize signalized intersections by mining street view images. *Cartography and Geographic Information Science* 49: 32–49. CrossRef

Li X, Ratti C, Seiferling I (2018) Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landscape and Urban Planning* 169: 81–91. CrossRef

Li X, Zhang C, Li W, Ricard R, Meng Q, Zhang W (2015) Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening* 14: 675. CrossRef

Lu Y (2019) Using Google Street View to investigate the association between street greenery and physical activity. *Landscape and Urban Planning* 191: 103435. CrossRef

Lu Y, Ferranti EJS, Chapman L, Pfrang C (2023) Assessing urban greenery by harvesting street view data: A review. *Urban Forestry & Urban Greening* 83: 127917. CrossRef

Rzotkiewicz A, Pearson AL, Dougherty BV, Shortridge A, Wilson N (2018) Systematic review of the use of Google Street View in health research: major themes, strengths, weaknesses and possibilities for future research. *Health & Place* 52: 240–246. CrossRef

Yang J, Rong H, Kang Y, Zhang F, Chegut A (2021) The financial impact of street-level greenery on New York commercial buildings. *Landscape and Urban Planning* 214: 104162. CrossRef

Ye C, Zhang F, Mu L, Gao Y, Liu Y (2021) Urban function recognition by integrating social media and street-level imagery. *Environment and Planning B: Urban Analytics and City Science* 48: 1430–1444. CrossRef

Yin L, Wang Z (2016) Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Applied Geography* 76: 147–153. CrossRef

Zhang F, Wu L, Zhu D, Liu Y (2019) Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. *ISPRS Journal of Photogrammetry and Remote Sensing* 153: 48–58. CrossRef

Zhang F, Zhang D, Liu Y, Lin H (2018) Representing place locales using scene elements. *Computers, Environment and Urban Systems* 71: 153–164. CrossRef

Zhang F, Zhou B, Liu L, Liu Y, Fung HH, Lin H, Ratti C (2018) Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning* 180: 148–160. CrossRef

Zhou B, Lapedriza A, Khosla A, Oliva A, Torralba A (2017) Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 40: 1452–1464. CrossRef

Zhou B, Zhao H, Puig X, Fidler S, Barriuso A, Torralba A (2017) Scene parsing through ADE20K dataset. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 633–641

© 2026 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Louvain-la-Neuve, Belgium. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

REGION
The Journal of ERSA
Powered by WU

Volume 13, Number 1, 2026, 61–94
DOI: 10.18335/region.v13i1.613

journal homepage: region.ersa.org
ISSN: 2409-5370

The General Urban Distortion Propagation Law – A Novel Dynamic Mathematical Framework for Modeling Shock Propagation and Recovery in Complex Urban Systems

Thaer Ayasreh[1]

[1] Independent Researcher, Jerash, Jordan

Received: 25 June 2025/Accepted: 9 March 2026

Abstract. Contemporary cities are increasingly exposed to cascading shocks that propagate through complex and interconnected urban networks, challenging the explanatory power of traditional risk models. This study introduces the General Urban Distortion Propagation Law (GUDPL), an innovative dynamic mathematical framework that integrates functional connectivity, effective distance, time delays, structural vulnerability, recovery dynamics, and irreducible residual effects. The model provides a unified representation of how urban shocks emerge, spread, and attenuate across interconnected systems, and can, in principle, support the identification of vulnerable urban nodes and resilience-oriented planning. To illustrate the operational behavior of the proposed framework, a proof-of-concept simulation based on a stylized cascading power outage scenario is conducted, highlighting delayed propagation, heterogeneous impacts, and recovery processes. GUDPL is adaptable to different crisis types and compatible with time-dependent and real-time modeling contexts, offering a flexible analytical foundation for urban resilience, sustainability, and crisis management research.

Key words: Urban shock propagation, Spatio-temporal modeling, General Urban Distortion Propagation Law (GUDPL), Intercity functional coupling, Cascading failures, Urban recovery models, Urban system resilience

1 Introduction

The twenty-first century is witnessing an unprecedented urban transformation, with cities becoming dominant centers of human activity and economic growth. However, this increasing concentration of population and resources has made urban areas more vulnerable to a variety of shocks. In recent decades, the challenges facing cities have been exacerbated by recurring crises such as natural disasters, pandemics, economic crises, and infrastructure failures. With cities becoming increasingly interconnected through complex networks of relationships, the effects of these shocks are no longer confined to their locations; they now spread across the network to distant cities, causing cascading failures that can lead to widespread paralysis in urban functions (Tang et al. 2025).

This increasing interconnectedness poses a new challenge in understanding the behavior and transmission of urban shocks, necessitating the development of models capable of representing the spatial and temporal dimensions of shocks to understand complex propagation patterns.

Over the past decades, contemporary literature has produced numerous models that have contributed to our understanding of urban interconnectedness. Gravity models have provided insights into patterns of interaction between cities (Fotheringham, O'Kelly 1989), while complex network theory has elucidated the structural characteristics of urban systems (Batty 2013). Simultaneously, research on resilience has enhanced our understanding of how urban systems respond to and recover from shocks (Meerow et al. 2016). The resilience of urban networks is not only related to physical endurance but is influenced by the level of functional connectivity and the ability to adapt to changes over time (Rahimi et al. 2025). Studies have shown that analyzing the resilience of complex urban networks requires integrating economic, structural, and social links into a multidimensional model (Zhang et al. 2022).

Despite significant progress in modeling urban relationships, most existing models do not capture the complex nature of urban relationships along with the temporal dynamics of shock propagation and lack the integration of recovery mechanisms. They often focus on specific aspects of urban interactions without capturing the multidimensional nature of urban connections. Moreover, most models are static or focus on steady-state conditions, failing to account for the dynamic processes of shock propagation and recovery.

In this context, this study aims to bridge this methodological gap by proposing the General Urban Distortion Propagation Law (GUDPL). GUDPL represents a new, flexible, and dynamic mathematical framework that simulates the transmission of urban distortion across interconnected cities. This model is based on the fundamental hypothesis that the impact of one city's collapse on another depends on the degree of spatial and functional connectivity between them, the magnitude of the gap in urban resilience, and the time required for the shock effect to be transmitted. The model is characterized by its flexibility in adapting to various crises and explaining and predicting the transmission of shocks and recovery across urban networks. It also incorporates essential elements such as structural vulnerability, cascading effects, time delays, and recovery functions, making it a potentially useful analytical framework for understanding urban risks and enhancing preventive planning.

This research includes a presentation of the theoretical foundations upon which GUDPL is based, an explanation of its development methodology, and a presentation of the mathematical formulation of the law and its main components. The research also discusses the potential theoretical and practical significance of the law, identifies current limitations, and suggests directions for future research.

1.1 Theoretical Foundations of Shock and Distortion Propagation in Urban Systems

The General Urban Distortion Propagation Law (GUDPL) is based on a multidisciplinary theoretical framework derived from developments in understanding urban networks, complex systems theory, gravity models, spatial interaction, and urban resilience research. This section aims to review the main theoretical foundations that form the conceptual basis of the law, with reference to relevant previous studies that paved the way for such an integrative model.

1.1.1 Urban Networks: From Hierarchical Structures to Complex Interconnected Systems

Urban network studies have evolved from classical hierarchical models—such as those by Christaller and Lösch, which emphasized the spatial distribution of urban functions—toward more dynamic frameworks focusing on inter-city interactions. The quantitative revolution, led by scholars like Berry (1964) and Pred (1977), expanded urban systems theory, while Wilson (1970) introduced entropy-maximization models that laid a more rigorous mathematical foundation. However, these approaches struggled to capture nonlinear dynamics.

Later, Batty (2005) conceptualized cities as complex adaptive systems, and Bettencourt et al. (2010) demonstrated how urban characteristics scale predictably with size.

In the last decade, big data enabled new empirical insights—Boeing (2021) analyzed global street networks, and Wang et al. (2023) mapped real-world urban interactions using mobile phone data—though such approaches still require integration into explanatory models.

Complex network theory now provides a key framework. The way connections are distributed between cities affects how shocks spread (Barabasi, Pósfai 2016). Concepts like centrality (Borgatti, Brass 2019), cascading failures (Valdez et al. 2020), and percolation thresholds (Dorogovtsev et al. 2008) help explain urban shock propagation. GUDPL builds on this by emphasizing functional coupling (Φ_{ij}) and network topology (Crucitti et al. 2006) to trace pathways of disruption and identify critical urban nodes.

1.1.2 Urban Systems Theory

Urban systems theory views cities as interconnected units engaged in flows of economy, population, services, and information—rather than isolated places. It emphasizes hierarchy (Batty 2005), specialization (Scott, Storper 2015), boundaries (Avni et al. 2022), and emergent properties (Allen 1997). Dynamic models support this theory by simulating how urban variables evolve, aiding GUDPL in analyzing the progression or containment of distortions.

Complex systems theory explains how minor shocks can trigger large-scale failures through cascading effects (Buldyrev et al. 2010). GUDPL integrates this by modeling collapse intensity (E_j), delay (τ_{ij}), and residual distortion ($E_{\min,j}$) to reflect nonlinear spread. Diffusion theory, rooted in geography and sociology (Hägerstrand 1967), also informs this framework. Though typically applied to economics (Zenou et al. 2018), innovation (Arieli et al. 2020), or epidemics (Alrasheed et al. 2024), its principles—like proximity and adoption thresholds—help explain urban shock transmission. GUDPL advances these models by incorporating recovery and structural vulnerability within the urban setting.

1.1.3 Gravitational Models and Urban Shock Transmission

Gravity models have long explained spatial flows by relating interaction strength to city "mass" (e.g., population, GDP) and inversely to distance (Wilson 1970, Fotheringham, O'Kelly 1989). Early contributions by Reilly (1931), Huff (1963), and Wilson (1970) introduced entropy-based refinements. Recent advances, supported by computing and data availability, include multidimensional distances (Tubadji, Rudkin 2025), mobile-data validation (Lenormand et al. 2016), and network effects (Wang et al. 2019).

Yet, these models remain limited in modeling urban shock propagation. They often assume symmetric, static interactions, which contradict the asymmetric and nonlinear nature of crises (Haynes, Fotheringham 1984, Hsu et al. 2021, Yuan et al. 2025). The lack of a temporal dimension further restricts their explanatory power (Fan et al. 2020).

Efforts to address these issues include the radiation model (Simini et al. 2012, Masucci et al. 2013), which removes symmetry and calibration constraints, and dynamic extensions (Yang et al. 2014). However, these remain partial solutions.

GUDPL builds on core gravity concepts—distance decay, attraction, and constraints (Taylor 2022)—but advances them through *effective distance* (d_{ij}), multidimensional decay (k), and integrated temporal-recovery functions, offering a more adaptive, crisis-sensitive framework.

1.1.4 Urban Resilience: From Resistance to Adaptive Transformation

Urban resilience research has evolved to address how cities resist, absorb, and recover from shocks. Definitions and frameworks now emphasize its multidimensional nature—physical, social, economic, and institutional (Meerow et al. 2016, Sharifi, Yamagata 2017). Empirical studies explored responses to terrorism (Coaffee, Lee 2016) and climate disasters (Lv et al. 2024), while quantitative tools such as composite indicators (Cutter et al. 2010), network metrics (Linkov et al. 2022), and integrated models (Markolf et al. 2018, Peiris 2024) have aimed to assess resilience at various scales.

Despite progress, most models lack a dynamic, network-based framework to simulate shock transmission and recovery over time. GUDPL addresses this by integrating key resilience parameters: recovery function $R_j(t-\tau_{ij})$, rate λ_j, and residual impact $E_{\min,j}$. It also incorporates structural vulnerability V_j in calculating collapse intensity E_j, reflecting each city's sensitivity to shocks based on social, environmental, and institutional factors—bridging the gap between resilience assessment and real-time urban shock modeling.

Despite extensive literature, most models remain limited—narrow in focus, static in nature, and lacking in realism or adaptability. They often fail to capture temporal dynamics, cascading effects, or urban variability. GUDPL addresses these gaps by offering a multidimensional, dynamic, and interactive framework that integrates various forms of interconnectedness, models nonlinear propagation and recovery, and adapts to diverse urban contexts. It thus contributes a robust and flexible tool for understanding shock transmission in complex urban systems.

2 Research Materials and Methods

The General Urban Distortion Propagation Law (GUDPL) was developed as a mathematical framework to model how shocks spread across interlinked urban systems. The methodology followed five main stages:

Identifying the Gap : Existing models lacked multidimensional connectivity, temporal dynamics, and asymmetry in representing urban shocks—prompting the need for a more comprehensive framework.

Conceptual Foundation : Core ideas such as functional coupling, effective distance, collapse intensity, recovery rate, and time delay were drawn from urban network theory, spatial interaction, complexity, and resilience literature.

Mathematical Translation : Key variables were formalized—for example, D_i for distortion, E_j for collapse intensity, Φ_{ij} for coupling strength, and $R_j(t - \tau_{ij})$ for recovery. Each component was designed to be flexible, allowing various functional forms based on context and data.

Equation Assembly : Components were integrated into a central equation modeling how distortion in each city evolves over time under intercity influences.

Assumptions and Limitations : Assumptions (e.g., measurability, data availability) and potential challenges (e.g., calibration, empirical validation) were clearly acknowledged.

GUDPL thus merges theoretical rigor with practical flexibility, offering a customizable foundation for analyzing diverse urban shock scenarios and advancing future research.

2.1 General Urban Distortion Propagation Law (GUDPL): Mathematical Model Formulation and Component Interpretation

2.1.1 Core Concept and Scope

The General Urban Distortion Propagation Law (GUDPL) aims to provide a dynamic mathematical framework for estimating the degree of distortion (D_i) in a specific city (i) at a given time (t), resulting from the effects of collapse or shock (E_j) originating from other cities (j) within an interconnected urban network. The law takes into account the strength of functional coupling between cities, effective distance, time delay, cities' recovery capacity, and underlying structural vulnerability.

2.1.2 Fundamental Assumptions

The model is built on three fundamental assumptions:

1. Cities are not independent entities but are interconnected through multiple functional relationships.

2. The impact of shocks decreases with increasing functional distance between cities.
3. Shock propagation follows patterns similar to physical attenuation laws, with non-linear dampening across distance.

2.1.3 Structural Flexibility

One of the main strengths of the law is its structural flexibility and adaptability to various types of urban crises. The model does not rely on predetermined fixed parameters but has been designed according to a flexible modeling structure that allows customization of its components based on the nature of the shock. Whether the crisis is caused by earthquakes, climate disasters, pandemics, infrastructure failures, or economic collapse, the internal variables of the model can be recalibrated based on specific indicators and scenarios.

2.1.4 Basic Equation of the GUDPL Model

Based on the previous assumptions, the static model of GUDPL has been formulated as follows:

$$D_i = \sum_{j=1}^{n} \left(\frac{E_j \Phi_{ij}}{d_{ij}^k} \right) + \varepsilon_i \tag{1}$$

2.1.5 Temporal model

In light of the limitations of the static model, a temporal model was constructed that takes into account the evolution of collapse intensity over time. The mathematical formulation of the dynamic temporal model is as follows:

$$D_i(t) = \sum_{j=1}^{n} \left(\frac{E_j(t)\Phi_{ij}(t)}{[d_{ij}(t)]^{k(t)}} \right) + \varepsilon_i(t) \tag{2}$$

where:

$D_i(t)$: The total degree of urban distortion in the city (i) at time t.

$E_j(t)$: The time-varying collapse intensity or shock severity in source city j at time t. It reflects the evolving magnitude of disruption in the source city and may be estimated through normalized indicators related to infrastructure failure, economic decline, service disruption, health-system stress, or other crisis-specific dimensions.

$\Phi_{ij}(t)$: The functional coupling factor between the city i and j at time t.

$d_{ij}(t)$: The effective distance between the city i and j at time t.

$k(t)$: The distance attenuation factor, a positive exponent that reflects how quickly the shock effect dissipates with increasing effective distance.

$\varepsilon_i(t)$: The random factor or 'chance' or 'external fluctuations in the model at time t'.

The basic components of the model

1. Degree of urban distortion ($D_i\,(t)$):

   A composite indicator reflecting the degree of disruption experienced by city i due to external shocks. Its composition can be adapted to the type of shock being studied.

2. Vulnerability-adjusted collapse severity (E_j)

   A component that can be customized according to the nature of the shock. It is calculated using a weighted average of a set of sub-indicators, as follows:

$$E_j = V_j \sum_{m=1}^{\mathrm{ME}} w_m I_{jm}$$

Here, V_j acts as an amplifier of collapse severity; higher structural vulnerability increases the realized collapse intensity for a given shock magnitude.

Where:

- E_j : The severity of the total collapse of the city j. It expresses the extent of the city's collapse in terms of the nature of the crisis, and the closer E_j is to 1, the greater the intensity of the collapse (or the opposite according to the definition).
- I_{jm} : The sub-indicator m that is selected and normalized according to the nature of the crisis; for example, collapse indicators suitable for an earthquake or flood (infrastructure + public services + health services), and in the case of a specialized crisis such as a power outage (energy availability + network efficiency), and economic collapse (GDP, unemployment, market activity, supply chains). The indicators I_{jm} are normalized according to the Min-Max Normalization Linear Normalization formula or by converting them to the range [0, 1] to avoid measurement bias.
- w_m : The weight of indicator m adjusted according to the importance of each indicator in the context of the crisis; for example, in the case of earthquakes, a higher weight may be given to infrastructure compared to the economy).
- M : The number of sub-indicators used in the assessment should preferably be ≥ 3 to cover multiple dimensions (economic, social, and basic functional).
- V_j : The structural vulnerability of city j is calculated as a weighted average of m normalized sub-indicators $V_{j\rho}$, reflecting the city's susceptibility to long-term or cascading distortions. These indicators cover domains such as infrastructure (e.g., decay rate), economy (e.g., sectoral dependency), governance (e.g., corruption levels), society (e.g., social cohesion), and health (e.g., sustainable coverage). Each dimension is assigned a weight (α_ρ) based on its relative importance. All indicators are scaled to [0, 1] and aggregated to form V_j, as follows:

$$V_j = \sum_{\rho=1}^{M_V} \alpha_\rho V_{j\rho}$$

3. Functional Coupling Factor (Φ_{ij})

Reflects the structural and functional connectivity between two cities. The basic linear formula:

$$\begin{aligned}\Phi_{ij} &= \omega_1 L_{ij}^{(1)} + \omega_2 L_{ij}^{(2)} + \ldots + \omega_R L_{ij}^{(R)} \\ &= \sum_{r=1}^{R} \omega_r L_{ij}^{(r)}\end{aligned}$$

where:

- Φ_{ij} : The total functional coupling index between city i and city j, taking a value in the range [0, 1], where a higher value indicates a stronger functional linkage.

L_{ij}^{r} : The degree of linkage between cities i and j for a given linkage type r is represented as a normalized value between 0 and 1, reflecting the strength of interdependence. Linkage types are selected based on the nature of the crisis, as different shocks influence networks in different ways. For instance, in a national power grid failure, relevant linkages may include megawatts of electricity exchanged (electrical linkage), number of trucks or supply lines (logistical linkage), cross-city use of services such as hospitals (service linkage), proportion of j's industrial electricity sourced from i (economic linkage). This modular structure allows GUDPL to adapt the definition and weight of linkages based on shock-specific dynamics.

ω_r : Each linkage type r is assigned a weight ω_r, representing its relative importance in the context of the specific crisis (Weighted Importance of Linkage Types). These weights are calibrated so that their total across all R types satisfies the condition:

$$\sum_{r=1}^{R} \omega_r = 1, \qquad \omega_r \geq 0$$

This weighting ensures that the composite functional coupling reflects not just the presence of connections, but their contextual relevance to the type of shock being analyzed.

R : The number of linkage dimensions used in constructing the index varies by crisis type. These may include economic, service, health, social, political, security, logistical, or electrical connections, among others—selected based on their relevance to the specific shock scenario.

The extended form (with nonlinear interactions)

To enhance the accuracy of the model, synergistic or competitive effects between types of connectivity can be integrated:

$$\begin{aligned} \Phi_{ij} &= \left(\omega_1 L_{ij}^{(1)} + \ldots + \omega_R L_{ij}^{(R)}\right) + \left(\lambda_{12} L_{ij}^{(1)} L_{ij}^{(2)} + \ldots + \lambda_{(R-1)R} L_{ij}^{(R-1)} L_{ij}^{(R)}\right) \\ &= \sum_{r=1}^{R} \omega_r L_{ij}^{(r)} + \sum_{\substack{r,s=1 \\ r \neq 1}}^{R} \lambda_{rs} L_{ij}^{(r)} L_{ij}^{(s)} \end{aligned}$$

where:

λ_{rs} : The coefficient λ_{rs} captures the interaction between linkage types r and s. A positive value ($\lambda_{rs} > 0$) indicates a reinforcing relationship – where one linkage amplifies the effect of the other – while a negative value ($\lambda_{rs} < 0$) indicates a competitive or inhibitory relationship. For example:

- In pandemics, increased social interaction can intensify health system strain.
- In security crises, political conflict may weaken intercity security coordination.
- In displacement scenarios, mass movement may overwhelm administrative capacities, reducing effective response across the network.

Effective Distance (d_{ij}): Concept and Formula

d_{ij} represents the effective distance between cities i and j, extending beyond mere geographic separation. It quantifies the level of *resistance* or *functional separation* between the two cities and can incorporate spatial, temporal, structural, and behavioral factors depending on the crisis context. For instance, high travel time, poor connectivity, or administrative barriers may increase d_{ij} even if the geographic distance is small. A flexible general formula for effective distance is:

$$d_{ij}^k = \left(\sum_{q=1}^{Md} \Phi_q X_{ij}^q \right)^k$$

where:

$X_{ij}^{(q)}$: Effective distance d_{ij} is constructed from a set of M sub-dimensions, each representing a specific form of separation between cities i and j, normalized to a [0, 1] scale. Relevant dimensions vary by crisis and may include (1) Geographic distance (e.g., kilometers), (2) Temporal distance (e.g., travel time), (3) Institutional distance (e.g., administrative divergence), (4) Structural distance (e.g., infrastructure or resource disparities), (5) Behavioral distance (e.g., social or cultural variation), (6) Economic distance, among others.

Φ_q : represents the relative weight assigned to each distance dimension m, indicating its significance in the model based on the nature of the crisis. The weights are normalized to satisfy the condition:

$$\sum_{q=1}^{Md} \Phi_q = 1, \ \ \Phi_q \geq 0$$

Md : Number of dimensions of distance. It is determined by the complexity of the model (usually between 3 to 5 dimensions).

In cases where distance dimensions are overlapping, strongly interdependent, or exhibit nonlinear interaction effects, an alternative formulation can be adopted:

$$d_{ij} = \sqrt{\sum_{q=1}^{Md} \Phi_q \left(X_{ij}^{(q)} \right)^2}$$

which captures overlapping or reinforcing distance effects. In the illustrative simulation presented in this study, distance dimensions are treated as independent, and the additive composite formulation is therefore employed.

This formulation corresponds to a weighted Euclidean distance and is included to highlight the structural flexibility of the GUDPL framework rather than to impose a mandatory modeling choice.

4. Attenuation Factor k

This factor determines how quickly the effect of the shock decreases with increasing effective distance:

$k = 1$: The effect decreases linearly with distance

$k > 1$: The effect decreases rapidly with distance (suitable for local crises)

$k < 1$: The effect decreases slowly with distance (suitable for global crises)

5. The Stochastic Error Term ε_i

   The term ε_i captures random fluctuations or unexplained external factors not accounted for by the core components of the model. These may include (1) Sudden political events (e.g., coups, border closures), (2) Additional natural disasters (e.g., local floods), (3) Behavioral anomalies or data distortion (e.g., social panic), (4) Measurement errors or reporting inconsistencies.

   In experimental modeling, ε_i is typically assumed to follow a normal distribution: $\varepsilon_i \sim N(0, \sigma^2)$. In simulation, a fluctuation range (e.g., $\pm 5\%$) or scenario-based assumptions may apply. In regression, it represents the residual error.

2.1.6 Time-Delayed Dynamic Model

While dynamic time models offer a more realistic view of collapse evolution, they often assume instantaneous shock transmission between cities – an oversimplification of real urban dynamics. In reality, transmission times vary based on functional distance, connectivity levels, and the type of shock. To address this, the temporal model integrates a delay factor (τ_{ij}), representing the time it takes for a shock to move from city j to city i. This spatial-temporal integration enhances the accuracy of simulating distortion spread across interconnected urban systems. The time-delayed model is expressed as:

$$D_i(t) = \sum_{j=1}^{n} \left(\frac{E_j(t - \tau_{ij}) * \Phi_{ij}(t - \tau_{ij})}{[d_{ij}(t)]^{k_i(t)}} \right) + \varepsilon_i(t) \tag{3}$$

where:

$E_j(t - \tau_{ij})$: The intensity of the collapse or the effective shock in city j at time $(t - \tau_{ij})$, taking into account the time delay for the impact to reach city i.

τ_{ij} : The time delay between the occurrence or development of the shock in city j and the arrival of its impact to city i. It depends on the nature of the crisis, the type of pathway through which the impact travels, and the effective distance.

τ_{ij} can be calculated based on the type of crisis. The most common, for example, are: in the case of a direct earthquake, $\tau_{ij}(t)$ is estimated based on the time it takes for emergency teams or shock waves to arrive (minutes – hours). In the case of a power outage, it is estimated by the time the electricity is cut off in city i after being cut off from j. In the case of an economic crisis, it is estimated by the time the market i is affected after a crisis in j (days – months). In the case of population displacement, it is estimated by the time city i starts receiving refugees or displaced persons from j (days – weeks). In addition, $\tau_{ij}(t)$ can be calculated based on historical data, if data from previous crises exists, determine the moment of collapse in city j, then moment m of the impact on city i, and then calculate the difference. The $\tau_{ij}(t)$ can also be calculated by dividing the effective distance between the two cities by the speed of wave propagation according to its type (for example: the speed of electrical outages, the speed of disease transmission, etc.).

2.1.7 The complete form of the GUDPL model with the recovery function

In real urban systems, shock impacts unfold over time rather than occurring in a single instant. Their duration and decay depend on system fragility, interconnectivity, and recovery capacity. To capture this, GUDPL incorporates an extended temporal model that simulates how a shock spreads after a time delay (τ_{ij}) and then gradually diminishes in the affected city through a recovery function—exponential, logistic, or linear—guided by a city-specific recovery rate (λ_j).

The model also includes a fixed minimum residual impact ($E_{\min,j}$), reflecting the irreversible structural damage. This formulation allows GUDPL to simulate not only the spread and peak of urban distortion, but also long-term stabilization, deterioration,

Table 1: The Structural Interpretation of $E_{\min,j}$

$E_{\min,j}$	Structural Interpretation
0	The city does not suffer from chronic structural weakness — its complete recovery after the shock is possible
0.2-0.4	Moderate structural weakness: some structures or sectors do not fully recover.
0.5-0.7	Significant or chronic weakness: the city is structurally fragile even in normal conditions.
0.8-1	The city is highly exposed, threatened by continuous collapse even after the shock has ended.

or partial recovery—enabling policy testing and comparative analysis of urban resilience based on varying recovery dynamics, as follows:

$$D_i(t) = \left[\sum_{j=1}^{n} \frac{[E_{\min,j} + (E_0(j) - E_{\min,j}) R_j(t - \tau_{ij}(t))] \, \Phi_{ij}(t - \tau_{ij}(t))}{[d_{ij}(t)]^{k_i(t)}} \right] * R_s(t - t_0) + \varepsilon_i(t) \tag{4}$$

where:

$D_i(t)$: The total urban distortion in city i at time t.

$E_0(j)$: The initial collapse intensity in city j.

τ_{ij} : The time delay in shock transmission from j to i.

$\Phi_{ij}(t - \tau_{ij})$: The functional coupling factor at the delayed time.

d_{ij} : The effective or functional distance between the two cities.

$k_i(t)$: The dynamic attenuation coefficient in city i.

$\varepsilon_i(t)$: The error coefficient or random external effects.

$E_{\min,j}$: Represents the irreducible level of distortion in city j, the portion of the shock impact that persists over time, even after recovery. Examples include chronic infrastructure damage or long-term social marginalization. The value ranges between 0 and 1 (see Table 1) and can be estimated through (a) a *composite index* based on structural vulnerability indicators, or *expert judgment*, depending on data availability and context.

$R_j(t - \tau_{ij})$: The standard recovery function for city j after the shock reaches it (post-delay τ_{ij}). It ranges between 1 (at the onset of impact) and 0 (full recovery), modeling how the residual collapse intensity decays over time due to emergency response, external support, and infrastructure restoration. The function reflects the city's dynamic recovery trajectory within the GUDPL framework.

Examples of types of $R_j(t)$ functions:

The recovery function $R_j(t)$ models how collapse intensity in city j decreases over time. Its form depends on the crisis context and data availability. Common types include:

- Exponential function $\longrightarrow R_j(t) = e^{-\lambda_j t}$
- Linear function $\longrightarrow R_j(t) = 1 - \gamma t$
- Logistic function $\longrightarrow R_{j(t)} = \frac{1}{1+ae^{\{-\lambda_j t\}}}$
- Function derived from real data $\longrightarrow R_j(t) = f_{\text{empirical}}(t)$

These formulations allow flexible modeling of diverse urban recovery patterns under different shock scenarios.

λ_j : represents the recovery rate of city j, indicating how quickly the city regains functionality after a shock. Its value depends on factors such as crisis type, institutional capacity, and the city's level of integration within the urban network. A higher λ_j implies faster recovery, while lower values reflect prolonged vulnerability.

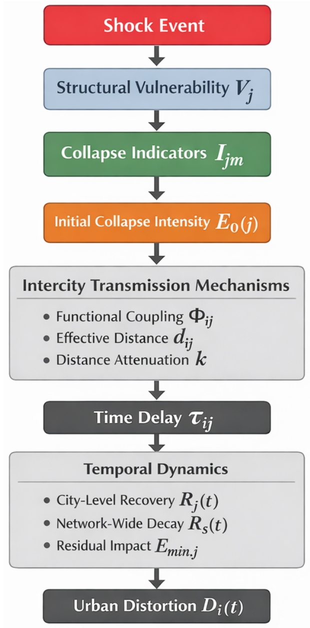


Figure 1: Conceptual flowchart of the General Urban Distortion Propagation Law (GUDPL)

$R_s(t - t_0)$: Represents the overall temporal regression of shock effects across the entire urban network, reflecting how systemic impact fades over time. Like city-level recovery functions, it ranges between 0 and 1 and can follow exponential, logistic, or linear decay. Yet it captures network-wide attenuation, not individual city recovery.

It reflects the "residual shadow" of the shock across the system and can be estimated using *time series* of affected cities, or *aggregated network response* data.

t_0 : The peak moment or the actual beginning of shock propagation in the network, either determines as a constant (for example, the date of the event) or is based on the largest time delay in the network.

Additional considerations and interpretations:

Iterative Computation : $D_i(t)$ is calculated iteratively, where the current distortion in city i depends on past distortions in other cities (including itself), adjusted for time delays. This requires *stepwise temporal simulation.*

Calibration and Validation : Parameter estimation (e.g., Φ_{ij}, V_j, k, τ_{ij}, λ_j, ω_r) is complex and demands *rich data*, advanced modeling techniques, or expert input. Model outputs should be validated against historical shock propagation patterns.

Flexibility : GUDPL allows *adaptive component functions.* Elements like Φ_{ij} or $R_j(t)$ can be tailored – or even made dynamic – to reflect evolving urban conditions or crisis types.

Figure 1 summarizes the sequential modeling stages of the GUDPL framework, from shock definition and city-level structural conditions to intercity transmission mechanisms, temporal dynamics, and the resulting urban distortion. The flowchart is intended to support interpretability of the mathematical formulation and does not replace the formal equations presented in Section 2.

2.2 Parameter Weighting and Calibration Strategy

The GUDPL framework incorporates multiple weighted components, including structural vulnerability weights, collapse severity weights, functional coupling weights, and effective-distance barrier weights. Given the multidimensional nature of the model, weight calibration is treated as a flexible and context-dependent process rather than a fixed universal specification.

At the model-development stage, the framework explicitly distinguishes between theory-informed, scenario-specific weighting and empirically calibrated weighting. In applied or empirical implementations, several established methodological pathways can be used to estimate or calibrate the weights associated with different GUDPL components, depending on data availability, scale, and research objectives.

First, expert-based multi-criteria decision methods, such as the Analytic Hierarchy Process (AHP), can be used to derive weights for vulnerability dimensions, collapse severity indicators, or distance barriers. AHP is particularly suitable when expert knowledge about causal precedence and relative importance is available, as is often the case in infrastructure resilience and disaster risk management. Pairwise comparison matrices can be constructed for each indicator set, and consistency ratios can be used to ensure internal coherence.

Second, data-driven statistical approaches can be employed when sufficient quantitative observations are available. Entropy-based weighting methods allow weights to be inferred from the variability and informational contribution of each indicator, assigning higher weights to dimensions with greater discriminatory power across cities. Similarly, principal component analysis (PCA) or related factor-analytic techniques can be used to reduce dimensionality and derive implicit weights based on explained variance, particularly for vulnerability or coupling indicators that exhibit strong correlations.

Third, in contexts where time-series or large-scale network data are available, machine learning–assisted calibration can be applied. Supervised learning models (e.g., regularized regression, gradient boosting) may be used to estimate weights by minimizing prediction error against observed impact or recovery trajectories from historical shock events. Regularization techniques (e.g., LASSO or ridge penalties) can mitigate multicollinearity and improve parameter identifiability. Unsupervised learning approaches can also support clustering and feature extraction prior to weight assignment.

Importantly, empirical calibration is not required to estimate all parameters simultaneously. In practice, a staged calibration strategy can be adopted, whereby vulnerability weights, coupling weights, distance barriers, and recovery parameters are estimated sequentially using distinct data sources. This staged approach reduces identifiability problems and aligns with best practices in complex systems modeling.

Overall, the GUDPL framework is designed to accommodate multiple calibration strategies while maintaining internal consistency. The choice of weighting method should be guided by the shock type, spatial scale, data availability, and analytical purpose, allowing the model to balance generalizability with contextual specificity.

2.3 Parameter Identifiability and Dependence Considerations

Given the multi-component structure of the GUDPL framework, not all parameters are intended to be estimated simultaneously in empirical applications. The model explicitly acknowledges potential identifiability challenges and high correlation among parameters, particularly between functional coupling Φ_{ij}, effective distance d_{ij}, vulnerability V_j, and recovery rates λ_j. Estimating all parameters jointly from a single dataset may lead to multicollinearity and unstable parameter estimates.

To address this issue, GUDPL is designed to support a staged and modular calibration strategy, whereby parameters are estimated sequentially or conditionally rather than simultaneously. Structural vulnerability parameters (V_j) and collapse severity components (E_j) can be calibrated independently using static or cross-sectional data prior to dynamic simulation. Functional coupling (Φ_{ij}) and effective distance components (d_{ij}) can then be inferred using network-based or flow data, while recovery parameters (λ_j) are estimated separately from post-shock time-series observations.

This staged approach reduces parameter interdependence and improves identifiability by leveraging distinct data sources and temporal windows for different parameter groups. Where residual correlation persists, regularization techniques (e.g., ridge or

LASSO penalties) or dimensionality-reduction methods (e.g., principal component analysis) can be applied to stabilize estimation without altering the structural form of the model.

Importantly, the proof-of-concept simulation presented in this study does not rely on empirical parameter estimation and therefore does not suffer from identifiability constraints. Instead, it serves to demonstrate the internal logic and dynamic behavior of the model under controlled, theory-consistent parameter settings. Full empirical identification and sensitivity analysis are explicitly deferred to future applied studies.

3 Research Results

The General Urban Distortion Propagation Law (GUDPL) offers an integrated mathematical framework for understanding how urban shocks propagate through complex, interconnected city networks. The results reported here derive from analytical inspection of the model structure and from a transparent proof-of-concept simulation designed to illustrate dynamic behavior under stylized conditions. These results demonstrate the model's capacity to overcome the limitations of classical, one-dimensional approaches by incorporating three essential dimensions:

1. Spatial dimension, represented by effective intercity distance.
2. Functional dimension, captured by intercity coupling across economic, social, and service layers (Φ_{ij}).
3. Temporal dimension, introduced through time-delay effects (τ_{ij}) and variable recovery functions (λ_j).

Furthermore, the model incorporates city-specific structural fragility V_j, which modulates the overall shock intensity based on latent vulnerabilities such as poor infrastructure, economic monocultures, or weak institutional capacities. This factor plays a decisive role in explaining why certain cities experience amplified distortions despite being geographically distant or less connected functionally.

The output D_i is a unit-less composite index that quantifies the cumulative distortion experienced by city i, resulting from shocks propagated across all functionally or spatially connected nodes. It is computed as a weighted aggregation of source intensities E_j, modulated by functional coupling Φ_{ij}, effective distance d_{ij}, and temporal dynamics – including propagation delays τ_{ij} and recovery rates λ_j – along with a stochastic term ε_i. Because its components span different scales and units, the analytical formulation of $D_i(t)$ may include stochastic disturbances. In empirical implementations, however, $D_i(t)$ can be reported as a nonnegative operational index (e.g., by truncation at zero or through min–max normalization). This reporting choice is applied solely at the presentation stage and ensures interpretability, comparability across cases, and consistency with the conceptual meaning of urban distortion. In practical interpretation, values of $D_i \approx 0$ indicate negligible or no impact, values around 1 typically reflect significant functional or structural disturbance, while values exceeding 2 (depending on empirical calibration) suggest the onset of cascading failure—particularly in structurally fragile or centrally positioned cities.

The illustrative simulation reveals that shock propagation is neither homogeneous nor linear. Instead, it follows complex trajectories shaped by the strength of functional ties, localized structural vulnerabilities, and temporal lags. Cities with high centrality within the network often emerge as critical nodes—acting either as conduits or buffers—depending on their resilience capacity.

The simulation illustrates the existence of regime-like transitions, where marginal increases in local shock intensity produce disproportionately larger systemic effects – suggesting the presence of critical thresholds, where small local shocks can escalate into widespread systemic disturbances via non-linear amplification mechanisms. This behavior aligns with principles from complexity theory and highlights the importance of early warning indicators.

Another key innovation lies in the dual-level recovery structure:

- Local recovery $R_j(t)$ based on city-specific characteristics.
- Network-wide recovery $R_s(t)$, which reflects the broader system's return to stability.

Importantly, GUDPL recognizes the existence of residual distortion ($E_{\min,j}$), indicating that some structural damage may persist even after recovery. This insight challenges the notion of full reversibility and underscores the need for adaptive planning strategies.

Finally, the model's modular architecture enables calibration and customization based on the nature of the crisis. Parameters such as ω_r, Φ, k, and λ_{rs}) can be adjusted using historical data or machine learning, allowing for broad applicability across different urban disruption scenarios.

3.1 Proof-of-Concept Simulation: Cascading Power Outage Scenario

To illustrate the dynamic behavior of the General Urban Distortion Propagation Law (GUDPL), a proof-of-concept simulation is conducted using a stylized cascading power outage scenario. This simulation is designed to demonstrate how urban distortion propagates through an interconnected urban network when transmission delays, functional coupling, distance-based attenuation, and recovery dynamics are jointly considered.

In this proof-of-concept, all model weights are intentionally specified using a theory-informed, scenario-specific approach. The purpose is not empirical estimation or predictive validation, but to demonstrate the internal coherence, dynamic behavior, and operational feasibility of the GUDPL framework under a stylized yet plausible shock scenario. Accordingly, the selected weights reflect established causal hierarchies documented in the urban resilience and cascading failure literature and are treated as illustrative rather than empirically optimized parameters. Future empirical applications may recalibrate these weights using the methods outlined in Section 2, without requiring structural modification of the underlying equations. This proof-of-concept should therefore be interpreted as a behavioral demonstration of the model, not as an empirical test of parameter values.

The simulated network consists of five cities (A–E) connected through directed functional linkages reflecting electricity exchange capacity, industrial dependence, and operational interconnection. City A serves as the primary shock origin, representing a major power system failure that initiates a cascading blackout. The simulation is conducted over a discrete time horizon of eight time steps, interpreted as hourly intervals.

Urban distortion is computed using the time-delayed GUDPL formulation with recovery, as defined in Section 2. Transmission delays explicitly govern the timing of shock arrival across cities, while distance-based attenuation and functional coupling regulate the magnitude of propagated impacts. City-level recovery processes gradually reduce collapse intensity over time, and a network-wide attenuation term captures system-level stabilization effects. For clarity of interpretation, stochastic disturbances are excluded in this illustrative run.

The illustrative outcomes of the simulation are presented in Figure 2. The distortion trajectories visually illustrate the characteristic behaviors captured by the GUDPL framework across the affected cities. Distortion emerges only after $t = 2$, reflecting the role of transmission delays. City B, characterized by strong functional coupling and low effective distance to the shock source, experiences a rapid and pronounced distortion peak. In contrast, Cities D and E exhibit cumulative and temporally staggered responses, reaching their peaks at $t = 3$ after aggregating impacts from multiple upstream sources. Over time, distortion levels decline as recovery unfolds, yet remain strictly positive due to the presence of irreducible residual effects ($E_{\min,j}$), capturing the long-tail persistence of structural impacts following the crisis.

The complete numerical operationalization of all model components, including parameter values, weighting rationales, and step-by-step calculations, is provided in the Appendix to ensure transparency and replicability.

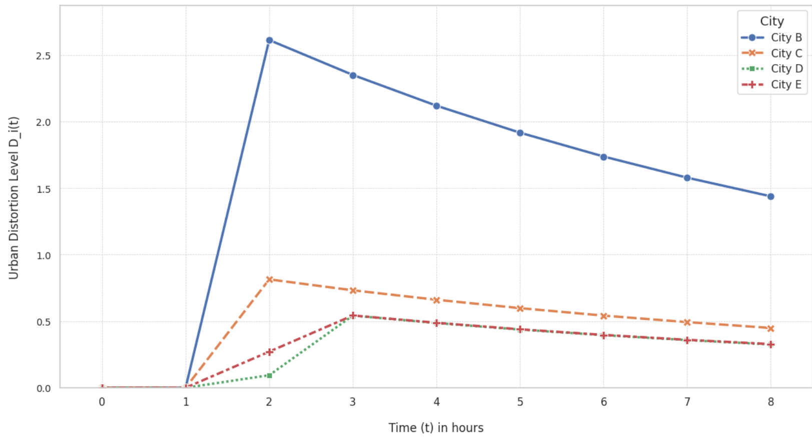


Figure 2: Urban Distortion Trajectories $D_i(t)$ for Affected Cities Following a Cascading Blackout

4 Discussion

The General Urban Distortion Propagation Law (GUDPL) represents an attempt to develop an integrated mathematical framework for understanding the dynamics of shock propagation in complex urban systems. This approach transcends the fundamental limitations of traditional gravity models identified by Haynes, Fotheringham (1984) and Hsu et al. (2021), which assume symmetrical and static interactions between cities. Instead, GUDPL provides a dynamic framework that allows for the representation of asymmetrical and time-varying interactions, aligning with recent trends in spatial interaction modeling.

The model also surpasses the limitations of radiation models developed by Simini et al. (2012) by introducing the concept of "functional coupling" (Φ_{ij}), which goes beyond the notion of a city's fixed "mass" to encompass multidimensional dynamic links. This development addresses the criticism raised by Fan et al. (2020) regarding the absence of a temporal dimension in traditional models, providing a more realistic representation of how urban systems evolve over time in response to disturbances.

GUDPL explicitly incorporates urban resilience concepts developed by Meerow et al. (2016) and Sharifi, Yamagata (2017). While previous studies focused on assessing resilience as a static property of individual cities, GUDPL offers a dynamic understanding of resilience as a process that evolves over time and is influenced by inter-city connectivity. This aligns with the trend identified by Rahimi et al. (2025) regarding the importance of understanding resilience in the context of functional connectivity between urban centers.

The "recovery function" ($R_j(t-\tau_{ij})$) in the model goes beyond simplistic approaches to linear recovery, reflecting the complex nature of urban recovery processes documented in case studies such as Coaffee, Lee (2016) and Lv et al. (2024). Similarly, the concept of "minimum residual distortion" ($E_{\min,j}$) reflects the reality noted in the literature about the impossibility of complete return to the original state after major shocks, acknowledging the permanent transformations that often follow significant disruptions.

GUDPL expands the applications of complex network theory in studying urban systems, moving beyond the static structural analyses presented by Batty (2013) and Bettencourt et al. (2010). By incorporating temporal dynamics and cascading effects, the model provides a framework for understanding how urban networks evolve in response to shocks, addressing an aspect insufficiently covered by previous models. The cascading failure mechanism in the model aligns with the theoretical work of Valdez et al. (2020) and Buldyrev et al. (2010), but adds a specific spatial and temporal dimension that reflects the uniqueness of urban systems.

Despite its theoretical strength, the practical application of the model faces significant challenges related to measuring its key variables. For instance, measuring "functional coupling" (Φ_{ij}) requires comprehensive data on economic, social, and informational flows between cities, which may not be readily available, especially in developing countries (Cutter et al. 2010).

Potential empirical operationalization of the GUDPL framework can draw on established international data sources. Urban structural and socioeconomic indicators relevant to vulnerability and recovery can be obtained from UN-Habitat, the World Bank (e.g., World Development Indicators), and national statistical offices. Data on infrastructure performance, energy systems, and service disruptions may be sourced from sectoral agencies and utility operators, while intercity flows and functional linkages can be approximated using transportation statistics, trade data, or mobility datasets. These sources provide a feasible empirical foundation for future calibration and validation efforts.

The GUDPL framework is inherently compatible with recent advances in computational modeling and network science. Its formulation allows direct integration with network metrics such as centrality, betweenness, and modularity to identify systemically important cities and critical transmission pathways. Functional coupling parameters (Φ_{ij}) can be informed by network-based measures derived from mobility, infrastructure, or economic flow data, while effective distance metrics can be refined using weighted or multilayer network representations. Moreover, artificial intelligence and machine learning techniques offer promising avenues for operationalizing GUDPL in applied contexts. Data-driven models can support parameter calibration, early-warning detection, and scenario exploration without altering the underlying mathematical structure of the law. Importantly, such techniques are viewed as complementary tools for estimation and monitoring rather than substitutes for the analytical core of the model. This compatibility positions GUDPL as a flexible bridge between theory-driven urban systems modeling and emerging computational approaches.

Determining accurate values for model parameters, such as the "distance decay coefficient" (k) and "recovery rate" (λ_j), requires intensive empirical studies. This aligns with observations by Markolf et al. (2018) regarding methodological challenges in modeling infrastructure resilience. The model also faces challenges in empirical verification and calibration due to the scarcity of comprehensive data on urban shock propagation, a challenge that aligns with observations by Yang et al. (2014) on the difficulties of modeling the temporal dynamics of crises.

GUDPL can be compared with other integrated frameworks proposed by researchers such as Peiris (2024) and Linkov et al. (2022). While these frameworks focus on specific aspects of resilience (such as infrastructure resilience or flood resilience), GUDPL provides a more general framework that can be applied to multiple types of shocks. However, this generality may come at the expense of precision in specific contexts, highlighting the trade-off between generalizability and contextual specificity in resilience modeling.

The application of the model could be enhanced through the utilization of big data techniques referenced by Boeing (2021) and Wang et al. (2023). For example, social media data and mobile phone data could be used to estimate functional coupling between cities, while satellite data could be used to track physical changes associated with shocks. These data sources offer new opportunities for operationalizing the model's theoretical constructs in real-world settings.

The model could be translated into practical decision support tools, similar to what Sharifi, Yamagata (2017) proposed. These tools could help planners and policymakers assess systemic risks and design resilience strategies that account for the interconnected nature of urban systems. The scope of the model's application could also be expanded to include different levels of spatial analysis, aligning with recent trends in studying multi-level urban systems, as in the work of Scott, Storper (2015).

5 Conclusion

The General Urban Distortion Propagation Law (GUDPL) represents a significant step toward a deeper understanding of shock propagation dynamics in complex urban systems. By developing a mathematical framework that integrates spatial, temporal, and functional dimensions of urban connectivity, this research makes an original contribution to the fields of urban planning, risk management, and resilience studies.

The findings of this research confirm that cities are not isolated entities but nodes in a functionally interconnected network, where shocks travel and amplify through multidimensional links. This understanding transcends traditional models that have focused on the individual resilience of cities, emphasizing instead the importance of understanding resilience as a property of the urban system as a whole. The interconnected nature of urban systems means that disturbances rarely remain localized but propagate through established functional connections, creating complex patterns of impact that cannot be predicted through simple linear models.

This research highlights the importance of the temporal dimension in understanding urban shock dynamics. Time delays in effect transmission, varying recovery rates between cities, and the evolving nature of functional links all contribute to shaping the temporal trajectory of distortion propagation and containment. This temporal dimension is essential for developing proactive strategies to mitigate risks and enhance resilience, moving beyond reactive approaches that often characterize traditional urban planning.

The study emphasizes the importance of recognizing variations in structural vulnerability and recovery capacity among different cities. Cities differ in their economic structures, infrastructure, and social institutions, and consequently in their ability to absorb shocks and recover from them. This variation must be taken into account when designing resilience policies and allocating resources, suggesting a need for context-specific rather than one-size-fits-all approaches to urban resilience.

The illustrative results indicate that shock propagation in urban systems follows a non-linear path, with critical thresholds at which small local shocks can cause widespread systemic effects. This finding underscores the importance of continuous monitoring of early distortion indicators and developing preventive strategies that can identify and address vulnerabilities before they reach critical thresholds. The non-linear nature of urban distortion propagation also suggests that traditional risk assessment methods may underestimate the potential for cascading failures in highly connected systems.

The General Urban Distortion Propagation Law provides a conceptual and analytical framework that can support systemic risk assessment, subject to empirical calibration and validation to assess systemic risks in urban networks and identify cities with "systemic importance" that may serve as gateways for shock propagation. This analysis can guide resilience enhancement efforts toward areas with the greatest impact on overall system stability, optimizing resource allocation in resilience planning. By identifying key nodes and connections within the urban network, planners can prioritize interventions that provide the greatest systemic benefits.

This research provides a framework for developing integrated resilience strategies that take into account the connectivity between cities. Rather than focusing solely on enhancing the resilience of individual cities, the findings emphasize the importance of diversifying functional links and reducing excessive dependence on specific cities. This network-based approach to resilience planning represents a significant shift from traditional urban planning paradigms that often treat cities as independent entities.

The study highlights the need for better coordination between different administrative levels in risk planning and crisis response. The interconnected nature of urban risks requires coordinated responses that transcend traditional administrative boundaries, suggesting new governance models that can address the complex, cross-jurisdictional nature of urban resilience challenges in an increasingly connected world.

The application of the model faces challenges related to the availability of comprehensive and accurate data on functional links between cities, structural vulnerabilities, and recovery rates. There is a need for further research to develop alternative indicators and innovative data collection methods, leveraging big data techniques to operationalize the model's theoretical constructs. These methodological advances could significantly enhance the practical applicability of the GUDPL framework.

Future research could expand the model to include how shocks propagate between different neighborhoods and sectors within a single city, following a multi-level systems approach. There is also a need for further empirical verification of the model through detailed case studies of actual shock events, which could help calibrate the model and improve the accuracy of its predictions. Additionally, future research could explore how the law can be integrated with other models, such as climate change or economic models, to develop a more comprehensive understanding of the multiple risks facing contemporary cities.

From a policy perspective, the GUDPL framework has the potential to support integrated risk assessment and resilience planning by enabling policymakers to identify systemically critical cities, vulnerable transmission pathways, and temporal windows for intervention. By translating complex intercity dependencies into an analytically tractable structure, the model can inform prioritization of infrastructure investment, emergency coordination, and resilience-enhancement strategies at regional and national scales. Moreover, the modular design of GUDPL facilitates future interdisciplinary applications, allowing integration with climate impact models, economic forecasting tools, and data-driven urban analytics platforms. This positions the framework as a flexible bridge between theoretical urban systems modeling and applied decision-support environments.

In a world of increasing connectivity and exposure to diverse shocks, understanding the dynamics of risk propagation across urban systems becomes critically important. The General Urban Distortion Propagation Law provides a theoretical and methodological framework that can contribute to the development of more resilient and sustainable cities and regions, offering a new lens through which to view and address the complex challenges of urban vulnerability in the 21st century.

References

Allen PM (1997) Cities and regions as evolutionary, complex systems. *Geographical systems* 4: 103–130

Alrasheed H, Alballa N, Al-Turaiki I, Almutlaq F, Alabduljabbar R (2024) City transmission networks: Unraveling disease spread dynamics. *ISPRS International Journal of Geo-Information* 13. CrossRef

Arieli I, Babichenko Y, Peretz R, Young HP (2020) The speed of innovation diffusion in social networks. *Econometrica* 88: 569–594. CrossRef

Avni N, Brenner N, Miodownik D, Rosen G (2022) Limited urban citizenship: The case of community councils in East Jerusalem. *Urban Geography* 43: 546–566. CrossRef

Barabasi AL, Pósfai M (2016) *Network Science*. Cambridge University Press, Cambridge, UK

Batty M (2005) *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models and Fractals*. MIT Press, Cambridge, MA

Batty M (2013) *The New Science of Cities*. The MIT Press, Cambridge, MA. CrossRef

Berry BJL (1964) Cities as systems within systems of cities. *Papers in Regional Science* 13: 147–163. CrossRef

Bettencourt LMA, Lobo J, Strumsky D, West GB (2010) Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* 5: 13541. CrossRef

Boeing G (2021) Street network models and indicators for every urban area in the world. *Geographical Analysis* 54: 519–535. CrossRef

Borgatti SP, Brass DJ (2019) Centrality: Concepts and measures. In: Brass DJ, Borgatti SP (eds), *Social Networks at Work*. Routledge, 9–22. CrossRef

Buldyrev T, Parshani R, Paul G, Eugene S, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464. CrossRef

Coaffee J, Lee P (2016) *Urban resilience: Planning for risk, crisis and uncertainty*. Palgrave, London/New York

Crucitti P, Latora V, Porta S (2006) Centrality measures in spatial networks of urban streets. *Physical review. E*. CrossRef

Cutter SL, Burton CG, Emrich CT (2010) Disaster resilience indicators for benchmarking baseline conditions. *Journal of homeland security and emergency management* 7. CrossRef

Dorogovtsev SN, Goltsev AV, Mendes JF (2008) Critical phenomena in complex networks. *Reviews of Modern Physics* 80: 1275–1335 10 1103 80 1275. CrossRef

Fan C, Jiang X, Mostafavi A (2020) A network percolation-based contagion model of flood propagation and recession in urban road networks. *Scientific Reports* 10: 13481. CrossRef

Fotheringham AS, O'Kelly ME (1989) *Spatial Interaction Models: Formulations and Applications*. Kluwer Academic

Haynes KE, Fotheringham AS (1984) *Gravity and Spatial Interaction Models*. Sage Publication, Beverly Hills, CA

Hsu C, Fan C, Mostafavi A (2021). Limitations of gravity models in predicting fine-scale spatial-temporal urban mobility networks. CrossRef

Huff DL (1963) A probabilistic analysis of shopping center trade areas. *Land Economics* 39: 81–90. CrossRef

Hägerstrand T (1967) *Innovation Diffusion as a Spatial Process. Postscript and translation by Allan Pred*. The University of Chicago Press, Chicago and London

Lenormand M, Bassolas A, Ramasco JJ (2016) Systematic comparison of trip distribution laws and models. *Journal of Transport Geography* 51: 158–169. CrossRef

Linkov I, Trump BD, Trump J, Pescaroli G, Hynes W, Mavrodieva A, Panda A (2022) Resilience stress testing for critical infrastructure. *International Journal of Disaster Risk Reduction* 82: 103323. CrossRef

Lv Y, Sarker MNI, Firdaus RR (2024) Disaster resilience in climate-vulnerable community context: conceptual analysis. *Ecological Indicators* 158: 111527. CrossRef

Markolf SA, Chester MV, Eisenberg DA, Iwaniec DM, Davidson CI, Zimmerman R, Chang H (2018) Interdependent infrastructure as linked social, ecological, and technological systems (SETSs) to address lock-in and enhance resilience. *Earth's Future* 6: 1638–1659. CrossRef

Masucci AP, Serras J, Johansson A, Batty M (2013) Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 88: 022812. CrossRef

Meerow S, Newell JP, Stults M (2016) Defining urban resilience: A review. *Landscape and Urban Planning* 147: 38–49. CrossRef

Peiris MTOV (2024) Assessment of urban resilience to floods: A spatial planning framework for cities. *Sustainability* 16: 9117. CrossRef

Pred A (1977) *City systems in advanced economies: Past Growth, Present Processes, and Future Development Options*. Wiley, London

Rahimi F, Sadeghi-Niaraki A, Ghodousi M, Choi SM (2025) Spatial-temporal modeling of urban resilience and risk to earthquakes. *Scientific Reports* 15: 8321. CrossRef

Reilly WJ (1931) *The law of retail gravitation*. Knickerbocker Press

Scott AJ, Storper M (2015) The nature of cities: The scope and limits of urban theory. *International journal of urban and regional research* 39: 1–15. CrossRef

Sharifi A, Yamagata Y (2017) Towards an integrated approach to urban resilience assessment. *APN Science Bulletin* 7. CrossRef

Simini F, González MC, Maritan A, Barabási AL (2012) A universal model for mobility and migration patterns. *Nature* 484: 96–100. CrossRef

Tang Y, Piao J, Wang H, Rajib S, Li Y (2025) Predicting cascade failures in interdependent urban infrastructure networks. *arXiv preprint*: 2503.02890. CrossRef

Taylor MO (2022) Spatial interaction modeling. In: Rey SJ, Franklin RS (eds), *Handbook of Spatial Analysis in the Social Sciences*. Edward Elgar Publishing, Chapter 12, 208–222,

Tubadji A, Rudkin S (2025) Cultural gravity and redistribution of growth through migration: Cohesion lessons from spatial econometrics and topological data analysis. *Papers in Regional Science* 104: 100064. CrossRef

Valdez LD, Shekhtman L, Rocca CE, Zhang X, Buldyrev SV, Trunfio PA, Havlin S (2020) Cascading failures in complex networks. *Journal of Complex Networks* 8: 013. CrossRef

Wang H, Zhao P, Yan XY (2023) Quantifying the overall characteristics of urban mobility considering spatial information. *arXiv preprint*: 2308.00975. CrossRef

Wang J, Kong X, Xia F, Sun L (2019) Urban human mobility: Data-driven modeling and prediction. *ACM SIGKDD explorations newsletter* 21: 1–19. CrossRef

Wilson AG (1970) *Entropy in Urban and Regional Modelling*. Pion Limited, Addison-Wesley Publishing, Boston

Yang Y, Herrera C, Eagle N, González MC (2014) Limits of predictability in commuting flows in the absence of data for calibration. *Scientific reports* 4: 5662. CrossRef

Yuan X, Geng H, Li Z (2025) Threshold effects of the interaction between urban development and atmospheric pollution. *Atmosphere* 16: 201. CrossRef

Zenou Y, Hodler R, Raschky P, Amarasinghe A (2018) Spatial diffusion of economic shocks in networks. CEPR Discussion Papers, https://ideas.repec.org/p/cpr/ceprdp/-12854.html

Zhang Y, Yang Y, Wei S, Ma Z, Tian M, Sun M, Nie J (2022) Research on spatial structure and resilience of complex urban network: A case study of Jing-Jin-Ji urban agglomeration. *Frontiers in Environmental Science* 10: 999124. CrossRef

Appendices: Numerical Operationalization of GUDPL Components (Proof-of-Concept)

These appendices provide the full numerical operationalization of the GUDPL components used in the proof-of-concept simulation presented in Section 3.1. All values are synthetic but designed to reflect plausible urban conditions. The purpose is to ensure transparency, replicability, and interpretability of the simulation results.

A Urban Network and Shock Context

Cities: A (shock origin), B, C, D, E

A.1 Shock origin (shock source city)

A.1.1 Shock origin city: City A

City A is the primary source node where the blackout initiates (e.g., major generation failure, substation cascade, or national dispatch collapse). It has the highest initial collapse intensity $E_0(A)$, and its downstream impacts propagate to other cities via directed functional couplings Φ_{ij} and time delays τ_{ij}.

A.2 Shock type

A.2.1 Shock type: Cascading power outage / grid failure (cascading blackout)

The shock propagates through electricity exchange capacity, industrial electricity dependence, and operational interconnection. These dimensions jointly define the functional coupling Φ_{ij}.

A.2.2 Collapse sub-indicators used to build $E_0(j)$:

I_{j1} : load-shedding ratio (power supply loss).
I_{j2} : Disruption of electricity-dependent critical services.
I_{j3} : Degradation in grid operational efficiency/stability.
All indicators are normalized to [0, 1] and aggregated using weights $\omega^{(E)}$.

A.2.3 The rationale for selecting the cascading blackout scenario as an illustrative case for the GUDPL model

The cascading power outage represents an ideal case study to illustrate the capabilities of the GUDPL model, as it embodies the core features the model is designed to capture. First, these shocks exhibit clear asymmetry in propagation, with influence flowing more strongly from cities with high generation capacity to those dependent on imports. Second, time delays τ_{ij} play a critical role, reflecting the operational and physical lags required for the disturbance to spread through the network. Third, the coupling between nodes Φ_{ij} depends on multiple dimensions (such as electricity exchange capacity, industrial dependence, and operational interconnectivity), enabling the representation of complex, nonlinear shock dynamics. Thus, this scenario provides a realistic and robust framework for testing the multidimensional, delayed, and asymmetric propagation mechanism that characterizes the GUDPL model.

A.3 Time horizon

Time horizon : $T = 8$ discrete steps. We simulate up to the final time index $T = 8$, i.e., $t \in \{0, \ldots, 8\}$. The horizon $T = 8$ is sufficient to capture the full propagation – attenuation – recovery cycle in the illustrative network, without loss of generality.
Time unit : *hours* (can be interpreted as hourly steps for blackout dynamics).
Simulation interval : $t = 0, 1, \ldots, T$ with $T = 8$.

A.4 Baseline attenuation (network-wide decay)

Baseline (system-wide) attenuation function :

$$R_s(t) = e^{-\lambda s t}$$

Chosen baseline attenuation rate : $\lambda_s = 0.04$

The value $\lambda_s = 0.04$ is selected for illustrative purposes, representing moderate system-wide stabilization dynamics commonly observed in national-scale blackout recovery scenarios.

Interpretation : This captures system-wide fading of shock intensity over time due to overall stabilization efforts, national-level interventions, and progressive restoration in the network.

A.5 City-level recovery (exponential):

$$R_j(\Delta t) = e^{-\lambda_j \Delta t}, \quad \Delta t \geq 0$$

A.6 Additional core attenuation setting (distance attenuation)

Because the model also includes distance-based attenuation, we state it explicitly: $k = 1.25$.

Interpretation: Shock influence decays sub linearly-to-moderately with effective distance; appropriate for a shock that is not purely local but still attenuates as functional barriers increase.

A.7 Important consistency note

Because the model correctly defined $E_{\min,j}$ as an irreducible residual, it must not exceed the initial collapse intensity in the same city. Therefore, we explicitly enforce $E_{\min,j} \leq E_0(j)$ implemented as $E_{\min,j} = \min\left(E^{\text{raw}}_{\min,j}, E_0(j)\right)$. This is especially relevant for cities with very small initial $E_0(j)$.

B Structural vulnerability V_j (city-level)

The vulnerability weights used in this appendix follow a theory-informed, scenario-specific approach appropriate for proof-of-concept modeling. These weights are not treated as fixed parameters and can be recalibrated in applied or empirical extensions of the GUDPL framework.

B.1 Vulnerability Sub-Indicators

We compute vulnerability from three normalized sub-indicators:

V_{j1} : Grid infrastructure fragility (higher = worse)
V_{j2} : Economic dependence on electricity-intensive sectors (higher = worse)
V_{j3} : Institutional/operational capacity weakness (higher = worse)

B.1.1 Weights: $w^{(V)} = (0.40, 0.35, 0.25)$, $\sum w^{(V)} = 1$

The weighting approach is grounded in urban resilience and critical infrastructure literature. This approach is appropriate at the model-development stage, where the objective is to establish a generalizable mathematical framework rather than perform empirical calibration. The selected weights reflect the relative causal importance of infrastructure, economic structure, and institutional capacity in shock propagation. In applied studies, these weights can be recalibrated using analytic hierarchy processes (AHP), expert elicitation, or data-driven techniques such as entropy or principal component analysis.

B.1.2 Scientific Basis for Selecting Vulnerability Weights $w^{(V)} = (0.40, 0.35, 0.25)$

1. Conceptual rationale (theoretical grounding)

The weights assigned to the vulnerability dimensions were determined based on established principles in urban resilience theory, critical infrastructure studies, and cascading failure literature, which consistently emphasize that structural and physical conditions form the primary channel through which shocks materialize and propagate, followed by economic exposure and then institutional response capacity.

Accordingly, the vulnerability index V_j was decomposed into three dimensions with descending structural precedence:

a. Infrastructure vulnerability
b. Economic structural dependence
c. Institutional and governance capacity

This ordering reflects the causal hierarchy of collapse mechanisms observed in complex urban systems.

2. Justification of each weight component

(a) Infrastructure vulnerability — weight 0.40

Scientific basis:

- Infrastructure represents the *physical substrate* of urban systems.

Table B.1: Vulnerability sub-indicators used to construct the structural vulnerability index V_j

City	V_{j1}	V_{j2}	V_{j3}
A	0.70	0.60	0.65
B	0.40	0.45	0.35
C	0.55	0.70	0.55
D	0.30	0.35	0.30
E	0.50	0.40	0.45

- Failures in power grids, transport networks, water systems, or communications often act as initial triggers and amplifiers of cascading shocks.
- Infrastructure damage is frequently non-linear and partially irreversible, making it the dominant vulnerability dimension.

Support from literature (conceptual):

- Critical infrastructure theory treats physical networks as first-order determinants of systemic collapse.
- Network-based failure models show that infrastructure nodes often have high betweenness and load, magnifying propagation effects.

Interpretation in GUDPL:

- A higher weight (0.40) reflects the fact that even strong institutions or diversified economies cannot function without minimally operational infrastructure.

(b) Economic structural vulnerability — weight 0.35

Scientific basis:

- Economic structure determines how strongly a city is exposed to functional disruptions, especially in crises such as blackouts, supply-chain failures, or financial shocks.
- Cities with high sectoral concentration or electricity-intensive industries experience faster and deeper distortion once infrastructure is disrupted.

Why slightly lower than infrastructure:

- Economic impacts are often mediated by infrastructure availability.
- Economic systems are typically more adaptive than physical systems in the short term (e.g., substitution, inventory buffering).

Interpretation in GUDPL:

- The weight 0.35 reflects a secondary but still major role, capturing how shocks translate from physical disruption into systemic urban distortion.

(c) Institutional / governance vulnerability — weight 0.25

Scientific basis:

- Institutional capacity governs response speed, coordination, and recovery, rather than the immediate physical onset of collapse.
- Governance failures tend to shape duration and recovery trajectories, not necessarily the initial shock magnitude.

Why lower than the other two:

- Strong institutions mitigate rather than prevent physical or economic collapse.
- Their effects are often time-delayed, aligning with recovery functions $R_j(t)$ rather than initial vulnerability.

Interpretation in GUDPL:

- Assigning 0.25 ensures institutional factors are explicitly included, but without overstating their influence at the moment of shock propagation.

3. Methodological classification of the weighting approach

Formally, the chosen weights fall under "Theory-informed normative weighting". This approach is widely accepted when a new model is introduced, empirical calibration is not yet available, or the objective is demonstration of model operability, not parameter estimation.

In the GUDPL framework, weights are not fixed constants, but are scenario-dependent parameters, explicitly designed to be recalibrated.

B.2 Aggregate Vulnerability Computation

$$V_j = \sum_{\rho=1}^{3} \alpha_\rho v_{j\rho}$$

$\alpha_\rho = (0.40, 0.35, 0.25)$, $\sum_{\rho=1}^{3} \alpha_\rho = 1$

- $V_A = 0.40(0.70) + 0.35(0.60) + 0.25(0.65) = 0.6525$
- $V_B = 0.40(0.40) + 0.35(0.45) + 0.25(0.35) = 0.405$
- $V_C = 0.40(0.55) + 0.35(0.70) + 0.25(0.55) = 0.6025$
- $V_D = 0.40(0.30) + 0.35(0.35) + 0.25(0.30) = 0.3175$
- $V_E = 0.40(0.50) + 0.35(0.40) + 0.25(0.45) = 0.4525$

Based on these values, City A can be considered the most vulnerable (0.6525), followed by City C (0.6025), while City D is the least vulnerable (0.3175). This pattern is consistent with the underlying data, as Cities A and C exhibit higher values across the vulnerability sub-indicators.

C Initial Collapse Intensity

C.1 Collapse Intensity Sub-Indicators

For the power outage context, collapse intensity at $t = 0$ is constructed from three normalized sub-indicators:

I_{j1} : Power supply loss ratio (share of load shed)
I_{j2} : Critical service disruption (hospitals/water/pumping)
I_{j3} : Grid efficiency degradation (frequency/voltage instability proxy)

C.1.1 Weights: $w^{(E)} = (0.45, 0.35, 0.20)$, $\sum w^{(E)} = 1$

Weights assigned to collapse severity components $w^{(E)} = (0.45, 0.35, 0.20)$ were determined using a theory-informed, scenario-specific approach. The highest weight was assigned to immediate functional loss, as it represents the primary trigger and strongest driver of cascading effects in power outage scenarios. Critical service disruption was assigned a slightly lower weight, reflecting its role in amplifying and prolonging urban impacts. System efficiency degradation received a lower weight, as it primarily affects stability and recovery rather than the initial propagation of the shock. These weights are illustrative and can be recalibrated in applied studies using expert-based or data-driven methods.

C.1.2 Scientific Basis for Selecting Collapse Severity Weights $w^{(E)} = (0.45, 0.35, 0.20)$

1. *Conceptual distinction (key clarification)*
   It is essential to distinguish between two different concepts in the GUDPL framework:
   - V_j represents the *structural susceptibility* of city j to collapse.
   - E_j represents the *realized collapse severity* resulting from an actual shock.

   Accordingly, the weights $w^{(E)}$ are not derived using the same logic as vulnerability weights $w^{(V)}$, but are instead based on a dynamic, operational rationale linked to the moment of shock occurrence and its propagation through the urban system.

2. *Definition of collapse severity components*
   In the selected shock scenario (cascading power outage), the collapse severity index E_j is constructed from three sub-components:
   - Immediate functional loss (e.g., power supply loss or load-shedding ratio)
   - Critical service disruption (e.g., hospitals, water pumping, and emergency services)
   - System efficiency degradation (e.g., grid instability, frequency or voltage disturbances)

   These components represent successive stages of collapse, rather than dimensions of equal influence.

3. Justification for individual weights

(a) Immediate functional loss — weight 0.45

Scientific basis:

Immediate functional loss represents the primary shock trigger (first-order shock). It is the most influential factor in:

- initiating cascading failure processes,
- determining the speed of shock propagation across the network,
- generating time delays τ_{ij}.

In power outage scenarios, loss of electricity supply is the event that initiates urban distortion and determines whether other systems are affected.

Reason for highest weight (0.45):

- It is the most time-sensitive component,
- and exhibits the highest potential for generating cascading effects.

(b) Critical service disruption — weight 0.35

Scientific basis:

This component reflects the transition from a technical failure to a system-wide urban crisis. While it is often temporally dependent on functional loss, it substantially amplifies and prolongs the shock's impacts.

Reason for slightly lower weight:

- It is causally downstream from immediate functional loss,
- yet more consequential than efficiency degradation in terms of social and humanitarian impact.

(c) System efficiency degradation — weight 0.20

Scientific basis:

System efficiency degradation captures the gradual structural deterioration of system performance. It is typically less visible at the moment of shock but is crucial for understanding the persistence of disruption.

Reason for lowest weight:

- It does not determine initial shock propagation,
- but primarily affects long-term stability and recovery trajectories, which are more directly associated with recovery functions $R_j(t)$.

4. Causal hierarchy underlying the weighting scheme

The ordering $0.45 > 0.35 > 0.20$ reflects a clear causal and temporal sequence:

1. Shock occurrence $\rightarrow$ immediate functional loss
2. Functional loss $\rightarrow$ critical service disruption
3. Prolonged disruption $\rightarrow$ system efficiency degradation

This hierarchy is not arbitrary, but mirrors both the order of occurrence and the order of influence in urban shock dynamics.

5. Methodological classification

The weighting scheme for $w^{(E)}$ falls under "Theory-informed, scenario-specific weighting". This approach is appropriate when the shock type is explicitly defined, the model is at a developmental stage, or the objective is to demonstrate dynamic behavior rather than empirical calibration.

C.2 Collapse Sub-Indicators and Raw Shock Magnitude S_j

We assume the initial shock originates in city A (primary collapse), while other cities have smaller initial disturbances.

Non-origin cities are assigned small baseline disturbances at $t = 0$ to reflect minor local susceptibility in blackout conditions. The dominant exogenous shock source remains city A.

These baseline disturbances do not represent independent shock sources, but latent instabilities that become relevant only through propagated coupling.

Table C.1: Collapse sub-indicators at $t = 0$ used to construct the raw shock magnitude S_j

City	I_{j1}	I_{j2}	I_{j3}
A	0.95	0.85	0.80
B	0.25	0.20	0.15
C	0.40	0.35	0.30
D	0.15	0.10	0.10
E	0.20	0.15	0.12

$$S_j = \sum_{m=1}^{3} \omega_m^{(E)} I_{jm}$$

$\omega_m^{(E)} = (0.45\ ,\ 0.35,\ 0.20), \quad \sum_{m=1}^{3} \omega_m^{(E)} = 1$

- $S_A = 0.45(0.95) + 0.35(0.85) + 0.20(0.80) = 0.885$
- $S_B = 0.45(0.25) + 0.35(0.20) + 0.20(0.15) = 0.2125$
- $S_C = 0.45(0.40) + 0.35(0.35) + 0.20(0.30) = 0.3625$
- $S_D = 0.45(0.15) + 0.35(0.10) + 0.20(0.10) = 0.1225$
- $S_E = 0.45(0.20) + 0.35(0.15) + 0.20(0.12) = 0.1665$

In the simulation, the values Sj are used directly as the initial collapse intensities $E_0(j)$.

Based on these values, The S_j represent the raw, unadjusted magnitude of the initial shock event at $t = 0$. The results clearly establish City A (0.885) as the epicenter of the crisis, with a collapse severity an order of magnitude higher than any other city. The small, non-zero values in other cities (e.g., City C at 0.363) reflect minor, localized instabilities inherent in the network even before the main shockwave propagates.

C.3 Realized Initial Collapse Intensity $E_0(j)$

$$E_0(j) = V_j * S_j$$

- $E_0(A) = 0.6525 * 0.885 = 0.5775$
- $E_0(B) = 0.405 * 0.2125 = 0.0861$
- $E_0(C) = 0.6025 * 0.3625 = 0.2184$
- $E_0(D) = 0.3175 * 0.1225 = 0.0389$
- $E_0(E) = 0.4525 * 0.1665 = 0.0753$

Based on these values, The final $E_0(j)$ represent the realized initial collapse intensity, factoring in both the shock's magnitude (S_j) and the city's inherent vulnerability (V_j). Although City A (0.578) remains the primary source, the high vulnerability of City C amplifies its minor initial disturbance into a more significant realized collapse ($E_0(C) = 0.218$), making it a secondary source of instability. In contrast, the low vulnerability of City D dampens its initial disturbance to a negligible level ($E_0(D) = 0.039$).

D Irreducible Residual Collapse

To reflect irreversible structural damage (e.g., transformer burnout), we set a parsimonious rule:

$E_{\min,j} = \min(\rho E_0(j), 0.30V_j)$, $\rho \in [0.2, 0.6]$. The parameter ρ controls the maximum fraction of the initial collapse that remains irreducible. In the illustrative simulation we set $\rho = 0.4$. Sensitivity can also be explored over $\rho \in [0.2, 0.6]$.

- $E_{\min,A} = 0.1958$
- $E_{\min,B} = 0.0344$
- $E_{\min,C} = 0.0874$
- $E_{\min,D} = 0.0156$
- $E_{\min,E} = 0.0301$

Table E.1: Directed functional coupling values Φ_{ij} between cities

$j \to i$	$L^{(1)}$	$L^{(2)}$	$L^{(3)}$	Φ_{ij}
$A \to B$	0.85	0.60	0.70	0.50(0.85)+0.30(0.60)+0.20(0.70)=0.745
$A \to C$	0.60	0.50	0.55	0.560
$A \to D$	0.35	0.30	0.25	0.315
$A \to E$	0.30	0.40	0.20	0.310
$C \to B$	0.55	0.50	0.45	0.515
$C \to D$	0.45	0.35	0.40	0.410
$C \to E$	0.40	0.30	0.35	0.360
$B \to D$	0.50	0.25	0.40	0.405
$B \to E$	0.35	0.30	0.30	0.325

Based on these values, The $E_{\min,j}$ quantify the permanent, irreversible damage that persists even after recovery efforts. The highest residual impact is assigned to City A (0.196), reflecting severe, lasting damage at the shock's origin. Notably, the value for City C (0.087) is also significant, indicating that its high vulnerability translates into a greater degree of permanent structural harm compared to less vulnerable cities like City D (0.016).

E Functional Coupling Φ_{ij}

For a blackout, functional coupling between j and i is built from three link types (normalized):

$L^{(1)}_{ij}$: Electricity exchange capacity (MW transfer potential, normalized)
$L^{(2)}_{ij}$: Industrial dependence of i on supply from j (normalized)
$L^{(3)}_{ij}$**:** Operational interconnection (shared substations/lines, normalized)

E.1 Weights

$\omega = (0.50, 0.30, 0.20)$, $\sum \omega = 1$

$$\Phi_{ij} = \sum_{r=1}^{3} \omega_r L^{(r)}_{ij}, \qquad \sum_{r=1}^{3} \omega_r = 1$$

In this proof-of-concept run, $\Phi_{j\to i}$ is treated as time-invariant for simplicity.

Note that $\Phi_{ij} \neq \Phi_{ji}$ in general, reflecting directional dependence and asymmetric intercity interactions.

E.2 Example Directed Linkage Values

In this appendix we use the directed notation $j \to i$ to indicate shock propagation from source city j to recipient city i. Accordingly, $j \to \Phi_i$, and $j \to \tau_i$ are defined on directed edges.

We list the edges used in the simulation (others assumed negligible/0 for compactness) in Table E.1.

All non-listed directed edges are set to 0 for compactness, yielding a sparse directed adjacency used in the proof-of-concept run.

Based on these values, The Φ_{ij} matrix quantifies the strength and direction of functional linkages, representing the pathways for shock propagation. The strongest coupling is from the shock origin to its immediate dependents, with $A \to B$ (0.745) being the most critical path. The matrix also reveals secondary propagation routes, such as $C \to B$ (0.515), indicating that City B is susceptible to shocks from multiple upstream sources. The sparse, directed nature of these values creates an asymmetric and realistic propagation network.

Table F.1: Effective distance components and aggregated effective distance d_{ij}

$j \to i$	$X_{ij}^{(1)}$	$X_{ij}^{(2)}$	$X_{ij}^{(3)}$	d_{ij}
$A \to B$	0.30	0.25	0.20	0.2575
$A \to C$	0.55	0.40	0.35	0.4475
$A \to D$	0.75	0.60	0.55	0.6475
$A \to E$	0.70	0.55	0.50	0.5975
$C \to B$	0.35	0.30	0.25	0.3075
$C \to D$	0.60	0.50	0.40	0.5150
$C \to E$	0.55	0.45	0.35	0.4650
$B \to D$	0.50	0.40	0.35	0.4275
$B \to E$	0.45	0.38	0.32	0.393

Note: To prevent unrealistic numerical amplification when some effective distances d_{ij} are very small (close to zero), a stability safeguard may be applied in general applications: $d_{ij}^{*} = \max(d_{ij}, \delta)$, with $\delta = 0.05$. In the present illustrative simulation, all $d_{ij} > \delta$. Therefore, results are unaffected.

F Effective Distance and Distance Attenuation

Effective distance is constructed from three normalized barrier dimensions (higher = harder for impact to travel):

- $X_{ij}^{(1)}$: restoration/response time barrier (normalized)
- $X_{ij}^{(2)}$: network constraint barrier (line redundancy weakness; barrier form)
- $X_{ij}^{(3)}$: institutional coordination barrier (normalized)

Weights: $\theta = (0.40, 0.35, 0.25)$, $\sum \theta = 1$

$$d_{ij} = \sum_{q=1}^{3} \theta_q X_{ij}^{(q)}$$

The distance attenuation exponent k is global for this illustrative run and set to $k = 1.25$. If "redundancy" is measured raw as Red_{ij} (higher = better), we normalize then invert:

$$\text{Red}_{ij}^{\text{norm}} = \frac{\text{Red}_{ij} - \text{Red}_{\min}}{\text{Red}_{\max} - \text{Red}_{\min}}, \quad X_{ij}^{(2)} = 1 - \text{Red}_{ij}^{norm}$$

F.1 Barrier Dimensions

$$d_{ij} = \sum_{q=1}^{3} \theta_q X_{ij}^{(q)} = 0.40 X_{ij}^{(1)} + 0.35 X_{ij}^{(2)} + 0.25 X_{ij}^{(3)}$$

Note: Values of d_{ij}^{k} are computed numerically and rounded to three decimals for readability.

The d_{ij} values represent the functional "friction" or resistance to shock transmission, moving beyond simple geography. The path $A \to D$ (0.648) has the highest effective distance, suggesting significant barriers (e.g., institutional, network constraints) that will slow and weaken the shock's impact. Conversely, the path $A \to B$ (0.258) has the lowest distance, indicating a highly efficient channel for shock propagation.

F.2 Distance Attenuation d_{ij}^{k}

Now compute:

$$d_{ij}^{k} = (d_{ij})^{1.25}$$

The approximate values (sufficient for an illustrative appendix) are shown in Table F.2.

Table F.2: Distance-attenuated effective distances $(d_{ij})^k$ used in shock propagation

$j \to i$	d_{ij}	$(d_{ij})^{1.25}$
$A \to B$	0.2575	0.184
$A \to C$	0.4475	0.367
$A \to D$	0.6475	0.582
$A \to E$	0.5975	0.528
$C \to B$	0.3075	0.229
$C \to D$	0.5150	0.437
$C \to E$	0.4650	0.386
$B \to D$	0.4275	0.344
$B \to E$	0.393	0.311

Table G.1: Time delays τ_{ij} derived from effective distances

$j \to i$	d_{ij}	$\tau_{ij} = \left\lceil \frac{d_{ij}}{\nu} \right\rceil$
$A \to B$	0.2575	$\lceil 1.03 \rceil = 2$
$A \to C$	0.4475	$\lceil 1.79 \rceil = 2$
$A \to D$	0.6475	$\lceil 2.59 \rceil = 3$
$A \to E$	0.5975	$\lceil 2.39 \rceil = 3$
$C \to B$	0.3075	$\lceil 1.23 \rceil = 2$
$C \to D$	0.5150	$\lceil 2.06 \rceil = 3$
$C \to E$	0.4650	$\lceil 1.86 \rceil = 2$
$B \to D$	0.4275	$\lceil 1.71 \rceil = 2$
$B \to E$	0.393	$\lceil 1.57 \rceil = 2$

G Time Delay Specification τ_{ij}

We model delay as proportional to effective distance

$$\tau_{ij} = \left\lceil \frac{d_{ij}}{v} \right\rceil$$

with $v = 0.25$ (distance units per hour-step).

The ceiling operator ensures that propagation delays are not underestimated in discrete-time simulations. The results are shown in Table G.1.

The τ_{ij} values translate effective distances into discrete time steps (hours) required for the shock to travel. Impacts from the origin (A) will take 3 hours to reach the most functionally distant cities (D and E). In contrast, impacts arrive at the closest cities (B and C) in just 2 hours. This staggered timing is critical for understanding the cumulative, multi-wave nature of the distortion as it unfolds across the network.

H Recovery Functions and Network-Wide Decay

City-level recovery (exponential):

$$R_j(\Delta t) = e^{-\lambda_j \Delta t}, \qquad \Delta t \geq 0$$

Choose recovery rates (illustrative) from Table H.1.

Table H.1: City-level recovery rates λ_j and exponential recovery functions

j	λ_j	$R_j(\Delta t)$
A	0.10	(slowest recovery: source city with major damage) $R_A(\Delta t) = e^{-0.10\Delta t}$
B	0.14	$R_B(\Delta t) = e^{-0.14\Delta t}$
C	0.12	$R_C(\Delta t) = e^{-0.12\Delta t}$
D	0.18	$R_D(\Delta t) = e^{-0.18\Delta t}$
E	0.16	$R_E(\Delta t) = e^{-0.16\Delta t}$

The λ_j values represent the intrinsic recovery capacity of each city. The shock's origin, City A (0.10), is assigned the slowest recovery rate, reflecting major structural damage. In contrast, the most resilient city, City D (0.18), is given the fastest recovery rate. This inverse relationship between vulnerability and recovery speed is a core assumption, modeling the principle that more fragile systems struggle more to bounce back from disruptions.

Network-wide temporal decay:

$$R_s(t) = e^{-\lambda_s t}, \quad \lambda_s = 0.04$$

The separation between city-level recovery R_j and network-wide decay R_s allows the model to distinguish local repair capacity from systemic stabilization.

Note on the choice of t_0: In the main model (Eq. 4), the network-wide decay is written as $R_s(t - t_0)$, where t_0 denotes the network shock onset (or peak reference) time. In this simulation, the shock starts at $t = 0$. Therefore $t_0 = 0$ and $R_s(t - t_0) = R_s(t)$.

I Full dynamic GUDPL model (equation 4)

We use the deterministic illustrative setting $\varepsilon_i(t) = 0$ to make computations transparent.

$$D_i(t) = \left[\sum_{j=1}^{n} \frac{(E_{min,j} + (E_0(j) - E_{min,j})\ R_j\,(t - \tau_{ij}))\ \ \Phi_{ij}}{(d_{ij})^k} \right] R_s(t)$$

Arrival condition: if $t - \tau_{ij} < 0$, the term from $j \to i$ is set to 0 (impact has not arrived yet).

Note: In this proof-of-concept simulation, source collapse trajectories $E_j(t) = E_{\text{min},j} + (E_0(j) - E_{\text{min},j})R_j(t)$ are treated as exogenous inputs governed solely by city-level recovery dynamics. Endogenous feedback – where $E_j(t)$ depends on distortions received from other cities – is left for empirical calibration and future extensions of the model.

J Worked Simulation Examples

To demonstrate the operational application of the full GUDPL model (Equation 4), this section provides a detailed, step-by-step computation for two representative target cities: City B and City D. These cities were selected as illustrative cases (worked examples) to showcase how the model handles different scenarios of shock propagation: City B represents a case of rapid, high-intensity impact from a few sources, while City D illustrates a more complex case of delayed, cumulative impact from multiple upstream sources. The final simulation results for all affected cities (B, C, D, and E) are summarized in the concluding table of this appendix.

J.1 Pre-computed Constants and Gains

For convenience define the "propagation gain":

$$G_{ij} = \frac{\Phi_{ij}}{d_{ij}^k}$$

G_{ij} is introduced purely for computational convenience and does not alter the underlying structure of Equation (4). The computations for major edges are shown in Table J.1.

Define effective collapse in source city j at delayed time:

$$E_j^{\text{eff}}(\Delta t) = \ E_{\text{min},j} + \ (\ E_0(j) - \ E_{\text{min},j})\, e^{-\lambda_j \Delta t}$$

The G_{ij} values synthesize the relationship between functional coupling (Φ_{ij}) and distance-based attenuation (d_{ij}^k) into a single "propagation gain" multiplier for each path. The path $A \to B$ exhibits by far the highest gain (4.049), indicating that shocks originating from A are heavily amplified as they travel to B due to strong coupling and

Table J.1: Propagation gains G_{ij} for major directed paths

$j \to i$	Φ_{ij}	d_{ij}^k	$G_{ij} = \frac{\Phi_{ij}}{d_{ij}^k}$
$A \to B$	0.745	0.184	4.049
$A \to C$	0.560	0.367	1.526
$A \to D$	0.315	0.582	0.541
$A \to E$	0.310	0.528	0.587
$C \to B$	0.515	0.229	2.249
$C \to D$	0.410	0.437	0.938
$C \to E$	0.360	0.386	0.933
$B \to D$	0.405	0.344	1.177
$B \to E$	0.325	0.311	1.045

low effective distance. Conversely, the path $A \to D$ has a very low gain (0.541), meaning the impact is significantly dampened along this route. Table J.1 is crucial as it reveals the most critical and impactful propagation channels in the network.

J.2 Example: $D_B(t)$

Using the pre-computed propagation gains $G_{j\to i}$, delays $\tau_{j\to i}$, recovery rates λ_j, and residual terms $E_{\min,j}$, we compute $D_B(t)$ as follows. Contributions are set to zero when $t - \tau_{j\to B} < 0$.

City B receives propagated shock impacts mainly from two upstream cities: $j \in \{A, C\}$. Using the directed notation $j \to i$ (shock propagation from source j to recipient i), the distortion in city B at time t is computed as:

$$D_B(t) = \left[\, G_{A\to B}E_A^{\text{eff}}(t \;-\; \tau_{A\to B}) + \; G_{C\to B}E_C^{\text{eff}}(t \;-\; \tau_{C\to B})\right] \; R_{s(t)}$$

Arrival condition: If $(t - \tau_{j\to B}) < 0$, the contribution from edge $j \to B$ is set to zero (impact has not yet arrived).
Delays: $\tau_{A\to B} = 2$, $\tau_{C\to B} = 2$.
Propagation gains: $G_{A\to B} = 4.049$, $G_{C\to B} = 2.249$.
Effective collapse in source city j at delayed time Δt is defined as:

$$E_j^{\text{eff}}(\Delta t) = E_{\min,j} + \left(E_0^{(j)} - \; E_{\min,j}\right) e^{-\lambda_j \Delta t}, \quad \Delta t \geq 0$$

Network-wide decay: $R_s(t) = \; e^{-\lambda_s t}$, with $\lambda_s = 0.04$.
Deterministic setting: $\varepsilon_B(t) = 0$.

At $t = 0$, 1 (No arrival yet):

$$D_B(0) = D_B(1) = 0$$

At $t = 2$ (arrival moment, $\Delta t = 0$):

$$E_A^{\text{eff}}(0) = E_0(A) = 0.5775$$
$$E_C^{\text{eff}}(0) = E_0(C) = 0.2184$$
$$R_s(2) = e^{(-0.08)} = 0.9231$$
$$D_B(2) = [4.049(0.5775) + 2.249(0.2184)](0.9231) = 2.612$$

At $t = 3$ ($\Delta t = 1$):

$$E_A^{\text{eff}}(1) = 0.1958 + (0.5775 - 0.1958)e^{(-0.10)} = 0.5412$$
$$E_C^{\text{eff}}(1) = 0.0874 + (0.2184 - 0.0874e^{(-0.12)} = 0.2036$$
$$R_s(3) = e^{(-0.12)} = 0.8869$$
$$D_B(3) = [4.049(0.5412) + 2.249(0.2036)](0.8869) = 2.349$$

At $t = 6$ ($\Delta t = 4$):

$$E_A^{\text{eff}}(4) = 0.1958 + 0.3817e^{(-0.40)} = 0.4516$$
$$E_C^{\text{eff}}(4) = 0.0874 + 0.1310e^{(-0.48)} = 0.1685$$
$$R_s(6) = e^{(-0.24)} = 0.7866$$
$$D_B(6) = [4.049(0.4516) + 2.249(0.1685)](0.7866) = 1.737$$

Interpretation: $D_B(t)$ peaks at the arrival time and then declines as recovery progresses, while remaining positive over the simulated horizon due to $E_{\min,j}$, while gradually attenuating under $R_s(t)$.

J.3 Example: $D_D(t)$

City D receives propagated shock impacts from three upstream cities:
$j \in \{A, C, B\}$
Accordingly, the distortion in city D at time t is computed as:

$$D_{D(t)} = \left[G_{A\to D}E_A^{\text{eff}}(t - \tau_{A\to D}) + G_{C\to D}E_C^{\text{eff}}(t - \tau_{C\to D}) + G_{B\to D}E_B^{\text{eff}}(t - \tau_{B\to D})\right] R_{s(t)}$$

At $t = 2$

Only the path $B \to D$ has arrived ($\tau_{B\to D} = 2$), while $A \to D$ and $C \to D$ have not yet arrived ($\tau_{A\to D} = \tau_{C\to D} = 3$).

$E_B^{\text{eff}}(0) = E_0(B) = 0.0861$, $Rs(2) = 0.9231$.

$$D_D(2) = [1.177 \times 0.0861] \times 0.9231 = 0.1013 \times 0.9231 = 0.094$$

At $t = 3$

Paths $A \to D$ and $C \to D$ arrive ($\Delta t = 0$), while Path $B \to D$ has $\Delta t = 1$.

$$E_A^{\text{eff}}(0) = 0.5775$$
$$E_C^{\text{eff}}(0) = 0.2184$$
$$E_B^{\text{eff}}(1) = 0.0344 + (0.0861 - 0.0344)e^{-0.14} = 0.080$$
$$Rs(3) = 0.8869$$
$$D_D(3) = [0.541 \times 0.5775 + 0.938 \times 0.2184 + 1.177 \times 0.080] \times 0.8869 = 0.542$$

At $t = 6$

$A \to D : \Delta t = 3$

$$E_A^{\text{eff}}(3) = 0.1958 + 0.3817e^{-0.3} = 0.4786$$
$$E_C^{\text{eff}}(3) = 0.0874 + 0.1310e^{-0.36} = 0.1788$$
$$E_B^{\text{eff}}(4) = 0.0344 + 0.0526e^{-0.56} = 0.0639$$
$$Rs(6) = 0.7866$$
$$D_D(6) = [0.541 \times 0.4786 + 0.938 \times 0.1788 + 1.177 \times 0.0639] \times 0.7866 = 0.395$$

Interpretation: City D experiences delayed but cumulative distortion driven by multiple upstream sources with heterogeneous arrival times. The distortion peaks shortly after the arrival of major contributors (A and C), then gradually declines due to recovery processes. However, $D_D(t)$ remains strictly positive because the irreducible residual term $E_{\min,j}$ prevents full disappearance of propagated effects, reflecting persistent structural damage within the network.

Table K.1: Final urban distortion trajectories $D_i(t)$ for all affected cities

Time (t)	$D_B(t)$ (Affected by A, C)	$D_C(t)$ (Affected by A)	$D_D(t)$ (Affected by A, C, B)	$D_E(t)$ (Affected by A, C, B)
0	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000
2	**2.612** (Peak)	**0.814** (Peak)	0.094	0.271
3	2.349	0.732	**0.542** (Peak)	**0.543** (Peak)
4	2.118	0.661	0.486	0.488
5	1.916	0.598	0.437	0.439
6	1.737	0.542	0.395	0.397
7	1.578	0.493	0.358	0.360
8	1.438	0.449	0.325	0.328

K Final Simulation Results

The complete, corrected distortion trajectories $D_i(t)$ for all affected cities were computed iteratively for each time step from $t = 0$ to $t = 8$. The table below summarizes these final results, which form the numerical basis for the graphical analysis presented in the main body of the paper. This summary provides a comprehensive overview of the heterogeneous propagation dynamics across the urban network.

The final simulation results reveal clear and heterogeneous urban distortion dynamics shaped by the combined effects of functional coupling strength, effective distance, and propagation delays. City B exhibits the highest distortion levels across the entire time horizon, reaching its maximum at t=2 immediately upon shock arrival. This behavior reflects the exceptionally strong and short-distance coupling from the primary shock origin ($A \to B$), resulting in rapid and amplified impact transmission. In contrast, City C experiences a moderate and relatively simple response pattern, with a single early peak at $t = 2$ driven solely by its direct linkage to the origin city A, followed by a smooth monotonic decline governed by recovery dynamics.

Cities D and E display delayed but cumulative distortion profiles. Both cities register only minor impacts at t=2, corresponding to early arrivals from secondary upstream sources, and reach their respective peaks at $t = 3$ when the dominant contributions from cities A and C arrive simultaneously. This temporal superposition of multiple delayed sources generates a pronounced cumulative effect, clearly illustrating the model's ability to capture multi-source, time-staggered shock propagation. After their peak, distortion levels in both cities gradually decrease as recovery mechanisms take effect. However, distortions remain strictly positive over the simulated horizon due to the presence of irreducible residual components, which represent persistent structural damage within the network. Overall, the proof-of-concept results illustrate how GUDPL can represent immediate amplification, delayed accumulation, and recovery-driven attenuation under stylized assumptions.

L Optional Stochastic Term

If included, $\varepsilon_i(t) \sim N(0, \sigma^2)$ or a bounded noise scenario (e.g., $\pm 5\%$ of the deterministic $D_i(t)$). For transparency, all illustrative computations above assume $\varepsilon_i(t) = 0$.

M Implementation Pseudo-Algorithm

Input: City set $\mathcal{C}$, time horizon T, weights $w(V), w(E)$, parameters ω, θ, ρ, exponent k, delays τ_{ij}, recovery rates λ_j, system decay λ_s
Output: Distortion levels $D_i(t)$ for all cities i over time

```
// Parameter domain
```
$\rho \in [0.2, 0.6]$;
Step 1: Compute vulnerability
foreach *city* $j \in \mathcal{C}$ **do**
  Compute V_j from sub-indicators V_{jm};
end
Step 2: Compute shock intensity
foreach *city* $j \in \mathcal{C}$ **do**
  Compute S_j from indicators I_{jm};
  Compute $E_0(j) = V_j \cdot S_j$;
end
Step 3: Compute minimum shock level
foreach *city* $j \in \mathcal{C}$ **do**
  $E_{\min,j} = \min\left(\rho \cdot E_0(j),\, 0.30 \cdot V_j\right)$;
end
Step 4: Compute coupling strength
foreach *directed pair* (j, i) **do**
  Compute Φ_{ij} from network connectivity;
end
Step 5: Compute effective distance
foreach *directed pair* (j, i) **do**
  Compute d_{ij}^k;
end
Step 6: Temporal propagation
for $t = 0$ **to** T **do**
  Compute $R_s(t) = e^{-\lambda_s t}$;
  foreach *target city* $i \in \mathcal{C}$ **do**
    Initialize *total* $\leftarrow 0$;
    foreach *incoming city* j *with* $\Phi_{ij} > 0$ **do**
      if $t < \tau_{ij}$ **then**
        contribution $\leftarrow 0$;
      else
        $E_j^{\text{eff}}(t - \tau_{ij}) \leftarrow E_{\min,j} + (E_0(j) - E_{\min,j}) \cdot e^{-\lambda_j (t - \tau_{ij})}$;
        contribution $\leftarrow \dfrac{E_j^{\text{eff}}(t - \tau_{ij}) \cdot \Phi_{ij}}{d_{ij}^k}$;
      end
      total $\leftarrow$ *total* + *contribution*;
    end
    $D_i(t) \leftarrow total \cdot R_s(t)$;
    ```
    // Optional noise term
    ```
    $D_i(t) \leftarrow D_i(t) + \varepsilon_i(t)$;
  end
end

Algorithm M.1: Computational Implementation of the GUDPL Model

© 2026 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Louvain-la-Neuve, Belgium. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

REGION
The Journal of ERSA
Powered by WU

Volume 13, Number 1, 2026, 95–114
DOI: 10.18335/region.v10i1.608

journal homepage: region.ersa.org
ISSN: 2409-5370

Disaster Risk and Property Valuation: A Study of Auction Assets in Earthquake-Prone Yogyakarta

Rahmat Saleh[1], Gea Dwi Asmara[1], Nurafifah Wulandari[2]

[1] Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2] Universitas Gadjah Mada, Yogyakarta, Indonesia

Received: 29 May 2025/Accepted: 2 March 2026

Abstract. This study aims to determine and identify the effects of property characteristics including exposure time, elevation adjustment, and topographic adjustment and disaster characteristics, such as seismicity, on the value of auctioned assets (collateral). The research employs multiple regression analysis using the ordinary least squares method. The findings indicate that property characteristics, including elevation, topography, exposure time, and seismicity-related disaster characteristics, significantly affect the value of auctioned assets. Specifically, the Topography and Seismic Adjustment variables have a positive and significant influence on the Auction Market Value of collateral property assets. This research introduces a novel approach to understanding asset auction dynamics in earthquake-affected areas by integrating economic valuation with disaster analysis. The study's originality lies in combining an economic analysis framework with disaster risk factors, an area seldom emphasized in previous asset auction research, which has typically addressed economic valuation and disaster impacts separately.

1 Introduction

In earthquake-prone regions like Yogyakarta, the potential for substantial financial loss in property and collateral asset auctions is significant. A major eruption could lead to auction value losses exceeding 30%, highlighting the grave stakes involved. Collateral asset auctions play a strategic role in the modern financial system, particularly as an effective solution for non-performing loans (Luck, Santos 2023). As a financial instrument, this process allows financial institutions to recover the value of assets associated with defaulted loans. It serves as a channel for distributing assets to individuals or entities that can productively utilize them. As such, collateral asset auctions contribute to market efficiency, create opportunities for the reutilization of economic resources, and support the overall stability of the financial system (de Roure, McLaren 2021).

Handling Non-Performing Loans (NPLs), based on data from national financial institutions, asset auctions account for an average recovery of up to 60-80% of collateral value, helping to reduce financial institutions' risk of loss. Moreover, in the most recent year, the total value of asset auction transactions in Indonesia exceeded IDR 30 trillion, reflecting high activity and market demand. The auction process serves as a mechanism for bank asset recovery and for redistributing assets to improve market efficiency (Uesugi et al. 2024). In a broader context, this activity supports financial system stability by helping to reduce the accumulation of NPLs, which, if unaddressed, can negatively impact credit growth and market liquidity (Zhang, Ibikunle 2023).

Collateral asset auction practices often face challenges stemming from inconsistent asset valuations, leading to disparities among parties, including financial institutions, independent appraisers, and bidders (Skočir, Lončarski 2024). In some cases, the valuation difference can be as high as 20-30%, depending on the method used, market conditions, or bias towards the asset's condition. These inconsistencies reduce transparency in the auction process and create conflicts of interest and legal risks, especially when one party feels aggrieved (Bellucci et al. 2021, Herweg, Schmidt 2019). A key issue overlooked in previous studies is the impact of disaster variables on valuation, which further exacerbates these gaps. To understand these valuation gaps, it's essential to recognize the differing motivations of stakeholders. Banks often prioritize asset recovery and may press for lower valuations to expedite sales, while appraisers aim for accuracy and market fairness. Bidders, on the other hand, seek undervaluation to maximize their investment potential. Highlighting these misalignments sets up the groundwork for using a risk-adjusted model as a tool to reconcile these differences and promote equitable asset listings. Gaps in asset valuation practices for property auctions have become a strategic issue in managing non-performing loans, particularly in the banking sector (Gaffeo, Mazzocchi 2018, Pandey 2019). These imbalances often arise from diverse valuation methods, non-uniform standards, and differing stakeholder interests (Maurin 2022). In property auction disputes, significant differences in asset values can delay credit settlement, trigger dissatisfaction among related parties, and even lead to lawsuits (Wardani 2020).

The symptoms that arise during collateral asset auctions reflect the complexity of the process, driven by internal and external dynamics (Anderson et al. 2023). Internal factors include the quality of asset valuation data, independent appraisers' competence, and financial institutions' internal mechanisms in determining collateral values (Cloyne et al. 2023, Donner 2020). Meanwhile, external factors include market fluctuations, inconsistent government regulations, and potential buyers' preferences and perceptions of asset value. These complexities often result in significant differences in the interpretation of asset values, leading to tensions among the parties.

Determining the value of an asset in an auction context relies on economic analysis and must also consider physical and environmental factors that affect the asset's intrinsic value (Calabrò et al. 2024). One important approach is to consider land elevation and topographic adjustments, as well as the potential for electoral seismicity, as additional valuation determinants (Follador et al. 2024). Specifically, elevation and topography serve as exposure factors influencing a property's vulnerability to seismic events. The potential for seismic disasters significantly impacts the economic value of an area, both in terms of physical assets and economic productivity (Amin et al. 2024). Seismicity represents a hazard intensity factor, shaping risk assessments and influencing investment decisions. Vulnerability to earthquakes can affect property valuations, infrastructure investments, and business decisions due to the risk of significant losses to strategic assets. In particular, regions with high seismic activity often experience reduced property values, higher insurance premiums, and greater investment in disaster mitigation technologies (Feliciano et al. 2023, Takeda, Inaba 2022, Xu et al. 2023). Linking seismic frequency directly to expected income loss and cap-rate shifts can provide clearer insights for investors. By translating geophysical risk factors into valuation levers, such as anticipated cash-flow disruptions and adjustments in expected returns, investors can make more informed decisions regarding the economic viability of investments in earthquake-prone areas.

For example, areas in active fault zones tend to have lower land values than more geologically stable areas despite having similar economic access and facilities (D'Apuzzo et al. 2022). Potential earthquakes also affect economic sectors such as tourism, industry, and trade, especially if key infrastructure is not designed to withstand disasters (Sousa et al. 2022). This research concerns potential volcanic seismic hazards, especially in areas around active volcanoes.

According to Oke et al. (2023), projections of climate change and population growth indicate that exposure to flood risk will increase across the United States, thereby increasing the potential for damage to human and economic systems. Flood risk assessment requires a comprehensive understanding of various spatially interdependent variables, including rainfall characteristics that cause flooding, the extent of the flooded area, the exposed population, and its vulnerability.

The paper by Hong et al. (2022) examines the impact of induced seismicity on the housing market in Pohang, South Korea, using a hedonic pricing model and a difference-in-differences approach. It highlights how perceived seismic risk affects regional housing values, with significant declines in property prices near the epicenter. The study emphasizes the importance of understanding spatial vulnerability in housing markets and shows that changes in risk perception can influence preferences for earthquake-resistant building structures, reflecting the principles of disaster economics.

The vulnerability of affordable housing to coastal flooding and sea-level rise: assessing risk through expected annual exposure metrics. While not directly addressing disaster economics or econometric approaches, this highlights the disproportionate exposure of affordable housing in various regional markets, particularly in states such as New Jersey, New York, and Massachusetts. The findings of Buchanan et al. (2020) underscore the need for strategic interventions to preserve affordable housing in high-risk coastal areas, reflecting broader implications for regional housing markets at risk.

Post-disaster economic resilience and recovery efficiency, particularly in the context of the Wenchuan Earthquake. Zhou et al. (2020) used econometric models, such as the ARIMA model and the Malmquist productivity index, to analyze economic recovery in disaster-affected areas. Although it does not specifically address regional housing markets or spatial vulnerability, it emphasizes the importance of understanding economic resilience and recovery processes to formulate effective disaster response strategies, which can, in turn, inform housing market dynamics in vulnerable areas.

Proactive acquisition of vacant land in flood-prone areas to reduce disaster economics by preventing future flood losses. Atoba et al. (2021) used a geodesign framework to assess at-risk regional housing markets, identifying vacant properties with high development potential in floodplains. This study uses cost-benefit analysis and predictive land-use modeling to evaluate spatial vulnerability, highlighting the economic implications of acquiring vacant land before development to reduce future flood-damage costs and increase community resilience.

Bui et al. (2022) integrates disaster economics by examining the economic impact of pluvial flooding on the regional housing market in Ho Chi Minh City, Vietnam. It uses a hedonic property model within a difference-in-differences framework and spatial econometric analysis to assess how flood risk affects house prices. The study highlights the spillover effects of flood risk on nearby properties, providing insights into spatial vulnerability and informing policymakers about flood insurance schemes and flood prevention policies.

Volcanic activity, such as eruptions or earthquakes caused by magma movement, can damage infrastructure, threaten residents' safety, and harm the natural resources that underpin the area's economy. The direct impacts of volcanic disasters include damage to the property, agriculture, and tourism sectors, which can significantly decrease asset values (You, Tesfamariam 2024). Electoral regions near active volcanoes, such as Mount Merapi in Indonesia's Special Region of Yogyakarta, often experience fluctuations in land and property values.

The study of spatial or geographical inequality is important for both scientific purposes and policy. From a scientific perspective, understanding spatial inequality enables researchers to uncover patterns and correlations that are vital to advancing knowledge across fields such as economics, public health, and environmental science. From a policy perspective, identifying and addressing geographical inequality is key to promoting social justice and economic development (Rey 2025). Furthermore, the varying auction discounts applied during the recovery of non-performing loans can significantly affect lower-income households, many of whom may be displaced as a result. These households often lack the resources to compete in the auction market, leading to potential social equity concerns. Addressing this issue can help ensure that policies in place contribute to the creation of more just cities, where economic recovery processes do not disproportionately affect the most vulnerable populations.

The electoral region of Yogyakarta, located near Mount Merapi, is one of the areas with significant volcanic seismic risk. Volcanic activity from Mount Merapi, one of the world's most active volcanoes, has a significant impact on the local economy (Pallister et al. 2019). Whenever a volcanic eruption or earthquake occurs, various economic sectors, such as property, agriculture, and tourism, can suffer significant losses. Infrastructure damage from lava flows, volcanic ash, and earthquakes can reduce the market value of land and property in the area around the mountain's base. However, this area also has significant economic value due to its proximity to Yogyakarta's city center (Rindrasih et al. 2024).

The Alert III level for Mount Merapi in Yogyakarta reflects a high level of vigilance against potential eruptions that could directly impact the social and economic life of surrounding communities. As Indonesia's most active volcano, Mount Merapi has a history of eruptions that often threaten residents' safety and damage infrastructure in the surrounding area. When Mount Merapi's status is at Alert Level III, there are significant risks to various economic sectors, including agriculture, tourism, and property. Regarding property values, an increase in alert status often leads to a decrease in land and property values in disaster-prone areas. Communities and investors tend to be reluctant to invest in areas at high risk of eruption (Houghton et al. 2021). However, some areas may experience increased land values due to disaster resilience and soil fertility generated by volcanic activity (Andreastuti et al. 2023, Martinez-Villegas et al. 2022).

Alert Level III for Mount Merapi signifies a heightened state of alert, indicating substantial volcanic activity. At this juncture, while a major eruption is highly probable, it remains impracticable to forecast the precise timing or the manifestation of the eruption. Typically, Alert Level III corresponds to substantial volcanic activity, encompassing phenomena such as minor eruptions, lava flows, or volcanic ash accumulation, which can impact the environs surrounding the mountain. The duration of this status is variable and contingent upon the development of volcanic activity. However, the occurrence of Alert Level III is not uniform. Mount Merapi, being one of the most active volcanoes in the world, can attain this status multiple times within a year; nevertheless, the duration and frequency are substantially influenced by the prevailing level of volcanic activity. An illustration of this phenomenon occurred in 2010, when Mount Merapi was at Alert Level III for several weeks. This was followed by a major eruption that caused substantial damage to the surrounding region, prompting mass evacuations and considerable material losses. However, in certain instances, Alert Level III can also be maintained for a shorter duration during periods of diminished volcanic activity, when there are only indications of minor eruptions or other disturbances that do not culminate in major disasters. Nevertheless, it is imperative to note that Alert Level III does not necessarily indicate a major eruption. In certain periods, this status merely signifies an escalation in gas activity, volcanic ash, or analogous volcanic phenomena that do not directly culminate in a major eruption. This substantiates the assertion that Alert Level III exhibits a high correlation with the potential for disaster. However, this relationship is not universally definitive, as there are instances in which the Alert Level III period concludes without any significant catastrophic events. Nevertheless, the risk persists, necessitating heightened vigilance. Mount Merapi, a highly active volcano, is situated on the border between the Special Region of Yogyakarta and Central Java. The mountain is a focal point of intensive volcanic monitoring due to its heightened activity. The Center for Volcanology and Geological Hazard Mitigation (PVMBG), under the Geological Agency of the Ministry of Energy and Mineral Resources, routinely monitors volcanic activity at Mount Merapi. Since November 5, 2020, Alert Level III has been sustained, indicating that seismic activity and lava dome growth persist at levels necessitating heightened vigilance. This state of elevated volcanic activity, characterized by significant seismic events and the growth of lava domes, does not inherently lead to a major eruption. The Alert Level III status, which signifies a period of heightened volcanic activity without mandating immediate large-scale evacuation measures, does not necessarily culminate in a major explosive eruption. There have been numerous instances in which Alert Level III was implemented in the absence of a major eruption; however, volcanic earthquakes and lava flows persisted. Nonetheless, Alert Level III functions as a crucial indicator for monitoring and anticipating heightened risk. While the correlation between Alert Level III and disaster events is not always definitive, it is considerable. For instance, the substantial eruption in 2010 that necessitated the evacuation of thousands of people and resulted in fatalities occurred after the implementation of Alert Level III. In many cases, Alert Level III marks a pre-eruption phase, indicating an escalation of volcanic activity that could culminate in a significant eruption.

As stated in the *Laporan Harian Gunung Api* (MAGMA Indonesia 2025), a substantial fluctuation in the number of earthquakes registered in Indonesia was documented *in 2024*. The highest number of earthquakes was recorded in April and August, with 54 and 53 events, respectively. These fluctuations indicate an increase in volcanic activity, a crucial indicator for determining alert status in the affected regions, particularly in the vicinity of Mount Merapi. The heightened frequency of earthquakes observed in these months often serves as the basis for implementing Alert Level III. Alert Level III is typically instituted when volcanic activity manifests significant patterns, such as acute fluctuations in seismic activity frequency in regions surrounding active volcanoes. In the given context, during April and August 2024, when a substantial increase in seismic activity was documented, it can be inferred that Mount Merapi's volcanic activity is at a heightened alert level. This directly correlates with an elevated potential for eruptions or other associated hazards.

Warnings associated with volcanic activity, such as Alert Level III on Mount Merapi, play an instrumental role in shaping public risk perception and economic behavior in disaster-prone regions. Communities surrounding Mount Merapi, particularly those in the zone closer to the crater, perceive these warnings not only as a signal of potential danger but also as an indicator to enhance preparedness. An escalation in seismic activity and changes in alert status often heighten public awareness of increased potential risks. Research indicates that when volcanic alert levels are elevated, communities tend to exhibit an enhanced perception of risk, often precipitating behavioral changes, including preparedness for potential disasters and evacuation. A study conducted in the aftermath of the 2010 eruption of Mount Merapi found a substantial increase in the community's perception of risk, particularly after the disaster's direct impact. This surge in risk perception prompted notable changes in the community's response to incoming warnings and in its adaptation to evolving volcanic hazards. The heightened perception of risk has been shown to significantly affect long-term attitudes, leading communities to exhibit increased vigilance toward potential hazards even after the alert status has been downgraded. Moreover, the economic behavior of local markets is also influenced by perceptions of volcanic risk (Medeiros et al. 2021). An increase in warning levels or the declaration of Alert III status has been shown to lead to a notable decline in investment interest in areas deemed high-risk for eruption, consequently resulting in a decline in property prices. This phenomenon occurs because investment and property market decisions are contingent on the perceived risk of potential damage (Hino, Burke 2021, Thompson et al. 2023). A number of studies have demonstrated that hazard notifications can engender short-term changes in property prices, with property values typically declining when Alert III status is implemented (Ikefuji et al. 2022, Miller, Pinter 2022, Shi, Naylor 2023). However, in the long term, some areas may experience increased property values if disaster risks are effectively managed or if infrastructure improvements reduce vulnerability. Consequently, public perception of volcanic warnings influences not only their awareness and preparedness for disasters, but also has a substantial impact on market behavior. Elevated volcanic alerts or warnings, such as Alert Level III, can affect the stability of local markets in the short term; however, their long-term impact can vary depending on mitigation factors and policies implemented in the region.

A study by Mani et al. (2021) underscores the need to enhance risk assessments that account for the cascading impacts of low-intensity volcanic eruptions (VEI 3-6) on critical infrastructure systems, particularly in the context of globalization. The present risk assessment proposes leveraging system mapping and evidence-based analysis methodologies, including horizon scanning and event tree analysis, to identify vulnerabilities in critical systems and thereby enhance resilience to volcanic risks. The paper underscores the need to draw on the expertise of professionals from a range of disciplines, including economic assessors, to formulate comprehensive strategies for managing volcanic risk. Additionally, it highlights the repercussions of volcanic eruptions on global supply chains, particularly in sectors reliant on critical infrastructure, such as technology and transportation, thereby emphasizing the urgency of more effective preparedness measures.

As stated by Jumadi et al. (2020), the present study employs risk modeling to manage disasters stemming from volcanic crises, with a particular focus on the community and regional levels, using Geographic Information System (GIS) technology. The incorporation of Agent-Based Modeling (ABM) facilitates the simulation of risk dynamics at the individual level during volcanic crises, encompassing the movement and behavior of people in response to hazards. The outcomes of this model are highly relevant to stakeholders in disaster management, as they offer valuable insights into the processes of evacuation decision-making and risk assessment during volcanic events.

A recent study by Epper, Fehr-Duda (2024) explores the relationship between risk tolerance and time discounting, shedding light on how they interact to shape the assessment of future prospects. Their research underscores the profound influence individual risk preferences exert on the evaluation of potential future outcomes. The aforementioned findings indicate that risk avoidance, discounting behavior, and preferences for resolving time uncertainty can influence decision-making in situations of uncertainty, as is frequently the case in the context of natural disasters. Consequently, these findings are pertinent to the development of risk assessment models in various scenarios, including disaster management, where uncertainty and rapid decision-making are paramount.

The study by Nesticò et al. (2023) examines the application of double discounting and declining methods in environmental economics, with a particular focus on their use in evaluating long-term investment projects with environmental impacts. This study underscores the significance of appraising country-specific discount rates, which are highly relevant to public policy and investment decision-making related to sustainable development goals. The conceptual framework put forth herein can be utilized in Cost-Benefit Analysis (CBA) for initiatives with substantial environmental ramifications, thereby facilitating a more precise evaluation of non-monetary ramifications. Furthermore, this paper underscores the importance of leveraging the Environmental Performance Index (EPI) to enhance the policy-making process and optimize the return on public investment in environmental quality. The methodology outlined in this paper is applicable across diverse consumption contexts, thereby demonstrating its versatility in evaluating consumption and its resulting environmental impacts.

The paper by Shi et al. (2020) underscores the significance of a methodical response model in disaster management, encompassing disaster risk assessment and modeling. The effectiveness of disaster response and risk management depends on a comprehensive understanding of the disaster system, encompassing hazards, the physical environment, and socioeconomic exposure. This framework facilitates precise measurement of disaster risk and underpins risk pricing. It is imperative to enhance risk management strategies and technologies to ensure effective disaster response and appropriate risk pricing.

Recent research (Lee et al. 2022) underscores the significance of incorporating social vulnerability into disaster risk pricing models. The researchers propose the Cost of Social Vulnerability to Disasters (CSVDM) model, which estimates the financial impact of disasters by accounting for multiple dimensions of social vulnerability. This framework enables governments to make more accurate estimates of disaster costs, allocate resources more efficiently, and manage the financial risks associated with disasters. Consequently, the model contributes to improving disaster management and developing superior recovery strategies.

The study by Mistry, Lombardi (2022) adopts a compound distribution approach, which integrates the frequency of disaster events and their severity. The Cox-Ingersoll-Ross model for interest rates and the non-exceedance loss curve from the vulnerability model are utilised within this framework, thereby facilitating the effective pricing of disaster bonds. This process enables municipal authorities to transfer the financial risk associated with earthquakes to the capital market. Consequently, this enhances urban resilience in the face of disasters.

The end-to-end risk modeling framework proposed in Cremen et al. (2022) emphasizes the importance of a comprehensive approach to disaster risk pricing. Cremen et al. (2022) integrates the physical and social impacts of disasters, thereby enabling the characterization of flexible risk metrics that go beyond mere asset losses. The proposed framework enhances the ability to inform disaster risk-related policy decisions by incorporating uncertainty in future urban environments and adopting people-centered participatory decision-making processes. Ultimately, this improvement in assessment models for disaster risk pricing is achieved.

Yogyakarta not only has a high seismic risk but is also a vital cultural and economic center in Indonesia. The city is renowned as an educational hub, with many universities and other institutions attracting thousands of students from across the region. In addition, the tourism sector also plays an important role, with many tourist destinations located around Mount Merapi. Therefore, property value fluctuations in this area are a very important indicator of how seismic risk affects sectors highly dependent on economic stability. Yogyakarta is highly relevant as a case study for collateral asset valuation, especially in the context of asset auctions and property valuation in disaster-prone areas. The frequently changing alert status of Mount Merapi, such as Alert Level III, influences market perceptions of investment risk in the area. Looking ahead, the continued expansion of Yogyakarta's educational and tourism sectors could further affect property values and auction outcomes. As the city grows and attracts more residents and visitors, demand for property may increase, potentially altering the coefficients currently used to assess valuation. Urban planners and stakeholders should consider these factors when predicting future market trends and investment opportunities.

Given the context, studying the valuation of seized collateral assets in Yogyakarta from both economic and disaster risk perspectives is essential. This research addresses the urgent need for an integrated approach to asset auction dynamics in seismic-prone areas by combining economic valuation with disaster analysis. The study aims to identify how property characteristics such as elevation, topography, and exposure time, along with disaster factors like seismicity, affect auction asset values. The originality of this research lies in merging economic analysis with disaster risk factors, a combination rarely emphasized in previous asset auction studies, which typically address these elements separately. Furthermore, the framework's adaptability to other types of disasters, like floods or cyclones, underscores its scalability, potentially broadening the application and attracting a wider readership interested in cross-hazard analysis.

2 Method

Data were collected using purposive, non-probability sampling, focusing on properties involved in transactions, offers, or auctions within 24 months before the valuation date. The study analysed 76 observations, determined using the Slovin formula, considering data access, cost, and time constraints. To ensure the representativeness of the sample and mitigate concerns over selection bias, a power analysis was conducted. This analysis confirmed that the 24% sample of the total 312 auctions is sufficient to infer trends within the population, considering the heterogeneity of market conditions in the Yogyakarta Special Region. Data were gathered from both primary and secondary sources over nine months, covering four districts and one city. Primary data were obtained through interviews with property owners, while secondary data included auction records, bank reports, and independent appraisals (KJPP). Volcanic activity data were also used to supplement and validate the findings. By providing a brief description of our sampling protocol, we invite other regions to replicate this method. This approach not only sets the groundwork for comparative analysis but also encourages a wider academic engagement to explore property valuation in disaster-prone areas.

This study uses cross-sectional data from 76 observations in the Yogyakarta Special Region. This approach enables detailed analysis of property market variables, including economic valuation and disaster characteristics such as seismic activity from Mount Merapi. The goal is to provide an accurate understanding of property market dynamics and disaster potential in the region.

Table 1: Operational Definition of Research Variable

Variable	Definition	Unit
	Dependent Variable	
Auction Market Value (LG_INP)	Indication of Market Value of Assets at the time of collateral auction process	Rupiah (Rp)
	Independent Variables	
Exposure Time (LG_WE)	The results of the quality rating method search related to the length of the exposure period for the sale of collateral auction assets	Day
Elevation Adjustment (LG_AE)	This study utilizes data from the working papers of the independent appraisal report of the public appraisal service office to examine asset value adjustments relative to the average level of elevation, equality, and depression.	Rupiah (Rp/cm)
Topography Adjusment (LG_T)	Indication of Asset Value Adjustment to the condition of flat/hilly/downhill/wavy level	Rupiah (Rp)
Seismicity (LG_K)	Potential number of seismic hazards of Mount Merapi in the form of volcanic/hybrid/volcanic clusters	Number of Occurrences

The present study employs multiple regression analysis with the Ordinary Least Squares (OLS) approach to examine the effect of independent variables on dependent variables. Statistical modelling and data analysis using the OLS approach were selected for their ability to generate efficient and consistent parameter estimates and to facilitate hypothesis testing to assess the significance of the variables' effects within the scope of this study. To justify the use of OLS assumptions and assure readers regarding the validity of the coefficient estimates, diagnostic tests such as the Breusch-Pagan test for heteroskedasticity and Moran's I test for spatial autocorrelation were conducted. These tests confirmed that our error terms behave well and that neighboring properties do not unduly bias estimates. The results of these tests, which passed, reinforce the reliability and robustness of the model used in the analysis.

This study uses statistical tests to evaluate the model's accuracy, including the F-test for overall significance, the t-test for the individual effect of independent variables, and the Coefficient of Determination (R^2) to explain the variation in the dependent variable. Classical Assumption tests, including normality, linearity, multicollinearity, autocorrelation, and heteroscedasticity, are also applied to ensure the model meets statistical criteria and produces reliable estimates.

The research approach shows that auction asset valuation cannot rely solely on economic market value but requires a multidimensional analysis that includes physical aspects of the environment and disaster potential, integrating elevation and topography adjustments, as well as disaster risk from Mount Merapi's seismic activity, into the valuation method. This research applies the Analytical Survey method to analyze the influence of the variables under study. The operational definitions of variables are formulated in detail to clarify the terms used and to set operational limits for implementing the following research concepts.

Table 1 presents a comprehensive overview of the research variables and their respective operational definitions. The Auction Market Value (LG_INP), measured in Rupiah (Rp), underwent a logarithmic transformation to normalize its distribution, given the extremely wide range in the original data. This transformation allows for the expression of variables in percentage terms, providing percentage elasticities valuable for investors conducting sensitivity analyses. This method was employed to minimize extreme fluctuations in market value and to clarify the relationship between auction market value and other factors. (For practitioners, this means they can interpret a 1% change in a variable as indicative of a corresponding shift in auction market value.) Exposure Time (LG_WE), which denotes the number of days required to sell or auction a property, is logarithmically transformed. This is done to mitigate the impact of uneven distribution and provide a more precise representation of the relationship between exposure time duration and property market value. Elevation Adjustment (LG_AE), which measures the

Table 2: Descriptive Statistic

Variable	Mean	Min	Max
Auction Market Value (LG_INP)	Rp994.913.255	Rp234.000.000	Rp4.252.125.000
Exposure Time (LG_WE)	237	180	360
Elevation Adjustment (LG_AE)	Rp91.882	Rp117	Rp504.000
Topography Adjusment (LG_T)	Rp146.826	Rp9.010	Rp590.635
Seismicity (LG_K)	36	9	54

impact of elevation changes on property value, is quantified in Rupiah per centimeter (Rp/cm). A logarithmic transformation is employed to achieve a uniform distribution of the data and streamline the analysis of the effect of elevation on auction-market value, as variations in elevation have been observed to influence a property's attractiveness. The concept of topography (LG_T) pertains to the influence of topographical conditions, including plains, hills, and valleys, on property value. This variable is measured in Rupiah. The incorporation of logarithms in this variable ensures the analysis of significantly varying topographical values in a more controlled manner, thereby clarifying the impact of geographical conditions on property market value. Seismicity (LG_K), which gauges the frequency of earthquakes or seismic activity over a given time period, is quantified as the number of events. By applying a logarithmic transformation, the impact of extreme fluctuations in earthquake frequency can be mitigated, thereby facilitating a more consistent analysis of the influence of seismic risk on property market value.

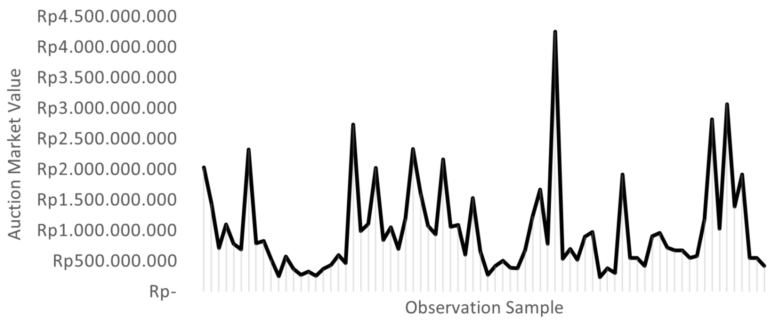
The Auction Market Value (LG_INP), measured in Rupiah (Rp), underwent a logarithmic transformation to normalize its distribution, given the extremely wide range in the original data. This method was employed to minimize extreme fluctuations in market value and to clarify the relationship between auction market value and other factors. Exposure Time (LG_WE), which denotes the number of days required to sell or auction a property, is logarithmically transformed. This is done to mitigate the impact of uneven distribution and provide a more precise representation of the relationship between exposure time duration and property market value. Elevation Adjustment (LG_AE), which measures the impact of elevation changes on property value, is quantified in Rupiah per centimeter (Rp/cm). A logarithmic transformation is employed to achieve a uniform distribution of the data and streamline the analysis of the effect of elevation on auction-market value, as variations in elevation have been observed to influence a property's attractiveness. The concept of topography (LG_T) pertains to the influence of topographical conditions, including plains, hills, and valleys, on property value. This variable is measured in Rupiah. The incorporation of logarithms in this variable ensures the analysis of significantly varying topographical values in a more controlled manner, thereby clarifying the impact of geographical conditions on property market value. Seismicity (LG_K), which gauges the frequency of earthquakes or seismic activity over a given time period, is quantified as the number of events. By applying a logarithmic transformation, the impact of extreme fluctuations in earthquake frequency can be mitigated, thereby facilitating a more consistent analysis of the influence of seismic risk on property market value.

3 Result

This study collected 76 observations of property offers and transactions in the Yogyakarta Special Region, focusing on data within 24 months of the valuation date. Data collection followed the research schedule and relied on verified, reliable sources, including direct communication and documentation.

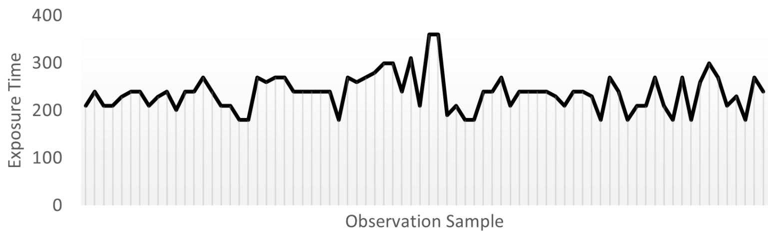
Descriptive statistics in Table 2 summarize the study data, including the mean, median, maximum, and minimum values for both independent and dependent variables.

Based on BNPB data on Mount Merapi's eruption history from the 17th to the 20th century, the number of fatalities from eruptions, both from hot clouds and lava, has exceeded 5,200. In the 2010 eruption, 198 people died directly from hot clouds, while 188 people died indirectly. In addition, the number of displaced people reached 400,000. The



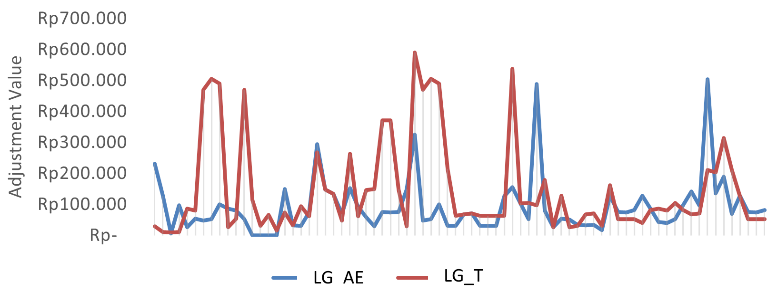
Notes: Collateral property auction values in the Yogyakarta Special Region range from Rp234,000,000 to Rp4,252,125,000, with an average of Rp994,913,255. The data indicate significant value disparities between districts and cities.

Figure 1: Trend chart of Auction Market Value



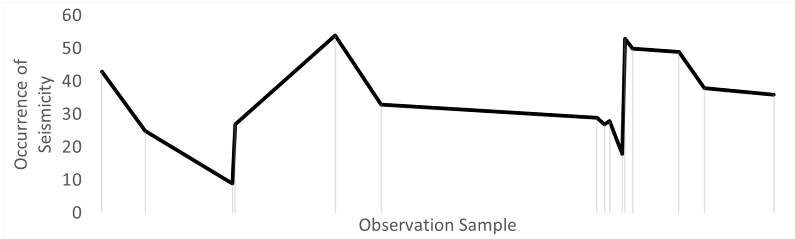
Notes: Exposure time for collateral auction asset sales ranges from 180 to 360 days, with an average of 237 days. Longer exposure periods are influenced by location, market conditions, property prices, and demand. Typically, more competitive markets require shorter exposure times.

Figure 2: Exposure Time Trend chart



Notes: Elevation and topography adjustments reflect changes in auction asset values based on elevation relative to the road and terrain type. Elevation adjustments range from Rp117 to Rp504,000, with an average of Rp91,882. These factors influence property attractiveness, accessibility, and potential use. Topography adjustments range from a minimum of Rp9,010 to a maximum of Rp590,635, with a mean value of Rp146,826. These adjustments account for various terrain conditions——such as flat, hilly, downhill, or undulating——and their subsequent impact on property functionality, accessibility, and development potential.

Figure 3: Trend graph of Elevation Adjusment (LG_AE) and Topography Adjusment (LG_T)



Notes: Seismicity data show a range of 9 to 54 seismic events, with an average of 36. The frequency of seismic events from Mount Merapi—including impact, hybrid, and volcanic types—can significantly affect auction asset values, especially for properties in disaster-prone areas. Seismic impact analysis is supported by risk maps, historical volcanic activity data, and assessments of property disaster resilience.

Figure 4: Seismic Trend graph

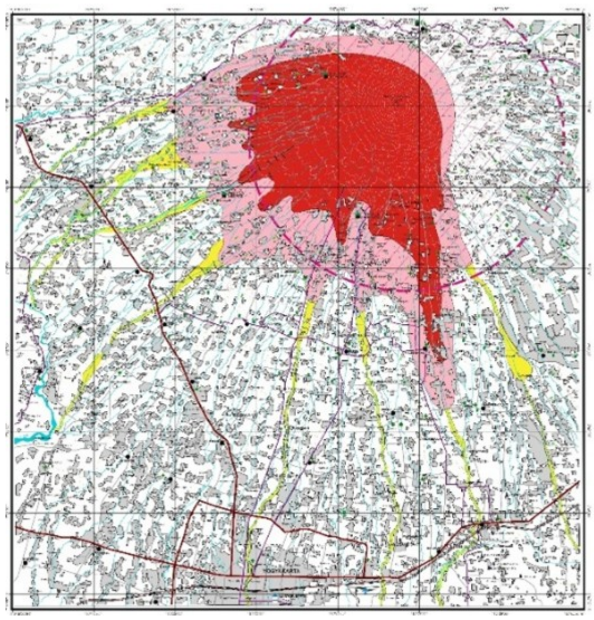


Figure 5: Volcanic hazard map of the study area (Sayudi et al. 2010), issued by the Center for Volcanology and Geological Hazard Mitigation (CVGHM)

hot clouds generated by the eruption damaged several villages, including Umbulharjo, Kepuharjo, Glagaharjo, Argomulyo, and Wukirsari in Cangkringan Sub-district, Sleman Regency, as well as Balerante in Kemalang Sub-district, Klaten Regency. The map of Mount Merapi Disaster-Prone Areas serves as a reference for local governments in designing regional spatial planning to reduce risk in disaster-prone areas. In addition, the surrounding community can use it as a guide for self-rescue efforts in the event of an eruption or lava flood. This map is also a reference in regional spatial planning related to the development and commercialization of property assets.

Table 3 presents the regression results, confirming the appropriateness of the log-transformed model based on the statistical criterion tests. The model yields an R-squared (R^2) value of 0.585, indicating that approximately 58.5% of the variability in auction market values is explained by the independent variables, namely exposure time, elevation adjustment, topography adjustment, and seismicity. Notably, the F-statistic of 7.792 is significantly greater than the F-table critical value of 2.20 ($F > 2.20$). This confirms that the independent variables collectively have a significant impact on the dependent variable. These findings suggest that the model effectively reduces pricing uncertainty, providing a more reliable framework for asset valuation and risk management in decision-making processes.

Table 3: Regression Results

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.444198	1.018419	8.291478	0.0000
LG_WE	-0.707430	0.454560	-1.556297	0.1241
LG_AE	0.013190	0.021493	0.613693	0.5414
LG_T	0.231408	0.069668	3.321577	0.0014
LG_K	0.610295	0.162773	3.749375	0.0004

R-squared	0.585089	Mean dependent var	8.894.737
Adjusted R-squared	0.265939	S.D. dependent var	0.289779
S.E. of regression	0.248275	Akaike info criterion	0.114967
Sum squared resid	4.376.476	Schwarz criterion	0.268304
Log likelihood	0.631272	Hannan-Quinn criter.	0.176248
F-statistic	7.792.841	Durbin-Watson stat	1.627.367
Prob(F-statistic)	0.000029		

Table 4: Descriptive Ordinary Least Squares

Multiple Regression Analysis	Test Statistics	Description
	T-test	
Exposure time	T-Stat value of -1.56 < T-tab 1.667 and prob value 0.1241 > 0.05	Not significant
Elevation adjustment	T-Stat value of 0.61 < T-tab 1.667 and prob value of 0.5414 > 0.05	Not significant
Topography adjustment	T-Stat value of 3.32 > T-tab 1.667 and prob value 0.0014 < 0.05	Significant
Seismicity	T-Stat value of 3.75 >T-tab 1.667 and prob value 0.0004 <0.05	Significant
	F-test	
Full model	F-stat 7.792841 > F-table 2.20	Significant
	Classical assumption tests	
Normality test	probability value 0.1873 > 0.05 and Jarue-Bera 3.3505	No indication of non-normality
Multicollinearity test	value R-square 0.5851 < 0.8	No indication of multicollinearity
Heteroscedasticity test	Probability value passed test 0.0662 > 0.05	No evidence of heteroscedasticity

Table 4 summarizes the statistical results of the hypothesis testing and classical assumption diagnostics. The t-test results indicate that topography adjustment and seismicity are the significant determinants of the dependent variable, while exposure time and elevation adjustment do not show statistical significance. Furthermore, the F-statistic of 7.792 confirms the overall significance of the model. Diagnostic tests demonstrate that the model adheres to the required classical assumptions. Specifically, the Jarque-Bera test yields a probability value of 3.3505 and ($0.1873 > 0.05$), indicating no violation of the normality assumption. Additionally, the multicollinearity test shows an R-square value of 0.5851, which is well below the 0.8 threshold, confirming no serious multicollinearity among the independent variables. No evidence of heteroscedasticity was found, ensuring the reliability of the OLS estimates.

4 Discussion

Based on the t-test results, the variables Topographic Adjustment (3.321577) and Seismicity (3.749375) have a positive and significant influence on the Auction Market Value of collateralized property assets. To assist property assessors in recognizing potential risks, specific thresholds can be highlighted. For example, it's observed that an exposure to more than 40 seismic events annually could lead to a reduction in property values by approximately 5%. Similarly, topographical changes that result in slope inclines exceeding 15% could act as red flags, suggesting potential depreciation in the asset's market

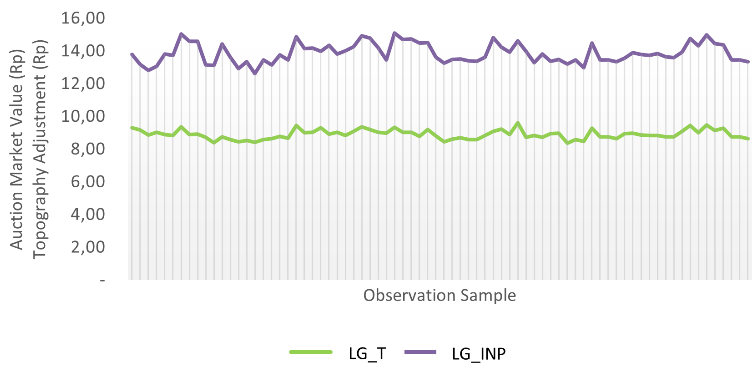


Figure 6: Trend chart of Auction Market Value (LG_INP) and Topography Adjusment (LG_T)

value. By translating these coefficients into simple rules of thumb, property assessors can more effectively gauge risk factors and make informed valuations.

Figure 6 illustrates the trend relationship between Auction Market Value (LG_INP) and Topographic Adjustment (LG_T). The regression analysis demonstrates that auction market value is significantly influenced by topographic adjustment, with a coefficient of 0.231408. This indicates that each Rp1 increase in topographic adjustment raises the Auction Market Value of collateral property assets by 0.2314%. This result supports the hypothesis that topography adjustment has a positive and significant effect on auction market value. Incorporating topographic characteristics into asset valuation is essential, as land topography—whether flat, hilly, downhill, or undulating—substantially affects a property's market value by influencing functionality, visual appeal, accessibility, and development costs. Flat land is generally considered ideal due to its facilitation of development and optimal accessibility, resulting in higher market values.

In contrast, while hilly land may offer aesthetic advantages, such as attractive views, it may entail additional engineering costs, such as soil stabilization, which may reduce its market value if there is no specific demand. Downhill and undulating conditions often require a greater investment in infrastructure, such as drainage or land leveling, which can affect asset value assessment. Properties with downhill topography also face risks such as waterlogging or inconvenient access, which can reduce their attractiveness in the auction market. Auction asset valuations that consider topography are conducted through a comparative market analysis, in which properties with similar characteristics are compared to assess the impact of topography on price. In addition, an analysis of remediation costs and land development potential was used to estimate the economic impact of less-than-ideal topographic conditions. The results of this adjustment show that topography that favors development and accessibility will significantly increase property values. In contrast, topography that creates obstacles can reduce market values unless other factors, such as a premium location or a unique landscape, compensate. As such, topography is an important variable that significantly affects the auction market value of property assets and should be thoroughly analyzed in the valuation process.

In the hedonic pricing literature, property valuation is conducted by identifying and measuring physical and environmental attributes that influence consumer perceptions of property value (Shi et al. 2022, Zheng, Wang 2024). Topography, as one factor in this model, affects property prices through various dimensions, including accessibility, land use, and aesthetics. This exposition elucidates the fundamental theory posited by Rosen (1974), which holds that real estate prices are determined by a suite of attributes that influence consumer preferences. Topography is thus considered an integral component of these attributes (Boto-García, Leoni 2022, Kovacs et al. 2022, Rosen 1974).

The hedonic pricing model serves as the foundation for property price assessment, assuming that property values are influenced by various attributes, including topography, which encompasses land conditions such as flat, hilly, sloped, or undulating. Topog-

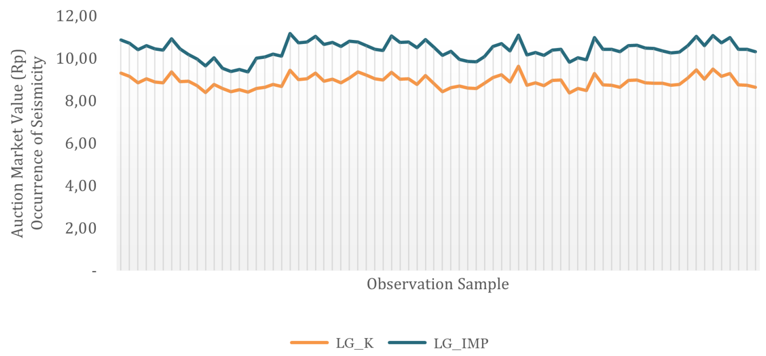


Figure 7: Trend graph of Auction Market Value (LG_INP) and Seismic Adjustment (LG_K)

raphy plays a significant role in determining property prices by affecting accessibility, construction costs, and aesthetics. However, when using the hedonic pricing model, it is imperative to acknowledge that topography must be examined alongside other salient factors, such as location and socio-economic characteristics, which collectively influence market prices. Hilly land is often appraised at a lower value due to the additional costs associated with soil stabilization and other infrastructure. However, its natural beauty can enhance its appeal to buyers who prioritize quality of life. Conversely, sloping or undulating land increases development costs, such as drainage or land leveling, which can diminish the property's market value. Nevertheless, consumer preferences and local demand can often mitigate the negative impact of poor topography. Furthermore, premium locations or unique landscapes have been demonstrated to offset physical obstacles arising from topographical conditions. Additionally, changes in government policy and infrastructure developments, such as road improvements or stormwater management, have been found to enhance property values, even when topography is suboptimal.

Secondly, the regression results demonstrate that the auction market value is significantly influenced by seismic adjustments, with a coefficient of 0.610295. Figure 7 illustrates the trend relationship between seismic activity (LG_K) and auction market value (LG_INP), visually confirming this positive correlation. This indicates that each additional seismic adjustment of 1 seismic event will increase the auction market value of collateral property assets by 0.6103%. This finding supports the hypothesis that seismic adjustments have a positive and significant effect on auction market value. However, this counter-intuitive result warrants further exploration to ensure robustness. One plausible alternative hypothesis is that the areas affected by seismic activity may attract a specific segment of buyers who are drawn to disaster-driven amenities, such as fertile land for agriculture, scenic landscapes, or enhanced infrastructure developed in response to frequent natural events. These factors can increase demand, possibly offsetting the risk perception typically associated with seismic zones. To further explore this hypothesis, examining whether demand, rather than risk, might drive the positive coefficient observed is a potential avenue for investigation. As a robustness check, instrumenting seismicity with exogenous eruption alerts could clarify the underlying factors contributing to this result. By exploring potential endogeneity, we aim to strengthen confidence in the observed effects. Seismic activity, including landslide, hybrid, and volcanic earthquakes, has been shown to exert a significant influence on market perception, investment feasibility, and property attractiveness, even when the property is situated in a high-risk zone. Geological phenomena, such as earthquakes and volcanic activity, have been shown to significantly influence real estate valuation. In areas deemed high risk, a decline in property prices has been observed, attributable to uncertainty and the potential for damage to infrastructure or buildings from disasters. This factor, compounded by the fear and heightened risk perception among prospective buyers, as well as the possibility of long-term management of these risks, results in lower property prices in vulnerable zones relative to safer zones. This issue is also related to higher insurance costs and the need for additional investment in building or infrastructure resilience in earthquake-prone areas. However, despite the potential for damage, some areas with high volcanic or seismic activity, such as around Mount Merapi, still have fairly high market demand. The reasons for this phenomenon require further investigation and understanding (Farquharson, Amelung 2022, Hardiansyah et al. 2020, Terry et al. 2022).

High demand persists in the area surrounding Merapi due to its highly supportive environment. For instance, the northern portion of Yogyakarta, closer to Merapi, offers fresh, clean air and natural beauty, appealing to individuals seeking a healthier, more comfortable living environment. In recent years, property buyers, particularly those from highly polluted metropolitan areas, have placed significant emphasis on air quality and environmental factors. Moreover, numerous regions have been outfitted with sufficient social and public facilities, including educational institutions, medical centers, and commercial hubs (Das et al. 2022, Jennings et al. 2021, Zhang et al. 2022). The presence of such amenities enhances the desirability of these locales, particularly for families seeking an improved quality of life. With access to quality educational facilities and healthcare services, many prospective buyers prioritize the quality of life they can attain in relatively tranquil areas, despite the potential for natural disasters. Despite the evident hazard of potential disasters, many contend that the risk of harm can be substantially reduced through the implementation of effective disaster mitigation systems and advanced technologies designed to fortify critical infrastructure. Consequently, the market value of properties situated in regions such as northern Yogyakarta, which lies in close proximity to Mount Merapi, can remain high or even continue to rise. This phenomenon can be attributed to a shift in the perceived risk associated with these areas. Recent studies indicate an increasing awareness among both communities and investors that disasters are unpredictable and cannot be precisely predicted. Moreover, when effective risk mitigation policies are in place—such policies include early warning systems, building reinforcement, and sustainable urban planning—a greater sense of security is engendered, leading to increased investment in these areas (Kurnio et al. 2021, Thamarapani, Rockmore 2022, Thompson et al. 2023). Additionally, the appeal of a healthy lifestyle in these areas has been shown to influence property market demand (Chen et al. 2022). The phenomenon of individuals and families choosing to reside in areas more closely associated with natural environments, even when these areas are near regions susceptible to natural disasters, has been widely documented (Taylor, Aalbers 2022). This phenomenon has been variously referred to as 'disaster gentrification', a term that denotes the increase in property prices in such areas despite heightened risk (Thompson et al. 2023).

In addition to environmental factors, the affordability of property in the northern part of Yogyakarta also contributes to rising property prices. The phenomenon of rapid urbanization in major cities has driven a growing demand for affordable, well-connected housing. Despite its designation as a disaster-prone area, the region surrounding Merapi often offers property prices lower than those in the densely populated, costly city center. The decision to invest in real estate in this region is often driven by the desire to acquire a residence at a comparatively lower cost while maintaining access to superior amenities and an enhanced quality of life. Moreover, improvements in infrastructure in high-risk areas significantly influence property value maintenance or appreciation. Local and national governments frequently allocate substantial financial resources to infrastructure development and disaster mitigation (Fischer et al. 2022, Saputra et al. 2021). Such initiatives include the construction of more robust roads, efficient drainage systems, and structures more resistant to seismic activity. These policies serve to indirectly reassure prospective buyers that the potential for damage from disasters can be effectively managed. It should be noted that even within risk zones, locations with strategic value or significant economic potential may still retain market appeal. Seismicity is a primary risk factor that affects property auction market values, particularly in regions susceptible to natural disasters. Accurate assessments must consider this risk through data-driven methodologies, geographic risk analysis, and an evaluation of available mitigation measures. Consequently, property market values can reflect the equilibrium between potential returns and seismic risk, thereby ensuring equitable transparency for all parties involved in the auction process.

5 Conclusion

Simultaneously, the independent variables collectively influence auction market value, but only topography adjustment (LG_T) and seismicity (LG_K) have a significant individual effect in Yogyakarta Province for 2023-2024. This research is limited to exposure time, elevation adjustment, topography, and seismicity. To enhance the predictive power of our model, incorporating insurance premium data could be particularly beneficial. This addition would provide a more comprehensive view of the financial risks associated with properties in disaster-prone areas, potentially improving the model's accuracy. Further studies should consider additional determinants and broader geographic coverage, as this study was constrained by time, cost, and location. A specific limitation imposed by the 24-month data window is its restriction on capturing longer-term price adjustments. This limitation may prevent a comprehensive understanding of the prolonged effects that seismic activity and other variables can have on property values. Acknowledging this trade-off openly enhances the study's credibility and sets the stage for future longitudinal research that can explore these impacts over a more extended period.

Based on an understanding of the creation of auction market value, an appropriate assessment education program is needed that is not only influenced by economic factors but also has significant seismic potential. The government must develop policies and programs that support disaster risk mitigation in vulnerable areas, including spatial planning, infrastructure strengthening, and property protection through insurance.

To enhance policy relevance, the findings suggest two actionable recommendations: first, integrate hazard maps into loan-to-value limits to better assess the risks associated with collateral in disaster-prone areas; second, implement mandatory risk assessments for properties located in high-seismic zones to ensure comprehensive valuation practices. Clear policies for disaster risk assessment can help maintain the stability of collateral market value, thereby providing a sense of security for collateral owners, financial institutions, and investors. As Yogyakarta continues to face seismic risks, robust policies in these areas can mitigate potential economic impacts and foster safer development practices

Acknowledgements

In 2024, LPPM Universitas Ahmad Dahlan supported our research through the Universitas Ahmad Dahlan internal research fund grant program. The author would like to thank you for all the support provided. Looking forward, we are eager to partner with practitioners and institutions to expand our dataset beyond the confines of UAD. We invite cross-sector collaborations that can enhance data sharing and support the scaling of our research model, driving innovation in asset valuation and disaster risk analysis.

References

Amin M, Nezami E, Kashani H (2024) The value of a statistical life in assessing seismic resilience enhancement policies: A stated preference binary choice experiment. *Journal of Safety Science and Resilience* 6: 150–162. CrossRef

Anderson G, Bahaj S, Chavaz M, Foulis A, Pinter G (2023) Lending relationships and the collateral channel. *Review of Finance* 27: 851–887. CrossRef

Andreastuti S, Paripurno E, Subandriyo S, Syahbana D, Prayoga A (2023) Volcano disaster risk management during crisis: Implementation of risk communication in Indonesia. *Journal of Applied Volcanology* 12: 3. CrossRef

Atoba K, Newman G, Brody S, Highfield W, Kim Y, Juan A (2021) Buy them out before they are built: Evaluating the proactive acquisition of vacant land in flood-prone areas. *Environmental Conservation* 48: 118–126. CrossRef

Bellucci A, Borisov A, Giombini G, Zazzaro A (2021) Estimating the relationship between collateral and interest rate: A comparison of methods. *Finance Research Letters* 43. CrossRef

Boto-García D, Leoni V (2022) The economic value of coastal amenities: Evidence from beach capitalization effects in peer-to-peer markets. *Environmental and Resource Economics* 84: 529–557. CrossRef

Buchanan MK, Kulp S, Cushing L, Morello-Frosch R, Nedwick T, Strauss B (2020) Sea level rise and coastal flooding threaten affordable housing. *Environmental Research Letters* 15: 124020. CrossRef

Bui N, Wen L, Sharp B (2022) House prices and flood risk exposure: An integration of hedonic property model and spatial econometric analysis. *The Journal of Real Estate Finance and Economics* 69: 100–131. CrossRef

Calabrò F, Giovanna E, Calabrò F, Minniti G, Fotia A, Pucinotti R (2024) Recovery of historic real estate: Cycle costing and economic feasibility. *Procedia Structural Integrity* 64: 1759–1766. CrossRef

Chen K, Lin H, You S, Han Y (2022) Review of the impact of urban parks and green spaces on residence prices in the environmental health context. *Frontiers in Public Health* 10: 993801. CrossRef

Cloyne J, Ferreira C, Froemel M, Surico P (2023) Monetary policy, corporate finance, and investment. *Journal of the European Economic Association* 21: 2586–2634. CrossRef

Cremen G, Galasso C, McCloskey J (2022) A simulation-based framework for earthquake risk-informed and people-centered decision making on future urban planning. *Earth's Future* 10: 2021 002388. CrossRef

Das R, Chatterjee T, Ivaldi E, Das R, Chatterjee T, Ivaldi E (2022) Nexus between housing price and magnitude of pollution: Evidence from the panel of some high-and-low polluting cities of the world. *Sustainability* 14. CrossRef

de Roure C, McLaren N (2021) Liquidity transformation, collateral assets and counterparties. *Central Bank Review* 21: 119–129. CrossRef

Donner H (2020) Determinants of a foreclosure discount. *Journal of Housing and the Built Environment* 35: 1079–1097. CrossRef

D'Apuzzo M, Evangelisti A, Rasulo A, Nicolosi V (2022) Simplified approach to integrate seismic retrofitting prioritization with social cost evaluation: A case study in central Italy. *Journal of Traffic and Transportation Engineering (English Edition)* 9: 447–460. CrossRef

Epper T, Fehr-Duda H (2024) Risk in time: The intertwined nature of risk taking and time discounting. *Journal of the European Economic Association* 22: 310–354. CrossRef

Farquharson J, Amelung F (2022) Volcanic hazard exacerbated by future global warming-driven increase in heavy rainfall. *Royal Society Open Science* 9. CrossRef

Feliciano D, Arroyo O, Cabrera T, Contreras D, Valcárcel Torres J, Gómez Zapata J (2023) Seismic risk scenarios for the residential buildings in the Sabana Centro province in Colombia. *Natural Hazards and Earth System Sciences* 23: 1863–1890. CrossRef

Fischer E, Biondo A, Greco A, Martinico F, Pluchino A, Rapisarda A, Fischer E, Biondo A, Greco A, Martinico F, Pluchino A, Rapisarda A (2022) Objective and perceived risk in seismic vulnerability assessment at an urban scale. *Sustainability* 14. CrossRef

Follador V, Donà M, Carpanese P, Saler E, D'Alpaos C, Porto F (2024) Seismic retrofit cost model for Italian masonry residential buildings to support territorial-scale risk analysis. *International Journal of Disaster Risk Reduction* 105: 1–23. CrossRef

Gaffeo E, Mazzocchi R (2018) “the price is right”: Using auction theory to enhance competition in the NPL market. *Journal of Banking Regulation* 20: 104–112. CrossRef

Hardiansyah M, I. B, C., Priyanto S (2020) Analysing vulnerability of road network and guiding evacuees to sheltered areas: Case study of Mt Merapi, Central Java, Indonesia. *Case Studies on Transport Policy* 8: 1329–1340. CrossRef

Herweg F, Schmidt K (2019) Procurement with unforeseen contingencies. *Management Science* 66: 2194–2212. CrossRef

Hino M, Burke M (2021) The effect of information about climate risk on property values. *Proceedings of the National Academy of Sciences of the United States of America* 118: 2003374118. CrossRef

Hong J, Jo H, Seo D, You S, Hong J, Jo H, Seo D, You S (2022) Impact of induced seismicity on the housing market: Evidence from Pohang. *Buildings* 12. CrossRef

Houghton B, Cockshell W, Gregg C, Walker B, Kim K, Tisdale C, Yamashita E (2021) Land, lava, and disaster create a social dilemma after the 2018 eruption of Kīlauea volcano. *Nature Communications* 12: 1223–. CrossRef

Ikefuji M, Laeven R, Magnus J, Yue Y (2022) Earthquake risk embedded in property prices: Evidence from five Japanese cities. *Journal of the American Statistical Association* 117: 82–93. CrossRef

Jennings V, Reid C, Fuller C (2021) Green infrastructure can limit but not solve air pollution injustice. *Nature Communications* 12: 4681. CrossRef

Jumadi J, Malleson N, Carver S, Quincey D (2020) Estimating spatio-temporal risks from volcanic eruptions using an agent-based model. *Journal of Artificial Societies and Social Simulation* 23. CrossRef

Kovacs K, West G, Nowak D, Haight R (2022) Tree cover and property values in the United States: A national meta-analysis. *Ecological Economics* 197: 107424. CrossRef

Kurnio H, Fekete A, Naz F, Norf C, Jüpner R (2021) Resilience learning and indigenous knowledge of earthquake risk in Indonesia. *International Journal of Disaster Risk Reduction* 62: 102423. CrossRef

Lee S, Dodge J, Chen G (2022) The cost of social vulnerability: An integrative conceptual framework and model for assessing financial risks in natural disaster management. *Natural Hazards* 114: 691–712. CrossRef

Luck S, Santos J (2023) The valuation of collateral in bank lending. *Journal of Financial and Quantitative Analysis* 59: 2038–2067. CrossRef

MAGMA Indonesia (2025) Api, Laporan Harian Gunung. Retrieved December 24, 2025, from https://magma.esdm.go.id/v1/gunung-api/laporan-harian/2024-08-23

Mani L, Tzachor A, Cole P (2021) Global catastrophic risk from lower magnitude volcanic eruptions. *Nature Communications* 12: 4756–. CrossRef

Martinez-Villegas M, Reniva P, Sanico L, Loza A, Seda R, Doloiras D, Pidlaoan A (2022) Perspectives on the 12 January 2020 Taal volcano eruption: An analysis of residents’ narrative accounts. *Frontiers in Earth Science* 10: 923224. CrossRef

Maurin V (2022) Asset scarcity and collateral rehypothecation. *Journal of Financial Intermediation* 52: 100992. CrossRef

Medeiros J, Carmo R, Pimentel A, Vieira J, Queiroz G (2021) Assessing the impact of explosive eruptions of Fogo volcano (Saõ Miguel, Azores) on the tourism economy. *Natural Hazards and Earth System Sciences* 21: 417–437. CrossRef

Miller R, Pinter N (2022) Flood risk and residential real-estate prices: Evidence from three US counties. *Journal of Flood Risk Management* 15: 12774. CrossRef

Mistry H, Lombardi D (2022) Pricing risk-based catastrophe bonds for earthquakes at an urban scale. *Scientific Reports* 12: 9729–. CrossRef

Nesticò A, Maselli G, Ghisellini P, Ulgiati S (2023) A dual probabilistic discounting approach to assess economic and environmental impacts. *Environmental and Resource Economics* 85: 239–265. CrossRef

Oke O, Dougherty E, Rasmussen K, Morrison R, Carter E (2023) Spatial distribution of socio-demographic and housing-based factors in relation to flash and slow-rise flooding hazards in the US. *Environmental Research Letters* 18: 054016. CrossRef

Pallister J, Wessels R, Griswold J, McCausland W, Kartadinata N, Gunawan H, Budianto A, Primulyana S (2019) Monitoring, forecasting collapse events, and mapping pyroclastic deposits at Sinabung volcano with satellite imagery. *Journal of Volcanology and Geothermal Research* 382: 149–163. CrossRef

Pandey A (2019) High bids and low recovery: A possible case for non-performing loan auctions in India. *IIMB Management Review* 31: 250–257. CrossRef

Rey S (2025) Spatial inequality. *REGION* 12: 19–45. CrossRef

Rindrasih E, Ratminto E, C. K, Silviani D (2024) Expert perspectives on disaster risk reduction strategies in the tourist area of Borobudur-Yogyakarta-Prambanan in Indonesia. *Progress in Disaster Science* 24: 100379. CrossRef

Rosen S (1974) Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82: 34–55. CrossRef

Saputra A, Gomez C, Delikostidis I, Zawar-Reza P, Hadmoko D, Sartohadi J (2021) Preliminary identification of earthquake triggered multi-hazard and risk in Pleret subdistrict (Yogyakarta, Indonesia. *Geo-Spatial Information Science* 24: 256–278. CrossRef

Sayudi S, Nurnaning A, Juliani D, Muzani M (2010) Volcanic hazard map of Merapi Volcano. Pusat Vulkanologi dan Mitigasi Bencana Geologi (PVMBG). Retrieved April 29, 2026, from https://volcanichazardmaps.org/map/?id=918

Shi D, Zhang Y, Liang Z (2022) The effects of park and sea landscape on property value in a tourist city. *Frontiers in Environmental Science* 10: 967094. CrossRef

Shi P, Ye T, Wang Y, Zhou T, Xu W, Du J, Wang J, Li N, Huang C, Liu L, Chen B, Su Y, Fang W, Wang M, Hu X, Wu J, He C, Zhang Q, Ye Q, Okada N (2020) Disaster risk science: A geographical perspective and a research framework. *International Journal of Disaster Risk Science* 11: 426–440. CrossRef

Shi S, Naylor M (2023) Perceived earthquake risk in housing purchases. *Journal of Housing and the Built Environment* 38: 1761–1787. CrossRef

Skočir M, Lončarski I (2024) On the importance of asset pricing factors in the relative valuation. *Research in International Business and Finance* 70: 102366. CrossRef

Sousa R, Silva V, Rodrigues H (2022) The importance of indirect losses in the seismic risk assessment of industrial buildings – An application to precast RC buildings in Portugal. *International Journal of Disaster Risk Reduction* 74: 102949. CrossRef

Takeda K, Inaba K (2022) The damage and reconstruction of the Kumamoto earthquake: An analysis on the impact of changes in expenditures with multi-regional input–output table for Kumamoto prefecture. *Journal of Economic Structures* 11: 20. CrossRef

Taylor Z, Aalbers M (2022) Climate gentrification: Risk, rent, and restructuring in Greater Miami. *Annals of the American Association of Geographers* 112: 1685–1701. CrossRef

Terry J, Goff J, Winspear N, Bongolan V, Fisher S (2022) Tonga volcanic eruption and tsunami, January 2022: Globally the most significant opportunity to observe an explosive and tsunamigenic submarine eruption since AD 1883 Krakatau. *Geoscience Letters* 9: 24. CrossRef

Thamarapani D, Rockmore M (2022) The stability and evolution of risk attitudes and time preferences after a disaster. *International Journal of Disaster Risk Reduction* 70: 102791. CrossRef

Thompson J, Wilby R, Hillier J, Connell R, Saville G (2023) Climate gentrification: Valuing perceived climate risks in property prices. *Annals of the American Association of Geographers* 113: 1092–1111. CrossRef

Uesugi I, Miyakawa D, Hosono K, Ono A, Uchida H (2024) The collateral channel versus the bank lending channel: Evidence from a massive earthquake. *Journal of Banking and Finance* 79: 1–55. CrossRef

Wardani Y (2020) Gugatan perbuatan melawan hukum terhadap proses lelang agunan debitur macet oleh bank. *ACTA DIURNAL Jurnal Ilmu Hukum Kenotariatan* 4: 96–113. CrossRef

Xu W, Wu J, Gao M (2023) Seismic hazard analysis of China's mainland based on a new seismicity model. *International Journal of Disaster Risk Science* 14: 280–297. CrossRef

You T, Tesfamariam S (2024) Spatial correlation in building seismic performance for regional resilience assessment. *Resilient Cities and Structures* 3: 57–65. CrossRef

Zhang L, You S, Zhang M, Zhang S, Yi S, Zhou B (2022) The effects of urbanization on air pollution based on a spatial perspective: Evidence from China. *Frontiers in Environmental Science* 10: 1058009. CrossRef

Zhang Z, Ibikunle G (2023) The market quality effects of sub-second frequent batch auctions: Evidence from dark trading restrictions. *International Review of Financial Analysis* 89: 102737. CrossRef

Zheng W, Wang M (2024) Three-dimensional land-use configuration and property prices: A spatially filtered multi-level modelling perspective. *Environment and Planning B: Urban Analytics and City Science* 51: 438–455. CrossRef

Zhou K, Liu B, Fan J (2020) Post-earthquake economic resilience and recovery efficiency in the border areas of the Tibetan plateau: A case study of areas affected by the Wenchuan ms 8.0 earthquake in Sichuan, China in 2008. *Journal of Geographical Sciences* 30: 1363–1381. CrossRef

© 2026 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Milan, Italy. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

REGION
The Journal of ERSA
Powered by WU

Volume 13, Number 1, 2026, 115–152
DOI: 10.18335/region.v13i1.603
journal homepage: region.ersa.org
ISSN: 2409-5370

Integration, ethnic concentration and migrant self-employment in EU countries

Carla Massidda[1], Romano Piras[1], Silvio Tunis[1]

[1] University of Cagliari, Cagliari, Italy

Received: 15 May 2025/Accepted: 15 April 2026

Abstract. This study contributes to a deeper understanding of the complex relationship between migration, integration, and self-employment across European Union countries. Migration presents multifaceted challenges for host societies, particularly in terms of social cohesion and inclusive development. The Action Plan on Integration and Inclusion (2021-2027) emphasises the importance of migrant entrepreneurship in enhancing livelihoods and promoting sustained economic growth. However, empirical evidence on how integration shapes migrant entrepreneurship remains scarce, especially concerning the role of ethnic diversity. This research aims to fill this gap by examining the link between self-employment, integration, and ethnic concentration, distinguishing between migrants originating from European Union and non-European Union countries and further disaggregating the analysis by gender. It also investigates whether ethnic concentration mediates the effects of integration on migrant self-employment. Findings suggest that the relationship between integration and self-employment is multifaceted, with ethnic concentration having a significant influence and sometimes reversing this correlation.

Key words: ethnic concentration, migrant integration, migrant self-employment, Zaragoza indicators

1 Introduction

Migration has become a significant and complex phenomenon in European societies, presenting substantial challenges for policymakers tasked with fostering social cohesion and building inclusive communities. As widely acknowledged in the literature, this complexity encompasses multiple interrelated dimensions. These include cultural norms, social networks, and language proficiency (Wang et al. 2018), all of which interact with the characteristics of the local context (Biagi et al. 2025). Such multidimensional barriers can significantly influence the extent and quality of migrants' integration into host societies. The European Union's Action Plan on Integration and Inclusion for 2021-2027, as part of the broader New Pact on Migration and Asylum, directly addresses these challenges and outlines the necessary steps for integrating migrants into host societies (European Commission 2020). A notable focus within this plan is the promotion of migrant entrepreneurship through targeted policies and programs (European Commission 2020, p. 12). The rationale behind this approach is that fostering migrant entrepreneurship via dedicated integration strategies can yield mutual benefits, enhancing migrants' quality of life while simultaneously driving economic growth in host countries. This initiative aligns with Europe's broader effort to establish a common set of integration indicators, known as the Zaragoza indicators, to facilitate international comparisons in the field of migrant integration.

However, a comprehensive understanding of how integration affects migrant entrepreneurship in Europe is far from being achieved due to the confluence of multiple factors and the scarcity of empirical studies that help distinguish among them. One of these factors concerns the ethnic diversity resulting from prolonged and heterogeneous immigration flows. As documented in the literature, the degree of ethnic diversity significantly affects migrants' decisions to turn to entrepreneurship (Awaworyi Churchill 2019) and, at the same time, influences their integration process into host societies. However, the interplay between integration and ethnic diversity in (dis)encouraging self-employment has not received considerable attention until now. To the best of our knowledge, while there is extensive literature on the relationship between ethnic diversity and entrepreneurship, empirical investigations on the relationship between integration and self-employment are scarce. Moreover, none of the existing works analyses the interplay between the integration process and ethnic diversity in pushing migrant entrepreneurship.

The present paper aims to contribute to the literature by focusing on migrant self-employment in European Union (EU) countries. Specifically, this study examines whether migrant integration and ethnic concentration influence migrant self-employment, and whether ethnic concentration acts as a mediator of the impact of integration. In addition, the analysis is conducted separately for migrants originating from other EU countries and migrants from countries outside the Union (NEU) and is further disaggregated by gender. This differentiation allows us to capture important sources of heterogeneity, as both country of origin and gender may shape access to resources, exposure to structural constraints, and the mechanisms through which integration and ethnic concentration affect self-employment outcomes.

To the best of our knowledge, we are the first to conduct this type of investigation from a macroeconomic perspective and to adopt the Zaragoza integration indicators for empirical analysis on this topic. Worth noting is that although our empirical analysis is conducted at the national level, the mechanisms we study (social and occupational integration, ethnic concentration, and sectoral specialisation) mirror regional and local dynamics. As widely acknowledged in the literature, national patterns are the aggregate outcome of processes that originate and evolve at subnational scales. Understanding these country-level relationships, therefore, helps to illuminate how regional disparities emerge and how territorial opportunity structures shape immigrant entrepreneurship across space. By framing our findings through this multiscalar lens, the paper connects directly with current regional development challenges.

The overall results confirm that the relationship between integration and entrepreneurship is highly complex, varying across different dimensions of integration. Furthermore, they reveal that ethnic concentration plays a significant role in shaping this relationship, potentially reversing the direction of the direct correlations.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature, first outlining the theoretical foundations and then summarising the main empirical evidence on migrant integration and the role of ethnic concentration in new business creation. Section 3 presents the methodological framework, describes the data and variables employed, and details the empirical strategy guiding the analysis. Section 4 discusses the results, distinguishing between migrants from EU and NEU countries, and includes a set of robustness checks. Section 5 offers a discussion of the key findings and their policy implications. Section 6 concludes the paper.

2 Literature review

2.1 Theoretical foundations of the relationship between migrant entrepreneurship, integration, and ethnic diversity

The relationship between integration and migrant entrepreneurship has been examined through different theoretical lenses. The Mixed Embeddedness Theory (Kloosterman et al. 1999) is often adopted as an overarching framework because it explicitly combines individual resources, institutional contexts, and market opportunity structures. In this perspective, other theories typically cited in the literature, such as Human Capital Theory (Becker 1962, Schultz 1961), Blocked Mobility Theory (Kim et al. 1989), and (Segmented) Assimilation (Gordon 1964), should not be seen as competing or independent frameworks. Rather, they correspond to the micro- and meso-level mechanisms that are embedded within Mixed Embeddedness. Human Capital Theory captures the role of individual endowments; Blocked Mobility accounts for structural constraints and labour-market segmentation; and (Segmented) Assimilation Theory introduces the dynamic evolution of social integration, institutional access, and opportunity over time. By situating these mechanisms within Mixed Embeddedness, existing studies identify two opposing dynamics linking integration and migrant entrepreneurship.

On one hand, migrants may be compelled to pursue entrepreneurship due to adverse conditions in host countries, such as high unemployment, low wages, political instability, corruption, and discrimination, factors often associated with inadequate integration. This pattern was particularly prevalent in the early 20th century when immigrant entrepreneurship largely consisted of necessity entrepreneurs who established small businesses to meet their financial needs (Borjas 1986). In contrast, integration into host countries may empower migrants to pursue entrepreneurship to capitalise on enhanced opportunities, such as higher incomes and improved living standards. In recent decades, some studies have contended that immigrant entrepreneurship is predominantly characterised by opportunity entrepreneurs who, due to their high skill levels, are adept at identifying and seizing business opportunities that native-born individuals may overlook (Dheer 2018).

Within this complex framework, another important factor often discussed in the literature is the role of ethnic-cultural diversity resulting from sustained and heterogeneous immigration flows. Ethnic-cultural diversity can foster entrepreneurship by blending diverse ideas, creativity, and innovation (Awaworyi Churchill 2019), while enclave living can create economic constraints, isolating immigrants and limiting access to external opportunities (Borjas 2000). From a different perspective, however, ethnic diversity can represent a threat, especially in developing countries, where homogeneous ethnic groups might compensate for a lack of social and physical capital and new innovative ideas (Appau et al. 2019). The central concept deriving from the Ethnic Enclave Theory (Portes, Jensen 1987) posits that residing and operating within homogeneous ethnic groups or enclaves can offer valuable resources via social networks and enhanced information on job prospects.

In relation to ethnic diversity, an additional dimension merits consideration. Ethnic diversity may indirectly affect migrant entrepreneurship by shaping migrants' integration and social cohesion within host societies (Putnam 2007). In social sciences, two contrasting theoretical views emerge regarding the impact of diversity on social cohesion (Putnam 2007). While the Contact Theory views diversity as a source of social cohesion, the Conflict Theory argues that it can foster distrust and in-group loyalty. Consequently, the link between integration and entrepreneurship may vary with levels of ethnic diversity, a relationship that remains insufficiently explored despite growing research on diversity and entrepreneurship (Yavuz, Bahadir 2022).

In summary, the theoretical framework addressing the relationship between integration and entrepreneurship, through the lens of ethnic diversity, remains complex and fragmented with no definitive consensus on whether and to what extent integration facilitates business creation (Brzozowski 2019, Massidda et al. 2024).

2.2 Empirical literature on the relationship between migrant entrepreneurship, integration, and ethnic diversity

When theoretical frameworks fail to provide a unified interpretation, empirical evidence typically helps to discriminate among competing explanations. However, in the context of the present study, empirical evidence offers limited guidance, as existing research on the topic remains scarce. To the best of our knowledge, only a few studies have jointly examined integration and ethnic diversity as independent determinants of migrant entrepreneurship. A recent systematic review by Massidda et al. (2024) analysed 1,922 articles published between 2002 and 2023. The identification phase began with a search in the WoS and Scopus databases using keywords related to migration (e.g., migrant, diaspora, ethnic), integration (e.g., integration, inclusion, assimilation), and entrepreneurial activity (e.g., firm, business, entrepreneurship, self-employment). Then, under the PRISMA protocol, the study isolates 33 contributions that adopted a quantitative approach. Of these, only 15 empirically investigated the effects of integration processes on migrant entrepreneurship in host countries, and 10 focused specifically on the role of ethnicity in immigrants' business formation decisions. A common feature among these contributions is their reliance on microdata. At the macro level, however, empirical research has been significantly constrained by the lack of consistent and longitudinal indicators of migrant integration, thereby limiting the scope for large-scale comparative analyses. The subsequent subsections examine the main findings of this pertinent literature. Table A.1 in Appendix A provides a concise overview of the studies considered.

2.2.1 Migrant integration and new business creations

In empirical research, direct measurements of migrants' integration in the destination country are not always available. Consequently, researchers often rely on indirect indicators to capture this complex phenomenon.

One common indirect measure of integration is the length of stay (Lofstrom 2002, Constant, Shachmurove 2006, Wang, Li 2007, Abada et al. 2014, Matricano, Sorrentino 2014, Andersson, Hammarstedt 2015, Bashko 2022, Sun, Fong 2022, Zhang et al. 2024). According to Mixed Embeddedness Theory (Kloosterman et al. 1999), longer residence helps migrants acquire the social norms and legal knowledge necessary for entrepreneurship. Sun, Fong (2022), Wang, Li (2007), Bashko (2022), and Zhang et al. (2024) confirm that extended stays increase the likelihood of entrepreneurship. However, other scholars identify non-linear patterns: Andersson, Hammarstedt (2015) and Lofstrom (2002) report an inverted U-shape, Constant, Shachmurove (2006) a U-shape, while Matricano, Sorrentino (2014) find no significant effect.

Language proficiency and, more generally, education serve as alternative measures of integration. Authors recall arguments, in particular, from Human Capital Theory to support the view that linguistic ability and educational skills are essential for immigrants' economic assimilation, fostering connections across ethnic groups, facilitating resource sharing, and augmenting their self-employment capacity. For instance, Zhang et al. (2024) and Abada et al. (2014) show that a lack of proficiency in one of Canada's official languages reduces self-employment rates. Similarly, Sun, Fong (2022) argue that language skills facilitate adaptation and access to business opportunities, a view supported by Bashko (2022) and Wang, Li (2007). Regarding education, Sun, Fong (2022) find that higher education increases both entrepreneurial success and self-employment, a result confirmed by Lofstrom (2002), Bashko (2022), Zhang et al. (2024), and Constant, Shachmurove (2006). Furthermore, Andersson, Hammarstedt (2015) find non-linear effects (secondary education positive, university negative), Abada et al. (2014) highlight intergenerational differences, Wang, Li (2007) show low-educated migrants enter self-employment out of necessity, and Matricano, Sorrentino (2014) detect no significant relationship.

Other indirect integration measures include unemployment, wages, and earnings differentials. Assimilation Theory and the Blocked Mobility Theory offer contrasting perspectives: labour market Assimilation Theory suggests barriers discourage self-employment, whereas Blocked Mobility Theory argues they may push individuals toward it. Abada et al. (2014) find group-specific unemployment positively affects fathers' self-employment, is insignificant for second-generation males, and negative for third-generation males, supporting Blocked Mobility mainly for fathers. Conversely, Andersson, Hammarstedt (2015) and Andersson (2021) report no significant effects. Regarding earnings differentials, Abada et al. (2014) find a stronger positive impact for sons than fathers, while Lofstrom (2002) finds no significant relationship.

Distinct from the other studies listed in Table A.1, Zhang et al. (2024) are the only authors to propose variables for directly measuring integration, including social interaction, social exclusion, and social identification, based on responses from the 2017 China Migrant Dynamic Survey. The authors contend that integration enhances networking and fosters a sense of identity, facilitating business operations within host societies.

2.2.2 Ethnic concentration and new business creation

Empirical research on ethnic concentration and entrepreneurship is largely grounded in ethnic enclave theory, which allows for both opportunity-enhancing and constraining mechanisms.

Most studies report a positive association. Lofstrom (2002), using U.S. data (1980–1990), finds that a higher share of co-nationals in metropolitan areas increases self-employment. Andersson, Hammarstedt (2015) show that ethnic enclaves in Sweden raise self-employment among Middle Eastern immigrants, although their ethnic network index-based on concentration weighted by average self-employment—has a negative effect. Andersson (2021) finds that co-ethnic self-employment and education increase refugees' probability of entering self-employment within five years, while co-ethnic unemployment is insignificant. Bashko (2022) reports that larger co-ethnic networks—measured by the number of co-ethnic contacts—enhance entrepreneurial opportunities among Chinese and Vietnamese immigrants in Warsaw. Zhang et al. (2024) show that migrant diversity promotes entrepreneurship in China, whereas ethnic diversity reduces it, highlighting the mediating role of social integration.

Other contributions find no significant effects. Abada et al. (2014) report that ethnic group population share does not affect immigrants' self-employment in Canada. Matricano, Sorrentino (2014) find ethnic economy variables insignificant for Ukrainian entrepreneurs in Italy, although living outside the enclave and being under 40 increase the likelihood of founding firms within it. Sun, Fong (2022) also detect no effect of co-ethnic population size in Hong Kong, stressing differences across ethnic groups, cohorts, and genders. Constant, Shachmurove (2006) and Wang, Li (2007), while not directly modelling ethnic concentration, highlight the role of demographic and local contextual factors in shaping immigrant self-employment.

2.3 Summary of the theoretical and empirical evidence

Overall, theoretical contributions do not exclude the possibility that both integration and ethnic diversity may exert either positive or negative effects on migrant entrepreneurship, depending on contextual factors. Moreover, the literature offers limited insights into the potential interaction between these two phenomena. From an empirical perspective, the evidence largely confirms this theoretical ambiguity. Based largely on microdata, studies examining the direct effects of integration and ethnic diversity on migrant entrepreneurship report mixed or inconsistent results. To date, no research has systematically investigated their interaction effects. Moreover, the lack of specific and comparable indicators of migrant integration across contexts limits the robustness and comparability of findings.

3 Methodology

3.1 Aims and research hypothesis

This study seeks to advance the literature on the relationship between migrant self-employment and integration by focusing on the context of EU countries. These countries provide a compelling scenario for investigation due to their increasing migratory flows, high rates of migrant entrepreneurship (Chodavadia et al. 2025), and the integration challenges that have attracted growing institutional attention. Worldwide, the latest available estimates report that in 2024, more than 304 million people lived in a country other than their country of birth. In Europe, the number of migrants is about 94 million. The corresponding figures in 1990 were 153 million and 50 million (McAuliffe, Oucho 2024). Another reason that makes the EU an interesting case study is the development of a harmonised set of country-level indicators measuring migrant integration. After the

Zaragoza Declaration (European Ministerial Conference 2010), the EU regularly monitors four distinct dimensions of the integration process: Education, Employment, Social Inclusion, and Active Citizenship (Eurostat 2020). Each dimension is operationalised through a set of multiple variables collected across different time spans, allowing for a dynamic assessment of migrant integration.

In more detail, this study aims to investigate three main research hypotheses. The first posits that migrant integration has a significant influence on migrant self-employment (H1). To investigate this hypothesis, we operationalise three of the four Zaragoza indicators, namely Education, Employment, and Social Inclusion (the fourth dimension, Active Citizenship, is excluded from our analysis due to data scarcity). With respect to Education, drawing on Human Capital Theory and the prevailing empirical evidence discussed above (see inter alia Sun, Fong 2022), we hypothesise that higher education integration increases the likelihood that migrants engage in self-employment (H1a).[1] In relation to Employment, building on Blocked Mobility Theory and prevailing empirical evidence (see inter alia Abada et al. 2014), we propose an inverse mechanism whereby greater integration into the labour market may lead to a decrease in migrant self-employment (H1b). Finally, regarding Social Inclusion, building on (Segmented) Assimilation Theory and previous empirical findings (Zhang et al. 2024), we assert a positive relationship between social inclusion and self-employment (H1c).

The second hypothesis (H2) posits that ethnic concentration influences migrant self-employment in a manner contingent upon the dominance of two opposing mechanisms. Drawing on Ethnic Enclave Theory and the prevailing empirical evidence, ethnic concentration can raise the likelihood of being self-employed (H2a). Conversely, concentration can discourage migrant entrepreneurship (H2b). Finally, the third hypothesis (H3) examines the potential moderating role of ethnic concentration in the relationship between integration and self-employment. As previously discussed, the relative strength of the Contact Theory versus the conflict hypothesis may shape how integration interacts with ethnic concentration. Given the lack of empirical evidence on this issue, we allow for both possibilities: a positive interaction (H3a) and its alternative (H3b).

Figure 1 summarises the hypotheses that we will empirically test, along with the theoretical frameworks underpinning these hypotheses. The thick solid lines illustrate the direct relationships between integration and ethnic diversity (H1 and H2) with self-employment, while the dashed line denotes the mediating role of ethnic diversity (H3).

In line with the conceptual framework presented in Figure 1, our research hypotheses are tested separately for migrants settled in EU countries and originating from other EU Member States, and those originating from NEU countries and settled in an EU country. The separate investigation of EU and NEU migrants appears especially promising in light of the marked structural differences between these two sub-samples. In this regard, it is crucial to note that NEU migrants are likely to be more diverse in background than EU migrants. In addition, migrants holding EU citizenship possess a set of legal entitlements and institutional protections that differ significantly from those granted to NEU migrants. These disparities encompass areas such as freedom of movement, access to the labour market, and eligibility for social services, all of which substantially shape their experiences within host societies. Consequently, such structural differences may contribute to, and indeed help elucidate, the divergent integration trajectories that we aim to analyse in this study.

For both EU and NEU sub-samples, we examine gender-specific patterns. A gender perspective deepens the analysis, as gender interacts with migration status to shape employment access, entrepreneurial opportunities, and social integration, enabling a more nuanced understanding of heterogeneous economic outcomes. Most migration studies overlook this dimension, implicitly treating migrants as homogeneous. Yet gender-disaggregated analysis is more appropriate. The common assumption that women migrate primarily as accompanying spouses is challenged by recent evidence showing diverse motivations and distinct integration trajectories (Yilmaz, Solano 2024).

[1] This hypothesis is further supported by empirical evidence on non-migrants (Van Der Sluis et al. 2008).

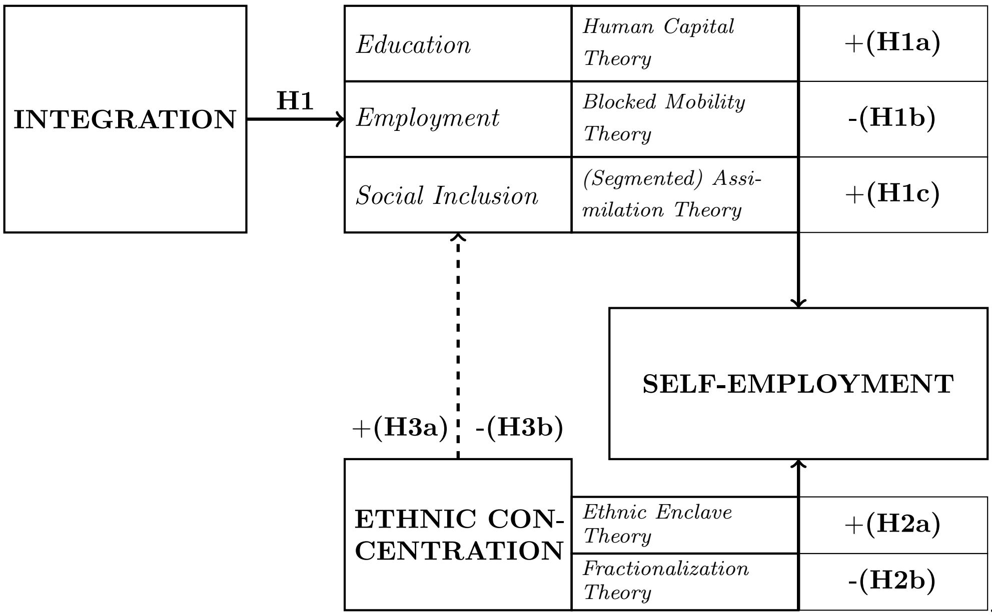


Figure 1: Conceptual framework

3.2 Variables description

3.2.1 Data and descriptive statistics

The primary data source for this study is Eurostat.[2] In the dedicated section on migrant integration, Eurostat reports the country/region of birth and, for each Member State, distinguishes between migrants coming from another EU country (excluding the reporting country) and migrants coming from a NEU country. Unfortunately, data availability is not fully homogeneous across country–year-gender observations. In some countries, data is entirely missing; in others, it is reported only for the total migrant population (aggregating male and female migrants), while disaggregated figures are available only for specific gender–origin combinations. To maximise country coverage and preserve the cross-sectional variation inherent in the dataset, we retain all countries for which relevant information is available for the specific sub-sample under consideration. Each specification is therefore estimated on the maximum set of country–year observations available for the corresponding origin–gender sub-sample. The use of unbalanced country–year panels driven by sub-sample-specific data availability is a common feature of macro-level analyses based on harmonised international databases, where coverage reflects reporting constraints rather than selective sample construction.

As a consequence, the time period and the list of countries included vary depending on the estimated model. The longest period spans from 2002 to 2021; the shortest, from 2010 to 2021. The number of countries ranges from 12 to 19. This lack of data stems from several factors, including the fact that 13 new Member States from Central and Eastern Europe joined the Union between 2004 and 2013[3], increasing the likelihood of missing observations in earlier years. Notice that the difference in country coverage consistently runs in one direction: the NEU sub-sample includes a few more countries than the corresponding EU sub-sample.

Table A.3 in the Appendix provides the descriptive statistics for the two sub-samples of self-employed immigrants in the EU, differentiating between individuals from other EU countries and those from outside the EU, with a further breakdown by gender for both female and male migrants. For each variable, descriptive statistics are computed based on the available data range. In both the EU and NEU samples, males exhibit a higher average percentage of self-employment compared to females. Furthermore, the

[2] See Table A.2 in the Appendix for details on variable definitions and sources.

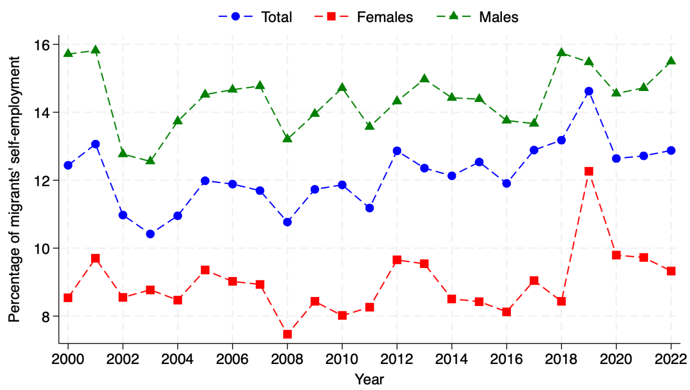
[3] In 2004, Cyprus, Estonia, Latvia, Lithuania, Malta, Poland, the Czech Republic, Slovakia, Slovenia and Hungary joined the EU, followed by Bulgaria and Romania in 2007 and by Croatia in 2013.

disparities between males and females are also evident in the key integration indicators associated with the primary independent variables. Overall, Table A.3 demonstrates a wide range of minimum and maximum values, indicating significant heterogeneity among countries.

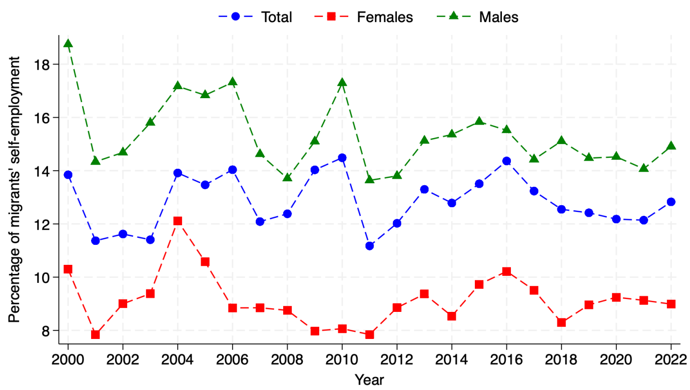
3.2.2 Dependent variable

Our dependent variable is the percentage of migrants self-employed over total employment (SE) in the age class 15-64. Figure 2 reports the percentage of migrants' self-employment as the average of migrants who are settled within an EU country. The top panel of the figure refers to migrants coming from an EU country, while the bottom panel refers to migrants coming from a NEU country.[4]

Within the EU migrants sub-sample, on average, the percentage of self-employed follows a fluctuating yet mildly increasing trend, with males consistently displaying higher rates than females. In the case of NEU migrants, the pattern of self-employment is confirmed: male migrants systematically exhibit a higher self-employment percentage than their female counterparts.



(a) EU migrants



(b) NEU migrants

Figure 2: Percentage of migrants' self-employment

3.2.3 Integration indicators

As previously anticipated, this analysis considers three out of the four dimensions of the Zaragoza indicators, namely Education, Employment, and Social Inclusion, which

[4]In Appendix A, Figure A.1 presents six informative maps showing the percentage of self-employed migrants (our dependent variable) across EU countries, differentiated by origin (EU and NEU) and gender. These maps correspond to the countries included in 2021 for which data were available.

are operationalised through multiple variables. We selected one representative variable for each dimension based on the longest time-span availability: for Education, we consider the intermediate level of education attainment ($\mathrm{EDUCATION}_{it}$) corresponding to ISCED levels 3 and 4, upper secondary and post-secondary non-tertiary education; for Employment, we consider the unemployment rate (UNRATE_{it}); for Social Inclusion we consider the mean equivalised net family income expressed in purchasing power standard (INCOME_{it}). Then, to convert these variables into proper integration indicators, we calculate the difference in the outcome of the target group (migrants from EU and NEU countries) against that of the reporting country (REP). The educational gap, $\mathrm{EDU\_GAP2}_{it}^{\mathrm{(N)EU}}$, is computed as the percentage-point difference between the percentage of migrants and the percentage of natives holding at least an upper-secondary degree in each country–year for the population aged 15–64 years. Since both variables are proportions bounded between 0 and 100, this yields a standardised and directly comparable measure of educational disparities across countries and over time. Analogously, the unemployment rate gap, $\mathrm{UNR\_GAP}_{it}^{\mathrm{(N)EU}}$, is computed as the percentage-point difference between the unemployment rate of migrants and the unemployment rate of natives for the population aged 15–64 years. Therefore, for both education and unemployment, we are using gaps in percentage-point terms, which is the convention adopted by the OECD and Eurostat in labour market monitoring. Regarding the income gap, $\mathrm{INC\_GAP}_{it}^{\mathrm{(N)EU}}$, we consider the mean equivalised net family income expressed in purchasing power standard. This is computed by Eurostat as the household's net disposable income, adjusted for household size using the modified OECD equivalence scale, and converted into purchasing power standards to ensure cross-country comparability of living standards. Given this definition, the absolute difference provides a transparent and policy-relevant measure of the magnitude of the economic disadvantage, commonly used in international monitoring (e.g. OECD, Eurostat).

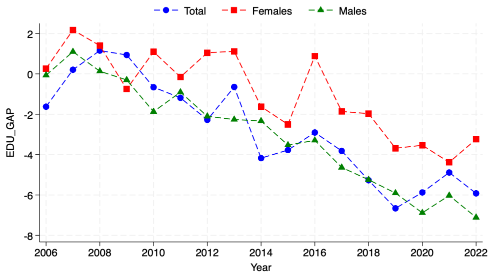
Specifically, the three indicators are calculated as follows:

$$\mathrm{EDU\_GAP2}_{it}^{\mathrm{(N)EU}} = \mathrm{EDUCATION}_{it}^{\mathrm{(N)EU}} - \mathrm{EDUCATION}_{it}^{\mathrm{REP}} \tag{1}$$
$$\mathrm{UNR\_GAP}_{it}^{\mathrm{(N)EU}} = \mathrm{UNRATE}_{it}^{\mathrm{(N)EU}} - \mathrm{UNRATE}_{it}^{\mathrm{REP}} \tag{2}$$
$$\mathrm{INC\_GAP}_{it}^{\mathrm{(N)EU}} = \mathrm{INCOME}_{it}^{\mathrm{(N)EU}} - \mathrm{INCOME}_{it}^{\mathrm{REP}} \tag{3}$$

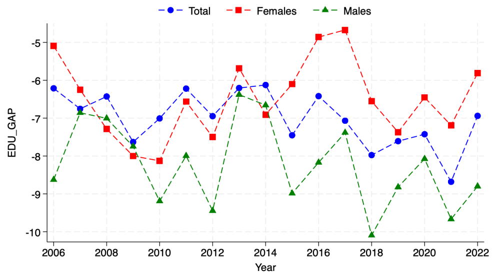
Given this definition, a *negative* gap for $\mathrm{EDU\_GAP2}_{it}^{\mathrm{(N)EU}}$ and $\mathrm{INC\_GAP}_{it}^{\mathrm{(N)EU}}$ and a *positive* gap for $\mathrm{UNR\_GAP}_{it}^{\mathrm{(N)EU}}$ indicate that EU and NEU migrants perform worse than natives. Figure 3 illustrates these gaps for EU and NEU migrants, as well as for total, female, and male migrants separately. Notice that since the data set is unbalanced, yearly information by country and gender is not fully covered. For some years and in different countries, only data for the total migrants are available (i.e., gendered data is not provided); in other circumstances, Eurostat provides data only for males or females. These are the reasons why the trendline for the total migrants does not always fall between the trendlines for females and males.

For $\mathrm{EDU\_GAP2}_{it}^{\mathrm{(N)EU}}$, panel (a) shows a general declining trend for EU migrants, which implies increasing educational gaps (i.e. lower integration), more pronounced for males. Panel (b) shows a less pronounced declining trend for NEU migrants. Recalling that citizens of the new Member Countries gained access to the EU labour market after a certain period, the decline in $\mathrm{EDU\_GAP2}_{it}^{\mathrm{(N)EU}}$ for EU migrants might be partially explained by less skilled workers coming from these new Member States.

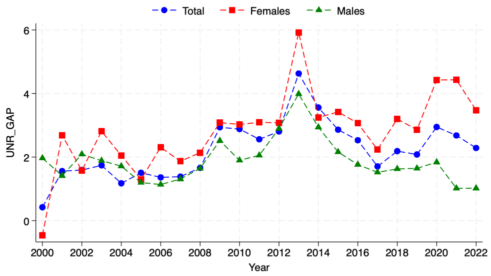
Considering $\mathrm{UNR\_GAP}_{it}^{\mathrm{(N)EU}}$, panel (c) shows relatively stable levels for EU migrants except for a peak in 2014 for females. Such a pattern does not occur for NEU migrants, as shown in panel (d), where the gaps slightly decrease for all sub-samples while remaining significantly higher than those of EU migrants. Finally, regarding $\mathrm{INC\_GAP}_{it}^{\mathrm{(N)EU}}$, the gap is substantially wider for NEU migrants, approximately 50%, than for their EU counterparts, but no clear trend emerges for the two sub-samples (panels (e) and (f), respectively).



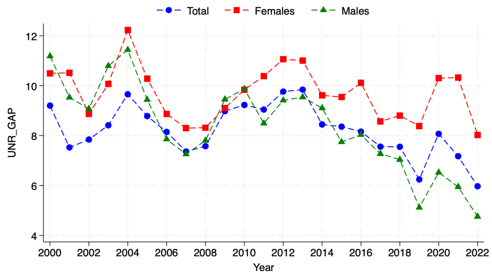
(a) EDU_GAP, EU migrants



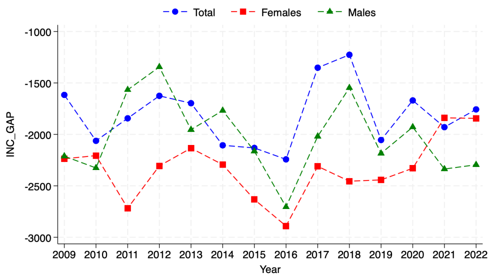
(b) EDU_GAP, NEU migrants



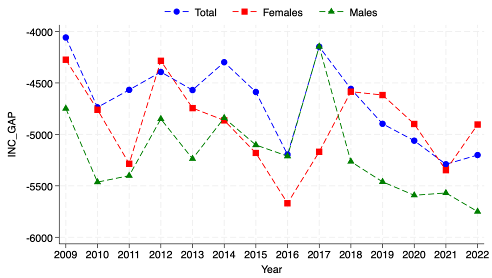
(c) UNR_GAP, EU migrants



(d) UNR_GAP, NEU migrants



(e) INC_GAP, EU migrants



(f) INC_GAP, NEU migrants

Figure 3: Integration indicators

Crucially, no comparable studies investigate the effects of these gaps on self-employment. However, the conceptual framework presented in Figure 1 and the empirical evidence summarised in Table A.1 provide some important suggestions on expected signs.

As reported in Figure 1, regarding $\text{EDU\_GAP2}_{it}^{(\text{N})\text{EU}}$, we expect a positive relationship between higher integration and self-employment (H1a). Being this gap negative (see Figure 3, panels 3a, 3b), its increase means that *algebraically*, it becomes higher going from, say, -5 to -4. However, in *absolute terms*, such a variation means a reduction of the gap, which implies higher integration among migrants. Therefore, to confirm H1a, we expect a positive estimated sign, indicating that migrants whose educational attainments are closer to those of their native peers are more likely to pursue self-employment. Concerning the role of education, we recognise that focusing exclusively on a single educational level entails a significant loss of information. Migrants' educational attainments are highly heterogeneous (see Table A.3 in the Appendix), and their motivations for pursuing self-employment may vary substantially depending on whether they are more or less skilled than natives. We will conduct additional analyses across different educational levels to address this limitation and present these additional results in Appendix D.

Regarding the unemployment gap, $\text{UNR\_GAP}_{it}^{(\text{N})\text{EU}}$, we expect an inverse relationship between integration and self-employment, which corresponds to H1b in Figure 1. Being this gap positive (see Figure 3, panels 3c, 3d), its increase, both *algebraically* and in *absolute terms*, implies lower migrants' integration. Therefore, to confirm an inverse relationship between labour market integration and self-employment (H1b), we expect a positive estimated coefficient.

Finally, about the income gap ($\text{INC\_GAP}_{it}^{\text{(N)EU}}$), we expect a positive relationship between higher integration and self-employment (H1c). Similarly to $\text{EDU\_GAP2}_{it}^{\text{(N)EU}}$, being this gap negative (see Figure 3, panels 3e, 3f), we expect a positive estimated sign to confirm H1c, which suggests that as the income gap increases (in absolute terms, indicates greater integration), migrants are more likely to engage in self-employment within the labour market.

3.2.4 The Ethnic Concentration Index

The fourth main explanatory variable is the Ethnic Concentration Index, measured by a Herfindahl–Hirschman-type index:

$$\text{ECI}_{it} = \sum_{j=1}^{N_j} \left(\frac{\text{mig}_{jit}}{\sum_{j=1}^{N_j} \text{mig}_{jit}} * 100 \right)^2 \tag{4}$$

where mig_{jit} is the number of migrants from country j residing in an EU country i at time t, $\sum_{j=1}^{N_j} \text{mig}_{jit}$ is the total number of migrants from N_j different nationalities residing in an EU country i at time t.[5] Hence, $\frac{\text{mig}_{jit}}{\sum_{j=1}^{N_j} \text{mig}_{jit}}$ is the share of migrants from country j in the overall migrant population of EU host country i in year t. It is important to note that ECI_{it} is always positive and varies between 0 and 10000; 0 being the minimum level of ethnic concentration (i.e. the maximum level of ethnic diversity), while 10000 represents the limit case of full concentration: a single foreign nationality j residing in the host country i. As previously discussed, the expected sign of ECI_{it} in relation to the percentage of self-employment among migrants remains ambiguous at a theoretical level. A positive coefficient would suggest that countries with a high concentration of ethnic groups provide greater motivation and a more conducive environment for migrant self-employment (H2a). This positive effect is plausible if homogeneous ethnic groups facilitate self-employment by compensating for deficiencies in capital and innovative ideas (Appau et al. 2019). Conversely, a negative coefficient would imply that ethnic concentration may impede the generation of new ideas (Awaworyi Churchill 2019), whereas ethnic diversity may enhance migrants' self-employment by fostering creativity and innovation (H2b). Consequently, the influence of ECI_{it} on entrepreneurship is contingent upon the prevailing mechanisms in host countries.

To validate the robustness of our findings, we construct two alternative measures of the Ethnic Concentration Index: one (ECI1) based on Eurostat migration stocks, and the other (ECI2) based on the bilateral series estimated by Standaert, Rayp (2022). While the two sources differ in data construction, they ultimately measure the same underlying phenomenon. Eurostat draws on official administrative records, whereas Standaert and Rayp fill data gaps by reconstructing missing observations using a state-space model. Despite methodological differences, both datasets capture the same latent construct.

Figure 4 illustrates both indices. With respect to ECI1, the index exhibits significant fluctuations, characterised by a notable increase in 2004 and the lowest values recorded between 2008 and 2011. The overall trend suggests periods of heightened ethnic concentration accompanied by reduced diversity, followed by intervals of markedly low ethnic concentration and increased diversity. It is essential to note, however, that, similar to the integration variables, the Eurostat data used for ECI1 contain missing values for certain years and/or countries, which may compromise its accuracy. In contrast, and given the absence of missing data (by construction), ECI2 appears less volatile, exhibiting a generally slight declining trend.

[5] Notice that the number of different nationalities residing in each host country j varies.

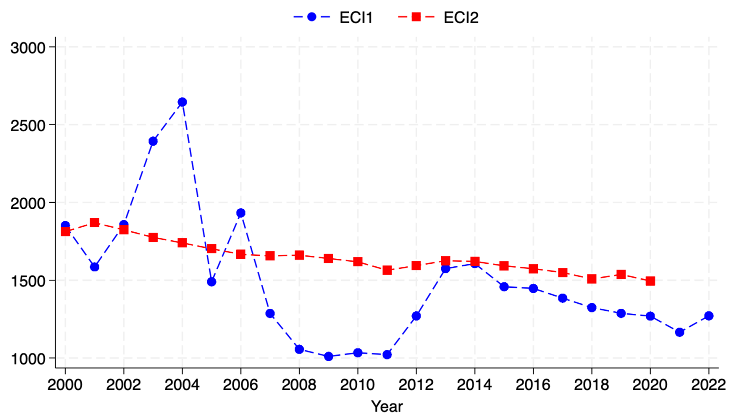


Figure 4: Ethnic Concentration Indexes

3.2.5 Control variables

The control variables incorporated in the analysis include the total native population (POP), real per capita gross domestic product (GDPPC), average full-time adjusted salary per employee (WAGE), domestic credit to the private sector as a percentage of GDP (CREDIT), and a regulatory quality index (RQI). The theoretical underpinnings behind the selection of the control variables are explained as follows. POP is included as an overall measure of the country's size. Intuitively, more populated countries are expected to boost general entrepreneurial activities and self-employment (Boudreaux 2020). Real per capita GDP (GDPPC) is included to account for the country's economic development. The relationship between economic development and self-employment is not established a priori. On the one hand, higher per capita GDP indicates a more developed economy with ample formal employment opportunities, potentially reducing the need for self-employment. On the other hand, it has been claimed (Wennekers et al. 2010) that in more developed economies, the desire for autonomy may render self-employment more appealing than wage employment. The average full-time adjusted salary per employee (WAGE) measures the opportunity cost of being self-employed; therefore, we expect a negative sign. CREDIT accounts for the ability to access financial resources, such as loans and working capital. Hence, we expect a positive sign. However, it is well known (Kariv, Coleman 2015) that migrants face challenges in obtaining creditworthiness from the banking system, and greater credit availability at the macroeconomic level does not necessarily imply that migrants could benefit from it. Lastly, RQI captures perceptions regarding the government's capacity to formulate and implement policies and regulations that facilitate and promote private-sector development (Kaufmann et al. 2010). It is posited that moderate levels of regulatory quality maximise entrepreneurial activity. Conversely, excessive or insufficient regulation can impede entrepreneurial initiatives (Polemis, Stengos 2020). Additionally, considering that well-regulated economies may provide a stable environment for formal employment, potentially reducing reliance on self-employment, we anticipate that RQI may be associated with a lower percentage of self-employment.

3.3 Empirical strategy

As previously anticipated, we estimate separate equations for each integration indicator, distinguishing between EU and NEU migrants. Additionally, separate estimations are performed for the total sample, as well as for female and male migrant sub-samples. In total, nine equations are estimated for each sub-sample of EU and NEU migrants. This methodology allows us to identify potentially varied behaviours that may lead to heterogeneous integration patterns in EU countries, depending on the migrants' country of origin (EU or NEU) and gender (total, females, and males).

To summarise the estimation of 18 models in a parsimonious style, we rely on three equations (one for each integration indicator) and use superscripts and subscripts to distinguish migrant origin (EU vs. NEU) and gender (total, female, male):

$$\begin{aligned}
\mathrm{SE}_{i,t}^{g,o} &= \beta_1^{\mathrm{EDU},g,o}\mathrm{ECI}_{i,t} + \beta_2^{\mathrm{EDU},g,o}\mathrm{EDU_GAP2}_{i,t}^{g,o} + \\
&\quad \beta_3^{\mathrm{EDU},g,o}\left(\mathrm{EDU_GAP2}_{i,t}^{g,o} * \mathrm{ECI}_{i,t}\right) + \boldsymbol{z}'_{i,t}\boldsymbol{\delta}^{\mathrm{EDU},g,o} + \\
&\quad \mu_i^{\mathrm{EDU},g,o} + \tau_t^{\mathrm{EDU},g,o} + \varepsilon_{i,t}^{\mathrm{EDU},g,o} \qquad (5)\\
\mathrm{SE}_{i,t}^{g,o} &= \beta_1^{\mathrm{UNR},g,o}\mathrm{ECI}_{i,t} + \beta_2^{\mathrm{UNR},g,o}\mathrm{UNR_GAP}_{i,t}^{g,o} + \\
&\quad \beta_3^{\mathrm{UNR},g,o}\left(\mathrm{UNR_GAP}_{i,t}^{g,o} * \mathrm{ECI}_{i,t}\right) + \boldsymbol{z}'_{i,t}\boldsymbol{\delta}^{\mathrm{UNR},g,o} + \\
&\quad \mu_i^{\mathrm{UNR},g,o} + \tau_t^{\mathrm{UNR},g,o} + \varepsilon_{i,t}^{\mathrm{UNR},g,o} \qquad (6)\\
\mathrm{SE}_{i,t}^{g,o} &= \beta_1^{\mathrm{INC},g,o}\mathrm{ECI}_{i,t} + \beta_2^{\mathrm{INC},g,o}\mathrm{INC_GAP}_{i,t}^{g,o} + \\
&\quad \beta_3^{\mathrm{INC},g,o}\left(\mathrm{INC_GAP}_{i,t}^{g,o} * \mathrm{ECI}_{i,t}\right) + \boldsymbol{z}'_{i,t}\boldsymbol{\delta}^{\mathrm{INC},g,o} + \\
&\quad \mu_i^{\mathrm{INC},g,o} + \tau_t^{\mathrm{INC},g,o} + \varepsilon_{i,t}^{\mathrm{INC},g,o} \qquad (7)
\end{aligned}$$

where the dependent variable $\mathrm{SE}_{i,t}^{g,o}$ is the percentage of migrants' self-employed on total migrants' employment of gender g (total, females, males), origin o (EU, NEU) in country i and year t and it is related to the ethnic concentration index (ECI_{it}), the three integration indicators ($\mathrm{EDU\_GAP2}_{i,t}^{g,o}$, $\mathrm{UNR\_GAP}_{i,t}^{g,o}$ and $\mathrm{INC\_GAP}_{i,t}^{g,o}$) and the set of control variables presented above represented by the vector $\mathbf{z'_{it}}$:

$$\mathbf{z'}_{i,t} = (\mathrm{POP}_{it}, \mathrm{GDPPC}_{it}, \mathrm{WAGE}_{it}, \mathrm{CREDIT}_{it}, \mathrm{RQI}_{it})$$

The vectors of parameters to be estimated are the β's and δ's. In addition, observe that both country and time fixed effects (different for gender, origin, and integration indicator) are included in all specifications. These fixed effects control for unobserved country-specific and unobserved year-specific effects. Finally, the ε's represent the residual error terms.

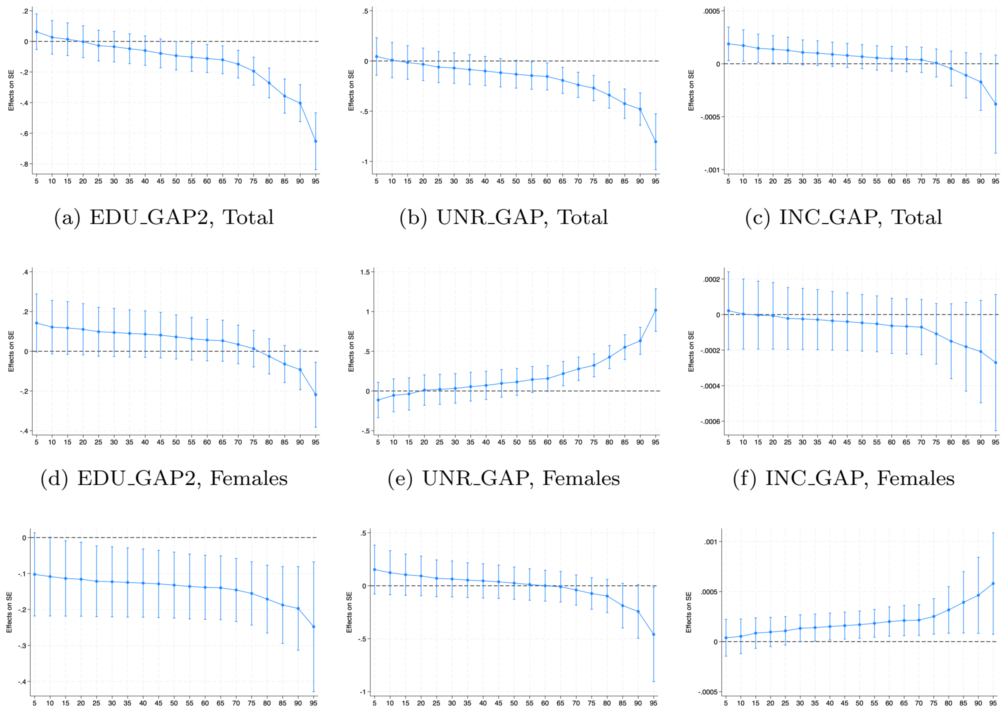
The simplest approach to estimate equation (2) is through ordinary least squares (OLS). However, a preliminary examination of the data revealed the presence of outliers (see Appendix B and Figures B.1 and B.2). To address this issue, various statistical techniques have been developed to mitigate the effects of outliers, employing robust regression estimators that down-weight observations with large residuals. Among the available robust estimators, we utilise iteratively reweighted least squares (IRLS). The IRLS estimator belongs to the class of M estimators and is highly efficient (Huber 1964). It first fits OLS, calculates the Cook's distance,[6] and drops those observations for which the Cook's distance is greater than one. Then, it applies an iterative algorithm that weighs absolute residuals. In doing that, cases with small residuals have weights equal to one; conversely, cases with larger residuals get progressively smaller weights.

4 Results

4.1 Migrants from EU countries

The regression results concerning migrants from EU countries are presented in Table 1. Columns (1) to (3) report estimates for the total number of self-employed migrants, while columns (4) to (6) and (7) to (9) pertain specifically to female and male migrants, respectively. A key finding is that, across all estimates, the primary effect of the Ethnic Concentration Index (ECI1) is negative, with its estimated coefficient consistently demonstrating high statistical significance. This finding aligns with previous literature indicating that foreign entrepreneurship is inhibited by ethnic concentration and, conversely, encouraged by ethnic diversity (Awaworyi Churchill 2019). Such results provide

[6] Cook's distance is the scaled change in fitted values and shows the influence of each observation on the fitted values. Basically, it measures how much the fitted values in the model change if the *i-th* observation is deleted. An observation with a large Cook's distance strongly affects the fitted values.



(a) EDU_GAP2, Total (b) UNR_GAP, Total (c) INC_GAP, Total

(d) EDU_GAP2, Females (e) UNR_GAP, Females (f) INC_GAP, Females

(g) EDU_GAP2, Males (h) UNR_GAP, Males (i) INC_GAP, Males

Figure 5: Marginal effects of integration indicators on the percentage of migrants' self-employment at different centiles of ECI1 (migrants from EU)

robust support for hypothesis H2b, suggesting that in countries characterised by high ethnic concentration, social interactions and economic relationships among different groups are diminished, thereby reducing the potential role of ethnic networking in promoting self-employment. In contrast, the patterns observed in the other covariates present a less uniform scenario across the groups.

Focusing on the total migrant population (columns 1 to 3), we examine the role of integration indicators. The primary results indicate that the direct effects of EDU_GAP2 and INC_GAP are both positive and statistically significant, thereby confirming hypotheses H1a and H1c. Namely, the positive coefficients suggest that an increase in these variables, which, by definition, reflect a reduction in the gaps between migrants and natives, would result in higher levels of self-employment (SE) among migrants. In contrast, the direct effect of UNR_GAP is not statistically significant, providing no support for H1b.

As for the interaction terms between any of these gaps and ECI1, they are consistently negative and statistically significant, indicating that increasing ethnic concentration negatively mediates (moderates) the direct effects of the gaps on self-employment (SE). Therefore, the estimates support hypothesis H3b. The marginal effects analysis provides a clearer picture of these dynamics.

Figure 5 illustrates the marginal effects of the three integration indicators on the self-employment of migrants from the EU across different centiles of the ECI1 distribution. The vertical bars in the figure represent the 90% statistical confidence interval. Panels (a) to (c) pertain to the total migrant population, while panels (d) to (f) focus on female migrants, and panels (g) to (i) concentrate on male migrants.

Panel (a) shows that at low values of ECI1, up to the 45^{th} centile, the marginal effect of EDU_GAP2 on SE is statistically insignificant; above the 45th centile, it is negative and statistically significant. Given the negative sign of the interaction term, the relationship between EDU_GAP2 and SE decreases as ECI1 increases. Consequently, the negative marginal impact of EDU_GAP2 on SE increases (*in absolute terms*) the higher ECI1 is. For example, at the 75^{th} centile, the marginal effect is -0.19, while at the 90^{th} centile it is

Table 1: The effect of ethnic concentration and integration indicators on the percentage of self-employment (migrants from EU)

		Total			Females			Males	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI1	-0.008373***	-0.002438***	-0.002211**	-0.004874***	-0.004667***	-0.002138*	-0.005520***	-0.004119***	-0.001760
	[0.001090]	[0.000399]	[0.000911]	[0.001194]	[0.001224]	[0.001120]	[0.001383]	[0.001189]	[0.001411]
EDU_GAP2	0.198594**			0.212050*			-0.076511		
	[0.090546]			[0.113982]			[0.088817]		
EDU_GAP2*ECI1	-0.000387***			-0.000196**			-0.000078		
	[0.000078]			[0.000084]			[0.000076]		
UNR_GAP		0.185769			-0.343551*			0.260350	
		[0.144379]			[0.175996]			[0.194373]	
UNR_GAP*ECI1		-0.000421***			0.000618***			-0.000327*	
		[0.000116]			[0.000132]			[0.000196]	
INC_GAP			0.000284**			0.000092			-0.000076
			[0.000142]			[0.000192]			[0.000175]
INC_GAP*ECI1			-0.000000*			-0.000000			0.000000
			[0.000000]			[0.000000]			[0.000000]
POP	0.000002***	0.000002***	0.000002***	0.000002***	0.000002***	0.000002***	0.000003***	0.000003***	0.000002***
	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000001]
GDPPC	0.000034	0.000007	0.000116**	-0.000031	-0.000045	0.000066	0.000047	-0.000032	0.000085
	[0.000060]	[0.000056]	[0.000050]	[0.000062]	[0.000068]	[0.000062]	[0.000070]	[0.000066]	[0.000078]
WAGE	-0.000195**	-0.000003	-0.000445***	-0.000068	0.000204*	-0.000222**	-0.000005	-0.000060	-0.000237*
	[0.000086]	[0.000074]	[0.000080]	[0.000100]	[0.000109]	[0.000106]	[0.000113]	[0.000106]	[0.000123]
CREDIT	-0.017892	-0.030929***	-0.005062	-0.037918***	-0.064410***	-0.016390	-0.027058*	-0.041234***	-0.021882
	[0.011180]	[0.009470]	[0.010050]	[0.012245]	[0.012512]	[0.013186]	[0.013836]	[0.011920]	[0.016401]
RQI	-1.999708**	-0.871328	-1.074626*	-0.242811	1.019061	0.132430	-2.126632**	-1.544828	-3.210566***
	[0.812248]	[0.715382]	[0.633099]	[0.874548]	[0.980415]	[0.851091]	[0.987547]	[0.935290]	[0.973259]
Observations	182	200	152	142	161	123	160	174	137
R-squared	0.92	0.83	0.92	0.71	0.71	0.75	0.90	0.89	0.90

Notes: Outlier robust regressions. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Country and year fixed effects included but not reported. The dependent variable is the percentage of migrants' self-employment on total employment. See the main text for more details.

-0.40. Thus, a one-unit increase in EDU_GAP2 would lower the percentage of migrants who are self-employed by 0.19 at the 75th centile and by 0.40 at the 90^{th} centile.

In summary, although the direct effect of EDU_GAP2 on SE is positive, the total effect becomes negative when accounting for the interaction effect with ECI1. Therefore, we conclude that higher levels of educational integration are associated with lower self-employment among migrants, particularly in countries with greater ethnic concentration. This finding suggests that migrants are more likely to perceive entrepreneurship as a necessity when their educational integration is low.

Panels (b) and (c) of Figure 5 present the marginal effects of UNR_GAP and INC_GAP on self-employment (SE), respectively. Regarding UNR_GAP, the marginal effect is statistically insignificant up to the 50th centile. Beyond this point, it becomes negative and statistically significant, with the negative impact intensifying as ECI1 increases. This finding indicates that in countries characterised by high ethnic concentration, a higher UNR_GAP (reflecting lower integration) is associated with reduced self-employment among migrants. In this context, the decision of migrants to engage in entrepreneurial activities does not appear to stem from necessity but rather from a desire to capitalise on improved labour opportunities. In contrast, the marginal effect of INC_GAP is positive up to the 25th centile and becomes statistically insignificant thereafter. This result suggests that in countries where ethnic concentration is low, specifically in highly fractionalised or ethnically diverse contexts, greater integration (as indicated by an increase in INC_GAP) promotes self-employment among migrants.

For female migrants (Table 1, columns (4) to (6)), the main effect of EDU_GAP2 is positive, confirming hypothesis H1a, whereas the effect of UNR_GAP is negative and therefore does not support hypothesis H1b. Additionally, no direct relationship is observed between INC_GAP and self-employment (SE). The impact of ECI1 is negative, thereby confirming hypothesis H2b. In terms of the interaction terms, a negative coefficient is estimated for EDU_GAP2 (confirming hypothesis H3b) and a positive coefficient for UNR_GAP (confirming hypothesis H3a). No direct effect is identified for INC_GAP or its interaction with ECI1. Given these results, panels (d) and (f) of Figure 5 illustrate that the overall impact of EDU_GAP2 is statistically insignificant across most centiles, with the exception of the highest 95th centile of ECI1. Conversely, panel (e) reveals that, above the 60th centile of ECI1, UNR_GAP exerts a positive, increasing, and statistically significant overall effect on SE. It is noteworthy that this finding contrasts with the results observed for the total migrant population. This discrepancy suggests that in countries with high ethnic concentration, female migrants may be compelled to pursue self-employment due to challenges associated with lower labour market integration, as indicated by higher levels of UNR_GAP. In this context, unlike the total sample, the decision of female migrants to start their own businesses appears to stem from necessity rather than a desire to capitalise on better opportunities.

For male migrants (columns (7) to (9) of Table 1), the Ethnic Concentration Index (ECI1) demonstrates negative and statistically significant coefficients (generally confirming hypothesis H2b). However, none of the main effects of the integration variables are statistically significant, nor are the interaction effects between EDU_GAP2 and INC_GAP with ECI1. The only exception is the interaction between UNR_GAP and ECI1, which is mildly statistically significant and negative, thereby confirming hypothesis H3b. Nevertheless, the marginal effects illustrated in panels (g) to (i) of Figure 5 reveal a more nuanced narrative. Specifically, the marginal effect of EDU_GAP2 (panel g) is consistently negative across all centiles of ECI1 and is statistically significant in nearly all instances. Conversely, the marginal effect of UNR_GAP is never statistically significant (panel h). In contrast, the marginal effect of INC_GAP exhibits an increasing trend as ECI1 rises, becoming statistically significant above the 35th centile (panel i). These findings suggest that the overall effect of higher educational integration, as measured by increasing EDU_GAP2, is associated with lower self-employment among male migrants, regardless of the level of ethnic concentration. However, once ECI1 exceeds the 35th centile, greater integration, as indicated by an increase in INC_GAP, is positively correlated with higher self-employment among male migrants.

Consequently, while there are no significant differences between males and females concerning EDU_GAP2, the data reveal markedly distinct scenarios for UNR_GAP and INC_GAP. For female migrants, lower labour market integration has a positive impact on self-employment (SE) at increasing levels of ECI1, whereas the effect of the income gap remains negligible at all levels of ECI1. In contrast, for male migrants, lower labour market integration does not influence SE at any level of ECI1; however, the effect of the income gap on SE becomes positive as ECI1 increases.

To complete the analysis of the sub-sample of migrants from EU countries, we will now briefly examine the control variables. The results presented in Table 1 indicate that the total native population (POP) consistently exhibits a positive coefficient that is highly statistically significant. The variable representing real per capita gross domestic product (GDPPC) shows a statistically significant positive coefficient only in column (3). When significant, the coefficients for domestic credit to the private sector as a percentage of GDP (CREDIT) and the regulatory quality index (RQI) are negative. The variable for average full-time adjusted salary per employee (WAGE) presents a somewhat less consistent pattern; it is negative in the total sample estimates reported in columns (1) and (3), as well as in the regressions for both females and males when considering the INC_GAP variable (columns 6 and 9, respectively). However, in column (5), the estimated coefficient is mildly statistically significant and positive.

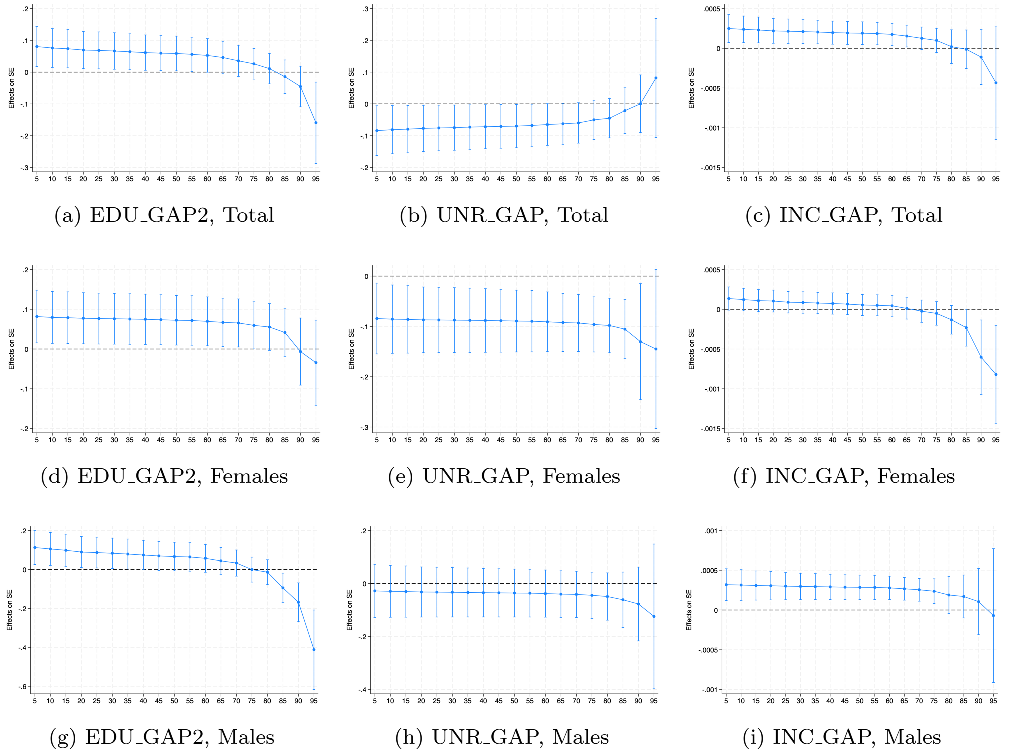
4.2 Migrants from NEU countries

Table 2 presents the regression results for the NEU migrant population, while Figure 6 illustrates the marginal effects of the integration variables.[7] Overall, our findings indicate that hypotheses H1a and H1c are confirmed across all sub-samples, while hypothesis H1b is not confirmed for either total or female migrants. Additionally, hypothesis H2b is typically confirmed, and the results for hypothesis H3 yield mixed outcomes.

To provide a concise analysis, we will focus on the marginal effects shown in Figure 6 and compare them with those for EU migrants. The analysis of total migrants presented in panel (a) of Figure 6 reveals that the marginal effect of EDU_GAP2 is positive up to the 60th centile. Beyond this point, the effect becomes statistically insignificant and ultimately exerts a negative impact on self-employment (SE) at the highest 95th centile of the ethnic concentration index (ECI). In contrast, panel (b) illustrates that the marginal effect of UNR_GAP remains negative up to the 60th centile and is statistically insignificant thereafter. Similarly, the marginal effect of INC_GAP, as shown in panel (c), diminishes with increasing ECI; however, the positive effect persists up to the 70th centile. These findings suggest that higher integration, as measured by increases in both EDU_GAP2 and INC_GAP, is correlated with greater self-employment among migrants, particularly in contexts of lower ethnic concentration. Conversely, lower integration, as indicated by increasing UNR_GAP, is associated with reduced self-employment among NEU migrants.

With respect to the self-employment of NEU female migrants, the marginal effect of EDU_GAP2, as depicted in panel (d), remains positive up to the 75th centile. Conversely, the marginal effect of UNR_GAP, illustrated in panel (e), is consistently negative throughout the range. In contrast, the marginal effect of INC_GAP, presented in panel (f), is negative and statistically significant only above the 85th centile. Excluding INC_GAP from consideration, the other two integration indicators demonstrate that increased educational integration—reflected in higher values of EDU_GAP2, correlates with greater self-employment among female migrants. Conversely, lower labour market integration, indicated by higher UNR_GAP, correlates with reduced self-employment among NEU female migrants. Notably, this presents a significant contrast to EU female migrants, where obstacles to labour market integration (as measured by higher UNR_GAP) appear to encourage self-employment, whereas NEU female migrants are deterred by similar challenges.

[7]Since the NEU regressions include a slightly larger set of countries than the EU specifications, we re-estimated the NEU models on the identical country sub-sample used for the EU analysis. The results are qualitatively unaffected, further supporting the robustness of our conclusions. The corresponding estimates are available upon request. We thank an anonymous reviewer for bringing this point to our attention.



(a) EDU_GAP2, Total (b) UNR_GAP, Total (c) INC_GAP, Total

(d) EDU_GAP2, Females (e) UNR_GAP, Females (f) INC_GAP, Females

(g) EDU_GAP2, Males (h) UNR_GAP, Males (i) INC_GAP, Males

Figure 6: Marginal effects of integration indicators on the percentage of migrants' self-employment at different centiles of ECI1 (migrants from NEU)

Regarding self-employment of NEU male migrants, the marginal effect of EDU_GAP2 remains positive up to the 40th centile of ECI1, as illustrated in panel (g). However, it is statistically insignificant between the 40th and 80th centiles, subsequently turning negative beyond the 80th centile. Conversely, the marginal effect of UNR_GAP, depicted in panel (h), consistently demonstrates statistical insignificance. In contrast, the marginal effect of INC_GAP, as shown in panel (i), remains positive up to the 80th centile of ECI1. The most significant distinctions when compared to NEU female migrants pertain to the overall effects of unemployment and income gaps. However, no notable differences emerge in comparison to EU male migrants.

As for the control variables, they exert a more pronounced impact on explaining self-employment among NEU migrants compared to their influence on EU migrants. In particular, GDPPC and CREDIT demonstrate a significant negative role. In contrast to the estimates for EU migrants presented in Table 1, the coefficients for these variables in Table 2 are consistently negative and highly statistically significant. Therefore, it appears that the general economic conditions of the countries provide a more robust explanation for the self-employment of NEU migrants.

4.3 Robustness check

To assess the robustness of the findings, we employ an alternative measure of the ethnic concentration index (ECI2), constructed utilising data from Standaert, Rayp (2022). Unlike ECI1, which contains gaps for certain years and/or specific EU countries, ECI2 does not suffer from missing observations and therefore offers broader coverage. The corresponding regression results and the marginal effects of the three integration indicators on self-employment are presented in Appendix C.

For EU migrants, the results in Table C.1 largely confirm those presented in Table 1. The same holds for the marginal effects depicted in Figure C.1, which show very similar trends to those observed in Figure 5. The only notable difference concerns the relationship between the unemployment gap (UNR_GAP) and self-employment (SE) among females.

Table 2: The effect of ethnic concentration and integration indicators on the percentage of self-employment (migrants from NEU)

	Total				Females			Males	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI1	-0.000348	-0.000184	-0.001134**	-0.000443**	-0.000286*	-0.001046**	-0.000591*	-0.000213	-0.001412**
	[0.000232]	[0.000127]	[0.000533]	[0.000194]	[0.000150]	[0.000442]	[0.000315]	[0.000191]	[0.000655]
EDU_GAP2	0.096998**			0.089382**			0.148984**		
	[0.042945]			[0.042738]			[0.060554]		
EDU_GAP2*ECI1	-0.000047**			-0.000022			-0.000103***		
	[0.000019]			[0.000014]			[0.000030]		
UNR_GAP		-0.094373*			-0.080506*			-0.021526	
		[0.053742]			[0.047819]			[0.067894]	
UNR_GAP*ECI1		0.000031			-0.000011			-0.000018	
		[0.000026]			[0.000022]			[0.000035]	
INC_GAP			0.000292**			0.000192*			0.000343**
			[0.000128]			[0.000101]			[0.000150]
INC_GAP*ECI1			-0.000000			-0.000000**			-0.000000
			[0.000000]			[0.000000]			[0.000000]
POP	0.000001***	0.000001***	0.000001	0.000001*	0.000001***	-0.000000	0.000001***	0.000001***	0.000001***
	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000001]
GDPPC	-0.000355***	-0.000424***	-0.000305***	-0.000460***	-0.000348***	-0.000416***	-0.000316***	-0.000368***	-0.000297***
	[0.000051]	[0.000047]	[0.000062]	[0.000048]	[0.000047]	[0.000052]	[0.000072]	[0.000065]	[0.000071]
WAGE	0.000093	0.000148**	0.000053	0.000081	0.000033	-0.000026	-0.000041	-0.000068	0.000031
	[0.000066]	[0.000058]	[0.000095]	[0.000077]	[0.000072]	[0.000088]	[0.000099]	[0.000090]	[0.000115]
CREDIT	-0.057761***	-0.069664***	-0.056624***	-0.078975***	-0.062287***	-0.078359***	-0.038500***	-0.049934***	-0.042776***
	[0.009256]	[0.008008]	[0.012193]	[0.009120]	[0.007815]	[0.010773]	[0.013315]	[0.010881]	[0.014118]
RQI	-1.904916***	-0.992543	-1.810920**	-1.293155*	-1.498114**	-0.977621	-3.382694***	-3.283133***	-2.426470**
	[0.692873]	[0.606884]	[0.834950]	[0.687858]	[0.653855]	[0.690543]	[0.952719]	[0.820580]	[0.962658]
Observations	236	261	198	171	188	153	223	236	188
R-squared	0.968005	0.967216	0.964299	0.964359	0.965292	0.933004	0.965780	0.965388	0.974037

Notes: Outlier robust regressions. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Country and year fixed effects included but not reported. The dependent variable is the percentage of migrants' self-employment on total employment. See the main text for more details.

Table 3: Synthesis of the estimated effects of the integration indicators on EU and NEU migrants' self-employment

	EU migrants	NEU migrants
EDU_GAP2: higher educational integration (secondary)	*discourages* total self-employment at high levels of ethnic concentration and male self-employment independently of ethnic concentration; *does not affect* female self-employment	*encourages* total, female and male self-employment, until very high levels of ethnic concentration
UNR_GAP: lower economic integration	*discourages* total self-employment at high levels of ethnic concentration; *encourages* female self-employment at high levels of ethnic concentration; *does not affect* male self-employment	*mildly discourages* total self-employment at low levels of ethnic concentration; *discourages* female self-employment independently of ethnic concentration; *does not affect* male self-employment
INC_GAP: higher social integration	*mildly favours* total self-employment at very low levels of ethnic concentration*; does not affect* female self-employment; *encourages* male self-employment at high levels of ethnic concentration	*encourages* total and male self-employment; *does not affect* female self-employment

In Figure 5, ECI1 appeared to negatively mediate this relationship, whereas in Figure C.1, ECI2 suggests a positive mediation effect.

For NEU migrants, the coefficients in Table C.2 are generally less statistically significant than in the baseline estimates. Nevertheless, Figure C.2 shows that, taking into account the interaction effects, the overall patterns remain largely consistent with those in Figure 6. Again, differences emerge mainly along gender lines: for females, the relationship between UNR_GAP and self-employment (SE) changes when using ECI2 (panel (e)), and, for males, a similar discrepancy appears in the link between the income gap (INC_GAP) and self-employment (panel (i)).

5 Discussion of the main findings and policy implications

The empirical analysis has been developed through several avenues, addressing the impact of integration indicators and ethnic concentration on migrants' self-employment. Concerning the direct effects of the main explanatory variable, the general key findings can be summarised as follows. Among both EU and NEU migrants, higher levels of educational and social integration encourage self-employment, supporting H1a and H1c, in line with findings from Sun, Fong (2022) and Zhang et al. (2024). Regarding the unemployment gap, in both EU and NEU sub-samples, its direct effect is almost negligible in line with Andersson (2021), among others, offering no strong support for H1b. The Ethnic Concentration Index (ECI1) consistently shows a negative effect across all specifications, supporting H2b and aligning with previous evidence that foreign entrepreneurship tends to be fostered by ethnic diversity (Awaworyi Churchill 2019).

The scenario becomes significantly more nuanced and complex when considering the interaction between integration indicators and ECI1. To manage this complexity, we summarise our results in Table 3.

In general, the picture shows heterogeneous patterns that vary depending on migrants' area of origin and gender. More in detail, three major outcomes are worth noting. The first concerns the effect of educational integration among EU migrants with respect to NEU migrants. When ethnic concentration is high, higher educational integration discourages self-employment for the total sample and for males, while it does not appear to influence females. The second result to highlight concerns the effects of an increase in the employment gap for females, specifically among EU migrants, where lower integration is associated with higher self-employment rates. These findings indicate that integration policies may have opposite effects across gender and nationality. Lastly, regarding the income gap, no significant differences are noted between the EU and NEU sub-samples or across genders. Greater integration generally supports self-employment among total and male migrants, and integration policies are likely to systematically favour self-employment by reducing uncertainty.

These results underscore the central finding of our analysis: the interaction between integration and ethnic concentration plays a crucial role in shaping migrants' self-employment. This implies that ethnic and cultural diversity should be carefully considered when designing policies aimed at promoting migrant entrepreneurship. Measures that enhance integration may have very different, potentially opposite, effects in areas with high ethnic concentration, and vice versa.

Moreover, our analysis highlights how the issue of immigrant entrepreneurship is influenced by two fundamental aspects of migration policies: the integration of those already residing in the host country and the encouragement or discouragement of new arrivals. Regarding individuals already settled in host countries, our findings suggest that policymakers should implement tailored policies focused on citizenship and gender, while also considering the degree of ethnic concentration, as this factor could significantly impact policy effectiveness. In terms of new arrivals, our results indicate that overly restrictive policies concerning the number and origin of migrants may hinder entrepreneurial growth in potential host countries.

Other important implications of our findings emerge when interpreted through the lens of current regional-level challenges in Europe. The heterogeneous interaction patterns observed across origin, gender, and integration dimensions offer useful insights for interpreting sub-national dynamics as well. Integration levels, labour-market conditions, and ethnic concentration vary widely across European regions, creating territorial opportunity structures that shape migrants' incentives for self-employment. The fact that the effects of educational, occupational, and social integration differ across groups and depend on ethnic concentration suggests that regional self-employment responses are strongly conditioned by local contexts. This is particularly relevant given the persistent disparities across European regions in labour-market inclusiveness, demographic composition, and sectoral specialisation. Our findings suggest that such differences may amplify or even reverse the effects of integration policies. Hence, integration and entrepreneurship support measures may operate differently in areas characterised by high ethnic diversity, tight labour markets, or strong sectoral concentration. Overall, the national patterns identified in our analysis help clarify how region-specific conditions mediate the relationship between integration and self-employment, as well as how regional challenges shape the potential contribution of migrant entrepreneurship to local development.

To conclude, it is important to recognise that the available economic policy options are challenging to implement and often uncertain in their effects. This uncertainty may explain why the literature has yielded contradictory results regarding the impact of integration policies on migrants' self-employment, with many studies reporting either scarce or nonexistent effects (Solano et al. 2022).

6 Concluding remarks

This study presents a new macro-comparative perspective on the relationship between integration and migrant entrepreneurship in EU countries, utilising harmonised integration indicators available across Member States. Despite extensive research on migrant entrepreneurship, empirical work that jointly examines integration outcomes and self-employment remains scarce, and this scarcity becomes even narrower when ethnic concentration and fragmentation are taken into account. To our knowledge, no macro-level comparative study has analysed this relationship using explicit and standardised measures of integration; existing micro-level contributions rely on indirect proxies and are not based on harmonised indicators. Against this backdrop, the paper makes three main contributions. First, it provides an integrated analysis of migrant entrepreneurship, integration, and ethnic concentration. Second, it adopts a macroeconomic perspective, offering a cross-country comparison largely absent from previous research. Third, it exploits harmonised indicators that ensure comparability across EU Member States and enable a systematic assessment of how different dimensions of integration relate to migrant self-employment.

While these contributions suggest the value of a macro-comparative approach, certain limitations must also be acknowledged. Macro-level data cannot capture the full complexity of migrant self-employment, including factors such as local context, residential mobility within the host country, length of stay, or motivations for migration. Addressing these mechanisms requires access to individual-level data, which represents a central avenue for future research. A further limitation concerns the temporal coverage of available indicators: although the dataset spans a long period and multiple countries, changes in migration policies or labour-market regulations within individual Member States cannot be fully isolated. Nevertheless, by establishing the first macro-comparative analysis based on standardised integration indicators, this study opens a space for systematic and comparable research on the economic and social incorporation of migrant entrepreneurs in Europe.

Acknowledgements

We thank the Editor Zoltán Elekes and two anonymous reviewers for useful and constructive comments and suggestions. All usual disclaimers apply.

Funding

This study was funded by the European Union – NextGenerationEU, Mission 4, Component 2, in the framework of the GRINS – Growing Resilient, INclusive and Sustainable project [grant number GRINS PE00000018 – CUP F53C22000760007]. The views and opinions expressed in this study are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

Abada T, Hou F, Lu Y (2014) Choice or necessity: Do immigrants and their children choose self-employment for the same reasons? *Work, Employment and Society* 28: 78–94. CrossRef

Andersson H (2021) Ethnic enclaves, self-employment, and the economic performance of refugees: Evidence from a Swedish dispersal policy. *International Migration Review* 55: 58–83. CrossRef

Andersson L, Hammarstedt M (2015) Ethnic enclaves, networks and self-employment among Middle Eastern immigrants in Sweden. *International Migration* 53: 27–40. CrossRef

Appau S, Awaworyi Churchill S, Farrell L (2019) Social integration and subjective well-being. *Applied Economics* 51: 1748–1761. CrossRef

Awaworyi Churchill S (2019) Firm financial performance in Sub-Saharan Africa: The role of ethnic diversity. *Empirical Economics* 57: 957–970. CrossRef

Bashko A (2022) Entrepreneurship and ethnic economy employment among Chinese and Vietnamese residents of Warsaw. *International Migration* 60: 177–192. CrossRef

Becker GS (1962) Investment in human capital: A theoretical analysis. *Journal of Political Economy* 70: 9–49. CrossRef

Biagi B, Lambiri D, Meleddu M (2025) Do local attitudes change with the exposure and the status of the migrants? *REGION* 12: 47–71. CrossRef

Borjas GJ (1986) The self-employment experience of immigrants. *The Journal of Human Resources* 21: 485–506. CrossRef

Borjas GJ (2000) Ethnic enclaves and assimilation. *Swedish Economic Policy Review* 7: 89–122

Boudreaux CJ (2020) Ethnic diversity and small business venturing. *Small Business Economics* 54: 25–41. CrossRef

Brzozowski J (2019) Entrepreneurship and economic integration of immigrants: A critical review of literature. *International Journal of Entrepreneurship and Innovation Management* 23: 584–604. CrossRef

Chodavadia S, Pekkala Kerr S, Kerr W, Maiden L (2025) Immigrant entrepreneurship: New estimates and a research agenda. *A Research Agenda for Migration and Innovation* 7: 101–127. CrossRef

Constant A, Shachmurove Y (2006) Entrepreneurial ventures and wage differentials between Germans and immigrants. *International Journal of Manpower* 27: 208–229. CrossRef

Dheer RJS (2018) Entrepreneurship by immigrants: A review of existing literature and directions for future research. *International Entrepreneurship and Management Journal* 14: 555–614. CrossRef

European Commission (2020) Action plan on integration and inclusion 2021–2027. European Commission, Luxemburg, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0758

European Ministerial Conference (2010) Declaration of the European Ministerial Conference on integration. Zaragoza. https://migrant-integration.ec.europa.eu/library-document/declaration-european-ministerial-conference-integration-zaragoza-15-16-april-2010_en

Eurostat (2020) Migrant integration statistics. Eurostats. https://ec.europa.eu/eurostat/web/products-statistical-books/-/ks-06-20-184

Gordon MM (1964) *Assimilation in American life: The role of race, religion, and national origins.* Oxford University Press, USA. CrossRef

Huber PJ (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35: 73–101. CrossRef

Kariv D, Coleman S (2015) Toward a theory of financial bricolage: The impact of small loans on new businesses. *Journal of Small Business and Enterprise Development* 22: 196–224. CrossRef

Kaufmann D, Kraay A, Mastruzzi M (2010) The worldwide governance indicators: Methodology and analytical issues. *World Bank Policy Research Working Paper* 5430. CrossRef

Kim KC, Hurth WH, Fernandez M (1989) Intra-group differences in business participation: Three Asian immigrant groups. *International Migration Review* 23: 73–95. CrossRef

Kloosterman R, Leun J, Rath J (1999) Mixed embeddedness: (in)formal economic activities and immigrant businesses in the Netherlands. *International Journal of Urban and Regional Research* 23: 252–266. CrossRef

Lofstrom M (2002) Labor market assimilation and the self-employment decision of immigrant entrepreneurs. *Journal of Population Economics* 15: 83–114. CrossRef

Massidda C, Tunis S, Piras R (2024) Migrant integration and entrepreneurship: A systematic literature review through the lens of bibliometrics. *Rivista Internazionale di Scienze Sociali* 132: 401–427. CrossRef

Matricano D, Sorrentino M (2014) Ukrainian entrepreneurship in Italy: Factors influencing the creation of ethnic ventures. *Journal of Innovation and Entrepreneurship* 3: article number 10. CrossRef

McAuliffe M, Oucho LA (2024) *World Migration Report 2024*. International Organization for Migration (IOM), Geneva. CrossRef

Polemis ML, Stengos T (2020) The impact of regulatory quality on business venturing: A semi-parametric approach. *Economic Analysis and Policy* 67: 29–36. CrossRef

Portes A, Jensen L (1987) What's an ethnic enclave? The case for conceptual clarity. *American Sociological Review* 52: 768–70

Putnam RD (2007) E pluribus unum: Diversity and community in the twenty-first century (The 2006 Johan Skytte Prize Lecture). *Scandinavian Political Studies* 30: 137–174. CrossRef

Schultz TW (1961) Investment in human capital. *The American Economic Review* 51: 1–17

Solano G, Yilmaz S, Huddleston T (2022) The link between migration policies and migration and migrant integration dynamics. An overview of the existing literature. (deliverable 8.1; HumMingBird project 870661 – H2020, https://www.researchgate.net/-publication/360335787, Leuven

Standaert S, Rayp G (2022) Where did they come from, where did they go? Bridging the gaps in migration data. UNU Institute on Comparative Regional Integration Studies. Working Paper Nr. 4, https://cris.unu.edu/sites/cris.unu.edu/files/WP22.04%20-%20Standaert%20and%20Rayp_0.pdf

Sun SB, Fong E (2022) The role of human capital, race, gender, and culture on immigrant entrepreneurship in Hong Kong. *Journal of Small Business and Entrepreneurship* 34: 363–396. CrossRef

Van Der Sluis J, Praag M, Vijverberg W (2008) Education and entrepreneurship selection and performance. A review of the empirical literature. *Journal of Economic Surveys* 22: 795–841. CrossRef

Wang Q, Li W (2007) Entrepreneurship, ethnicity and local contexts: Hispanic entrepreneurs in three U.S. Southern metropolitan areas. *GeoJournal* 68: 167–182. CrossRef

Wang Z, Graaff T, Nijkamp P (2018) Barriers of culture, networks, and language in international migration: A review. *REGION* 5: 73–89. CrossRef

Wennekers S, Stel A, Carree M, Thurik R (2010) The relationship between entrepreneurship and economic development: Is it U-shaped? *Foundations and Trends in Entrepreneurship* 6: 167–237. CrossRef

Yavuz RI, Bahadir B (2022) Remittances, ethnic diversity, and entrepreneurship in developing countries. *Small Business Economics* 58: 1931–1952. CrossRef

Yilmaz S, Solano G (2024) Migrant integration. In: *The Palgrave Handbook of Global Social Problems*. Palgrave Macmillan, Cham. CrossRef

Zhang P, Wei X, Mao G (2024) Cultural diversity, social integration, and migrant entrepreneurship—evidence from the China migrants dynamic survey. *Small Business Economics* 62: 1135–1155. CrossRef

© 2026 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Milan, Italy. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Appendices

A Basic Information

This appendix provides supplementary material referenced in the main text. Table A.1 reviews the empirical literature discussed in Section 2, summarising the main characteristics and key findings of each study. Table A.2 lists all variables used in the analysis, together with their definitions and data sources. Table A.3 reports summary statistics separately for the EU and NEU subsamples, with an additional breakdown by gender. Finally, Figure A.1 presents six maps illustrating the geographical distribution of migrant self-employment rates across EU Member States in 2021, disaggregated by country of origin (EU and NEU) and gender.

Table A.1: Literature review

Author(s)	Dependent variable	Integration variable(s)	Ethnic dimension	Countries/City	Period	Methodology	Main findings
Lofstrom (2002)	Probability of being self-employed	- Length of stay (LoS) - Education - Earnings differentials	Proportion of immigrants from the same country and living in the same metropolitan area	USA	1980-1990	OLS/Probit	- LoS: *inverted_U-shaped* - Education: *positive* - Earning differentials: *no effect* - Ethnic groups: *positive*
Constant and Shachmurove (2006)	Probability of being self-employed	- Length of stay (LoS) - Education	Analysis divided by ethnic groups	Germany	2000	Binomial logit/Logistic	- LoS: *U-shaped* - Education: *positive* - Age in different ethnic groups: *positive*
Wang and Li (2007)	Probability of Hispanic immigrants being self-employed	- Length of stay (LoS) - Language proficiency (English) - Education	Characteristics of Hispanic Entrepreneurs in the USA in three metropolitan areas	USA	2000	Logistic	- LoS: *positive* - Language proficiency: *positive* - Education: *negative*
Abada et al. (2014)	Probability of being self-employed	- Length of stay (LoS) - Language proficiency - Education - Unemployment - Earnings differentials	- Ethnic group population share	Canada	1981 and 2006 census data	Probit	- LoS: *positive* - Language proficiency: *positive* - Education: *differences in generations* - Unemployment: *positive for fathers* - Earning differentials: *positive (stronger for sons than for fathers)* - Group population share: *no effect*
Matricano and Sorrentino (2014)	Probability of Ukrainians creating a new ethnic venture within the enclave	- Length of stay (LoS) - Unemployment - Education	- Co-ethnic employees - Inputs from co-ethnic suppliers - Ethnic output for existing ethnic markets	Caserta (Italy)	2009-2010	Logistic	- LoS: *no effect* - Education: *no effect* - Ethnic dimension: *no effect*

Table A.1: Literature review (cont.)

Author(s)	Dependent variable	Integration variable(s)	Ethnic dimension	Countries/City	Period	Methodology	Main findings
Andersson and Hammarstedt (2015)	Probability of Middle Eastern immigrants being self-employed by	- Length of stay (LoS) - Education - Unemployment	- Ethnic enclave (group concentration) - Ethnic network	Sweden	2011	Probit	- LoS: *inverted-U* - Education: *non-linear effect* - Unemployment: *no effect* - Ethnic enclave: *positive* - Ethnic networks: *negative*
Bashko (2022)	Probability of being self-employed	- Length of stay (LoS) - Language proficiency - Education	- Self-declared number of co-ethnic friends who reside in Poland	Warsaw (Poland)	2017	Logistic	- LoS: *positive* - Language proficiency: *positive* - Education: *positive* - Network size: *positive*
Andersson (2021)	Probability of refugees being self-employed (within 5 years)	- Unemployment	- Co-ethnic self-employment - Co-ethnic education	Working-age foreign-born adults who were given a residence permit in Sweden during 1990 or 1991	1990-2014	Linear probability model	- Unemployment: *no effect* - Co-ethnic self-employment: *positive* - Co-ethnic education: *positive*
Sun and Fong (2022)	Probability of being self-employed	- Length of stay (LoS) - Language proficiency - Education	Co-ethnic size	Hong Kong	2001 and 2016 census data	Linear probability model	- LoS: *positive* - Language proficiency: *positive* - Education: *positive* - Co-ethnic size: *no effect*
Zhang et al. (2024)	Probability of being self-employed	- Length of stay (LoS) - Social interaction - Social exclusion - Social identification - Education	- Cultural diversity - Migrant diversity - Ethnic diversity	China	2017	Logit	- LoS: *positive* - Education: *positive* - Food diversity: *negative* - Migrant diversity: *positive* - Ethnic diversity: *negative*

Table A.2: Variable definitions and sources

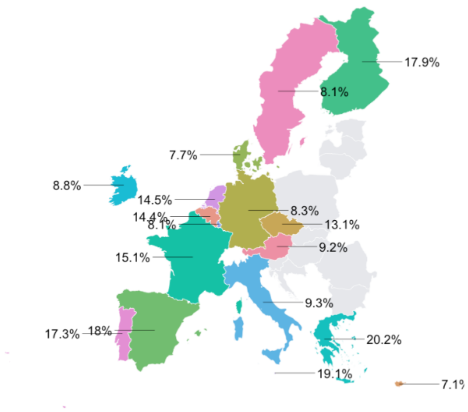
Name	Definition	Source
Dependent variables		
SE	Percentage of migrants' self-employment in total employment	Eurostat
Main independent variables		
EDU_GAP1	Primary school educational gap (migrants vs. native population)	Eurostat
EDU_GAP2	Secondary school educational gap (migrants vs. native population)	Eurostat
EDU_GAP3	Tertiary school educational gap (migrants vs. native population)	Eurostat
UNR_GAP	Unemployment rate (migrants vs. native population)	Eurostat
INC_GAP	Income gap (migrants vs. native population)	Eurostat
ECI1	Ethnic concentration index	Authors' elaboration based on Eurostat data
ECI2	Ethnic concentration index	Authors' elaboration based on Standaert and Rayp (2022) data
Control variables		
POP	Total native population	Eurostat
GDPPC	Real per capita gross domestic product	Eurostat
WAGE	Average full-time adjusted salary per employee	Eurostat
CREDIT	Domestic credit to the private sector as a % of GDP	Eurostat
RQI	Regulatory quality index	World Bank

Table A.3: Descriptive statistics

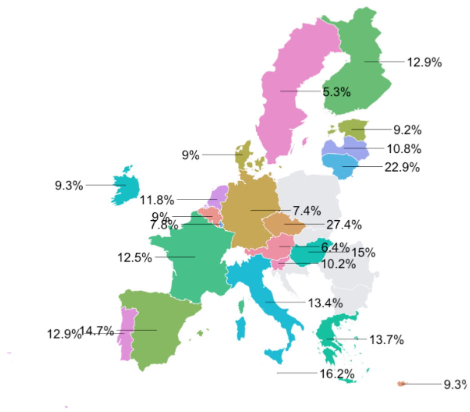
	EU sub-sample											
	Total				Females				Males			
	Mean	Sd	Min	Max	Mean	Sd	Min	Max	Mean	Sd	Min	Max
SE	12.23	4.84	3.00	45.20	9.03	4.17	2.40	56	14.45	5.74	3.4	45.30
EDU_GAP1	-1.71	11.89	-30.60	25.00	-1.78	11.20	-33.70	26.30	0.04	12.73	-33.20	26.80
EDU_GAP2	-2.91	11.44	-32.90	23.70	-1.01	11.03	-19.90	24.30	-3.13	11.83	-30.80	24.00
EDU_GAP3	6.18	12.61	-19.40	65.50	4.29	11.46	-21.60	45.90	5.79	13.92	-17.7	70.80
UNR_GAP	2.33	2.67	-6.90	17.30	2.99	3.60	-13.80	32.20	1.92	2.43	-3.80	13.30
INC_GAP	-1805	3079	-7751	10139	-2330	3071	-8893	10864	-2022	3215	-7328	11243
	NEU sub-sample											
	Total				Females				Males			
	Mean	Sd	Min	Max	Mean	Sd	Min	Max	Mean	Sd	Min	Max
SE	12.87	8.18	2.60	55.00	9.06	5.22	3.40	38.70	15.22	9.86	3.60	71.10
EDU_GAP1	9.70	13.89	-20.90	35.70	11.49	14.10	-19.20	46.70	9.45	14.65	-23.40	49.80
EDU_GAP2	-7.01	12.19	-30.40	22.10	-6.49	13.26	-33.40	26.90	-8.24	12.39	-35.00	35.10
EDU_GAP3	-0.07	12.55	-25.30	39.60	-2.60	12.53	-28.30	35.70	1.67	13.67	-23.30	43.70
UNR_GAP	8.17	6.20	-6.70	26.60	9.65	7.52	-10.00	35.50	8.14	6.05	-3.80	27.60
INC_GAP	-4683	3533	-15524	9037	-4893	3364	-15854	5975	-5185	3710	-15123	12167

Table A.3: Descriptive statistics (cont.)

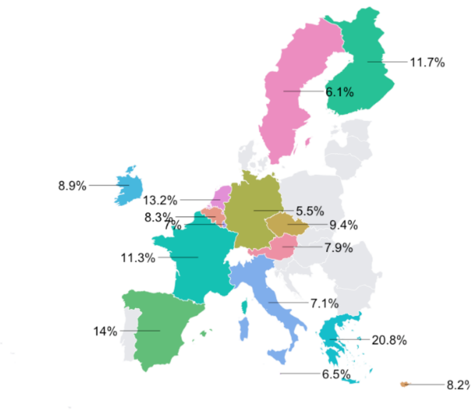
	Common variables across EU and NEU sub-samples			
	Total			
	Mean	Sd	Min	Max
ECI1	1446	1546	304	10000
ECI2	1649	1429	244	7827
POP(thousands)	14900	19900	277	75200
GDPPC	24726	16790	2990	88250
WAGE	24341	14926	1919	72247
CREDIT	83.48	42.79	0.19	254.67
RQI	0.13	0.96	-2.08	1.93



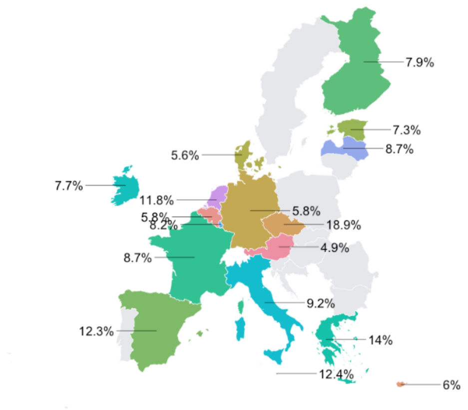
(a) EU migrants (Total)



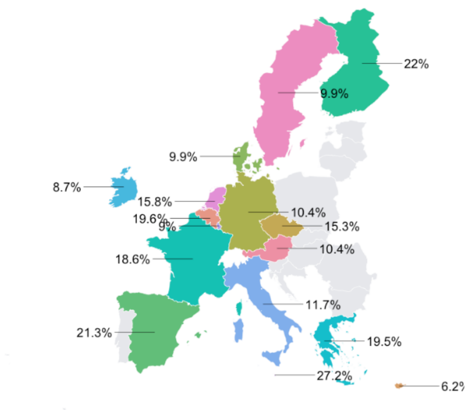
(b) NEU migrants (Total)



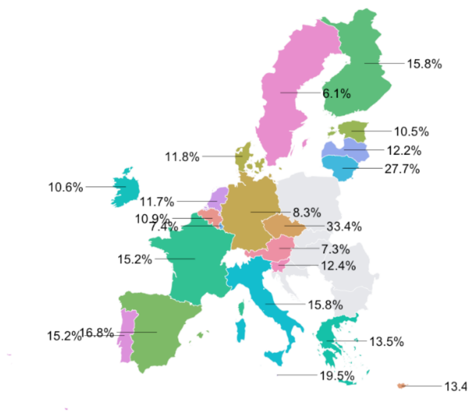
(c) EU migrants (Females)



(d) NEU migrants (Females)



(e) EU migrants (Males)



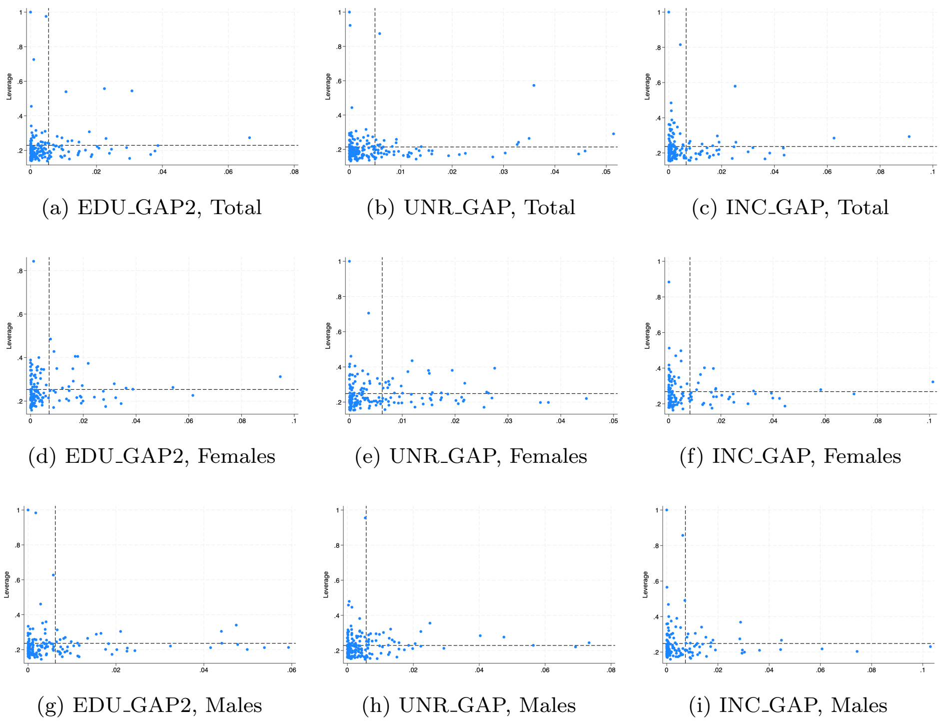
(f) NEU migrants (Males)

Figure A.1: Maps of the dependent variables (2021)

B Outliers in ordinary least squares (OLS) regressions

Outliers are observations that are inconsistent with the rest of the data, and in such situations, OLS estimates lead to inefficiency and result in low statistical power. They can have a strong adverse impact on the estimates and undermine OLS results should they remain unnoticed.

A simple and useful method for determining how a given point affects OLS regressions is illustrated in Figures B.1 (EU sub-sample) and B.2 (NEU sub-sample). In these figures, each panel represents the scatterplot of the leverage against the normalised residuals squared for the OLS regressions of equation (2). In this type of figure, the most problematic points have a high leverage and a high residual (located in the upper right of the plot). Somehow less problematic but still worrisome are points in the upper left of the plot that have a high leverage and a small, normalised residual squared. For both the EU and NEU sub-samples, Figures B.1 and B.2 show that all OLS regressions display problematic points (in both the upper right and upper left of the plot).

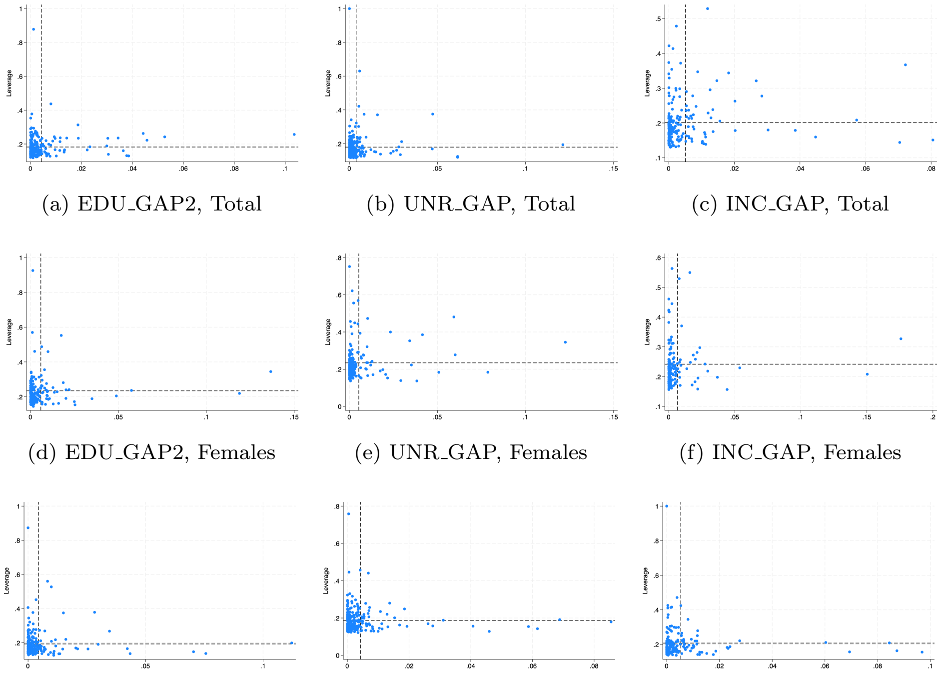


(a) EDU_GAP2, Total (b) UNR_GAP, Total (c) INC_GAP, Total

(d) EDU_GAP2, Females (e) UNR_GAP, Females (f) INC_GAP, Females

(g) EDU_GAP2, Males (h) UNR_GAP, Males (i) INC_GAP, Males

Figure B.1: Leverage against the normalised residuals squared. OLS regressions (EU sub-sample)



(a) EDU_GAP2, Total (b) UNR_GAP, Total (c) INC_GAP, Total

(d) EDU_GAP2, Females (e) UNR_GAP, Females (f) INC_GAP, Females

(g) EDU_GAP2, Males (h) UNR_GAP, Males (i) INC_GAP, Males

Figure B.2: Leverage against the normalised residuals squared. OLS regressions (NEU sub-sample)

C Additional results with ECI2

This appendix presents results based on the alternative ethnic concentration measure, ECI2, derived from the bilateral migration series estimated by Standaert, Rayp (2022). Tables C.1 and C.2 report the regression estimates for the EU and NEU subsamples, respectively, while Figures C.1 and C.2 illustrate the corresponding marginal effects across the ECI2 distribution.

Table C.1: The effect of ethnic concentration (ECI2) and integration indicators on the percentage of self-employment (EU sub-sample)

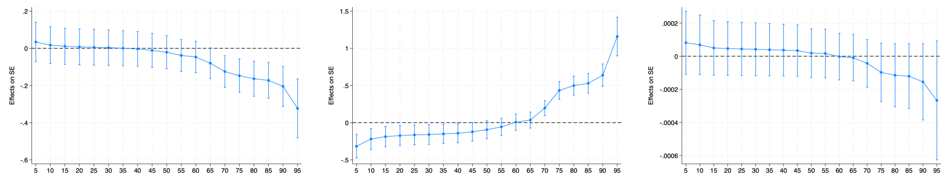
	Total			Females			Males		
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI2	-0.002956**	-0.000400	-0.000416	0.000288	-0.001464*	-0.000272	-0.001970	-0.000797	-0.000469
	[0.001481]	[0.000703]	[0.000664]	[0.001669]	[0.000786]	[0.000820]	[0.001274]	[0.001167]	[0.001057]
EDU_GAP2	0.124196			0.055296			0.001603		
	[0.082019]			[0.089659]			[0.094582]		
EDU_GAP2*ECI2	-0.000206***			0.000013			-0.000132**		
	[0.000074]			[0.000077]			[0.000060]		
UNR_GAP		-0.263455*			-0.456945***			-0.174206	
		[0.152902]			[0.151728]			[0.160293]	
UNR_GAP*ECI2		0.000340**			0.000592***			0.000264*	
		[0.000157]			[0.000136]			[0.000141]	
INC_GAP			-0.000148			-0.000474**			0.000034
			[0.000168]			[0.000183]			[0.000131]
INC_GAP*ECI2			0.000000			0.000000*			0.000000
			[0.000000]			[0.000000]			[0.000000]
POP	0.000002***	0.000002***	0.000002***	0.000002***	0.000001***	0.000001***	0.000002***	0.000002***	0.000002***
	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]
GDPPC	0.000111*	0.000106**	0.000119***	-0.000011	0.000107**	0.000061	0.000168**	0.000030	0.000166***
	[0.000058]	[0.000045]	[0.000045]	[0.000065]	[0.000051]	[0.000057]	[0.000073]	[0.000067]	[0.000057]
WAGE	-0.000096	-0.000060	-0.000242**	-0.000102	0.000056	-0.000234*	-0.000050	0.000068	-0.000250**
	[0.000076]	[0.000068]	[0.000104]	[0.000099]	[0.000083]	[0.000119]	[0.000114]	[0.000091]	[0.000115]
CREDIT	-0.003343	-0.011218	0.001859	-0.012168	-0.015765	-0.008085	-0.001467	-0.026788**	0.006839
	[0.008884]	[0.007091]	[0.008193]	[0.010224]	[0.010221]	[0.010807]	[0.011100]	[0.011155]	[0.009362]
RQI	-1.656295**	-0.039395	-2.126871***	-0.365856	0.147193	-1.479939	-0.757378	-1.315411	-1.739406
	[0.761142]	[0.687834]	[0.743523]	[0.838370]	[0.906072]	[0.972409]	[1.127216]	[0.950530]	[1.066503]
Observations	231	248	189	181	199	154	204	212	169
R-squared	0.89	0.82	0.85	0.59	0.60	0.63	0.86	0.86	0.86

Notes: Outlier robust regressions. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Country and year fixed effects included but not reported. The dependent variable is the percentage of migrants' self-employment on total employment. ECI2 is computed using data from Standaert, Rayp (2022). See the main text for more details.

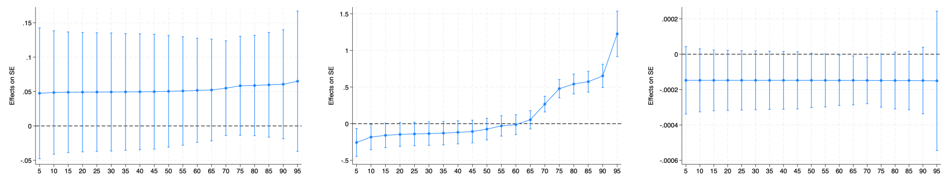
Table C.2: The effect of ethnic concentration (ECI2) and integration indicators on the percentage of self-employment (NEU sub-sample)

		Total			Females			Males	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI2	0.000158	-0.001133	-0.000977	-0.001002	-0.001841	0.000230	-0.001571*	-0.000352	-0.002123***
	[0.000871]	[0.000876]	[0.000764]	[0.000921]	[0.001348]	[0.001398]	[0.000913]	[0.000995]	[0.000815]
EDU_GAP2	0.181489***			0.066551			0.216519***		
	[0.060673]			[0.064969]			[0.074359]		
EDU_GAP2*ECI2	-0.000079***			-0.000103**			-0.000153***		
	[0.000030]			[0.000043]			[0.000036]		
UNR_GAP		0.032600			-0.069248			0.150706*	
		[0.057589]			[0.074552]			[0.077786]	
UNR_GAP*ECI2		-0.000029			0.000112			-0.000069**	
		[0.000026]			[0.000074]			[0.000035]	
INC_GAP			-0.000030			0.000193			0.000057
			[0.000194]			[0.000155]			[0.000204]
INC_GAP*ECI2			0.000000			-0.000000			0.000000
			[0.000000]			[0.000000]			[0.000000]
POP	0.000002***	0.000001***	0.000001***	0.000001***	0.000001***	0.000000	0.000002***	0.000001**	0.000001***
	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000001]
GDPPC	-0.000354***	-0.000262***	-0.000256***	-0.000183***	-0.000199***	-0.000163***	-0.000304***	-0.000271***	-0.000195***
	[0.000070]	[0.000046]	[0.000048]	[0.000053]	[0.000052]	[0.000051]	[0.000082]	[0.000080]	[0.000069]
WAGE	0.000126	-0.000052	-0.000024	-0.000059	0.000016	-0.000058	0.000060	-0.000097	-0.000054
	[0.000094]	[0.000063]	[0.000100]	[0.000080]	[0.000081]	[0.000097]	[0.000125]	[0.000102]	[0.000133]
CREDIT	-0.046345***	-0.029435***	-0.031940***	-0.025976***	-0.025203***	-0.015255*	-0.023893	-0.025014	-0.000404
	[0.011204]	[0.007366]	[0.009935]	[0.008190]	[0.007997]	[0.008991]	[0.016324]	[0.015816]	[0.013728]
RQI	-2.665938***	-2.578807***	-3.003048***	-1.848037*	-1.104395	-1.563450	-5.258018***	-4.462711***	-6.181287***
	[0.881467]	[0.727736]	[0.896842]	[1.082735]	[1.079102]	[1.085529]	[1.423490]	[1.273426]	[1.328958]
Observations	298	313	241	214	228	186	275	284	226
R-squared	0.90	0.93	0.90	0.88	0.87	0.86	0.90	0.90	0.91

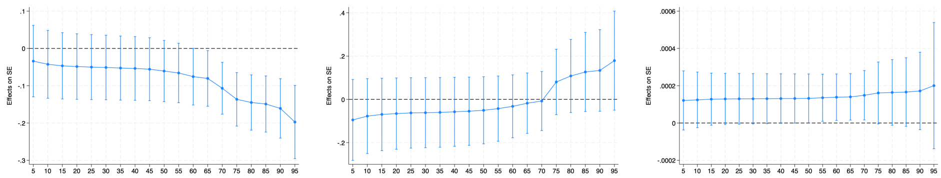
Notes: Outlier robust regressions. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Country and year fixed effects included but not reported. The dependent variable is the percentage of migrants' self-employment on total employment. ECI2 is computed using data from Standaert, Rayp (2022). See the main text for more details.



(a) EDU_GAP2, Total (b) UNR_GAP, Total (c) INC_GAP, Total

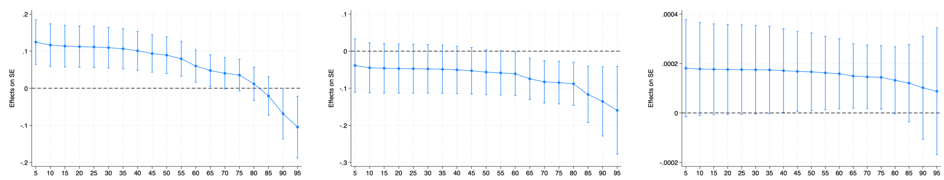


(d) EDU_GAP2, Females (e) UNR_GAP, Females (f) INC_GAP, Females

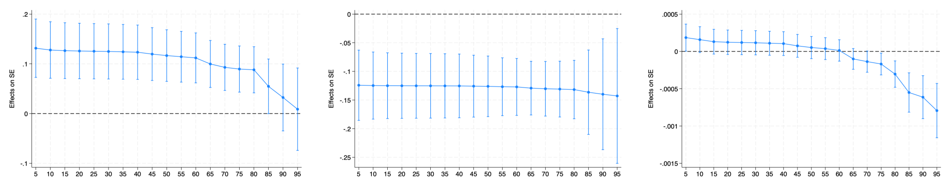


(g) EDU_GAP2, Males (h) UNR_GAP, Males (i) INC_GAP, Males

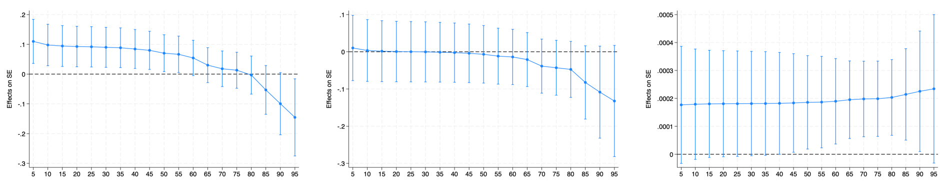
Figure C.1: Marginal effects of integration indicators on the percentage of migrants' self-employment at different centiles of ECI2 (EU sub-sample)



(a) EDU_GAP2, Total (b) UNR_GAP, Total (c) INC_GAP, Total



(d) EDU_GAP2, Females (e) UNR_GAP, Females (f) INC_GAP, Females



(g) EDU_GAP2, Males (h) UNR_GAP, Males (i) INC_GAP, Males

Figure C.2: Marginal effects of integration indicators on the percentage of migrants' self-employment at different centiles of ECI2 (NEU sub-sample)

D A deeper analysis of the role of human capital

We have seen in the main text that for the sub-sample of EU migrants, higher educational integration (higher EDU_GAP2) is an obstacle to total self-employment at high levels of ethnic concentration. However, it does not affect male self-employment independently of ethnic concentration, nor does it affect female self-employment. For the sub-sample of NEU migrants, increasing educational integration (higher EDU_GAP2) encourages total, male and female self-employment, particularly at low levels of ethnic concentration.

In this Appendix, we aim to delve deeper into the role of human capital by examining the impact of primary (EDU_GAP1) and tertiary (EDU_GAP3) educational gaps on migrants' self-employment.[8] Our estimates (Tables D.1 and D.2) reveal highly heterogeneous responses to migrants' self-employment. To be concise, we comment on these results, looking at the marginal effects of the three educational gaps on SE.

Figures D.1 and D.2 report the average marginal effects of the three educational gaps on self-employment for EU and NEU migrants, respectively. To better appreciate the differences across all educational level gaps, Figures D.1 and D.2 also report the marginal effects of EDU_GAP2 for EU and NEU migrants already shown in Figures 5 and 6, respectively. The results are striking.

Looking at total EU migrants, EDU_GAP1 does not affect SE (Panel (a) of Figure D.1), while EDU_GAP3 exerts a positive effect on SE above the 60$^{\text{th}}$ centile (Panel (c) of Figure D.1), increasing at higher levels of ECI1. Notice that this result is the opposite of what was found for EDU_GAP2, as depicted in Panel (b) of the same figure. As far as female EU migrants are concerned (Panels (d)-(f) of Figure D.1), no statistically significant role emerges for any educational gap. The results regarding male EU migrants are particularly interesting, as they suggest a U-shaped relationship between migrants' educational level and their self-employment. Indeed, the overall effect of EDU_GAP1 is positive up to the 80$^{\text{th}}$ centile of ECI1, whereas the overall effect of EDU_GAP3 is positive above the 45$^{\text{th}}$ centile of ECI1. We interpret this result as suggesting the existence of distinct motivations to enter self-employment: low-skilled migrants are likely to become self-employed out of necessity to escape unemployment; conversely, high-skilled migrants are likely to become self-employed out of opportunity, as they are better equipped to navigate the complexities of business-related issues. In this interpretation, ethnic concentration plays a crucial role.

Moving to the NEU sample, total migrants' self-employment is neither affected by EDU_GAP1 nor by EDU_GAP3 (Panels (a) and (c) of Figure D.2); EDU_GAP2 positively influences it below the 60$^{\text{th}}$ centile of ECI1 (Panel (b) of Figure D.2). As regards females, only EDU_GAP2 encourages SE below the 80$^{\text{th}}$ centile of ECI1, while neither EDU_GAP1 nor EDU_GAP3 influences SE. Finally, the behaviour of male NEU migrants is similar to that of their EU counterparts regarding EDU_GAP3, albeit at higher values of ECI1 (above the 75$^{\text{th}}$ centile compared with the 45$^{\text{th}}$ centile for EU). Still, it is different for the primary and secondary educational gaps. Indeed, the total effect of EDU_GAP1 on SE is positive up to the third quartile of the ECI1 distribution, while the total effect of EDU_GAP2 turns out from positive (below the 40$^{\text{th}}$ centile) to negative (above the 80$^{\text{th}}$ centile).

[8] More precisely, we consider ISCED levels 0-2 corresponding to "Less than primary, primary and lower secondary education" and ISCED levels 5-8 corresponding to "Tertiary education".

Table D.1: The effect of ethnic concentration (ECI1) and integration indicators (EDU_GAP1, EDU_GAP2 and EDU_GAP3) on the percentage of self-employment (EU sub-sample)

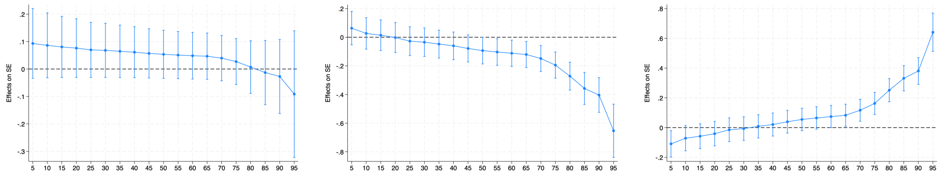
		Total			Females			Males	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI1	-0.009362***	-0.008373***	-0.005915***	-0.004203***	-0.004874***	-0.003099***	-0.010472***	-0.005520***	-0.002539*
	[0.001661]	[0.001090]	[0.000710]	[0.001188]	[0.001194]	[0.001158]	[0.002014]	[0.001383]	[0.001342]
EDU_GAP1	0.127941			-0.097418			0.355680***		
	[0.106986]			[0.079284]			[0.097085]		
EDU_GAP1*ECI1	-0.000100			0.000104*			-0.000243***		
	[0.000102]			[0.000062]			[0.000089]		
EDU_GAP2		0.198594**			0.212050*			-0.076511	
		[0.090546]			[0.113982]			[0.088817]	
EDU_GAP2*ECI1		-0.000387***			-0.000196**			-0.000078	
		[0.000078]			[0.000084]			[0.000076]	
EDU_GAP3			-0.251825***			0.002294			-0.037697
			[0.065889]			[0.107030]			[0.094203]
EDU_GAP3*ECI1			0.000406***			0.000009			0.000191**
			[0.000051]			[0.000106]			[0.000092]
POP	0.000003***	0.000002***	0.000003***	0.000002***	0.000002***	0.000002***	0.000004***	0.000003***	0.000003***
	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]
GDPPC	0.000140**	0.000034	0.000082	-0.000006	-0.000031	-0.000008	0.000201***	0.000047	0.000009
	[0.000069]	[0.000060]	[0.000065]	[0.000063]	[0.000062]	[0.000061]	[0.000074]	[0.000070]	[0.000073]
WAGE	-0.000456***	-0.000195**	-0.000207**	-0.000097	-0.000068	-0.000029	-0.000424***	-0.000005	-0.000009
	[0.000101]	[0.000086]	[0.000086]	[0.000103]	[0.000100]	[0.000107]	[0.000114]	[0.000113]	[0.000112]
CREDIT	0.004769	-0.017892	-0.008544	-0.032877***	-0.037918***	-0.035370***	0.003908	-0.027058*	-0.033497**
	[0.012420]	[0.011180]	[0.012322]	[0.012278]	[0.012245]	[0.012506]	[0.013759]	[0.013836]	[0.014605]
RQI	-2.433851***	-1.999708**	-2.065733**	-0.241203	-0.242811	0.245080	-2.784294***	-2.126632**	-2.161109**
	[0.892043]	[0.812248]	[0.858302]	[0.889588]	[0.874548]	[0.911342]	[0.937169]	[0.987547]	[0.992567]
Observations	178	182	183	142	142	143	158	160	161
R-squared	0.838992	0.921092	0.915251	0.705771	0.705683	0.923012	0.894756	0.898617	0.918041

Notes: Outlier robust regressions. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Country and year fixed effects are included but not reported. The dependent variable is the percentage of migrants' self-employment on total employment. EDU_GAP1 is the primary school education gap, EDU_GAP2 is the secondary school education gap (see also Table 1 in the main text), EDU_GAP3 is the tertiary school education gap. See the main text for more details.

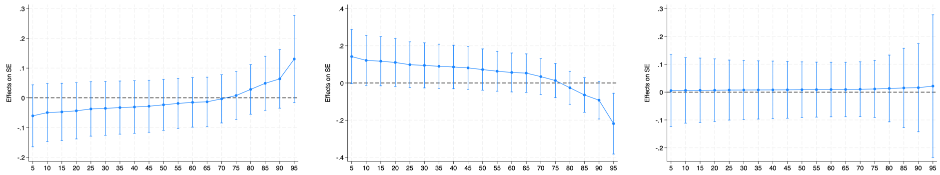
Table D.2: The effect of ethnic concentration (ECI1) and integration indicators (EDU_GAP1, EDU_GAP2 and EDU_GAP3) on the percentage of self-employment (NEU sub-sample)

		Total			Females			Males	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI1	0.000002	-0.000348	-0.000233	-0.000206	-0.000443**	-0.000541**	-0.001140**	-0.000591*	-0.001133**
	[0.000296]	[0.000232]	[0.000497]	[0.000263]	[0.000194]	[0.000246]	[0.000546]	[0.000315]	[0.000512]
EDU_GAP1	-0.062287			-0.061321			-0.176721***		
	[0.045803]			[0.044546]			[0.049916]		
EDU_GAP1*ECI1	0.000009			0.000015			0.000088***		
	[0.000019]			[0.000022]			[0.000027]		
EDU_GAP2		0.096998**			0.089382**			0.148984**	
		[0.042945]			[0.042738]			[0.060554]	
EDU_GAP2*ECI1		-0.000047**			-0.000022			-0.000103***	
		[0.000019]			[0.000014]			[0.000030]	
EDU_GAP3			-0.039210			-0.038973			-0.000285
			[0.049891]			[0.050870]			[0.054156]
EDU_GAP3*ECI1			0.000052			0.000013			0.000056**
			[0.000032]			[0.000013]			[0.000025]
POP	0.000001***	0.000001***	0.000001***	0.000001**	0.000001*	0.000000	0.000002***	0.000001***	0.000002***
	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]	[0.000000]
GDPPC	-0.000367***	-0.000355***	-0.000355***	-0.000467***	-0.000460***	-0.000443***	-0.000307***	-0.000316***	-0.000309***
	[0.000053]	[0.000051]	[0.000053]	[0.000049]	[0.000048]	[0.000049]	[0.000064]	[0.000072]	[0.000075]
WAGE	0.000082	0.000093	0.000115	0.000088	0.000081	0.000077	-0.000123	-0.000041	-0.000082
	[0.000073]	[0.000066]	[0.000075]	[0.000084]	[0.000077]	[0.000082]	[0.000091]	[0.000099]	[0.000108]
CREDIT	-0.056234***	-0.057761***	-0.060678***	-0.076936***	-0.078975***	-0.079975***	-0.030078**	-0.038500***	-0.039632***
	[0.010277]	[0.009256]	[0.009737]	[0.009504]	[0.009120]	[0.009316]	[0.012185]	[0.013315]	[0.013729]
RQI	-1.496990**	-1.904916***	-1.718289**	-1.144186	-1.293155*	-1.318031*	-2.711016***	-3.382694***	-3.577352***
	[0.707950]	[0.692873]	[0.703473]	[0.709553]	[0.687858]	[0.715090]	[0.848240]	[0.952719]	[0.987334]
Observations	224	236	234	171	171	171	215	223	222
R-squared	0.965672	0.968005	0.967394	0.962602	0.964359	0.962942	0.973754	0.965780	0.964164

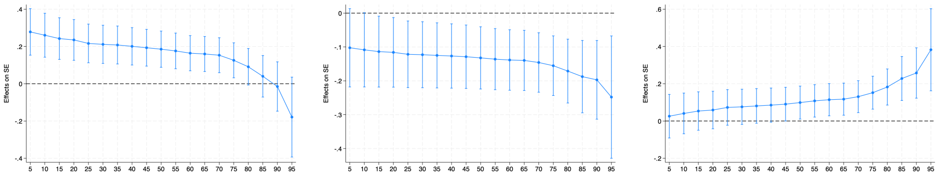
Notes: Outlier robust regressions. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Country and year fixed effects are included but not reported. The dependent variable is the percentage of migrants' self-employment on total employment. EDU_GAP1 is the primary school education gap, EDU_GAP2 is the secondary school education gap (see also Table 1 in the main text), EDU_GAP3 is the tertiary school education gap. See the main text for more details.



(a) EDU_GAP1, Total (b) EDU_GAP2, Total (c) EDU_GAP3, Total

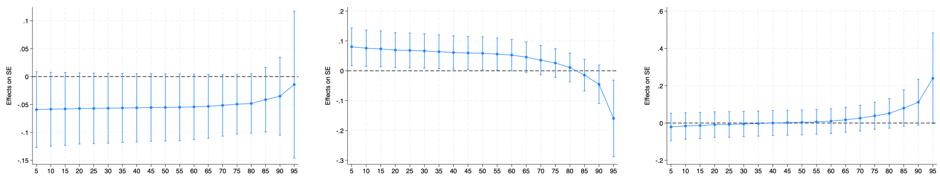


(d) EDU_GAP1, Females (e) EDU_GAP2, Females (f) EDU_GAP3, Females

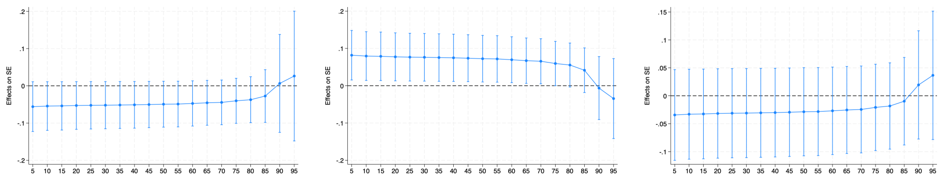


(g) EDU_GAP1, Males (h) EDU_GAP2, Males (i) EDU_GAP3, Males

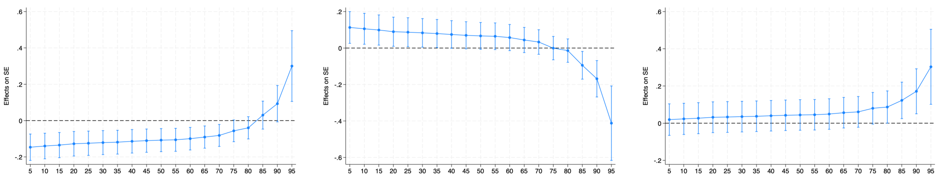
Figure D.1: Marginal effects of EDU_GAP1, EDU_GAP2 and EDU_GAP3 on the percentage of migrants' self-employment at different centiles of ECI1 (EU sub-sample)



(a) EDU_GAP1, Total (b) EDU_GAP2, Total (c) EDU_GAP3, Total



(d) EDU_GAP1, Females (e) EDU_GAP2, Females (f) EDU_GAP3, Females



(g) EDU_GAP1, Males (h) EDU_GAP2, Males (i) EDU_GAP3, Males

Figure D.2: Marginal effects of EDU_GAP1, EDU_GAP2 and EDU_GAP3 on the percentage of migrants' self-employment at different centiles of ECI1 (NEU sub-sample)

REGION

The Journal of ERSA
Powered by WU

Volume 13, Number 1, 2026, 153–170
DOI: 10.18335/region.v13i1.658

journal homepage: region.ersa.org
ISSN: 2409-5370

Bank Access and Cultural Entropy across the Regions of India

Govindapuram Suresh[1]

[1] Krea University, Sri City, Andhra Pradesh, India

Received: 11 December 2025/Accepted: 3 March 2026

Abstract. Holding a bank account represents access to an important bank service for achieving the formal financial inclusion of an individual. At present, people living in different regions of India seem to experience different levels of access to this basic bank service and are therefore experiencing different levels of financial market exclusion. This paper studies to what extent social and cultural barriers hinder effective banking access across the regions of India. Normative caste-based exclusion can be modified by the rigidity of the local context in following inherited norms. This overall cultural rigidity of the context can be termed cultural openness. The Culture Based Development (CBD) approach to quantify cultural openness through the Cultural Entropy measure is employed here. Using data from the National Family Health Survey (NFHS) and National Sample Survey Organization (NSSO), I conduct a probit model estimation to test how Cultural Entropy (CE) modifies the caste discrimination effect in access to bank services. I find that there is strong evidence that regional Cultural Entropy modifies significantly the individual caste discrimination experience across the regions of the country.

Key words: bank services, regional accessibility, Culture Based Development, Cultural Entropy, India

1 Introduction

Having a formal bank account is still a major issue in most developing countries. In India, still people lack access to formal bank account even though India has demonstrated substantial progress in financial inclusion, with account ownership rising from 77.5% in 2021 to 89% in 2024 (DRG 2025). A persistent challenge is the high proportion of inactive accounts (16%), which remains considerably above the global average of 6%. Opening a bank account is important but using it is more important. However, most alarming is the fact that the improvement in bank access is not equitably distributed across space and different regions remain left behind in terms of bank access in comparison to others that manage more successfully to integrate their incumbent population.

Generally, the individual differences in access to bank services have been well studied. The reasons for poor access relate to factors such as distrust towards the formal institutions, gender-based mobility constraints, and low financial literacy, which all contribute to self-exclusion and limited bank account use. These factors reinforce reliance on informal financial networks (money lenders and shop keepers) despite high levels of account ownership. The Government of India has launched many schemes and programmes (such as Jan Dhan Yojana, MUDRA, etc.) that aim to correct this self-exclusion behaviour.

There is however another aspect of the exclusion from bank services, and this is the discrimination by society and the system towards individuals from certain social groups. Numerous studies document the role of discrimination based on caste, religion, gender, race and other socio-economic variables in determining the access to bank accounts at the household level (Kumar, Venkatachalam 2019, Ghosh, Vinod 2017, Govindapuram et al. 2023).

However, none of the above can answer the question why the same social groups experience more discrimination in some regions of India and are less discriminated and better included in others. To address this gap, the present study hypothesizes that local cultural context affects financial inclusion by influencing how strongly individuals adhere to inherited social norms versus their openness to change. This regional cultural openness is hypothesized to moderate the effect of caste discrimination and thus results in unequal levels of financial system exclusion across the regions of India.

To operationalise this hypothesis, the paper employs the Culture Based Development (CBD) paradigm (Tubadji 2025a), and specifically its notion of Cultural Entropy (CE) (Tubadji 2025b). According to CBD, the local cultural openness can be understood in terms of wiring of the brain of the local population through the means of local aesthetic experience of either inherited (cultural heritage) nature, or experiences and participation including ideas and attitudes shaped contemporaneously by the society (i.e. what CBD terms living culture). CBD measures the overall open-mindedness of the local milieu as a balance between the level of exposures to cultural heritage (CH) and level of exposures to living culture (LC) experiences by the population. If both CH and LC are evenly present as experience locally, the locality will not exert cultural bias, and the choice locally will be rational. However, an imbalance in the proportion between CH and LC will generate cultural bias among the different social groups locally. A greater imbalance between CH and LC is associated with a higher likelihood of cultural discrimination across social groups. The balance between CH and LC is quantified using the Shannon Entropy Index for unrelated variety between CH and LC as components of the local cultural capital i.e. the power of the local cultural milieu to bias choice on the regional level. Hence, I will test whether the Cultural Entropy across the regions of India modifies the average individual effects of class discrimination towards castes, thus creating inequitable financial inclusion across the regions of India.

The rest of the paper is structured as follows. Section 2 presents my literature review of financial access in India and the role of caste and local culture from a CBD perspective. Section 3 discusses my data. Section 4 offers my estimation strategy. Section 5 presents the results and Section 6 concludes.

2 Literature Review

This section is divided into three main sub-topics. It starts with an overview of the literature on access to bank services in India, continues with an overview of the literature on caste-based discrimination in the country, and finishes with the Cultural Based Development (CBD) paradigm. This helps to distinguish between the caste norms and the local cultural variation across the regions of India that modifies the stickiness (i.e. severity) with which the caste discrimination affects the individual users of the bank services.

2.1 Access to Bank Services in India

Inclusive economic development is one of the major objectives of the Indian government. After the independence in 1947 it started with the Five-year plans, nationalization of banks and other private sector entities, which clearly shows the efforts of the government towards inclusion and Sustainable Development Goals (SDGs).

Access and usage of banking is an important public good, that enables all aspects of finance such as savings, credit, payments, insurance and risk management. It has a significant impact on reducing poverty and inequality and advances inclusive growth (Demirgüç-Kunt, Klapper 2013). India experienced a steep progress in financial inclusion, but the barriers remain the same such as geography, economic status, caste, religion, gender, and digital divide (Sarma, Pais 2011, Govindapuram et al. 2023).

India's financial inclusion journey has evolved through several phases: bank nationalization in 1969, the expansion of rural branch networks (between 1970s to 1990s), the introduction of self-help groups (SHGs) and microfinance (2000s), Aadhaar-enabled direct benefit transfers (2010s), and the recent digital infrastructure revolution enabled by India Stack, UPI, and Jan Dhan Yojana (IMF 2021). Although account ownership has increased dramatically, especially after PMJDY (Pradhan Mantri Jan Dhan Yojana), account usage, quality of services, and consumer trust remain uneven across states, gender groups, and rural-urban boundaries (RBI 2020, Data For India 2023).

Access to banking in India has expanded through nationalisation-driven bank branch growth that reduced rural poverty (Burgess, Pande 2005), though liberalisation later slowed progress, sustaining exclusion (Sarma, Pais 2011). Microfinance and SHG-Bank Linkage models improved outreach but remained uneven and limited in transformative impact (Harper 2002).

Digital public infrastructure has significantly reshaped financial access in India, with Aadhaar reducing documentation barriers and enabling low-cost, paperless KYC processes. Together, Aadhaar, e-KYC, Jan Dhan accounts, and UPI created an interoperable digital architecture which has expanded access and reduced transaction costs for underserved populations (Mehta 2021). Aadhaar-enabled Payment Systems further extended last-mile service delivery through business correspondents, while UPI's rapid scaling beyond card and wallet transactions reflected widespread adoption of low-cost digital payments (NPCI 2022). However, digital divides persist due to gaps in smartphone ownership, gendered access to technology, and uneven digital literacy, limiting effective participation among women, elderly users, and rural households (Roy, Singh 2025).

Socio-economic factors remain strong predictors of financial inclusion. Income, education, and social identity shape account ownership and usage (Demirgüç-Kunt, Klapper 2013). Gender constraints, caste-based discrimination, and rural infrastructural deficits continue to restrict engagement with formal banking services (Jayachandran 2021, Thorat 2010, Sarma, Pais 2011, Govindapuram et al. 2023).

Supply-side constraints significantly shape the extent and quality of financial service delivery in India. High operational costs make rural branches financially unviable for many commercial banks, particularly in sparsely populated or low-income regions (RBI 2020). This has historically limited branch expansion despite substantial demand for formal financial services. The Business Correspondent (BC) model, introduced to enhance last-mile outreach at lower cost, has delivered mixed results. While it has improved access in remote areas, many BC agents face challenges such as low and irregular commissions, inadequate training, liquidity shortages, and operational risks associated with connectivity issues or fraud (Kochhar 2020). Product design also remains a structural weakness. Standardised savings and credit products are poorly suited to the irregular cash flows of low-income households, whose financial needs centre on frequent small deposits, emergency loans, and informal savings mechanisms (Collins et al. 2009). Weak consumer protection frameworks and inadequate grievance redressal systems further reduce trust in formal financial institutions, a concern exacerbated by rising cyber-fraud in digital transactions, particularly among digitally inexperienced populations (RBI 2020).

Policy interventions have been critical in shaping India's financial inclusion trajectory. The Pradhan Mantri Jan Dhan Yojana (PMJDY), launched in 2014, led to one of the fastest expansions of account ownership globally, increasing coverage from about 53 percent to over 80 percent of adults within a few years (World Bank 2020). However, studies highlight that a considerable share of these accounts remains inactive or maintains zero balances, raising questions about sustained usage and the depth of inclusion achieved (RBI 2020). The integration of Jan Dhan accounts with Aadhaar and mobile phones (the JAM trinity) substantially strengthened the implementation of Direct Benefit Transfers (DBT). This system streamlined welfare delivery, reduced leakages, and increased account activity, particularly among low-income households (Chattopadhyay, Roy 2021). Financial literacy initiatives by the Reserve Bank of India and public agencies aim to improve financial knowledge, responsible usage, and digital awareness. Yet their effectiveness varies considerably across regions, gender, and socio-economic groups (Aggarwal et al. 2022). Regulatory innovations such as licensing Small Finance Banks and Payments Banks and promoting fintech-bank collaborations have diversified the financial ecosystem and expanded the provision of technology-driven solutions for underserved populations (Shah 2022).

Overall, the literature indicates that India has made substantial progress in expanding access to banking and financial services, particularly through digital infrastructure and state-led interventions. However, persistent gaps in usage, quality of service delivery, and equitable access remain. Scholars broadly agree that the central challenge has shifted from simply opening accounts to achieving meaningful financial inclusion that enhances financial resilience, economic participation, and long-term welfare for all citizens.

2.2 Caste Discrimination in India

Caste discrimination in India has been extensively studied (e.g. Thorat, Attewell 2007, Fuller 1996, Jodhka 2015) across the social sciences, with scholars repeatedly emphasizing that caste remains one of the most enduring and influential structures shaping social, economic, and political life. Early social sciences work positioned caste as a rigid hierarchy rooted in ideas of purity, pollution, and hereditary occupation (Dumont 1970). Although modernization, urbanization, and constitutional safeguards were expected to undermine caste boundaries, researchers argue that caste has shown remarkable resilience and adaptability, continuing to influence life chances in both rural and urban contexts (Fuller 1996, Jodhka 2015). The literature consistently demonstrates that caste persists not only as a cultural identity but also as a structural determinant of inequality.

One of the most widely documented domains of caste discrimination is economic participation. Empirical work shows that caste influences access to jobs, wages, and economic mobility. Using audit experiments, Thorat, Attewell (2007) provided compelling evidence that job applicants with Dalit-identifying (lower caste) names received significantly fewer callbacks than upper-caste applicants with identical qualifications. Labour market studies build on these findings to demonstrate that wage gaps between Dalits and upper castes persist even after accounting for education, experience, and location, suggesting that the disparities largely reflect discrimination rather than productivity differences (Madheswaran, Attewell 2007). Earlier research had already highlighted occupational segregation, with Dalits being concentrated in manual, low-wage, and stigmatized forms of labour (Banerjee, Knight 1985). The persistence of landlessness and unequal asset ownership further reinforces caste hierarchies. Rawal (2008) notes that Dalits remain disproportionately represented among landless households, and this structural deprivation limits their bargaining power and economic autonomy.

Education, often conceptualized as a route to mobility, is similarly shaped by caste-based barriers. Research on primary and secondary schooling, documents subtle and overt forms of discrimination, including biased teacher attitudes, segregation within classrooms, bullying, and lowered expectations for Dalit students (Nambissan 2010). Scholars find that such experiences affect learning outcomes, self-confidence, and school retention. In higher education, discrimination becomes more institutionalized. Sukumar (2013) observes that Dalit students in universities often face micro-aggressions, stereotyping, and social exclusion, undermining their academic performance and mental well-being. Although affirmative action policies have improved access to professional education and government employment, their impact is constrained by the persistence of prejudice and by the fact that many spaces of elite education remain dominated by upper castes (Deshpande, Yadav 2006).

Beyond economic and educational spaces, caste continues to shape everyday social interactions, neighbourhood patterns, and access to public goods. Ethnographic studies reveal that many villages still exhibit clear spatial segregation, with Dalits living in separate hamlets and facing restrictions on access to wells, temples, and common spaces (Shah et al. 2006). Urban spaces, long viewed as more anonymous and egalitarian, also show caste-based residential patterns. Sidhwani (2015) finds that informal rental markets often discriminate against Dalits, reinforcing patterns of exclusion that resemble rural marginalization. Everyday discrimination also manifests in social practices such as refusal to share food, reluctance to enter Dalit households, or denial of services by barbers, priests, and other service providers (Thorat, Newman 2012). While such practices may appear subtle, they reflect deep-rooted prejudices that shape social hierarchies.

Violence remains an extreme but persistent expression of caste inequality. The literature documents that crimes against Scheduled Castes often arise in contexts where Dalits challenge traditional hierarchies, such as attempts to access common land, inter-caste marriage, or assertion of political rights (Jangam 2021). Scholars argue that the threat or use of violence functions as a mechanism to reinforce caste dominance and discourage upward mobility (Vithayathil, Singh 2012). Despite legal safeguards such as the Scheduled Castes and Scheduled Tribes (Prevention of Atrocities) Act, implementation gaps and local power structures often limit justice for victims (Kapur 2020). Research shows that police response to caste-based violence is frequently inadequate, reflecting systemic bias and bureaucratic inertia.

The political sphere illustrates both the transformation and continuity of caste. Democratic participation has created new opportunities for lower-caste groups to mobilize and secure representation. Jaffrelot (2003) argues that the rise of Dalit and backward-caste political movements have altered India's political landscape, enabling marginalized groups to assert their rights and gain visibility. However, scholars caution that symbolic political representation does not always translate into substantive empowerment, especially in rural areas where upper-caste elites continue to dominate local governance (Pai 2002). Even where Dalits hold elected positions, social pressure and economic dependence can constrain their autonomy.

Recent literature expands the study of caste into modern and digital domains. With the proliferation of technology, caste identities have acquired new forms of expression. Studies on matrimonial websites show that caste remains a central criterion in marriage selection, demonstrating the continued relevance of endogamy (Banaji 2018). Social media has also become a space where caste prejudice is reproduced through trolling, abuse, and hate speech targeted at Dalits and other marginalized groups (Paik 2022). Online platforms, often thought to be democratizing spaces, thus also reflect offline hierarchies. Scholars argue that digital environments magnify certain caste biases because anonymity emboldens discriminatory behaviour that might be restrained in face-to-face settings.

Across this wide body of literature, a consistent theme emerges. Although India has enacted constitutional protections, progressive legislation, and affirmative action policies, caste discrimination continues to shape opportunities, interactions, and institutional outcomes. The nature of discrimination has evolved, shifting from overt exclusion to more subtle, indirect, or institutionalized forms, but its effects remain profound. Caste operates not only as a social identity but as a system of economic and political power. Researchers increasingly highlight the need to analyse caste as a dynamic, context-dependent force that interacts with class, gender, region, and emerging technologies. The growing recognition that caste persists even in spaces assumed to be modern and meritocratic, such as universities, corporate workplaces, and digital platforms, which suggests that it remains a durable form of inequality requiring multidimensional interventions.

2.3 Culture Based Development (CBD): Cultural Entropy

The concept of culture-based development (CBD) has emerged as a distinctive analytical framework that foregrounds the role of culture, its norms, values, symbolic systems, and inherited social structures, in shaping economic choices and regional development trajectories. Classical development theories tended to privilege capital accumulation, technological progress, and institutional design while treating culture as peripheral. However, contemporary scholarship argues that culture acts as a foundational determinant of economic behaviour and institutional performance, thereby shaping long-term development patterns (North 1990, Sen 1999). Culture influences the aspirations, preferences, and constraints faced by individuals and communities, and these, in turn, guide processes of social cooperation, resource allocation, and innovation. Thus, culture does not merely accompany development; it actively structures it.

Throsby (2001) was among the first to conceptualise cultural capital as an asset with both economic and intrinsic value. His framework highlights that cultural practices, artistic traditions, and shared meanings contribute to human well-being and economic productivity in ways that are not fully captured by market transactions. This view aligns with the influential theory of cultural capital by Bourdieu (1986) which identifies cultural knowledge, competencies, and socialized behaviours as critical resources that shape educational attainment, labour-market opportunities, and social mobility. Combined, these perspectives shift the analysis of development from material resources alone to the distribution and reproduction of cultural resources across social groups.

Building on this foundation, Tubadji (2013, 2015) formalised the Culture-Based Development (CBD) model, which posits that culture functions as a contextual "background factor" influencing economic choices. The CBD model emphasizes that cultural biases, shaped by history, collective memory, social norms, and identity, play an important role in influencing individual behaviour, innovation, institutional efficiency, and regional economic development. In this framework, Tubadji (2013, 2015) distinguishes between"cultural bias", which affects social preferences, and "cultural capital", which reflects a community's capacity to generate economic value. According to this model, regions characterized by supportive cultural norms, such as trust, learning orientation, openness, and cooperation, tend to experience higher levels of social cohesion, better governance, and stronger economic performance. Conversely, regions marked by exclusionary cultural norms, rigid hierarchies, or discriminatory practices often face persistent underdevelopment despite their economic potential.

The CBD paradigm developed by Tubadji offers a robust theoretical and empirical framework for understanding the role of culture in shaping socio-economic outcomes (Tubadji 2012, 2013, 2025a). Unlike conventional approaches that treat culture as a residual factor or approximate it through narrow indicators such as religion, language, or ethnicity, CBD reconceptualises culture as a foundational element of socio-economic organisation. In this framework, culture is understood as a proto-institution: a locally embedded system of shared attitudes that shapes how individuals and communities assign value, form preferences, and interpret choices. Rather than directly constraining behaviour, culture operates as a cognitive and behavioural code that precedes formal institutions and structures the valuation processes underlying all socio-economic decisions (Tubadji 2021).

To operationalise this cultural valuation system empirically, CBD identifies two fundamental and exhaustive components of cultural capital: Cultural Heritage (CH) and Living Culture (LC) (Tubadji 2021, 2025a). Cultural Heritage represents inherited, identity-preserving attitudes that are transmitted across generations. These attitudes reflect deeply rooted symbolic meanings, norms, and traditions that provide continuity and stability within a locality. Living Culture, by contrast, captures contemporaneously formed and adaptive attitudes that emerge through ongoing cultural participation, aesthetic engagement, and social interaction. LC reflects the dynamic aspect of culture, responding to changing economic, social, and institutional environments. Together, CH and LC capture culture's dual role as both a mechanism of persistence and a source of adaptation, avoiding the under-specification that arises when culture is measured through isolated or arbitrarily selected indicators. Building on this dual structure, CBD introduces the concept of Cultural Entropy (CE), defined as the balance between Cultural Heritage and Living Culture (Tubadji 2025a,b). Cultural Entropy does not measure the quantity of culture but its internal configuration. It indicates whether a local cultural system is dominated by preservation-oriented attitudes, adaptation-oriented attitudes, or maintains a balanced combination of both. CE captures the degree of openness, coherence, and flexibility within a cultural system and predicts the direction and strength of behavioural bias at the local level. Empirical evidence suggests that CE often provides a more stable and interpretable predictor of socio-economic outcomes than CH or LC considered independently, particularly in contexts requiring cooperation, institutional adoption, or innovation (Tubadji 2013). A key empirical strength of the CBD framework lies in its approach to cultural measurement. Rather than relying on normative assumptions about which cultural traits are relevant, CBD constructs CH and LC using a broad set of cultural participation and behavioural indicators, aggregated through dimensionality-reduction techniques such as principal component analysis (PCA)

(Tubadji 2021). This value-neutral methodology minimises both under-measurement and over-interpretation, addressing two persistent challenges in empirical cultural research: the reduction of culture to simplistic proxies and the conflation of cultural effects with socio-economic characteristics.

CBD has been applied across diverse empirical settings, including regional economic performance, innovation, migration, and institutional quality (Tubadji 2013, Tubadji, Nijkamp 2015). These studies demonstrate that culture, when rigorously operationalised, exerts an independent and systematic influence on behavioural and spatial outcomes. This makes CBD particularly relevant for analysing ecological public goods and collective action problems, where outcomes depend on shared norms, cooperation, and long-term commitment.

In the Indian context, where cultural diversity is extensive and cannot be adequately captured through single dimensions such as caste or religion, the CBD framework provides a uniquely comprehensive approach to measuring cultural milieu (Tubadji et al. 2024). By capturing both inherited and adaptive cultural attitudes and summarising their balance through cultural entropy, CBD enables a systematic analysis of how local cultural capital shapes access to formal credit across regions in India.

Empirical literature reinforces these claims. Putnam's comparative study of Italian regions (Putnam 1993) demonstrated that social capital rooted in shared cultural norms of trust and reciprocity strongly predicts institutional performance and developmental success. Similar findings have been observed across diverse settings, where cultural contexts shape the effectiveness of local governance, collective action, and public-goods provision. In the Indian context, prevailing cultural structures, especially caste hierarchies, gender norms, and localised community identities, have been shown to influence educational outcomes, economic aspirations, labour-market participation, and entrepreneurship (Deshpande 2011, Basu 2017). These patterns illustrate that cultural norms can both empower communities and entrench disadvantage, depending on their nature and historical evolution.

CBD literature also highlights the economic potential embedded in cultural assets. Throsby (2010) and UNESCO (2013) argue that cultural heritage, creative industries, and traditional knowledge systems can serve as drivers of local development by generating employment, promoting tourism, and fostering innovation rooted in local identity. Cities that have invested in cultural clusters, heritage conservation, and creative economies often observe positive spillover effects on urban regeneration and social cohesion. However, critical scholarship warns that the commodification of culture must be approached with caution. Without inclusive governance, cultural-economy policies may lead to gentrification, exploitation of indigenous communities, or the reinforcement of existing hierarchies (Harvey 2001).

Taken together, the literature on culture-based development challenges the notion that development is culturally neutral. Instead, it reveals that cultural contexts shape preferences, institutions, and opportunities, influencing both micro-level decision-making and macro-level developmental trajectories. The CBD approach integrates cultural analysis with economic modelling, offering a multidimensional explanation of why regions with similar economic endowments can experience divergent growth paths. As globalisation and digitalisation reshape social identities and cultural interactions, culture-based development frameworks provide valuable analytical tools for understanding how deep-rooted beliefs and evolving cultural dynamics interact with economic transformations.

I have chosen to employ the CBD paradigm in my research, due to its comprehensive definition and measurement of local cultural as a factor for behaviour and choice. Namely, I choose to employ CBD in order to account more comprehensively for the cultural milieu effect beyond the particular caste effect that the literature already knows about. This can reveal a more complete picture about the cultural bias on the access to financial services in the regions of India.

3 Data

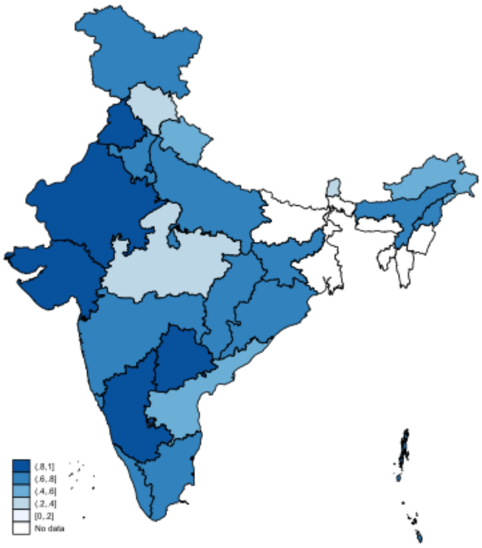
To operationalise my model, I use a rich dataset of about 630,000 individuals. The data is obtained from two main data sources: (i) the National Sample Survey Organisation (NSSO) 71st round on the Survey on Literacy and Culture of India 2014, used to operationalise regional Cultural Entropy and (ii) the National Family Health Survey (NFHS) 5th round (2019-21), used to obtain the information on bank access, demographic controls and among them the caste belonging of the individual. Both data sets are sample surveys of the Government of India, and are publicly available.

The main outcome variable is the access to bank account. In the NFHS Survey, respondents are asked *do you have a bank account?* The variable is a binary in nature (yes or no). This justifies the use of a probit estimator for testing our hypothesis. The 0-1 coding of the outcome provides clear identification of who has access and who does not have access, i.e. is excluded from access to basic bank services.

The main explanatory variables reflect differences in the local cultural milieu, particularly the persistence of caste norms and the extent to which they continue to shape social interactions and behaviour. From the Survey on Literacy and Culture of India (NSSO - 2014), I utilize 18 variables capturing diverse forms of cultural participation and practices related to inherited cultural traditions. These include activities such as reading books, attending musical events, undertaking travel motivated by religious beliefs, and listening to religion-related mass media programmes. These variables are employed to construct composite indicators – factor variables – that represent the local cultural context,commonly referred to as the cultural milieu. In the framework of Culture-Based Development (CBD), the local cultural milieu is studied as a source of cultural bias on individual choices. The power of the milieu to create this bias is termed local cultural milieu. The magnitude of this power is defined as a product of the composition of cultural capital, specifically by the level of its two components: cultural heritage (CH; inherited local attitudes) and living culture (LC; currently shaped cultural attitudes). The degree of stickiness, flexibility to change of the local milieu is termed Cultural Entropy as measured in terms of the balance between the level of CH and LC locally. CBD calculates the balance between CH and LC employing the Shannon Entropy index of unrelated variety between CH and LC in the composition of the local milieu. The Cultural Entropy (CE) index can vary between 0 and 1, these being the only mathematically possible extremes for Shannon entropy for two classes. When CE is close to 0, the local cultural milieu exudes very little source of bias on local choices and the latter are optimal and rational locally or at least do not vary across space in the degree of their cultural bias in the country that follows the same formal laws. Hence, through principal component factor analysis I obtain the CH and LC for the regions of India and use them to calculate the CE for the regions of India (see Appendix A for the details). This CE will be my main explanatory variable. I will interact it with the individual caste, expecting that the caste-discrimination experienced by the individual will be moderated by the stickiness to tradition that the local cultural milieu exudes. The higher the Cultural Entropy, the lower the experienced discrimination due to the caste is expected to be, in line with the CBD paradigm theory and preceding research.

I have plotted these main dependent and independent variables on two maps. Figure 1 shows Cultural entropy across Indian states, while Figure 2 shows the banking access across Indian states. Both figures show significant inter-state heterogeneity and highlight the spatial variation in social-cultural diversity and financial penetration across India. Figure 1 shows that North-eastern states and large urban states have larger CE. It may be because of the more linguistic, ethnic and migrant populations. Further, Western and Southern states shows that there is a higher diversity probably due to opportunities in urban areas and the internal migration. Figure 2 shows that formal financial services are not equally distributes across Indian states. South and Western states have higher banking access as compared to other areas, it may be due the better educational levels and the stronger government policies. North-Eastern and some eastern states have low access, may be because of the geographic and institutional challenges.

Control Variables come from the NFHS. They include wealth of the individuals, num-



Source: DataMeet (State Boundary shapefiles); NFHS and NSSO data
Notes: authors' calculations

Figure 1: Cultural Entropy across Indian States

ber of children in the family, age of the household head, gender of the household head, location and the year of interview. More details on the descriptive statistics for each variable can be seen in Appendix B.

4 Estimation Strategy

My testable working hypothesis can be stated as follows:

> *H01*: The regional Cultural Entropy modifies the degree of caste discrimination experienced by an individual in terms of access to banking services in India.

To test this hypothesis I have to implement two steps. The first regards quantifying comprehensively the regional context using the CBD paradigm. The second comprises of testing the hypothesis. The estimation strategy for each step is detailed below.

First, I obtain the Cultural Entropy index according to the CBD paradigm procedures. All eighteen mono-dimensional variables available from the dataset are used in a principal component factor analysis. Thus, I obtain the two components of the local cultural capital: Cultural Heritage (CH) and Living Culture (LC). I use their sum and the percentage that each represents in the sum to calculate the Shannon Entropy index that gives the Cultural Entropy, as defined by CBD (see Tubadji 2025a).

Second, I use the following Probit model (Equation 1) to test my working hypothesis *H01*:

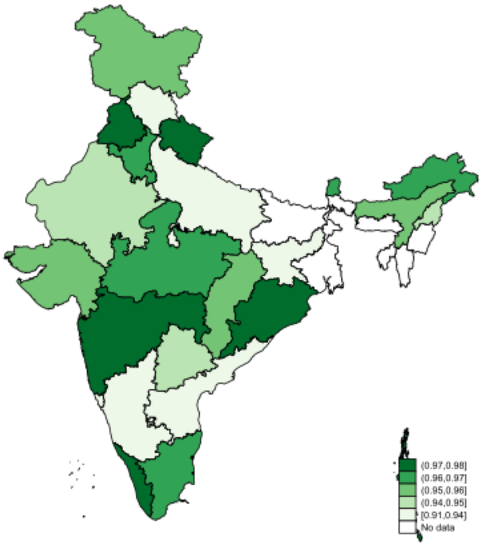
$$\text{Access\_Bank} = f(\text{Caste}, \text{Cultural Entropy}, \text{Caste*Cultural Entropy}, Z, \varepsilon) \qquad (1)$$

The components of the above stated model (Equation 1) are defined as follows:

Access_Bank – having a bank account (yes/no, 0 or 1)

Caste – caste of the individual

Cultural Entropy – the openness (i.e. stickiness) of the regional cultural context, expressed as the balance between CH and LC, calculated using a Shannon Entropy index, according to the CBD methodology (Tubadji 2025a,b).



Source: DataMeet (State Boundary shapefiles); NFHS and NSSO data
Notes: authors' calculations

Figure 2: Banking Access across Indian States

$\boldsymbol{Z}$ – a battery of individual and household control variables

ε – error term

The main test for hypothesis *H01* here consists in investigating whether the interaction between individual caste and regional Cultural Entropy is statistically significant. The direction of the effect is a matter of the particular case.

5 Results

Using the Probit model, I estimated step wise results. Before obtaining the parametric main results, I can clearly see in Table 1 that there are different strength and direction of the basic correlation between the individual castes and the Cultural Entropy of the regions. Also, the bank access is weakly and negatively associated with Cultural Entropy and wealth follows the same pattern but with higher magnitude. Apparently, I can expect to find that places that are more open and equitable as a local cultural milieu (having higher Cultural Entropy) end up concentrating poorer population which consequently has lower access to financial services due to one's economic exclusion.

Table 1: Correlation Coefficients

	Cultural Entropy	bank_acc	wealth index	caste_1	caste_2	caste_3	location
Cultural Entropy	1						
bank_acc	-0.0145	1					
wealth index	-0.0689	0.0705	1				
caste_1	0.0943	-0.0355	-0.2717	1			
caste_2	-0.0288	0.0224	0.0641	-0.401	1		
caste_3	-0.0404	0.0151	0.2701	-0.2568	-0.4053	1	
location	0.0149	0.0127	-0.477	0.133	-0.0308	-0.1204	1

Source: Authors calculations using NFHS and NSSO data

At a next step, I can proceed with the actual parametric hypothesis testing; in particular whether the caste discrimination in bank access is moderated by the local cultural openness (i.e., the cohesiveness and inclusivity-propensity of the local cultural milieu). To test this hypothesis, I use a probit model that clearly distinguishes and identifies exclusion by not having a bank account. The variable is 0-1 coded in the survey, hence it is also statistically appropriate to use a probit model in this instance. The results are presented in Table 2.

Table 2: Caste Discrimination for Bank Services Access, moderated by Regional Cultural Entropy – Probit model estimations

Variables	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
Cultural Entropy	-0.0911***	-0.149***	-0.0646***	-0.120***
	(0.0104)	(0.0216)	(0.0107)	(0.0225)
caste_1	-0.105***	-0.231***	-0.0564***	-0.140***
	(0.00853)	(0.0232)	(0.00866)	(0.0240)
caste_2	0.0951***	0.0852***	0.0565***	0.0300
	(0.00786)	(0.0200)	(0.00803)	(0.0207)
caste_3	0.0973***	0.0435*	-0.00415	-0.0593**
	(0.00908)	(0.0231)	(0.00955)	(0.0241)
wealth_index			0.157***	0.157***
			(0.00277)	(0.00277)
No_child			0.0179***	0.0179***
			(0.00394)	(0.00394)
Age_hh			0.00134***	0.00135***
			(0.000222)	(0.000222)
Sex_hh			-0.114***	-0.114***
			(0.00721)	(0.00721)
2.location			0.321***	0.320***
			(0.00786)	(0.00786)
2020.year_interview			0.0154**	0.0115*
			(0.00682)	(0.00692)
2021.year_interview			0.143***	0.138***
			(0.00754)	(0.00765)
inter_ce_caste_1		0.179***		0.120***
		(0.0312)		(0.0324)
inter_ce_caste_2		0.0149		0.0395
		(0.0275)		(0.0286)
inter_ce_caste_3		0.0815**		0.0829**
		(0.0323)		(0.0337)
Constant	1.749***	1.788***	1.129***	1.168***
	(0.00921)	(0.0157)	(0.0205)	(0.0245)
Observations	603,348	603,348	603,348	603,348

Source: Authors calculations using NFHS and NSSO data
Notes: Dependent Variable = Having access to Bank account, Independent variables include: Cultural Entropy, Caste 1= STs (Schedule Tribes), Caste 2 = SCs (Schedule Castes) and Caste 3 = OBCs (Other Backward Classes) and General caste category is the base category. wealth_index = Wealth Index of the Individuals (which measures the income across the individuals). No_child = Number of Children in the family. Age_hh = Age of the Household head. Sex_hh = Gender of the Household Head, it may male or female. Location is Urban or Rural. NFHS Survey Interviews – 2019, 2020 and 2021, 2019 is the base category. I considered this Year of the interviews to check whether COVID-19 pandemic have some influence on the selected variable. Further, I have done interaction between the Cultural Entropy (CE) with the all the Caste categories step wise. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2 gives the estimation of probit results for four alternative specifications. Specification 1 tests for the presence of association between individual caste, regional Cultural Entropy and the bank access of the individual. Specification 2 adds the main interaction between caste and CE, which is the core of hypothesis *H01*. Specification 3 adds a battery of controls to what is tested in Specification 1. Specification 4 adds the same battery of controls to what is tested in Specification 2. Thus, Specification 4 is my most rigorous test of *H01* in the presence of all available controls.

There is negative and significant relationship between the cultural factors and the bank access in all the model specifications. Caste 1, that is Scheduled Tribe (ST) caste have lower access to bank account as compared to the Scheduled caste (SCs) and other castes. (Caste 2 and Caste 3). The wealth of the individual is positively impacting bank access, as higher wealth will lead to more bank access, in turn more financial inclusion. This result is clearly expected and indicated that the model performs adequately. Number of children in the family, age of the head of the households and the location are positively and significantly influencing bank access. This might be due to the correlation between wealth and decision to have a family and what size this family can afford to be in terms of number of kids.

The main test of *H01* is contained in the statistical significance of the interaction terms in Specification 2 & 3 in Table 2. Statistically speaking, both CE, caste and their interaction are statistically significant with and without additional controls. This means that the statistical significance of the interaction is indeed very powerful. *H01* cannot be rejected based on my data and estimation strategy.

The economic meaning of these results is that regional cultural context has a comprehensive characteristic as a milieu that moderates the level to which the inherited social norms affect individuals. This is a core claim and expectation of the CBD market for values (see Tubadji 2025a). It seems that this expected relationship is present across the regions of India. There is a comprehensive local characteristic of the cultural milieu that moderates the inherited social norms and the discrimination that different cultural clubs will experience locally.

6 Conclusion

This study aims to employ the CBD paradigm in order to explain the regional inequality in access to bank services among the regions of India. We know that there is caste discrimination, but castes are present everywhere across the territory of the country and some localities experience worse inclusion than others. This effect is not entirely compositional – in terms of having segregation in castes per regions. Hence, it is very relevant to hypothesize that the different cultural milieux in the regions of India contribute to different degrees of caste discrimination which results in different average levels of exclusion from the banking services.

The study uses a big dataset, with over 600,000 observations regarding the banking services and augments this data with high quality survey data about local cultural participation across the regions of India. The latter allows for implementing the CBD over-identification of culture and measure most precisely the aggregate cultural milieu in every region.

The results clearly demonstrate that individual level caste discrimination is moderated by the overall regional cultural entropy of the region. This is a highly encouraging insight because it opens an avenue for intervention in the unwanted inequality in financial integration. By rebalancing the cultural participation of the lagging behind regions, the necessary cultural milieu can be ensured that will decrease the regional cultural bias and allow these regions to catch up in financial integration of all their social groups.

The strong statistical significance of the interaction between caste and regional Cultural Entropy suggest that the above policy advice is likely to be reliable. Further CBD-employing research, confirming the role of aesthetic education for moderating local inequality in discrimination and financial integration is therefore highly relevant and needed.

References

Aggarwal S, Jain T, Sharma A (2022) Financial literacy and digital financial behaviour in India: Evidence from household surveys. *Journal of Financial Economic Policy* 14: 402–420. CrossRef

Banaji S (2018) *Young citizens and new media: Learning and democratic engagement.* Routledge, London

Banerjee B, Knight J (1985) Caste discrimination in the Indian urban labour market. *Journal of Development Economics* 17: 277–307. CrossRef

Basu A (2017) *Culture, social norms, and development in India.* Routledge, London

Bourdieu P (1986) The forms of capital. In: Richardson J (ed), *Handbook of theory and research for the sociology of education.* Greenwood Press, Westport, CT, 241–258

Burgess R, Pande R (2005) Do rural banks matter? Evidence from the Indian social banking experiment. *American economic review* 95: 780–795. CrossRef

Chattopadhyay S, Roy S (2021) Aadhaar-enabled welfare delivery and financial inclusion in India. *Economic & Political Weekly* 56: 45–53

Collins D, Morduch J, Rutherford S, Ruthven O (2009) *Portfolios of the poor: How the world's poor live on $2 a day.* Princeton University Press, Princeton, NJ. CrossRef

Data For India (2023) Access to banking. https://www.dataforindia.com/access-to-banking/

Demirgüç-Kunt A, Klapper L (2013) Measuring financial inclusion: Explaining variation in use of financial services across and within countries. *Brookings papers on economic activity* 2013: 279–340. CrossRef

Deshpande S (2011) *The grammar of caste: Economic discrimination and social mobility in contemporary India.* Oxford University Press, Oxford, UK. CrossRef

Deshpande S, Yadav Y (2006) Redesigning affirmative action: Castes and benefits in higher education. *Economic and Political Weekly* 41: 2419–2424

DRG – Development Research Group, Finance and Private Sector Development Unit (2025) *The Global Findex Database 2025: Connectivity and Financial Inclusion in the Digital Economy.* World Bank, Development Data Group. CrossRef

Dumont L (1970) Religion, politics, and society in the individualistic universe. *Proceedings of the Royal Anthropological Institute of Great Britain and Ireland*: 31–41. CrossRef

Fuller C (1996) *Caste today.* Oxford University Press, Oxford, UK

Ghosh S, Vinod D (2017) What constrains financial inclusion for women? Evidence from Indian micro data. *World development* 92: 60–81. CrossRef

Govindapuram S, Bhupatiraju S, Sirohi R (2023) Determinants of women's financial inclusion: Evidence from India. *Annals of Public and Cooperative Economics* 94: 131–158. CrossRef

Harper M (2002) Promotion of self help groups under the SHG bank linkage programme in India. In: *Seminar on SHG-bank Linkage programme at New Delhi.* 25–6

Harvey D (2001) *Spaces of capital: Towards a critical geography.* Routledge, London. CrossRef

IMF – International Monetary Fund (2021) India's approach to open banking: Some implications for financial inclusion. IMF Working Paper No. 21/52. International Monetary Fund, Washington, DC. CrossRef

Jaffrelot C (2003) *India's silent revolution: The rise of lower castes in North India*. Columbia University Press, New York, NY

Jangam C (2021) Understanding caste violence in contemporary India. *Economic and Political Weekly* 56: 45–52

Jayachandran S (2021) Social norms as a barrier to women's employment in developing countries. *IMF Economic Review* 69: 576–595. CrossRef

Jodhka S (2015) *Caste in contemporary India*. Routledge India, New Delhi. CrossRef

Kapur A (2020) *Violence, law, and justice: The state and caste atrocities in India*. Oxford University Press, Oxford, UK

Kochhar R (2020) Business correspondents and the last-mile challenge in India's financial inclusion architecture. *Reserve Bank of India Occasional Papers* 41: 1–25

Kumar S, Venkatachalam R (2019) Caste and credit: A woeful tale? *The Journal of Development Studies* 55: 1816–1833. CrossRef

Madheswaran S, Attewell P (2007) Caste discrimination in the Indian urban labour market: Evidence from the National Sample Survey. *Economic and Political Weekly* 42: 4146–4153

Mehta P (2021) Aadhaar and UPI integration for secure digital payments in India. *Journal of Financial Technology Policy* 8: 45 57

Nambissan G (2010) Exclusion and discrimination in schools: Experiences of Dalit children. *Indian Institute of Dalit Studies Working Paper Series* 1: 1–34

North D (1990) *Institutions, institutional change, and economic performance*. Cambridge University Press, Cambridge, UK. CrossRef

NPCI – National Payments Corporation of India (2022) Annual report 2022. National Payments Corporation of India

Pai S (2002) *Dalit assertion and local democracy*. Oxford University Press, Oxford, UK

Paik S (2022) *The subalterns and technology: Dalit and Adivasi encounters with the digital*. Oxford University Press, Oxford, UK

Putnam R (1993) *Making democracy work: Civic traditions in modern Italy*. Princeton University Press, Princeton, NJ

Rawal V (2008) Ownership holdings of land in rural India: Putting the record straight. *Economic and Political Weekly* 43: 43–47

RBI – Reserve Bank of India (2020) National strategy for financial inclusion 2019-2024. https://www.rbi.org.in/commonman/upload/english/content/pdfs/english_16042021.pdf

Roy S, Singh A (2025) Digital financial inclusion among self help group women in india: A study on adoption, ability and attitude. In: Malik FA, Mahajan S, Yadav DK, Lone NA, Amin S (eds), *Financial Resilience and Environmental Sustainability: Global South Perspectives*. Springer Nature Singapore, Singapore, 167–188. CrossRef

Sarma M, Pais J (2011) Financial inclusion and development. *Journal of international development* 23: 613–628. CrossRef

Sen A (1999) *Development as freedom*. Oxford University Press. CrossRef

Shah A (2022) Regulatory reforms and the evolution of India's digital financial ecosystem. *Reserve Bank of India Occasional Papers* 43: 1–28

Shah G, Mander H, Thorat S, Deshpande S, Baviskar A (2006) *Untouchability in rural India*. SAGE Publications, Thousand Oaks, CA

Shannon C (1948) A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423. CrossRef

Sidhwani P (2015) Spatial inequality and residential segregation in India. *Economic and Political Weekly* 50: 55–63

Sukumar P (2013) Caste discrimination and exclusion in Indian higher education: Dalit experiences. *International Journal of Humanities and Social Science Invention* 2: 23–29

Thorat S (2010) Caste, social exclusion, and poverty linkage: Concept, measurement, and empirical evidence. *Indian Journal of Human Development* 4: 1–13

Thorat S, Attewell P (2007) The legacy of social exclusion: A correspondence study of job discrimination in India. *Economic and Political Weekly* 42: 4141–4145

Thorat S, Newman K (2012) *Blocked by caste: Economic discrimination in modern India*. Oxford University Press, Oxford, UK

Throsby D (2001) *Economics and culture*. Cambridge University Press, Cambridge, UK. CrossRef

Throsby D (2010) *The economics of cultural policy*. Cambridge University Press, Cambridge, UK. CrossRef

Tubadji A (2012) Culture-based development: Empirical evidence for Germany. *International Journal of Social Economics* 39: 690–703. CrossRef

Tubadji A (2013) *Culture-based development and regional growth: An analytical framework*. Springer, Berlin

Tubadji A (2015) *Culture and regional economic performance: Theory and evidence*. Routledge, London

Tubadji A (2021) Ceteris paribus and fixed effects in regional and cultural economics. In: Suzuki S, Patuelli R (eds), *A Broad View of Regional Science: Essays in Honor of Peter Nijkamp*. Springer Singapore, Singapore, 175–198. CrossRef

Tubadji A (2025a) Cultural entropy, innovation, and growth. *Politics & Policy* 53: 70050. CrossRef

Tubadji A (2025b) Culture based development (CBD) in a nutshell. In: Tubadji A (ed), *Culture Based Development*. Edward Elgar Publishing, Cheltenham, UK, 226–233. CrossRef

Tubadji A, Jain Y, Asproudis E (2024) Culture based development and pro-circular-economy behaviour in India. *Scienze Regionali* 24: 203–236

Tubadji A, Nijkamp P (2015) Cultural gravity effects among migrants: A comparative analysis of the EU15. *Economic Geography* 91: 343–380. CrossRef

UNESCO (2013) *Creative economy report 2013: Widening local development pathways*. United Nations Development Programme (UNDP), New York, NY

Vithayathil T, Singh D (2012) Spaces of discrimination: Caste, geography, and the role of the state in India. *Environment and Planning A* 44: 2130–2147. CrossRef

World Bank (2020) *The Global Findex Database 2020: Financial inclusion, digital payments, and resilience in the age of COVID-19*. World Bank, Washington, DC

A Appendix: Obtaining the CBD Cultural Entropy Measure

The 18 variables used in this study are drawn from the NSSO - Survey on Literacy and Culture (2014) and capture diverse forms of cultural participation and inherited cultural norms. These include membership in clubs and cultural groups, engagement with reading books, viewing television programs, radio, cinema, and video shows, as well as attendance at dramas, dance performances, musical programs, and religious programs. Participation in sports and games, magic shows, puppet shows, museum or exhibition visits, educational and recreational activities, and religiously motivated travel are also included. Collectively, these variables provide a comprehensive measure of individual and household involvement in cultural activities, serving as indicators to construct factor variables that reflect the local cultural milieu or cultural capital.

A principal component factor analysis is implemented with these 18 variables. This identifies a CH and LC factor variable as elements of the local cultural capital. The rotated loadings of the variables in factors are presented in Table A.1.

Figure A.1 shows the Scree plot of eigenvalues after factor loading, which means a decline in eigenvalues after the fourth factor, indicating that four factors are sufficient to represent the underlying data structure and variation.

Table A.1 shows a significant Bartlett's test and a KMO value of 0.642 suggest adequate intercorrelation among variables and support the use of factor analysis.

Table A.2 clearly indicate that six factors have eigenvalues greater than one and together explain approximately 45.2% of the total variance in the data.

Table A.3 shows the contribution of each variable to the extracted latent factors, highlighting the underlying dimensions captured by the factor analysis.

Table A.1: Bartlett's Test of Sphericity and Kaiser–Meyer–Olkin (KMO) Measure of Sampling Adequacy

Bartlett test of sphericity H0: variables are not intercorrelated	Kaiser-Meyer-Olkin Measure of Sampling Adequacy
p-value = 0.000 Chi-square = 5.78e+05 Degrees of freedom = 153	KMO = 0.642

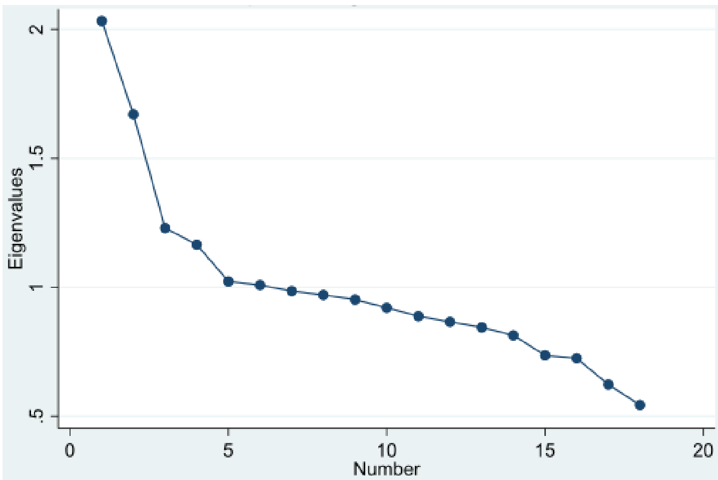


Figure A.1: Scree plot of eigenvalues after factor

Table A.2: Eigenvalues and Variance Explained by Retained Factors

Factor	Variance	Difference	Proportion	Cumulative
LC_Performance	1.89637	0.45321	0.1054	0.1054
LC_Telecast	1.44317	0.09933	0.0802	0.1855
SC	1.34384	0.09556	0.0747	0.2602
CH_relig	1.24828	0.10098	0.0693	0.3295
LC_shows	1.1473	0.09608	0.0637	0.3933
Sports	1.05122	.	0.0584	0.4517

Notes: LR test: independent vs. saturated: $\text{chi}^2(153) = 5.8\text{e}{+}05$ Prob>chi2 = 0.0000

Table A.3: Factor Loadings and Uniqueness of Variables from Factor Analysis

Variable	LC_Performance	LC_Telecast	SC	CH_relig	LC_shows	Sports	Uniqueness
Club	0.2327	0.0971	0.205	0.6209	0.0218	-0.1772	0.4771
Cultural Group	0.2505	-0.0006	0.1878	0.6071	-0.107	-0.1292	0.5052
Reading Groups	0.2214	0.6818	-0.0926	-0.0452	0.2207	0.0749	0.4212
Radio/Tape recorder	0.1798	0.6008	-0.4467	0.0768	-0.0448	0.1124	0.3866
TV Programmes	0.2096	0.5265	0.4972	-0.2549	0.0694	-0.0338	0.3608
Cinema	0.1953	0.3266	-0.4014	0.176	-0.3318	0.2019	0.5123
Video Show	0.2394	0.2874	0.1932	-0.124	0.5085	0.0779	0.5427
Drama	0.4196	-0.1139	-0.0325	0.1152	-0.0705	0.0429	0.7899
Dance	0.5445	-0.1991	-0.1016	-0.1365	-0.0024	-0.0696	0.6301
Recreation/Musical	0.2953	-0.0735	0.0485	0.1363	0.2024	0.2729	0.7711
Religious Programmes	-0.0391	-0.3786	0.0143	0.197	0.5258	0.2761	0.4634
Sports & Games	0.1174	0.13	0.1846	0.1724	0.1844	-0.3116	0.7744
Circus/Magic show	0.5618	-0.2217	-0.115	-0.2662	0.0338	-0.0628	0.5461
Puppet/ other show	0.6341	-0.2377	-0.1188	-0.1726	0.0033	-0.1128	0.4847
Museum/Exhibition	0.5653	-0.1777	-0.0346	-0.0556	-0.0554	-0.0308	0.6405
Educational	0.2253	-0.0315	0.0474	0.1712	-0.1291	0.1287	0.8834
Recreational	0.1623	0.0769	0.6047	-0.1915	-0.4411	0.0435	0.3689
Religious	0.0781	-0.0789	0.2331	0.0577	-0.1226	0.7769	0.3114

Source: Authors collation based on the NSSO data
Note: Using 18 cultural variables from the NSSO (2014), I conducted a factor analysis and extracted six underlying factors. Each factor was labelled based on its constituent variables, as indicated by the factor loadings. These factor scores were then aggregated at the regional level to generate a cultural profile for each region. The resulting regional summaries provide a quantitative description of the local cultural context in which individuals observed in the NFHS dataset are embedded.

B How is the Cultural Entropy calculated?

Cultural Entropy is calculated according to Tubadji (2025a). To quantify the local cultural environment, this study adopts the Culture-Based Development (CBD) framework, which conceptualizes culture as a complex entity composed of two core components: Living Culture (LC) and Cultural Heritage (CH). CBD argues that any local cultural system can be understood through the balance between these two elements, where LC captures every day, contemporary cultural practices, and CH reflects historically rooted or traditional cultural expressions.

To operationalize this balance, CBD defines cultural entropy (CE) as a measure of the variability between LC and CH within the cultural capital of a locality. This measure is derived by adapting the Shannon Entropy formula (Shannon 1948):

$$CE = H(p_i) = -\sum_i p_i \log(p_i) \tag{2}$$

where:

i denotes the cultural class, which in this framework comprises two categories: LC and CH.

p_i represents the probability that a cultural activity in a given locality falls into category i.

H denotes the Shannon Entropy, interpreted here as the variability of the cultural capital composition.

A higher CE value indicates a more even balance between LC and CH, suggesting a diverse and dynamic cultural environment. Conversely, a lower CE value reflects the dominance of one cultural component, signalling a more homogeneous cultural landscape. Although Shannon Entropy can theoretically range from 0 to infinity, with two classes it is effectively bounded by the distribution of LC and CH probabilities.

To construct these probabilities empirically, I use 18 cultural participation variables from the NSSO (2014) survey, representing a wide spectrum of cultural activities such as reading, watching cultural programs, attending performances, sports, religious travel, museum visits, and other cultural engagements.

A factor analysis (FA) is conducted on these 18 variables, producing six underlying factor dimensions, each labelled according to its dominant cultural characteristics based on factor loadings. These factor scores are then aggregated at the regional level, generating quantitative summaries that capture regional cultural profiles.

Next, these regional cultural scores are mapped onto the two CBD cultural classes (LC and CH), allowing the computation of probability distributions p_i for each region. These distributions are used to calculate cultural entropy (CE) for each region.

Finally, the regional CE scores are merged with the National Family Health Survey (NFHS) dataset. Each individual in the NFHS is thus situated within a quantified description of their local cultural context. This enables analysis of whether and how variations in local cultural entropy influence household behaviours, social outcomes, or developmental indicators captured in the NFHS.

Table B.1: Summary Statistics for the Main Variables in the Analysis

Variable	Obs	Mean	Std. Dev.	Min	Max
bank_acc	6,34,731	0.956353	0.204308	0	1
CE	6,34,868	0.661338	0.280253	0.000779	1
caste_1	6,03,491	0.202606	0.401942	0	1
caste_2	6,03,491	0.387608	0.487205	0	1
caste_3	6,03,491	0.206013	0.40444	0	1
inter_ce_c~1	6,03,478	0.144648	0.309233	0	1
inter_ce_c~2	6,03,478	0.25228	0.364477	0	1
inter_ce_c~3	6,03,478	0.131579	0.28752	0	1
wealth_index	6,34,881	2.814297	1.39628	1	5
no_child	6,34,881	0.435746	0.757291	0	11
age_hh	6,34,881	49.62948	14.11599	12	98
sex_hh	6,34,881	1.172283	0.377693	1	3
location					
urban	6,34,881	0.251591	0.433927	0	1
rural	6,34,881	0.74841	0.433927	0	1
year_inter~w					
2019	6,34,881	0.476655	0.499455	0	1
2020	6,34,881	0.279197	0.448605	0	1
2021	6,34,881	0.244148	0.429581	0	1

Source: Authors collation based on the NFHS and NSSO data
Notes: Dependent Variable = Having access to Bank account, Independent variables includes: Cuntural Entropy, Caste 1= STs (Schedule Tribes), Caste 2 = SCs (Schedule Castes) and Caste 3 = OBCs (Other Backward Classes) and General cate category is the base category. wealth_index = Wealth Index of the Indivuals (which measures the income across the individuals). No_child = Number of Children in the family. Age_hh = Age of the Household head. Sex_hh = Gender of the Household Head, it may male or female. Location is Urban or Rural. NFHS Survey Interviews – 2019, 2020 and 2021, 2019 is the base category.

© 2026 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Milan, Italy. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Funded by

ersa

WU
WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

FWF