Journal of Geographical Systems
https://doi.org/10.1007/s10109-026-00492-9

ORIGINAL ARTICLE

# "What is a city but [what it offers to] the people?": spatial interaction gravity models on internal migration within Tuscany

**Francesco Viviani**[1,2] · **Leonardo Piccini**[3] · **Chiara Bocci**[2]

Received: 1 October 2024 / Accepted: 17 March 2026


## Abstract

In recent decades, the Gravity model has proved to be an essential tool in understanding and studying migration patterns and trends, becoming a widely used framework for researchers in this area. This paper contributes to the literature on the subject by applying a "Hurdle Negative Binomial Gravity Model" in order to investigate migration flow patterns within Tuscany from 2012 to 2020. In particular - taking into account the flow matrices between all the 273 municipalities of the Tuscan region - the main objectives are to define and investigate the factors that drive people to migrate from one municipality to another and, given the ever increasing importance of technology and digitisation, to focus on the role that the digitisation level of a municipality has in attracting new flows. The results show the deterrent effect of distance, confirming Ravenstein's idea that "migrants move only short distances", and the positive effect of the municipalities' digitisation level, confirming the idea that people prefer to live in more digitised cities.

**Keywords** Spatial interaction data · Gravity model · Hurdle model · Flow data · Internal migration

**JEL classification codes** C21 · R23 · C10

✉ Francesco Viviani
f.viviani5@campus.unimib.it

Leonardo Piccini
leonardo.piccini@irpet.it

Chiara Bocci
chiara.bocci@unifi.it

[1] Dipartimento di Giurisprudenza, University of Milano-Bicocca, Milan, Italy

[2] DISIA - Dipartimento di Statistica, Informatica, Applicazioni “Giuseppe Parenti”, University of Florence, Florence, Italy

[3] I.R.P.E.T - Istituto Regionale Programmazione Economica della Toscana, Florence, Italy

Published online: 02 May 2026

## 1 Introduction

In recent years, the significant increase in the availability of bilateral data on migrations has led to an always increasing interest in the application of the gravity model as an option to investigate the main factors that drive people to migrate, both in relation to international migration (flows between nations) and to internal migration (flows between regions, provinces, municipalities, etc.) (Ramos, 2016).

In this literature, migration from an origin *i* to a destination *j* is typically explained by (i) characteristics of the origin and destination—often described as *push* and *pull* factors—and (ii) bilateral mobility frictions such as distance, travel time, and other barriers to movement (Manzoor et al. 2021; Tamaş 2017; Karemera et al. 2000). At the same time, recent contributions have emphasised that empirical results in gravity settings crucially depend on modelling choices related to zero flows, heteroskedasticity and unobserved heterogeneity across locations (Beine et al. 2016; Martin and Pham 2020).

The gravity model has a long tradition in migration studies. Early qualitative insights already suggested that spatial frictions matter: for instance, Ravenstein's pioneering discussion of migration regularities highlighted the importance of proximity in shaping mobility patterns (1889). Building on this intuition, gravity models have been employed in a variety of contexts to quantify the role of spatial frictions and location-specific drivers. Examples range from educational migration (e.g., student flows) to other forms of mobility where distance and destination attributes jointly affect choices (Alm and Winters 2009; Florax et al. 2004; Tamaş 2017). While these applications differ in setting and outcomes, they consistently illustrate the core gravity intuition: flows depend on the attractiveness of destinations, the characteristics of origins, and the costs of moving between them.

In the context of international migration, a broad literature has investigated standard economic and demographic determinants, such as income differentials, labour market conditions, networks and cultural proximity (Karemera et al. 2000; Beine et al. 2019; Belot and Ederveen 2012). In parallel, an expanding body of work has studied internal migration, where recent studies showed that within-country migration responds to a rich set of drivers, including local labour-market conditions, amenities, infrastructure and accessibility (Bunea 2012; Serlenga and Shin 2021), together with a growing body of evidence of studies showing that internal flows react to shocks and broader structural forces, such as environmental events and climate-related changes (Pirani et al. 2019; Reimann et al. 2023; Wajdi et al. 2017). For Italy, gravity-based studies at aggregated spatial scales (e.g., provinces or regions) have highlighted the role of unemployment, per-capita income and structural characteristics in shaping internal migration (Basile and Causi 1991; Casacchia et al. 2021; Piras 2016).

Despite this extensive literature, two gaps remain particularly relevant for policy and for understanding spatial dynamics. First, evidence at a highly granular spatial scale is still limited, even though many determinants of attractiveness (public services, local governance, accessibility and amenities) vary substantially across municipalities and are directly influenced by local policies (Basile and Causi 1991). Second, the role of *local digitisation* as a potential driver of internal mobility has received comparatively little attention in gravity applications, especially at the municipality level

(Beine et al. 2016). Digitisation may, in fact, matter through several channels: it can reduce information and transaction costs in the interaction with public administrations, improve access to services, and facilitate remote or hybrid work; at the same time, it may capture broader differences in local administrative capacity and service quality (Szedmák et al. 2025).

This paper contributes to filling these gaps by analysing bilateral migration flows among all 273 municipalities in Tuscany over 2012–2020, and by evaluating whether the digitisation level of the destination municipality is associated with the ability to attract new inflows. Tuscany provides a suitable setting because it features substantial heterogeneity across municipalities in terms of size, accessibility, demographic structure, economic conditions and service provision, within a relatively compact geographic area.

## 2 Methodology

### 2.1 The gravity model

As introduced in the previous paragraph, Gravity models are among the best known and most widely used when dealing with bilateral data. Ravenstein (1889), by assessing the idea that "most migrants move only short distances", was the pioneer of the concept of a Gravity model, but it was not until the work of Tinbergen (1962), Nobel Prize winner for Economics in 1966, who, inspired by Newtonian physics, first theorised the concept of a Gravity model for the study of international trade flows. Soon after, in an important research entitled "An Econometric Study of World Trade Flows", Linneman (1966), who was a Tinbergen's student, extended the work of Tinbergen and theorised a more detailed and specific definition of the Gravity model.

The above-mentioned authors took inspiration from Newton and his law of universal gravity, suggesting that patterns of spatial interaction, such as international trade or migratory phenomena, could be predicted and elucidated by making an analogy with the law of gravity formulated by Newton in the late 1600 s (Bailey and Gatrell 1995), which states that the attractive force between two objects *i* and *j* is given by the formula (Wolver et al. 2018):

$$F_{ij} = G \cdot \frac{M_i \cdot M_j}{D_{ij}^2} \tag{1}$$

Where:

- *i* and *j* denote the two interacting bodies.
- $F_{ij}$ is the attractive force between two bodies.
- $M_i$ and $M_j$ are the masses of the two bodies.
- $D_{ij}^2$ is the distance squared between the two masses.

- *G* is the constant, called the gravitational constant, which depends on the units of the two masses.

As can be seen in Eq. (1), the formula theorised by Isaac Newton suggests that the attraction of two bodies, $i$ and $j$, is directly proportional to the heaviness of their masses and inversely proportional to their distances (Newton 1687, in Wolver et al. 2018). In the same way, Tinbergen's concept was to utilize the same formula in order to show how the volume of trade flows between two nations is directly related to their attractiveness and inversely proportional to the distance between the two economic centers. Tinbergen's formula is now known as 'The Traditional Gravity Equation,' and it may be represented as follows (Bailey and Gatrell 1995; Wolver et al. 2018):

$$Y_{ij} = \frac{X_i \cdot X_j}{D_{ij}^{\gamma}} \tag{2}$$

where:

- $Y_{ij}$ is the total volume of interaction between $i$ and $j$, such as for example, the amount of flows between an origin $i$ and a destination $j$.
- $X_i$ are the origin-specific variables.
- $X_j$ are the destination-specific variables.
- $D_{ij}$ is the distance between $i$ and $j$.

The formula shown in Eq. (2) has provided a functional framework that can be adapted to study flow data, such as migration patterns, where the interaction between an origin and a destination depends not only on the characteristics of both locations but also on the spatial distance between them.

### 2.2 Modelling flows with excess of zeros: the gravity hurdle model

Flow data refers to the measurement of movement or transfer of entities such as people, goods or services between different spatial locations. In many cases, it is possible to treat flow data as count data because of the fact that they represent, for example, the amount of people migrating from one city or country to another, the amount of goods traded from a country to another and so on (O'Kelly 2009). This means having to deal with a discrete and non-negative response variable, which are the two major characteristics of count data models. The most utilised models in literature that are able to deal with that kind of data are the Poisson and the Negative Binomial regression models (Cameron and Trivedi 2013), where the choice between the two depends on the distribution of the data. In particular, when over-dispersion occurs, i.e the variance exceeds the mean, the Negative Binomial model is preferred over the Poisson model.

Moreover, count data can be sometimes affected by what is called "excess of zeros", namely that the response variable shows an higher frequency of zeros than the frequency predicted by standard distributions.

In order to model count data that show a high frequency of zeros, Zero-Inflated and Hurdle Models are both valid options (Cameron and Trivedi 2013). The main differences within the two approaches lie in the conceptualization of zeros as well as the interpretation of the parameters of the model. A Zero-Inflated model assumes that zeros arise from a combination of two distributions: one where certain subjects always produce zeros ("Structural zeros") and one where subjects are exposed to the outcome but did not report the experience and generate zeros ("Sampling zeros") (Feng 2021). A Hurdle model, on the other hand, considers all zero data as "structural zeros" and the positive values arise from a truncated positive distribution. Therefore this model consists of two parts: a model for the binary outcome "zero" or "not zero" values, and a model for the positive counts. The choice between the Zero-Inflated model and the Hurdle model depends on the nature of the data and on the process leading to the excess of zeros.

This study focuses on Hurdle models, which can be viewed as two-part mixture models, where the first part is a binary model, for example a Logistic or a Probit model, which determines whether the response outcome is zero or positive, while the second part is instead related only on positive outcomes and employs a truncated model that adjusts a standard distribution by considering only non-negative outcomes (Agresti 2015). Given $Y$ a discrete non negative response variable, $\boldsymbol{X}$ and $\boldsymbol{Z}$ two covariates matrices and calling the probability of observing a count of zero for $Y$ as $P(Y=0|\boldsymbol{x}_i)=\pi_i$, a generic Hurdle model for each individual $i$ can be defined as:

$$P(Y=y_i|\boldsymbol{x}_i,\boldsymbol{z}_i)=\begin{cases}\pi_i & \text{if} y_i=0\\ (1-\pi_i)P(Y=y_i|Y>0,\boldsymbol{z_i}) & \text{if} y_i>0\end{cases} \tag{3}$$

Let $\mu_i=E[Y_i|\boldsymbol{z}_i]$ the model can be then specified according to the distribution of Y. For example, if the data follow a Poisson distribution, the model becomes a Hurdle Poisson Model that can be specified as:

$$P(Y=y_i|\boldsymbol{x_i},\boldsymbol{z_i})=\begin{cases}\pi_i & \text{if } y_i=0\\ (1-\pi_i)\exp(-\mu_i)\frac{\mu_i^{y_i}}{(1-\exp(-\mu_i))y_i!} & \text{if } y_i>0\end{cases} \tag{4}$$

On the other hand, a Negative Binomial distribution is preferred when dealing with situations where the Poisson's assumption of equality between mean and variance is not plausible. In this case a Hurdle Negative Binomial Model is usually the most used solution. Starting from Eq. (3), a specification is given as:

$$P(Y=y_i|\boldsymbol{x_i},\boldsymbol{z_i})=\begin{cases}\pi_i & \text{if } y_i=0\\ \frac{(1-\pi_i)}{1-(\frac{\theta}{\theta+\mu_i})^\theta}\frac{\Gamma(y_i+\theta)}{\Gamma(\theta)\Gamma(y_i+1)}\left(\frac{\theta}{\theta+\mu_i}\right)^\theta\left(\frac{\mu_i}{\mu_i+\theta}\right)^{y_i} & \text{if } y_i>0\end{cases} \tag{5}$$

When $Y=\{y_{ij}\}$ represents flows from origin $i$ to the destination $j$, it is possible to model the two components of the Hurdle model with gravity models, obtaining a gravity Hurdle model, with:

1. A gravity regression model for $\pi_{ij}$, the probability of observing a zero flow between area $i$ and area $j$.
2. A gravity regression model for $\mu_{ij}$, the conditional mean of the positive flow component.

While the hurdle negative binomial gravity model provides a convenient way to decompose the extensive margin (zero vs. positive flows) and the intensive margin (size of positive flows), recent contributions in the gravity literature stress that flow equations should also be estimated in their multiplicative form in order to avoid biases induced by log-linearization under heteroskedasticity and to naturally accommodate the large mass of zero observations that typically characterizes bilateral matrices (Silva et al., 2004). In particular, Silva et al. (2004) show that, when the variance of the error term depends on the covariates, the conventional log-linear OLS specification may yield inconsistent estimates of the elasticities, and they propose a Poisson pseudo-maximum-likelihood (PPML) estimator as an alternative that can be applied even in the presence of zeros. Moreover, practitioners' guides to gravity models of migration emphasize that the dyadic nature of bilateral data calls for econometric approaches that exploit the panel dimension and allow for high-dimensional fixed effects, while explicitly addressing the prevalence of zero flows through appropriate estimators such as PPML (Beine et al., 2016).

Formally, letting $Y$ denote the number of migrants moving from municipality $i$ (origin) to municipality $j$ (destination) in year $t$, we specify the conditional mean as:

$$\mathbb{E}[Y_{ijt} \mid \boldsymbol{z}_{ijt}] = \mu_{ijt} = \exp\left(\beta' \boldsymbol{z}_{ijt} + \alpha_i + \tau_t\right), \tag{6}$$

where $\boldsymbol{z}_{ijt}$ includes bilateral mobility frictions (travel time) and a set of origin/destination covariates, such as the destination digitisation index and standard pull controls such as destination population and income per capita, as well as additional controls adopted in the baseline specification. The fixed effects $\alpha_i$ and $\tau_t$ capture, respectively, time-invariant origin-specific factors and common year shocks. Estimation is performed by PPML and standard errors are clustered at the municipality-pair (dyad) level to account for within-pair dependence over time.

## 3 Case study: the effect of digitisation on italian internal migration flows

### 3.1 Data

The response variable is constructed from the survey "Registrations and Cancellations in the Population Register Due to Change of Residence", which is conducted annually by the Italian National Institute of Statistics (ISTAT) (Table 1).

Migration data are collected on the day that the request for registration in the new municipality of residence begins, but are recorded when the migration process has ended. Furthermore, ISTAT provides not only information from the population register, but also an Origin–Destination Matrix at the municipality level. The matrix includes the exact amount of individuals who have relocated from their original municipality to their destination during the year. Each Origin–Destination Matrix comprises a total of 74,529 observations ($273 * 273$) and has led to the creation of the response variable denominated "Flows", which is defined as $Y_{ijt}$ = "Amount of

**Table 1** Description of variables used in the analysis

| Variable | Description | Source |
|---|---|---|
| Flows | Number of migrants from origin $i$ to destination $j$ | ISTAT |
| Driving_Time | Driving time from the origin $i$ to the destination $j$ | ISTAT |
| Digit_Dest | Digitisation index of the destination $j$: computed as the ratio between households with access to > 100 Mbps and total number of households | AGCOM |
| Pop_Ori | Population in the origin $i$ | ISTAT |
| HouseRatio | Ratio between the average house price in the origin $i$ and the average house price in the destination $j$ | OMI |
| Class_Area_Dest | Area classification of the destination $j$ (1: city, 2: metropolitan area, 3: periphery) | ISTAT |
| Prov_Cap_Dest | Dummy variable of the destination $j$ (1: provincial capital; 0: otherwise) | ISTAT |
| Prov_Cap_Ori | Dummy variable of the origin $i$ (1: provincial capital; 0: otherwise) | ISTAT |
| Income_pc_Dest | Income per capita of the destination $j$ | ISTAT |
| Old_Ori | Old Age Index of the origin $i$ | ARS |

*Notes:* ISTAT = Italian National Institute of Statistics; AGCOM = Italian Communications Regulatory Authority; OMI = Real Estate Market Observatory; ARS = Regional Health Agency

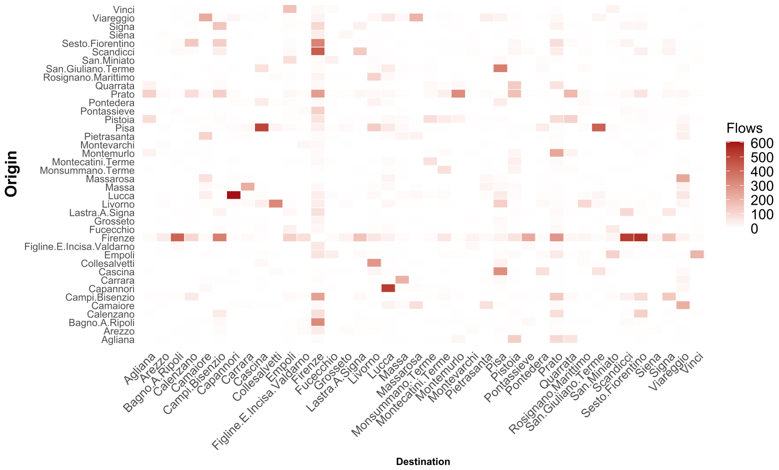


**Fig. 1** Origin/Destination Matrix for the first 40 municipalities (2020)

people who have moved from the municipality $i$ to the municipality $j$", where $t$ = "2012; 2013;...;2020".

The heatmap in Fig. 1 gives an example of the Origin/Destination matrix for the year 2020, showing the changes in municipalities of people in Tuscany. The heatmap gives the reader an easier and better visualisation of the matrix: the darker the colour, the higher the number of people who decided to move from the origin $i$ to the destination $j$ and vice-versa.

The analysis has been conducted using a dataset made of 670,761 observations between municipalities. The response variable is the variable “Flows”, while Table2 provides a general outlook of the dataset and of the explanatory variables with a short description and their relative sources Table 1.

### 3.2 The model specification

Given the spatial interaction data approach and the characteristics of the response variable - which is a discrete, non-negative variable with an excess of zeros problem - the choice of the model was that of a ’Hurdle Negative Binomial Gravity Model’. A key motivation for using a Negative Binomial specification in the count component is the presence of overdispersion in migration flows. Under the Poisson model, the conditional variance equals the conditional mean, an assumption that is strongly violated in our data. Consistently, the Negative Binomial model allows the conditional variance to increase more than proportionally with the conditional mean,

$\mathrm{Var}(Y \mid z_{ij}) = \mu_{ij} + \alpha\mu_{ij}^2$, and the estimated dispersion parameter is statistically significant in the hurdle Negative Binomial models.

The above-mentioned model allows to address two main challenges simultaneously. On one hand, that of modelling spatial interaction data using the Gravity Model, that takes into account distance and the specific covariates related to the ori-

**Table 2** Descriptive Statistics and Multicollinearity Diagnostics

| Variable | Min | Mean | Median | Max | VIF |
|---|---|---|---|---|---|
| Flows (number of migrants) | 0.00 | 0.93 | 0.00 | 756.00 | $< 2.5$ |
| Driving_Time (minutes) | 2.63 | 104.01 | 94.40 | 475.74 | $< 2.5$ |
| House-Ratio (origin/destination) | 0.07 | 1.18 | 1 | 14.8 | $< 2.5$ |
| Digit_Dest (index 0–1) | 0.00 | 0.42 | 0.45 | 0.95 | $< 2.5$ |
| Pop_Ori (number of inhabitants) | 391.00 | 13716.10 | 6345.00 | 377395.00 | $< 2.5$ |
| Income_pc_Dest (euros) | 5729.69 | 13338.12 | 13112.51 | 36475.40 | $< 2.5$ |
| Old_Ori (old-age index) | 106.70 | 230.19 | 212.45 | 760.29 | $< 2.5$ |
| Prov_Cap_Ori (dummy) | 0.00 | 0.0366 | 0.00 | 1.00 | $< 2.5$ |
| Prov_Cap_Dest (dummy) | 0.00 | 0.0366 | 0.00 | 1.00 | $< 2.5$ |
| Class_Area_1 (dummy) | 0.00 | 0.0623 | 0.00 | 1.00 | $< 2.5$ |
| Class_Area_2 (dummy) | 0.00 | 0.5790 | 1.00 | 1.00 | $< 2.5$ |
| Class_Area_3 (dummy) | 0.00 | 0.3590 | 0.00 | 1.00 | $< 2.5$ |

gin and destination; on the other hand, it also allows to model the response variable 'Flows', which is characterised by an excess of zeros problem, using the hurdle part. Thanks to this approach, it is in fact possible to divide the dataset into two distinct parts: one part that models the zeros using a binomial regression, and another that models the count part employing a negative binomial regression. The model is specified according to the probability distribution already showed in eq. (5), where the response variable $y_{ij}$ is:

$$y_{ij} \sim \begin{cases} 0 & \text{with probability } \pi_{ij} \\ \text{TNB}(\mu_{ij}, \theta) & \text{with probability } 1 - \pi_{ij} \end{cases} \tag{7}$$

and TNB stands for a "Truncated Negative Binomial" distribution.

The model for the first part (investigating the probability of being zero) is a binary model and is specified as:

$$\text{Zero Part Model} \tag{8}$$

$$\begin{aligned} \text{logit}(\pi_{ij}) = \beta_0 &+ \beta_1 \cdot log(Driving\_Time_{ij}) + \\ &+ \beta_2 \cdot Prov\_Cap\_Ori_i + \beta_3 \cdot Old\_Ori_i \\ &+ \beta_4 \cdot (Class\_Area\_Dest_j = 2) + \beta_5 \cdot (Class\_Area\_Dest_j = 3) \end{aligned}$$

while the second part, specified a negative binomial regression model with a log-link function, is given as:

$$\text{Count Part Model} \tag{9}$$

$$\begin{aligned} \log(\mu_{ij}) = \beta_0 &+ \beta_1 \cdot log(Driving\_Time_{ij}) + \beta_2 \cdot Digit\_Dest_j + \\ &+ \beta_3 \cdot log(Pop\_Ori_i) + \beta_4 \cdot HouseRatio_{ij} \\ &+ \beta_5 \cdot (Class\_Area\_Dest_j = 2) + \beta_6 \cdot (Class\_Area\_Dest_j = 3) \\ &+ \beta_7 \cdot Prov\_Cap\_Dest_j + \beta_8 \cdot log(Income\_pc\_Dest_j) \end{aligned}$$

In addition to the hurdle negative binomial specification, and in line with recent contributions in the gravity literature, we also estimate a Poisson Pseudo-Maximum Likelihood (PPML) gravity model on the full municipality-pair panel (2012–2020) (Silva et al., 2004). The PPML approach is particularly suitable for spatial interaction matrices characterised by a large share of zero flows and by heteroskedasticity, as it allows the model to be estimated in its multiplicative form without log-transforming the dependent variable (Silva et al., 2004). Moreover, recent methodological discussions of migration gravity models emphasise the importance of exploiting the longitudinal structure of bilateral data through fixed effects and of adopting estimators that appropriately handle zeros (Beine et al., 2016).

Let $Y_{ijt}$ denote the number of migrants moving from origin municipality *i* to destination municipality *j* in year *t*. The PPML gravity model specifies the conditional mean of $Y_{ijt}$ as:

$$\mathbb{E}[Y \mid \boldsymbol{z}_{ijt}] = \mu_{ijt} = \exp\left(\eta_{ijt}\right), \tag{10}$$

with linear predictor:

$$\begin{aligned}\eta_{ijt} = \beta_0 &+ \beta_1 \cdot log(Driving\_Time_{ij}) + \beta_2 \cdot Digit\_Dest_{jt} \\ &+ \beta_3 \cdot log(Pop\_Ori_{it}) + \beta_4 \cdot log(Income\_pc\_Dest_{jt}) \\ &+ \beta_5 \cdot HouseRatio_{ijt} + \beta_6 \cdot Old\_Ori_{it} \\ &+ \beta_7 \cdot Prov\_Cap\_Dest_{jt} + \beta_8 \cdot (Class\_Area\_Dest_{jt} = 2) \\ &+ \beta_9 \cdot (Class\_Area\_Dest_{jt} = 3) + \alpha_i + \tau_t,\end{aligned}$$

where $\alpha_i$ denotes origin fixed effects and $\tau_t$ year fixed effects.

## 4 Results

### 4.1 Model results

The statistical analysis has been conducted with the two main objectives of implementing a Gravity Model capable of explaining the phenomenon of internal migration in Tuscany in the period spanning from 2012 to 2020 and identifying what in literature are called the "push and pull factors", i.e. the factors that make a destination more attractive and the factors that instead lead people to move away from their municipality of residence. A particular emphasis is given to the impact that the digitisation level of a municipality had in attracting new flows in Tuscany throughout all the reference period.

Consequently, nine different models, one for each year of reference, have been formulated, using always the same set of covariates already shown in Eq. (8) and Eq. (9). This approach allowed to investigate the effect of explanatory variables' coefficients across the reference period, studying at the same time, both the evolution of the impact of explanatory variables' on the response variable, and the general effect of covariates in every year of reference. In particular the attention was on the explanatory variable denominated "Digit_Dest", with the aim of investigating the impact that the digitisation level of a municipality had in attracting new residents.

Table 3 presents descriptive statistics, while Table 3 presents the results of the models, one for each year of reference, providing estimations, standard errors, and significance levels of the analysis.

Examining the count component, which captures conditional associations between covariates and migration flows, provides useful evidence on how flows co-vary with the included explanatory variables among Tuscan municipalities from 2012 to 2020. As expected, traditional variables such as the population and the area classification of the destination play a significant role. Variables such as the driving time and distance between municipalities are negative and statistically significant across all models. In line with theoretically expectations of both the Gravity model and migration theory, an increase in travel time between *i* and *j* reduces the quantity of flows between locations, as longer times of travel usually has a deterrent effect in migrations and discourage movements. Moreover, the estimated coefficient on the destination digitisation index (Digi_Dest) is consistently positive and statistically significant across specifications. This suggests that, conditional on the included controls, higher levels of digitisation in the destination municipality are associated with larger bilateral migration flows from municipality i to municipality j within Tuscany. Given the central role of this variable in the study, the next paragraph provides a more detailed discussion of how this association evolves over time.

In addition, following the recent gravity literature and in order to exploit the longitudinal (panel) structure of the municipality-pair dataset, we estimate a Poisson Pseudo-Maximum Likelihood (PPML) gravity model on the full 2012–2020 panel, including origin fixed effects and year fixed effects, with standard errors clustered at the municipality-pair (dyad) level (Silva et al., 2004; Beine et al., 2016). Table 4 reports the PPML estimates.

The PPML estimates confirm the expected negative role of mobility frictions and the positive association of destination characteristics with inflows. Importantly, the coefficient on $Digit\_Dest$ is positive and statistically significant, supporting the hypothesis that higher levels of municipal digitisation are associated with larger expected bilateral migration flows, conditional on the included controls and fixed effects.

### 4.2 The effect of digitisation

Within all the several variables that the study takes into account, the digitisation index is the one of greatest interest. In fact, as already explained in the previous paragraph, one of the principal objective of this study is to assess the impact that digitisation had on attracting new flows. By running nine different Hurdle Negative Binomial Gravity models, each one corresponding to a precise year, and using always the same set of covariates, Fig. 2 illustrates the evolution of the effect that the digitisation index had on internal migration flows within Tuscan municipalities in the period 2012-2020.

As can be seen from the figure, the coefficients are always positive and statistically significant, confirming the idea that, on average, municipalities with an higher level of digitisation tend to attract more flows compared to municipalities with a lower level. On the other side, the main trend of the coefficients' evolution did not show an increase in the period 2012-2020, suggesting that there has been no significant improvement in the effect of digitisation in recent years. In fact, the effect remains steady over time, consistently oscillating around similar values.

Furthermore, given the relatively short time period of the study, it is plausible that the period may not be long enough to detect an improvement in the effect.

### 4.3 Digitisation as a way to repopulate small villages

After analysing the association between digitisation and migration flows across all municipalities in Tuscany, we then focus on a specific subset of movements: flows from municipal hubs to peripheral municipalities. The aim of this section is to assess whether, conditional on the included controls, migrants relocating from hubs tend to choose peripheral destinations with higher levels of digitisation. This exercise is

**Table 3** Hurdle negative binomial gravity estimates (2012–2020)

| | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|
| Count Equation | | | | | | | | | |
| Intercept | −13.267*** | −11.536*** | −13.433*** | −10.234*** | −10.284*** | −8.748*** | −10.603*** | −0.235 | −10.518*** |
| | (1.462) | (1.459) | (1.535) | (1.404) | (1.504) | (1.535) | (1.543) | (1.181) | (1.613) |
| Driving_Time | −1.978*** | −1.958*** | −1.973*** | −1.927*** | −1.926*** | −1.965*** | −2.006*** | −1.918*** | −1.958*** |
| | (0.024) | (0.025) | (0.025) | (0.025) | (0.024) | (0.025) | (0.025) | (0.024) | (0.024) |
| Digit_Dest | 0.250*** | 0.427*** | 0.371*** | 0.419*** | 0.450*** | 0.410*** | 0.314*** | 0.492*** | 0.418*** |
| | (0.062) | (0.063) | (0.063) | (0.064) | (0.061) | (0.063) | (0.063) | (0.061) | (0.062) |
| Income_pc_Dest | 1.702*** | 1.541*** | 1.714*** | 1.387*** | 1.388*** | 1.222*** | 1.435*** | 0.333** | 1.388*** |
| | (0.153) | (0.152) | (0.160) | (0.146) | (0.157) | (0.160) | (0.160) | (0.122) | (0.168) |
| Pop_Ori | 0.536*** | 0.493*** | 0.531*** | 0.498*** | 0.511*** | 0.523*** | 0.519*** | 0.507*** | 0.531*** |
| | (0.013) | (0.013) | (0.013) | (0.014) | (0.013) | (0.013) | (0.013) | (0.013) | (0.014) |
| Prov_Cap_Dest | 1.135*** | 0.906*** | 0.975*** | 0.958*** | 0.909*** | 0.943*** | 0.915*** | 0.966*** | 0.940*** |
| | (0.076) | (0.077) | (0.075) | (0.076) | (0.074) | (0.075) | (0.075) | (0.074) | (0.074) |
| Class_Area 2 | −0.405*** | −0.491*** | −0.395*** | −0.432*** | −0.487*** | −0.454*** | −0.444*** | −0.458*** | −0.418*** |
| | (0.066) | (0.070) | (0.067) | (0.068) | (0.067) | (0.067) | (0.066) | (0.065) | (0.066) |
| Class_Area 3 | −0.072 | −0.210** | −0.031 | −0.141* | −0.146* | −0.151* | −0.120 | −0.205** | −0.126 |
| | (0.079) | (0.082) | (0.080) | (0.081) | (0.080) | (0.080) | (0.080) | (0.079) | (0.079) |
| HouseRatio | −0.073** | −0.009 | −0.115** | −0.101** | −0.121** | −0.118** | −0.103** | −0.127*** | −0.057* |
| | (0.037) | (0.040) | (0.040) | (0.038) | (0.039) | (0.038) | (0.036) | (0.033) | (0.033) |
| Log(theta) | −0.729*** | −0.669*** | −0.506*** | −0.644*** | −0.546*** | −0.617*** | −0.636*** | −0.745*** | −0.623*** |
| | (0.052) | (0.053) | (0.049) | (0.054) | (0.051) | (0.052) | (0.052) | (0.053) | (0.051) |

**Table 3** (continued)

| Zero Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | 10.455*** | 10.365*** | 10.251*** | 10.450*** | 10.527*** | 10.579*** | 10.691*** | 10.613*** | 10.542*** |
| | (0.122) | (0.124) | (0.124) | (0.125) | (0.125) | (0.125) | (0.125) | (0.122) | (0.124) |
| Driving_Time | −2.519*** | −2.530*** | −2.490*** | −2.518*** | −2.525*** | −2.530*** | −2.539*** | −2.495*** | −2.484*** |
| | (0.026) | (0.027) | (0.027) | (0.027) | (0.027) | (0.027) | (0.027) | (0.026) | (0.026) |
| Prov_Cap_Ori | 2.214*** | 2.183*** | 2.158*** | 2.107*** | 2.176*** | 2.059*** | 2.076*** | 2.045*** | 2.150*** |
| | (0.050) | (0.051) | (0.051) | (0.052) | (0.051) | (0.051) | (0.051) | (0.050) | (0.050) |
| Class_Area 2 | −1.811*** | −1.783*** | −1.875*** | −1.850*** | −1.816*** | −1.831*** | −1.775*** | −1.818*** | −1.794*** |
| | (0.041) | (0.042) | (0.042) | (0.042) | (0.042) | (0.042) | (0.042) | (0.041) | (0.042) |
| Class_Area 3 | −2.015*** | −2.015*** | −2.067*** | −2.053*** | −2.046*** | −2.043*** | −1.970*** | −2.013*** | −1.987*** |
| | (0.047) | (0.049) | (0.049) | (0.049) | (0.049) | (0.048) | (0.048) | (0.047) | (0.047) |
| Old_Ori | −0.001*** | −0.001*** | −0.002*** | −0.002*** | −0.002*** | −0.002*** | −0.002*** | −0.002*** | −0.002*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |

Notes: Standard errors in parentheses. N=670,761. *** p<0.01, ** p<0.05, * p<0.10

**Table 4** PPML Gravity Model with Origin and Year Fixed Effects

| Variable | Estimate | Std. Error |
|---|---|---|
| $Driving\_Time$ | $-2.654^{***}$ | 0.029 |
| $Digit\_Dest$ | $0.820^{***}$ | 0.073 |
| $Income\_pc\_Dest$ | $1.070^{***}$ | 0.187 |
| $Pop\_Ori$ | $1.533^{***}$ | 0.252 |
| $Prov\_Cap\_Dest$ | $0.861^{***}$ | 0.077 |
| $(Class\_Area\_Dest)2$ | $-0.686^{***}$ | 0.062 |
| $(Class\_Area\_Dest)3$ | $-0.836^{***}$ | 0.074 |
| *HouseRatio* | $-0.298^{***}$ | 0.059 |
| $Old\_Ori$ | $-0.00021$ | 0.00043 |
| Origin fixed effects | Yes | |
| Year fixed effects | Yes | |
| Dyad-clustered s.e | Yes | |

Notes: N = 670,761. *** p<0.01, ** p<0.05, * p<0.10

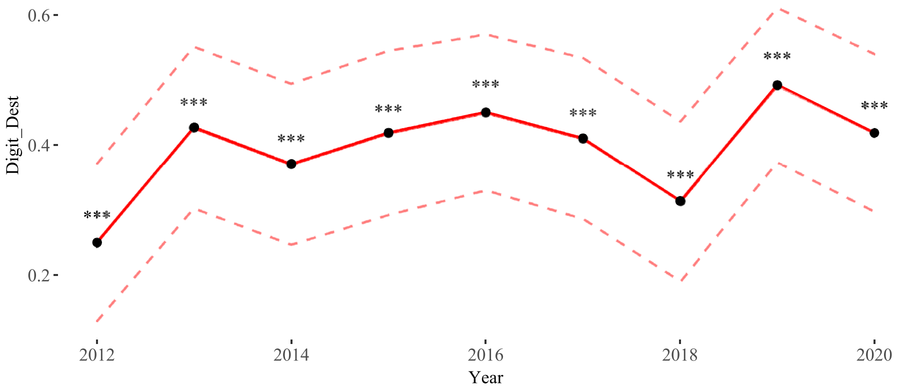


**Fig. 2** Evolution of the digitisation coefficient (2012-2020)

intended to inform the discussion on whether digital infrastructure may be relevant when considering policies aimed at mitigating peripheral depopulation. To do so, we employ the same Hurdle Negative Binomial Gravity Model specification, but we restrict the sample to hub-to-periphery flows only.

Even though, as expected, the highest number of peripheries are characterised by low levels in digitisation, an initial descriptive analysis already shows the strong impact that the digitisation level had in attracting flows between municipal hubs and peripheries. Figure 3 provides a map that describes migration pattern from municipal hubs to peripheral municipalities: The larger is the circle, the higher is the number of people that have moved from a municipal hub to that particular peripheral municipality, while the darker the colour of the circle is, the higher is the digitisation level of the municipality defined as low (0−0.4); medium (0.4 - 0.6); high (0.6-1).

Looking at the figure it appears quite clear that even if there are a lot of bright bubbles (defining municipalities with a low digitisation index), the largest bubbles are also the darkest, confirming the idea that when people move from a big city to

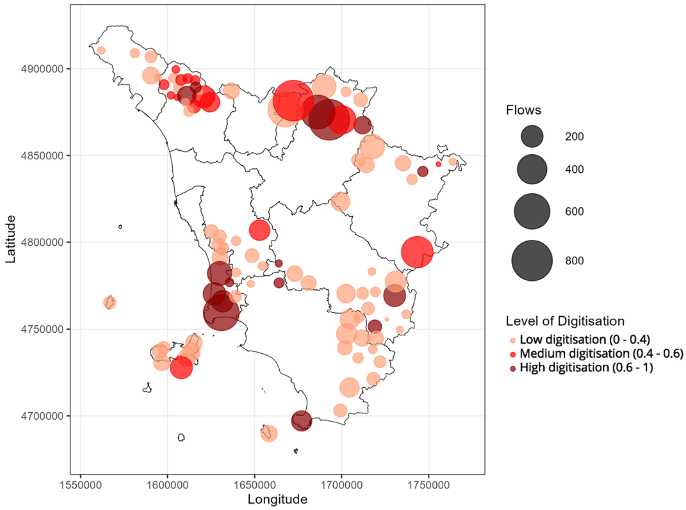


**Fig. 3** Migrations between cities and peripheries and digitisation level

**Table 5** Hurdle negative binomial gravity model

| Variable | Estimation | Std.Error | p-value | Model |
|---|---|---|---|---|
| (Intercept) | −15.0621 | 2.6198 | *** | Count |
| Driving_Time | −1.6283 | 0.0605 | *** | Count |
| Digit_Dest | 0.2092 | 0.1106 | * | Count |
| Income_pc_Dest | 1.2750 | 0.2716 | *** | Count |
| Pop_Ori | 0.8987 | 0.0420 | *** | Count |
| HouseRatio | −0.3553 | 0.0400 | *** | Count |
| (Intercept) | 3.3710 | 0.3060 | *** | Zero |
| Driving_Time | −1.5217 | 0.0513 | *** | Zero |
| Old_Ori | 0.0070 | 0.0010 | *** | Zero |
| Prov_Cap_Ori | 1.2632 | 0.0480 | *** | Zero |

Notes: N = 14,841. *** p<0.01, ** p<0.05, * p<0.10

a small periphery they, on average, tend to prefer peripheral municipalities that are characterised by a higher level of digitisation if compared to peripheral municipalities with a lower level. This general idea is also confirmed by the implementation of the Hurdle Model provided in Table 5.

As expected, this result indicates once again that, also within the subset of hub-to-periphery movements, peripheral municipalities with more developed digital infrastructure tend to attract relatively more inflows. In other words, when individuals relocate from major urban centres to peripheral areas, destinations with higher digitisation levels are more strongly associated with the observed migration patterns. While these estimates should be interpreted as conditional associations rather than causal effects, they remain consistent with the idea that digital infrastructure may represent a relevant feature of peripheral municipalities in the context of policies aimed at addressing depopulation dynamics.

## 5 Discussion

This study, employing flow matrices from 2012 to 2020, and implementing a "Hurdle Negative Binomial Gravity Model", has explored and investigated the dynamics of internal migration within Tuscany, posing a specific focus on the factors that have influenced people's migration choices. Results have shown that, while traditional variables still play a significant role, the level of digitisation emerged as a new influential factor in migration attractiveness. The evolution of the coefficient provided in Fig. 2 shows in fact that, for every year of reference, the digitisation level of a municipality has played a positive and statistically significant role, confirming the idea that, on average, municipalities with higher levels of digitisation attract more people than those with lower levels. This trend is also confirmed in movements from cities to peripheries, where citizens who moved from a city to a more peripheral municipality often opted for peripheries characterised by higher levels in digitisation. In conclusion, looking back at the question posed in the Introduction, even if it is certainly true what Shakespeare said - "a city is nothing but the people who live in it" - researchers and policy-makers should always consider that it is also true that people prefer to live in cities characterised by advanced technology and higher levels of digitisation.

**Funding** Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement.

### Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.



## References

Agresti A (2015) Foundations of linear and generalized linear models. Wiley Series in Probability and Statistics, Hoboken, New Jersey

Alm J, Winters J (2009) Distance and intrastate college student migration. Econ Educ Rev 28(6):728–738

Bailey TC, Gatrell AC (1995) Interactive spatial data analysis. Prentice Hall, Harlow Essex, England

Basile R, Causi M (2005) Le determinanti dei flussi migratori nelle province italiane: 1991-2001. Department of Economics - University Roma Tre, Departmental Working Papers of Economics - University 'Roma Tre' XLI

Beine M, Bertoli S, Fernández-Huertas Moraga J (2016) A practitioners' guide to gravity models of international migration. The World Economy 39(4):496–512

Beine M, Bourgeon P, Bricongne JC (2019) Aggregate fluctuations and international migration. Scand J Econ 121(1):117–152. https://doi.org/10.1111/sjoe.12258

Belot M, Ederveen S (2012) Cultural barriers in migration between oecd countries. J Popul Econ 25:1–29. https://doi.org/10.1007/s00148-011-0356-x

Bunea D (2012) Modern gravity models of internal migration. The Case of Romania. Theoretical and Applied Economics pp 127–144

Cameron A, Trivedi P (2013) Regression analysis of count data. Cambridge University Press, Cambridge

Casacchia O, Reynaud C, Strozza S et al (2021) Internal migration patterns of foreign citizens in italy. Int Migr 60. https://doi.org/10.1111/imig.12946

Feng CX (2021) A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. J Stat distrib and Appl 8(1):8

Florax R, Rietveld P, Sá C (2004) Determinants of the regional demand for higher education: A gravity model approach. Reg Stud 38:375–392. https://doi.org/10.2139/ssrn.384861

Karemera D, Oguledo V, Davis B (2000) A gravity model analysis of international migration to north america. Appl Econ 32(13):1745–1755. https://doi.org/10.1080/000368400421093

Linneman H (1966) An econometric study of international trade flows. North Holland Publishing Company, Amsterdan, Holland

Manzoor W, Safdar N, Mahmood HZ (2021) A gravity model analysis of international migration from bric to oecd countries using poisson pseudo-maximum likelihood approach. Heliyon 7(6):e07357. https://doi.org/10.1016/j.heliyon.2021.e07357

Martin W, Pham CS (2020) Estimating the gravity model when zero trade flows are frequent and economically determined. Appl Econ 52(26):2766–2779

Newton I (1687) The mathematical principles of natural philosophy. Londini, Jussu Societatis Regiæ ac Typis Josephi Streater. Prostat apud plures Bibliopolas

O'Kelly M (2009) Spatial Interaction Models, Elsevier, pp 365–368. https://doi.org/10.1016/B978-008044910-4.00529-0

Pirani E, Marino MF, Petrucci A (2019) Correlates of inter-districts migrations in tanzania a gravity-type modeling approach. Statistica (Bologna) 79(2):201–221

Piras R (2016) A long-run analysis of push and pull factors of internal migration in italy. estimation of a gravity model with human capital using homogeneous and heterogeneous approaches. Papers in Regional Science 96:n/a–n/a. https://doi.org/10.1111/pirs.12211

Ramos R (2016) Gravity models: A tool for migration analysis. IZA World of Labor https://doi.org/10.15185/izawol.239

Ravenstein EG (1889) The laws of migration. J Roy Stat Soc 52(2):241–305

Reimann L, Jones B, Bieker N et al (2023) Exploring spatial feedbacks between adaptation policies and internal migration patterns due to sea-level rise. Nat Commun. https://doi.org/10.1038/s41467-023-38278-y

Serlenga L, Shin Y (2021) Gravity models of interprovincial migration flows in canada with hierarchical multifactor structure. Empir Econ 60:365–390

Silva S, João M, Tenreyro S (2004) The log of gravity. Rev Econ Stat 88:641–658

Szedmák B, Varga L (2025) Szabó RZ (2025) Digital transformation of public services: The case of the document management application. Int J Public Adm 10(1080/01900692):2520522

Tamaș A (2017) Romanian erasmus students' flows-gravity model approach. J Eastern Eur Res Busin and Econ 17:1–11

Tinbergen J (1962) Shaping the world economy; suggestions for an international economic. The Twentieth Century Fund

Wajdi N, Adioetomo SM, Mulder CH (2017) Gravity models of inter-regional migartion in indonesia. Bull Indones Econ Stud 53(3):309–320. https://doi.org/10.1080/00074918.2017.1298719

Wolver A, Burgard JP, Bresslein M (2018) Gravity models in r. Austr J Stat 47:16–38. https://doi.org/10.17713/ajs.v47i4.688