Geographical accessibility to bank branches and its relationship to new firm formation in Sweden via multiscale geographically weighted regression

Cynthia Sin Tian Ho · Mats Wilhelmsson

Accepted: 23 February 2022 / Published online: 22 March 2022 © The Author(s) 2022

Abstract The geographical accessibility to banks and its relationship with new firm formation can vary across space due to spatial contexts. With increasing bank branch closures in Sweden, it is even more critical to understand where interventions are needed and at which scale, to implement effective policy. Thus, spatial context is incorporated into the analysis with the use of the multiscale geographically weighted regression (MGWR) model. A two-stage least squares (2SLS) model is employed in which two instrumental variables are utilised to instrument the proximity to the nearest bank branch in 2013. The MGWR results show that the geographical distance to the nearest bank branch has a negative association with new firm formation across all Swedish municipalities. The results also show that the relationship between the geographical distance to bank branches and new formation is not spatially varying across space. A policy implication from the analysis shows it is ideal to focus on the geographical accessibility to bank branches as it would continue to play an important role in the financing of new firms, regardless of the location.

Keywords New firm formation · Bank branches · Accessibility · Multiscale · MGWR

Classification codes C31 · C36 · G21 · M13

Availability of data and material Not applicable

Code availability Not applicable

Cynthia Sin Tian Ho
cstho@kth.se

1 Department of Real Estate and Construction Management, KTH—The Royal Institute of Technology, Stockholm, Sweden
1 Introduction

The research on the relationship between new firm formation and geographical accessibility to banks has regained popularity in recent years (Backman 2015; Nguyen 2019; Ho and Berggren 2020; Agostino et al. 2021). However, an area that has received considerably less attention is the geographical accessibility to bank branches and its relationship with new firm formation across space in Sweden. Although bank loans are one of the most popular forms of finance for Swedish small and medium-sized enterprises (SMEs), almost half of Sweden’s local bank branches have closed between 1990 and 2010, especially in rural areas (Swedish Agency for Economic and Regional Growth 2017; Backman and Wallin 2018).

This paper adds to the limited literature on the geographical accessibility to banks and its relationship with new firm formation. In this paper, we analyse the relationships between the geographical accessibility to the nearest bank branch and new firm formation across all 290 Swedish municipalities in 2013 and by using another time frame in 2007 as a robustness test, with the use of multiscale geographically weighted regression (MGWR). This paper also addresses the issue of endogeneity by using a 2SLS model to instrument the physical accessibility to the nearest bank branch in 2013 with two instrumental variables in 2000. The fitted values from the 2SLS model are then used in the MGWR modelling.

By comparing local parameter estimates across space using the MGWR model instead of a global model or a geographically weighted regression (GWR) model, multiple spatial scales can be applied simultaneously to characterise spatial contexts. Moreover, MGWR has not yet been applied to model new firm formation determinants. The results of the MGWR models in this paper reveal how the relationship between the proximity to bank branches and new firm formation varies across each Swedish municipality.

We hypothesized that the relationship between new firm formation and the proximity to the nearest bank branch differs across space, with more localised bank influence in rural areas. We also argued that spatial contexts can influence the relationship between the proximity to the nearest bank branch and new firm formation, which causes it to vary across space. The results of this paper show empirical evidence of a negative association between the proximity to the nearest bank branch and new firm formation in all Swedish municipalities, regardless of the location. However, the results also show that the relationship between the geographical distance to bank branches and new formation is not spatially varying across space.

This paper is organized as follows. Section 2 presents a review of studies that investigates the relationship between geographical accessibility to bank branches and new firm formation. A discussion of the methodology follows in Section 3. Section 4 presents the data and the variables used in this paper. Section 5 presents the empirical results, namely the results obtained from a 2SLS model, GWR, and MGWR models. Section 6 concludes.
2 Relationship between geographical accessibility to bank branches and new firm formation

Geographical accessibility to banks at the local level has been argued by some researchers to be less important due to technological developments which reduce the “distance-related diseconomies” (Petersen and Rajan 2002; Berger 2003). Berger (2003) argued that advances in technologies help banks in loan monitoring between the firm and the bank at long distances, allowing banks to issue credits at a greater distance. Technological progress also helps banks in carrying out traditional banking services which require physical proximity. For example, the use of credit scoring to screen a potential loan applicant without the need for geographical proximity or local knowledge. Bank managers may also find it easier to monitor and control staff located at a long distance, with the use of technology.

Several studies examined the relationship between the geographical accessibility to banks and lending to new firms in a region (Petersen and Rajan 2002; Brevoort et al. 2010; Backman 2015; Nguyen 2019; Kärnä et al. 2021). A study by Petersen and Rajan (2002) examined the distance between small businesses and their bank lenders from 1973 to 1993. They found out that the average distance between the small firms and their bank lenders has increased and attributed the increase in distance to improvements in bank productivity. Advances in computing and communications have increased the availability and timeliness of hard information, which facilitates impersonal and distant lending. Hence, distant firms do not necessarily need to have the highest credit quality and even informationally opaque firms that are located far away can secure a loan from the bank (Petersen and Rajan 2002).

However, an updated study by Brevoort et al. (2010) compares the results from the same data source used by Petersen and Rajan (2002), using data in the previous decade from the Surveys of Small Business Finances in 1993, 1998 and 2003. The findings by Brevoort et al. (2010) provide conflicting evidence with the findings by Petersen and Rajan (2002). The results from the updated study also show that firms with higher credit quality and more experienced ownership realize larger gains in distance compared with other firms (Brevoort et al. 2010). Another finding from the updated study shows that the growth in the distance between relatively old firms and their lenders shows a greater increase over time compared with younger firms, which are more informationally opaque than older firms (Brevoort et al. 2010). Hence, the updated study shows that even as the distance between small firms and their lenders has increased overall, the importance of geographical distance in small business lending cannot be underestimated, especially for young firms. A Swedish study also shows that the interest rates increase with the distance to the lender while the loan size decreases with the distance (Kärnä et al. 2021).

Thus, some researchers argue that geographical accessibility to banks is still important at the local level to new firm lending and that the advancement in communication technology does not necessarily mean an increased exchange of quality information (Degryse and Ongena 2004). Loan officers have limited incentives to transfer information and even with better communication platforms like emails, they might find it difficult to harden the soft information or transmit the information which is too sensitive to move through the bank branch organisation (Liberti 2011).
Soft information is often harder to communicate to others as it requires multiple face-to-face meetings and trust, which are often developed through a long-term relationship between the banks and the borrowers. It takes time for the loan officer to gather knowledge about the borrower’s personality, the quality of his or her firm’s management and his or her relationships with customers (Uchida et al. 2012). Therefore, proximity between the banks and borrower firms gives the banks an informational advantage by reducing search and monitoring costs (Geanakopoulos and Milgrom 1991; Becker and Murphy 1992; Bolton and Dewatripont 1994). Nguyen’s (2019) study also highlights that geographical distance matters not only due to the improvement of accessibility but also in reducing the costs of information transmission.

Despite the advances in communication technology and Internet banking, banks continue to rely on personal relationships with borrower firms as it helps to facilitate monitoring and screening activities. Proximity helps in the forging of a long-term relationship between the lenders and the borrowers, and the closures of bank branches can have a disruptive effect on lender-specific relationships that have already been formed over a period of time (Nguyen 2019). The search for another lender located farther away and the formation of a new relationship with another lender takes a long time and effort. Whether the firm would be able to secure a loan with similar rates with another lender also involves uncertainty as the new lender does not know the borrower as well as the previous lender did. Moreover, soft information on distant firms is hard to obtain unless the loan officers who were previously in charge of the loan, transfer this soft information to the new lender (Sharpe 1990; Rajan 1992). However, previous loan officers have no incentives to do so due to the proprietary nature of soft information (Drexler and Schoar 2014).

Proximity also enables banks to access knowledge about the local market; distance erodes lenders’ ability to acquire valuable proprietary intelligence about potential borrowers in the local market (Berger and Udell 1995; Agarwal and Hauswald 2010). Several studies also emphasize the importance of a well-developed financial system at the local level, which can enhance the ability of a small firm to gain access to external financing (Cole et al. 2004; Arcuri and Levratto 2020). Small firms generally suffer from great difficulty in applying for loans due to the lack of track record (Cole et al. 2004; Arcuri and Levratto 2020).

In general, spatial contexts can influence the importance of geographic distance to bank branches on new firm formation but it is unclear in the literature which spatial scale the relationship between geographic distance to bank branches and new firm formation applies. For instance, the different spatial scales in which the relationship applies can be on a local, regional or global scale.

A limitation of previous studies is the assumption that each new firm formation determinant operates at the same spatial scale and it is more likely that the relationship may vary at different scales due to complex social, economic, and demographic factors (Backman 2015; Kärnä et al. 2021). For example, geographical accessibility to banks might be more important for new firm formation in rural areas than in urban areas. One factor for this difference in the magnitude of importance is the existence of alternative external financing in urban regions. The majority of the venture capital market is located in the two largest metropolitan cities in Sweden: Stockholm and
Gothenburg (SVCA 2017). New firms in urban regions have accessibility to a wider range of finance such as formal and informal venture capital and business capital while new firms in rural regions have fewer financing options (Avdeitchikova 2009). Therefore, firms in urban regions should face lesser credit constraints as they have easier access to finance, while firms in remote regions have to rely on a few local banks.

Furthermore, a Swedish study finds that in non-metropolitan areas, both soft and hard information is treated equally important by the loan officer for the credit risk assessment of a firm. While in metropolitan areas, the loan officer focuses a lot more on quantitative information like the annual financial statement (Silver 2001). This is because, in smaller towns, it is easier to receive and verify soft information from another actor in the social network (Agarwal and Hauswald 2010). Furthermore, Silver (2001) also finds that bank branch managers in non-metropolitan areas make greater use of social network forums to get hold of interesting prospects and to maintain their relationship with their customers. This local context influences the importance of geographical distance to bank branches in the credit risk assessment process, which can affect the chances of a new firm getting a bank loan.

Another factor that contributes to the local context is the industrial composition of the economy in the local region. Capital-intensive sectors such as transport, agriculture and manufacturing have a larger demand for external capital, hence creating more need for bank capital, which means that it is even more important to have good accessibility to a bank branch (Landström 2017). Furthermore, the nature of the industry is also another factor that contributes to the importance of acquiring soft information from the firm, which is then related to the importance of physical accessibility to banks (Silver 2001). In the metropolitan areas where the service industry takes the bulk of the industry composition, a firm visit by the loan officer usually meant less for firms in the knowledge-intensive service industries. Moreover, in smaller communities, the loan officer probably finds it easier to understand the impact of the local environment on a firm due to the prior local knowledge he possesses.

The hypothesis resulting from this section is as follows:

**H1** The influence of the physical accessibility to bank branches on new firm formation is spatially-varying across different Swedish municipalities, with higher importance in rural municipalities.

### 3 Methodology

#### 3.1 GWR model

A conventional global regression model assumes constant relationships across a study area and calculates the average relationship between the dependent variable and the independent variables across space. However, many spatial processes vary in reality, which makes the conventional global regression model unfit for estimation
The GWR model relaxes the spatial homogeneity assumption by allowing the parameter estimates to vary locally across the spatial units of analysis. The GWR model is represented in this paper by:

\[ y_r = \beta_{0r}(u_r, v_r) + \sum_{k=1}^{p} \beta_{kr}(u_r, v_r) x_{kr} + \varepsilon_r \]  

where \( y_r \) is new firm formation per capita calculated by the number of new firms formed per capita (NFFPC) at municipality \( r \), \( (u_r, v_r) \) represents the coordinates of municipality \( r \)'s centroid, \( \beta_{0r}(u_r, v_r) \) and \( \beta_{kr}(u_r, v_r) \) represent the local intercept and the coefficient of each independent variable in each municipality respectively. \( \varepsilon_r \) is the random error at municipality \( r \).

For the calibration of a GWR model, a spatial weighting matrix is calculated with three different elements, namely the weighting matrix, its bandwidth, and the type of distance matrix (Brunsdon et al. 1996). The weighting matrix assumes that areas that are closer to municipality \( r \) have more influence on municipality \( r \) than areas that are farther away (Chasco et al. 2008). Therefore, more weight is given to observations near municipality \( r \), and less weight is given to observations that are farther away. Two popular choices of weighting matrices are Gaussian and Bi-square kernel functions, both of which are spatial kernel functions (Brunsdon et al. 1998).

The bandwidth is calculated such that a certain proportion of observations that are nearest to municipality \( r \) will be included in the local regression for each municipality. An optimal bandwidth parameter is selected by minimizing a corrected Akaike information criterion (AICc), which strikes a balance between model variance and bias.

The adaptive Gaussian and Bi-square weighting functions are defined formally as:

**Gaussian:**

\[ w_{rq} = \exp \left( -\frac{1}{2} \left( \frac{\text{DIST}_{rq}^2}{h_r^2} \right) \right) \]  

**Bi-square:**

\[ w_{rq} = \begin{cases} \left[ 1 - \left( \frac{\text{DIST}_{rq}/h_r}{h_r} \right)^2 \right]^2, & \text{if } \text{DIST}_{rq} < h_r \\ 0, & \text{otherwise} \end{cases} \]  

where \( w_{rq} \) represents the weight assigned to the data for municipality \( q \) when considering the calibration of the model in municipality \( r \). \( \text{DIST}_{rq} \) is the distance between municipalities \( r \) and \( q \), and \( h_r \) represents the different bandwidths that consider the same proportion of municipalities to be included in the estimation of the regression model for each municipality \( r \). As for the distance matrix, GWR models usually use Euclidean distance in practice to measure spatial proximity.
3.2 Multiscale geographically weighted regression model (MGWR)

However, a drawback of the GWR is that the same spatial scale is applied to all the local relationships within the model, which means that the same bandwidth is used (Oshan et al. 2020). A fixed spatial scale is not valid in the case where the phenomena involve multiple spatial processes with distinct spatial scales (Mansour et al. 2021).

MGWR extends the GWR model by allowing each local relationship in the model to vary at a unique spatial scale and is calculated as follows:

\[ y_r = \sum_{j=1}^{m} \beta_{b_wj}(u_r, v_r)x_{rj} + \epsilon_r \]  

where \( b_wj \) represents the bandwidth for each of the spatial processes being modelled and is calibrated by using a back-fitting algorithm (Fotheringham et al. 2017). In comparison to GWR, MGWR has much less restrictive assumptions as the relationship between the response variable and a covariate is allowed to vary locally, vary regionally, or not vary at all (Oshan et al. 2019). By allowing each relationship in the model to vary at a unique spatial scale, eliminates over-fitting, decreases the bias in the parameter and mitigate collinearity issues (Oshan et al. 2019). Therefore, MGWR has been recommended by many scholars to be used in place of GWR in investigating process spatial heterogeneity and scale (Fotheringham et al. 2017; Oshan et al. 2019).

4 Empirical data

4.1 Spatial unit of analysis

Sweden is divided into 290 municipalities, 5984 demographical areas, and 227,235 included areas. The included areas are represented by square areas of 250 m by 250 m in urban areas and 1000 m by 1000 m in rural areas as rural areas are usually more sparsely distributed. Fig. 1 shows the representation of a Swedish municipality—Stockholm in different spatial units of reference. The empty spaces as shown in Fig. 1 are uninhabited areas such as forests, nature reserves, and lakes.

The municipal level is chosen as the spatial unit of analysis in this study as municipalities are the smallest local authorities in Sweden, where policies are being carried out. Hence, the municipal level is considered to be localised and there can be significant differences across municipalities in the same region (Eliasson 2016). A larger spatial unit of reference relative to the municipal level like the functional region carries a larger risk of aggregation error. Furthermore, Backman (2015) found no significant relationship between the number of bank branches per capita and new firm formation on the functional regional level.

Smaller spatial units of reference like the demographical area and included areas can help to capture a higher degree of localness as compared to the municipal level. However, it is harder to obtain the data on a finer spatial level; some data

\[ \text{Springer} \]
are protected by personal data protection laws due to confidentiality issues, making it not possible to compare data sources from different units of reference (Eurostat 2021). Moreover, the primary data at the granular level consists of sparse counts of new firm formation and other variables, which makes it impractical to study small-scale geographical variation.

Thus, we utilize municipal level data in this study. The accessibility to bank branches is calculated by making use of the small included areas contained within the municipalities. In this way, we account for the spatial distribution of the population inside the municipality and minimise aggregation error when calculating the geographical proximity to bank branches on the municipal level.

4.2 Variables

The locations of all the local bank branch offices are captured in the form of geographical coordinates with data gathered from Argomento GIS & IT. Data on new firm formation are gathered from the Swedish Agency for Growth Policy Analysis, which do not include existing firms that are reorganised. Data related to the population are acquired from Statistics Sweden and they include information on the human capital level, changes in housing prices and the proportion of immigrants on the municipal level. From the Swedish Public Employment Service’s website, unemployment data are also obtained. Information about the firms’ perception of public attitudes towards entrepreneurship in each municipality is also obtained from an annual survey conducted by the Federation of Swedish Enterprise.
Table 1: Description of variables

| Variable                                                | Description                                                                 | Expected sign |
|----------------------------------------------------------|-----------------------------------------------------------------------------|---------------|
| New firm formation per capita (NFFPCLabour)              | Number of new firms per 1000 people in the labour force in the municipality | Dependent sign |
| Weighted mean distance to banks, weighted by the labour population (WeightedDistLabour) | Weighted mean distance to the nearest bank branch in the municipality weighted by the labour population | –             |
| Weighted mean distance to banks, weighted by the firm population (WeightedDistFirms) | Weighted mean distance to the nearest bank branch in the municipality weighted by the firm population | –             |
| Firm density (FirmDensity)                              | Number of firms divided by the population in the municipality                | +             |
| Establishment size (EstSize)                            | Number of employees divided by the number of firms in the municipality       | –             |
| Human capital level (HumanCap)                          | Proportion of the labour force with at least three or more years of university education, in the municipality | +             |
| Unemployment rate (UnempRate)                           | Proportion of the labour force that is unemployed in the municipality       | –             |
| Theil index (TheilIndex)                                | Measure of diversity of the industries in the municipality                  | +             |
| Specialization index (SpecIndex)                        | Measure of specialisation of the industries in the municipality             | +             |
| Percentage of immigrants (ImmigrantsShare)              | Number of immigrants who are living in the municipality divided by the total population in the municipality (in %) | +             |
| Change in housing prices (ChangeHP)                     | Percentage change in housing price in 2013 and that of four years ago       | +             |
| Income growth rate (IncomeGrowth)                       | Average rate of change in annual income per capita between 2013 and that of two years ago | +             |
| Entrepreneurial attitudes (EntrepreneurialAttitudes)    | Perception of public attitudes towards entrepreneurship                      | +             |

Detailed calculations for TheilIndex, SpecIndex and IncomeGrowth can be found in the appendix.

Table 1 summarises the description and expected signs of all the variables used in the models.

4.2.1 Dependent variable: New firm formation per capita

NFFPC is calculated by the number of new firms formed per capita and NFFPC can be calculated using two different approaches. NFFPC calculated using the labour market approach (NFFPCLabour) is the number of new firms formed per capita in the labour population per 1000 people since firms are usually started by people from the labour population and in areas close to where they live (Audretsch and Fritsch 1994a). The age group between 16 and 64 years old is used as a proxy for the labour population. Fig. 2 shows the NFFPCLabour calculated using the labour market approach in 2013. The map shows that the NFFPCLabour is higher in metropolitan municipalities and in municipalities that are popular for tourism.
An alternative approach to calculate $NFFPC$ is the ecological approach which standardises the number of new firms by the number of existing firms ($NFFPCEco$) (Audretsch and Fritsch 1994b). However, the ecological approach has been criticised by several researchers as it gives off higher values in areas with a larger mean firm size as the number of new firms are standardised by a smaller number of firms that are larger (Garofoli 1994). As a robustness test, the analysis is also conducted using $NFFPCEco$ as a dependent variable.

4.2.2 Independent variable: Weighted mean distance to the nearest bank branch

To measure the accessibility to bank branches, the following formula is used to calculate the weighted mean distance to the nearest bank branch for each municipality $y$:

$$\text{WeightedDist}_y = \left( \frac{\sum (D_{xy} \cdot P_x)}{\sum P_{xy}} \right),$$

where $D_{xy}$ is the driving distance to the first nearest bank branch from each included area $x$ in municipality $y$, and $P_x$ represents the population in the included area $x$ and $\sum P_{xy}$ represents the total population in all the included areas $x$ in municipality $y$. Street network analysis is used to identify the shortest driving route from each area to the nearest bank branch.
included area \( x \) to its first nearest bank branch. By weighing the distance to the first nearest bank branch with the population, the distribution of the population within the municipality is considered. Moreover, this measure assumes that the higher the number of members in the population, the higher the number of potential new firms that would apply for a bank loan. Hence, the relative location of the population and bank branches will determine the pattern of spatial accessibility to the bank branches.

Similar to the labour market and ecological approaches in calculating \( NFFPC \), the labour market population and the existing firms’ population is considered in the calculation of the \( WeightedDist \) as different denominators in Eq. 5. The labour population is considered as bank loans are often applied by individuals from the labour market where they live since they are most likely to start a new firm.

The existing firm population is considered as the establishment of new firms might not necessarily follow the residential pattern of the labour force population but might instead follow the pattern of the locations with high firm density. A higher firm density in the region means that there is a potentially favourable environment for setting up a business, which encourages potential entrepreneurs to set up their business in the same region. Hence, there could be a higher demand for bank loans by these entrepreneurs. A high firm density can also be advantageous for new firm formation due to a higher probability in the match between firms, labour, suppliers and customer, which lowers the transaction costs (Krugman 1991b, a). Furthermore, banks are often profit-driven and they may locate more branches in locations of high firm density (Okeahalam 2009). The weighted mean distance to the nearest bank branch weighted by the number of existing firms is included in the robustness test.

Fig. 3 shows the weighted mean distance to the nearest bank branch weighted by labour population (\( WeightedDistLabour \)) while Fig. 6 in the appendix shows the weighted mean distance to the nearest bank branch weighted by the number of existing firms (\( WeightedDistFirms \)). It is observed that the weighted mean distance to the nearest bank branches in both Fig. 3 and 6 is generally higher in northern Sweden than in southern Sweden as bank branches in the north are more sparsely distributed.

4.2.3 Reverse causality problem

However, we faced the problem of reverse causality as it was not possible to establish the direction of causality between the accessibility to the nearest bank branch variable and new firm formation. If the accessibility to the nearest bank branch is endogenous due to reverse causality, the estimate of the effect of the accessibility to the nearest bank branch on new firm formation will be biased and inconsistent. To address this endogeneity concern, an instrumental variable approach is adopted in this paper by manually estimating the MGWR model in the same way a global two-stage least squares (2SLS) is performed.

In the first stage, the accessibility to the nearest bank branch is regressed on two instrumental variables \( z_1 \) and \( z_2 \), and the other exogenous independent variables:
Fig. 3  WeightedDistLabour in 2013 in m (“Natural breaks” classification)

Regress WeightedDistLabour$_r$ on $z_1$, $z_2$, and $\sum_{k=3}^{p+2} x_{rk}$

$$\text{WeightedDistLabour}_r = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \sum_{k=3}^{p+2} \gamma_k (u_r, v_r) x_{rk}$$ \hspace{1cm} (6)

where $\gamma_k(u_r, v_r)x_{rk}$ is the coefficient for each exogenous independent variable $k$ ($k = 3, ..., p+2$) in municipality $r$.

Second stage of 2SLS: NFFPCLabour is regressed on the fitted values of WeightedDistLabour$_r$:

$$y_r = \beta_0 (u_r, v_r) + \beta_1 (u_r, v_r) \text{WeightedDistLabour}_r$$
$$+ \sum_{k=1}^{p} \beta_k (u_r, v_r) x_{rk} + v_r$$ \hspace{1cm} (7)
where $y_i$ is the $NFFPCLabour$ and $v_i$ is a composite error term that is uncorrelated with $WeightedDistLabour$.

The first instrumental variable of the $WeightedDistLabour$ in 2013 is a temporal lag of the $WeightedDistLabour$ variable in 2000 ($WeightedDistLabour_{2000}$). The second instrumental variable measures the weighted mean distance to the nearest savings bank weighted by the labour population in 2000 ($WeightedDistSavingsLabour_{2000}$). As pointed out by Backman (2015), the historical localisation of savings banks provides insight into the present localisation of primary savings banks and other banks. There is a high bivariate correlation of 0.401 between the $WeightedDistSavingsLabour_{2000}$ and $WeightedDistLabour$ in 2013. The original purpose of establishing the savings banks historically was to encourage people to save up for their retirement and to lower the cost of almshouses, which did not follow the strict logic of profit (Backman 2015). Furthermore, the historical localisation of savings banks is unrelated to the present $NFFPCLabour$ in 2013, which is supported by its low bivariate correlation of –0.054.

The manual estimation of Eq. 6 and 7 is not ideal as the standard errors of the locally varying parameter estimates that are obtained from a manual second-stage of the 2SLS MGWR model are likely to be incorrect. However, in the current MGWR framework, there is no known solution that can deal with both endogeneity and incorrect standard errors of a manual 2SLS simultaneously (Bilgel 2020). There are three possible options as explained by Bilgel (2020) who weighed the different options. One option is to ignore endogeneity by estimating a conventional MGWR model as in Eq. 4 (i.e., locally varying but biased and inconsistent parameter estimates if the accessibility to the nearest bank branch is truly endogenous). Another option is to estimate a manual 2SLS MGWR model as in Eq. 6 and 7 (i.e., locally varying and unbiased but imprecise parameter estimates due to incorrect standard errors.). The third option is to estimate a global 2SLS model (i.e., unbiased but erroneously fixed and monotonic estimates with correct standard errors). Since the objective of this paper is to estimate the effect of the accessibility to bank branches on new firm formation for each municipality, we proceed with a manual 2SLS MGWR model that has a small bias due to the potentially incorrect standard error. Since the MGWR is linear, the manual estimation by plugging the fitted values from the first stage leads to a consistent estimator.

The first stage results in the 2SLS model for $WeightedDistLabour$ is shown in Table 5. Both the instruments $WeightedDistLabour_{2000}$ and $WeightedDistSavingsLabour_{2000}$ fulfill the relevance condition because they are correlated with the endogenous variable, which is verified by significant estimates in the first-stage equation, high pairwise correlations, and high F-value (above 10), leading to a rejection of the null hypothesis of weak instruments. The validity condition implies that the instruments are not correlated with the error term. Similar instruments have been utilised in previous empirical work (Guiso et al. 2004; Alessandrini et al. 2010; Backman 2015).

The Sargan test is also conducted for testing of instrument exogeneity using overidentifying restrictions since there are more instrumental variables than endogenous variables and rejecting the null hypothesis means that at least one of our instruments is invalid. The null hypothesis of the Sargan test is not being rejected as the $p$-
value of this test are above 5 per cent for the estimated model. By using a Hausman test, we reject the null hypothesis that there is a systematic difference between the coefficients in the OLS and 2SLS models at the 5% level, supporting the use of the 2SLS model.

4.2.4 Control variables: other determinants of new firm formation

Several additional explanatory variables were also obtained and included in the model explaining new firm formation as they were shown to determine regional differences in \textit{NFFPC} in previous studies. These include firm density, establishment size, human capital level, unemployment rate, industry diversity index, industry specialization index, the percentage of immigrants, change in housing prices, income growth and entrepreneurial attitudes. The formulas for the calculation of the industry diversity index, industry specialisation index and income growth rate are found in the appendix.

5 Estimation procedure and empirical results

Log transformations are applied to all the variables to make the distribution of the variables more normal except for housing price change as it contains negative values. The dependent variable used in the models in this section is \textit{NFFPCLabour}. The independent variables used are the \textit{WeightedDistLabour}, firm density (\textit{Firm-Density}), establishment size (\textit{EstSize}), human capital (\textit{HumanCap}), unemployment rate (\textit{UnempRate}), regional industry diversity index (\textit{TheilIndex}), specification index (\textit{SpecIndex}), the share of immigrants in the municipality (\textit{ImmigrantsShare}), change in housing price (\textit{ChangeHP}), income growth (\textit{IncomeGrowth}) and entrepreneurial attitudes (\textit{EntrepreneurialAttitudes}). All the variables are measured in 2013.

5.1 Findings

Results from the second stage of the 2SLS model in Table 2 are first reported to provide context for the GWR and MGWR results. The independent variables do not pose the issue of multicollinearity based on their global variance inflation factors (VIFs) when evaluated against one another (i.e., less than 5) (O’Brien 2007). Overall, 8 out of 11 variables have significant relationships with \textit{NFFPCLabour} with \textit{p}-values less than 0.1 in Table 2. Table 2 also shows that \textit{WeightedDistLabour} has a significant negative relationship with \textit{NFFPCLabour}.

The global 2SLS model has a moderately low adjusted R-squared of 0.468, which indicate that about 54% of the variation in \textit{NFFPCLabour} across Swedish municipalities cannot be explained by the variables chosen in this study. One of the unknown factors that can increase the explanation of the variation in \textit{NFFPCLabour} is the spatial non-stationarity that may exist in the relationships between the dependent and the independent variables. As the global 2SLS model assumes that the relationships between the dependent and the independent variables are not spatially varying, the
Table 2  Second stage of 2SLS model

|                      | NFFPCLabour |
|----------------------|-------------|
| Intercept            | 3.986*** (0.587) |
| WeightedDistLabour (ln) | -0.109*** (0.039) |
| FirmDensity (ln)     | 0.171** (0.068)  |
| EstSize (ln)         | -0.143*** (0.046) |
| HumanCap (ln)        | 0.312*** (0.049)  |
| UnempRate (ln)       | -0.181*** (0.044) |
| TheilIndex (ln)      | 0.639*** (0.245)  |
| SpecIndex (ln)       | -0.005 (0.036)   |
| ImmigrantsShare (ln) | 0.14*** (0.032)   |
| ChangeHP             | 0 (0.001)       |
| IncomeGrowth (ln)    | -0.107* (0.062)  |
| EntrepreneurialAttitudes (ln) | -0.042 (0.163) |

Number of observations 290
P-value 0.000
Adjusted R-squared 0.468

Robust standard errors in parentheses
*, **, and *** show significance levels at 10%, 5% and 1%, respectively.

The calibration of an MGWR model creates a set of optimal bandwidths which describe the scale for each spatial varying process in the model. As compared to the single bandwidth of 192 nearest neighbours in the GWR models, the bandwidths for each explanatory variable are listed in Table 4. The MGWR models using the Gaussian and Euclidean weighting matrices have a much larger improvement in the adjusted R-squared of 0.723 and 0.683 respectively.

It is observed that for both MGWR models which are based on a Euclidean distance matrix, the Gaussian and Bi-square weighting functions give the same set of bandwidths for each process in the model. Though the parameter estimates from both models 3 and 4 in Table 4 are the same in terms of magnitudes and signs, model 3 which is based on a Gaussian weighting function has the lower AICc of the two models, which means that it has a better fit. Therefore, the subsequent discussion of the results is devoted to the MGWR model based on a Gaussian weighting function.
### Table 3  GWR Results

|                                      | Model 1 (Gaussian, Euclidean) | Model 2 (Bi-square, Euclidean) |
|--------------------------------------|-------------------------------|--------------------------------|
|                                      | Min   | Median | Max   | Bandwidth | Min   | Median | Max   | Bandwidth |
| Intercept                            | -0.015 | 0.008  | 0.019 | 192       | -0.289 | -0.040 | 0.146 | 192       |
| WeightedDistLabour (ln)              | -0.217 | -0.186 | -0.151 | 192       | -0.401 | -0.227 | 0.022 | 192       |
| EstSize (ln)                         | -0.272 | -0.231 | -0.194 | 192       | -0.352 | -0.161 | 0.023 | 192       |
| FirmDensity (ln)                     | 0.184  | 0.226  | 0.253  | 192       | -0.049 | 0.304  | 0.542 | 192       |
| HumanCap (ln)                        | 0.384  | 0.455  | 0.474  | 192       | 0.194  | 0.413  | 0.533 | 192       |
| UnempRate (ln)                       | -0.265 | -0.210 | -0.179 | 192       | -0.381 | -0.179 | 0.049 | 192       |
| TheilIndex (ln)                      | 0.131  | 0.158  | 0.179  | 192       | -0.001 | 0.161  | 0.391 | 192       |
| SpecIndex (ln)                       | -0.048 | -0.019 | 0.022  | 192       | -0.175 | -0.020 | 0.171 | 192       |
| ImmigrantsShare (ln)                 | 0.217  | 0.262  | 0.301  | 192       | -0.019 | 0.132  | 0.511 | 192       |
| ChangeHP (ln)                        | -0.005 | 0.011  | 0.020  | 192       | -0.087 | -0.001 | 0.058 | 192       |
| IncomeGrowth (ln)                    | -0.096 | -0.082 | -0.056 | 192       | -0.207 | -0.062 | 0.066 | 192       |
| EntrepreneurialAttitudes (ln)       | -0.064 | -0.043 | 0.007  | 192       | -0.107 | -0.008 | 0.145 | 192       |
| Number of observations               | 290 municipalities            | 290 municipalities            |
| Adjusted R-squared                   | 0.496                          | 0.559                         |
| AIC                                  | 643.231                        | 624.627                       |
| AICc                                 | 645.841                        | 638.740                       |

### Table 4  MGWR Results

|                                      | Model 3 (Gaussian, Euclidean) | Model 4 (Bi-square, Euclidean) |
|--------------------------------------|-------------------------------|--------------------------------|
|                                      | Min   | Median | Max   | Bandwidth | Min   | Median | Max   | Bandwidth |
| Intercept                            | -0.631 | 0.012  | 0.906 | 38        | -0.631 | 0.012  | 0.906 | 38        |
| WeightedDistLabour (ln)              | -0.130 | -0.129 | -0.124 | 288       | -0.130 | -0.129 | -0.124 | 288       |
| EstSize (ln)                         | -0.145 | -0.135 | -0.128 | 288       | -0.145 | -0.135 | -0.128 | 288       |
| FirmDensity (ln)                     | -0.012 | 0.445  | 0.790 | 68        | -0.012 | 0.445  | 0.790 | 68        |
| HumanCap (ln)                        | 0.352  | 0.372  | 0.378 | 288       | 0.352  | 0.372  | 0.378 | 288       |
| UnempRate (ln)                       | -0.178 | 0.032  | 0.097 | 154       | -0.178 | -0.032 | 0.097 | 154       |
| TheilIndex (ln)                      | -0.075 | 0.021  | 0.122 | 172       | -0.075 | 0.021  | 0.122 | 172       |
| SpecIndex (ln)                       | -0.087 | -0.034 | 0.008 | 262       | -0.087 | -0.034 | 0.008 | 262       |
| ImmigrantsShare (ln)                 | -0.247 | 0.160  | 0.395 | 70        | -0.247 | 0.160  | 0.395 | 70        |
| ChangeHP (ln)                        | -0.130 | -0.034 | 0.115 | 160       | -0.130 | -0.034 | 0.115 | 160       |
| IncomeGrowth (ln)                    | -0.092 | -0.085 | -0.024 | 284       | -0.092 | -0.085 | -0.024 | 284       |
| EntrepreneurialAttitudes (ln)       | 0.000  | 0.004  | 0.074 | 288       | 0.000  | 0.004  | 0.074 | 288       |
| Number of observations               | 290 municipalities            | 290 municipalities            |
| Adjusted R-squared                   | 0.723                          | 0.683                         |
| AIC                                  | 472.154                        | 539.816                       |
| AICc                                 | 475.640                        | 566.080                       |
5.1.1 Local effects of WeightedDistLabour on NFFPCLabour

Fig. 4 maps the coefficients of the GWR and MGWR for the WeightedDistLabour variable, with the parameter estimate surfaces for GWR on the left and MGWR model on the right. The grey shaded regions on the maps are not statistically different from zero as their confidence intervals overlap with zero. In both the GWR and MGWR models, similar patterns are exhibited for the WeightedDistLabour in explaining the spatial distribution of the NFFPCLabour at the municipal level as the coefficient estimates of the WeightedDist variable do not vary much across the study area, as observed in Fig. 4. All statistically significant parameter estimates of WeightedDistLabour on NFFPCLabour are of the same sign, which indicates spatial monotonicity. The negative sign indicates that a long distance to the nearest bank branch is associated with a lower new firm formation in municipalities where the effect of WeightedDistLabour on NFFPCLabour is statistically significant.

5.1.2 Local effects of other variables on NFFPCLabour

Based on the results of the MGWR model with Gaussian weighting function, six relationships occur at an effectively global scale (WeightedDistLabour, EstSize, HumanCap, SpecIndex, IncomeGrowth, and EntrepreneurialAttitudes) with large bandwidths indicating almost all the municipalities are included in each local subset. Five processes seem to occur at a regional scale (FirmDensity, UnempRate, Theil-Index, ImmigrantsShare, and ChangeHP) with relatively smaller bandwidths of more than 50 nearest neighbours and only one process varies locally (the intercept) with a relatively small bandwidth of 38 nearest neighbours.
Overall, MGWR provides a better model fit than GWR as it provides a lower AIC and AICc value than GWR and a much higher R-squared than GWR. MGWR is also less prone to multicollinearity problems, which is supported by the lower local condition numbers compared to GWR, which are all well below the rule-of-thumb of 30, as shown in Fig. 5 (Wheeler and Tiefelsdorf 2005; Wheeler 2007).

5.2 Robustness check

Three different robustness tests are conducted. The first robustness test is by analysing with the use of the data from 2007, the second robustness test is conducted by using the WeightedDistFirms variable along with other data in 2013, and the third robustness test is by utilising an ecological approach for calculating NFFPC in 2013 as the dependent variable along with other data in 2013. All three robustness tests are conducted with a Gaussian weighting function and a Euclidean distance matrix. The results of the MGWR models for all three robustness tests are reported in Table 6. The results are mapped for visualisation of patterns in Fig. 7, with the parameter estimate surfaces for GWR on the left and MGWR model on the right.

For all three robustness tests, the relationship between the weighted mean distance to the nearest bank branch variable and NFFPC occur at an effectively global scale with a bandwidth of 288 which include almost all the municipalities in each local subset, which is consistent with the results in Table 4. However, when inspecting the MGWR parameter estimate surfaces on the right side of Fig. 7, they exhibit different patterns in all three robustness tests. The second robustness test show results which are consistent with model 3 from Table 4.
By comparing the results between the GWR model and MGWR model in the second robustness test, the MGWR surfaces for \textit{WeightedDistFirms} display little-to-no spatial heterogeneity and are statistically non-zero, as observed in Fig. 4. Moreover, the MGWR parameter estimates for \textit{WeightedDistFirms} in the second robustness tests are between $-0.177$ to $-0.167$, which are higher than the MGWR parameter estimates for \textit{WeightedDistLabour} in model 3 from Table 4.

The MGWR parameter estimates for \textit{WeightedDistLabour} for the other two robustness tests hardly show any statistical significance. For the first robustness test where the data in 2007 is used, the MGWR parameter estimates for \textit{WeightedDistLabour} in the first robustness test are between $-0.047$ to $-0.010$, which are 3 times lower than the parameter estimates for \textit{WeightedDistLabour} in model 3 from Table 4. The number of bank branches has decreased drastically in Sweden from 2007 to 2013, especially in the countryside where about 13% of the bank branches has shut down over the years. Thus, possible reasons could be the drastic decrease in the number of bank branches and the financial crisis in 2008 which have resulted in a stronger association between the weighted mean distance to bank branches and \textit{NFFPCLabour} in 2013. A finding by a recent paper shows that an external event such as a financial crisis can affect the relative importance of location factors (Cruz and Teixeira 2021).

As for the third robustness test using \textit{NFFPCEco} as a dependent variable, the MGWR parameter estimates for \textit{WeightedDistLabour} are between $-0.08$ to $-0.071$ which are also lower than the MGWR parameter estimates for \textit{WeightedDistLabour} in model 3 from Table 4. Only some rural municipalities show statistically non-zero estimate surfaces for \textit{WeightedDistLabour} in the third robustness test. The calculation of \textit{NFFPC} using an ecological approach can however be misleading due to a small denominator problem in regions with a small number of new firm formations and an even smaller number of existing firms, which might bias the results (Garofoli 1994).

6 Conclusion

This paper demonstrates the potential of MGWR to improve our understanding of how the weighted mean distance to the nearest bank branch influences the \textit{NFFPC}. The results show that a mix of global and local processes can best model new firm formation, which provides a richer quantitative representation of the determinants compared to both GWR and global models. The MGWR model is also able to overcome the problems of GWR due to multicollinearity, which has also a higher model fit. As MGWR is a multiscale method, it may be useful for facilitating the development of more specific policy development by framing new firm formation determinants through a mix of global, regional, and local spatial contexts (Oshan et al. 2020).

It is important to note that the purpose of the MGWR model is not to claim causality but to explore the spatial variation of the relationship between new firm formation and its independent variables, which can have important policy implications. The results of the MGWR model in this paper reject our hypothesis that the

\[\text{Equation}\]
relationship between new firm formation and physical accessibility to banks differs across space, which means that regardless of the location, the effects of the proximity to the nearest bank branch on new formation is the same.

A policy implication resulting from this finding is that it is important to focus on the geographical accessibility to bank branches in all the Swedish municipalities. A state-owned company called Almi, provides loans, venture capital and advice for start-ups and established companies. They have 16 regional subsidiaries in Sweden, as of February 2022. The business loan that Almi provides only acts as a complement to the company’s financing solution, where the bank is expected to be the other financier (Almi 2021). Almi also charges a higher interest rate which is above the average bank interest rate. They usually put together the entire financing solution in collaboration with the bank. Given the negative association between the weighted mean distance to the nearest bank branch and new firm formation in all Swedish municipalities and the complementary nature of the business loan provided by Almi, this means that the geographical accessibility to bank branches would continue to play an important role in the financing of firms in Sweden, regardless of the location. Furthermore, it would be beneficial to new companies if Almi can lower their interest rates so that new companies would not be intimidated by the high interest rate charged by Almi.

One of the three robustness tests also shows consistent results in terms of the negative global association while showing insignificant results in the other two robustness tests. The robustness test which uses 2007 data shows that the relationship between \( \text{WeightedDistLabour} \) and \( \text{NFFPCLabour} \) can change over time due to external changes such as the increasing closures of bank branches. Another finding in the paper is that some determinants vary in the parameter estimates over space. For instance, \( \text{FirmDensity}, \text{UnempRate}, \text{TheilIndex}, \text{ImmigrantsShare} \) and \( \text{ChangeHP} \) show regional and local variation in the parameter estimates over space. Hence, localized policy interventions can be designed for these variables.

To extend this research in the future, additional determinants can be identified to explore within MGWR models of new firm formation. Furthermore, the outcome of this research could be operationalized by collaborating with policymakers to formulate, deploy and evaluate specific policies to increase entrepreneurship and increase the accessibility to financing for entrepreneurs. Lastly, similar MGWR model specifications can be applied in other developed countries which are reliant on banks such as Germany and Norway to validate the conclusion obtained in this paper and to compare the results for other developed countries. These would allow us to increase our understanding of the multiscale processes like the proximity to banks on new firm formation and increase our ability to plan policy and increase entrepreneurship in the long run.
7 Appendix

7.1 Equations

The regional industrial diversity index \((\text{TheilIndex})\) for municipality \(r\) is:

\[
\text{TheilIndex}_r = \sum_i \frac{E_{ir}}{E_r} \log \frac{E_r}{E_{ir}},
\]

where \(E_{ir} = \frac{\sum \text{emp}_{ir}}{\sum \text{emp}_{ir}}, E_r = \frac{\sum \text{emp}_{ir}}{\sum \text{emp}_{ir}}, E_{ir} = \frac{\text{emp}_{ir}}{\sum \text{emp}_{ir}}\) and \(\sum \sum E_{ir} = 1\).

\(\text{emp}_{ir}\) is the number of people employed in the industrial sector \(i\) in municipality \(r\). Six broadly defined sectors are defined and they are namely, “primary”, “manufacturing”, “trade, hotels and restaurants”, “business services”, “public and personal services” and “others”. If there is only one industrial sector in municipality \(r\), the Theil index is given a value of 0. If all six industrial sectors have an equal number of employed people in municipality \(r\), the Theil index is given a value of \(\log(6)\). Thus, the Theil index ranges from 0 to \(\log(6)\). To bound this index in the \([0, 1]\) interval, the Theil index obtained from Eq. 8 is further divided by \(\log(6)\).

The formula for the specialisation index \((\text{SpecIndex})\) is as follows:

\[
\text{SpecIndex}_r = \frac{1}{2} \sum_i \left| \frac{\text{emp}_{ir}}{E_r} - \frac{\text{emp}_{in}}{E_n} \right|,
\]

where \(\text{emp}_{ir}\) is the number of employees in industrial sector \(i\) in municipality \(r\) and \(\text{emp}_{in}\) is the number of employees in industrial sector \(i\) on the national level. \(E_r\) and \(E_n\) represent the total employment at the municipal and national level, respectively. The specialization index ranges from 0 to 1. If the industrial sector structure of the municipality is the same as the industrial sector structure on the national level, the specialization index is given a value of 0. If there only exists one sector in the municipality, the specialization index is given a value of 1.

Income growth \((\text{IncomeGrowth})\) is calculated as follows:

\[
\text{IncomeGrowth}_t = \sqrt{\frac{\text{MeanIncome}_t}{\text{MeanIncome}_{t-2}}} - 1,
\]

where \(\text{IncomeGrowth}_t\) is the average change in the mean income during the last two years.

Note 1 All GWR and MGWR models are built using the \textit{mgwr} library in the Python Spatial Analysis Library (PySAL).

2 Natural breaks classification is used to accurately portray trends found in the data with high variance (Slocum et al. 2008).
Table 5  
First stage of 2SLS

| Variable                                  | WeightedDistLabour (ln) |
|-------------------------------------------|-------------------------|
| Intercept                                 | 4.313*** (0.673)        |
| WeightedDistLabour_2000 (ln)              | 0.633*** (0.04)         |
| WeightedDistSavingsLabour_2000 (ln)       | 0.039*** (0.014)        |
| FirmDensity (ln)                          | 0.324*** (0.085)        |
| EstSize (ln)                              | 0.094 (0.074)           |
| HumanCap (ln)                             | -0.177** (0.07)         |
| UnempRate (ln)                            | -0.011 (0.065)          |
| TheilIndex (ln)                           | 0.367 (0.357)           |
| SpecIndex (ln)                            | -0.005 (0.048)          |
| ImmigrantsShare (ln)                      | -0.055 (0.041)          |
| ChangeHP                                  | 0.002* (0.001)          |
| IncomeGrowth (ln)                         | 0.191** (0.081)         |
| EntrepreneurialAttitudes (ln)             | -0.42* (0.215)          |
| Number of observations                    | 290                     |
| F-test                                    | 91.92                   |
| P-value                                   | 0.000                   |
| Adjusted R-squared                        | 0.791                   |

Robust standard errors in parentheses

*, **, and *** show significance levels at 10%, 5% and 1%, respectively
Table 6   MGWR Results for the three robustness tests

|                           | 1st robustness test: 2007 data | 2nd robustness test: use of WeightedDistFirms | 3rd robustness test: NFFPC\(\text{Eco}\) as dependent variable |
|---------------------------|--------------------------------|-----------------------------------------------|----------------------------------------------------------|
|                           | Min   | Median | Max       | Min         | Median   | Max       | Min         | Median   | Max       | Min         | Median   | Max       | Bandwidth |
| Intercept                 | -0.634 | -0.081 | 0.840     | -0.754      | -0.023   | 1.059     | -0.435      | 0.004    | 0.620     | -0.071      | 288      |
| WeightedDistLabour (ln)   | -0.047 | -0.038 | -0.010    | -0.080      | -0.072   | -0.071    | 288         |          |
| WeightedDistFirms (ln)    | -0.302 | 0.508  | 74        | -0.149      | -0.146   | -0.130    | 288         | -0.058   | -0.057    | -0.050      | 288      |
| EstSize (ln)              | 0.380  | 0.442  | 0.531     | 0.339       | 0.355    | 0.359     | 288         | 0.249    | 0.258     | 0.260       | 288      |
| FirmDensity (ln)          | 0.235  | 0.289  | 0.294     | 0.339       | 0.355    | 0.359     | 288         | 0.249    | 0.258     | 0.260       | 288      |
| HumanCap (ln)             | -0.183 | 0.068  | 0.265     | -0.221      | 0.010    | 0.087     | 142         | -0.178   | -0.010    | 0.044       | 122      |
| UnempRate (ln)            | -0.149 | -0.015 | 0.089     | -0.053      | 0.041    | 0.141     | 172         | -0.026   | 0.042     | 0.099       | 166      |
| TheilIndex (ln)           | -0.018 | 0.006  | 0.007     | -0.065      | -0.007   | 0.030     | 262         | -0.054   | -0.011    | 0.020       | 254      |
| SpecIndex (ln)            | 0.132  | 0.238  | 0.324     | 0.128       | 0.133    | 0.159     | 288         | 0.128    | 0.130     | 0.142       | 288      |
| ImmigrantsShare (ln)      | 0.001  | 0.068  | 0.133     | -0.118      | -0.018   | 0.146     | 150         | -0.066   | -0.015    | 0.066       | 160      |
| ChangeHP (ln)             | -0.008 | 0.046  | 0.063     | -0.085      | -0.079   | -0.017    | 284         | -0.113   | -0.048    | 0.077       | 112      |
| EntrepreneurAttitudes (ln)| -0.052 | 0.005  | 0.009     | -0.027      | -0.023   | 0.055     | 288         | -0.008   | -0.006    | 0.046       | 288      |
| Number of observations    | 290 municipalities | 290 municipalities | 290 municipalities |
| Adjusted R-squared        | 0.636  | 0.730  | 0.893     | 468.368     | 199.147  |
| AIC                       | 549.524 | 464.853 | 199.912 |
| AICc                      | 552.336 | 468.368 | 199.912 |
Fig. 6  *WeightedDistFirms* in 2013 in m ("Natural breaks" classification)
Fig. 7  Maps for GWR (a) and MGWR (b) parameter estimates surfaces for $WeightedDistLabour$ in first robustness test which uses data in 2007 (top), second robustness test which uses $WeightedDistFirms$ in 2013 (middle), third robustness test which uses $NFFPCEco$ in 2013 (bottom). Grey regions are not statistically different from zero.
Funding  The funding was provided by Bostad 2.0 and Handelsbanken Forskningsstiftelse.

Author Contribution  Cynthia Ho conducted the analysis, built the models and wrote the paper. Mats Wilhelmsson provided supervisions and feedback for the paper.

Funding  Open access funding provided by Royal Institute of Technology.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Declarations

Conflict of interest  C.S.T. Ho and M. Wilhelmsson declare that they have no competing interests.

Ethical standards  For this article no studies with human participants or animals were performed by any of the authors. All studies mentioned were in accordance with the ethical standards indicated in each case. Consent to participate: Not applicable. Consent for publication: Not applicable.

References

Agarwal S, Hauswald R (2010) Distance and private information in lending. Rev Financ Stud 23(7):2757–2788. https://doi.org/10.1093/rfs/hhq001
Agostino M et al (2021) Do cooperative banks matter for new business creation? Evidence on Italian manufacturing industry. Ann Public Coop Econ. https://doi.org/10.1111/apce.12342
Alessandrini P, Presbitero AF, Zazzaro A (2010) Bank size or distance: What hampers innovation adoption by SMEs? J Econ Geogr 10(6):845–881. https://doi.org/10.1093/jeg/lbp055
Almi (2021) A complement to your total financing solution. https://www.almi.se/vara-tjanster/tjanster/lan-och-ovrig-finansiering/foretagslan/. Accessed 27 Aug 2021
Arcuri G, Levratto N (2020) Early stage SME bankruptcy: does the local banking market matter? Small Bus Econ 54(2):421–436. https://doi.org/10.1007/s11187-018-0042-4
Aurdretsch D, Fritsch M (1994a) On the measurement of entry rates. Empirica 21(1):105–113. https://doi.org/10.1007/BF01383974
Aurdretsch D, Fritsch M (1994b) The geography of firm births in Germany. Reg Stud 28(4):359–365. https://doi.org/10.1080/00343409412331348326
Avdeitchikova S (2009) False expectations: reconsidering the role of informal venture capital in closing the regional equity gap. Entrepreneursh Reg Dev 21(2):99–130. https://doi.org/10.1080/08985620802025962
Backman M (2015) Banks and new firm formation. J Small Bus Enterp Dev 22(4):734–761. https://doi.org/10.1108/JSBED-03-2013-0035
Backman M, Wallin T (2018) Access to banks and external capital acquisition: perceived innovation obstacles. Ann Reg Sci 61(1):161–187. https://doi.org/10.1007/s00168-018-0863-8
Becker GS, Murphy KM (1992) The division of labor, coordination costs, and knowledge. Q J Econ 107(4):1137–1160. https://doi.org/10.2307/2118383
Berger A (2003) The economic effects of technological progress: evidence from the banking industry. J Money Credit Bank 35(2):141–176. https://doi.org/10.1353/mcb.2003.0009
Berger A, Udell G (1995) Relationship lending and lines of credit in small firm finance. J Bus 68(3):351–381. https://doi.org/10.1086/296668
Bilgel F (2020) Guns and homicides: a multiscale geographically weighted instrumental variables approach. Geogr Anal 52(4):588–616. https://doi.org/10.1111/gean.12227
Bolton P, Dewatripont M (1994) The firm as a communication network. Q J Econ 109(4):809–839. https://doi.org/10.2307/2118349
Bolton KP, Wolken JD, Holmes JA (2010) Distance still matters: the information revolution in small business lending and the persistent role of location, 1993–2003. Finance and Economics Discussion Series. Washington, D.C. (https://ssrn.com/abstract=1782053)
Brunsdon C, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr Anal 28(4):281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x
Brunsdon C, Fotheringham S, Charlton M (1998) Geographically weighted regression-modelling spatial non-stationarity. J Royal Stat Soc Ser D 47(3):431–443. http://geoinformatics.wp.st-andrews.ac.uk/files/2012/09/Geographically-Weighted-Regression-Modelling-Spatial-Nonstationarity.pdf.
Chasco C, García I, Vicéns J (2008) Modeling spatial variations in household disposable income with geographically weighted regression. Estadística Española 50(168):321–360
Cole RA, Goldberg LG, White LJ (2004) Cookie cutter vs. character: the micro structure of small business lending by large and small banks. J Financial Quant Analysis 39(2):227–251 (http://www.jstor.org/stable/30031854)
Cruz SCS, Teixeira AAC (2021) Spatial analysis of new firm formation in creative industries before and during the world economic crisis. Ann Reg Sci 67:385–413. https://doi.org/10.1007/s00168-021-01052-3
Degryse H, Ongena S (2004) The impact of technology and regulation on the geographical scope of banking. Oxif Rev Econ Policy 20(4):571–590 (http://www.jstor.org/stable/23609636)
Drexler A, Schoar A (2014) Do relationships matter? Evidence from loan officer turnover. Manage Sci 60(11):2722–2736 (http://www.jstor.org/stable/24550450)
Eliasson J (2016) Inter-industry differences in local banks’ effect on new firm formation. Jönköping University. http://hj.diva-portal.org/smash/get/diva2:930040/FULLTEXT02.pdf. Accessed 9 Mar 2022
Eurostat (2021) Statistical confidentiality and personal data protection. https://ec.europa.eu/eurostat/web/microdata/statistical-confidentiality-and-personal-data-protection. Accessed 24 Aug 2021
Fotheringham AS, Yang W, Kang W (2017) Multiscale geographically weighted regression (MGWR). Ann Am Assoc Geogr 107(6):1247–1265. https://doi.org/10.1080/24694452.2017.1352480
Garofoli G (1994) New firm formation and regional development: the Italian case. Reg Stud 28(4):381–393. https://doi.org/10.1080/00343409412331348346
Geanakoplos J, Milgrom P (1991) A theory of hierarchies based on limited managerial attention. J Jpn Int Econ 5(3):205–225. https://doi.org/10.1016/0889-1583(91)90011-E
Guiso L, Sapienza P, Zingales L (2004) Does local financial development matter? Q J Econ 119(3):929–969. https://doi.org/10.1162/00220510412331348346
Ho CST, Berggren B (2020) The effect of bank branch closures on new firm formation: the Swedish case. Ann Reg Sci 65(2):319–350. https://doi.org/10.1007/s00168-020-00986-4
Kärnä A, Manduchi A, Stephan A (2021) Distance still matters: Local bank closures and credit availability. Int Rev Finance 21:1503–1510. https://doi.org/10.1111/irfi.12329
Krugman P (1991a) History and industry location: the case of the manufacturing belt. Am Econ Rev 81(2):80–83 (https://www.jstor.org/stable/2006830%0A)
Krugman P (1991b) Increasing returns and economic geography. J Polit Econ 99(31):483–499. https://doi.org/10.1086/261763
Landström H (2017) Advanced introduction to entrepreneurial finance. In: Advanced introduction to entrepreneurial finance. Elgar.
Liberti JM (2011) Initiative, incentives and soft information: how does delegation impact the role of bank relationship managers? SSRN Journal. https://doi.org/10.2139/ssrn.537583
Mansour S et al (2021) Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR). Sustain Cities Soc 65:102627. https://doi.org/10.1016/j.scs.2020.102627
Nguyen H-LQ (2019) Are credit markets still local? Evidence from bank branch closings. Am Econ Journal: Appl Econ 11(1):1–32. https://doi.org/10.1257/app.20170543
O’Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41(5):673–690. https://doi.org/10.1007/s11135-006-9018-6
Okeahalam C (2009) Bank branch location: a count analysis. Spatial Econ Anal 4(3):275–300. https://doi.org/10.1080/17421770903114695
Oshan TM et al (2019) MGWR: a python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. Int J Geo Inform 8(6):1–31. https://doi.org/10.3390/ijgi8060269
Oshan TM, Smith JP, Fotheringham AS (2020) Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. Int J Health Geogr 19(1):1–17. https://doi.org/10.1186/s12942-020-00204-6

Petersen MA, Rajan RG (2002) Does distance still matter? The information revolution in small business lending. J Finance 57(6):2533–2570. https://doi.org/10.1111/1540-6261.00505

Rajan RG (1992) Insiders and outsiders: the choice between informed and arm’s-length debt. J Finance 47(4):1367–1400. https://doi.org/10.1111/j.1540-6261.1992.tb04662.x

Sharpe SA (1990) Asymmetric information, bank lending, and implicit contracts: a stylized model of customer relationships. J Finance 45(4):1069–1087. https://doi.org/10.2307/2328715

Silver L (2001) Credit risk assessment in different contexts—The influence of local networks for bank financing of SMEs. Uppsala University.

Slocum TA et al (2008) Thematic cartography and Geovisualization. Prentice Hall series in geographic information science

SVCA (2017) Swedish private equity market. Sweden. https://www.svca.se/wp-content/uploads/2017/07/The-Swedish-Private-Equity-Market-A-footprint-analysis-Copenhagen-Economics-SVCA-July-2017.pdf. Accessed 9 Mar 2022

Swedish Agency for Economic and Regional Growth (2017) Finansiering i svenska företag. Stockholm. https://tillvaxtverket.se/vara-tjanster/publikationer/publikationer-2018/2018-03-13-finansiering-i-svenska-foretag.html. Accessed 9 Mar 2022

Uchida H, Udell GF, Yamori N (2012) Loan officers and relationship lending to SMEs. J Financial Intermediation 21(1):97–122. https://doi.org/10.1016/j.jfi.2011.06.002

Wheeler D (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environ Plan A 39(10):2464–2481. https://doi.org/10.1068/a38325

Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J Geograph Syst 7(2):161–187. https://doi.org/10.1007/s10109-005-0155-6

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.