Accounting for unobserved individual heterogeneity in spatial stochastic frontier models: the case of Italian innovative start-ups

Accounting for Unobserved Individual Heterogeneity in Spatial Stochastic Frontier Models: the case of Italian Innovative Start-ups

Federica Galli
University of Bologna
Department of Statistical Sciences "Paolo Fortunati"
Via delle Belle Arti 41, Bologna, Italy
federica.galli14@unibo.it

Abstract: Spatial correlation and individual effects are two key sources of heterogeneity that should be handled in empirical applications investigating the productive performance of firms, especially when dealing with start-up activity. Therefore, in this work, we propose a fixed-effects spatial autoregressive stochastic frontier model for unbalanced panel data, we test the finite sample properties of our spatial estimator and we provide an empirical application on Italian innovative start-ups. The results of our analysis indicate that Italian start-ups are characterized by a significant level of global spatial dependence, especially overall and among firms belonging to the information and communication sector.

Keywords: Spatial stochastic frontier models, Fixed effects, Spatial dependence, Italian innovative start-ups.

JEL Codes: C2, C50, O31, R30.

Declaration of Interests: none.

1 Introduction

Stochastic frontier models (SF) are the parametric tool that is most commonly used by researchers for productivity and efficiency analysis in firm-level microdata applications, in the business-strategy framework, and in the regional sciences literature. The main feature of the baseline SF model concerns the possibility of distinguishing between two different kinds of disturbances influencing firms' performance: random errors resulting from favourable or unfavourable external events not controllable by firms and deviations in firms' productivity level resulting from factors under the

firm's control, such as technical and economic inefficiency. Recent advancements in stochastic frontier model literature mainly concentrated on extending the baseline SF model to a spatial setting (Glass, Kenjegalieva, and Sickles; Orea and Alvarez; Galli) or on incorporating fixed or random effects to account for unobserved individual characteristics (Greene; Wang and Ho). Indeed, spatial correlation and unobserved individual effects are two key sources of heterogeneity that should be handled in empirical applications dealing with spatial panel data. Nevertheless, in stochastic frontier model literature, limited attention has been devoted to controlling for both individual effects and spatial dependence shaping firms' production or cost functions.

Therefore, in this work, we develop a fixed effects spatial autoregressive stochastic frontier model for unbalanced panel data that allows considering both global spatial dependence related to the frontier function (i.e. productivity spillovers) and unobserved individual heterogeneity. In the empirical section, we take advantage of the spatial SF model with fixed effect introduced in this paper to analyse the productive performance of Italian start-ups. Indeed, to our knowledge, while several works examined start-up innovation processes and performance by using SF analysis (Matricano, Candelo, and Sorrentino; Chung, Jo, and Lee; Charoenrat and Harvie), there are no studies considering together start-ups' heterogeneous characteristics and spatial effects in an SF setting.

Our model specification considering both fixed effects and spatial dependence is particularly suitable for start-ups' productivity analysis for several reasons. First, newly established firms and small firms in general benefit to a greater extent from knowledge spillovers compared to larger firms (Audretsch and Link; Audretsch and Belitski; Audretsch, Belitski, and Caiazza). Indeed, the major part of them are managed by young entrepreneurs who tend to be very attentive to market opportunities and to have a high propensity to invest (Barboza and Capocchi), leading to an increased willingness to take risks and to absorb new knowledge from external sources in order to create new innovative products (Acs and Audretsch; Caiazza, Belitski, and Audretsch). Therefore, when dealing with firms' productive performance, considering spatial dependence occurring across neighbouring units is a fundamental aspect and this is particularly true for start-ups due to the importance of external inputs in the first stages of a business (Audretsch, Belitski, and Caiazza). Second, the population of start-ups is very heterogeneous (Colombelli), the radical nature of the innovations they advance is highly localized and context-specific (Antonietti and Gambarotto), and the entrepreneurs' characteristics such as managerial skills, psychological traits, and motivation for achievement play a key role in the performance and survival of these types of organizations (Fritsch and Wyrwich). Thus, unobserved individual heterogeneity may have a relevant role in shaping start-ups' performance, making it necessary to include firm-specific effects in empirical models to avoid biased estimates due to omitted variables. Finally, innovative start-ups tend to be engaged in new and cutting-edge projects especially related to the knowledge-intensive and high-tech sectors. However, a large number of innovative start-ups fail within the first three years of activity mainly due to the lack of appropriate institutional, technological and market conditions (Acs et al.). When

focusing on innovative start-ups, it is thus difficult to narrow the analysis to a balanced panel due to the high number of start-ups entering the market each year and the high probability of not surviving the first few years of business. This feature requires the use of statistical tools that are capable of handling unbalanced panel data like the one introduced in this study.

Among the different sources of spillovers, in this model we consider spatial effects influencing firms' frontier function since, in most empirical applications, incorporating the spatial lag of the dependent variable through a spatial autoregressive (SAR) term is to be preferred to spatial error structures because global spillovers in a SAR specification have a structural economic interpretation and allow to capture direct, indirect, and total impacts of the independent variables on Y differently from spatial error models. For example, in the context of regional sciences, while the SAR term has a strong economic justification given by the abundant evidence of significant feedback processes at the regional level, *"it is probably more difficult to give an economic interpretation to the spatial correlation of the inefficiency terms because the observations in a regional setting are not individual firms making similar decisions to other firms"* (Orea and Alvarez, p.557). In our particular case study, global productivity spillovers related to the frontier function are highly relevant in determining new ventures' productive performance due to the transmission of best practices between peers, collective and emulation behaviours, learning from others, and firms' adoption of new similar technologies (Skevas and Lansink). Moreover, as it is well known in spatial econometric literature (Elhorst), spatial error models can be regarded as specific parameter configurations of spatial models including the spatial lags of Y and X. The proposed model specification can be easily extended in order to consider local spatial dependence related to the X variables adding the exogenous spatial lag of the input variables (SLX). Indeed, the SLX terms can be considered as part of the set of the explanatory variables of the model and thus, the spatial Durbin SF model controlling for fixed effects is a straightforward extension of our spatial specification.

The proposed model can be estimated using ML techniques applying a within transformation approach as proposed by Wang and Ho. Then, direct, indirect, and total effects can be computed following LeSage and Pace. The model introduced in this work can be very useful in empirical application because it allows to (i) capture the overall level of spatial dependence; (ii) control for unobserved individual heterogeneity; (iii) model the systematic differences in inefficiency using some exogenous determinants. In addition, while current contributions to spatial SF models literature only apply to strongly balanced panels, which is a quite restrictive condition in practice, our spatial specification fits unbalanced panel data. Finally, rather than selecting a production or a cost frontier, we broaden our specification by taking both approaches into account.

The remainder of the paper is organized as follows: Section 2 discusses previous SF model literature, Section 3 introduces the model specification, the estimation procedure and the derivation of the marginal effects, Section 4 shows the results of some Monte Carlo simulations, Section 5 provides an empirical application on Italian innovative start-ups, and Section 6 concludes.

2 Literature Review

In the last decades, several advancements have been proposed by authors in order to expand the classic SF model introduced by Aigner, Lovell, and Schmidt. The first strand of the literature concentrated on distinguishing individual unobserved heterogeneity from inefficiency when panel data are available. Indeed, if individual effects are not taken into account, firm-specific factors affecting firms' productive outcomes are absorbed in the inefficiency component even if they are unrelated to inefficiency (Greene). In this framework, Kumbhakar is the first contribution separating technical inefficiency from firm-specific and time-specific effects specifying a random effects model. Afterwards, Greene proposed a fixed effect model including individual dummies for the fixed effects and estimated the large number of parameters in the model by advanced numerical maximization algorithms. Despite the complexity of the model as N increases, simulation studies performed by Greene demonstrated that the incidental parameters problem (i.e. possible inconsistency due to the number of parameters growing with the number of firms) does not affect the slope coefficients of SF models, but the variance parameters are more likely to be biased especially if the number of time periods is small. However, since estimated firms' efficiency scores rely on the error variance, the incidental parameters problem cannot be ignored. Wang and Ho solved this issue by model transformation. Specifically, the authors estimated a fixed-effects SF model through ML techniques using either a first difference or a within-transformation approach to remove time-invariant individual effects. Chen, Schmidt, and Wang extended Wang and Ho's approach by assuming that the inefficiency component varies randomly over time taking advantage of results in the statistical literature on the closed skew normal distribution. Belotti and Ilardi further developed Chen, Schmidt, and Wang's model, allowing inefficiency to be heteroskedastic and to follow a first-order autoregressive process.

The second strand of literature focused on introducing some spatial components in SF models in order to consider spatial cross-sectional dependence. Indeed, biased estimates and invalid statistical inference can occur in evaluating the level of productivity of firms using SF models if spatial correlation is not taken into account (Fusco and Vidoli; Glass, Kenjegalieva, and Sickles). Until now scholars have introduced spatial terms in SF models following two different paths: evaluating global and local spatial spillovers affecting the frontier function introducing the spatial lag of the dependent variable and/or of the input variables (Adetutu et al.; Glass, Kenjegalieva, and Sickles; Ramajo and Hewings; Gude, Alvarez, and Orea; Tsukamoto) or considering spatial dependence in the inefficiency and/or in the random error term (Schmidt et al.; Areal, Balcombe, and Tiffin; Fusco and Vidoli; Tsionas and Michaelides; Orea and Alvarez; Galli). While the first approach allows to directly measure direct and indirect impacts related to the frontier function, the second one is useful to consider unobserved but spatially correlated variables such as environmental and climatic conditions or to capture spatial correlation deriving from the specific attributes that commonly

characterize all the producers in the same area. The only contribution that merges all the different spatial structures described before obtaining a comprehensive specification that takes four different sources of spatial dependence into account is Galli.

Despite the growing number of contributions to SF model literature in recent years, a SF model handling both cross-sectional spatial dependence related to the frontier function and unobserved individual heterogeneity by introducing fixed effects has not yet been developed since it is not a straightforward extension. Indeed, while removing individual fixed effects by model transformation in linear models with normal errors leads to the same maximum likelihood estimator for the β parameters as the model in which the fixed effects are estimated as parameters, in SF models the MLE-based on the deviations from the means is not the same as the full MLE (Chen, Schmidt, and Wang). Therefore, existing spatial estimators for panel data models with fixed effects such as the one introduced by Lee and Yu do not apply to the SF literature. However, in stochastic frontier analysis, the omission of the spatial lag of the dependent variable and unmodelled individual-specific effects may result in biased estimates and invalid statistical inference. To our knowledge, existing papers in SF models literature considering both spatial dependence and individual heterogeneity are Glass, Kenjegalieva, and Sickles, Carvalho, Belotti, Ilardi, and Mortari, and Lai and Tran. Glass, Kenjegalieva, and Sickles introduce both the spatial autoregressive term and the fixed effects, however, the authors consider a deterministic frontier function without distinguishing between the inefficiency component and the random error. Carvalho develops a stochastic cost frontier model that incorporates spatial dependence in the inefficiency component and accounts for individual heterogeneity through random effects and estimates it using Bayesian techniques. Lai and Tran propose a random effect spatial stochastic frontier model that includes the spatial lag of the dependent variable following a simulated maximum likelihood approach for the estimation. Finally, Belotti, Ilardi, and Mortari model individual heterogeneity through fixed effects and spatial dependence by including a spatial autoregressive structure related to the inefficiency error term implementing a simulated likelihood estimator. In this paper, we extend previous literature on SF models by introducing a stochastic frontier model with fixed effects for unbalanced panel data that considers spatial dependence related to the frontier function using a spatial autoregressive specification. The model can be estimated by standard maximum likelihood techniques through model transformation as described in the following section.

3 The Model

3.1 Model Specification

We specify our spatial stochastic frontier model with individual fixed effects as in Eq.(1)-(6) for individual $i, j = 1, ..., N$ $(i \neq j)$ and $t = 1, ..., T_i$, where T_i is the individual i's number of time observations in the data. We keep the specification as general as possible, considering a cost

frontier for $c = -1$ and a production frontier for $c = 1$.

$$Y_{it} = \alpha_i + \rho \sum_{j=1}^{N} w_{ijt} Y_{jt} + \boldsymbol{X}_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{1}$$

$$\varepsilon_{it} = v_{it} - c u_{it} \tag{2}$$

$$v_{it} \sim \mathcal{N}(0, \sigma_v^2) \tag{3}$$

$$u_{it} = h_{it} u_i^* \tag{4}$$

$$h_{it} = f(\boldsymbol{z}_{it}\boldsymbol{\delta}) \tag{5}$$

$$u_i^* \sim \mathcal{N}^+(\mu, \sigma_u^2) \tag{6}$$

Specifically, Y_{it} is the logarithm of the productive output of firm i at time t in the case of a production frontier or the logarithm of the firm's total costs for a cost frontier, the α_is are the individual fixed effects capturing unobserved individual heterogeneity, and $\boldsymbol{X}_{it}$ is a $(1 \times K)$ vector of exogenous input variables or exogenous cost determinants respectively for a production and a cost frontier with associated parameter vector $\boldsymbol{\beta}$ $(K \times 1)$. We include in the model specification the spatial lag of the dependent variable to consider the overall level of spatial dependence between neighbouring producers. For each year t, the spatial structure of the data is defined by the row-standardized spatial weight matrix W_t having zero values on the main diagonal and non-negative spatial weights w_{ijt} to identify each pair of neighbouring spatial units $i, j = 1, ..., N$ $(i \neq j)$ in year t. Following Egger, Pfaffermayr, and Winner and Baltagi, Egger, and Pfaffermayr, since the panel is unbalanced, the spatial weight matrix changes every year and, starting from the "full" $(N \times N)$ version of W_t, every year, rows and columns related to missing observations are deleted. For the whole panel, the spatial weighting matrix W is a block-diagonal matrix of dimension $(NT \times NT)$ with blocks W_t, where $NT = \sum_{i=1}^{N} T_i$. The overall level of spatial dependence is captured by the spatial autoregressive parameter ρ. The spatial structure of this model can be further extended in order to consider also local spatial dependence associated with the input variables, obtaining a spatial Durbin specification. Including the spatial lag of the input variables (SLX term) in Eq.(1) coincides with enlarging the set of the exogenous explanatory variables of the model, and thus, all the steps below straightforwardly apply to a spatial Durbin stochastic frontier model with individual fixed

effects.

Considering the error component, we model ε_{it} as the sum of two independent components v_{it} and u_{it}, following the classic structure for the error term proposed by Aigner, Lovell, and Schmidt. In particular, the random error term v_{it} results from favourable or unfavourable external events such as luck, climate, topography, and machine performance and it is assumed to be distributed as a normal random variable with zero mean and variance σ_v^2 while the inefficiency component u_{it} captures the deviation in firm's productivity or cost level resulting from factors under the firm's control, such as technical, economic or cost inefficiency. If a production frontier is specified, the positive error component u_{it} has to be subtracted from the SF production function reflecting the fact that each firm's output must lie on or below the frontier due to inefficiency, while, in the case of a cost function, u_{it} represents the cost increase due to inefficiency and thus, it has to be summed and not subtracted to the SF cost function. Following Wang and Ho, in order to obtain a tractable form for the likelihood function, we specify the inefficiency component u_{it} as the product of a scaling function h_{it} and a basic distribution u_i^*, as proposed by Wang and Schmidt (further details on the scaling property and on the related advantages can be found therein). While h_{it} depends on a positive function f of M exogenous variables $\boldsymbol{z}_{it}$ with associated parameter vector δ $(M \times 1)$, u_i^* is a time-invariant firm-specific inefficiency error term distributed as a truncated normal distribution with mean μ and variance σ_u^2.

For the model specified in Eq.(1)-(6), the following assumptions must hold: (i) the spatial autoregressive parameter ρ must be within the range $\left(\frac{1}{\omega_{min}}, 1\right)$, where ω_{min} is the smallest eigenvalue of W and the upper bound equals 1 for row-normalized spatial weight matrices; (ii) the cross-sectional correlation must approach zero as the distance separating two spatial units increases (Kelejian and Prucha). The first condition ensures that $(I_{NT} - \rho W)$ is a non-singular matrix and the second one guarantees that the row and columns sums of W and $(I_{NT} - \rho W)^{-1}$, before row-normalization, are uniformly bounded in absolute value when N goes to infinity. In particular, assumption (i) is always satisfied if W is a symmetric matrix while assumption (ii) is always satisfied as long as W is a binary contiguity matrix while for inverse distance spatial weight matrices it should be introduced a cut-off point in the distance measure (Elhorst). For the identifiability of the model, neither the $\boldsymbol{x}_{it}$ nor the $\boldsymbol{z}_{it}$ must contain constant intercepts. Moreover, the scaling function h_{it} is required to be not constant. This condition is satisfied when z_{it} contains at least one variable which changes over time. Finally, as common in the stochastic frontier model literature, imposing v_{it} and u_{it} independently and identically distributed and $E(v_{it}u_{it}) = 0$ ensures that the two error components of ε_{it} can be identified. It is important to remark that, as discussed by Belotti and Ilardi, despite the use of a fixed-effects removing transformation, the distributional assumptions allow the correct identification of the parameters.

To estimate the model in Eq.(1)-(6) removing individual fixed-effects, a within transformation can be used. In the section below we adapt the within-transformed SF model developed by Wang and

Ho to a spatial setting, deriving the corresponding loglikelihood function and taking into account the endogeneity arising from the inclusion of the spatial lag of Y_{it}. Moreover, we extend Wang and Ho's approach for balanced panels to an unbalanced panel setting. When dealing with unbalanced spatial units, the asymptotic properties of the spatial estimators may become problematic if the reason why data are missing is unknown (Elhorst). Thus, further research is needed to deal with possible endogeneity related to stayers and quitters.

3.2 Estimation Procedure

In order to remove the time-invariant individual fixed effects, a within transformation can be used by subtracting the sample mean of each panel from the related individual observations. In particular, we use the following notation: $q_{i.} = (1/T_i)\sum_{t=1}^{T_i} q_{it}$ and $q_{it.} = q_{it} - q_{i.}$. Thus, for a given i, $\widetilde{q}_{i.} = (q_{i1.}, q_{i2.}, ..., q_{iT_i.})'$. Since the panel is unbalanced, the length of each generic vector $\widetilde{q}_{i.}$ changes for each ith observation depending on the number of available time periods. The minimum required number of time periods for each individual unit is two. The within transformed model is shown in Eq.(7)-(10) for $i = 1, ..., N$, where $\widetilde{WY}_{i.}$ is a vector containing the within transformed weighted values of Y in the neighbouring units of the ith individual. In particular, $WY_{it.} = \sum_{j=1}^{N} w_{ij}Y_{jt} - WY_{i.}$, where $WY_{i.} = (1/T_i)\sum_{t=1}^{T_i}\sum_{j=1}^{N} w_{ij}Y_{jt}$ for $j = 1, ..., N$, and the stacked vector of $WY_{it.}$ related to the neighbouring observations of the ith spatial unit is defined as $\widetilde{WY}_{i.} = (WY_{i1.}, WY_{i2.}, ..., WY_{iT_i.})'$. However, $\widetilde{WY}_{i.}$ corresponds to $W\widetilde{Y}_{i.}$, i.e. multiplying the vector of within transposed Y_{it} for the spatial weight matrix W.

$$\widetilde{Y}_{i.} = \rho\widetilde{WY}_{i.} + \widetilde{\boldsymbol{X}}_{i.}\boldsymbol{\beta} + \widetilde{v}_{i.} - c\widetilde{u}_{i.} \tag{7}$$

$$\widetilde{v}_{i.} \sim MVN(0, \Pi_{T_i}) \tag{8}$$

$$\widetilde{u}_{i.} = \widetilde{h}_{i.}u_i^* \tag{9}$$

$$u_i^* \sim \mathcal{N}^+(\mu, \sigma_u^2) \tag{10}$$

As demonstrated by Wang and Ho, within transforming the model introduces correlation in the variance-covariance matrix Π_{T_i} of the $\widetilde{v}_{i.}$s within the ith panel. In particular, a $(T_i \times T_i)$ matrix with values $\sigma_v^2(1 - 1/T_i)$ on the main diagonal and $\sigma_v^2(-1/T_i)$ elsewhere is obtained. Since we adopt an unbalanced panel data setting, the variance-covariance matrix Π_{T_i} changes dimension for each panel depending on the number of ith spatial units available in the whole period.
The loglikelihood function associated with the within transformed model is shown in Eq.(11) for a

generic panel i, where Φ represents the standard normal cumulative distribution function and μ_*, σ_*^2 and $\widetilde{\varepsilon}_{i.}$ are defined as in Eq.(12)-(14). The notation $\Pi_{T_i}^-$ indicates the generalized inverse of Π_{T_i} because Π_{T_i} is a non-invertible matrix (further details on this solution can be found in Wang and Ho).

$$\begin{aligned}\log L_i = &-\frac{1}{2}(T_i-1)\log(2\pi) - \frac{1}{2}(T_i-1)\log(\sigma_v^2) - \frac{1}{2}\widetilde{\varepsilon}_{i.}'\Pi_{T_i}^-\widetilde{\varepsilon}_{i.} + \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) \\ &+ \log\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \log\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right)\end{aligned} \tag{11}$$

$$\mu_* = \frac{\dfrac{\mu}{\sigma_u^2} - c\widetilde{\varepsilon}_{i.}'\Pi_{T_i}^-\widetilde{h}_{i.}}{\widetilde{h}_{i.}'\Pi_{T_i}^-\widetilde{h}_{i.} + \dfrac{1}{\sigma_u^2}} \tag{12}$$

$$\sigma_*^2 = \frac{1}{\widetilde{h}_{i.}'\Pi_{T_i}^-\widetilde{h}_{i.} + \dfrac{1}{\sigma_u^2}} \tag{13}$$

$$\widetilde{\varepsilon}_{i.} = \widetilde{Y}_{i.} - \rho\widetilde{W}\widetilde{Y}_{i.} - \widetilde{\boldsymbol{X}}_{i.}\boldsymbol{\beta} \tag{14}$$

The final loglikelihood function equals $\log L = \log|I_{NT} - \rho W| + \sum_{i=1}^N \log L_i$, where $NT = \sum_{i=1}^N T_i$ and W is the block-diagonal spatial weight matrix associated to the whole panel. In order to consider the endogeneity coming from the introduction of the spatial lag of Y, the determinant of the Jacobian of the transformation from $\widetilde{\varepsilon}$ to $\widetilde{Y}$ is added in the final loglikelihood function since $\frac{\partial\widetilde{\varepsilon}}{\partial\widetilde{Y}} = (I_{NT} - \rho W)$ given $\widetilde{\varepsilon} = (I_{NT} - \rho W)\widetilde{Y} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}$. In an unbalanced panel data setting, the only issue in the computation of the log determinant consists of considering a block diagonal $(NT \times NT)$ spatial weight matrix W with blocks W_t that vary across the different years depending on the number of spatial units available. Thus, rows and columns corresponding to the missing spatial units should be deleted from the full spatial weight matrix corresponding to the whole sample.

The unknown parameters are simultaneously estimated by numerically maximising the loglikelihood function using standard software [1]. To satisfy the first assumption described before, i.e. $(I_{NT} - \rho W)$ non-singular, the spatial autoregressive parameter ρ is bounded to the interval $\left(\frac{1}{\omega_{min}}, 1\right)$, where ω_{min} is the smallest eigenvalue of W and the upper bound equals 1 for row-standardized spatial weight matrices (LeSage and Pace). Finally, some attention should be paid to the computation of the term $|I_{NT} - \rho W|$, which is the determinant of a large matrix and it needs to be recalculated at each iteration of the optimization procedure, leading to a computationally expensive and time-consuming

routine. One potential solution to speed up the computation of the log-determinant, as suggested by Pace and Barry, is to use a vector of values for the log-determinant corresponding to different values of ρ in the interval $\left(\frac{1}{\omega_{min}}, 1\right)$, calculated before optimization. This involves calculating the vector of log-determinants only once during the estimation and an interpolated value for the log-determinant can be used considering a sufficiently fine grid for ρ (for a review on these numerical approaches see LeSage and Pace).

3.3 Marginal Effects

The model in Eq.(1)-(6) can be written in reduced form as in Eq.(15) using matrix notation. Specifically, Y is an $(NT \times 1)$ vector including the N individual observations stacked over time, X and Z are $(NT \times K)$ and $(NT \times M)$ matrices containing respectively K and M exogenous variables, I_{NT} is an $(NT \times NT)$ identity matrix, W is the $(NT \times NT)$ block-diagonal spatial weight matrix, $D = I_N \bigotimes i_{T_i}$ is an $(NT \times N)$ matrix, v is the $(NT \times 1)$ random error term, and α and u^* are $(N \times 1)$ vectors containing respectively the N time-invariant individual fixed effects and inefficiency error terms, respectively. Moreover, the positive function f is specified as an exponential function, ρ is a scalar parameter capturing the overall level of spatial dependence, β and δ are $(K \times 1)$ and $(M \times 1)$ vectors of parameters and c is a scalar that equals 1 for a production function and -1 for a cost function.

$$Y = (I_{NT} - \rho W)^{-1}(D\alpha + X\beta + v - c\exp(Z\delta)Du^*) \tag{15}$$

In spatial models including the spatial autoregressive term, the β parameters cannot be meaningfully interpreted because, as shown in Eq.(16), they do not coincide with the first partial derivatives of Y with respect to the generic regressor X_r. The marginal effects of X_r on Y are contained in the $S_r(W)$ matrix $(NT \times NT)$.

$$\frac{dY}{dX_r} = (I_{NT} - \rho W)^{-1}(I_{NT}\beta_r) = S_r(W) \tag{16}$$

Starting from $S_r(W)$, LeSage and Pace proposed to calculate synthetic measures for the marginal effects differentiating into direct, indirect and total effects. In particular, the direct effects are calculated as the average of the elements in the main diagonal of $S_r(W)$, the indirect effects are calculated as the average of the sum of the off-diagonal elements of $S_r(W)$ and the total effects equal the sum of the direct and the indirect effects. The associated standard errors or t-values can be computed using the delta method or by simulation as proposed by LeSage and Pace.
Considering the marginal effects of the determinants of firms' inefficiency on the inefficiency level

u_{it}, the first derivative of $u = \exp(Z\delta)Du^*$ with respect to the generic determinant Z_r is shown in Eq.(17). In particular, applying the logarithm to u as suggested by Wang and Schmidt generates a straightforward expression for the impact of the generic determinant Z_r on $\log(u)$. Moreover, it should be noticed that we obtain the identical interpretation of δ_r even considering the unconditional definition of u which, starting from the reduced form in Eq.(15), includes the frontier feedback effects through the spatial filter $(I_{NT} - \rho W)^{-1}$.

$$\frac{\partial \log(u)}{\partial Z_r} = \frac{\partial \log(\exp(Z_r\delta_r)Du^*)}{\partial Z_r} = \frac{\partial(\log(\exp(Z_r\delta_r)) + \log(Du^*))}{\partial Z_r} = \delta_r \tag{17}$$

Therefore, the interpretation of the δ parameters is straightforward, as they represent the marginal effects of the Z variables on $\log(u)$ even if the spatial lag of Y is included in the model.

4 Monte Carlo Simulations

In order to test the finite sample properties of our novel spatial estimator, in this section, we conduct some Monte Carlo experiments. We start by examining the performance of our within transformed spatial stochastic frontier estimator for different numbers of spatial units, time periods, and ratios of missing data. In particular, we concentrate on the impact of setting different percentages of spatial units as missing at random after having generated the data on the bias, the mean squared error and the standard error of the estimates. Afterwards, we compare our spatial within transformed model (SSF with FE) with the non-spatial within transformed stochastic frontier specification (SF with FE), the spatial autoregressive stochastic frontier model without individual fixed effects (SSF), and the non-spatial stochastic frontier model without fixed-effects (SF).

To perform our simulation studies, we select a production function approach ($c = 1$) and we generate N individual observations over T time periods as follows. The fixed effects α_i are defined as random draws from a $[0, 1]$ uniform distribution, the input variable X_{it} is generated as a random draw from a $N(\alpha_i, 1)$, and the exogenous determinant of firms' efficiency Z_{it} from a $N(0, 1)$. Starting from the longitudes and latitudes of a sample of N Italian innovative start-ups located in the northwest of Italy, we define the spatial structure of the data using an $(NT \times NT)$ block-diagonal contiguity matrix. Moreover, u_i^* is generated as a random value from a truncated normal distribution with mean μ and variance σ_u^2 and the v_{it}s as random values from a normal distribution with zero mean and variance σ_v^2. Finally, our dependent variable Y_{it} is defined as shown in Eq.(18).

$$Y_{it} = (I_{NT} - \rho W)^{-1}(\alpha_i + X_{it}\beta + v_{it} - c\exp(Z_{it}\delta)u_i^*) \tag{18}$$

After having generated the data, for each year t, we set a percentage of spatial units as missing at

random. Thus, the rows and columns referring to missing data are also deleted from the spatial weight matrix W_t referring to year t. Afterwards, for each iteration, the model is estimated using only the available information and therefore, the sample mean computed to apply the within transformation is based only on non-missing spatial units.

We evaluate the performance of our spatial within transposed estimator for different numbers of time periods ($T = 5, 10$) and individual observations ($N = 100, 200, 300$) considering different percentages of missing data: $0\%, 10\%, 20\%, 30\%, 40\%$. The baseline parameters are fixed at $\{\beta = 0.50,\ \rho = 0.40,\ \delta = 0.50,\ \mu = 0.50\ \sigma_u^2 = 0.10\ \sigma_v^2 = 0.20\}$. For each experiment, the total number of replications is equal to 1000 and we use the same α_i, X_{it}, Z_{it}, thus only u_i^*, v_{it} and the missing spatial units are redrawn each time. For each simulation study, we report the average estimated coefficients (*Mean*), the mean squared errors (*MSE*) and the standard errors in parenthesis. The simulation results are shown in Table 1 for $T = 5$ and in Table 2 for $T = 10$. Moreover, in Table 3 we compare our spatial within transposed model with other spatial and non-spatial stochastic frontier models described above for $N = 100, 200, 300$, considering $T = 10$.

Insert Table 1. Simulation results for different sample sizes with T=5

Insert Table 2. Simulation results for different sample sizes with T=10

Table 1 and Table 2 show how the mean estimates, the mean squared errors and the standard errors (in brackets) vary depending on different sample sizes (N=100,200,300 and T=5,10) and percentages of missing values (pm). We find that the bias is very near to zero also considering small samples and high percentages of missing values. Moreover, it rapidly decreases for larger sample sizes and having 10 time periods instead of 5 highly improves the unbiasedness of the estimates. Considering the standard error, μ is the parameter with the larger variance relative to the value of the parameter itself. This result is in line with Wang and Schmidt who find that μ is the harder parameter to estimate even though it does not cause any problem for the parameters of main interest like β and δ. In addition, μ is also the parameter that tends to be mostly affected by the presence of missing values since it is slightly overestimated at high rates of missing values. However, considering an unbalanced panel instead of a balanced one mainly impacts the standard error and the mean squared error of the estimates but they tend to remain unbiased. Overall, the results in Table 1 and Table 2 indicate that the statistical properties of the estimated parameters tend to be satisfactory even for unbalanced panels.

Additional simulation results are contained in Appendix A. In particular, we investigate the finite sample properties of our proposed estimator for different parameter vectors. The estimation results contained in Tables A1 and A2 are in line with the results of Wang and Schmidt (2002) and Wang and Ho (2010). Indeed, we find that (i) μ is a difficult parameter to estimate and its estimation becomes even more difficult when μ is positive and large, (ii) larger σ_u^2 values make

μ and σ_u^2 to be estimated less precisely. For more details on the technical reasons behind these aspects see Wang and Schmidt (2002) and Wang and Ho (2010). A further interesting aspect is that the finite sample properties of the proposed estimator are satisfactory also in the case of no spatial dependence.

Insert Table 3. Performance of spatial and non-spatial SF models with and without fixed effects

Table 3 shows the simulation results on the ability of different SF specifications to estimate the unknown parameters consistently and efficiently when both spatial dependence and unobserved individual heterogeneity are present. In particular, we compare the fixed effects spatial autoregressive SF model proposed in this paper (SSF with FE) considering a balanced sample, the non-spatial fixed effects SF model (SF with FE), the spatial autoregressive SF model without individual fixed effects (SSF), and the baseline SF model that takes into account neither spatial dependence nor individual unobserved heterogeneity (SF). As demonstrated in the previous experiments, simulation results concerning the spatial FE model confirm the satisfactory performance of our spatial SF specification removing individual fixed effects by model transformation, especially as the number of individual observations increases. Passing to the non-spatial FE model, we find that, if spatial dependence is not adequately handled, the biggest issue concerns the estimation of the variance parameter σ_v^2. Indeed, besides detecting small biases in the estimates of β, μ, and σ_u^2, the results in Table 3 indicate that σ_v^2 is largely overestimated. Thus, the random error term absorbs unmodelled spatial dependence, leading to biased estimates for the error variance and consequently, also for firms' efficiency scores. On the other hand, considering spatial dependence in SF models but ignoring unobserved individual heterogeneity leads to a higher bias associated with the β, ρ, δ and μ estimates compared to the two variance parameters. Indeed, results related to the spatial SF specification indicate that while σ_u^2 and σ_v^2 are only slightly overestimated, unmodelled individual heterogeneity generates upward estimates of β and δ and downward estimates of μ and ρ. In particular, we find that the overall level of spatial dependence captured by ρ through the introduction of the SAR term is downwardly biased if fixed effects are not included in the model specification. Finally, not considering either spatial dependence or time-invariant individual heterogeneity leads to (i) overestimation of the variance parameter σ_v^2 due to unmodelled spatial dependence absorbed by the random error term; (ii) overestimation of the variance parameter σ_u^2 depending on the fact that individual-specific effects are not distinguished from the inefficiency component; (iii) upward bias of β and μ and downward bias of δ. In sum, while the baseline SF model suffers from many issues if individual units are affected by a significant level of spatial dependence and if firm-specific effects have a relevant role in shaping firms' performance, introducing the spatial autoregressive term helps in obtaining almost unbiased estimates at least for the two variance parameters. On

the other hand, considering individual fixed effects leads to fairly good estimates for all parameters with the exception of σ_v^2.

A final insight from Table 3 concerns how the finite sample properties of the different SF models considered in this simulation study are affected by the number of individual observations available yearly. Our findings indicate that the mean square errors and the standard errors of the estimated parameters tend to decrease as N increases and thus, efficiency rises with N. However, we find that the bias associated with β and δ tends to raise as N increases while the estimation results related to the two variance parameters are quite constant across the three different simulation studies considering 100, 200, and 300 individual observations. Therefore, omitted variable bias and distortions resulting from uncontrolled spatial dependence may worsen in applied works based on very large samples. This insight highlights even more the need to control for time-invariant firm-specific effects and spatial correlations in empirical applications.

Even if we do not formally derive the asymptotic properties, it is worth noting that our Monte Carlo simulations suggest consistency of our proposed estimator. Indeed, as highlighted by Wang and Ho, the within-transformation used to remove the individual effects makes the estimator immune to the incidental parameters problem and thus, consistency of the estimator is obtained by $N \to \infty$. Furthermore, as demonstrated by Jin and Lee for spatial autoregressive SF models, when inefficiency exists, all parameter estimators are $\sqrt{n}$ consistent and asymptotically normal. However, in the case of no inefficiency, some parameter estimators have slower convergence rates than $\sqrt{n}$ (for more details see Jin and Lee) but this case is seldom met in practice.

5 Application to Italian Innovative Start-ups

5.1 Data and Empirical Model

In this section, we illustrate the use of the spatial SF model with fixed effects for unbalanced panel data introduced in the previous sections by analysing Italian innovative start-ups' productive performance. Indeed, when analyzing the productive performance of start-ups, both spatial effects and unobserved individual heterogeneity play a relevant role due to existing spatial feedbacks occurring among firms (Audretsch and Link) and the heterogeneous characteristics of start-ups such as the typology of the product proposed, the specific context in which they operate, the managerial skills of the entrepreneur, etc. (Colombelli). Moreover, it's hard to dispose of balanced panels since start-ups tend to continuously enter and exit the market.

Insert Figure 1. Innovative start-ups map

Figure 1 Alt Text: Geographic location of the 2237 start-ups considered in the analysis with information on the number of temporal observations available for each one.

The data used in this application are collected from the AIDA Bureau Van Dijk database, which provides information on the consolidated accounts of Italian companies as well as their geographical location. After having adequately cleaned the data, our finite sample provides information on 2237 innovative start-ups located in Italy in the time period 2016-2020 [2]. In particular, 291 individual observations are available for the year 2016, 800 for 2017, 1437 for 2018, 2197 for 2019, and 2092 for 2020. Figure 1 shows the geographical location of the 2237 spatial units of our sample indicating the number of temporal observations available for each of them. In order to perform the within transformation, the minimum required number of time periods is two and thus, observations with only one time occurrence have been deleted. Overall, only 237 units are observed for all the years of the analysis and the average number of available time periods is 3.05.

We adopt a Translog production function for our empirical specification due to its higher flexibility compared to a Cobb-Douglas functional form. Specifically, for $i, j = 1, ..., N$ $(i \neq j)$ and $t = 1, ..., T_i$, we define

$$Y_{it} = \alpha_i + \rho \sum_{j=1}^{N} w_{ij} Y_{jt} + \beta_L L_{it} + \beta_K K_{it} + \beta_{LL} L_{it}^2 + \beta_{KK} K_{it}^2 + \beta_{LK} L_{it} K_{it} + v_{it} - h_{it} u_i^*, \tag{19}$$

$$h_{it} = \exp(\delta_{Size} Size_{it} + \delta_{Lev} Lev_{it} + \delta_{Liquid} Liquid_{it} + \delta_{Int} Int_{it} + \delta_{Pat} Pat_{it} + \delta_{R\&D} R\&D_{it}), \tag{20}$$

where Y_{it} is the logarithm of the value added of firm i at time t, α_i is the individuals' i fixed effect, and L_{it} and K_{it} are the logarithms of total wages and gross fixed assets, representing labour and capital inputs respectively. In order to control for spatial dependence, we include the endogenous spatial lag of the dependent variable, obtaining a measure of global productivity spillovers through ρ. The spatial structure of the data is defined by an inverse distance spatial weight matrix W truncated at 50km, with elements equal to zero on the main diagonal and generic off-diagonal elements w_{ij}. To identify neighbouring observations we consider a 50km radius around each spatial unit because spillover effects mainly originate from closest neighbours due to emulation, face-to-face interactions, local cooperation, and individual contact (Griliches). However, in Appendix C we provide several robustness checks considering different definitions for the spatial weight matrix. Since the panel is unbalanced, W is defined as a $(NT \times NT)$ block-diagonal matrix with blocks W_t, where W_t is the spatial weight matrix of the available units in year t, obtained by deleting the rows and columns referring to the missing observations in that year. Concentrating on the error structure, the random variable v_{it} is assumed to follow a zero-mean normal distribution while the inefficiency component u_{it} is defined as the product of the scaling function $h_{it} = \exp(\boldsymbol{Z_{it}\delta})$ and u_i^* following a half-normal distribution. We model h_{it} as a function of several variables that reflect the economic structure and innovative activity of Italian start-ups. In particular, we consider firms'

size as the logarithm of the number of employees (*Size*), and firms' financial flexibility and leverage proxied respectively by the logarithm of the ratio of liquid assets to total assets (*Liquid*) and the logarithm of long debts over net equity (*Lev*). Moreover, we measure firms' innovative activity through *Int*, *Pat*, and *R&D*, where *Int* indicates the ratio of intangible capital to total capital measuring firms' propensity to invest in intellectual capital, IT products, technological know-how, etc., *Pat* is the logarithm of patent rights, and *R&D* refers to the logarithm of R&D expenditure. Table B1 in Appendix B describes the meaning, definition and unit of measure of the variables included in the model. In addition, Figure B1 shows that firm capital and the share of intangible investments are characterized by a very weak association.

5.2 Estimation Results

Table 4 shows the estimation results of our spatial autoregressive SF model controlling for fixed effect by within transformation. The model in Eq.(19)-(20) is estimated considering both the overall sample (N=2237) and some specific industrial sectors. In particular, we concentrate on the manufacturing sector (Manuf), the information and communication sector (Inf&Comm), and the scientific and technological sector (Sci&Tech), which are the most numerous sectors in our sample. Firms in the manufacturing sector (N=309) are primarily involved in the manufacture of metal products, computers, electronic and optical products, and electrical equipment. Start-ups belonging to the Inf&Comm group (N=1110) mainly work on programming, IT consultancy and information service activities, while those in the Sci&Tech sector (N=580) perform scientific research and development and other professional, scientific and technical activities. The remaining start-ups in our sample (N=238) mainly carry out travel agency activities and booking services, retail and wholesale, electricity, gas, steam and air conditioning supply, and training. In this case study, we choose to estimate the model by subsamples in order to obtain some specific insights on the overall level of spatial dependence among start-ups belonging to specific industrial sectors as well as on sectoral-specific input spillovers. However, in the robustness check subsection, we check the robustness of the effects detected for the inefficiency determinants by estimating the overall model with some interaction terms.

Insert Table 4. Empirical model: Estimation results

The estimation results in Table 4 indicate that the overall level of global spatial dependence is positive and significant both in the comprehensive model (0.224) and for all sub-samples. In particular, among the three sectors considered in this empirical application, the manufacturing sector is less affected by global productivity spillovers (0.098) while the information and communication and the scientific and technological industries highly benefit from positive spatial feedback (0.210 and 0.189, respectively). However, productivity spillovers are found to be more effective between

firms operating in different industries. Indeed, the variety of industries in a geographical cluster fosters knowledge spillovers and firms' innovative activity (Jacobs) and thus, diversification rather than specialization contributes to generating greater positive agglomeration externalities.
Concerning the determinants of firms' inefficiency, we find a positive effect associated with *Size* and *Lev* and mainly a negative effect (on inefficiency) related to firms' innovative activity in all the estimated models. On the other hand, the impact of *Liquid* changes depending on the industrial sector considered. In particular, our results indicate that larger start-ups tend to be more inefficient compared to smaller ones, especially those belonging to the manufacturing sector. This finding is in line with the results of Matricano, Candelo, and Sorrentino, finding that a greater number of employees negatively affects Italian start-ups' performance. Indeed, greater firm size can slow managers' competitive moves and agreements over a firm's strategy (Hambrick, Cho, and Chen; Iaquinto and Fredrickson) and it requires greater monitoring and coordination costs (Liang, Rajan, and Ray). Thus, especially in the early stages of a business, a larger team size can make communication, coordination, decision-making, and the allocation of tasks more difficult, resulting in decreased efficiency levels. Similarly, financial leverage increases the inefficiency level of start-ups, due to greater financial limitations linked to indebtedness. Considering firms' innovative activity, we find that higher R&D expenditure and patent investments lead to increased efficiency levels both overall and for the specific industrial sectors. Indeed, innovative activity performed through R&D allows the development of new products and services and contributes to improving business processes and operations while patenting allows start-ups to protect the newly developed products as trade secrets, giving the innovative firms a competitive advantage (Helmers and Rogers). Finally, while the Inf&Comm and the Sci&Tech sectors as well as overall, Italian start-ups benefit from a greater share of investments in intangible capital over total capital, *Int* has an opposite effect on the manufacturing sector. Indeed, intangible capital is commonly believed to play a positive role in firms' performance due to its link with inventions of new processes and products, enhancements of employees' skills, business reputation and brand recognition. However, in the manufacturing sector, fixed assets, machinery and equipment may have a major role in the production process compared to intangibles due to the peculiarities of this industry.

5.3 Marginal Effects and TE Scores

Since the endogenous spatial lag of the dependent variable is included in the model specification, the marginal effects of the input variables do not correspond to the β estimates and thus, direct, indirect and total effects of labour and capital are shown separately in Table 5.

Insert Table 5. Empirical model: Marginal effects

The results in Table 5 indicate that labour has a greater direct impact on start-ups' profitability

compared to capital both overall and for the different industrial sectors. In particular, while labour holds a leading role in the manufacturing industry (0.689), the Sci&Tech sector is the one that benefits most from fixed assets (0.362). Taking only direct effects into account, we find that all the sectors considered are characterized by decreasing returns to scale. Concentrating on the indirect effects, we detect positive and significant input spillovers in all the industries (standard errors are computed using the delta method). In line with the magnitudes of ρ in the different estimated models, our results indicate that labour and capital spillovers are more effective at the global level and in the Inf&Tech sector. Moreover, the indirect effect of labour always exceeds the one of capital indicating that being located in an area with a great endowment of trained and specialized workers generates positive returns to neighbours likely due to labour pooling (i.e. a greater availability of skilled staff) (Marshall).

Analysing the technical efficiency scores, we find that the average TE score of Italian innovative start-ups in the time period considered is 0.679. However, manufacturing firms are those obtaining the worst result (0.273) while start-ups belonging to the Inf&Comm and Sci&Tech sectors reach higher average efficiency scores (0.494 and 0.542, respectively). Figure D1 in Appendix D shows the geographical distribution of the TE scores for the different industrial sectors. In particular, denser efficiency clusters are primarily located in the areas around Milan, Rome, and Naples, especially considering the Inf&Comm sector.

5.4 Robustness Checks

As a first robustness check, we estimate the model in Eq.(19)-(20) using different specifications for the spatial weight matrix W. First, we use different truncation criteria by specifying a dense inverse distance matrix with no truncation point or considering the 200 or 400 nearest neighbours. Second, we use spectral normalization (i.e. each element of W is divided by the modulus of the largest eigenvalue of the matrix) as an alternative standardization method to row normalization. Third, we consider different channels rather than geographical proximity through which spillovers may occur. Inspired by Santi et al., we define the spatial weight matrix as a function of the distance between units in terms of characteristics (i.e., values of regressors). In particular, we consider technological distance as a relevant factor in shaping spillovers among innovative start-ups and we build the spatial weight matrix as a function of firms' innovative capacity proxied by the ratio of intangible investments over total investments. The estimation results are shown in Table C1 of Appendix C. Higher levels of global spatial dependence are detected by modelling spillovers based on technological distance instead of geographical distance (0.420) as well as considering a larger number of neighbours through a dense W matrix (0.270) or a 400 nearest neighbours matrix (0.256). However, the other parameters do not vary considerably for changes in the spatial weight matrix confirming the robustness of our estimates to different specifications of W.

As a second robustness check, we consider different model specifications. In particular, we estimate

the overall model with a dense W using a non-spatial fixed-effect specification, a spatial SF model without fixed effect, and a baseline non-spatial SF specification. The results in Table C2 of Appendix C indicate that when using non-spatial specifications, the β coefficients related to labour and capital are overestimated compared to spatial models. Moreover, non-considering individual unobserved heterogeneity leads to an underestimation of ρ (0.196) and to greater variability in firms' inefficiency level σ_u^2. Finally, non-controlling for individual fixed effects, we obtain a positive effect of *Size* on firms' efficiency level indicating that the negative impact of the number of employees previously detected only occurs when controlling for unobserved individual characteristics such as managerial abilities, skills, talents, psychological traits, and motivation.

Third, we check the robustness of our estimates to different sample compositions. Specifically, rather than considering start-ups after the first year of activity (N=2237), we select firms with at least two (N=1540) or three (N=858) years of activity in order to provide more robust evidence on the effect of the covariates, especially those included in the inefficiency model. The results shown in Table C3 in Appendix C provide consistent insights with those obtained using the complete sample. The most remarkable difference concerns the level of global spatial dependence captured by ρ which drastically decreases considering fewer observations, as expected.

Finally, since the magnitudes of the estimates by subsector shown in Table 4 may depend on the different sample sizes, we estimate the overall model including some interaction terms between sectoral typology and the inefficiency determinants. The results shown in Table C4 in Appendix C are in line with those of Table 4 in terms of signs and significance leading to analogous conclusions.

6 Conclusion

Previous studies in stochastic frontier model literature mainly concentrated on introducing some spatial components in order to consider spatial cross-section dependence or on controlling for individual unobserved heterogeneity through fixed or random effects. However, in empirical case studies, it is fundamental to consider both aspects because, on one hand, firms located in neighbouring locations tend to interact and influence each other and on the other, unobserved individual-specific characteristics may determine substantial differences in firms' production processes. Therefore, in this work, we develop a novel spatial stochastic frontier model with fixed effects aiming to control both aspects. In particular, we include the spatial autoregressive term in the model specification to consider global productivity spillovers and following the proposal of Wang and Ho, we remove fixed effects by within transformation. The resulting specification may be estimated using ML techniques and direct, indirect and total effects of the input variables on firms' output may be computed as proposed by LeSage and Pace. A further characteristic of our approach concerns the possibility of handling unbalanced panel data, which is a noteworthy aspect when dealing with real data.

The results of our simulation studies indicate that the proposed spatial within transposed SF model

can be estimated with negligible bias even with small sample sizes and high percentages of missing (at random) data. Moreover, we find that the variance parameter σ_v^2 absorbs unmodelled spatial dependence if spatial correlation is not adequately handled. On the other hand, not controlling for individual heterogeneity leads to upward estimates of the β parameters and downward estimates of the spatial parameter ρ. Thus, basic SF models suffer from many issues related to unmodelled spatial interactions and/or individual heterogeneity leading to biased estimates and decreased statistical efficiency.

The proposed model is used to investigate the productive performance of Italian innovative start-ups taking into account spillover effects occurring among neighbouring producers and uncontrolled individual characteristics. In particular, we estimate the spatial within transposed SF model using balance sheet data on 2237 Italian innovative start-ups in the period 2016-2020. The results of our analysis indicate that Italian start-ups are characterized by a significant level of global spatial dependence, especially overall and among firms belonging to the information and communication sector. Accordingly, we detect positive and significant spillover effects related to labour and capital, indicating that being located in an area with greater availability of workers, assets, and skills generates positive returns to neighbouring units. Moreover, we find that innovative activity proxied by patent rights, R&D expenditure and the share of intangibles on total capital is strongly related to firms' efficiency. On the other hand, larger team size and financial leverage tend to decrease firms' efficiency, mainly for start-ups in the manufacturing industry. In terms of efficiency scores, the mean efficiency level of Italian innovative start-ups equals 0.679 in the time period considered while concentrating on the specific industrial sectors, the manufacturing industry results to be the less efficient one whereas the scientific and technological sector reaches the highest efficiency score. Our findings may suggest relevant insights for policy makers. First, networking and collaboration among neighbouring entrepreneurs should be encouraged due to positive productivity returns arising from spatial interactions. In particular, synergies among firms belonging to different industrial sectors may benefit Italian start-ups more than intra-sectoral cooperation. Second, it is fundamental to further stimulate firms' innovative activity due to its positive link with start-up performance. Thus, government incentives should be aimed at supporting networking and personal interactions as well as internal research activity in order to create a fertile competitive environment that facilitates learning and knowledge transfer.

Further developments of this study may consider including additional spatial terms related to the error components in order to take into account spatial dependence related to the inefficiency or to the random error term. Moreover, it should be introduced the possibility of controlling for possible sources of endogeneity using an instrumental variable approach.

References

Acs, Z.J., D.B. Audretch, E.E. Lehmann, and G. Licht. "National systems of entrepreneurship". *Small Business Economics* 46 (2016): 527–535. https://doi.org/https://doi.org/10.1016/j.respol.2013.08.016.

Acs, Z.J., and D.B. Audretsch. "Innovation and small firms". *Mit Press* (1990).

Adetutu, M., A.J. Glass, K. Kenjegalieva, and R.C. Sickles. "The effects of efficiency and TFP growth on pollution in Europe: A multistage spatial analysis". *Journal of Productivity Analysis* 43 (2015): 307–326. https://doi.org/https://doi.org/10.1007/s11123-014-0426-7.

Aigner, D., C.K. Lovell, and P. Schmidt. "Formulation and estimation of stochastic frontier production function models". *Journal of Econometrics* 6, no. 1 (1977): 21–37. https://doi.org/https://doi.org/10.1016/0304-4076(77)90052-5.

Antonietti, R., and F. Gambarotto. "The role of industry variety in the creation of innovative start-ups in Italy". *Small Business Economics* 54 (2018): 561–573. https://doi.org/https://doi.org/10.1007/s11187-018-0034-4.

Areal, F.J., K. Balcombe, and R. Tiffin. "Integrating Spatial Dependence into Stochastic Frontier Analysis". *Agricultural and Resource Economics* 56, no. 4 (2012): 521–541. https://doi.org/https://doi.org/10.1111/j.1467-8489.2012.00597.x.

Audretsch, D.B., and M. Belitski. "The missing pillar: The creativity theory of knowledge spillover entrepreneurship". *Small Business Economics* 41, no. 4 (2013): 819–836. https://doi.org/https://doi.org/10.1007/s11187-013-9508-6.

Audretsch, D.B., M. Belitski, and R. Caiazza. "Start-ups, Innovation and Knowledge Spillovers". *The Journal of Technology Transfer* 46 (2021): 1995–2016. https://doi.org/https://doi.org/10.1007/s10961-021-09846-5.

Audretsch, D.B., and A.N. Link. "Entrepreneurship and knowledge spillovers from the public sector". *International Entrepreneurship and Management Journal* 15, no. 1 (2019): 195–208. https://doi.org/https://doi.org/10.1007/s11365-018-0538-z.

Baltagi, B.H., P. Egger, and M. Pfaffermayr. "Estimating models of complex FDI: Are there third-country effects?" *Journal of Econometrics* 140 (2007): 260–281. https://doi.org/https://doi.org/10.1016/j.jeconom.2006.09.009.

Barboza, G., and A. Capocchi. "Innovative startups in Italy. Managerial challenges of knowledge spillovers effects on employment generation". *Journal of Knowledge Management* 24, no. 10 (2020): 2573–2596. https://doi.org/https://doi.org/10.1108/JKM-08-2019-0436.

Belotti, F., and G. Ilardi. "Consistent inference in fixed-effects stochastic frontier models". *Journal of Econometrics* 202, no. 2 (2018): 161–177. https://doi.org/https://doi.org/10.1016/j.jeconom.2017.09.005.

Belotti, F., G. Ilardi, and A.P. Mortari. "Estimation of stochastic frontier panel data models with spatial inefficiency" (2019). http://papers.ssrn.com/paper.taf?abstract_id=3396404.

Caiazza, R., M. Belitski, and D.B. Audretsch. "From latent to emergent entrepreneurship: the knowledge spillover construction circle". *The Journal of Technology Transfer* 45, no. 3 (2020): 694–704. https://doi.org/https://doi.org/10.1007/s10961-019-09719-y.

Carvalho, A. "Efficiency spillovers in Bayesian stochastic frontier models: application to electricity distribution in New Zealand". *Spatial Economic Analysis* 13, no. 2 (2018): 171–190. https://doi.org/https://doi.org/10.1080/17421772.2018.1444280.

Charoenrat, T., and C. Harvie. "The efficiency of SMEs in Thai manufacturing: A stochastic frontier analysis". *Economic Modelling* 43 (2014): 372–393. https://doi.org/https://doi.org/10.1016/j.econmod.2014.08.009.

Chen, Y.Y., P. Schmidt, and H.J. Wang. "Consistent estimation of the fixed effects stochastic frontier model". *Journal of Econometrics* 181 (2014): 65–76. https://doi.org/https://doi.org/10.1016/j.jeconom.2013.05.009.

Chung, W.Y., Y. Jo, and D. Lee. "Where should ICT startup companies be established? Efficiency comparison between cluster types". *Telematics and Informatics* 56 (2021): 101482. https://doi.org/https://doi.org/10.1016/j.tele.2020.101482.

Colombelli, A. "The impact of local knowledge bases on the creation of innovative start-ups in Italy". *Small Business Economics* 47 (2016): 383–396. https://doi.org/https://doi.org/10.1007/s11187-016-9722-0.

Egger, P., M. Pfaffermayr, and H. Winner. "An unbalanced spatial panel data approach to US state tax competition". *Economics Letters* 88 (2005): 329–335. https://doi.org/https://doi.org/10.1016/j.econlet.2005.03.002.

Elhorst, J.P. "Applied spatial econometrics: raising the bar". *Spatial Economic Analysis* 5 (2010): 9–28. https://doi.org/https://doi.org/10.1080/17421770903541772.

Elhorst, J.P. *Spatial panel data models*. Fischer, M.M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis. New York: Springer, 2010.

Fritsch, M., and M. Wyrwich. "Regional knowledge, entrepreneurial culture, and innovative start-ups over time and space an empirical investigation". *Small Business Economics* 51 (2018): 337–353. https://doi.org/https://doi.org/10.1007/s11187-018-0016-6.

Fusco, E., and F. Vidoli. "Spatial stochastic frontier models: controlling spatial global and local heterogeneity". *International Review of Applied Economics* 27, no. 5 (2013): 679–694. https://doi.org/https://doi.org/10.1080/02692171.2013.804493.

Galli, F. "A spatial stochastic frontier model including both frontier and error based spatial cross-sectional dependence". *Spatial Economic Analysis* 18, no. 2 (2023): 239–258. https://doi.org/https://doi.org/10.1080/17421772.2022.2097729.

Galli, F. "A spatial stochastic frontier model introducing inefficiency spillovers". *Journal of the Royal Statistical Society Series C: Applied Statistics* 72, no. 2 (2023): 346–367. https://doi.org/https://doi.org/10.1093/jrsssc/qlad012.

Glass, A., K. Kenjegalieva, and R.C. Sickles. "Estimating efficiency spillovers with state level evidence for manufacturing in the US". *Economic Letters* 123 (2014): 154–159. https://doi.org/https://doi.org/10.1016/j.econlet.2014.01.037.

Glass, A.J., K. Kenjegalieva, and R.C. Sickles. "A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers". *Journal of Econometrics* 190, no. 2 (2016): 289–300. https://doi.org/https://doi.org/10.1016/j.jeconom.2015.06.011.

Greene, W. "Fixed and Random Effects in Stochastic Frontier Models". *Journal of Productivity Analysis* 23 (2005a): 7–32. https://doi.org/https://doi.org/10.1007/s11123-004-8545-1.

Greene, W. "Reconsidering heterogeneity in panel data estimators of the stochastic frontier model". *Journal of Econometrics* 126 (2005b): 269–303. https://doi.org/https://doi.org/10.1016/j.jeconom.2004.05.003.

Griliches, Z. "The Search for R&D Spill-Overs". *Scandinavian Journal of Economics* 94 (1992): 29–47. https://doi.org/https://doi.org/10.2307/3440244.

Gude, A., I. Alvarez, and L. Orea. "Heterogeneous spillovers among Spanish provinces: a generalized spatial stochastic frontier model". *Journal of Productivity Analysis* 50 (2018): 155–173. https://doi.org/https://doi.org/10.1007/s11123-018-0540-.

Hambrick, D., T. Cho, and M. Chen. "The influence of top management team heterogeneity on firms' competitive moves". *Administrative Science Quarterly* 41, no. 4 (1996): 659–684. https://doi.org/https://doi.org/10.2307/2393871.

Helmers, C., and M. Rogers. "Does patenting help high-tech start-ups?" *Research Policy* 40, no. 7 (2011): 1016–1027. https://doi.org/https://doi.org/10.1016/j.respol.2011.05.003.

Iaquinto, A., and J. Fredrickson. "Top management team agreement about the strategic decision process: A test of some of its determinants and consequences". *Strategic Management Journal* 18, no. 1 (1998): 63–75. https://doi.org/https://doi.org/10.1002/(SICI)1097-0266(199701)18:1<63::AID-SMJ835>3.0.CO;2-N.

Jacobs, J. *The Economy of Cities*. London: Jonathan Cape, 1969.

Jin, F., and L. Lee. "Asymptotic properties of a spatial autoregressive stochastic frontier model". *Journal of Spatial Econometrics* 1, no. 2 (2020): 1–40. https://doi.org/https://doi.org/10.1007/s43071-020-00002-z.

Kelejian, H.H., and I.R. Prucha. "A generalized moments estimator for the autoregressive parameter in a spatial model". *International Economic Review* 40, no. 2 (1999): 509–533. https://doi.org/https://doi.org/10.1111/1468-2354.00027.

Kelejian, H.H., and I.R. Prucha. "A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance". *Journal of Real Estate Finance and Economics* 17 (1998): 99–121. https://doi.org/https://doi.org/10.1023/A:1007707430416.

Kumbhakar, S.C. "Estimation of technical inefficiency in panel data models with firm- and time-specific effects". *Economics Letters* 36, no. 1 (1991): 43–48. https://doi.org/https://doi.org/10.1016/0165-1765(91)90053-N.

Lai, H., and K.C. Tran. "Persistent and transient inefficiency in a spatial autoregressive panel stochastic frontier model". *Journal of Productivity Analysis* 58 (2022): 1–13. https://doi.org/https://doi.org/10.1007/s11123-022-00638-z.

Lee, L.F., and J. Yu. "Estimation of spatial autoregressive panel data models with fixed effects". *Journal of Econometrics* 154 (2010): 165–185. https://doi.org/https://doi.org/10.1016/j.jeconom.2009.08.001.

LeSage, J.P., and R.K Pace. *Introduction to Spatial Econometrics*. Boca-Raton: Taylor & Francis, 2009.

Liang, P., M. Rajan, and K. Ray. "Optimal team size and monitoring in organizations". *The Accounting Review* 83, no. 3 (2008): 789–822. https://doi.org/https://doi.org/10.2308/accr.2008.83.3.789.

Marshall, A. *Principles of Economics*. London: Macmillan, 1890.

Matricano, D., E. Candelo, and M. Sorrentino. "Start-ups' innovation processes and performance in the food industry: a stochastic frontier analysis". *British Food Journal* 124, no. 3 (2022): 936–950. https://doi.org/https://doi.org/10.1108/BFJ-10-2020-0944.

Orea, L., and I.C. Alvarez. "A new stochastic frontier model with cross-sectional effects in both noise and inefficiency terms". *Journal of Econometrics* 213 (2019): 556–577. https://doi.org/https://doi.org/10.1016/j.jeconom.2019.07.004.

Pace, R.K., and R. Barry. "Quick Computation of Spatial Autoregressive Estimators". *Geographical Analysis* 29, no. 3 (1997): 232–247. https://doi.org/https://doi.org/10.1111/j.1538-4632.1997.tb00959.x.

Ramajo, J., and G.J. Hewings. "Modelling regional productivity performance across Western Europe". *Regional Studies* 52, no. 10 (2018): 1372–1387. https://doi.org/https://doi.org/10.1080/00343404.2017.1390219.

Santi, F., M.M. Dickson, G. Espa, E. Taufer, and A. Mazzitelli. "Handling spatial dependence under unknown unit locations". *Spatial Economic Analysis* 16, no. 2 (2021): 194–216. https://doi.org/https://doi.org/10.1080/17421772.2020.1769171.

Schmidt, A.M., A.R.B. Moreira, S.M. Helfand, and T.C.O. Fonseca. "Spatial stochastic frontier models: accounting for unobserved local determinants of inefficiency". *Journal of Productivity Analysis* 31 (2009): 101–112. https://doi.org/https://doi.org/10.1007/s11123-008-0122-6.

Skevas, I., and A.O. Lansink. "Dynamic Inefficiency and Spatial Spillovers in Dutch Dairy Farming". *Journal of Agricultural Economics* 71, no. 3 (2020): 742–759. https://doi.org/https://doi.org/10.1111/1477-9552.12369.

Tsionas, E.G, and P.G. Michaelides. "A Spatial Stochastic Frontier Model with Spillovers: Evidence for Italian Regions". *Scottish Journal of Political Economy* 63, no. 3 (2016): 243–257. https://doi.org/https://doi.org/10.1111/sjpe.12081.

Tsukamoto, T. "A spatial autoregressive stochastic frontier model for panel data incorporating a model of technical inefficiency". *Japan and the World Economy* 50 (2019): 66–77. https://doi.org/https://doi.org/10.1016/j.japwor.2018.11.003.

Wang, H.J., and C.W. Ho. "Estimating fixed-effect panel stochastic frontier models by model transformation". *Journal of Econometrics* 157, no. 2 (2010): 286–296. https://doi.org/https://doi.org/10.1016/j.jeconom.2009.12.006.

Wang, H.J., and P. Schmidt. "One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels". *Journal of Productivity Analysis* 18, no. 2 (2002): 129–144. https://doi.org/https://doi.org/10.1023/A:1016565719882.

Notes

[1]Matlab codes are available in the Github repository https://github.com/FedericaGalli17/Spatial-Stochastic-Frontier-Models.git

[2]By law 179/2012 art.25, Italian start-ups are defined as innovative if they meet at least one of the following requirements: having R&D expenses equal to at least 15% of the higher value between the cost and the total value

of the production; employing highly qualified personnel (at least 1/3 of employees with research doctorates, doctoral students or researchers, or at least 2/3 of workers with a master's degree); being the owner, custodian or licensee of at least one patent or registered software.

Table 1: Simulation results for different sample sizes with T=5

		pm=0%		**pm=10%**		**pm=20%**		**pm=30%**		**pm=40%**	
N=100	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.499	0.001	0.500	0.001	0.502	0.008	0.502	0.001	0.501	0.001
		(0.022)		(0.026)		(0.023)		(0.029)		(0.031)	
ρ	0.4	0.384	0.002	0.390	0.001	0.388	0.002	0.390	0.002	0.383	0.004
		(0.034)		(0.041)		(0.035)		(0.043)		(0.057)	
δ	0.5	0.501	0.003	0.499	0.004	0.499	0.005	0.504	0.004	0.502	0.009
		(0.051)		(0.058)		(0.063)		(0.051)		(0.083)	
μ	0.5	0.505	0.021	0.510	0.023	0.508	0.030	0.496	0.003	0.511	0.054
		(0.126)		(0.139)		(0.143)		(0.141)		(0.191)	
σ_u^2	0.1	0.101	0.003	0.104	0.004	0.102	0.005	0.097	0.004	0.104	0.010
		(0.033)		(0.051)		(0.030)		(0.031)		(0.036)	
σ_v^2	0.2	0.199	0.000	0.198	0.000	0.199	0.004	0.198	0.000	0.200	0.001
		(0.014)		(0.017)		(0.017)		(0.018)		(0.020)	
N=200	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.501	0.000	0.500	0.000	0.501	0.000	0.501	0.001	0.500	0.001
		(0.017)		(0.017)		(0.018)		(0.022)		(0.023)	
ρ	0.4	0.389	0.001	0.389	0.001	0.390	0.001	0.389	0.001	0.388	0.002
		(0.023)		(0.025)		(0.027)		(0.031)		(0.036)	
δ	0.5	0.500	0.001	0.500	0.002	0.499	0.002	0.499	0.003	0.497	0.005
		(0.035)		(0.038)		(0.042)		(0.055)		(0.065)	
μ	0.5	0.503	0.008	0.501	0.010	0.506	0.012	0.507	0.016	0.515	0.028
		(0.081)		(0.093)		(0.099)		(0.132)		(0.174)	
σ_u^2	0.1	0.100	0.001	0.102	0.002	0.100	0.002	0.101	0.003	0.102	0.004
		(0.025)		(0.036)		(0.024)		(0.047)		(0.063)	
σ_v^2	0.2	0.200	0.000	0.199	0.000	0.199	0.000	0.200	0.000	0.200	0.000
		(0.011)		(0.012)		(0.012)		(0.014)		(0.017)	
N=300	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.500	0.000	0.500	0.000	0.500	0.000	0.500	0.000	0.500	0.000
		(0.013)		(0.015)		(0.016)		(0.015)		(0.019)	
ρ	0.4	0.389	0.001	0.391	0.001	0.391	0.001	0.391	0.001	0.388	0.001
		(0.019)		(0.022)		(0.023)		(0.027)		(0.030)	
δ	0.5	0.501	0.001	0.500	0.001	0.499	0.001	0.499	0.002	0.500	0.003
		(0.030)		(0.034)		(0.040)		(0.049)		(0.054)	
μ	0.5	0.499	0.006	0.503	0.007	0.504	0.009	0.506	0.012	0.503	0.015
		(0.071)		(0.082)		(0.090)		(0.117)		(0.128)	
σ_u^2	0.1	0.099	0.001	0.102	0.001	0.102	0.001	0.099	0.002	0.099	0.002
		(0.023)		(0.026)		(0.024)		(0.045)		(0.047)	
σ_v^2	0.2	0.200	0.000	0.200	0.000	0.200	0.000	0.200	0.000	0.201	0.000
		(0.009)		(0.010)		(0.011)		(0.011)		(0.015)	

Table 2: Simulation results for different sample sizes with T=10

		pm=0%		**pm=10%**		**pm=20%**		**pm=30%**		**pm=40%**	
N=100	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.500	0.000	0.501	0.000	0.500	0.000	0.500	0.000	0.501	0.000
		(0.016)		(0.011)		(0.017)		(0.018)		(0.021)	
ρ	0.4	0.397	0.001	0.396	0.001	0.396	0.001	0.397	0.001	0.399	0.001
		(0.022)		(0.026)		(0.025)		(0.031)		(0.039)	
δ	0.5	0.499	0.001	0.498	0.001	0.500	0.001	0.499	0.001	0.499	0.002
		(0.027)		(0.030)		(0.030)		(0.035)		(0.046)	
μ	0.5	0.502	0.005	0.505	0.007	0.502	0.008	0.504	0.009	0.508	0.011
		(0.071)		(0.081)		(0.082)		(0.090)		(0.113)	
σ_u^2	0.1	0.100	0.001	0.099	0.001	0.100	0.001	0.099	0.002	0.099	0.002
		(0.023)		(0.038)		(0.042)		(0.021)		(0.038)	
σ_v^2	0.2	0.199	0.000	0.199	0.000	0.199	0.000	0.199	0.000	0.198	0.000
		(0.010)		(0.011)		(0.011)		(0.012)		(0.015)	
N=200	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.500	0.000	0.500	0.000	0.500	0.000	0.500	0.000	0.500	0.000
		(0.011)		(0.012)		(0.013)		(0.013)		(0.015)	
ρ	0.4	0.395	0.000	0.395	0.000	0.397	0.000	0.397	0.000	0.398	0.001
		(0.017)		(0.017)		(0.019)		(0.022)		(0.025)	
δ	0.5	0.501	0.000	0.499	0.001	0.500	0.001	0.501	0.001	0.500	0.001
		(0.020)		(0.021)		(0.023)		(0.027)		(0.031)	
μ	0.5	0.498	0.003	0.502	0.004	0.502	0.004	0.499	0.005	0.501	0.006
		(0.054)		(0.058)		(0.058)		(0.068)		(0.084)	
σ_u^2	0.1	0.099	0.000	0.100	0.001	0.099	0.001	0.100	0.001	0.099	0.001
		(0.023)		(0.021)		(0.019)		(0.025)		(0.050)	
σ_v^2	0.2	0.200	0.000	0.200	0.000	0.200	0.000	0.200	0.000	0.199	0.000
		(0.007)		(0.007)		(0.008)		(0.008)		(0.009)	
N=300	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.501	0.000	0.500	0.000	0.500	0.000	0.500	0.000	0.500	0.000
		(0.009)		(0.010)		(0.010)		(0.011)		(0.012)	
ρ	0.4	0.396	0.000	0.397	0.000	0.397	0.000	0.397	0.000	0.398	0.000
		(0.013)		(0.014)		(0.016)		(0.017)		(0.019)	
δ	0.5	0.500	0.000	0.499	0.000	0.500	0.000	0.501	0.001	0.501	0.001
		(0.017)		(0.018)		(0.021)		(0.023)		(0.026)	
μ	0.5	0.500	0.002	0.503	0.002	0.502	0.003	0.500	0.003	0.497	0.004
		(0.043)		(0.048)		(0.053)		(0.059)		(0.065)	
σ_u^2	0.1	0.100	0.000	0.100	0.000	0.099	0.000	0.099	0.000	0.099	0.001
		(0.015)		(0.021)		(0.021)		(0.026)		(0.022)	
σ_v^2	0.2	0.200	0.000	0.200	0.000	0.200	0.000	0.200	0.000	0.200	0.000
		(0.006)		(0.006)		(0.007)		(0.007)		(0.008)	

Table 3: Performance of spatial and non-spatial SF models with and without fixed effects

		SSF with FE		SF with FE		SSF		SF	
N=100	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.499	0.112e-03	0.506	0.168e-03	0.520	0.523e-03	0.521	0.551e-03
		(0.011)		(0.014)		(0.011)		(0.013)	
ρ	0.4	0.396	0.351e-03	-	-	0.332	0.005	-	-
		(0.019)				(0.014)			
δ	0.5	0.501	0.507e-03	0.500	0.592e-03	0.574	0.006	0.439	0.004
		(0.024)		(0.028)		(0.010)		(0.006)	
μ	0.5	0.499	0.004	0.525	0.005	0.285	0.048	0.692	0.039
		(0.068)		(0.079)		(0.045)		(0.030)	
σ_u^2	0.2	0.198	0.002	0.217	0.003	0.215	0.001	0.299	0.012
		(0.054)		(0.064)		(0.038)		(0.045)	
σ_v^2	0.1	0.100	0.025e-03	0.163	0.004	0.112	0.167e-03	0.167	0.005
		(0.005)		(0.008)		(0.005)		(0.007)	
N=200	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.500	0.059e-03	0.519	0.442e-03	0.538	0.002	0.536	0.001
		(0.008)		(0.010)		(0.008)		(0.010)	
ρ	0.4	0.398	0.153e-03	-	-	0.304	0.009	-	-
		(0.012)				(0.010)			
δ	0.5	0.500	0.226e-03	0.502	0.238e-03	0.586	0.007	0.477	0.568e-03
		(0.015)		(0.019)		(0.007)		(0.005)	
μ	0.5	0.502	0.002	0.528	0.003	0.238	0.070	0.575	0.007
		(0.043)		(0.050)		(0.029)		(0.021)	
σ_u^2	0.2	0.200	0.760e-03	0.211	0.001	0.213	0.782e-03	0.279	0.007
		(0.028)		(0.033)		(0.021)		(0.027)	
σ_v^2	0.1	0.100	0.012e-03	0.168	0.005	0.119	0.361e-03	0.167	0.005
		(0.004)		(0.006)		(0.004)		(0.005)	
N=300	True Values	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*	*Mean*	*MSE*
β	0.5	0.500	0.043e-03	0.522	0.052e-02	0.542	0.002	0.541	0.002
		(0.007)		(0.008)		(0.007)		(0.008)	
ρ	0.4	0.398	0.112e-03	-	-	0.307	0.009	-	-
		(0.010)				(0.008)			
δ	0.5	0.500	0.154e-03	0.497	0.018e-02	0.579	0.006	0.470	0.001
		(0.013)		(0.015)		(0.005)		(0.004)	
μ	0.5	0.501	0.001	0.548	0.004	0.269	0.054	0.617	0.014
		(0.036)		(0.042)		(0.025)		(0.018)	
σ_u^2	0.2	0.200	0.551e-03	0.215	0.981e-03	0.224	0.001	0.296	0.001
		(0.026)		(0.032)		(0.023)		(0.027)	
σ_v^2	0.1	0.100	0.008e-03	0.168	0.005	0.119	0.371e-03	0.168	0.005
		(0.003)		(0.005)		(0.003)		(0.004)	

Note: SSF with FE: spatial stochastic frontier with fixed effects; SF with FE: non-spatial stochastic frontier with fixed effects; SSF: spatial stochastic frontier; SF: non-spatial stochastic frontier.

Table 4: Empirical model: Estimation results

	OVERALL		**MANUF**		**INF&COMM**		**SCI&TECH**	
	Coeff	*SD*	*Coeff*	*SD*	*Coeff*	*SD*	*Coeff*	*SD*
β_L	0.692***	0.018	0.802***	0.050	0.674***	0.026	0.548***	0.038
β_K	0.424***	0.022	0.184***	0.052	0.450***	0.034	0.482***	0.043
β_{LL}	0.001**	0.004	0.007	0.011	0.022***	0.006	0.039***	0.009
β_{KK}	-0.008*	0.005	0.030***	0.011	-0.009	0.008	-0.015*	0.010
β_{LK}	-0.068***	0.004	-0.061***	0.009	-0.092***	0.006	-0.0676***	0.008
ρ	0.224***	0.009	0.099***	0.020	0.210***	0.013	0.189***	0.016
δ_{Size}	0.953***	0.089	2.123***	0.294	1.346***	0.129	0.509***	0.092
δ_{Lev}	0.050***	0.006	0.098***	0.020	0.014*	0.008	0.042***	0.007
δ_{Liquid}	-0.003	0.000	0.007	0.024	0.052***	0.015	-0.025**	0.009
δ_{Int}	-1.237***	0.151	1.564***	0.416	-0.555***	0.172	-0.779***	0.155
δ_{Pat}	-0.099***	0.014	-0.146***	0.039	-0.169***	0.023	-0.034**	0.012
$\delta_{R\&D}$	-0.079***	0.015	-0.386***	0.071	-0.132***	0.026	-0.043*	0.026
σ_u^2	0.195		0.033		0.121		0.937	
σ_v^2	0.801		0.789		0.778		0.759	
LL	6147.05		850.88		2939.71		1590.71	
TE	0.679		0.273		0.494		0.542	

$***: pvalue \leq 0.01$; $**: pvalue \leq 0.05$; $*: pvalue \leq 0.10$

Table 5: Empirical model: Marginal effects

		OVERALL		**MANUF**		**INF&COMM**		**SCI&TECH**	
		Coeff	*SD*	*Coeff*	*SD*	*Coeff*	*SD*	*Coeff*	*SD*
L	*Direct*	0.607***	0.026	0.689***	0.035	0.621***	0.018	0.523***	0.029
	Indirect	0.131***	0.015	0.056***	0.012	0.123***	0.010	0.091***	0.01
	Total	0.738***	0.034	0.745***	0.039	0.744***	0.023	0.614***	0.034
K	*Direct*	0.299***	0.038	0.129***	0.038	0.278***	0.024	0.362***	0.033
	Indirect	0.063***	0.010	0.011***	0.004	0.053***	0.005	0.062***	0.008
	Total	0.363***	0.046	0.140***	0.041	0.331***	0.028	0.425***	0.039

$***: pvalue \leq 0.01; \quad **: pvalue \leq 0.05; \quad *: pvalue \leq 0.10$

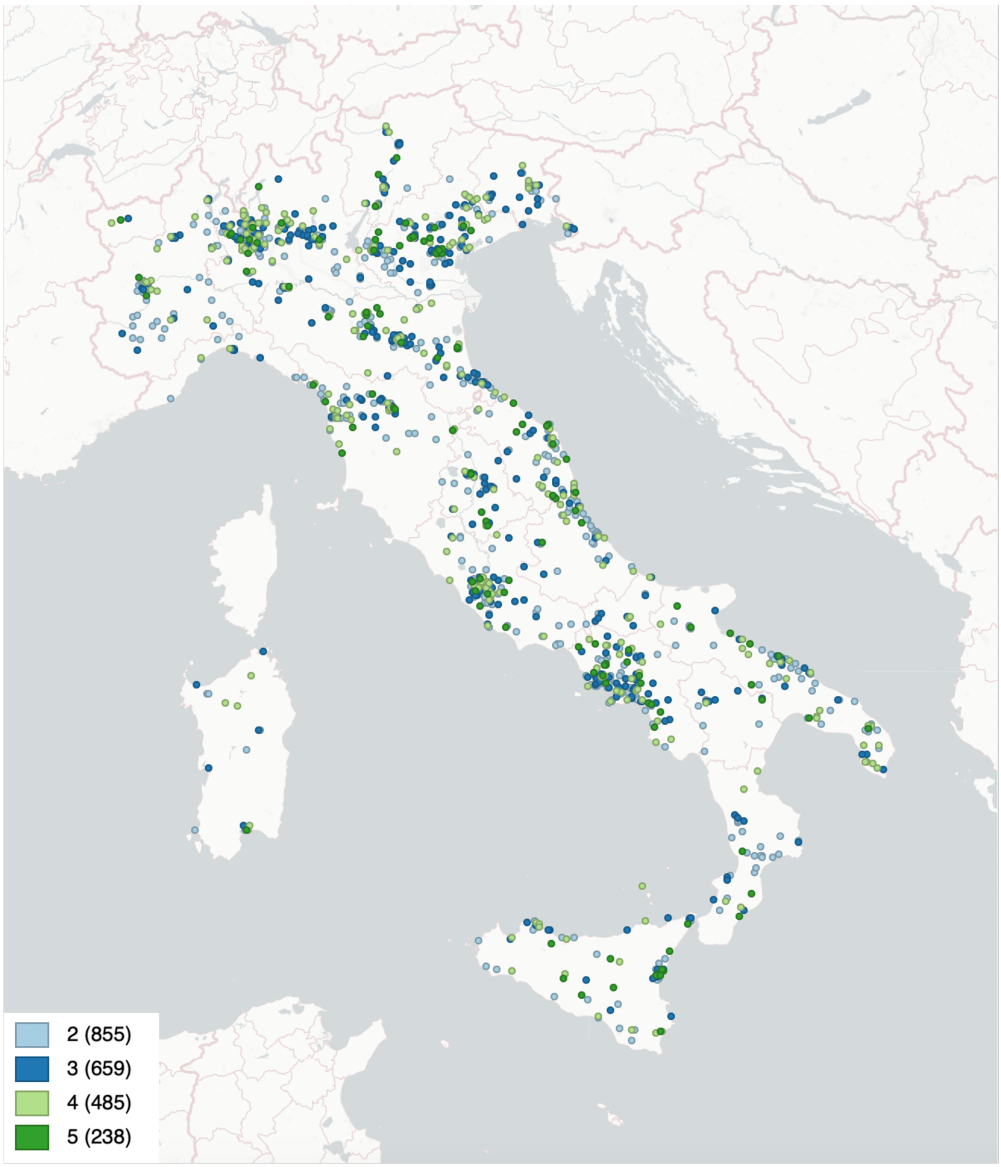

Figure 1. Innovative Start-ups Map