Tesis para optar al grado de
Master en Economía

Título:

"Using the Similarity of Trade Patterns Across Proximate
Countries to Identify the Presence of Technological Spillovers"

Autor: Sergio Petralia

Evaluadores: Dr. Facundo Albornoz
Dr. Marcos Herrera Gómez
Dr. Enrique Kawamura

Victoria – Buenos Aires
Agosto 2013

Using the Similarity of Trade Patterns Across Proximate Countries to Identify the Presence of Technological Spillovers

Sergio Petralia

1 Introduction:

International technology diffusion has received considerable attention by both academic researchers and policy makers. The degree and extent to which technology diffuses determines the global technology frontier and may contribute reducing the gap between developed and developing economies.

In this paper we will try to determine the existence of agglomeration effects on countries' trade structures after the effect of factor endowments is netted out. Agglomeration effects would induce similarity of trade patterns among nearby countries beyond what can be explained by factor endowments. We will study the observed pattern of agglomeration and postulate a process through which it may have emerged. We will inspect these effects taking into account the intensity in the use of technology by industries to later argue that agglomeration may be caused by technological spillovers.An appealing feature of our approach is that we will incorporate industry heterogeneity as an essential part of the argument. As opposed to the existing literature we will argue that agglomeration effects, if present, may differ by sector, stressing the importance of incorporating sector heterogeneity in the analysis of technological diffussion.

Adopting new technologies requires absorbing knowledge. Firms acquire this knowledge from several sources, being one of them the interaction with other agents. Interactions between firms are needed because of the partially "tacit" nature of technology, as only some components of a new technology can be completely codified. Transmission of this non-codified knowledge is costly, requiring sometimes even face-to-face interactions. It is reasonable to conjecture that this sort of technological diffusion may be affected by geography.

If this "tacit" component of technology exits we should see two patterns emerging. First, technology diffusion should be incomplete, creating regional clusters of technology adoption. Second, as the "tacit" component of a technology increases it will be more likely to see regional/clustered patterns of diffusion.

Technology is not directly measurable, making any attempt to estimate its impact very difficult. However, the literature has tried to circumvent this problem by analyzing technological

spillovers indirectly. In general it is possible to categorize papers dealing with technology diffusion into three, not mutually exclusive, groups[1]. First we have those which try to approach the problem by measuring what are considered inputs for technological progress and innovation. This set of papers usually employ measures of resources devoted to Research and Development (R&D) to evaluate how these efforts, which may be undertook within and/or outside the sector/country, affect performance. Examples are Scherer (1984),Coe and Helpman (1995), Griliches and Mairesse (1995), Keller (2000 and 2001), Branstetter (2001), Peri (2002) and Savvides and Zachariadis (2005). See Keller (2001) for an exhaustive literature review on that sort of models.

The second group covers papers which focus their attention on measures of outputs of the technological progress, like patents. For instance Eaton and Kortum (1996 and 1999) use data on patents to estimate a model of technology diffusion and productivity growth. Jaffe et al. (1993) and Jaffe and Trajtenberg (1998) studied the diffusion of knowledge by looking at patent's citations.

The third category includes papers which study technology diffusion by investigating its effects on the economic activity. Most of the analysis has been carried out by measuring the impact of technology on total factor productivity (TFP) of countries/sectors. For instance, Griliches (1984) uses TFP measures and data on R&D expenditures to study technical change. Another example is Katayama, Lu, and Tybout (2003). Our paper can be included in the third group as we will analyze to what extent technological spillovers affect countries' trade structures.

Two of the most studied channels for technology diffusion are trade in intermediate goods and Foreign Direct Investment (FDI). If trade in intermediate goods disseminates knowledge, nearby countries could take advantage of technological progress somewhere else by importing embodied technology. Hence, progress made in one country could improve productivity of industries located upstream on the productive chain, generating a more similar pattern of comparative advantages among countries that adopt this technology. We should expect that this diffusion is geographically localized because of the widely studied fact that trade falls with distance (see Rivera-Batiz and Romer (1991), Grossman and Helpman (1991), and Eaton and Kortum (2002) for models incorporating trade in intermediate goods).

International mobility of technology through FDI is incomplete in the sense that a share of it is non-codifiable or "tacit". Consider the case of a multinational spreading some sort of technology to an affiliate through FDI. This will tend to increase productivity of a particular sector in the affiliate's country. Also, as the transmission of the non-codified part of this knowledge is costly, diffusion may be affected by geography. Therefore, we can expect regional clusters of technology dissemination influencing trade patterns (see Keller and Yeaple (2009) for a model in which a multinational's production vary according to the codifiability of their technology).

We will conduct an empirical analysis using bilateral trade data at the SITC 2-digit level for 80 countries. There are 62 industries at that level of aggregation. As mentioned before, we aim to document the existence of agglomeration effects on countries' trade structures after the effect of factor endowments is netted out. Also, whether the intensity of these effects is higher on more technologically advanced industries. If the "tacit " component of knowledge plays indeed a crucial role in shaping the pattern of technological diffusion, the magnitude of this diffusion should differ across sectors. In particular, if the share of non-codifiable knowledge is higher in more technologically advanced industries we should expect a stronger pattern of agglomeration

[1] There exist also the literature of learning by exporting, examples are Clerides, Lach, and Tybout (1998), and Albornoz and Ercolani (2008).

in these sectors. We will carry out our analysis by grouping sectors into two broad categories, manufacture and agriculture.

We will implement a two step econometric approach. The first step will entail netting out the effect of factor endowments from countries' trade structures while the second step will test the presence of agglomeration effects.

QUAERERE VERUM
Universidad de
SanAndrés

2 Data:

In order to test our hypothesis we will need data on bilateral trade, country endowments, and distance between countries. This paper makes use of data coming from six different sources. First, bilateral trade data was taken from the dataset compiled by Feenstra et al (2005), which is available at the National Bureau of Economic Research (NBER). The dataset is organized by 4-Digit SITC categories (revision 2). World trade flows are drawn from United Nations (UN) data sources and are based primarily on reporter-country import data, supplemented by reporter-country export data where import data gaps occur. For the present study trade data was aggregated to the 2-Digit SITC level[2]

Regarding countries' endowments we relied on several sources to finally collect 6 different variables, capital, labour and skilled labour intensity along with measures of crop, arable land and mineral resources. Land data was obtained from FAOSTAT available at the Food and Agriculture Organization (FAO) of the UN. The data on minerals production was taken from the U.S. Geological Survey (USGS), provided in the Minerals Yearbook Volume I[3].

Information about labour and skilled labour was built using data coming from the Penn World Tables (PWT), available at the Center for International Comparisons at the University of Pennsylvania (CICUP) and from EdStats, published by The World Bank. The former contains information about population by country while the later contains data on educational attainment for 142 countries from 1955-2000. For further references see Barro and Lee (2000). Data on physical capital stock and bilateral distance was taken from The Global Development Network database and from Andrew Rose's web page, respectively.

The notion of distance among countries is a key feature of the analysis we will develop on the next sections is. In what follows we will measure proximity between countries as the distance in kilometers between their capitals[4].

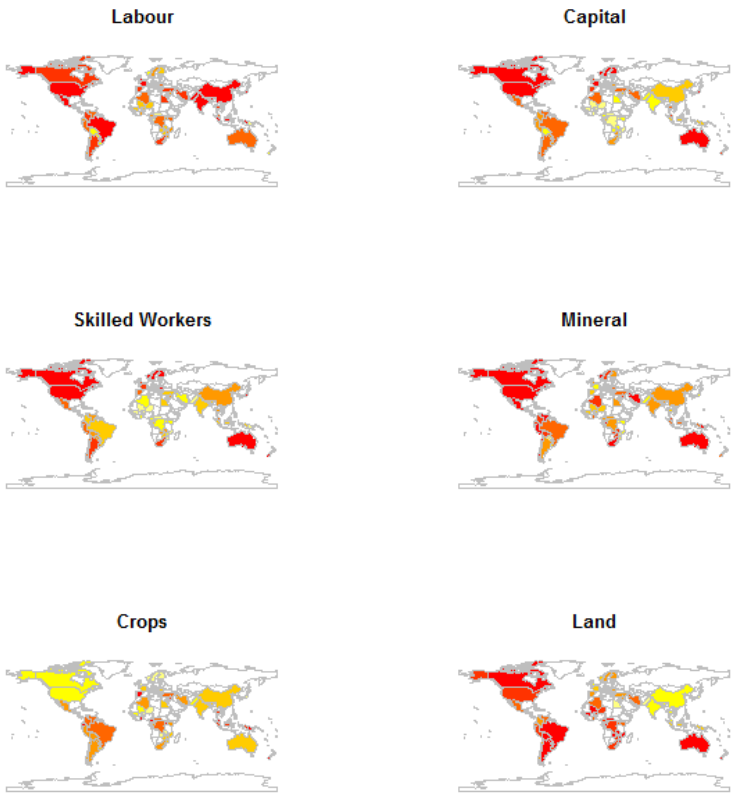
Several aspects of the data at hand are worth pointing out. First, we are going to work with a cross section of data for the year 1996. Information about minerals production by countries was available only for 1996, limiting the scope of the present study to that year. Second, the SITC trade data will be classified into manufactured and non manufactured goods following Adrian Wood (2000).Then the manufacturing industries are those corresponding to SITC codes larger than 50, except for the 90's sectors and sector 68. The 90's sectors are excluded from the sample. Third, sectors 65 and 84 were merged and treated as a single sector reducing the total number of industries to 61

After merging the data from these different sources we end up with a dataset consisting of bilateral trade data for 61 industries in 80 countries. We count with 6 measures of factor endowments: capital, labour, skilled labour, crop land, arable land, and mineral resources.

Figure 1 below contains information about countries' factor endowments. Maps display countries' relative positions in terms each factor endowment by assigning a darker colour to countries where the factor is relatively more abundant. Countries for which information was not available appear in white.

[2] In section 5 we will perform a more disaggregated analysis at the 3-Digit level.
[3] See the appendix for more details on how these measures were constructed.
[4] Results are mantained also if we consider distance in kilometers between countries' geographic centroids.

Figure 1: Countries' Factor Endowments



Universidad de San Andrés

Note: Maps display countries' relative positions in terms each factor endowment by assigning a darker colour to countries where factors are relatively more abundant. Countries for which information was not available appear in white.

Not surprisingly our measures for capital and skilled labour are higher in regions such as North America and Western Europe while they reach their lower levels in Africa. Latin America appears to be in between these two extreme cases. With respect to crop and land measures this pattern reverses, with Africa and Latin America being relatively more abundant than Europe and North America.

In the next section we will carry out our analysis using net trade (exports minus imports) over GDP as a way of characterizing countries' trade patterns. This measure has been used frequently in the trade literature to compare countries' trade patterns. It has the advantage that considers both exports and imports and corrects for country size to get rid of scale effects. In section 5 we will test if results are maintained as we use exports over GDP.

3 Methodology:

As previously mentioned, the main purpose of this paper is to determine the existence of agglomeration effects on countries' trade structures. These effects would induce similarity of trade patterns among nearby countries beyond what can be explained by factor endowments. In order to answer this question we opted to implement a two step econometric approach that has the advantage of being both parsimonious and very intuitive.

3.1 First Step: Netting out the effect of factor endowments

The first step towards identifying agglomeration effects entails netting out the effect of factor endowments from countries' trade patterns. For each of the 61 2-digit SITC sectors, we estimate a cross-country regression of net trade as a function of factor proportions, and compute the residuals from those regressions. The following is the proposed expression for the relationship between net trade and factor proportions in sector s:

$$\frac{N_s^c}{GDP^c} = \sum_{f=1}^{F} g_f^s \left(\ln(\frac{V_f^c}{V_O^c})\right) + \varepsilon_s^c \qquad s = 1, ..., 61 \quad (1)$$

where N_s^c is country c's net trade (exports minus imports) in sector s and GDP^c is country c's Gross Domestic Product. The variables V_f^c and V_f^o are, respectively, country c's endowments of factor f and factor O (We will use Labour as the numeraire). Ideally, we would like to have enough information in order to estimate non-parametrically the relationship between factor proportions and net trade. However the data at hand makes it unfeasible. We face the challenge of capturing as tightly as possible the influence of factor proportions on trade patterns without proposing a model that may overfit the data. For the baseline case we will use a specification for $g_f(.)$ that includes a third-degree polynomials in each variable[5]:

$$\frac{N_s^c}{GDP^c} = \sum_{f=1}^{F} \alpha_{1,f} \left(\ln(\frac{V_f^c}{V_o^c})\right) + \alpha_{2,f} \left(\ln(\frac{V_f^c}{V_o^c})\right)^2 + \alpha_{3,f} \left(\ln(\frac{V_f^c}{V_o^c})\right)^3 + \varepsilon_s^c \quad (2)$$

On the right-hand-side, we express factor proportions relative to labor, which we take as the numeraire factor. We use the logarithm of these ratios as our regressors. Taking logs has the main advantage of making our formulation invariant to which endowment is used as numeraire. The final specification includes five factor proportions: physical capital (K/L), human capital (H/L), cropland (C/L), non-crop land (T/L), and minerals (M/L).

The error term ε_s^c captures the component of trade not explained by countries' factor proportions and it is assumed to be uncorrelated with the regressors. On the left-hand-side, sectorial net exports are divided by the country's GDP to normalize for country size.

Once first-stage regressions are estimated, we can compute the residuals ε_s^c for each sector in each country .The residuals will capture the component of countries' net trade that cannot

[5] In section 5 we test our results proposing alternative specifications for the first stage.

be explained by factor endowments. If our hypothesis is true, the existence of technological spillovers would induce similarity of trade patterns among nearby countries beyond what can be explained by factor endowments. The error term being spatially correlated would indicate the existence of agglomeration effects, while stronger spatial dependence on more technologically advanced industries would suggest the existence technological spillovers.

Even if it is true that the existence of technological spillovers should induce the error term to be correlated across space, there are also other circumstances that might affect ε_s^c in such a way that it displays spatial dependence. From an econometric point of view our argument relies heavily on the assumption that we are able to completely net out the effect of factor endowments. This assumption may be false if either the model is misspecified or we have ommited variables. In section 5 we dealt with the former problem by including alternative specifications.

Consider the problem of omitting variables for this particular case, any endowment not considered in the regression may bias the results if it is spatially correlated. Which means that the observed spatial correlation or "agglomeration effect" is not such. Additionally, if this endowment is used more intensively in technologically advanced industries, our arguments about technological spillovers as a plausible explanation may be invalid. In section 6 we propose a way to deal with this problem by incorporating the time dimension.

From a theoretical point of view there are many arguments that can invalidate our results. Consider for instance climate shocks. It is possible that unusual weather conditions like high/low temperatures or excessive/insufficient rainfall may affect economic activity in such a way that trade patterns are altered. Those sorts of shocks usually affect entire regions and not only particular countries, meaning that ε_s^c will be spatially correlated beyond what can be explained by factor endowments. However, climate shocks are likely to have a stronger impact in agriculture than in manufacture sectors. Which contradicts our initial hypothesis that spillovers should be higher in more technologically advanced sectors.

Now consider the effect of regional shocks. Suppose there is a region in which, for a particular sector, countries differ in their pattern of trade. Assume that some countries exhibit positive net trade while others negative. In this example countries don't display a similar trade pattern. Now assume that a shock affects all countries in the region generating an excess of demand in their domestic markets, also suppose the shock is strong enough such that all countries end up with the same trade pattern (positive or negative). In this case a regional demand shock will certainly affect the trade pattern for that industry, and consequently our estimates. However, note that this effect will be significant only if the shock is strong enough so as to reach many sectors at the same time[6]. Recall that we are considering two broad groups of sectors, manufacture and agriculture, not individual industries. There are numerous regional shocks (including political decisions), others than technological shocks that can be influencing the results. In section 5 we outline different approaches to tackle these issues.

Finally consider what happens when countries have access to intermediate inputs in foreign markets. From a theoretical perspective, the fact that some country specializes in the production of certain goods may stimulate nearby countries to specialize in the production of inputs for those goods, as the production of the final good induces a derived demand of inputs. In this context, a potential disadvantage of the classification at the 2-digit level is that one sector that produces a final good and another that produces an input of that good may be included in the same industry although they actually belong to two different industries. if the producer of the final good and the intermediate input are in different countries, the estimates may be affected

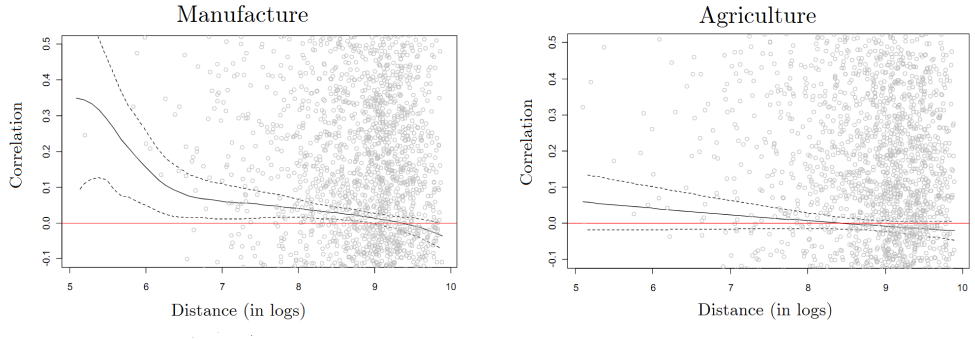
[6] The same effect could be reached by having many sector specific shocks of the same sort.

by the level of aggregation. In section 5 we address this issue.

Before to the second stage of our procedure we would like to motivate it by approaching the problem intuitively. Note that for each country c we have a vector of residuals containing the shocks for each sector, $\varepsilon_s^c = [\varepsilon_1^c, ..., \varepsilon_{61}^c]$. Therefore for each country-pair $\{a, b\}$, we can compute the correlation of the vector of residuals, i.e. $corr(\varepsilon_s^a; \varepsilon_s^b)$, and then estimate the relationship between this correlation and the distance between the two countries.

The existence of agglomeration effects would imply that the first-stage residuals of two given countries will be more correlated the smaller is the bilateral distance. To inspect such relationship, we estimated a non-parametric regression of $corr(\varepsilon_s^a; \varepsilon_s^b)$ on bilateral distance $d(a; b)$[7]. Where the distance between two given countries a and b is defined as the distance in km. between their capitals. Results are displayed below:

Figure 2: Similarity of trade structures and distance



Universidad de San Andrés

Results confirm our hypothesis, the correlation between countries' trade structures after controlling for factor endowments show an inverse relationship with distance, suggesting the presence of a regional spatial dependence. Also, the intensity of these spillovers are higher on more technologically advanced industries. The estimated expected correlation for manufacturing sectors is significantly different from zero up to a log distance of 9.1 (9000 km. approximately). This effect seems not to be present on agriculture sectors.

Next, we provide an additional quantitative measure of the differences between manufacturing and agricultural sectors' estimated correlations, which complements the information conveyed by the figures. For both manufacturing and agricultural sectors, Table 1 displays the differences in median correlations among various relevant log distance intervals:

[7] Note that in order to implement these kind of procedures several decisions need to be made regarding the type of kernel, bandwidth selection, and so on. We opted for a local linear regression with a Second-Order Epanechnikov kernel since it is one of the most common specifications, it has very well known properties in dealing with boundaries, and because of its superior finite sample properties. A fixed bandwidth was endogenously selected by Least Squares Cross-Validation.

Table 1: Median correlation

Log Distance	5-6 (148-403)	5-7 (148-1097)	7-9 (1097-8103)	9-10 (8103-22026)
Manufacture	0.535	0.200	0.045	-0.026
Agriculture	0.172	0.082	0.008	0.013

Note: Table shows median correlations for each category and several distance intervals. Values in parentheses are kilometers.

The results reported in the table above are in line with the results displayed in figure 2. Median correlation ("representative correlation") is higher in manufacturing sectors and decays with distance.

3.2 Second Step: Measuring spatial dependence

So far we have dealt with documenting, intuitively, the existence of regional dependence. In what follows we are going to specify the second stage procedure from which all the results will be derived. The goal is then to conduct a more elaborated analysis of the spatial dependence that may arise among countries due to the presence of technological spillovers. In order to do that we will impose a particular spatial structure on the unexplained component of trade, i.e. ε_s^c. Consider the following model:

$$\frac{N_s^c}{GDP^c} = \sum_{f=1}^{F} \alpha_{1,f} \left(\ln(\tfrac{V_f^c}{V_o^c})\right) + \alpha_{2,f} \left(\ln(\tfrac{V_f^c}{V_o^c})\right)^2 + \alpha_{3,f} \left(\ln(\tfrac{V_f^c}{V_o^c})\right)^3 + \varepsilon_s^c \quad (3)$$

$$\varepsilon_s^c = \lambda W \varepsilon_s^c + u_s^c \quad (4)$$

Where W is an $n \times n$ spatial weighting matrix of known constants, λ is the spatial autoregressive parameter, ε_s^c is the vector of regression disturbances, $g_{fs}(.)$ is defined as before, and u_s^c is a vector of innovations with mean zero and variance σ^2. This type of econometric model is known as the spatial error model. Cross sectional spatial regression models are usually specified in such a way that they incorporate interdependence between observations. One form of spatial dependence arises when the disturbance term is presumed to be spatially autoregressive, as in this case.

Note that the matrix W will weight neighboring observations so that We_s^c will become a spatially lagged vector of disturbances with a typical element of the form $\sum_{j=1}^{n} W_{ij} e_j^*$, where $e^* = e_s^c$. As W will be row normalized the spatially lagged vector of disturbances is a simple average of neighboring values. The scalar parameter λ describes the strength of spatial dependence, reflecting the average level of dependence. It captures the degree of spatial autocorrelation in measurement errors or omitted variables, which in our case can be interpreted as unobservable spillovers.

The weighting matrix (W) is an exogenous parameter and plays a crucial role in the specification of the type and degree of the spatial correlation. We constructed this matrix in a very intuitive way using the geographic distance among countries, specifically:

$$W_{i,j} = \max\{(\overline{D} - Distance_{i,j}), 0\} \quad (5)$$

Where subscripts i and j index countries i and j respectively, and $\overline{D}$ represents a threshold distance. In this way the near the countries are the higher the value of $W_{i,j}$, meaning that closer countries receive a higher weight when evaluating spatial dependence (as usual, the matrix W is later row-normalized and diagonal elements are set to zero). Introducing the exogenous threshold $\overline{D}$ gives us the possibility of evaluating spatial correlation "locally" by including only observations or countries within a certain range of distance.

When it comes to the estimation of this sort of models two different procedures are usually employed. The first estimates the parameters using the maximum likelihood estimation method (MLE). By making use of first-order conditions the log-likelihood function can be reduced to a non-linear function of the spatial parameter λ. If we are willing to assume that $f_\varepsilon(\varepsilon|X)$ $\tilde{}N(0, \sigma^2 I)$ then the log-likelihood function can be constructed as follows:

$$l(\lambda, \sigma^2, g_f^s(.)) = -n/2\ln(\sigma^2) + \ln|I - \lambda W| - 1/2\sigma^2\{(I - \lambda W)[Z]\}^T\{(I - \lambda W)[Z]\} \quad (7)$$

Where $g_f^s(.)$ is defined as in section 3.1 and $Z = \frac{N_s^c}{GDP^c} - \sum_{f=1}^{F} g_f^s\left(\ln(\frac{V_f^c}{V_O^c})\right)$ .This log-likelihood function can be maximized to obtain pseudo ML estimates of the parameters. Estimation steps and expressions for the asymptotic variance can be found in LeSage (1999).See Anselin (1988) for a detailed description of MLE of spatial models.

An alternative way of proceeding is to approach the problem by matching moments as in Kelejian and Prucha (2010). In this paper authors introduce a class of Generalized Methods of Moments (GMM) estimator for the autoregressive parameter of a spatially autoregressive disturbance process that allows for heteroskedastic innovations.

The following three equation system, which is based on moment conditions, can be solved to obtain by setting it up as a NLLS minimization problem.

$$2\lambda n^{-1}\textstyle\sum_i \varepsilon_i W\varepsilon_i - \lambda^2 n^{-1}\sum_i (W\varepsilon_i)^2 + n^{-1}\sum_i u_i^2 = n^{-1}\sum_i \varepsilon_i$$

$$2\lambda n^{-1}\textstyle\sum_i W\varepsilon_i W^2\varepsilon_i - \lambda^2 n^{-1}\sum_i (W^2\varepsilon_i)^2 + n^{-1}\sum_i (Wu_i)^2 = n^{-1}\sum_i (W\varepsilon_i)^2$$

$$2\lambda n^{-1}\textstyle\sum_i [W\varepsilon_i W^2\varepsilon_i + (W\varepsilon_i)^2] - \lambda^2 n^{-1}\sum_i W\varepsilon_i (W^2\varepsilon_i)^2 + n^{-1}\sum_i u_i (Wu_i)^2 = n^{-1}\sum_i \varepsilon_i (M\varepsilon_i)$$

Where in this case i indexes $i = \varepsilon_{s=1}^{c=1}, \varepsilon_{s=1}^{c=2}, ..., \varepsilon_{s=2}^{c=1}, \varepsilon_{s=2}^{c=2}, ..., \varepsilon_{s=S_{M,A},}^{c=80}$. S_M and S_A represent the number of sectors in manufacture and agriculture categories respectively.

We opted to use the GMM procedure. It is well-know that MLE spatial models have some problems when the number of observations increases as it becomes more difficult to compute the log determinant of the Jacobian. Additionally, the GMM based procedure needs less assumptions as it does not require assuming normality of u_s^c. For a detailed comparison between MLE versus GMM procedures in spatial econometrics see Walde, Larch, and Tappeiner (2008), Egger, Larch, Pfaffermayr, and Walde (2009), and Bivand (2010)

4 Results:

Recall that one of the main goals of this paper is documenting the existence of technological spillovers and to what extent they are global or local. The introduction of the additional parameter $\overline{D}$, presented in the previous section, will allows us to test how the strength of spatial dependence changes as distance increases.

As previously mentioned the weight matrix (W) is an exogenous parameter and plays a crucial role in the specification of the type and degree of the spatial correlation. We constructed this matrix in a very intuitive way using the geographic distance among countries, specifically:

$$W_{i,j} = \max\{(\overline{D} - Distance_{i,j}), 0\} \quad (8)$$

Where subscripts i and j index countries i and j respectively, and $\overline{D}$ represents a threshold distance. In this way the near the countries are the higher the value of $W_{i,j}$, meaning that closer countries receive a higher weight when evaluating spatial dependence (as usual, the matrix W is later row-normalized and diagonal elements are set to zero). Introducing the exogenous threshold $\overline{D}$ gives us the possibility of evaluating spatial correlation "locally" by including only observations or countries within a certain range of distance. . Table 2 below is a summary of the results for different values of $\overline{D}$.

Table 2: Summary of the results

Threshold Distance (Km.)	Lambda (Manufacture)	Lambda (Agriculture)	Difference
250	0.2856***	0.095***	0.189
500	0.2378***	-0.086***	0.324
750	0.1267***	0.099***	0.026
1000	0.0689***	0.068***	0
1250	0.0817***	0.044*	0.037
1500	0.0818***	0.008	0.073
1750	0.0634**	-0.004	0.068
2000	0.0444	-0.016	0.060
2250	0.0271	-0.026	0.053
2500	0.0077	-0.053	0.061
2750	0.0014	-0.042	0.043
3000	-0.0020	-0.036	0.034
4000	-0.0089	-0.037	0.028
5000	-0.0034	-0.039	0.036
6000	-0.0014	-0.038	0.036
7000	-0.0017	-0.036	0.034
8000	-0.0014	-0.035	0.033
9000	-0.0021	-0.029	0.027
10000	-0.0038	-0.025	0.021

*** significant at 1%; ** significant at 5%; * significant at 10%.

Results confirm the findings of the previous section. Once we netted out the effect of factor endowments in countries' trade structures, ε_s^c is spatially correlated. Moreover, this correlation

seems to be regional in the sense that the effect is stronger for neighboring countries and is dying out with distance. Results show that in manufacture there is a positive spatial dependence up to a distance of 1750 km. while in agriculture only up to 1000 km[8]. Additionally, the agglomeration effect is always greater in manufacture than in agriculture. Which supports the hypothesis that technological spillovers may be a plausible explanation. [9].

There are two issues worth pointing out about the interpretation of the results. First note that for a given value of the threshold $\overline{D}$, the coefficient λ summarizes the average spatial correlation of an observation with the weighted average of neighboring observations. Meaning that it represents an average effect.

Second, as the threshold increases more neighboring observations fall within the range of analysis and the average is therefore taken over more observations. which means that any increment of the threshold $\overline{D}$ followed by λ falling by, let's say 10%, does not represent a marginal effect of distance over the spatial correlation parameter.

Third, results are not invariant to the notion of distance used. The matrix W is intended to capture the influence of neighboring observations by giving weights to them depending on the distance between countries. In our case, weights are decreasing linearly as distance increases and are homogeneous across sectors because they depend only on country to country distance. however, we could have arguments to prefer a weighting scheme that decreases exponentially with distance or that incorporates proximity between industries (for instance looking at transactions on Input-Output tables). In that case, results will differ without any obvious reason to prefer one specification over the other.

Therefore we found that after netting out the effect of factor endowments on countries' trade structures, the remaining unexplained component (ε_s^c), or technological component, is spatially correlated. The fact that the correlation is higher in more technologically advanced industries suggests that the effect may be caused by technological spillovers. Additionally, this correlation dies out with distance indicating that the effect is regional.

Universidad de SanAndrés

[8] A negative spatial autocorrelation indicates that neighboring values are dissimilar or, equivalently, that similar values are dispersed.

[9] In the next section we will show results using an alternative categorization of sectors according to their technological intensity.

5 Robustness Checks:

In the previous section we reported a summary of our estimates for the baseline case. While specifying our model we set many of the parameters arbitrarily. This section is an attempt to test our results against different possible specifications. In what follows we will test the robustness of our findings to the change of the weighting matrix W, the definition of what we considered as a representation of countries' trade patterns, the level of aggregation of our data, the specification of the model in the first stage of the econometric procedure, and the classification of sectors according to their use of technology.

5.1 Alternative Definition of W :

In the baseline case we have defined neighboring countries as those which fall within a certain range of distance. As we mentioned before, the weighting matrix is an exogenous parameter and results are usually very sensitive to changes on it. In what follows we will consider an alternative choice of W. Note that in our case we are dealing with an irregular grid, in the sense that the number of neighbors considered will differ depending on the geographical conditions each country is surrounded by. Typically, European countries will have more neighbors falling within any range of distance specified by a given threshold $\overline{D}$ than Latin American countries. Therefore one might be concerned that any measure based on contiguity will yield observations with an uneven number or even no connections, casting doubts about the validity of the results.

In what follows we will evaluate the presence of regional technological spillovers by weighting the $k-$Nearest observations. Choosing the $k-$Nearest neighbors ensures a constant number of neighbors for each country, with no country having zero observations in any range. Additionally, the weighting matrix W will be row-normalized using weights inversely proportional to distance. Results are summarized on the table below:

Table 3: Alternative Weighting Matrix

Nearest Neighbors (Weighted)	Lambda (Manufacture)	Lambda (Agriculture)	Difference
1	0.131***	0.074***	0.057
2	0.107***	0.015	0.091
3	0.092***	-0.009	0.101
4	0.090***	-0.013	0.104
5	0.098***	-0.015	0.113
6	0.090***	-0.013	0.104
7	0.080***	-0.019	0.100
8	0.076***	-0.018	0.095
9	0.073***	-0.019	0.093
10	0.070***	-0.022	0.093
11	0.069***	-0.026	0.096
12	0.068***	-0.024	0.093
13	0.065***	-0.024	0.089
14	0.0630**	-0.023	0.086
15	0.0620**	-0.023	0.085

*** significant at 1%; ** significant at 5%; * significant at 10%.

Results are consistent with previous findings. There exist a component of countries' trade patterns, left unexplained by factor endowments, which is spatially correlated and decreasing in distance (number of neighbors considered). Also, results show that in manufacture there is a much stronger spatial dependence than in agriculture.

5.2 Alternative Characterization of Trade Patterns:

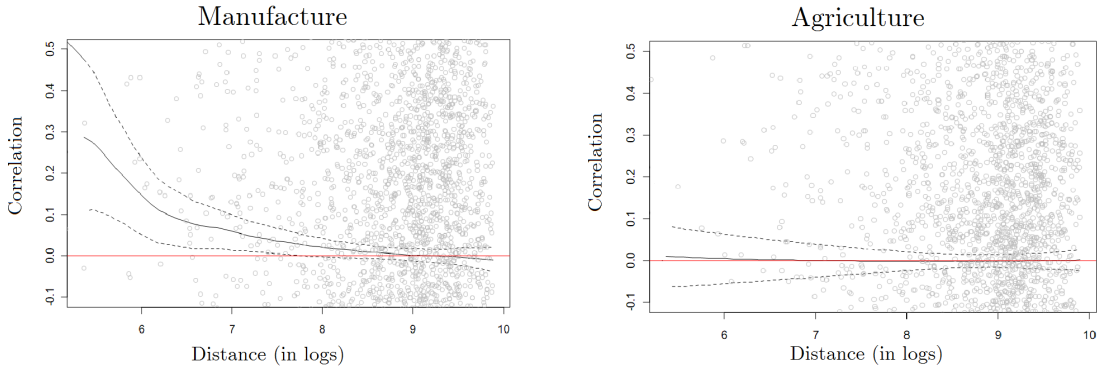
Recall that for the baseline case we characterized countries' trade patterns using sectorial net exports over GDP($\frac{N_s^c}{GDP^c}$) as our dependent variable. There exist reasons to believe that a variable containing only sectorial exports could be also appropriate to capture/measure countries' trade patterns. Changing the dependent variable will change the results from the very beginning as it will affect our first stage results. Consider the modified version of the model:

$$\frac{E_s^c}{GDP^c} = \sum_{f=1}^{F} g_{fs} \left(\ln(\frac{V_f^c}{V_f^o}) \right) + \varepsilon_s^c \quad (9)$$

$$\varepsilon_s^c = \lambda W e_s^c + u_s^c \quad (10)$$

Where E_s^c takes positive values if country c registered exports in sector s and zero otherwise. Let the rest of the procedure including methodology and exogenous parameters like W remain fixed. Now lets calculate the correlation of the vector of residuals, i.e. $corr(\varepsilon_s^a; \varepsilon_s^b)$, as we did in section 3 to later estimate the relationship between this correlation and the distance between the two countries. A non-parametric regression of $corr(\varepsilon_s^a; \varepsilon_s^b)$ on bilateral distance $d(a; b)$ is depicted below[10]:

Figure 3: Similarity of trade structures and distance



Results look very similar to those obtained with net exports as dependent variable. The correlation between countries' trade structures in manufacturing sectors after controlling for

[10] In order to compare as closely as possible this results with those obtained in the baseline case we used the same specification for kernel selection, keeping also the bandwidth fixed.

factor endowments decays with distance, suggesting the presence of regional spillovers. At first sight it is clear that in this case the effect dies out earlier, around a log distance of 8 while in the baseline case it was statistically different from zero up to a log distance of 9.1. Again, the effect seems not to be present on agriculture sectors.

Now lets consider the spatial econometric approach. We will proceed by showing the new estimates that are analogous to those summarized on table 2.

Table 4: Exports as Dependent Variable

Threshold Distance (Km.)	Lambda (Manufacture)	Lambda (Agriculture)	Difference
250	0.261***	0.147***	0.114
500	0.136***	-0.047**	0.183
750	0.093***	0.100***	-0.006
1000	0.082***	0.068***	0.013
1250	0.066***	0.049**	0.017
1500	0.055**	0.020	0.034
1750	0.035	0.003	0.032
2000	0.016	-0.010	0.026
2250	-0.001	-0.014	0.012
2500	-0.028	-0.033	0.004
2750	-0.034	-0.032	-0.002
3000	-0.035	-0.030	-0.005
4000	-0.033	-0.035	0.002
5000	-0.025	-0.036	0.010
6000	-0.022	-0.035	0.013
7000	-0.019	-0.032	0.012
8000	-0.017	-0.029	0.012
9000	-0.017	-0.024	0.007
10000	-0.018	-0.021	0.002

*** significant at 1%; ** significant at 5%; * significant at 10%.

Universidad de San Andrés

Results show similar patterns to those obtained before. Manufacture sectors tend to show a higher degree of spatial dependence and this effect is declining with distance. However, note that in this case the effect on agriculture sectors is higher than before.

The table 5 below summarizes the coefficient estimates for the case in which $k-$ Nearest neighbors is used. Note that results are consistent with those obtained on table 3 but with the difference that the effect seems to be more localized than before, in line with the findings of the intuitive approach of this subsection.

Table 5: Exports as Dependent Variable (NN)

Nearest Neighbors (Weighted)	Lambda (Manufacture)	Lambda (Agriculture)	Difference
1	0.107***	0.070***	0.037
2	0.072***	0.026	0.046
3	0.066***	0.008	0.057
4	0.059**	0.009	0.049
5	0.062**	0.017	0.044
6	0.062**	0.011	0.050
7	0.052*	0.000	0.051
8	0.046	0.002	0.044
9	0.040	-0.000	0.041
10	0.040	-0.005	0.045

*** significant at 1%; ** significant at 5%; * significant at 10%.

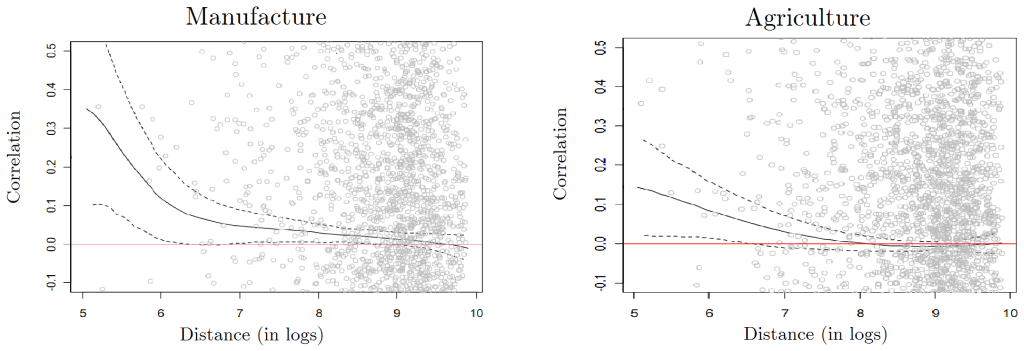
5.3 Agglomeration Effects:

An important exercise to assess the robustness of the effect of technological spillovers is to perform the previous analysis with trade data at the 3-digit level of aggregation. The main purpose of this exercise is to provide a direct way to test the hypothesis that the similarity of trade patterns among nearby countries is in effect due to technological spillovers and not to agglomeration effects. From a theoretical perspective, the fact that some country specializes in the production of certain goods may stimulate nearby countries to specialize in the production of inputs for those goods as the production of the final good induces a derived demand of inputs. In this context, a potential disadvantage of the classification at the 2-digit level is that one sector that produces a final good and another one that produces an input of that good may be included in the same industry, although they are actually two distinct industries.

A more refined classification, such as that at the 3-digit level, could enable us to tackle this problem. If the influence of agglomeration effects on the similarity of trade patterns were actually a relevant factor, we should observe that the difference in mean correlations for manufacturing sectors at the 3-digit level is significantly smaller than that at the2-digit level.

Consider again the following non-parametric regression of the correlation of trade patterns on bilateral distance:

Figure 4: Similarity of trade structures and distance



Results show that there is no significant difference using more disaggregated data. For the spatial approach conclusions are almost identical. The table below summarizes the estimates at 3-Digit level.

Table 6: Results at 3-Digit Level

Threshold Distance (Km.)	Lambda (Manufacture)	Lambda (Agriculture)	Difference
250	0.271***	0.086***	0.185
500	0.243***	0.071***	0.172
750	0.134***	0.080***	0.053
1000	0.061***	0.051***	0.009
1250	0.061***	0.041*	0.019
1500	0.061***	0.009	0.052
1750	0.063**	-0.005	0.068
2000	0.052**	-0.011	0.063
2250	0.046**	-0.008	0.054
2500	0.016	-0.014	0.029
2750	0.001	-0.021	0.022
3000	-0.002	-0.019	0.017
4000	-0.007	-0.014	0.007
5000	-0.005	-0.010	0.005
6000	-0.001	-0.015	0.014
7000	-0.001	-0.012	0.011
8000	-0.001	-0.012	0.011
9000	-0.001	-0.013	0.012
10000	-0.001	-0.014	0.013

*** significant at 1%; ** significant at 5%; * significant at 10%.

Universidad de San Andrés

5.4 An alternative to the first stage estimator:

In the first stage we obtained our residuals by estimating an OLS regression including powers and interactions of the variables. By doing this we were able to allow for some flexibility and at the same time keep the number of coefficients within a certain range. The econometric procedure used to estimate the autorregresive parameter described in Kelejian and Prucha (2010). only requires that residuals are estimated consistently, disregarding the estimator used.

We propose an alternative approach to the problem of netting out the effect of factor endowments on countries' trade structures by using a Generalized Additive Model (GAM) on the first stage. This procedure is particularly handy for dealing with potential outliers as it allows for nonlinear relationships between factor endowments and trade patterns. The only restriction is that it imposes additivity of the regressors. The alternative model is of the form:

$$\frac{N_s^c}{GDP^c} = \sum_{f=1}^{F} g_{fs}\left(\ln(\frac{V_f^c}{V_f^o})\right) + \varepsilon_s^c \quad (11)$$

$$\varepsilon_s^c = \lambda W e_s^c + u_s^c \quad (12)$$

where now $g_{fs}(.)$ is any derivable function. GAM require specifying some parameters too. We used splines to approximate $g_{fs}(.)$ letting the degree of the polynomial be chosen endogenously. Also we imposed a Gaussian link function. For a detailed explanation of GAM see Hastie and ibshirani (1990).

Results are consistent with those obtained before and are virtually the same although less precise, meaning that standard error are on average 20% higher.

QUAERERE VERUM
Universidad de
SanAndrés

5.5 Alternative Characterization of Sectors:

We divided sectors into two broad categories, manufacture and agriculture. The main idea behind that arbitrary division has to do with the idea of testing whether technological spillovers are a plausible explanation for the agglomeration effects..

We opted to present an alternative division of sectors. In particular, we will consider the OECD methodology to classify countries' industrial sectors by level of technology as in Hatzichronoglou (1997).

According to this classification there are four categories of technological intensity: High, Medium High, Medium-Low, and Low. The technological intensity reflects the amount of R&D spent compared to the value added of that industry. We end up then with 8 sectors as Hi-Tech, 19 as Medium-High-Tech (which include Hi-Tech sectors), 32 as Low-Tech sector, and 42 as Medium-Low-Tech Sectors (which include Low Tech sectors).

Consider the following results which are the analogous of those reported on table 1, the only difference is on how sectors are grouped.

Table7: Mean Correlation

	5-6 (148-403)	5-7 (148-1097)	7-9 (1097-8103)	9-10(8103-22026)
Hi Tech	0.465	0.208	0.026	-0.028
Med Hi Tech	0.409	0.132	0.015	-0.019
Med Low Tech	0.101	0.096	0.009	-0.009
Low Tech	0.060	0.059	0.021	-0.010

Note: Table shows mean correlations for each category and several distance intervals. Values in parentheses are kilometers.

Table8: Median Correlation

	5-6 (148-403)	5-7 (148-1097)	7-9 (1097-8103)	9-10(8103-22026)
Hi Tech	0.818	0.424	0.074	-0.039
Med Hi Tech	0.595	0.203	0.049	-0.054
Med Low Tech	0.160	0.129	0.013	-0.007
Low Tech	0.152	0.081	0.019	-0.018

Note: Table shows median correlations for each category and several distance intervals. Values in parentheses are kilometers.

Results show, as in section 3, that the correlation between countries' trade structures after controlling for factor endowments decays with distance. Additionally, we see that the more technologically advanced the sectors are the higher the effect of technological spillovers.

QUAERERE VERUM
Universidad de
SanAndrés

6 Conclusion:

In this paper we documented the existence agglomeration effects on countries' trade structures after the effect of factor endowments was netted out. Agglomeration effects would induce similarity of trade patterns among nearby countries beyond what can be explained by factor endowments. We studied the observed pattern of agglomeration and postulated a process through which it may have emerged. We inspected these effects taking into account the intensity in the use of technology by industries to later argue that agglomeration may be caused by technological spillovers.

This findings are consistent with the literature of technology diffusion. Previous studies have mainly focused their attention on two channels through which technology diffusion can take place, trade in intermediate goods and FDI flows. Consider trade as a diffuser, if trade in intermediate goods disseminates knowledge, nearby countries could take advantage of technological progress somewhere else by importing embodied technology. Hence, progress made in one country could improve productivity of industries located upstream on the productive chain, generating a more similar pattern of comparative advantages among countries that adopt this technology. Additionally, technology diffusion can take place trough FDI flows.

Our findings suggest that these spillovers are regional, in the sense that they decay with distance. Results are intuitive as they don't contradict previous findings. If spillovers take place trough trade we should expect that this diffusion is geographically localized because of the widely studied fact that trade falls with distance. If on the other hand technology diffusion takes place trough FDI flows, the fact that technology is not completely codifiable will make transmission of knowledge costly and likely to be affected by geography.

By dividing sectors according to their intensity on the use of technology we showed that spillovers are higher on more technologically advanced industries.

The natural step forward would be to extend our analysis including a temporal dimension. A spatial panel would allow us to include regional fixed effects to control for regional demand shocks and time variant effects.

Universidad de SanAndrés

7 Appendix:

Recall that in section 2 we briefly described which variables were to form part of the analysis without giving much details about how they were constructed. In this section we will describe how we built those variables. Most of the information used was taken directly from the sources described in section 2 and didn't need any additional treatment. Those variables are capital, labour, skilled labour, crops, bilateral trade, and geographic distance. Data on minerals and land endowments were constructed as follows:

Minerals endowment of a country was approximated as the sum of the production for 1996 of the following minerals: Antimony, Asbestos, Barite, Bauxite, Beryllium, Bismuth, Boron, Bromine, Chromium, Cobalt, Copper, Diamond, Fluorspar, Gold, Iodine, Iron Ore, Lead, Manganese, Mercury, Molybendum, Nickel, Niobium, Platinum, Palladium, Potash, Rhenium, Silver, Tantalum, Tin, Tungsten, and Zinc.

Regarding land endowments recall that we performed our econometric analysis using two variables, Crops and Land. The former was taken directly from from data while the second was created comprising information about land resources. This variable grouped data corresponding to measures of countries' availability of arable land, pastures, and forest. The idea behind this decision was to maintain the number of regressor within a certain range and avoid overfitting the data.

We did that by performing a principal component analysis on all the possible combinations of three vectors. Results showed that by comprising information of pasture, forest, and arable land we are able to explain a 76.5% of the variance of the three vectors, losing only less than 25 % of the information contained.

QUAERERE VERUM
Universidad de
SanAndrés

8 References:

- Anselin, Luc. Spatial econometrics: methods and models. Vol. 4. Kluwer Academic Pub, 1988.

- Albornoz, Facundo, and Marco Ercolani. "Learning by exporting: do firm characteristics matter? Evidence from Argentinian panel data." Evidence from Argentinian Panel Data (September 2007) (2007).

- Barro, Robert J., and Jong-Wha Lee. "International data on educational attainment: updates and implications." Oxford Economic Papers 53.3 (2001): 541-563.

- Bivand, Roger. "Comparing estimation methods for spatial econometrics techniques using R." NHH Dept. of Economics Discussion Paper 26 (2010).

- Branstetter, Lee G. "Are knowledge spillovers international or intranational in scope?: Microeconometric evidence from the US and Japan." Journal of International Economics 53.1 (2001): 53-79.

- Clerides, Sofronis K., Saul Lach, and James R. Tybout. "Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco." The Quarterly Journal of Economics 113.3 (1998): 903-947.

- Coe, David T., and Elhanan Helpman. "International r&d spillovers." European Economic Review 39.5 (1995): 859-887.

- Eaton, Jonathan, and Samuel Kortum. "Technology, geography, and trade." Econometrica 70.5 (2002): 1741-1779.

- Egger, Peter, et al. "Small sample properties of maximum likelihood versus generalized method of moments based tests for spatially autocorrelated errors." Regional Science and Urban Economics 39.6 (2009): 670-678.

- Feenstra, Robert C., et al. World trade flows: 1962-2000. No. w11040. National Bureau of Economic Research, 2005.

- Griliches, Zvi. "R&D, Patents, and Productivity". U. Chicago for NBER.

- Griliches, Zvi, and Jacques Mairesse. Production functions: the search for identification. No. w5067. National Bureau of Economic Research, 1995.

- Grossman, Gene M., and Elhanan Helpman. "Trade, knowledge spillovers, and growth." European Economic Review 35.2 (1991): 517-526.

- Hatzichronoglou, Thomas. Revision of the high-technology sector and product classification. No. 1997/2. OECD Publishing, 1997.

- Henderson, Rebecca, Adam B. Jaffe, and Manuel Trajtenberg. "Universities as a source of commercial technology: A detailed analysis of university patenting, 1965-1988." Review of Economics and Statistics 80.1 (1998): 119-127.

- Katayama, Haijime, Shihua Lu, and James Tybout. Why plant-level productivity studies are often misleading, and an alternative approach to interference. No. w9617. National Bureau of Economic Research, 2003.

QUAERERE VERUM
Universidad de San Andrés

- Kelejian, Harry H., and Ingmar R. Prucha. "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances." Journal of Econometrics 157.1 (2010): 53-67.

- Kelejian, Harry H., and Ingmar R. Prucha. "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances." The Journal of Real Estate Finance and Economics 17.1 (1998): 99-121.

- Keller, Wolfgang. Geographic localization of international technology diffusion. No. w7509. National Bureau of Economic Research, 2000.

- Keller, Wolfgang. International technology diffusion. No. w8573. National Bureau of Economic Research, 2001.

- Keller, Wolfgang, and Stephen R. Yeaple. The gravity of knowledge. No. w15509. National Bureau of Economic Research, 2009.

- Peri, Giovanni. Knowledge Flows and Knowledge Externalities. No. 765. CESifo Group Munich, 2002.

- Rivera-Batiz, Luis A., and Paul M. Romer. "Economic integration and endogenous growth." The Quarterly Journal of Economics 106.2 (1991): 531-555.

- Savvides, Andreas, and Marios Zachariadis. "International technology diffusion and the growth of TFP in the manufacturing sector of developing economies." Review of Development Economics 9.4 (2005): 482-501.

- Scherer, Frederic. "Using Linked Patent and R&D Data to Measure InterindustryTechnology Flows." R & D, Patents, and Productivity. University of Chicago Press, 1984. 417-464.

- Walde, Janette, Mario Larch, and Gottfried Tappeiner. "Performance contest between MLE and GMM for huge spatial autoregressive models." Journal of Statistical Computation and Simulation 78.2 (2008): 151-166.

- Wood, Adrian, and Jörg Mayer. "Africa's export structure in a comparative perspective." Cambridge Journal of Economics 25.3 (2001): 369-394.

Universidad de San Andrés