The Spatial Diffusion of Technology*

Diego Comin
Harvard University

Mikhail Dmitriev
Boston College

Esteban Rossi-Hansberg
Princeton University

May 17, 2012

Abstract

We study the patterns of technology diffusion across countries and time. We find significant evidence that technology diffuses slower to locations that are farther away from adoption leaders. This effect is stronger across rich countries and also when measuring distance along the south-north dimension. We propose a simple theory of human interactions that accounts for these empirical findings. The theory suggests that the effect of distance should vanish over time, a hypothesis that we confirm in the data. We then structurally estimate, for each technology, the two parameters of the model: the spatial decay in interactions and the frequency of interactions. The parameter governing the frequency of interactions is larger for newer and network-based technologies. Overall, the evidence we uncover points to a significant role of geography in determining the level of technology adoption across countries.

*We thank Walker Hanlon, Stefania Garetto, and seminar participants at various institutions for useful comments.

1 Introduction

Technology disparities are critical to explain cross-country differences in per capita income.[1] Despite being non-rival in nature,[2] technology diffuses slowly both across and within countries resulting in significant lags between the time of invention and the time when a technology is initially used in a country. Even when a technology has arrived in a country, it takes years and even decades before it has diffused to the point of having a significant impact on productivity. These observations have led economists to study why does technology diffuse slowly, and what explains cross-country differences in its speed of diffusion.

Existing empirical studies on technology adoption have treated adoption units (e.g. countries, cities, or firms) as independent.[3] Consequently, they have tried to link a country's technology adoption patterns to the country's characteristics (e.g. human capital, institutions, policies, adoption history, etc.).[4] This empirical approach to the drivers of technology adoption ignores the possibility of cross-country interactions in the adoption process. This assumption might be restrictive. Adopting a technology requires acquiring knowledge[5] which often comes from interactions with other agents. The frequency and success of these interactions is likely to be shaped by geography. Technological knowledge is likely to be more easily transmitted between agents in countries that are close than between agents located far apart. Similarly, the payoff to adopting a given technology (e.g. railways) may be affected by the adoption experience of neighboring countries. These mechanisms may generate correlated adoption patterns across nearby countries. To explore the empirical importance of these mechanisms, in this paper we explore the presence of cross-country interactions in technology adoption that are mediated by geographical distance. In particular, we study empirically the diffusion of technology across time and space.

A clear impediment to collecting evidence on the presence of geographic interactions in technology adoption is the lack of comprehensive datasets that directly document the diffusion of specific technologies across countries. In this paper, we study the diffusion over time and space of 20 major technologies in 161 countries over the last 140 years using data from the CHAT dataset (Comin and Hobijn 2004, 2010). Our interest lies in exploring the presence of cross-country correlations in technology adoption that are mediated by geographical distance. To this end, we measure how far a country is from the high-density points in the distribution of technology adoption in the other countries. We denote this measure the spatial distance from other country's technology or, to abbreviate, the spatial distance from technology (SDT).[6] After controlling for country and time fixed

[1] See Klenow and Rodriguez-Clare (1997), Hall and Jones (1999), Comin and Hobijn (2010) and Comin and Mestieri (2010) among many others.

[2] See, for example, Romer (1990). This property steams from the fact that once invented, the use of a technology by one producer does not preclude others from using it.

[3] See, for example, Griliches (1957) and Mansfield (1961).

[4] See Comin and Hobijn (2004) and Comin, Easterly and Gong (2010).

[5] Potential users of the technology may first learn about its existence and properties, then, they need to learn how to use it, and finally they need to figure out how to apply it to an existing production process or to new ones.

[6] Formally, the SDT of a country is defined as the scalar product of the (log) adoption levels in the rest of the countries and the distance to each of these countries.

effects, a negative correlation between SDT and adoption implies that countries that are further away from those where the technology diffuses faster tend to experience a slower diffusion of the technology.[7]

In Section 2, we present a battery of empirical findings. We estimate a strong and significant negative partial correlation between SDT and a country's adoption, after controlling for per capita income and technology-specific country and time fixed effects. We also explore whether the spatial correlations in adoption are purely driven by the spatial correlation in income or in other variables strongly correlated with income. To this end, we control for a measure of the spatial distance from other countries per capita income (SDI), constructed in a way parallel to SDT. We find that the sign and magnitude of SDT in our diffusion equations is unaffected by the inclusion of SDI. Of course, the correlated adoption patterns we document could result from the spatial correlation of other drivers of technology adoption independent of income.[8] It is hard to imagine what these drivers could be but, in any case, our methodology cannot rule out this possibility. Our more modest goal is to describe for the first time the spatial patterns of technology diffusion across countries and to make the case that these patterns can be parsimoniously rationalized by models of spatial technology diffusion.

To this end, we first explore the robustness of the association between SDT and technology adoption to various specifications and samples. We find robust and significant effects of SDT on technology diffusion across sectors and income levels as well as when we use distinct country samples to compute SDT. Similarly, we take on Jared Diamond's hypothesis that technologies diffuse along latitudes.[9] We construct measures of SDT based on latitude and longitude distances, and run a horse race between the two. We find that, consistent with Diamond (1997), SDT across latitude distances has a stronger association with technology adoption than SDT across longitude distances. This finding is remarkable provided that our sample does not have any technology where climatic reasons might suggest that distance across latitudes is a larger impediment than distance across longitudes for their diffusion. These exercises, we believe, provide a richer characterization of the spatial patterns of technology diffusion.

To further explore the mechanisms that drive these spatial diffusion patterns, we develop a simple model which borrows from the literature on external effects and contagion[10] as well as recent papers that have emphasized the importance of individual knowledge exchanges for growth (e.g. Eaton and Kortum, 1999, Lucas, 2009, and Lucas and Moll, 2011). In particular, our stylized model is based on two key assumptions. First, technology is diffused by interacting with adopters. Second, interactions are random and more likely between agents located nearby. The model is

[7] The construction of SDT might raise concerns of endogeneity, since adoption is a function of all other countries' adoption rates. In the next section we argue that these concerns are minor if we have many countries and so the contribution of any one country to the distribution of adoption is minor. In Appendix B we also argue that even with a small sample of countries the upper bound of the bias generated by endogeneity is rather small and certainly irrelevant for our substantive results.

[8] This possibility is related to the reflection problem emphasized by Maski (1993).

[9] See "Guns, Germs and Steel," Diamond (1997).

[10] See, for example, Fujita and Thisse (2002) and the survey in Duranton and Puga (2004).

consistent with our econometric findings. It also yields new testable predictions. In particular, it predicts that the geographic interactions in adoption as measured by the effect of SDT on technology adoption should diminish as technology diffuses. Going back to the data, we document that this prediction holds in a large majority of the technologies and samples studied. Furthermore, we provide a parsimonious statistical characterization that fits well the time variation in the effect of SDT on technology adoption.

The time patterns of geographic interactions in adoption are so stark that we use them to estimate structurally the two parameters that characterize the model. Using a simulated method of moments (SMM) estimator, we show that our simple model can generate time-varying interaction effects that resemble the data for fourteen out of the twenty technologies in the sample. Our estimates of the structural parameters of the model help us understand better the spatial diffusion process. In particular, they show that the frequency of interactions has been higher for newer than for older technologies. We also explore a three-parameter version of the model with heterogeneity in the initial adoption levels of followers. This version of the model fits well the observed patterns for the time-varying interaction effects for nineteen of the twenty technologies in our sample.

Despite the intuitive appeal of cross-country interactions in technology adoption, the literature has not been able to document its presence and to assess their contribution to the large cross-country differences in technology adoption. Some strands of the literature have explored the presence of knowledge spillovers associated with research and development activities. Broadly speaking, this approach has been pursued in two different ways. One has used patent citations data mostly within rich countries (Jaffe, Trajtenberg, and Henderson, 1993, Almeida, 1996, Thompson and Fox-Kean, 2005). Another has used cross-country data to study the effects of a country's R&D expenditures on other nearby countries' TFP (see Keller ,2004, for a comprehensive survey). However, innovation and adoption are distinct phenomena and it is unclear whether the knowledge and factors relevant to adopt a technology are related to those that matter for innovating. Furthermore, to explain cross-country differences in adoption it seems more appropriate to rely on cross-country spillovers than within-country spillovers. In addition, given the typical length of gestation lags, a positive correlation between Solow residuals and R&D expenditures may just reflect international cyclical co-movement rather than international technology diffusion. A final strand of the literature has studied directly adoption using micro-level data for some simple agricultural technologies such as new crops or high yield seeds (e.g. Foster and Rosenzweig,1995, and Bandeira and Rasul, 2006). These studies have found evidence of spatial correlations in adoption patterns across individuals. Nevertheless, the nature of these studies does not inform us about the applicability of these findings for other technologies (e.g. in other sectors, more complex, or more capital intensive) and the significance of the identified interactions for cross-country differences in adoption.

The rest of the paper is organized as follows. In Section 2, we present our non-structural empirical approach, the data, and the findings from our non-structural exploration. In Section 3 we develop our simple stylized model and analyze it deriving some testable predictions. We contrast them with the data in Section 4. In Section 5, we estimate the model structurally. Section 6 concludes.

2 Empirical exploration

In this section, we make a first pass in our investigation of the role of geographic interactions on technology diffusion. We start by imposing minimal structure to uncover general and robust patterns in the data.

2.1 Methodology

Our empirical methodology is based on the following econometric specification:

$$x^j_{ct} = \beta^j_{1c} I^j_c + \beta^j_{2t} I^j_t + \beta^j_3 y_{ct} + \beta^j_4 x^j_{-ct} + \beta^j_5 y_{-ct} + \epsilon^j_{ct} \tag{1}$$

The dependent variable, x^j_{ct}, is the level of adoption of a technology j in country c in year t. Technology adoption measures come from the cross-country historical adoption technology (CHAT) dataset (Comin and Hobijn, 2004, 2009, and 2010). To maximize the country representation of the sample, we focus on 20 major technologies, listed in Table 1.[11] Broadly speaking, the technologies studied belong to three sectors, transportation, communication and industry. They cover, in an unbalanced way, technology diffusion in 161 countries going back until 1825. For each technology measure, (e.g. tons*kilometer transported by rail per capita), we take logarithms and add a technology-specific constant that ensures that x^j_{ct} is always positive.[12] Adding a constant is inconsequential for the dependent variable, but it is relevant for the interpretation of the SDT term that we introduce below.

Given the time series length for some of our technologies, they may eventually become dominated by newer technologies. Because our interest is on the phase in which technologies diffuse, we censor the time series to eliminate the obsolescence phase. We achieve this by censoring the data once the adoption per capita of the leader (i.e. the country with higher adoption per capita) starts to decline.

Income affects the demand for the goods and services that embody new technologies. This mechanism is orthogonal to the forces we explore in this paper and we control for it by including log of domestic income per capita (y_{ct}) as an independent variable. Controlling for income also takes care of the potential effects of foreign business cycles on domestic technology adoption. This is the case, to the extent that international business cycles affect the domestic economy (only) through the domestic income level. One obvious, yet worthwhile remark, is that we can control for domestic income because the dependent variable in our analysis is a direct measure of technology adoption (as opposed to something much closer to income such as TFP). We use Madison (2005) data to construct the (log) of per capita GDP (in 1990 dollars).

[11] We select the 20 technologies with observations for the largest number of countries and that are relevant for a variety of sectors.

[12] In particular, we add the minimum of x^j_{ct} along c and t, for the years used in the regressions.

Table 1: List of Technologies

Sector	Name
Transportation	Aviation Passengers*Km
	Aviation Tons*Km
	Cars
	Rail Line Km
	Rail Passengers*Km
	Rail Tons*Km
	Ships
	Trucks
Communication	Cellphone
	Computer
	Internet
	Radio
	Telegram
	Telephone
	TV
Industry	ATM
	Electricity
	Steel Tons from Blast Oxygen
	Steel Tons from Electric-arc
	Tractors

In all our regressions we include technology-country dummies, I_c^j, and technology-year dummies, I_t^j. Country dummies capture country-specific factors that affect technology diffusion and that are relatively constant over the time-span in which the technology diffuses. These might include geographical variables (e.g., remoteness, size of the country, density, ruggedness, climate,...), institutional variables (e.g. political regime, expropriation risk,...) or historical endowment (e.g., familiarity with related technologies, education system, ...). Note that we allow country dummies to differ across technologies to capture the possibly different effect that persistent factors have on different technologies.

The inclusion of time and country dummies affects the identification of the estimated coefficients. The country fixed effects imply that the estimates reflect correlations of the change in the dependent variable with the change in the adoption level, x_{ct}^j. That is, with the diffusion of the technology. Technology-year dummies remove the average evolution of the diffusion process for each technology which may vary across technologies for a variety of factors largely orthogonal to our analysis.[13] As a result, the estimated coefficients capture the differential effect on technology diffusion of the dependent variables in a country relative to the rest.

[13] These may include the nature of the technology, its capital intensity, when the technology was invented (Comin and Hobijn, 2010), etc.

The centerpiece of our exploration of the presence of geographical interactions in adoption is the spatial distance from other countries' technology (SDT). Intuitively, SDT is just an interaction between the (log) of adoption in other countries and how distant they are. In principle, there are many different ways to construct these interactions. In Appendix A, we present several alternatives and show the robustness of the basic empirical findings to these various specifications of SDT. Our baseline way to compute the interaction between technology and distance is as the scalar product of a vector of (log) adoption levels in the other countries and the vector of distances (thousands of kilometers) to these other countries. Formally,

$$x^j_{-ct} = \sum_{\forall k \neq c} d_{ck} x^j_{kt}$$

where d_{ck} is the distance between countries c and k.

Note that, when the number of countries is large, the vector of adoption measures in the rest of the world, x^j_{kt}, is almost the same across countries, and the cross-country variation in SDT comes from differences in the matrix of distances. Therefore, in the cross-section, SDT is highly correlated with the remoteness of the country. Because the matrix of distances is constant over time (other than due to changes in the sample composition) this direct effect of remoteness on adoption is captured by the technology-specific country fixed effects. Therefore it does not affect the identification of β^j_4.

Since the matrix of distances is constant over time, time variation in SDT is generated by the diffusion of technology (i.e., from x^j_{kt}). As technology diffuses, x^j_{-ct} increases slowly in countries located close to places where technology diffuses faster. Conversely, x^j_{-ct} increases faster in places that are farther from countries where technology diffuses faster. Therefore, if being close to adoption leaders is beneficial for the diffusion of technology, we should observe that x^j_{-ct} is negatively correlated with adoption, x^j_{ct}. This is the logic behind the identification of β^j_4. Note that, because of the time dummies, the identification of β^j_4 comes from the relative change of SDT in countries that are close to adoption leaders vs. those that are far (not from absolute changes in SDT).[14]

It is important to be aware that there is a potential endogeneity bias concern in the estimation of the regression in (1). Specifically, adoption in country c enters in the construction of the SDT of the other countries. If SDT affects adoption (i.e. $\beta^j_4 \neq 0$), then the adoption levels of the other countries will also be affected by adoption in c. But because SDT in c is computed using adoption in the other countries, it will indirectly be affected also by adoption in c. If in reality β^j_4 is negative, the endogeneity of SDT is likely to introduce a negative bias in β^j_4. This is the case because a higher adoption in country c, x^j_{ct}, increases SDT in the other countries which, because β^j_4 is negative, should result in smaller adoption, x^t_{kt}, which in turn leads to a smaller SDT for country c.

There are two reasons why this bias is not significant concern in practice. First, when the number of countries in the sample is large, the effect of a country's adoption on the other countries SDT is

[14] That is the reason why our findings are robust to various specifications of SDT as shown in Appendix A.

negligible. Second, under the null (i.e. $\beta_4^j = 0$) there is no endogeneity bias and so the standard test to reject the null is still valid. Still, in Appendix B we conduct some back-of-the-envelope calculations and show that even in regressions where we use smaller samples, the endogeneity bias generates less than 0.3% of the standard deviation in SDT and can account for less than 3% of the magnitude of the estimated coefficients. We conclude that the estimates of β_4^j reported below are not significantly affected by an endogeneity bias.

Of course, geographic interactions may take place along variables other than technology. Trade is an obvious example. However, the terms that arise in standard gravity equations used in international trade[15] are all captured in the regressors included in (1) independently of our variable of interest x_{-ct}^j.[16] A literature in political science (e.g. Simmons et al., 2007) has also emphasized the international diffusion of institutions and markets. These other forms of geographic interactions may, in principle, affect the adoption dynamics in a country. To increase our confidence that the geographic interactions we are identifying with SDT occur through technology and not through these alternatives mechanisms, we introduce another control that we call spatial distance from (other countries') income (SDI). SDI is defined in an analogous way to SDT but rather than computing it with other countries' adoption, we use other countries' (log) per capita income. Formally, SDI is defined as follows:

$$y_{-ct} = \sum_{\forall k \neq c} d_{ck} y_{kt}.$$

The controls we add in equation (1) are a way of addressing, in an imperfect way, the reflection problem (see Maski, 1993) that arises in our specification. Of course, there might be other geographic interactions that affect the adoption dynamics in a country. Given the scope of our study in terms of number of countries, technologies and time, it is extremely hard (and impossible given our data) to identify exogenous changes in SDT that identify its effect on adoption dynamics. Hence, we have to rely on the argument that it is hard for us to think of variables that affect diffusion, that are geographically correlated, that change over time according to the patterns we uncover below, and that are not correlated with income (and therefore captured via SDI). Still, if such a variable exited, it could be influencing our results and we could be confounding the effect of diffusion with the effect of this other variable. In Sections 4 and 5 we show that a simple parsimonious model

[15] See Anderson, 2004, and Anderson and van Wincoop, 2003, among many others.

[16] The typical gravity equation (reproduced in 2) has proven an accurate way to predict bilateral trade flows, $TR_{c,c'}$, between countries.

$$TR_{c,c't} = \beta * \frac{Y_{ct} * Y_{c't}}{d_{c,c'}} * \epsilon_{c,c't}. \tag{2}$$

Taking logs in (2) and adding across all other countries, we obtain

$$tr_{ct} = \alpha + \beta_1 * y_{ct} + \beta_2 * \sum_{c'} y_{c't} + \beta_3 * \sum_{c'} d_{c,c'} + \epsilon_{ct}. \tag{3}$$

Note that the regressors in (3) are captured by the controls in (1). In particular, the log income term controls for the effect of own income, the country fixed effect controls for the distance term and, when there are many countries, the time dummies basically capture the average income of the other countries. Hence, SDT identifies effects distinct from standard gravity effects.

Table 2: Descriptive Statistics		
	Standard Deviation	Residual after Removing Country and Time*Technology FE Standard Deviation
X (technology)	2.849	2.74
Y (income)	1.00	0.99
Distance Interaction	3246.45	2731.098
Income Interaction	416.29	263.62
Distance Interaction (latitude)	1461 .37	1240.60
Distance Interaction (longitude)	3533.42	3130.42
Distance Interaction (abs. latitude)	943.21	767.37
N	53579	53579

of technology diffusion can capture well the pattern we find in the data. This, we believe, lends credibility to our interpretation of the technology diffusion patterns we document in the data.

Table 2 presents descriptive statistics of the variables used in the empirical exploration. We report the standard deviations for the raw variables and also for the residuals after regressing each variable against a full set of technology-specific country and time dummies.

We consider four possible specifications of (1) which differ on the restrictions imposed on parameters β_4^j and β_5^j. In our baseline specification (i.e. Specification 1) β_4^j is the same across technologies while we allow β_5^j to differ. In Specification 2 both β_4^j and β_5^j are the same across technologies. In Specification 3, we allow β_4^j to differ across sectors, though not within sectors, and β_5^j varies across technologies. Finally, in the fourth specification, both β_4^j and β_5^j differ across sectors but not across technologies in the same sector.

2.2 Empirical findings

Our empirical approach is flexible and we shall take advantage of this flexibility in several ways. We first investigate the presence and strength of geographic interactions in technology adoption in various sectors and country samples. We are also able to disentangle the nature of geographic interactions in adoption by decomposing the SDT variable along several dimensions.

2.3 Pooled regressions

We start by running regression (1) in our full sample of countries. Table 3 reports the estimates of the effects of SDT on technology adoption in our four specifications. The column labeled Specification 1 reports the estimate of β_4^j in the first specification, where only the effect of SDT is constant across technologies. That is the most flexible specification.[17] We find a negative, significant (at the 1% level) effect of SDT on a country's adoption. As discussed above, this suggests that countries

[17] Note that, in Specifications 1 and 3, we are not reporting the 20 coefficients β_5^j estimated from the SDI terms are also included in the regressions.

that are far from adoption leaders tend to adopt new technologies more slowly than countries that are close. From the statistics in Table 2, it follows that the magnitude of this effect is economically relevant. In particular, a reduction of one standard deviation in SDT leads to an increase in adoption by 17% of one standard deviation.

In the first column, labeled Specification 0, we report the estimate of β_4^j in a regression that does not include the SDI control (recall that in Specification 1 SDI is included but it is allowed to vary by technology). Comparing the first two columns it seems clear that controlling for SDI does not diminish the estimate of geographic interactions in technology. Columns 3 through 5 show that the effect of SDT on technology adoption is robust across the four specifications we explore. Columns 3 and 4 explore the sectoral variation in geographic interactions in adoption. Transportation is the default option. Therefore, the coefficient of SDT for transportation technologies corresponds to the first row. The rows labeled "SDT Com." and "SDT Ind." report the differential coefficient of SDT for communication and industry technologies, respectively, relative to the coefficient for those in transportation. The ranking of the coefficients of SDT on diffusion across technologies is not robust. In Specification 3, the strongest effect is in communication technologies, while in Specification 4, the strongest effect of SDT on technology diffusion is found in transportation technologies.

Table 3: Pooled Regressions

	Specification				
	0	1	2	3	4
SDT	*-.000147****	*-.000171****	*-.000126****	*-.000109****	*-.000080****
	($4.50e^{-6}$)	($8.00e^{-6}$)	($6.82e^{-6}$)	($1.68e^{-5}$)	($1.30e^{-5}$)
SDI	No		.000659***		-.000300***
			($4.56e^{-5}$)		($7.30e^{-5}$)
SDT Com.				*-.000089****	*.000070****
				($1.90e^{-5}$)	($1.60e^{-5}$)
SDT Ind.				*-.000053****	*.000043*
				($2.80e^{-5}$)	($4.30e^{-5}$)
SDI Com.					.000770***
					($1.00e^{-4}$)
SDI Ind.					.000450***
					($1.40e^{-4}$)
# Obs.	53579	53579	53579	53579	53579

Note: Each column corresponds to one specification of regression (1). In Specifications 1 and 3, the coefficient of SDI is allowed to vary by technology. Specification 0 is the only one that does not include the SDI controls.

2.4 The importance of geography for rich and poor countries

After showing the significance of geographic interactions in adoption dynamics, one may wonder whether their relevance is uniform across countries. To explore this question, we split the countries in our sample in two groups depending on whether in 1990 their income per capita was above or below 8000 dollars according to the estimates in Madison (2005).[18] Then we run regression (1) separately in both subsamples.

Table 4: Rich Countries

	Specification			
	1	2	3	4
SDT	*-.000397**** ($3.80e^{-5}$)	*-.000317**** ($3.37e^{-5}$)	*-.000560**** ($6.74e^{-5}$)	*-.000423**** ($5.46e^{-5}$)
SDI		.000459*** ($6.44e^{-5}$)		.000390*** ($1.02e^{-4}$)
SDT Com.			*-.000175**** ($8.38e^{-5}$)	*.000099**** ($7.20e^{-5}$)
SDT Ind.			*-.000597**** ($1.36e^{-4}$)	*.000469**** ($1.19e^{-4}$)
SDI Com.				.000145 ($1.44e^{-4}$)
SDI Ind.				-.000181 ($2.02e^{-4}$)
# Obs.	20151	20151	20151	20151

Note: Estimates of β_4^j for countries with income in 1990 higher than 8000 USD. The first and third specifications allow for technology-specific SDI coefficients.

Tables 4 and 5 report the estimates for the sample of rich and poor countries, respectively. We observe the following findings from this exercise. First, the effect of SDT on a country's adoption is significant both for rich and poor countries. Still, it is significantly larger for rich than for poor countries implying that distance from adoption leaders slows down adoption more in rich countries than in poor ones. There are also interesting differences in the sectors where geographic interactions in adoption are most relevant. For poor countries, we observe stronger effects of SDT on a country's adoption of technologies in communications. For rich countries, it depends on the specification. In Specification 3, the estimates of β_4^j are highest in industry while in the fourth specification they are highest in transportation.

[18] With this split, there are 36 rich countries and 125 poor countries in our sample.

Table 5: Poor Countries				
	Specification			
	1	2	3	4
SDT	*-.000260**** $(1.40e^{-5})$	*-.000250**** $(1.00e^{-5})$	*-.000130**** $(2.90e^{-5})$	*-.000140**** $(2.40e^{-5})$
SDI		-.000510*** $(5.80e^{-5})$		-.000810*** $(9.00e^{-5})$
SDT Com.			*-.000190**** $(3.50e^{-5})$	*-.000150**** $(2.80e^{-5})$
SDT Ind.			*-.000096*** $(4.80e^{-5})$	*-.000100**** $(4.00e^{-5})$
SDI Com.				.000600 $(1.40e^{-4})$
SDI Ind.				.000250* $(1.60e^{-4})$
# Obs.	33428	33428	33428	33428

Note: Estimates of β_4^j for countries with income in 1990 lower than 8000 USD. The first and third specifications allow for technology-specific SDI coefficients.

2.5 Decomposing SDT

A natural next step consists in exploring whether the characteristics of the countries with whom a country interacts with also matter. To investigate this possible dependence, we decompose the SDT variable in two parts. Namely,

$$x_{-ct}^{jRICH} = \sum_{\forall k \neq c \ \& \ k \in RICH} d_{ck} x_{kt}^{j},$$

and

$$x_{-ct}^{jPOOR} = \sum_{\forall k \neq c \ \& \ k \in POOR} d_{ck} x_{kt}^{j},$$

where x_{-ct}^{jRICH} captures the geographic interactions in adoption with rich countries and x_{-ct}^{jPOOR} the interactions with poor countries. Note that, for all countries, $x_{-ct}^{j} = x_{-ct}^{jRICH} + x_{-ct}^{jPOOR}$. The SDI variable can be decomposed in an analogous way.

Table 6 estimates regression (1) allowing for different coefficients in the rich and poor components of SDT. The data again speaks clearly. Adoption interactions with rich countries affect technology adoption between four and five times more than interactions with poor countries. Interestingly, including a similar decomposition for SDI does not affect significantly the estimates of the effects of the two SDT terms. Since adoption leaders are rich countries, we interpret these findings as evidence that it is particularly detrimental to be far from adoption leaders.

Table 6: Rich and Poor Countries

	Specification 1	Specification 2
SDT Rich	*-.000530**** ($1.50e^{-5}$)	*-.000670**** ($1.80e^{-5}$)
SDT Poor	*-.000110**** ($5.12e^{-6}$)	*-.000140**** ($5.07e^{-6}$)
SDI Rich		.001530*** ($1.17e^{-4}$)
SDI Poor		-.000136* ($4.00e^{-5}$)
# Obs.	53579	53579

Note: In the first column SDI Rich and SDI Poor vary by technology. In the second column they are constant across technologies.

2.6 Early adopters

We focus next on the countries that adopt each technology relatively early. For each technology, we define early adopters as the 15 countries with earliest data on adoption. Specifically, we limit the left hand side variable to observations from early adopters and we compute the SDT and SDI variables using only information from early adopter countries.[19] This exercise is relevant because, by design, the panel used in the estimation and in constructing the interactions variables is roughly balanced (there is still the possibility that a country drops from the sample, but this is not a significant concern in CHAT). Therefore, this exercise may provide reassurance that the geographic interactions in adoption we have uncovered are robust to controlling for the sample of countries considered. In addition, early adoption dynamics may be interesting in themselves.

Table 7 reports the estimates of (1) for the early adopters. Qualitatively the results are the same as when studying the full sample. The coefficient of SDT is negative and significant, and it is largest for transportation technologies. However, there are significant quantitative differences between the estimates reported in Tables 3 and 7. The estimates of the geographic interactions in adoption for early adopters (with other early adopters) are four times larger than the equivalent effects for the

[19] The smaller sample of countries makes the potential endogeneity problem a more relevant concern. In Appendix B we calculate a bound on the effect of this endogeneity bias. We find that the true coefficient can only be larger than the coefficient reported by 0.000022. Given the magnitude of the estimated coefficients reported in Table 7, this proves the endogeneity bias essentially irrelevant in practice.

full sample. This should not be surprising since early adopters are rich countries, given that we have already established that (i) rich countries are more sensitive to geographic interactions and that (ii) geographic interactions with rich countries have a larger impact on a country's adoption.

Table 7: Early Adopters

	Specification			
	1	2	3	4
SDT	*-.000700****	*-.001100****	*-.000850****	*-.001600****
	($1.10e^{-4}$)	($1.00e^{-4}$)	($1.60e^{-4}$)	($1.30e^{-4}$)
SDI		.000480***		.000720***
		($5.90e^{-5}$)		($8.60e^{-5}$)
SDT Com.			*.000270*	*.000870****
			($2.40e^{-4}$)	($2.10e^{-4}$)
SDT Ind.			*.000600**	*.000140****
			($4.00e^{-4}$)	($4.00e^{-4}$)
SDI Com.				-.000510***
				($1.40e^{-4}$)
SDI Ind.				-.000510***
				($1.60e^{-4}$)
# Obs.	12540	12540	12540	12540

Note: Each column corresponds to a specification of the regression in (1) for the balanced sample of early adopters. SDT and SDI are computed only with early adopters. In Specifications 1 and 3, the coefficient of SDI can vary by technology.

2.7 Longitude vs. latitude

Jared Diamond conjectured in his 1997 best-seller book "Guns, Germs and Steel" that, due to the specificity of crops to particular climates, technologies have diffused along a given latitude rather than across latitudes. Our simple econometric framework can be adapted to test Diamond's hypothesis. In particular, we can compute separate SDT variables using distances in the east-west dimension (SDT EW) and in the north-south dimension (SDT NS). Diamond's hypothesis is that distance along the north-south axis slows down technology diffusion more than distance along the east-west axis. Therefore, if Diamond's hypothesis is correct, we should observe a higher effect of SDT NS on adoption than of SDT EW.

We start by introducing separately the two SDT terms in Table 8. We find that the coefficient on the SDT across latitudes (i.e. SDT NS) is higher than the SDT across longitudes (i.e. SDT EW). In all regressions we include SDI terms that use the same measures of distance as the corresponding SDT term. The absolute and relative size of the effects of SDT NS and SDT EW on adoption is robust to whether the coefficient of the SDI terms varies or not across technologies.

Table 8: Longitude and Latitude Individually

	Specification Longitude		Specification Latitude	
	1	2	1	2
SDT	*-.000046****	*-.000069****	*-.000310****	*-.000230****
	($6.10e^{-6}$)	($4.91e^{-6}$)	($1.30e^{-5}$)	($1.20e^{-5}$)
SDI		-.000027		-.000480***
		($3.50e^{-5}$)		($7.00e^{-5}$)
# Obs.	53579	53579	53579	53579

Note: Each column corresponds to either Specification 1 or 2 of the regression in (1). SDT and SDI are computed using distance either along longitude (first two columns) or along latitude (third and fourth columns).

In Table 9 we compare the spatial distance interactions across latitudes and longitudes. The outcome is quite clear. As hypothesized by Diamond, being in a distant latitude is a higher barrier to the diffusion of technologies than being in a distant longitude. The estimates imply that distance across latitudes slows down adoption approximately forty seven times more than distance across longitudes.

Table 9: Longitude and Latitude Simultaneously

	Specification	
	1	2
SDT NS	*-.005310****	*-.006700****
	($1.50e^{-5}$)	($1.80e^{-5}$)
SDT EW	*-.000110****	*-.000140****
	($5.10e^{-6}$)	($5.07e^{-6}$)
SDI NS		.001500***
		($1.20e^{-4}$)
SDI EW		.000140***
		($4.00e^{-5}$)
# Obs.	53579	53579

Note: SDT NS and SDI NS are computed using distance along latitudes. SDT EW and SDI EW are computed using distance along longitudes. In Column 1, SDI NS and SDI EW vary by technology.

Confirming the Diamond hypothesis in a sample of technologies without any crops in it is somewhat surprising. Actually, other than tractors, our sample does not contain any agricultural technology. Clearly, Diamond's rationale for the greater importance of latitude distances for technology diffusion is not relevant for technologies such as cars or telephones which can work equally well at different latitudes or longitudes. Providing and testing alternatives explanations for this finding is beyond the scope of this paper. However, we can advance one hypothesis that may be worthwhile investigating in future work. Namely, the diffusion of early agricultural technologies could have created a series of networks and trade routes along latitudes that then have been used for the diffusion of more modern technologies. Clearly, the empirical relevance of this hypothesis remains a topic for future research.

3 The simplest model

3.1 Description

We now present a simple mechanical model to represent and analyze the forces we have uncovered so far. A model can also help us parametrize the effects we find in the data and can point to some new hypothesis to test. For these purposes we want to propose the simplest theory of human interactions that can accommodate the temporal and geographic effects that are present in the data. The theory we propose is a theory of social interactions in which agents meet randomly with other agents and adopt new technologies if the agents they meet have already adopted (similar to the mechanism in Eaton and Kortum, 1999, and Lucas, 2009). We also assume that agents meet more frequently agents that are close by. The model specifies stochastically who do agents meet and when do they adopt new technologies. Agents make no decisions. The result is a mechanical, mathematical description of adoption rates over time. Adoption dynamics are governed by the rate of meetings among agents (α) and the decay in the meeting probability over space (δ). These are the two key parameters we estimate for each technology in Section 5.

3.2 Formalization

Consider an economy where a mass N of agents are located uniformly in space. Space is given by the closed interval $[0,1]$. Time, $t = 0, 1, ...$ is discrete. We consider the diffusion of a technology that is first adopted at time $t = 0$. Let $G(0, \ell, t)$ denote the fraction of agents at location ℓ and time t that have not adopted the technology. The fraction of agents that have adopted is, therefore, given by $G(1, \ell, t) = 1 - G(0, \ell, t)$.

Agents meet randomly with α agents per period. We assume that the new technology strictly dominates the old one, so if an agent meets someone that has adopted the new technology already, he adopts immediately. A meeting between two agents that have not adopted does not lead to any technology upgrading. The parameter α governs the frequency of meetings and therefore determines the speed of technology adoption.

Agents meet more frequently with agents that locate close to them. In particular, the probability that an agent at location ℓ meets an agent at location r is $e^{-\delta|\ell-r|}$ times lower than the probability of meeting an agent that lives at ℓ. The parameter δ governs the importance of space for technology adoption. A high δ implies that agents meet with agents far away from them very infrequently and therefore that diffusion is very localized.

The probability of not adopting in period $t+h$ conditional on not having adopted in period t at location r is then given by

$$G(0,r,t+h) = G(0,\ell,t)\left[\frac{\int_0^1 G(0,\ell,t)\, e^{-\delta|\ell-r|}d\ell}{\int_0^1 e^{-\delta|\ell-r|}d\ell}\right]^{\alpha h},$$

which implies, taking limits as $h \to 0$, that

$$\frac{\partial \ln G(0,r,t)}{\partial t} = \alpha \ln\left(\int_0^1 G(0,\ell,t)\, e^{-\delta|\ell-r|}d\ell\right) - \alpha \ln\left(\int_0^1 e^{-\delta|\ell-r|}d\ell\right). \tag{4}$$

The above equation implies that if $G(0,\ell,0) < 1$ for some interval of positive Lebesgue measure $L \in [0,1]$, $G(0,\ell,t) < 1$ for all ℓ and t and $G(0,\ell,t)$ is increasing over time for all ℓ. That is, if a non-trivial number of agents adopted in period $t=0$, then the technology diffuses to all locations and adoption increases over time at all locations.

The effect of geography enters the model only through the distribution of the first adopters, $G(0,\cdot,0)$. To illustrate this, consider an example without geography where $G(0,\ell,0) = g < 1$ for all ℓ. So the same fraction of agents in all locations start adopting at time zero. Then $\partial \ln G(0,\ell,t)/\partial t = \alpha \ln G(0,\ell,t)$ so $\partial \ln G(0,\ell,0)/\partial t = \alpha \ln g$. One can then guess and verify that the solution of the differential equation is given by $G(0,\ell,t) = e^{e^{\alpha t}\ln g} = g^{e^{\alpha t}}$.[20] In this example space plays no role. Technology diffuses slowly and uniformly and eventually all agents adopt, since $\lim_{t\to\infty} G(0,\ell,t) = \lim_{t\to\infty} g^{e^{\alpha t}} = 0$.

The example above eliminates the importance of space using two assumptions. First, assuming that the number of meetings is independent of the location (α is constant). An assumption we will maintain throughout. Second, it assumes that the density of adoption at $t=0$ is uniform. This second assumption is unrealistic and should be modified. Initial adoption is in general concentrated geographically. For example, it is probably concentrated close to the inventor of the new technology. Therefore, a natural way to add geography is to add heterogeneity in initial conditions. In Section 5 we do this in a variety of ways that help us fit the data. However, to illustrate the implications of the model, the simplest way is to start with an interval of locations that adopts initially, while all other areas start with no adoption whatsoever. Formally, the initial conditions now are

$$G(0,\ell,0) = \begin{cases} g < 1 \text{ for } \ell \in [0,a] \\ g = 1 \text{ otherwise} \end{cases}.$$

[20] Guess that $G(0,\ell,t) = e^{\lambda(t)\ln g}$, so $\ln G(0,\ell,t) = \lambda(t)\ln g$ and so

$$\frac{\partial \ln G(0,\ell,t)}{\partial t} = \alpha\frac{\partial \lambda(t)}{\partial t}\ln g.$$

Hence, $\lambda(t) = \alpha\partial\lambda(t)/\partial t$ and so $\lambda(t) = e^{\alpha t}$. This implies that $G(0,\ell,t) = e^{e^{\alpha t}\ln g} = g^{e^{\alpha t}}$.

The resulting dynamics are more complicated than before and cannot be fully solved analytically. However, since $g < 1$,

$$\frac{\partial \ln G(0,r,0)}{\partial t} = \alpha \ln \left(g \int_0^a e^{-\delta|\ell - r|} d\ell + \int_a^1 e^{-\delta|\ell - r|} d\ell \right) - \alpha \ln \left(\int_0^1 e^{-\delta|\ell - r|} d\ell \right) < 0$$

for all ℓ and so for $a < \ell < \ell'$, $\partial \ln G(0,\ell,0)/\partial t < \partial \ln G(0,\ell',0)/\partial t$. Since $G(0,\ell,0)$ is decreasing in ℓ, this implies that $\partial \ln G(0,\ell,t)/\partial t < \partial \ln G(0,\ell',t)/\partial t$, and thus

$$\frac{\partial^2 \ln G(0,\ell,t)}{\partial t \partial \ell} > 0, \text{ for all } t \text{ and all } \ell > a.$$[21]

The previous arguments result in the following two implications:

Implication 1: *The fraction of non-adopters is lower in locations closer to the source of innovation.*

Implication 2: *The fraction of non-adopters declines proportionally faster in locations closer to the source of innovation.*

Since this process implies that in the limit all locations adopt fully so $G(1,\ell,t) = 1$ for all ℓ, we can also conclude that:

Implication 3: *The effect of distance on the level of adoption vanishes over time.*

The parameters α and δ affect the growth in the fraction of adopters as well as their level. It is easy to conclude from equation (4) that

$$\frac{\partial^2 \ln G(0,\ell,t)}{\partial t \partial \alpha} < 0.$$

Therefore, the larger α the faster adoption grows over time. It is harder to draw analytically other conclusions on the effects of α and δ on the evolution of adoption. However, we can illustrate them with the help of numerical examples. Figures 1 to 3 show three examples with $\alpha = 0.05$ and $\delta = 10$, $\alpha = 0.05$ and $\delta = 20$, and $\alpha = 0.01$ and $\delta = 10$.[22] The left panel represents the fraction of adopters over space in 10 different time periods, with equal intervals between them. The right panel represents the density of adopters over time for 10 points in space (again, equally spaced).

The results are clear, intuitively, and expected: First, the density of adopters decreases as we move away from ℓ, and the slope (in logs) decreases with time. The slope increases in absolute value with δ. Second, in all locations adoption increases monotonically over time, with the fraction of non-adopters falling proportionally slower in locations farther away from the initial innovation (in the examples $\ell = 0$). Finally, the growth rate of adoption increases with the number of meetings per period, α.

[22] We simulate the model for the case where $a = 1/1000$, $g = .99$ and two levels of α and δ. The fraction of adopters is plotted in log scale. We use a grid of 1000 points for space and run the model over 300 periods.

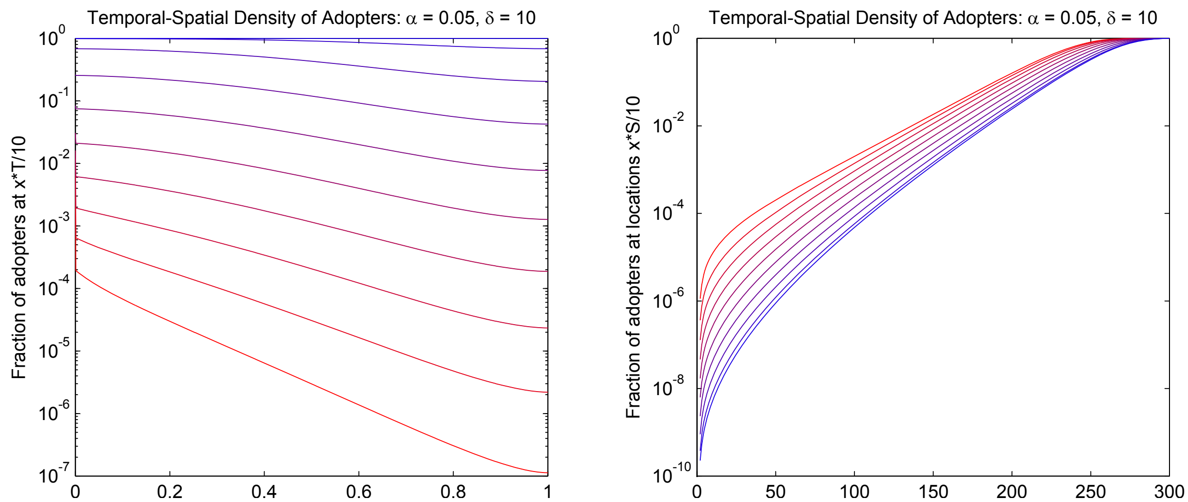

Figure 1: $\alpha = 0.05$ and $\delta = 10$

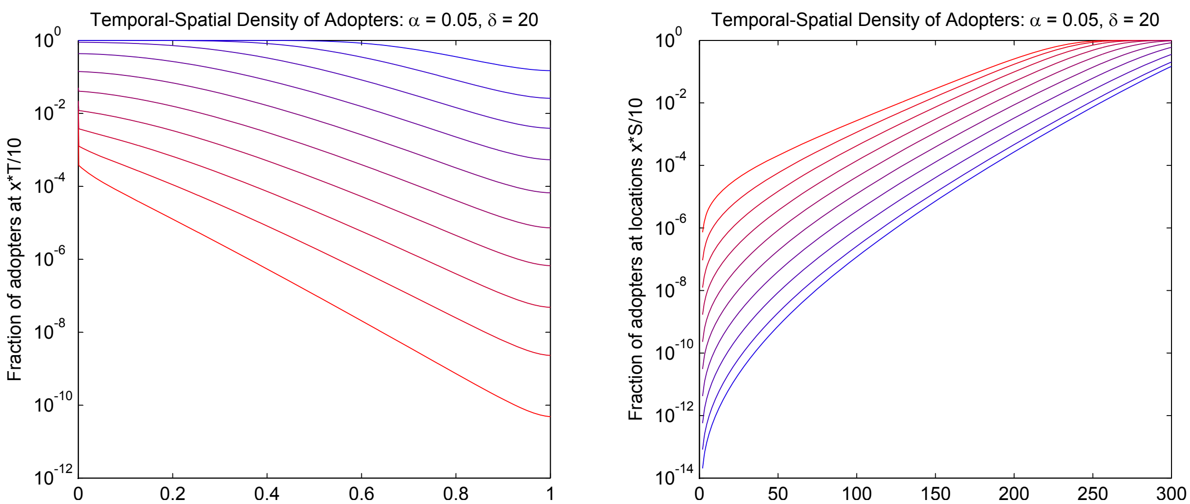

Figure 2: $\alpha = 0.05$ and $\delta = 20$

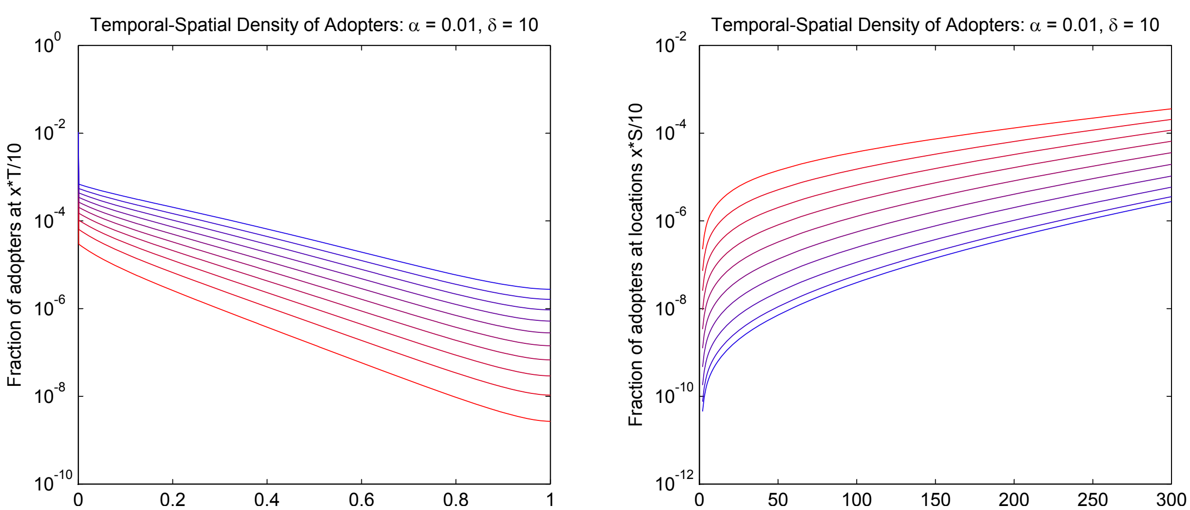

Figure 3: $\alpha = 0.01$ and $\delta = 10$

The regressions presented in Section 2 show that Implications 1 and 2 are consistent with the data. In particular we found that the coefficient on SDT in (1) is negative and significant. We now proceed to contrast the other prediction. In particular, we are interested in Implication 3, which tells us that the effect of SDT on adoption should vanish over time.

4 Exploring the model's predictions

To contrast Implication 3 with the data we proceed in two steps. First, we modify the specification in (1) to allow for time varying coefficients of β_4^j and β_5^j. The new specification is given by

$$x_{ct}^j = \beta_{1c}^j I_c^j + \beta_{2t}^j I_t^j + \beta_3^j y_{ct} + \beta_{4t}^j x_{-ct}^j + \beta_{5t}^j y_{-ct} + \epsilon_{ct}^j. \tag{5}$$

In a second step, for each technology, we take the series of estimates of β_{4t}^j, and fit them the following three-parameter non-linear specification

$$\beta_{4t}^j = c^j + e^{-b^j(t-t_0)}(a^j - c^j) + \tilde{\epsilon}_t^j, \tag{6}$$

where $\tilde{\epsilon}_t^j$ is a residual, and t_0 is the initial adoption year. The parameter a^j determines the initial level of β_{4t}^j, and, according to our model, it should be negative. The parameter b^j determines the rate of increase of β_{4t}^j, and should be positive according to our theory. When b^j is positive, c_j is the long run level of β_{4t}^j.

We apply this two-stage procedure both for the balanced (15 countries) and unbalanced (161 countries) samples. Figures 4 and 5 plot, for each technology, the estimates of β_4^j in the unbalanced (Figure 4) and balanced (Figure 5) samples together with the fitted curves from (6). The first observation is that in a large majority of the technologies the model predictions are borne by the data. In particular, in 13 out of 20 technologies in the unbalanced sample, and in 19 out of 20 in the balanced sample, we observe estimates of β_{4t}^j that are initially negative and increase over time.

Tables 10 and 11 report, for each technology, the estimates and standard errors of a^j, b^j and c^j, together with the R^2 of regression (6) for the unbalanced (Table 10) and balanced (Table 11) samples. The table also reports the year of invention of the technology. The point estimates confirm that the data conforms to the model's predictions. The point estimates of a large majority of technologies in both the unbalanced and balanced samples have negative estimates of a^j, positive estimates of b^j and estimates of c^j that often are close to zero and are almost always smaller (in absolute value) than the point estimates of a^j.

Note also the goodness of fit of the three-parameter specification (6) to the time-varying estimates β_4^j. Both in the unbalanced and (especially) in the balanced samples the R^2's are very high. The median R^2 for the unbalanced sample is 0.71 and for the balanced sample it is 0.98.

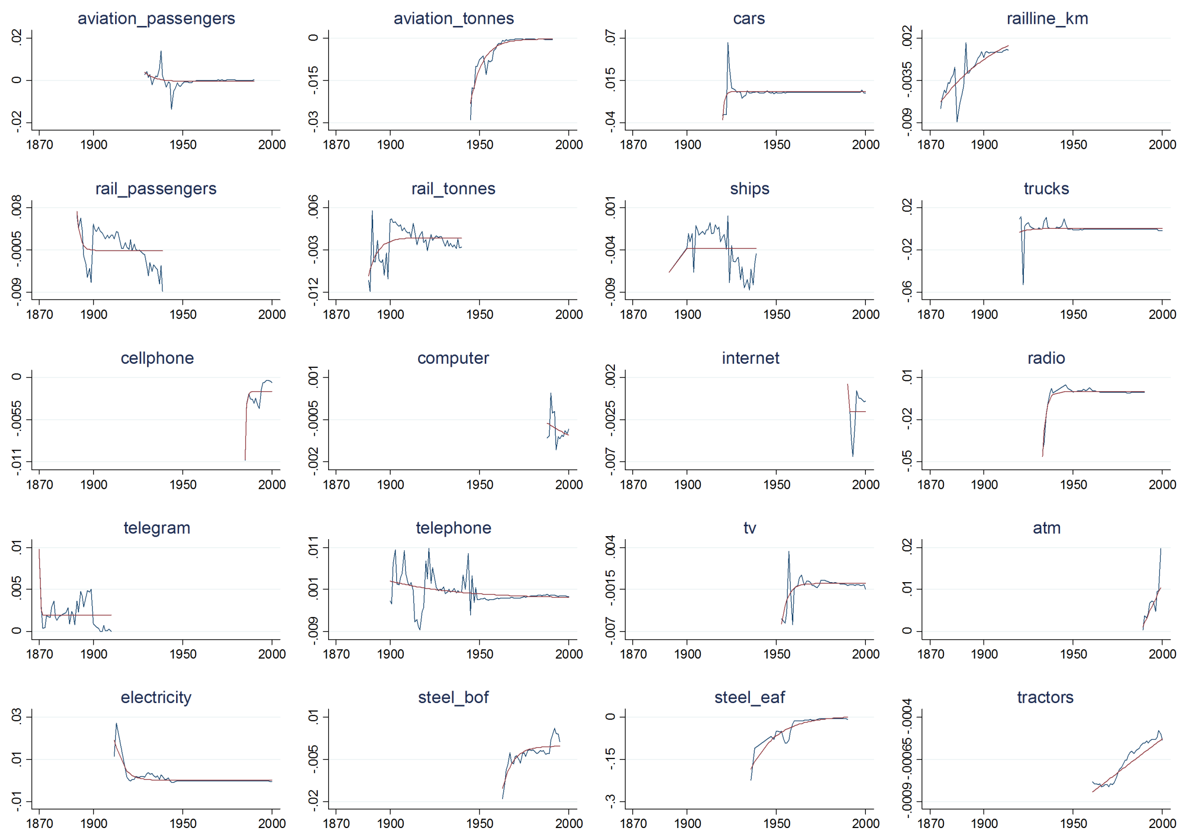

Figure 4: Estimates of β_{4t}^{j} in the unbalanced sample; fitted lines from the regression in (6)

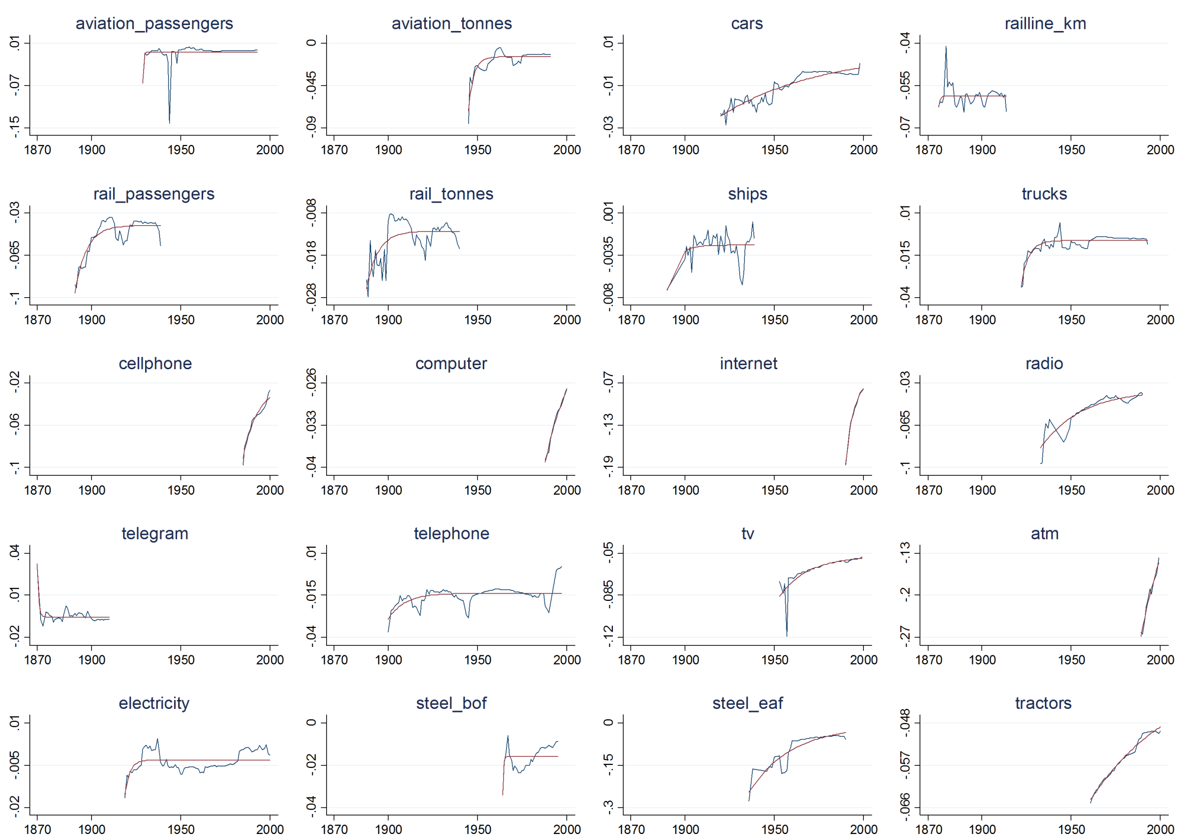

Figure 5: Estimates of β_{4t}^{j} in the balanced sample; fitted lines from the regression in (6)

Table 10: Unbalanced Sample										
Sector	Technology	Year	a^j	*s.e.* a^j	b^j	*s.e.* b^j	c^j	*s.e.* c^j	R^2	# Obs.
Trans.	Aviation Pass.	1903	0.0037	0.0021	*0.1772*	0.1500	-0.0005	0.0005	0.09	62
	Aviation Ton.	1903	-0.0234	0.0014	*0.1507*	0.0174	-0.0004	0.0004	0.92	47
	Cars	1885	-0.0365	0.0091	*0.9820*	0.5466	0.0001	0.0010	0.19	81
	Rail Line Km	1825	-0.0063	0.0008	*0.0270*	0.0243	0.0052	0.0065	0.79	39
	Rail Pass.	1825	0.0073	0.0033	*0.5296*	0.4084	-0.0006	0.0005	0.14	49
	Rail Ton.	1825	-0.0085	0.0023	*0.1950*	0.0950	-0.0004	0.0005	0.35	53
	Ships	1776	-0.0067	0.0025	*2.6075*	0.0000	-0.0039	0.0004	0.03	41
	Trucks	1903	-0.0031	0.0050	*0.2088*	0.5213	0.0003	0.0008	0.01	81
Comm.	Cellphone	1973	-0.0107	0.0013	*1.6287*	0.7547	-0.0018	0.0004	0.88	16
	Computer	1973	-0.0006	0.0004	*0.0240*	0.6814	-0.0024	0.0447	0.75	13
	Internet	1983	0.0013	0.0023	*25.779*	0.0000	-0.0016	0.0007	0.15	11
	Radio	1920	-0.0464	0.0020	*0.5203*	0.0413	0.0004	0.0003	0.94	52
	Telegram	1835	0.0098	0.0015	*2.5420*	2.5098	0.0019	0.0002	0.74	41
	Telephone	1876	0.0030	0.0012	*0.0257*	0.0258	-0.0012	0.0015	0.12	101
	TV	1927	-0.0060	0.0012	*0.2978*	0.1059	-0.0007	0.0002	0.58	47
Industry	ATM	1971	0.0011	0.0021	*0.0095*	0.2052	0.1031	2.0728	0.86	11
	Electricity	1882	0.0191	0.0016	*0.2144*	0.0328	0.0000	0.0002	0.69	84
	Steel Bof	1950	-0.0154	0.0019	*0.1711*	0.0459	-0.0003	0.0008	0.81	32
	Steel Eaf	1907	-0.1843	0.0137	*0.0699*	0.0098	0.0033	0.0065	0.89	47
	Tractors	1903	-0.0008	0.0000	*0.0000*	0.0099	0.3239	126.38	0.87	40

Note: Estimates of a^j, b^j and c^j from regression (6), and goodness of fit.

5 Structural estimation

After exploring the presence of geographic interactions in adoption in the data without using a theoretical structure, it is informative to study the spatial diffusion of technology in a more structured way. In particular, a structural estimation may serve two purposes. First, it will allow us to understand whether a model as stylized as ours can do a decent job in fitting the patterns uncovered in the data. Second, it will allow us to identify the values of the deep parameters that in our model govern the frequency of interactions between agents and how geographic distance affects the probability of a successful interaction. These parameter values can in turn be used to quantify spatial growth models, as in Desmet and Rossi-Hansberg (2011).

We consider a sample of 15 countries and locate them evenly spaced in the unit interval so that their locations can be indexed by $j = \{1, 2, ..., 14, 15\}$.[23] As we have seen in the simulations,

[23] This sample size corresponds to the balanced panel we have used above. In our structural execises we focus on this sample because simulating the unbalanced sample has the additional complexity of countries entering the sample at different times.

geography matters for the diffusion of technology, and in particular, where the adoption leader is located affects the diffusion dynamics. We place the leader in the middle of the unit interval (i.e. $j = 8$).

Table 11: Balanced Sample

Sector	Technology	Year	a^j	*s.e.* a^j	b^j	*s.e.* b^j	c^j	*s.e.* c^j	R^2	# Obs.
Trans.	Aviation Pass.	1903	-0.0666	0.0181	*3.1571*	6.9464	-0.0070	0.0023	0.27	65
	Aviation Ton.	1903	-0.0723	0.0056	*0.3367*	0.0586	-0.0146	0.0011	0.93	47
	Cars	1885	-0.0246	0.0011	*0.0177*	0.0049	0.0060	0.0046	0.96	79
	Rail Line Km	1825	-0.0628	0.0039	*1.1609*	2.6294	-0.0586	0.0007	0.99	39
	Rail Pass.	1825	-0.0962	0.0049	*0.1556*	0.0250	-0.0404	0.0014	0.98	49
	Rail Ton.	1825	-0.0258	0.0025	*0.1554*	0.0525	-0.0123	0.0007	0.94	53
	Ships	1776	-0.0072	0.0014	*0.1948*	0.1232	-0.0024	0.0003	0.80	41
	Trucks	1903	-0.0338	0.0023	*0.2414*	0.0339	-0.0063	0.0004	0.92	72
Comm.	Cellphone	1973	-0.0914	0.0035	*0.1287*	0.0354	-0.0244	0.0080	0.99	16
	Computer	1973	-0.0391	0.0004	*0.0667*	0.0233	-0.0173	0.0053	0.99	13
	Internet	1983	-0.1841	0.0019	*0.2240*	0.0165	-0.0659	0.0035	0.99	11
	Radio	1920	-0.0837	0.0031	*0.0439*	0.0105	-0.0359	0.0041	0.99	52
	Telegram	1835	0.0324	0.0030	*1.3812*	0.2852	-0.0054	0.0005	0.87	41
	Telephone	1876	-0.0294	0.0030	*0.1070*	0.0346	-0.0138	0.0006	0.91	95
	TV	1927	-0.0860	0.0041	*0.0629*	0.0232	-0.0528	0.0038	0.99	46
Industry	ATM	1971	-0.2677	0.0069	*0.0464*	0.0497	0.0575	0.2780	0.99	11
	Electricity	1882	-0.0155	0.0028	*0.3644*	0.1450	-0.0032	0.0004	0.63	82
	Steel Bof	1950	-0.0340	0.0051	*2.0971*	2.2733	-0.0159	0.0010	0.92	31
	Steel Eaf	1907	-0.2466	0.0181	*0.0433*	0.0100	-0.0121	0.0194	0.94	47
	Tractors	1903	-0.0644	0.0003	*0.0209*	0.0043	-0.0364	0.0040	0.99	40

Note: Estimates of a^j, b^j and c^j from regression (6), and goodness of fit.

Simulation for a given α and δ $-$ To bring the model to the data, we recognize that while diffusion in the model is measured by the percentage of adopters, CHAT variables measure the amount of output produced with the technology (per capita) or the number of units of the technology (per capita). The difference between adoption measures in the model and data is that the data includes an intensive margin (i.e. number of units of technology per adopter) that in the model is absent. We make the model and data comparable by adding an intensive margin to the model. The simplest way to model the intensive margin is as a log-linear function of income. However, because the baseline regression in (1) already controls for log income, the intensive margin that we add to the simple model presented above is just a technology-specific constant. We compute this constant from the leader's adoption (in CHAT) in the terminal period, T. In particular, as time goes to infinity, the fraction of adopters goes to 1. Therefore, the (log) intensive margin, $\bar{x}^j$, is equal to

$$\bar{x}^j = \max_i x_{iT}^j.$$

Before simulating the model, we need to determine the initial condition, namely, the fraction of initial adopters in the leader and in the rest of the countries. We use the initial adoption measures in CHAT to calculate these two moments. In particular, given our calibration of the intensive margin, $\bar{x}^j$, the initial (log) fraction of adopters in the leading country is given by

$$\max_i \log G_{i0}^j = \max_i x_{i0}^j - \bar{x}^j. \tag{7}$$

Similarly, we set the initial (log) fraction of adopters in the remaining 14 countries to match the average adoption level among the followers in the first year in which the adoption data is available in CHAT. Namely,

$$\log G_{k0}^j = \frac{\sum_i x_{i0}^j}{15} - \bar{x}^j$$

Then we use our model of diffusion to generate time-series associated with a given α and δ for the share of adopters, G_{it}^j. For each technology, the time series have the same length as the CHAT time series. We then construct the model adoption levels as

$$\hat{x}_{it}^j = \log G_{it}^j + \bar{x}^j.$$

Estimation of α and $\delta-$ Once we generate the adoption data for a given pair (α, δ) we are ready to apply the estimation procedure. For each technology the estimates of α and δ are those values that minimize the distance in the time-varying coefficients, β_{4t}^j, between the model and the data.

To compute the model' s counterpart to β_{4t}^j we proceed as follows. First, we compute the model time-series for SDT ($SDTM$) as follows:

$$SDTM_{it}^j = \sum_{k \neq i} \hat{x}_{it}^j d_{ik}.$$

Then, the model's counterpart to β_{4t}^j is given by the $\tilde{\beta}_t^j$ from the following regression:

$$\hat{x}_{it}^j = I_i^j + I_t^j + \sum_t \tilde{\beta}_t^j SDTM_{it}^j + \epsilon_{it}. \tag{8}$$

We choose α and δ to minimize the sum of squared distances between the series of data estimates β_{4t}^j and the series of model estimates $\tilde{\beta}_t^j$.

Estimation results $-$ Figure 6 plots, for each technology, the estimates of β_{4t}^j from the data and the estimates of $\tilde{\beta}_t^j$ associated with the optimal α and δ. The model does a reasonably good job in fitting the evolution of the effect of SDT on adoption for 14 out of the 20 technologies.[24] For

[24] As shown in Table 12, the R^2 for some of these technologies is negative. Mechanically, we may obtain negative R^2's because the standard deviation of the error terms are larger than the standard deviation of the estimates $\tilde{\beta}_t^j$.

these technologies, $\tilde{\beta}_t^j$ is initially negative, then it starts increasing, and it ends at a less negative level. These pattern reflects the presence of geographic interactions in adoption and the decline in their intensity as technology diffuses and adoption levels become more uniform across countries. The fact that typically $\tilde{\beta}_t^j$ does not converge to zero is due either to the fact that technologies have not fully diffused or that, with country fixed effects, $\tilde{\beta}_t^j$ does not need to asymptotically converge to zero.[25]

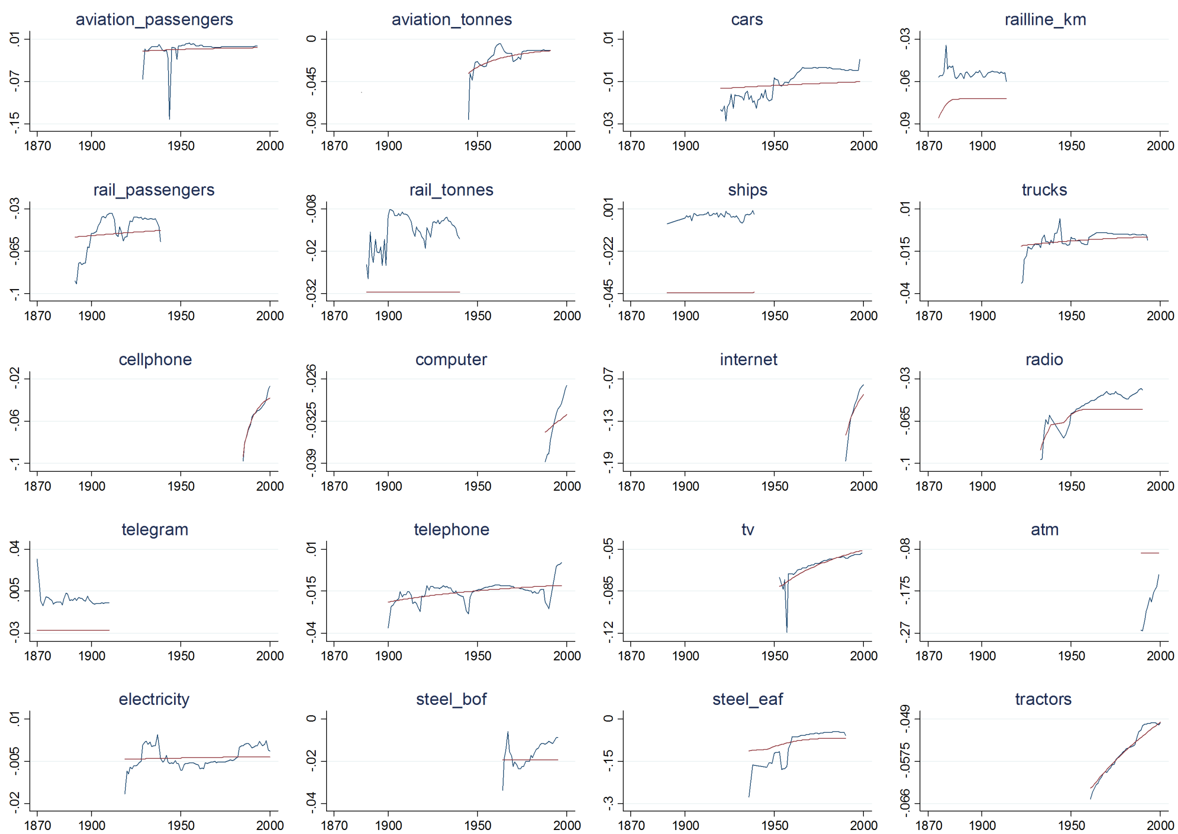

Figure 6: Estimates of geographic interactions in adoption in the balanced sample (data in blue and model in red).

For five of the six technologies where the model fails to generate the pattern of β_4^j observed in the data,[26] the estimates $\tilde{\beta}_t^j$ are flat and below the series of β_{4t}^j observed in the data. Why is the model unable to generate the observed patterns of geographic interactions? Intuitively, for these technologies, the geographic distribution of the followers is such that, with a constant initial adoption among followers, the SDT variable has a very small dispersion (relative to the dispersion in adoption). As a result, the coefficients $\tilde{\beta}_t^j$ are higher (in absolute value) than what we compute in the data (i.e. β_4^j). In this case the α and δ that minimize the distance between the series of $\tilde{\beta}_t^j$ and β_4^j are those that generate a series for $\tilde{\beta}_t^j$ that is less negative. This typically is the case when

[25] Intuitively, the country fixed effects introduce positive (asymptotic) dispersion on the RHS. Therefore, $\tilde{\beta}_t^j$ cannot converge to zero as the dispersion of the LHS goes to zero. $\tilde{\beta}_t^j$ needs to converge to some negative value to undo the dispersion introduced on the RHS by the country fixed effects.

[26] These are rail-lines, rail-tons, ships, telegrams and blast-oxygen steel.

α is very low and δ is very high. Table 12 reports the estimates of α and δ for each technology. We can see how, for the technologies where the model does badly and the model's estimates are below the data coefficients, α is equal to 0, and δ is high. What this implies, effectively, is that in the model these technologies do not diffuse because there are no contacts and interactions occur only with adopters in very near locations. Since technology (almost) does not diffuse, we are effectively recreating the initial conditions every period, which leads to $\tilde{\beta}_t^j$'s that are roughly constant.[27]

Table 12: Structural Estimates of α and δ[28]

Sector	Technology	Year	α	δ	R^2
Transport	Aviation Passengers	1903	0.117	3.900	0.042
	Aviation Tons	1903	0.156	0.563	0.488
	Cars	1885	0.015	3.900	0.202
	Rail Line Km	1825	0.346	0.119	-26.2
	Rail Passengers	1825	0.008	3.900	0.113
	Rail Tons	1825	0.000	10.000	-13.9
	Ships	1776	0.000	3.500	-703.8
	Trucks	1903	0.040	0.626	0.270
Communication	Cellphone	1973	0.577	0.137	0.943
	Computer	1973	0.046	1.082	0.384
	Internet	1983	0.318	0.227	0.788
	Radio	1920	0.232	0.105	0.405
	Telegram	1835	0.000	2.700	-17.0
	Telephone	1876	0.036	0.662	0.269
	TV	1927	0.046	0.483	0.599
Industry	ATM	1971	0.001	0.100	-8.676
	Electricity	1882	0.018	0.666	0.050
	Steel Bof	1950	0.000	3.900	-0.226
	Steel Eaf	1907	0.139	3.498	0.362
	Tractors	1903	0.038	1.500	0.960

[27] Since with this configuration of α and δ we are recreating a very similar cross-location diffusion pattern every period, the location fixed effects capture most of the variation in adoption. Therefore, the coefficients $\tilde{\beta}_t^j$ are closest to 0.

[28] The R^2's we report are calculated as

$$R^2 = 1 - \frac{\sum_t (\beta_{4t}^j - \tilde{\beta}_t^j)^2}{\sum_t (\beta_{4t}^j - \overline{\beta_t^4})^2},$$

where

$$\beta_t^4 = \frac{\sum_t \beta_{4t}^j}{T}.$$

The case of ATMs is different because for this technology the geographic interactions in adoption observed in the data are so high that the model is unable to come close to them. That may be the case either because the geographic distance between the advanced economies, that entirely form the ATM sample, is small and therefore SDT has a small dispersion in the data; or because the technological knowledge needed to adopt ATMs flows intensely across countries. Either way, the model cannot generate such a strong covariance between adoption and SDT. The best fit for this technology is achieved with a very small α, and a very low δ. This is the case because if technology diffused fast across locations, there would be an even larger dispersion in SDT across countries and the estimated coefficients for $\tilde{\beta}_t^j$ would be lower than what is reported in Figure 6.

The point estimates of α and δ provide valuable information about the spatial and temporal diffusion processes. The estimates vary significantly by technology. α is highest for cell phones, and the internet and lowest for rail tons, ships, telegrams and blast oxygen steel. Note that, there seems to be a pattern by which α is increasing in the invention date of a technology. Figure 7A confirms this observation. The correlation between invention date and α is 0.34 though it is insignificant at a 5% level. δ is highest for some transportation technologies such as rail tons, rail passengers, cars and aviation passengers, and it is lowest for some communication technologies such as radio, cell phone or the internet. For δ, there seems to be also a correlation with the invention date. Figure 7B shows that newer technologies have lower estimates of δ. The correlation between these two series is -0.47 and it is statistically significant at the 5% level.

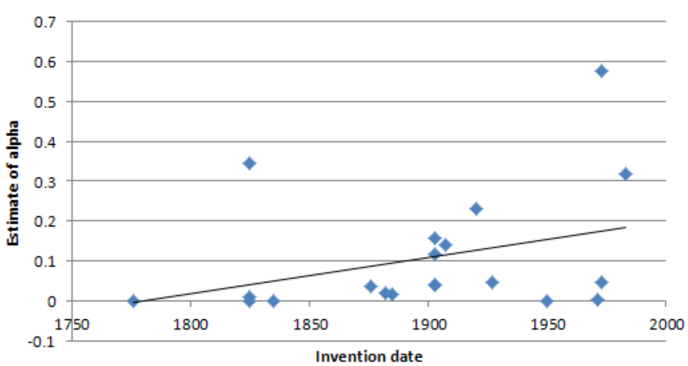

Figure 7A: α vs. invention date.[29]

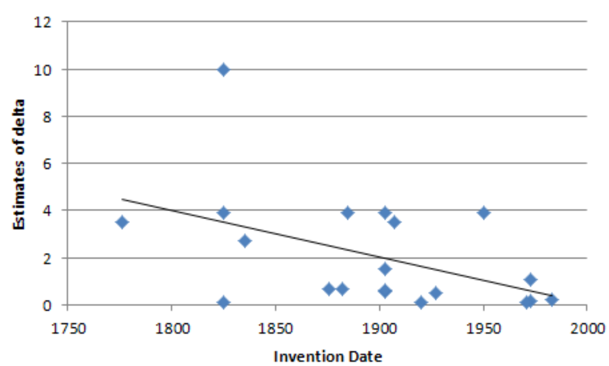

Figure 7B: δ vs. invention date.[30]

5.1 Heterogeneity in initial conditions

In the structural estimation of the model, we have assumed that the initial adoption in all the non-leading countries is the same. Of course, this is not the case in reality. We conclude our exploration of the model by relaxing this assumption. As before, we parameterize the initial fraction of adopters in the leading country as in (7). For the followers the fraction of initial adopters shall depend on their position in the unit interval. For those with location index $j \in [1, 7]$, we set the (log) of the

[29] The regression line has a slope equal to 0.0009 with s.e. 0.0007.

[30] The regression line has a slope equal to -0.0198 with s.e. 0.0095.

share of initial adopters to

$$\log G_{k0}^{j} = \frac{\sum_i x_{i0}^{j}}{15} - \sigma_x + 2\sigma_x \frac{j-1}{6} - \bar{x}^{j},$$

and for countries with index $j \in [9, 15]$, we set it to

$$\log G_{k0}^{j} = \frac{\sum_i x_{i0}^{j}}{15} - \sigma_x + 2\sigma_x \frac{15-j}{6} - \bar{x}^{j}.$$

Note that, with this parameterization of the log of initial adoption, σ_x is equal to the standard deviation of initial adoption across followers. In addition to estimating α and δ, we now also estimate structurally σ_x. By allowing σ_x to vary across technologies, we allow the model to capture cross-country differences in the initial share of adopters, as well as other geographic differences in the sample that affect the diffusion dynamics (e.g. distribution of distance between countries). As before, the length of the simulated series is equal to the number of years for which we have data in CHAT for each specific technology.

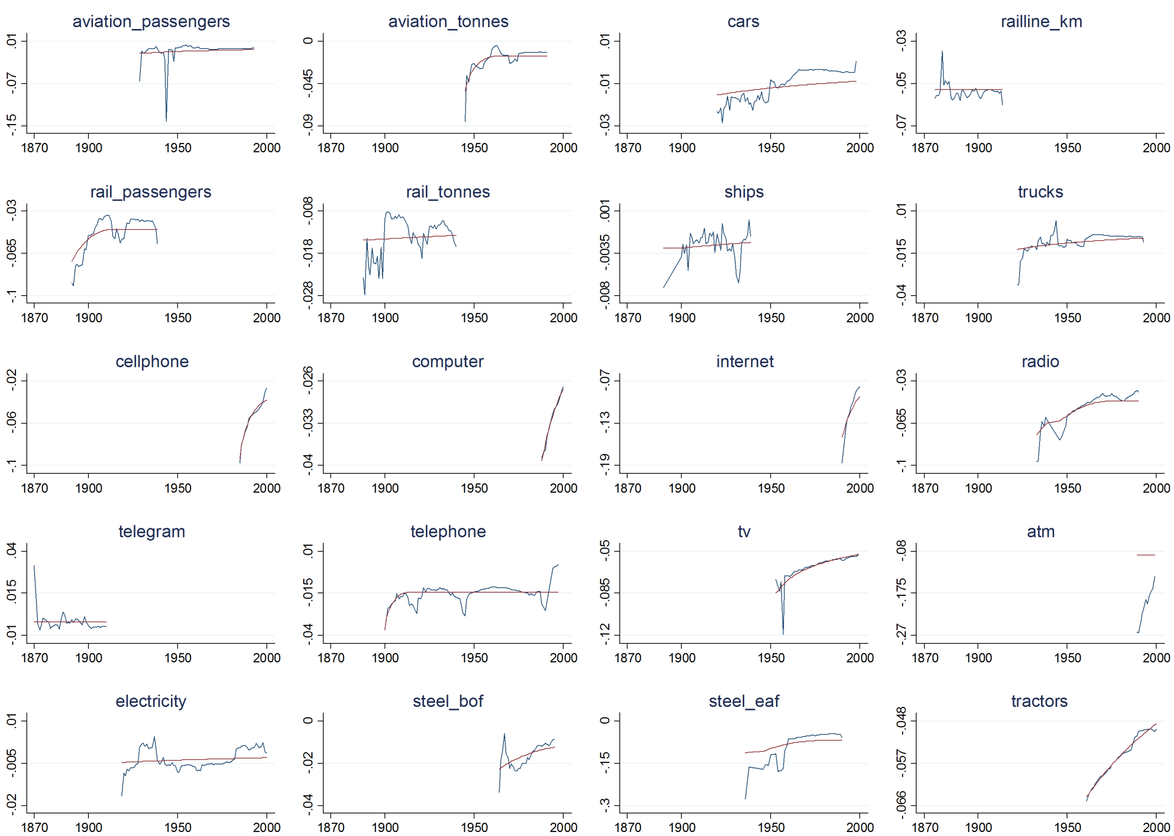

Figure 8: Estimates of geographic interactions in adoption with heterogeneity in the balanced sample (data in blue and model in red).

Figure 8 plots, for each technology the time series for β_{4t}^{j} and the series of $\tilde{\beta}_{t}^{j}$ from regression (8) for the trio (α, δ, σ_x) that minimizes the squared distance between β_{4t}^{j} and $\tilde{\beta}_{t}^{j}$. It is quite evident that the model fit has improved significantly. Now the model seems to fit the pattern of geographic interactions in adoption observed in the data reasonably well for all the technologies with the exception of ATMs. Table 13, which contains the point estimates of α, δ, and σ_x as well as the R^2s, confirms this observation.

Table 13: Structural Estimates of α, δ and σ with Initial Heterogeneity

Sector	Technology	Year	α	δ	σ	R^2
	Aviation Passengers	1903	0.144	0.497	1.170	0.047
	Aviation Tons	1903	0.777	0.476	1.270	0.702
	Cars	1885	0.040	3.900	2.490	0.391
Transport	Rail Line Km	1825	0.001	0.210	1.010	-0.093
	Rail Passengers	1825	0.160	3.900	0.235	0.620
	Rail Tons	1825	0.008	0.396	1.610	0.054
	Ships	1776	0.003	0.097	2.540	0.010
	Trucks	1903	0.058	0.416	1.380	0.330
	Cellphone	1973	0.577	0.137	0.000	0.943
	Computer	1973	0.309	0.418	2.520	0.985
	Internet	1983	0.318	0.227	0.000	0.789
Communication	Radio	1920	0.121	1.160	0.280	0.754
	Telegram	1835	0.000	0.096	2.110	-0.001
	Telephone	1876	0.679	0.010	1.640	0.290
	TV	1927	0.028	0.016	0.020	0.623
	ATM	1971	0.001	0.100	0.000	-8.675
	Electricity	1882	0.034	0.449	1.080	0.080
Industry	Steel Bof	1950	0.070	0.145	2.400	0.352
	Steel Eaf	1907	0.139	3.500	0.000	0.362
	Tractors	1903	0.046	0.770	0.205	0.982

Intuitively, heterogeneity in initial adoption helps the model fit the observed pattern of geographic interactions in adoption because it increases the dispersion in the SDT variables. As a result, the model does not need such large coefficients, $\tilde{\beta}_t^j$, to match the average β_{4t}^j. Because the average $\tilde{\beta}_t^j$ is, on average higher now (i.e. less negative), the model does not need to resort to corner solutions ($\alpha = 0$, or very high δ) to match the data. With higher α and lower δ, technologies diffuse and the model generates the patterns of geographic interaction in adoption we observe in the data. That is, $\tilde{\beta}_t^j$ is more negative initially and it increases towards a higher long run level.

Heterogeneity does not help the model improve the fit of ATMs significantly. This is the case because unlike the other technologies, for ATM's the geographic interactions in the data are so large that the predicted $\tilde{\beta}_t^j$ are above the actual series of β_{4t}^j, not below. Not surprisingly, the estimate of σ_x for ATMs is zero.

There is a positive correlation, albeit insignificant,[31] between the estimated σ_x and the standard deviation of initial adoption among followers in CHAT which suggests that, at least in part, the variation in σ_x captures observed initial heterogeneity in adoption among followers. However, there seems to be room for other factors not captured by our simple model to generate cross-country dispersion in SDT. Our estimates of σ_x are likely to reflect, at least in part, these missing dimensions.

6 Conclusions

In this paper we have used a new data set of direct measures of technology to study technology diffusion across time and space. Our findings indicate that understanding technology diffusion over space is crucial to understand the speed of technology diffusion. Countries that are far away from the adoption leaders benefit less rapidly from these technologies. The spatial effects we identify vanish over time. For most technologies this implies that the effect of geography is initially strong, decays over time, and for most technologies it eventually disappears. As far as we know, this is the first paper to document these patterns in adoption rates for a large number of technologies and countries.

The empirical pattern of technology adoption over time and space is well accounted for by a simple model of random interactions. The model determines a pattern of adoption for each technology given two key parameters. The frequency of interactions (governed by α) and the spatial decay in the probability of interactions (given by δ). Our structural estimation of the model provides estimates of these parameters for each technology. These estimates show that interactions are more frequent for more recent technologies. Perhaps more important, our paper provides estimates of structural parameters that can be used to inform spatial theories of growth (as in Eaton and Kortum, 1999, and Desmet and Rossi-Hansberg, 2011). The speed and spatial scope of technology diffusion is a key component to the quantitative implications of these type of theories. Thus, we hope that the evidence on the significance of the spatial and temporal links in technology adoption we estimate proves helpful to stimulate future research in these areas.

[31] The point estimate is 0.12.

7 Appendix A: Robustness to other specifications of SDT

This appendix illustrates the robustness of our basic results to different specifications of the SDT variable. To this end, we compute the technology distance interactions, SDT, using the following two alternative specifications. Define SDT2 and SDT3 as

$$SDT2 = \sum_{\forall k \neq c} \frac{x_{kt}}{d_{ck}},$$

and

$$SDT3 = \log(\sum_{\forall k \neq c} e^{(x_{kt})^{-\gamma d_{ck}}}),$$

where, as above, x_{kt} denotes the log of adoption per capita in country k at time t, and d_{ck} denotes the distance (in thousands of km) between countries c and k, and γ is a parameter that we calibrate. As our baseline specification, these alternatives are also sensible ways to capture the interaction between adoption in other countries and how far they are from c. One important difference is that the presence of geographic interactions in adoption *would lead to positive coefficients* (rather than negative) of $SDT2$ or $SDT3$ on country c technology level.

We use the alternative specification to define SDI for each of these formulations, and run Specification 1 of the pooled regression (1) for each of the three different measures of SDT. Table 14 presents the coefficients of SDT. (We denote our baseline specification $SDT1$.) It is clear that we obtain significant geographic interactions from adoption regardless of the specification. In the context of $SDT3$, we have tried various values of γ and the results are not sensitive to its value.

Table 14: Pooled Regressions

	Specification		
	SDT1	SDT2	SDT3 ($\gamma = 1$)
SDT	*-.0001468**** ($4.50e^{-6}$)	*.0013544**** ($1.04e^{-4}$)	*.2873901**** ($5.01e^{-3}$)
# Obs.	53579	53579	53579

Table 15 reproduces the estimates of the coefficients of the interaction terms when decomposing them between the latitude vs. longitude parts for the two specifications of SDT presented above. Again, it is clear from the table that, as in our baseline specification of SDT, interactions along latitudes have stronger effects on adoption than along longitudes.

Table 15: Longitude and Latitude with Alternative Measures of SDT				
	Specification: SDT2		Specification: SDT3	
	1	2	1	2
SDT NS	*1.87*e^{-6}*****	*2.23*e^{-6}*****	*.1603851****	*.1765056****
	(1.65e^{-7})	(1.78e^{-7})	(1.10e^{-2})	(1.07e^{-2})
SDT EW	*-1.50*e^{-6}*****	*-3.24*e^{-7}	*.1468938****	*.1505348****
	(1.85e^{-7})	(1.95e^{-7})	(5.24e^{-3})	(5.24e^{-6})
SDI NS		7.97e^{-7}		-.0810472***
		(1.15e^{-6})		(1.83e^{-2})
SDI EW		-2.99e^{-6}		.1269351*
		(1.61e^{-6})		(1.36e^{-2})
# Obs.	52731	52731	53579	53579

8 Appendix B: Upper bound of endogenity bias

In the appendix we detail the back-of-the envelope calculations about the impact of the endogeneity of the SDT variable on the estimates of β_4^j. To this end, let's suppose that country $c's$ adoption level increases by one standard deviation (2.46 in the balanced sample). Since the average distance in the balanced sample is 7.5 (thousands km), the SDT of the other countries will increase on average by 18.45. Since the coefficient β_4^j is -0.0007 (from Table 7, Column1), this should lead to an average reduction in the adoption for the 14 countries other than c of 0.0129 ($= -0.0007 * 18.45$). If the average country is 7.5 thousand km's from country c, then these declines in adoption will reduce the SDT of country c by 1.35 ($= 0.0129 * 7.5 * 14$). Since the standard deviation of SDT in the balanced sample is 385, the endogenous increase of SDT represents just 0.35% (= 1.35/385) of the observed dispersion of the independent variable (i.e. SDT).

The small share of the dispersion of SDT generated by its endogeneity limits the magnitude of the bias this has on the estimate of β_4^j. To get an approximate bound on the size of the bias, suppose that SDT can be decomposed between the exogenous (SDT^x) and the endogenous (SDT^n) components as follows:

$$SDT = SDT^x + SDT^n \tag{9}$$

To get a back-of-the-envelope bound on the effect of SDT^n on β_4^j, let's consider a univariate version of regression (1) where adoption (x) is the dependent variable and SDT the independent one. In that case,

$$\hat{\beta}_4^j = \frac{Cov(SDT, x)}{\sigma^2(SDT)}$$

where Cov stands for covariance and $\sigma^2(.)$ is the variance. Using (9) and some straightforward manipulations, β_4^j can be decomposed between the exogenous and the endogenous components as follows:

$$\begin{aligned} \hat{\beta}_4^j &= \overbrace{\frac{Cov(SDT^x, x)}{\sigma^2(SDT)}}^{\text{Exogenous}} + \overbrace{\frac{Cov(SDT^n, x)}{\sigma^2(SDT)}}^{\text{Endogenous}} \\ &= \frac{Cov(SDT^x, x)}{\sigma^2(SDT)} + Corr(SDT^n, x)\frac{\sigma(SDT^n)}{\sigma(SDT)}\frac{\sigma(SDT)}{\sigma(x)} \end{aligned}$$

With the information we have, it is possible to bound the endogenous component (i.e. the second term). $Corr(SDT^n, x)$ must be higher than -1. From our previous calculations, $\frac{\sigma(SDT^n)}{\sigma(SDT)} = 0.0035$. And from the descriptive statistics in the balanced sample, $\frac{\sigma(SDT)}{\sigma(x)} = \frac{2.46}{385} = 0.0064$. Therefore, the endogenous component of $\hat{\beta}_4^j$ is higher than -0.000022. This represents 3% of the estimate we obtain for β_4^j which is -0.0007.

9 References

Almeida, Paul. 1996. "Knowledge Sourcing by Foreign Multinationals: Patent Citation Analysis in the U.S. Semiconductor Industry." *Strategic Management Journal* 17(Winter Special Issue):421-31

Jaffe, Adam, & Manuel Trajtenberg, and Rebecca Henderson. 1993. "Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108(3):577–98.

Anderson, James & Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* 93(1):170–192.

Anderson, James & Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature* 42(3):691–751.

Bandeira, Oriana & Imran Rasul. 2006. "Social Networks and Technology Adoption in Northern Mozambique." *Economic Journal* 116:869–902.

Basu, Susanto. 1996. ""Procyclical Productivity: Increasing Returns or Cyclical Utilization?" *Quarterly Journal of Economics* 111(3):719–751.

Burnside, Craig, Martin Eichenbaum & Sergio Rebelo. 1995. "Capital Utilization and Returns to Scale." in Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Annual.* Cambridge, MA: MIT Press. pp. 67-110.

Comin, Diego & Bart Hobijn. 2004. "Cross-Country Technology Adoption: Making the Theories Face the Facts." *Journal of Monetary Economics* 51(1):39–83.

Comin, Diego & Bart Hobijn. 2009. The CHAT Dataset. NBER WP 15319.

Comin, Diego & Bart Hobijn. 2010. An Exploration of Technology Diffusion." *American Economic Review* 100(5):2031–59.

Comin, Diego, William Easterly & Erick Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics* 2(3):65–97.

Desmet, Klaus & Esteban Rossi-Hansberg. 2011. "Spatial Development." Working Paper, Princeton University.

Diamond, Jared. 1997. *Guns, Germs, and Steel: The Fates of Human Societies.* Norton, W. W. & Company, Inc.

Duranton, Gilles & Puga, Diego. 2004. "Micro-foundations of Urban Agglomeration Economies," in: J. V. Henderson & J. F. Thisse (ed.), *Handbook of Regional and Urban Economics*, Ed. 1, 4, 48, 2063-2117, Elsevier.

Eaton, Jonathan & Samuel Kortum. 1999. "International Technology Diffusion: Theory and Measurement." *International Economic Review* 40(3):537–70.

Foster, Andrew & Mark Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy* 103(6):1176–1209.

Fujita, Masahisa & Jacques-Francois Thisse. 2002. *Economics of Agglomeration.* Cambridge University Press.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25(4):501–522.

Hall, Robert E. & Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?," *Quarterly Journal of Economics*, 114(1): 83-116.

Keller, Wolfgang. 2004. "International Technology Diffusion." *Journal of Economic Literature* 42(3):752–782.

Klenow, Peter J. & Andres Rodriguez-Clare. 1997. "Economic growth: A review essay," *Journal of Monetary Economics*, 40(3): 597-617.

Lucas, Robert & Benjamin Moll. 2011. "Knowledge Growth and the Allocation of Time." NBER WP 17495.

Lucas, Robert. 2009. "Ideas and Growth." *Economica* 76: 1-19.

Maddison, Angus. 2005. "Measuring And Interpreting World Economic Performance 1500-2001." *Review of Income and Wealth* 51(1):1–35.

Mansfield, Edwin. 1961. "Technical Change and the Rate of Imitation." *Econometrica* 29(4):741–766.

Manski, Charles. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60(3):531–542.

Simmons, Beth, Frank Dobbin & Geoffrey Garrett. 2007. *The Global Diffusion of Markets and Democracy.* Cambridge University Press, First edition, Cambridge, UK.

Thompson, Peter & Melanie Fox-Kean. 2005. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review* 95(1):450–60.