Knowledge Diffusion in the Network of International Business Travel

Michele Coscia[1,2,*], Frank Neffke[1], and Ricardo Hausmann[1,3]

[1]Center for International Development, Harvard University, 79 JFK St, Cambridge, MA 02138 USA
[2]ITU Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen, Denmark
[3]Senior Fellow Mastercard Center for Inclusive Growth, 2000 Purchase Street, Purchase, NY 10577 USA
[*]To whom correspondence should be addressed; E-mail: michele_coscia@hks.harvard.edu.

We use aggregated and anonymized information based on international expenditures through corporate payment cards to map the network of global business travel. We combine this network with information on the industrial composition and export baskets of national economies. The business travel network helps predict which economic activities will grow in a country, which new activities will develop and which old activities will be abandoned. In statistical terms, business travel has the most significant impact among a range of bilateral relations between countries, such as trade, foreign direct investments and migration. Moreover, our analysis suggests that this impact is causal: business travel from countries specializing in a specific industry causes growth in that economic activity in the destination country. Our interpretation of this is that business travel helps diffuse knowledge and we use our estimates to assess which countries contribute or benefit most from the knowledge diffusion

through global business travel.

Introduction

Globalization has led to a tremendous increase in international business travel. Its growth has outstripped the growth of the world economy by a wide margin: whereas nominal, USD denominated, global GDP has risen by 0.7% per year between 2011 and 2016 (World Bank data), our data suggest that nominal expenditures related to business travel have grown at an annualized rate of 8.3% over the same period (see Methods section). This growth coincides with tremendous improvements in the availability, quality and costs of long-distance communication technologies. From conference calls to online collaboration platforms, new technologies have made it easier for businesses to connect across the globe without the need for costly and time-consuming travel. So why do we still need to travel so much? What is it that face-to-face interaction on business trips can achieve that other means of communication cannot?

Business scholars have argued that, without face-to-face communication, some knowledge is hard to transmit (*1*). Accordingly, one can think of knowledge as consisting of three components. The first component is knowledge that is codified in production recipes, algorithms, textbooks, blueprints and the like (*2*). This knowledge component consists of know-what and know-why (*3*): knowledge about facts – e.g., the physical dimensions of a product – and about well-understood causal mechanisms, like laws of physics. The second component consists of knowledge that is embedded in physical artifacts, such as machines, tools or intermediate products (*4*). Nowadays, either component can be easily transferred: the costs and speed with which machines, tools and semi-finished products can be shipped has never been lower and, with internet-based technology, code and textbooks can be transmitted almost instantaneously at high fidelity.

The third component of knowledge, however, is stickier. It consists of knowledge of how to expertly carry out certain tasks oneself or of where to find someone who possesses this knowledge. The former is often referred to as know-how, the latter as know-who (*3*). This third component of knowledge resides in people, teams of people and in the relations between these teams (*5*). Large parts of such knowledge cannot be easily articulated by its carriers or would be tremendously costly to codify (*6*), let alone transmitted in digital form. It has therefore been described as tacit (*7*). Although tacit knowledge is typically associated with physical and artisanal skills, sociologists of science have shown that tacit knowledge also plays an important role in science and technology (*8, 9*). Moreover, its transfer, which is indispensal in the training of scientists and engineers, typically involves repeated interaction, imitation and on-the-job training and is often organized in an apprenticeship-like relation between the experienced scientist and her trainee, as evident in common practice in doctoral training programs and artfully illustrated by MacKenzie and Spinardi's (*10*) case study on nuclear weapons design.

Given that in modern economies, knowledge about even the most mundane production technologies is too complex for any single individual to comprehend in full, know-how often needs to be complemented by know-who: knowledge of how and where to access experts in a field. Know-who is particularly important in inter-firm relations, as it helps identify and forge alliances with customers and suppliers (*11*). Like know-how, know-who tends to be more tacit, embedded in people's understanding of the social network that surrounds them and where in this network reliable and trustworthy expertise can be tapped.

In light of this, one plausible explanation for why business travel has not only endured, but expanded, in spite of the increasing availability of substitutes in the form of new communication technologies is that these new technologies are still inadequate when it comes to transmitting tacit know-how or to establishing the trust-based relations associated with know-who. By temporarily relocating know-how through moving the individuals that carry it, business travel enables face-to-face contacts through which tacit know-how can diffuse and trust and social networks can develop.

In this paper, we map the pattern of global business travel and explore how it affects the growth of economic activity. Our underlying hypothesis is that business travel enables the diffusion of know-how and know-who across countries. It is therewith related to prior work that has used business travel as an explanation for bilateral trade links (*12*) (and in Poole's 2010 "Business travel as an input to international trade"), innovation (*13*), and increases in productivity (*14*).

Our hypothesis is that business travel helps to diffuse tacit knowledge. As pointed out by Krugman (*15*), knowledge flows are notoriously hard to quantify: "Knowledge flows [...] are invisible; they leave no paper trail by which they may be measured and tracked." Subsequent authors accepted the implicit challenge, uncovering the imprint that knowledge flows leave in patent citations (*16*) or in academic collaborations (*17*). This work has shown that geographical distance is still a formidable impediment to the flow of knowledge and that knowledge diffuses first locally and only later across longer distances. Moreover, when knowledge diffuses spatially, it does so primarily through the social networks of mobile skilled individuals (*18, 19*). By analyzing patents and scientific publications, these studies focus on knowledge diffusion in the exceptionally knowledge-intensive domains of innovation and scientific progress. However, academic debates often operate at some distance from the economy and patents shed light on only a subset of industries, mostly in the manufacturing sector. Moreover, patents and academic publications mainly reflect codified knowledge, not the tacit know-how and know-who that we hypothesize is transferred through business travel. Research efforts based on patents and scientific publications are therefore constrained in their scope compared to the totality of the global economy.

Yet, business travel flows are not necessarily a reflection of knowledge flows. The concern that Krugman raised is still valid: by its very nature, it is difficult, if not impossible, to observe tacit knowledge directly. However, its presence will be manifest in what an economy is able to produce (*20*). Consequently, knowledge flows should reveal themselves in the way they help economies grow and diversify. Therefore, instead of trying to measure knowledge flows themselves, we use a strategy previously employed in (*21*) and look for indirect evidence for the existence of these flows by studying how business travel relates to changes in the economic structure of a country. In particular, if business travel helps transfer tacit know-how, we should observe that countries grow in economic activities that coincide with the economic specializations of the places from which they receive business travelers.

To test this hypothesis, we use aggregated and anonymized data made accessible for the duration of the research by the Mastercard Center for Inclusive Growth ("the Center"). These data shed light on corporate credit or debit card (henceforth: "card" or "corporate card") foreign spend in business travel for the period 2011-2016.

Credit and other payment card data have previously been used to shed light on socioeconomic phenomena (*22–24*). Typically, these studies are restricted to expenditures in a single city or country. In contrast, the aggregated and anonymized data underlying our study relate to foreign expenditures in 127 different countries.

Each observation in our data consists of the number of trips from a country of origin to a country of destination in a given year. That is, it counts the unique number of cards that were issued in the country of origin and that made payments in the country of destination. For the purpose of this research we refer to these card counts as the number of "business travelers" – which in no way relates to identifiable individuals. Because the only cards considered herein are issued to a company or employer, foreign spend associated with them generally reflects work-related travel. We use these data to derive a business travel network that connects 127 different countries. These travel flows turn out to be representative of aggregate and bilateral flows recorded in official statistics (see Sec. S3).

To infer the most likely industrial content of the know-how that is transmitted through this network, we use information on the industrial composition of national economies. To avoid contaminating the business travel information with priors taken from other sources, we apply the principle of maximum ignorance and assume that business travelers are drawn at random from the population of establishments in their country of origin. Subsequently, we use this estimate to predict how national economies change their mix of economic activities: which activities grow? which new activities emerge? and which old activities disappear? We find that predictions based on business travel compare favorably to the ones based on alternative bilateral relations between countries, such as migrant stocks, foreign direct investments (FDI), trade, and similarities in countries' histories, cultures or languages. More importantly, we use instrumental variables estimation to show that the estimated effects of business travel on economic growth can plausibly be considered causal.

Results

Mapping business travel

Expenditures related to international business travel have expanded markedly: according to our data, the value of these expenditures rose by 47% between 2011 and 2016 and the number of travelers went up by 38% over the same period. Among the fastest growing origins of business travel were China and India; rapidly growing destinations are found in eastern Europe and central America (see Methods section). In spite of the explosive growth in the volume of business travel, its network structure has remained remarkably stable (see Sec. S4). To determine this, for each year, we collect the number of business travelers between each country of origin and each country of destination in yearly matrices, B_t, where rows represent origins and columns destinations. The correlation of the log-transformed elements of these matrices is $\rho(9, 485) = 0.96$ ($p < 0.001$, 95% Confidence Interval = [0.958-0.962]) when comparing consecutive years, (B_t, B_{t+1}), and still reaches $\rho(9,485) = 0.90$ ($p < 0.001$, 95% Confidence Interval = [0.896-0.904]) when comparing flows in 2011 to flows in 2016. To avoid observations with $\log(0)$, we add 1 to each element before taking logs.

To minimize noise, we average the matrices across years into a matrix $\bar{B}$. This business travel matrix is sparse: 43% of its elements are equal to zero and 65% of all business travelers can be accounted for by country pairs in the top 1% of bilateral flows. The sparse structure suggests analyzing this matrix as a complex network (*25, 26*). To do so, we first extract a noise-corrected backbone that retains only statistically significant links using (*27*). In practice, this means that we only retain links with travel intensities that have a below 1% likelihood of having emerged at random. The result is a business travel network that contains 667 of the original 9,529 nonzero links, depicted in Fig. 1A.

The business travel network is far from homogeneous. Its edges can to some extent be explained by a simple gravity model: countries are strongly connected if they are geographically close and if they are large in terms of their Gross Domestic Products (GDP). In a simple log-linear model, geographical distance and origin and destination GDP can together account for more than half the variance in business travel flows ($R^2 = 60.7\%$, F-statistic (3,112) = 189.8, $p < 0.001$). However, business travel is also associated with several bilateral relations between countries. Fig. 1B reports the statistical association between the volume of business travel and a variety of bilateral links between countries: foreign direct investment (FDI), cross-ownership of corporations (equity links), trade, migration and sharing a colonial history. If we use all of these covariates simultaneously in an ordinary least squares (OLS) regression to predict business travel, we obtain an $R^2 = 73.9\%$ (F-statistic(9,112) = 304.9, $p < 0.001$) (Fig. 1C, see also Sec. S4). Most of the variance in business travel is explained by equity and FDI relations (46.4%). The remaining factors – including geography and GDP – explain 22.4% of the variation, with 26.1% left unexplained.

The Travel Quantity Index

Our main question of interest is whether business travel can explain the growth of economic activities in a country. To explore this, we need to assign business travelers to the industry they are most likely to represent. We do this by decomposing the business travel flows that emanate from a country into industry-specific subflows. Given that we have no information on the industry from which a flow originates, we apply the principle of maximum ignorance and assume that any establishment in a country of origin is equally likely to dispatch a business traveler. That is, we assume that a business traveler that is observed to travel from country o to country d is equally likely to come from any of the establishments in country o. Consequently, the expected size of the business travel flow from a particular industry in a country is proportional to the industry's number of establishments in that country.

To be precise, consider an activity matrix, A, with elements A_{ci} that record for each country $c \in C$ (be it origin or destination) – with C the set of countries in our dataset – the share of establishments classified under industry i. For instance, in a stylized example where 50% of Germany's economic establishments produce cars, the corresponding entry in this matrix is 0.5. Furthermore, consider a business travel matrix, B, with elements B_{od} that record the number of business travelers from country of origin o that visit country of destination d. We now estimate the number of business travelers from an industry i that visit country d by $Q_{di} = \sum\limits_{o \in C} A_{oi} B_{od}$. In other words, we guess a traveler's industrial origins by assuming that all establishments in o are equally likely to have sent the traveler. We refer to the elements in matrix Q as the Travel Quantity Index (TQI).

Because we will assess how business travel relates to industrial growth between 2011 and 2016, we use the business travel matrix of the base year 2011. Moreover, we restrict industries to industries that produce tradable products, which can export their products beyond the national market, and drop natural-resource based industries (see Sec. S2). Both restrictions ensure that the industries we focus on are not limited in their location choice by the presence of natural resources or of a large local market.

Fig. 2A illustrates this operation: the total flow of business travelers from a country of origin (Germany or the US) to a country of destination (Austria, Italy or Portugal) in 2011 is split into three industry-specific subflows – one for cars, one for chemicals and one for agriculture – according to each industry's share of establishments in the economy of the country of origin. In our earlier example, if Germany sends six business travelers to Austria, three of them (6×0.5) are assumed to represent car manufacturers.

Note that this approach does not aim for the most accurate estimate of a business traveler's industry affiliation. In fact, we could improve our estimate by adding, for instance, industry-specific information on FDI or trade flows. However, the resulting industry-specific business travel flows would now mix information on business travel with information from these other flows. This would make it impossible to unambiguously associate any subsequently estimated parameters with business travel per se, as opposed to, in the examples above, FDI or trade.

The end result is an estimate of which countries will have access to which industries' knowledge bases. To illustrate this, Fig. 2B depicts the degree to which an industry is estimated to be overexpressed in the incoming business travel flow of China, Mexico and Germany. China is estimated to have easy access to manufacturing know-how, Mexico to know-how related to business services and Germany to know-how of a more balanced array of industries.

The differences between these three countries are driven by the fact that China, Mexico and Germany are connected to countries with vastly different industrial specialization patterns. Take for instance the three most overexpressed industries in the business travel flows to China: machinery for paper making, carbon paper and inked ribbons, and foreign bank branches. These industries represent the top 3 economic activities in which the neighboring Republic of Korea specializes. Likewise, Mexico's top 3 can be explained by the specialization pattern of the

US economy (ranks 1, 2 and 7), whereas Germany's top 3 consists of the top specialization of Poland, the Czech Republic and Denmark. To calculate how specialized a country is in an industry we use the number of establishments in country c and industry i, P_{ic}. To be precise, we define the degree of specialization as: $\frac{P_{ic}/\sum\limits_{i'\in I} P_{i'c}}{\sum\limits_{c'} P_{ic'}/\sum\limits_{i''\in I, c''\in C} P_{i''c''}}$.

Does TQI contain information on international knowledge flows? To assess this, we examine whether these flows leave traces in the transformation of the industrial composition of the recipient country's economy. In particular, we explore if the TQI predicts how fast different industries will grow in a country. In these analyses, we drop small countries with a population below 2.5 million inhabitants from the data, leaving us with a sample of 96 countries. Results for the full set of countries are shown in Sec. S6.

Tab. 1 reports outcomes of OLS regressions with different sets of control variables. Observations are composed of industry-country pairs. The first model shows the effect of only business travel and a mean-reversion term. In the second model, we add variables that control for the global size of the industry and the size and wealth of the country, all of which will affect both, business travel flows and economic growth. In the third model, we replace these control variables by industry and country fixed effects. These fixed effects will control nonparametrically for any industry or country specific confounding factors.

In all regression models, a higher TQI is associated with a higher growth rate. Our preferred model is model 3, which controls for any idiosyncrasies that affect growth rates at the level of the country or of the industry. The point estimates can be interpreted as elasticities: in our preferred model, a 10% increase in the number of incoming business travelers is associated with a .6$\%$ (t-statistic(95) = 2.86, $p = .005$, effect size = 0.06, 95$\%$ Confidence Interval = [.02-.10]) increase in an industry's growth. Note that the estimates in Tab. 1 are limited to industries that already existed in the country. However, in Sec. S6, we show that business travel is also associated with increased entry rates of new industries.

However, as shown in Fig. 1B, business travel is also correlated with a number of alternative bilateral relations between countries. To assess the predictive performance of the TQI, relative to these other relations, we construct a series of XQI indicators using a procedure identical to the one that generated TQI, while replacing the business travel matrix, B, by matrices that capture other bilateral relations. When estimating the effects of each indicator separately, TQI features the highest estimated elasticity. When modeled jointly, the variables are too collinear to allow for the inclusion of country and industry fixed effects. However, if we instead use the specification of model 2 in Tab. 1, TQI remains statistically significant at $p < .05$ (t-statistic(93) = 2.26, p = $.025$, effect size = 0.08, 95% Confidence Interval = [$.01$-$.14$]) and outperforms all alternative indicators (Sec. S6).

Does the level of industrial aggregation matter? The models in Tab. 1 assume that there are no spill-overs from business travel across industries. In other words, business travelers from a narrowly defined industry such as "Envelopes" (SIC 2677) would only affect the growth rate of, in this case, envelope-making, but not of the manufacturing of other stationary products. We can relax this assumption by gradually aggregating industries into higher-level sectors, such that the estimated coefficients internalize more of the potential spillovers between narrowly defined industries. In Sec. S6.2, we show that, as the industry classification coarsens, the point estimates for the effect of TQI become successively stronger. This supports the idea that knowledge that is diffused by business travel may be shared beyond the narrowly defined industries of Tab. 1.

Finally, we test how robust our results are when describing national economies, not in terms of the size of their industries, but in terms of the product-specific volumes of their exports. To do so, we replace the data on economic establishments by data on global trade patterns (Sec. S6.3). This exercise shows that TQI not only predicts the evolution of the industry mix of a country's economy, but also the change in the country's export profile ($p < 0.001$). That is, business travel associated with a certain export product predicts the emergence and growth of exports of this product in recipient countries.

Together, these analyses document a robust statistical association between business travel and the transformation of a country's economic activity mix. How should we interpret this finding? It is important to note that the uncovered statistical associations do not necessarily mean that business travel causes economic growth. For instance, business travel may be just one among a number of other steps firms undertake to find new suppliers or to establish new branch plants abroad. Moreover, economic growth itself may attract business travelers, suggesting that the direction of causation may be reversed.

To probe into this question more deeply, we apply instrumental variables (IV) estimation. Note that TQI_{id} represents a weighted sum of industry i's shares in the economies of origin countries. It thus essentially represents the weighted size of the industry in the global economy, where weights are given by origin countries' business travel flow to the destination country, d. Because business travel may both cause and be caused by the growth prospects of industry i in the destination, these weights may be endogenous. Therefore, we instrument TQI, using information on bilateral visa regimes. Given that visa-related travel restrictions only affect the flow of people, this provides plausible exogenous variation in the business travel volume between countries. A detailed description of the construction of these instruments can be found in Sec. S7.

The IV regressions yield causal effects that are, if anything, even somewhat larger than the effects derived from OLS regressions. However, it is unlikely that business travel is an ultimate cause of growth: we would not expect that randomly dispatched business travelers would, by themselves, spur much growth in recipient countries. Instead, business travel is more likely to enable economic growth by facilitating other interactions between countries, such as FDI and trade. In line with this, we find that business travel has a strong causal effect on the intensity of bilateral FDI, equity and trade (see Sec. S7). In other words, trade and investment relations are typically accompanied by business travel (see Fig. 1B) and depend on it for their success. This, in turn, suggests that international travel helps initiate or maintain such relations. Given the substantial monetary and time expenditures associated with business travel, firms apparently put a high value on the face-to-face interactions that travel buys them. In light of this, the causal effect of business travel on the growth of economic activities related to such travel corroborates our hypothesis that business travel helps transfer know-how and know-who. The mechanism through which business travel achieves this is by enabling successful investment and trade links.

Know-how Index

How valuable is global business travel? To provide an estimate of this, we ask two questions. First, how many jobs in a given economy can we attribute to incoming business travel? Second, how much value do these jobs create?

To answer these questions, we will have to make some drastically simplifying assumptions. First, we assume that the effects of business travel on job creation are causal. The IV models in Sec. S7 warrant such an assumption. However, in this section we will use OLS estimates, which are more precisely estimated – especially in the presence of country and industry fixed effects – and more conservative than the IV estimates. To be precise, we will fit a regression model that predicts the employment of each industry in a country. Next, using the fitted parameters of this model, we estimate how much higher or lower the employment in an industry would have been, had the country received some counterfactual number of business travelers. Second, we use information on industry-differentials in establishments' sales per worker as a proxy for differentials in productivity per worker. Unfortunately, in our data set this information is unreliably captured for establishments outside the US. Therefore, we proceed in two steps. First, we estimate the average sales per worker in an industry using US figures only. Next, we correct for productivity differences between the country and the US by multiplying this number by the ratio of the country's GDP per capita to the US GDP per capita. A country's predicted output is now simply the predicted number of jobs in an industry, multiplied by the predicted productivity of workers in these jobs, summed across all industries. Note that the accuracy of this calculation will depend on the correlation between value added and sales in an industry and on how similar the inter-industry productivity ratios in a country are to the ones observed in the US.

We use this procedure to construct two indices. First, we predict a destination country's output, once with its current inflow of business travel and once with the expected inflow, had business travel been distributed proportionally according to countries' populations and industries' global employment sizes. The ratio of these two predictions answers the question: How much larger or smaller would a country's economy have been, had it received its "fair share" of business travel? We convert this into a percentage by subtracting 1 from the ratio and multiplying the result by 100. We call the resulting quantity the Incoming Know-how Index (IKI).

Second, for each country of origin, we compare the predicted size of the global economy with and without the country's business travelers. This tells us by how much we would predict the world economy to shrink, were a country to withdraw from global business travel and stop sending travelers abroad. Once again, we convert the ratio of the two predictions into a percentage and call the resulting quantity the Outgoing Know-how Index (OKI). Details of the calculations, as well as full rankings by country are provided in Sec. S7.

Fig. 3 plots the distribution of both indices across countries. The predicted benefits of business travel (3A) are distributed unequally across the world. Moreover, they tend to favor rich countries and their neighbors. This is unsurprising, given that business travel follows a law of gravity: it tends to connect each country to nearby large economies. The origins of global know-how flows are even more concentrated. They are dominated by a small number of wealthy countries, like Germany, Canada, the US, the UK and the Republic of Korea. These countries not only send the largest number of business travelers abroad, they also tend to send them from productive industries and to other advanced economies. As a result, they are not only predicted to create a large number of jobs, but also jobs in highly productive activities and in highly productive economies.

Fig. 4 illustrates this point by showing how predicted output losses associated with the withdrawal of specific business travel origins are felt across the globe. The upper row compares the withdrawal of Germany, the top contributor to OKI, to the withdrawal of the US. Whereas US business travel concentrates in North and South America, German business travel is more spread out. By contrast, a withdrawal of Japan or the Republic of Korea has quite similar predicted impacts, mostly concentrated in East Asia. Finally, the comparison of Spain's and Portugal's predicted contributions to the global diffusion of know-how shows the importance of speaking the same language in business travel: whereas, outside Europe, Spanish business travelers frequent the Spanish speaking parts of Latin America, the most notable destinations for Portuguese business people are Brazil and Angola. Further explorations of individual countries' contribution to the OKI can be accessed through the interactive visualizations available at `https://growthlab.cid.harvard.edu/academic-research/business-travel`.

Discussion

In spite of improving communication technologies, firms still rely heavily on business travel to establish and maintain the international links that keep the global economy connected. Business travel has an important impact on economic growth: countries are more likely to expand existing and enter new industries if these industries are already well-developed in places from which they receive most of their business travelers. However, the causal effect of business travel on economic growth materializes through a variety of economic mechanisms, such as international trade and investment. The dependence of trade and investment on business travel suggests that business travelers help establish and grow relations in ways that other forms of communication cannot. This corroborates the hypothesis that countries rely on business travel to connect to know-how that exists elsewhere in the world.

A different but complementary explanation of our findings is that business travel helps businesses to identify and connect to a local network of suppliers and customers. In this case, business travel helps establish trust and share know-who embedded in personal relations rather than know-how. We leave the question of which of these explanations provides the dominant rationale for business travel for future research.

There are a number of limitations to this study. First, we measure business-travel mediated knowledge spillovers only indirectly, through the relation between the growth of industries in a destination and the most likely industrial origins of the arriving business travelers. Second, due to the nature of our data, the estimated business travel network should be regarded as a rough approximation of actual business travel connections. For instance, lodging and local transportation expenses are increasingly paid through online booking and ride-sharing platforms and therefore not captured in a traveler's expenditures abroad. Similarly, our data predominantly capture the know-how and know-who shared by corporations whose employees routinely use corporate payment cards. To the extent that business travel initiated by other companies or of employees without such cards differs, its effect will not be adequately captured in our business travel metrics. Third, the spatial granularity of the business flows, as well as their industrial origins, could be improved. Knowing which sectors business travelers work in and between which cities (as opposed to countries) they travel would enrich future analyses.

However, a robust conclusion is that the network of business travel is far from egalitarian, favoring a small number of already well-developed economies and their geographical neighbors. This has important consequences for which countries contribute most to the global diffusion of knowledge and which are best positioned to benefit from it. Finally, although the business travel network is stable in the period of our study, over longer time horizons, it is likely to change. Moreover, the locus of the global technological frontier may shift, as it has repeatedly throughout human history. The evolving business travel network may offer exciting ways to analyze how such shifts percolate through the social network and face-to-face contacts between firms. Further analysis of this network may therefore yield new insights on the development path of the world economy.

Methods

The paper uses two main datasets: an aggregated and anonymized data set relating to corporate payment card ("card") foreign spend, a dataset containing information on the location of economic establishments across the world and data on global exports. The first aggregated and anonymized dataset was made accessible, subject to strict privacy and data protection safeguards, by the Mastercard Center for Inclusive Growth (henceforth referred to as "the Center"), while the second was purchased from Dun & Bradstreet. Due to the proprietary nature of the data, we cannot make them publicly available. The third dataset is based on United Nations Comtrade data, which can be downloaded here: `http://atlas.cid.harvard.edu/engage#data-download`. In this section, we detail the characteristics of these datasets and how we clean them to build the business travel network and vectors that describe the industrial composition of national economies.

Data on indexed foreign expenditures. The business travel data are based on indexed expenditures by corporate payment cards in countries other than where the cards were issued. That is, they relate to expenditures in one country made by cards issued in another country. Below, we will refer to a country-of-origin×country-of-destination×year combination as a data "cell". In data cells with less than a certain minimum threshold of recorded cards, the exact number of card holders had been replaced by ranges (see below) – as a privacy and data protection protocol. Moreover, the dollar value of expenditures was indexed, i.e., expressed in a unit-less index. However, this index retains the relative differences between cells across years and countries of origin as well as destination.

We accessed these data in aggregated form at the country-of-origin by country-of-destination level and inflated to correct for variations in the data collaborator's (undisclosed) market shares. Because of this adjustment, the data reflect not just the indexed expenditures by payment cards related to our data collaborator's share of the market, but the indexed expenditures in the entire payment market.

Note that the aforementioned indexed and aggregated expenditures include withdrawals by, for instance, debit cards from local ATMs. This means that business travel to predominantly cash-based economies need not be unreliably captured, as long as business travelers rely on ATMs to withdraw local currency. Furthermore, the data are not necessarily limited to the data collaborator's own payment network, but include international expenditures that were cleared by it for other card networks. However, expenditures by business travelers who use personal instead of corporate cards will not be included. Therefore, it is important to check how representative our estimates of global business travel compared to alternative sources on this phenomenon. We provide a detailed analysis of this in Sec. S2. Moreover, the estimated business travel flows are necessarily imperfect. Therefore, our business travel variable will be measured with a certain amount of error. We will come back to this when discussing causal estimation in Sec. S7.

Country List. The first cleaning operation involves the creation of a list of countries of interest. The Center provided access to information on payment cards issued in 135 countries and territories (henceforth: countries), although indexed expenditures are recorded in 236 countries.

We were allowed access to the data for one-year intervals, between 2011 and 2016. Foreign indexed expenditures by corporate payment cards reflect the footprint of business travel that forms the basis of our analysis. We focus this analysis on the 135 countries that are observed both as an origin and as a destination of such international indexed expenditures. Henceforth, we will refer to this dataset on business travel as the BT data.

The second dataset is taken from the Dun & Bradstreet index of productive establishments ("D&B"). The dataset lists over 120 million economic establishments across the world. For each establishment, it records the corporate parent (if any), an estimate of the number of employees and up to six economic activities ("industries"). We acquired the D&B dataset for the years 2011 and 2016. For the latter year, the data also contain information on the dollar value of the establishments sales ("output"). However, output figures are only consistently available for US establishments. In other countries, this variable is populated more erratically. Of the 135 origin/destination countries in the BT data, seven do not appear in the Dun & Bradstreet data. The seven missing codes are:

- Puerto Rico, US Virgin Islands: aggregated to the United States in the D&B dataset.
- Mongolia: not covered by D&B.
- Montenegro: aggregated to Serbia in the D&B dataset.
- Netherlands Antilles: converted into Curaćao.
- Turkmenistan: no business travel data before 2015.
- QZZ, ZAR: codes are used by the Center, but they are either outdated or custom-defined.

As a result, our final dataset includes 127 countries. Supplementary Table 1 reports the full list. Out of these 127 countries, 96 meet the 2.5 million population threshold for the main analyses. This list includes most of the largest economies in the world and accounts for 87% of the world's population and 97% of global GDP.

Imputing Censored Cells. Our unit of observation – henceforth, "cell" – is a combination of a country of origin, a country of destination, and a year. As a privacy and data protection control, the Center provided us with an exact card count only in cells with more than a minimum threshold of cards. For cells below this threshold, we just know that the number of cards falls into one of five of predefined intervals. To still be able to use these cells in our analysis, we impute an expected number of cards for each censored interval. We use this estimate everywhere where data are censored.

The distribution of the number of cards per cell is heavily skewed (Supplementary Figure 1(left)). As a consequence, using the midpoint of each interval as an estimate would not be optimal. The number-of-cards distribution is broad and exhibits a power-law with exponential cutoff. This makes it challenging to find an accurate fit of the entire distribution of cells. However, because we are only interested in filling in the censored cells, which are located in the head of the distribution, we can limit ourselves to small cells where the power law approximately holds.

Supplementary Figure 1(right) shows the cumulative distribution function of this subsection of the data. Note that in the censored range, we only know the cumulative number of cards at the end of each interval. It is easy to find a log-linear curve that accurately fits this distribution ($R^2 \sim 99.9\%$, F-statistic (1,27): 29270 , $p < 0.001$). We use this fitted curve to interpolate how many cells will contain any of the integer-valued number of cards – ranging from 1 to the start of the uncensored section of the distribution – and estimate the weighted average number of cards in each of the censored intervals.

Imputing Long-Stay Cards. One complication of using corporate payment cards to understand business travel is that card holders may complete transactions en route to their final destination. To correct for this, we aim to base our business travel network only on cards that stayed in the country of destination for more than one day. On average, two out of three cards stayed only one day in their destination, meaning they were en route. This is high, but not implausible. For instance, consider the case of Dubai, one of the largest transit hubs in the world connecting Europe to Asia. In our data, 63% of trips to the Arab Emirates only stayed one day, which we interpret as that they were connecting to go somewhere else. The real number of connecting travelers may be somewhat higher, given that the legs of connecting flights may fall in two consecutive days. Counting all those travelers as business people visiting the Arab Emirates would clearly be problematic.

Due to privacy and data protection requirements, for a number of countries, we cannot determine how many cards stayed for more than one day. Rather than dropping these countries from the analysis – the list is long and includes countries that significantly contribute to the world's economy such as the Republic of Korea, Austria, Israel, Australia, Turkey, and Russia – we decide to impute the most likely number of cards that stay for more than a day from available data. Note that the number of rows we have to impute is high – 55% – but they contain relatively little traffic, just 13% of the total number of trips.

To do so, we predict the number of cards staying for more than one day from the total number of cards in the cell using a using a regression model in which the dependent and independent variables first undergo a logarithmic transformation. We fit this model for countries for which the length-of-stay information is available. The estimated parameter is close to 1 with $R^2 \sim 95\%$ (F-statistic (1,38653): 807390, $p < 0.001$). Given this excellent fit, we use this model to predict out-of-sample how many cards stayed for more than one day in the other countries.

After making these corrections, we map the top origins and destinations of international business travel in our sample of 127 countries (Supplementary Fig. 2). Moreover, Supplementary Fig. 3 shows the growth of global business travel in terms of indexed expenditures and number of travelers in our data and compares this to the growth in global GDP as reported by the World Bank. While the indexed business travel expenditures grew by 49% and the number of travelers by 38% between 2011 and 2016, global GDP had grown by just 3.5% in the same period.

D&B Data. To associate business travel to the growth of industries in a country of destination, we rely on the Dun & Bradstreet global industrial index ("D&B data", `http://www.dnb.com/`) retrieved in 2011 and 2016. D&B data are derived from official government sources and direct investigation by D&B. They have worldwide coverage and have been widely used in research on FDI (*28*). Ultimately, we will aggregate these data to the country-industry level but here we describe some of the characteristics of the unaggregated data set.

For each establishment, the D&B data record their location, up to six industry codes ("SIC codes"), as well as an estimate of the number of employees and their corporate parent organization. In the 2016 data, we also have for some establishment an estimate of the value of the establishments' sales output. Industries are coded at the four-digit level of the SIC87 industrial classification system, which distinguishes among 1,005 industry codes. We consider only the main activity of the establishment, namely the first SIC code.

We believe that this main industry code captures most of the operational know-how embedded in the establishment. Moreover, the average number of SIC codes per establishment is low. Supplementary Fig. 4 shows the distribution of the number of SIC codes per establishment: while over 100 million establishments have only one SIC code, fewer than 20 millions have two.

For each establishment, D&B provides either an estimate of the number of employees working in that particular physical location. Supplementary Fig. 5(left) shows that the distribution of the number of employees per establishment is broad, spanning five orders of magnitude. For the year 2016, we also have an estimate of the total sales value produced in an establishment – Supplementary Fig. 5(right). This allows us to calculate the establishment's productivity. However, sales information is mostly missing or less reliable outside the United States: Supplementary Figure 5(right) shows that the mode of the distribution is zero. For this reason, we calculate an industry's average productivity using exclusively US establishments.

Data Availability

The aggregated and anonymized dataset made accessible by the Mastercard Center for Inclusive Growth is available for the duration of this research, after which this dataset and any existing copies will be permanently destroyed. The aggregated data used in our growth estimations, along with the data on economic establishments (from Dun & Bradstreet), are provided as a Supplementary Data zip file, hosted by the Mastercard Center for Inclusive Growth, which can be reached at the URL: `https://growthlab.cid.harvard.edu/academic-research/business-travel`. This archive contains the processed business travel network as shown in Fig. 1, countries' trade profiles as well as the number of establishments, employment and estimated productivity by industry and country. We added random noise to the business travel information to preserve confidential information held by the data providers. The remaining publicly available datasets can be downloaded at `http://atlas.cid.harvard.edu/engage#data-download` and at `http://www.michelecoscia.com/?page_id=1612`.

References

1. Nonaka, I. & Takeuchi, H. *The knowledge-creating company: How Japanese companies create the dynamics of innovation* (Oxford university press, 1995).

2. Auerswald, P. E. *The code economy: a forty-thousand year history* (Oxford University Press, 2017).

3. Johnson, B., Lorenz, E. & Lundvall, B.-Å. Why all this fuss about codified and tacit knowledge? *Industrial and corporate change* **11**, 245–262 (2002).

4. Arthur, W. B. *The nature of technology: What it is and how it evolves* (Simon and Schuster, 2009).

5. Kogut, B. & Zander, U. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization science* **3**, 383–397 (1992).

6. Cowan, R., David, P. A. & Foray, D. The explicit economics of knowledge codification and tacitness. *Industrial and corporate change* **9**, 211–253 (2000).

7. Polanyi, M. *The tacit dimension* (University of Chicago press, 1967).

8. Collins, H. M. & Pinch, T. *The golem: What you should know about science* (Cambridge University Press, 1998).

9. Collins, H. & Pinch, T. *The golem at large: What you should know about technology* (Cambridge University Press, 1999).

10. MacKenzie, D. & Spinardi, G. Tacit knowledge, weapons design, and the uninvention of nuclear weapons. *American journal of sociology* **101**, 44–99 (1995).

11. Lissoni, F. Knowledge codification and the geography of innovation: the case of brescia mechanical cluster. *Research policy* **30**, 1479–1500 (2001).

12. Cristea, A. D. Buyer-seller relationships in international trade: Evidence from us states' exports and business-class travel. *Journal of International Economics* **84**, 207–220 (2011).

13. Hovhannisyan, N. & Keller, W. International business travel: an engine of innovation? *Journal of Economic Growth* **20**, 75–104 (2015).

14. Piva, M., Tani, M. & Vivarelli, M. Labour mobility through business visits as a way to foster productivity. *Labour* 004 (2017).

15. Krugman, P. R. *Geography and trade* (MIT press, 1993).

16. Jaffe, A. B., Trajtenberg, M. & Henderson, R. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics* **108**, 577–598 (1993).

17. Hoekman, J., Frenken, K. & Van Oort, F. The geography of collaborative knowledge production in europe. *The Annals of Regional Science* **43**, 721–738 (2009).

18. Breschi, S., Lissoni, F. *et al. Mobility and social networks: Localised knowledge spillovers revisited* (Università commerciale Luigi Bocconi, 2003).

19. Agrawal, A., Cockburn, I. & McHale, J. Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography* **6**, 571–591 (2006).

20. Hausmann, R. *et al. The atlas of economic complexity: Mapping paths to prosperity* (Mit Press, 2014).

21. Bahar, D., Hausmann, R. & Hidalgo, C. A. Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics* **92**, 111–123 (2014).

22. Lenormand, M. *et al.* Influence of sociodemographic characteristics on human mobility. *Scientific reports* **5**, 10075 (2015).

23. Sobolevsky, S. *et al.* Cities through the prism of people's spending behavior. *PloS one* **11**, e0146291 (2016).

24. Di Clemente, R. *et al.* Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nature communications* **9** (2018).

25. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics* **74**, 47 (2002).

26. Newman, M. E. The structure and function of complex networks. *SIAM review* **45**, 167–256 (2003).

27. Coscia, M. & Neffke, F. M. Network backboning with noisy data. In *Data Engineering (ICDE), 2017 IEEE 33rd International Conference on*, 425–436 (IEEE, 2017).

28. Alfaro, L. & Charlton, A. Intra-industry foreign direct investment. *American Economic Review* **99**, 2096–2119 (2009).

Acknowledgments

The authors gratefully acknowledge their data collaboration with the Mastercard Center for Inclusive Growth (the Center), which generously provided access to aggregated and anonymized data – for the duration of this research – on expenditures through foreign corporate cards. We would like to thank the following people for valuable comments and material help in preparing the manuscript: Anna Stansbury, Dany Bahar, Bruno Zuccolo, and Sid Ravinutala. The authors received no specific funding for this work.

Author Contributions

M.C. and F.N. collected the data. M.C. and F.N. developed the model, conducted analyses and wrote the manuscript. All three authors (M.C., F.N., and R.H.) conceived and designed the study, reviewed, and approved the paper.

Competing interests

The Mastercard Center for Inclusive Growth provided access, subject to strict privacy and data protection safeguards, to an aggregated and anonymized dataset relating to corporate credit card foreign spend. The Center reviewed the paper to ensure that it complied with these privacy and data protection safeguards. However, the Center did not determine any parts of the design, execution or interpretation of the research: all opinions, findings, and conclusions or recommendations expressed in this paper and its supplementary materials are those of the authors and do not necessarily reflect the views or opinions of Mastercard. Ricardo Hausmann is a senior fellow at the advisory council of the Mastercard Center for Inclusive Growth. The authors declare no conflict of interest.

Figure Legends

Figure 1: The topology of international business travel. (A) Network representation of business travel structure. Link width is proportional to the number of business travelers between countries. The arrow represents the direction of the flow. Link color is proportional to the statistical significance of the connection when observed links are set against a random benchmark in which flows are proportional to nodes' total in- and outflows of business travelers. Node size is proportional to the country's economic size (GDP). Node color represents the world region in which the country is located. (B) Scattergrams representing the relationship between the number of business travelers (always on the y axis) and six possible explanations for business travel (top to bottom, left to right): geographical distance (red), migration (blue), foreign direct greenfield investments (green), equity links in number of employees (purple) and establishments (brown), and trade (orange). (C) The contributors to the R^2 ($R^2 = 73.9\%$, F-statistic (9,112) = 304.9, $p < 0.001$) of a regression explaining the network's connectivity by each of those elements, as well as by the GDP of the country of origin and destination.

Figure 2: A: Visual representation of the Travel Quantity Index calculation. The color of the flows first represents the country of origin, then the industry of origin. The width of the flow is proportional to its intensity. B: Overrepresentation of business travelers by industry for China, Mexico and Germany. Overrepresentation of industry i in country of destination d is calculated as the share of business travelers attributed to industry i as a percentage of the total inflow of business travelers in country d, divided by the industry's share in global business travel: $\frac{TQI_{id}/\sum_{i' \in I} TQI_{i'd}}{\sum_{d'} TQI_{id'}/\sum_{i'' \in I, d'' \in C} TQI_{i''d''}}$.

Figure 3: Know-how Index. A: Incoming: Map depicts the percentage by which a country's

economy is predicted to be larger or smaller than what we would have predicted, had its industries received their proportional share of incoming business travel. B: Outgoing: Map depicts the percentage by which the world economy would be predicted to shrink, were the country to stop sending business travelers abroad. (Made with Natural Earth)

Figure 4: Distribution of predicted know-how losses associated with the withdrawal of specific countries. (A) USA, (B) Germany, (C) Japan, (D) Korea, (E) Spain, (F) Portugal. (Made with Natural Earth)

Tables

VARIABLES	(1) $\Delta\log(\#\text{est.})$	(2) $\Delta\log(\#\text{est.})$	(3) $\Delta\log(\#\text{est.})$
log(TQI)	0.059*	0.066	0.060
	(0.016)	(0.028)	(0.021)
	0.03 - 0.09	0.01 - 0.12	0.02 - 0.10
	(0.001)	(0.023)	(0.005)
log(#est. in 2011)	-0.033	-0.051	-0.039
	(0.018)	(0.021)	(0.020)
	-0.07 - 0.00	-0.09 - -0.01	-0.08 - 0.00
	(0.071)	(0.017)	(0.055)
log(GDP/cap)		-0.090	
		(0.050)	
		-0.19 - 0.01	
		(0.078)	
log(pop)		0.080	
		(0.030)	
		0.02 - 0.14	
		(0.010)	
log(#est. in ind.)		0.058	
		(0.031)	
		-0.00 - 0.12	
		(0.060)	
country FE	no	no	yes
industry FE	no	no	yes
R^2	0.014	0.062	0.341
# obs.	37,596	37,596	37,596
F-stat	8.9	16.0	4.6
d.o.f. F-stat	(2, 95)	(5, 95)	(2, 95)
Prob. $>$ F	0.000	0.000	0.012

Statistics: Point estimate, (standard errors), 95% two-sided confidence interval, (p-value)
Standard errors two-way clustered by country and industry: * p$<$0.001

Table 1: Business travel and industry growth. Dependent variable: Change in the logarithm of establishments. Independent variables: mean-reversion term ($\log(\#\text{est. in }2011)$, models 1-3), logarithms of the country's GDP per capita and population and logarithm of the industry's number of establishment worldwide (model 2), dummy variables for the country of origin and for the country of destination (model 3). Full model results replicated in Supplementary Table 25.