WORKING PAPER Nº IDB-WP-01712

Skills and Multinational Production

Andrea Lassmann
Christian Volpe Martincus

Inter-American Development Bank
Productivity, Trade and Innovation Sector

May 2025

Skills and Multinational Production

Andrea Lassmann
Christian Volpe Martincus

Inter-American Development Bank
Productivity, Trade and Innovation Sector

May 2025

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Lassman, Andrea.
Skills and multinational production / Andrea Lassmann, Christian Volpe Martincus.
p. cm. — (IDB Working Paper Series ; 1712)
Includes bibliographical references.
1. English language-Economic aspects-Mathematical models. 2. Investments, Foreign-Mathematical models. 3. International business enterprises-Mathematical models. 4. Production (Economic theory)-Mathematical models. 5. Life skills-mathematical models. I. Volpe Martincus, Christian. II. Inter-American Development Bank. Productivity, Trade and Innovation Sector. III. Title. IV. Series.
IDB-WP-1712

Keywords: Multinational Firms; Foreign Language; Digital Skills.
JEL-Codes: F23, J24

http://www.iadb.org

Skills and Multinational Production*

Andrea Lassmann
University of Mainz and CESifo

Christian Volpe Martincus
Inter-American Development Bank and CESifo

This version: May 2025

Abstract

This paper provides new evidence on the role of human capital in shaping economies' participation in multinational production. In particular, we primarily examine whether and to what extent the level of English language mastery among countries' inhabitants affects the presence of multinational firms in their territories. To do so, we combine data on multinational firms' foreign subsidiaries worldwide and data on English proficiency of possible host countries' populations. Our estimates suggest that countries whose populations have higher levels of English proficiency attract more multinational firms. The same holds for economies with higher shares of individuals with advanced digital skills.

Keywords: Multinational Firms; Foreign Language; Digital Skills.
JEL-Codes: F23, J24

*We would like to thank Education First (EF) for kindly providing data on English proficiency. We are also grateful to Peter Schott and participants at the ETSG Conference held in Surrey for valuable comments and suggestions as well as to Richard Schulz for assistance. The views and interpretations in this paper are strictly those of the authors and should not be attributed to the Inter-American Development Bank, its executive directors, or its member countries. Other usual disclaimers also apply. Andrea Lassmann: andrea.lassmann@uni-mainz.de; Christian Volpe Martincus: christianv@iadb.org.

1. Introduction

Evidence suggests that foreign direct investment (FDI) and specifically participation in multinational production can contribute to countries' growth and sustainable and inclusive development (Wang and Blomstrom, 1992, Borensztein et al., 1998, Hanson, 2001, Alfaro et al., 2004, Blalock and Gertler, 2008, Alfaro, 2016) (Blalock and Gertler, 2002).[1] FDI and multinational production are, however, spatially concentrated in relatively few economies (Ramondo et al., 2015). This naturally leads to the question of what the factors are that determine such spatial distribution.

This paper provides new evidence on the role of one of such factors: skills. To capture countries' skill endowments, we first and primarily use the level of foreign language proficiency of their populations. In particular, we focus on English proficiency because this language is the prevalent *lingua franca* in modern international relationships. Importantly, an increasing number of multinational firms (MNEs) have adopted English as their official language (Neeley, 2012, Ryder, 2014). Higher levels of English mastery can therefore help lower information and communication frictions, which are a major obstacle for international economic activities (see, e.g., Rauch, 1999, Anderson and van Wincoop, 2004, Oldenski, 2012, Allen, 2014). Consistent with this, existing empirical evidence shows that English proficiency has trade-promoting effects and that proficiency in that language is more important for trade than proficiency in other spoken languages (see Hutchinson, 2002, Ku and Zussman, 2010, Fidrmuc and Fidrmuc, 2016).

More precisely, we examine whether and to what extent English proficiency shapes multinational production patterns as proxied by the number MNEs in host economies.[2] In our baseline analysis, we assess the impact of population's English mastery on the extensive margin within established bilateral multinational production links (i.e., the number of MNEs from given home countries operating in given sectors in given host economies in which they were already present). In an extension, we examine the effect of such an skill on the overall extensive margin (i.e., the presence of MNEs from given home countries operating in given sectors in given host economies in which they were not present and thus there were no previous bilateral multinational production links).

To assess the effects of English proficiency on multinational production, we combine data on MNEs' foreign activity worldwide from Dun&Bradstreet's WorldBase and

[1] Several micro-level studies uncover the various channels through which FDI and multinational production can positively impact host economies, including through demonstration and competition effects, labor turnover, and buyer–supplier linkages (e.g., Rodríguez-Clare, 1996, Aitken et al., 1997, Blomstroem and Kokko, 1998, Alfaro and Rodríguez-Clare, 2004, Javorcik, 2004, Balsvik, 2011, Harding and Javorcik, 2012, Muendler et al., 2012, Poole, 2013, Alfaro-Ureña et al., 2022, Carballo et al., forthcoming)

[2] "Home countries" can be alternatively referred to as "origin countries" or "source countries", and "host countries" as "destination countries".

data on English proficiency of possible host countries' populations from Education First (EF). Using these data, we regress the (log) change of the number of MNEs from a home country in a sector in a host country between 2012 and 2019 on the host country's initial (2012) level of English proficiency along several host country-specific covariates (e.g., the size and level of development of these economies) and home country-sector fixed effects. This long-difference specification allows us to control for largely time invariant bilateral multinational production drivers (e.g., distance and, noteworthy, common native language) and potentially relevant confounding host country-, home country-, and sector-specific factors.

Still, admittedly, English proficiency can be endogenous for several reasons including measurement error, simultaneity (i.e., participation in multinational production and skills can be jointly determined by policies or firms' decisions), and reverse causality (i.e., the arrival of new MNEs can result in improved curricula in host countries).[3] To address these endogeneity concerns, we resort to an instrumental variables approach, whereby the initial level of English proficiency is instrumented with two linguistic proximity indices that were constructed in Melitz and Toubal (2014) and are standard in the literature.

Moreover, to provide further support to our identification strategy, we additionally assess whether, as expected, the so-estimated impacts are in line with English proficiency acting as a mechanism that lowers information and communication costs. We accordingly allow effects to vary across sectors depending on the severity of information frictions associated with their goods and services and the intensity of communication needs and complexity of the tasks required to produce them, and examine whether the size of these effects corresponds to the level of these information and communication costs.

Also important, English proficiency is likely to interact with other skills in shaping multinational production patterns. This is particularly the case with digital skills. We proxy these skills with the share of individuals who have used software for electronic presentations and individuals who have written computer codes as reported by the OECD.[4] To explore whether and how digital skills affect the spatial distribution of multinational production along English proficiency, we use a generalized propensity framework (Hirano and Imbens, 2004, Imai and van Dyk, 2004). The reason is twofold. First, it is challenging to find a suitable IV for digital skills. While it does not entirely preclude them, this estimator helps alleviate potential endogeneity con-

[3]The latter two concerns apply to both the initial level of multinational production and the dependent variable as such in the presence of high autocorrelation.

[4]While clearly relevant proxies, the exact choice among the alternatives has been at least partially motivated by data coverage reasons.

cerns. Second, this approach allows us to estimate the joint effect of English proficiency and digital skills on the extensive margin of multinational production while addressing non-linearities.

Our IV (and OLS) estimates suggest that English proficiency has a positive and significant effect on economies' participation in multinational production: countries whose populations are more proficient in English attract more MNEs. According to these estimates, the elasticity of the change in the number of MNEs with respect to the EPI ranges between 0.67% and 1.16% and averages 0.89% in our preferred benchmark specification that includes home country-sector fixed effects. Hence, a doubling of the population's English proficiency is associated with almost a doubling of the growth in the number of MNEs from a home country active in a sector in the respective host country. Thus, if Chile (at the 10th percentile of English proficiency) had attained Norway's level (90th percentile) in 2012, we predict an increase of 0.272 log points in MNE presence, equivalent to a 31.3% rise. This would translate to approximately 2 MNEs per industry rather than the observed average of 1.5. In aggregate terms, Chile's total MNE count would have increased from 870.5 to approximately 1143 firms, representing a substantive economic difference attributable to language proficiency.

Further, our results reveal that that these effects are stronger in sectors facing more information frictions and that are more intensive in interactive communication, thus pointing to English proficiency as a means to lower information and communications costs of doing business. In particular, impacts are larger in industries that manufacture differentiated goods, provide services, and have relative high shares of occupations that involve tasks whose completion requires frequent communication or are complex and demand higher levels of creativity. In addition, estimates indicate that English proficiency does not only contribute to determine a country's (change in the) level of involvement in multinational production, but also its participation altogether.

Finally, our estimation results reveal that higher shares of advanced digital skills in the population are associated with more multinational production in the country, whereas the effect of English proficiency is about linear. These results inform education policies, specifically those fostering foreign language (for an overview, see Ginsburgh and Weber, 2020) and digital skills, and how these may impact FDI, involvement in multinational production, and, ultimately, growth (Rodrik, 2007).

Our paper relates to three main strands of literature. The first, rooted in the exploration of the determinants of bilateral trade within the gravity framework, has shown that native and spoken languages impact international trade (e.g., Melitz, 2008, Melitz and Toubal, 2014) and FDI (e.g., Feng et al., 2019). Language naturally embodies various aspects, such as cultural aspects inherent in *native* language (e.g., Egger and Lassmann, 2015, Ginsburgh and Weber, 2020), and the facilitation of information and communication, which results from the acquisition and knowledge of foreign languages (common *spoken* language, see Egger and Toubal, 2016). It therefore affects bilateral trade by shaping preferences and/or reducing trade frictions.

The second strand of literature focuses on the link between FDI and human capital (e.g., Blomstroem and Kokko, 2002). An emerging body of papers examines the role of language and digital skills in the organization of multinational firms and the transfer of knowledge within these firms.[5] For example, using experimental data for Myanmar, Guillouet et al. (2024) find that English language barriers between foreign and domestic managers can impede these knowledge transfers.[6] Relatedly, there is a broader literature on the importance of communication in multinational firms (e.g. Defever, 2012, Cristea, 2015, Gumpert, 2018, Bahar, 2020).

Finally, the paper adds to a series of studies that explore the conditions under which FDI is conducive to economic growth. Thus, previous literature has suggested that this has been primarily the case when the host country has a minimum level of human capital (Borensztein et al., 1998), which ensures the required absorptive capacity to benefit from it. More recently, Bénétrix et al. (2022) indicate that, from the 1990s onwards, FDI positively correlates with growth in countries with both high GVC activity growth and low initial levels of human capital or financial development.

We contribute to these branches of the literature by assessing the role of acquired foreign language knowledge (as opposed to common native or official languages) in shaping the multinational production's extensive margin. This is relevant because the extensive margin accounts for a large share of the variation of this production across countries and is responsible for most of given multinational firms' expansion over time (see Ramondo et al., 2015, Garetto et al., 2019). Importantly in this regard, to properly measure such a margin, we use microdata on multinational firms' location decisions, instead of relying on FDI data that primarily captures financial transactions. Furthermore, we explore the mechanisms of the effect of population's English proficiency on the multinational production extensive margin and, different from previous studies, we also examine its interplay with other relevant skills such as digital skills. This is not only of academic interest but also valuable from a policy point of view as our results provide insights on the appropriate policy mix to increase countries' participation in

[5]It has been estimated that two countries that share a common language have 65% more bilateral affiliates than their counterparts with different languages (Ramondo et al., 2015).

[6]There is also a small literature in business management. For instance, Welch and Welch (2008) describe the cultural aspect of language and its role in the knowledge transfer process. Tenzer et al. (2013) study how language barriers influence trust formation in multinational teams for some German automotive firms. Peltokorpi (2015) examines host country corporate language proficiency and reverse knowledge transfer for Japan.

multinational production.

The paper is structured as follows. Section 2 describes the data and presents descriptive evidence. Section 3 explains the estimation strategy. Section 4 reports and discusses the main estimation results. Section 5 presents the results of extensions that aim at establishing the mechanisms of the effects of interest. Section 6 concludes.

2. Data

Our dataset consists of two main databases and data from several complementary sources. First, we use data on MNEs from Dun&Bradstreet's WorldBase. This data is collected from various sources including chamber of commerce registers, telephone directory and insolvency records, websites, and dedicated investigations. The data quality is verified centrally through multiple checks (e.g., Alfaro and Chen (2012). Comparisons with other databases such as those of UNCTAD and US' BEA suggest that the WorldBase can be considered one of the best estimations of the global population of multinational firms (see Alfaro and Charlton, 2009).

We specifically work with all (global ultimate) parent firms that, at some point of the sample period (2012-2019), have at least one subsidiary (or branch) in a different country (i.e., roughly 200,000 firms). For these MNEs, the WorldBase furnishes us with data on source country, year of establishment and (ISIC 4-digit) sector of activity as well as data on location (i.e., host country), year of establishment, and (ISIC 4-digit) sector of activity for each of its foreign affiliates.

We limit the sample to the manufacturing and main services sectors: manufacturing (ISIC Rev.4 divisions 10-33), transportation (ISIC divisions 49-53), ICT (ISIC divisions 58-63), financial (ISIC divisions 64-66), and real estate and business services (ISIC divisions 68-82). We exclude all other divisions because these industries are relatively special in terms of nature of their activities, market structure, or tradeability of output. [7]

Based on these data, we compute the *number of MNEs* and, in addition, the *number of their foreign affiliates* by home country-sector-host country triplets over time.

The distribution of the number of MNEs (and that of their foreign affiliates) across countries in 2019 is shown in Figure A1 and Figure A2 in Appendix A. This distribution is strongly skewed, more so by home country than by host country, so that the number of these firms is very large in a few countries and relatively low in most countries. The same holds for industries.[8]

[7]This is the case with agriculture, mining, electricity, construction, wholesale and retail trade, public administration, etc. Still, the main results are robust to including all sectors as well as to also excluding the financial and transportation sectors. These results are available from the authors upon request.

[8]*Financial services* and *Business support activities* are among the largest sectors in terms of the number of

In the analysis below, we use a subset of these data, namely, those corresponding to countries for which we also have information on the main explanatory variable, i.e., English proficiency. The coverage of the resulting data – separately by host and home country – is presented in Figure A3 in Appendix A. Because we have no information about English-speaking countries, these are not included as host countries, a point that we clarify in more detail below. The largest number of hosted multinational firms and foreign affiliates is observed in Europe (the top host country is Germany), China, Mexico and Russia. In terms of home countries, the geographic pattern is similar, but more pronounced towards developed economies. The top three home countries are the US, the United Kingdom, and Germany.

Second, we use data on English proficiency from 2012 to 2019 from Education First (EF), who publishes the *English Proficiency Index (EPI)* at the country-level (see `https://www.ef.com/wwen/epi/`). This index is compiled from annual samples. In 2019, the EPI was based on reading and listening skills of around 2.3 million individuals from 100 countries, who took part in the EF Standard English Test (EF SET) or one of the English placement tests in the previous year. The test correlates strongly with TOEFL iBT scores (as used in Ku and Zussman, 2010) and IELTS Academic Test scores (correlation coefficient of 0.8 and 0.74, respectively). It is worth noting that countries where English is the majority language, such as Australia, Canada, Ireland, New Zealand, the United Kingdom, or the United States are not included (coverage is indicated in Figure A4 in the Appendix).In our baseline, we therefore exclude majority English-speaking countries from the set of host countries.[9] The score ranges from 300 to 700.

In addition, we use two different measures of linguistic proximity, LP1 and LP2, to address potential endogeneity of the EPI as explained in Section 3 (see Melitz and Toubal, 2014). The first measure, LP1, is based on the *Ethnologue* classification of language trees between trees, branches, and sub-branches (Laitin, 2000, Fearon, 2003). A language tree refers to a language family (e.g., Indo-European), and (sub-)branches to separate nodes in a language tree.LP1 takes four distinct values: it is coded as 0 in case any two languages belong to separate language trees, 0.25 if they adhere to different branches in the same tree, 0.5 if they adhere to the same branch, and 0.75 if they belong to the same sub-branch. The second measure, LP2, is based on the lexical similarity between at least 40 words compiled by the Automated Similarity Judgment Program (ASJP) project. LP2 differentiates better in cases where two languages belong to dif-

MNEs and foreign affiliates, whereas *Veterinary activities* and *Printing and reproduction of recorded media*are among those with the lowest number of these firms.

[9] As the index is not provided for countries in which English is the majority language, we would have to assign an own value to these countries, which is not trivial given the calculation of the score. Even if it is not fully correct that English proficiency is 100% in these countries, we can safely assume that in their location decision, MNEs assume to find fully proficient employee there and thus exclude these countries.

ferent trees and will therefore used as our baseline. For both measures, higher values mean closer proximity. The two variables were then normalized for comparability (for a detailed description see Melitz and Toubal, 2014).[10]

Moreover, we rely on data on host countries' income and human capital measures to account for the possible positive correlation between country's English proficiency and level of development and compensation or general levels of education (see Konara and Wei, 2019). These measures include: GDP and GDP per capita (in constant 2015 USD), average (monthly) earnings of employees (across economic activities), gross secondary school enrollment in percentage, and gross tertiary school enrollment in percentage, from the World Bank Development Indicators and ILO's ILOSTAT. Note that gross school enrollment variables can exceed 100% because they are defined as "the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown" (`https://data.worldbank.org/`). These variables are available at the host country level and their coverage is shown in Figure A4 in Appendix A.

We also resort to data on investment promotion policies in the host countries to control for the reduction in information costs that MNEs experience in considering locations when they receive assistance through the services provided by the respective host countries' investment promotion agencies. Such reduction in investment-related information costs may be potentially correlated with the host country's English proficiency either because these agencies' facilitation activities make it easier to actually take advantage of such skills by helping foreign firms find and recruit individuals who are proficient in English or because these agencies' policy advocacy activities contribute to improve the population's skills and specifically English knowledge by informing the government and the public opinion about the need to do so to attract more foreign firms. Hence, we estimate alternative specifications of our baseline equation whereby we include their budget along with the country and sector prioritization strategies of host countries' investment promotion agencies using data from Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025). With the exception of that capturing sector prioritization, which is at the host country-sector level, these variables are at the country level. The geographical coverage of these variables is presented in Figure A4 in Appendix A.

Finally, we take into account the characteristics of the tasks involved in sectors' occupations to allow for the possibility that the relative importance of language skills varies depending on the extent to which these are actually needed to perform those tasks.

[10] The correlation between LP1 and LP2 is 0.93. An alternative measure could be linguistic distance as described in Isphording and Otten (2013).

More precisely, we use data on the share of communication intensive and non-routine tasks in occupations from the US Department of Labor's Occupational Information Network (O*NET) aggregated at the sector-level (see Oldenski, 2012).[11]

Merging these databases results in a dataset with information on 53 host and 177 home countries in 251 ISIC Rev. 4 industries (52 2-digit industries) and 8 years (2012-2019).[12] Panel A of Table 1 presents descriptive statistics for the dependent variables. We average over country pairs and ISIC Rev.4 4-digit industries. The corresponding average number of multinational firms per country-pair-industry is 3.6, whereas that of their foreign affiliates is 4.7. This means that a typical multinational firm from a given home country has little more than one affiliate operating in a given sector in given a host country. Looking next at average changes over time, these numbers increase by 1 and 1.5 between the beginning (2012) and the end (2019) of the sample period, respectively.

Panel B of Table 1 provides information on the explanatory variables. The average EPI amounts to 547, with considerable variation across countries and, as indicated by its difference, also over time.[13] LP1 and LP2 are normalized such that their value extend the unit interval. They range from 0 to roughly 4, with higher values meaning closer language proximity. Regarding the remaining variables, the average host country in our sample has a GDP per capita of 29,099 USD (9.9 for log GDP per capita), average log earnings of 7.9, a secondary school enrollment rate of 108%, and a tertiary school enrollment rate of 69%. Furthermore, the average 2016 IPA budget amounted to 1.9 (in logs).

[11]We thank Lindsay Oldenski very much for kindly sharing this sector-level data with us.

[12]This lower number of host countries is the result of using EPI data, which is not available for all countries for which we have information on multinational activity. Our sample period is also limited by data constraints. While the data on MNEs to which we had access for this study would have allowed us to consider a longer period backwards, the coverage of this data finished in 2019. On the other hand, data on English proficiency was available until 2021, but only starting in 2012.

[13]While, on average, the differences over time are positive some countries seem to have experienced a worsening in the English proficiency of their population over the last decade. These include United Arab Emirates, Indonesia, Iran, Jordan, Libya, Sri Lanka, Latvia, Oman, Pakistan, Qatar, Turkey, and Venezuela.

Table 1: SUMMARY STATISTICS I

	N	Mean	SD	Min	Max
A. Dependent variables					
Number of MNEs	53513	3.627	14.444	1	1631
Ln(number of MNEs)	53513	0.560	0.823	0	7
ΔNumber of MNEs	53513	0.977	7.373	-25	849
ΔLn(number of MNEs)	53513	0.130	0.303	-2	3
Number of affiliates	53513	4.730	23.926	1	2184
Ln(number of affiliates)	53513	0.667	0.900	0	8
ΔNumber of affiliates	53513	1.439	12.805	-28	1552
ΔLn(number of affiliates)	53513	0.147	0.326	-2	3
B. Independent variables					
English Proficiency Index (EPI)	53513	547.662	59.804	388	644
ΔEPI	53513	23.359	24.395	-37	82
ln(EPI, 2012)	53513	6.257	0.102	5.986	6.444
LP1	52023	1.678	1.383	0.000	3.746
LP2	52023	1.308	0.959	0.000	3.563
Ln GDP (constant 2015 USD)	52449	27.383	1.217	23.975	30.291
Ln GDP per capita (constant 2015 USD)	52449	9.931	0.936	7.305	11.39
Ln earnings	42211	7.879	0.645	6.272	9.186
School enrollment, secondary (% gross)	46602	107.791	19.059	44.868	156.081
School enrollment, tertiary (% gross)	46978	68.592	19.781	12.221	115.042
Log IPA Budget, 2016	32564	1.942	1.585	-2.285	5.715
IPA Country Prioritization	53513	0.113	0.317	0	1
IPA Sector Prioritization	53513	0.099	0.299	0	1
Number of host countries	53				
Number of home countries	177				
Number of ISIC Rev.4 industries	251				
Number of 2-digit industries	52				

Sources: Number of MNEs and foreign affiliates, Authors' calculations based on Dun&Bradstreet's WorldBase; English Proficiency Index (EPI), EF (Education First); LP1 and LP2, CEPII based on Melitz and Toubal (2014); GDP, GDP per capita, secondary and tertiary school enrollment, World Bank; Earnings, ILOSTAT; Investment Promotion (IPA) variables, Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025). Differences are constructed between 2012 and 2019. All variables are averaged over country-pairs and ISIC Rev.4 4-digit industries as applicable and relevant.

3. Empirical Approach

We estimate the effect of English proficiency on countries' participation in multinational production using the following baseline specification:

$$\Delta \ln y_{ijkt} = \beta_1 \ln EPI_{it_0} + \beta_2 \Delta \mathbf{X}_{it} + \delta_{jk} + \Delta \varepsilon_{ijkt}, \tag{1}$$

where i denotes the host country, j the home (headquarter) country, k the industry, t time, and Δ corresponds to the change between 2012 and 2019. y is the number of multinational firms and EPI_{it_0} is the English proficiency index in the host country in 2012.[14] $\mathbf{X}_{it}$ includes control variables (i.e., host country GDP, GDP per capita, earnings, secondary and tertiary education shares), and ε_{ijkt} is an error term. Equation (1) is first estimated by OLS. Note that first-differencing absorbs all unobserved time-invariant factors both at the host country-home country level such as distance, common (native) language, and common colonial history and at the host country-sector-home country level such as systematic relative specialization patterns. Alternatively, Equation (1) is expanded to include home country-industry fixed effects, the latter at the 2-digit ISIC level. Standard errors are clustered at the host country level for inference purposes.

There may be several sources of potential endogeneity bias. These include: (i) English proficiency may be measured with error because the proxy we use, the EPI measure, is based on national samples of language tests instead of on comprehensive administrative records; (ii) omitted relevant variables such as productivity, which may be correlated with both the EPI (this is explicitly indicated in the annual reports by EF) and the number of MNEs; and, especially in the presence of strong serial correlation, (iii) simultaneity, whereby the number of multinational firms and English proficiency impact on each other at the same time and are thus concurrently determined; and (iv) reverse causality, whereby the increased presence of multinational firms in a host country leads to improvements in education curricula, in general, and English proficiency, in particular.[15]

To address these endogeneity concerns, we instrument EPI with the linguistic proximity measure LP2 and, as a robustness check, with the alternative measure LP1. This is similar to the approach in Ku and Zussman (2010), who instrument English TOEFL

[14] We have also used the log number of foreign affiliates as dependent variable (see Section 4). The corresponding estimates are reported in Appendix A, Table A3. These are in line with the baseline presented here. Furthermore, we have estimated the equation using $\Delta \ln EPI_{it}$ instead of the log EPI in 2012. These alternative results convey a similar message and are available from the authors on request.

[15] One way to mitigate concerns associated with the second potential source of endogeneity bias could be to include productivity. However, this would produce estimates of β_1, which are exclusive of productivity. Because English skills may improve productivity, thereby leading to the establishment of more multinational firms, we are interested in an effect inclusive of productivity.

scores with a measure of linguistic distance. Specifically, in order to measure proximity to the English language, we calculate the linguistic proximities (LP2 and, alternativly, LP1) of all host countries in the sample, to the United Kingdom and the United States. We then take averages across these two countries, because spoken language families differ by country, and thus the linguistic proximities to British and American English are not always the same.[16] This IV can be expected to be relevant: more proximity to the English language is likely to be strongly associated with higher English proficiency (see Kim and Lee, 2010, Ku and Zussman, 2010). We present first evidence for this in Figure A5 in Appendix A and complement it in Section 4. Regarding the exclusion restriction, first differencing along with the covariates (in $\mathbf{X}_{it}$) and the sets of home country-sector fixed effects should account for most possible channels through which its violation could occur. For instance, it could be argued that countries with close proximity to the English language may also have developed similar institutions, which, in turn, could affect location choice. However, this would be controlled for by first differencing and inclusion of covariates such as GDP per capita (one of whose growth determinants are institutions), along with the fixed effects.

4. The Effect of English Proficiency on Multinational Production's Extensive Margin

We first report OLS estimates of Equation (1) in Table 2, both without home country-sector fixed effects (Panel A) and with home country-sector fixed effects (Panel B).[17] Across columns, we control for alternative sets of host country-level covariates including: log change of earnings and GDP (Column 2); log change of GDP and GDP per capita (Column 3); log change of GDP and GDP per capita and change in secondary and tertiary education enrollment (Column 4); and log change of GDP and GDP per capita and variables capturing host country investment promotion policies–i.e, the log budget and the country and sector priorities of the respective national investment promotion agency (Column 5). Regardless of the control variables, OLS estimates indicate that EPI has a positive and significant relationship with the growth of the number of MNEs in the respective host countries. The average estimated elasticity across specifications is 0.027, i.e., a 1% increase in 2012's host country average EPI is related to a 0.027% higher growth in the number of MNEs active in its territory (see Panel A).[18]

[16] This list could be augmented by Australia, Ireland, New Zealand and different islands. When doing so, the constructed variable remains the same.

[17] The number of observations stems from host country $\times$ home country $\times$ sector, for which data on English proficiency and the number of MNEs is available (see column 1 of Table 1). The number of observations in other columns are generally lower due to limitations in the data on the included covariates.

[18] The sign of the coefficient on the change in log GDP is negative. Because we simultaneously include the change in log GDP per capita, the change in log GDP captures population growth. Since the latter

This elasticity remains significant and becomes larger in magnitude when the specifications are expanded to additionally include home country-sector fixed effects.[19] It ranges between 0.251 and 0.554 and averages 0.398. Thus, when identification come from variation across host countries within home country-sector pairs, an increase in 2012's average EPI by 1% is associated with an increase in the growth of number of MNEs by 0.398% (see Panel B).[20]

As discussed in Section 3, English proficiency can be potentially endogenous to multinational firms' activities. To address this concern, we implement the IV approach described in that section and report the respective results using LP2 as instrument in Table 3, based on both the specifications that do not include home country-sector fixed effects (Panel A) and the specifications that include these fixed effects (Panel B). We show results for LP1 in Appendix Table A1. These are similar to those reported here. This is not surprising given the high correlation between LP1 and LP2.

The p-values for the Kleibergen-Paap LM statistic of underidentification are small, and the Kleibergen-Paap Wald rk F statistics of weak identification are reasonable.In consonance with what was observed in Figure A5, these statistics indicate that our IV has a high conditional correlation with the EPI and hence are relevant.

The estimated effects are positive and significant across all alternative specifications. The estimated elasticity averages 0.048 across specifications without home country-sector fixed effects (see Panel A). Similar to Table 2, estimates increase in magnitude when we resort to more stringent identification strategy and include home country-sector fixed effects (Panel B). In this case, the elasticity averages 0.890 across across specifications.[21] This implies that a 1% increase in 2012's average EPI would be associated with 0.89% increase in the growth rate of MNEs between 2012 and 2019, thus doubling the initial mean EPI would have resulted in a 89% increase in the number of MNEs (i.e., in terms of the mean indicated in Table 1, the growth rate would rise by 12 percentage points, from 0.13% to 0.25%). Overall, these IV estimates are larger than their OLS counterparts reported in Table 2. This difference is likely to reflect measurement error in the English proficiency, which leads to downward bias in the OLS estimates.[22]

is negatively correlated with development, the negative sign is thus not surprising. Furthermore, the coefficient on the change in log GDP per capita has a positive sign and is significant.

[19]The same holds when the estimated specification includes separate home country and industry fixed effects. These alternative estimation results are available from the authors upon request.

[20]Since the independent variables are correlated, we also run a specification curve analysis (Simonsohn et al., 2015, 2020). Figure A6 in Appendix A reports the estimated coefficient on EPI for several model specifications. Overall, the change in the coefficient is minor.

[21]Note that estimates are less precise and the IV is weaker than when such fixed effects are not included.

[22]Our control variables could also be potentially endogenous to changes in the number of MNEs due to simultaneity or reverse causality. Our IV estimates without these covariates are similar to those including them, suggesting that, if present, such an issue does not seem to significantly affect our results.

Table 2: THE EFFECT OF ENGLISH PROFICIENCY ON THE NUMBER OF MNEs (OLS)

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			Δlog number of MNEs		
A. Without home country-sector fixed effects					
lnEPI_{it_0}	0.021***	0.027***	0.032***	0.029***	0.025***
	(0.003)	(0.005)	(0.005)	(0.004)	(0.006)
Δln(earnings)		0.079			
		(0.060)			
Δln(gdp)		-0.262+	-1.084***	-1.244**	-0.498
		(0.137)	(0.229)	(0.362)	(0.425)
Δln(gdp pc)			0.923***	1.168*	0.603
			(0.224)	(0.445)	(0.384)
Δsecondary ed.				0.001	
				(0.000)	
Δtertiary ed.				-0.000	
				(0.002)	
lnIPA Budget					0.009
					(0.012)
IPA Country Prioritization					-0.053+
					(0.027)
IPA Sector Prioritization					0.057*
					(0.025)
N	53513	36692	52449	36158	32212
Adj. R-sq.	0.157	0.177	0.176	0.176	0.204
Home-industry FE	No	No	No	No	No
B. With home country-sector fixed effects					
lnEPI_{it_0}	0.438***	0.554***	0.360**	0.251	0.389*
	(0.117)	(0.126)	(0.114)	(0.162)	(0.161)
Δln(earnings)		0.103*			
		(0.043)			
Δln(gdp)		-0.130	-0.744**	-1.256**	-0.298
		(0.097)	(0.253)	(0.351)	(0.488)
Δln(gdp pc)			0.648*	1.180*	0.335
			(0.247)	(0.441)	(0.432)
Δsecondary ed.				0.000	
				(0.001)	
Δtertiary ed.				0.001	
				(0.002)	
lnIPA Budget					0.008
					(0.011)
IPA Country Prioritization					-0.032
					(0.027)
IPA Sector Prioritization					0.052*
					(0.022)
N	52413	35636	51353	35105	31240
Adj. R-sq.	0.126	0.150	0.134	0.143	0.136
Home-industry FE	Yes	Yes	Yes	Yes	Yes

Notes: Linear regressions in first differences with standard errors clustered at the host country level in parentheses, $^{+}(p < 0.1)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$. Differences refer to the difference of variables for years 2019 and 2012. *Sources*: Number of MNEs, Dun&Bradstreet WorldBase; log English Proficiency Index (ln EPI) in the year 2012, EF (Education First); Earnings, ILOSTAT; GDP, GDP per capita, secondary and tertiary school enrollment, World Bank; IPA (Investment Promotion Activity) variables 2016, Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025). Source-industry fixed effects at the 2-digit ISIC industry level.

It is worth mentioning that findings for the number of affiliates are equivalent. We report them in Table A3 in Appendix A.[23]

Tables 4 and A2 reports the first-stage estimates that correspond to Tables 3 and A1. These estimates consistently indicate that both IVs are strong across all alternative specifications. Furthermore, their effects do not vary much across types of linguistic proximity (LP1 versus LP2).

Next, we examine the relationship between English proficiency and the likelihood of a country pair to engage in multinational production in a given industry. We refer to this as the *overall extensive margin*. To do so, we use a binary indicator that takes the value of one if the number of MNEs in any triple home country-sector-host country in 2019 is positive and zero otherwise, and regress the change in this indicator on the log EPI in 2012 along with the alternative sets of covariates included in previous estimations. OLS and IV estimates of these specifications are reported in Panels A and B of Table A4, respectively. As before, the IV coefficients are larger than their OLS counterparts. According to the IV estimates, the coefficients range from 0.0003 to 0.001, i.e., doubling the 2012 EPI would be approximately associated with a 0.02–0.07 percentage point increase in the likelihood of a country hosting MNEs from a given home country and sector. Given that the unconditional probability is 0.47%, this would amount to a 4-14% increase in the likelihood of a country hosting MNEs from a given home country and sector.[24] Hence, better English proficiency in a host country has a relatively small overall but substantial effect in particular industries on the likelihood of multinational production taking place therein.

5. Mechanisms and Interplay with Digital Skills

In this section, we first shed light on potential mechanisms that drive the impact of English proficiency on multinational activity. Better knowledge of the *lingua franca* can facilitate information dissemination and communication when conducting such an activity (e.g. Guillouet et al., 2024). Accordingly, we explore whether EPI has heterogeneous effects across sectors depending on the severity of the information and communication frictions that they face. We capture these frictions through the degree of differentiation of the goods they produce and sell (differentiated goods vs. non-differentiated goods), the nature of their output (goods vs. services), and the type of tasks associated with their more important occupations (communication intensive vs. non-intensive,

[23] We therefore abstain from showing them separately in the following analysis, but they are available from the authors upon request.

[24] These estimation results are available from the authors upon request.

Table 3: The effect of English proficiency on the number of MNEs (IV)

	(1)	(2)	(3)	(4)	(5)
Dependent variable:		Δlog number of MNEs			
A. Without home country-sector fixed effects					
$\ln EPI_{it_0}$	0.027^{***}	0.039^{***}	0.046^{***}	0.042^{***}	0.066^{**}
	(0.004)	(0.008)	(0.008)	(0.011)	(0.022)
Δln(earnings)		0.021			
		(0.114)			
Δln(gdp)		-0.480^{*}	-1.782^{***}	-1.369^{***}	-2.538^{*}
		(0.188)	(0.397)	(0.349)	(1.372)
Δln(gdp pc)			1.339^{***}	0.976^{*}	1.742^{*}
			(0.339)	(0.498)	(0.959)
Δsecondary ed.				-0.000	
				(0.001)	
Δtertiary ed.				-0.004	
				(0.003)	
lnIPA Budget					-0.022
					(0.018)
IPA Country Prioritization					-0.060
					(0.038)
IPA Sector Prioritization					0.036
					(0.034)
N	52023	35202	50959	34668	32212
F-stat.	86.541	33.838	55.174	22.095	16.403
Home-industry FE	No	No	No	No	No
B. With home country-sector fixed effects					
$\ln EPI_{it_0}$	0.953^{*}	1.164^{**}	0.879^{**}	0.665^{*}	0.787^{*}
	(0.372)	(0.349)	(0.325)	(0.286)	(0.309)
Δln(earnings)		0.116^{+}			
		(0.06cl8)			
Δln(gdp)		-0.051	-0.429	-1.307^{**}	-0.219
		(0.158)	(0.399)	(0.366)	(0.660)
Δln(gdp pc)			0.353	1.166^{*}	0.204
			(0.397)	(0.469)	(0.553)
Δsecondary ed.				-0.001	
				(0.001)	
Δtertiary ed.				0.003	
				(0.002)	
lnIPA Budget					0.007
					(0.012)
IPA Country Prioritization					-0.003
					(0.037)
IPA Sector Prioritization					0.061^{*}
					(0.022)
N	50926	34136	49866	33610	31240
F-stat.	23.519	15.998	25.635	11.105	28.446
Home-industry FE	Yes	Yes	Yes	Yes	Yes

Notes: Linear IV regressions in first differences with standard errors clustered at the host country level in parentheses, $^{+}(p < 0.1)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$. Language proximity (LP2) as IV. Differences refer to the difference of variables for years 2019 and 2012. *Sources*: Number of MNEs, Dun&Bradstreet WorldBase; English Proficiency Index (ln EPI) in the year 2012, EF (Education First); Earnings, ILOSTAT; GDP, GDP per capita, secondary and tertiary school enrollment, World Bank; IPA (Investment Promotion Activity) variables 2016, Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025). LP2, CEPII. The F-statistic reports the Kleibergen-Paap F statistic for weak identification. Source-industry fixed effects at the 2-digit ISIC industry level.

Table 4: FIRST STAGE OF IV REGRESSIONS

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			$\ln \text{EPI}_{it_0}$		
A. Without home country-sector fixed effects					
LP2	3.131***	2.209***	1.923***	1.843***	1.106***
	(0.337)	(0.380)	(0.259)	(0.392)	(0.273)
Δln(earnings)		1.745			
		(2.443)			
Δln(gdp)		12.491***	23.949***	-4.110	23.413*
		(3.137)	(6.036)	(10.110)	(10.777)
Δln(gdp pc)			-9.695	24.023*	-6.222
			(7.400)	(10.839)	(11.298)
Δsecondary ed.				-0.011	
				(0.044)	
Δtertiary ed.				0.195***	
				(0.042)	
lnIPA Budget					0.539***
					(0.163)
IPA Country Prioritization					0.902
					(0.672)
IPA Sector Prioritization					0.543+
					(0.320)
N	52023	35202	50959	34668	32212
Adj. R-sq.	0.659	0.886	0.898	0.914	0.937
Home-industry FE	No	No	No	No	No
B. With home country-sector fixed effects					
LP2	0.059***	0.071**	0.073***	0.070***	0.084**
	(0.012)	(0.018)	(0.014)	(0.021)	(0.016)
Δln(earnings)		-0.016			
		(0.060)			
Δln(gdp)		0.065	-0.569*	-0.286	-0.940**
		(0.160)	(0.243)	(0.316)	(0.281)
Δln(gdp pc)			0.799**	0.755+	1.287***
			(0.3278)	(0.431)	(0.281)
Δsecondary ed.				0.001	
				(0.001)	
Δtertiary ed.				-0.002	
				(0.002)	
lnIPA Budget					0.005
					(0.007)
IPA Country Prioritization					-0.011
					(0.021)
IPA Sector Prioritization					-0.005
					(0.011)
N	50926	34136	49866	33610	31240
Adj. R-sq.	0.358	0.413	0.460	0.612	0.679
Home-industry FE	Yes	Yes	Yes	Yes	Yes

Notes: First stage of linear IV regressions in first differences with standard errors clustered at the host country level in parentheses, $^{+}(p < 0.1)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$. Differences refer to the difference of variables for years 2019 and 2012. *Sources*: English Proficiency Index (ln EPI) in the year 2012, EF (Education First); LP2, CEPII; Earnings, ILOSTAT; GDP, GDP per capita, secondary and tertiary school enrollment, World Bank; IPA (Investment Promotion Activity) variables 2016, Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025). Source-industry fixed effects at the 2-digit ISIC industry level.

routine vs non-routine). Second, we examine whether and how English proficiency interacts with other relevant skills (i.e., digital skills) in shaping countries' engagement in multinational production.

5.1. Mechanisms: The Heterogeneous Effects of English Proficiency across Industries

We first assess whether the EPI effects varies with the degree of differentiation of the goods that the sectors produce. We accordingly allow the impact of EPI to differ for differentiated goods sectors and non-differentiated goods sectors using the classification proposed by Rauch (1999).[25] Estimates reported in column 1 of Table 5 reveal that English proficiency impact differently multinational productions in these two types of sectors. More precisely, the effect is substantially stronger for sectors producing differentiated goods than for their counterparts producing non-differentiated goods.[26]

We next examine whether English proficiency has differential impacts on multinational production depending on the nature of their output. More specifically, we allow such impacts to differ between manufacturing and services industries. Column 2 of Table 5 shows the respective estimates. These suggests that English proficiency is significantly more important in services than in manufacturing. This is precisely what it would be expected if MNE-provided services are, on average, more skill-intensive, more personalized, and thus more differentiated, and therefore require more interactive communication for their provision than goods and the costs of such a communication are actually lowered by higher levels of English mastery (see, e.g. Blinder, 2006, Andrenelli et al., 2018). Our data also support this, as the average values of CRTV (the importance of complex relative to importance of other tasks in an industry) and COMM (the importance of communication outside of the organization relative to the importance of other tasks) are substantially larger in services (0.72 for CRTV and 0.94 for COMM) than in manufacturing (0.21 and 0.02, respectively).

Finally, we explore whether the relative importance of language skills as a multinational production determinant varies depending on the extent to which these are actually needed to perform the tasks involved in main sectors' occupations.[27] To do so, we use data on the characteristics of tasks in occupations from the US Department of Labor's Occupational Information Network (O*NET) aggregated at the sector-level (see Oldenski, 2012, for a detailed description). These characteristics are proxied with two variables: (1) the importance of communicating outside of the organization rela-

[25] We specifically subsume homogeneous and reference-priced sectors under non-differentiated sectors.

[26] Main effects are estimated but omitted in the estimations whose results are reported in Panel A.

[27] This also broadly relates to the literature on the importance of communication costs in the organization of knowledge and multinational investment (e.g., Defever, 2012, Cristea, 2015, Gumpert, 2018, Bahar, 2020).

Table 5: Robustness Checks: Industries and Sectors

	(1)	(2)	(3)	(4)
Dep. var.:	Δlog number of MNEs			
A. Without home country-sector fixed effects				
$\ln\text{EPI}_{it_0}$	0.011***	0.016***	0.017***	0.020***
	(0.002)	(0.003)	(0.004)	(0.004)
$\ln\text{EPI}_{it_0}\times$ Differentiated	0.978**			
	(0.362)			
$\ln\text{EPI}_{it_0}\times$ Services		1.149**		
		(0.386)		
$\ln\text{EPI}_{it_0}\times$ COMM			1.145**	
			(0.386)	
$\ln\text{EPI}_{it_0}\times$ CRTV				1.132**
				(0.399)
N	38379	52023	49679	49679
F-stat.	12.840	13.595	13.479	12.519
Home-industry FE	No	No	No	No
B. With home country-sector fixed effects				
$\ln\text{EPI}_{it_0}$	0.422*	0.691*	0.719*	0.791*
	(0.205)	(0.308)	(0.312)	(0.316)
$\ln\text{EPI}_{it_0}\times$ Differentiated	0.523*			
	(0.222)			
$\ln\text{EPI}_{it_0}\times$ Services		0.484**		
		(0.170)		
$\ln\text{EPI}_{it_0}\times$ COMM			0.429**	
			(0.156)	
$\ln\text{EPI}_{it_0}\times$ CRTV				0.314^{+}
				(0.164)
N	37384	50926	48590	48590
F-stat.	11.111	10.372	10.532	11.494
Home-industry FE	Yes	Yes	Yes	Yes

Notes: IV regressions in first differences with standard errors clustered at the destination level in parentheses, $^{+}(p < 0.1)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$. IV: LP2. Differences refer to the difference of variables for years 2019 and 2012. *Sources*: Number of MNEs, Dun&Bradstreet WorldBase; English Proficiency Index (ln EPI) in the year 2012, EF (Education First); LP2, CEPII. The p-value refers to the p-value of the Kleibergen-Paap LM statistic of underidentification. The F-statistic reports the Kleibergen-Paap F statistic for weak identification. Source-industry fixed effects at the 2-digit ISIC industry level. Differentiated refers to goods classified as differentiated according to Rauch (1999) based on ISIC industries. Services industries (compared to manufacturing industries) according to ISIC. COMM refers to the importance of communicating outside of the organization relative to the importance of other tasks; CRTV refers to the importance of complex tasks, relative to the importance of other tasks (Oldenski (2012) and O*NET).

tive to the importance of other tasks (i.e., relative communication intensity), and (2) the importance of complex tasks, relative to the importance of other tasks (i.e., non-routineness intensity).[28] Estimates are shown in Columns 3 and 4 of Table 5. In line with English proficiency serving as a communication cost reducing mechanism, the estimated effects are significantly stronger on multinational production in sectors whose occupations involve a higher degree of interaction with customers and relevant partners (than in sectors with whose occupations require less frequent interactions with external counterparts) and in sectors whose occupations primarily consists of creative, complex, non-routine tasks (than in sectors whose occupations mainly consist of routine tasks).[29]

5.2. English Proficiency and Digital Skills

So far, we have focused on English proficiency as the sole specific skill-related determinant of multinational production. However, MNEs tend to consider a broader set of skills that are likely to interact with language skills in driving their location decisions.

To assess the incidence of these additional specific skills that can interact and specifically complement language skills in shaping multinational production geographical patterns, we turn to digital skills. More precisely, we use variables measuring availability of different digital skills across countries from the OECD ICT Access and Usage by Households and Individuals database. We distinguish between *advanced digital skills* as captured by the ability to modify software applications, write computer codes, or install and replace operating systems; and *basic digital skills* as captured by the ability to work with spreadsheets, produce electronic presentations, or electronic transfer skills. In particular, for data coverage reasons, we focus on the following two specific skills: the share of individuals who have who have written computer codes and the share of individual who have used software for electronic presentations.

Given that we lack an appropriate IV for digital skills, we conduct bivariate analysis of English language proficiency with these two digital skills, H1D (Individuals who have used software for electronic presentations) and H1K (individuals who have written computer code), by way of the generalized propensity score (GPS) framework

[28]The importance score is computed as follows: $M_{sz} = \sum_c \alpha_{zc} \ell_{sc}$, where tasks are denoted by s, occupations by c and industries by z, α_{zc} is the occupational share in a given industry and ℓ_{sc} is an index of the importance of tasks by occupation. This is scaled by the sum of scores for each task in each industry, i.e., $\sum_c M_{sz}$. Specifically, the variables we use stem from the O*NET measures "working with the public" and "creative thinking", respectively, and are scaled in the unit interval as described in Oldenski (2012).

[29]We have also explored whether EPI has heterogeneous effects on the overall extensive margin of multinational production for sectors facing different levels of information and communication frictions. Estimates indicate that the effects are stronger for sectors producing differentiated versus non-differentiated goods, services versus manufacturing sectors, and industries with higher importance of communication and complex relative to other tasks. The estimates are available upon request.

(Hirano and Imbens, 2004, Imai and van Dyk, 2004). This allows for a flexible joint impact of English language proficiency and digital skills on the number of MNEs, whereby we assume selection of the two continuous treatment variables on observable joint determinants–which are the covariates in the specifications whose estimates are reported in the previous sections. Appendix B includes a detailed explanation of this empirical approach.

Estimates indicate that better English proficiency together with advanced digital skills as measured by the share of individuals who have written computer code exhibit a positive and significant impact on the number of MNEs. Graphical analysis that allows for nonlinearities in the joint effect of English proficiency and digital skills specifically suggests that high shares of digital skills in the population boost multinational production, whereas the effect of English proficiency is more or less linear. In contrast, basic digital skills as proxied by the share of individuals that have used software for electronic presentations do not have a significant average effect on the extensive margin of multinational production. Appendix B discusses the estimation results more in-depth.

6. Concluding Remarks

Human capital in general and skills in particular are important determinants of multinational production and the associated benefits in terms of inclusive growth. In this paper, we have examined the impact of English language proficiency on the extensive margins of multinational activity using an IV approach whereby such a proficiency is instrumented by measures of proximity between the domestic and English languages. We find that host countries with higher levels of English mastery tend to attract more MNEs. This effect is greater in sectors with higher degree of differentiation and in sectors whose main occupations primarily involve tasks that require more interactive communication and are less routine, thus suggesting that English proficiency acts as a mechanism that helps reduce information and communication frictions. In addition, our results indicate the higher shares of advanced digital skills in the population are associated with more multinational production in the country, whereas the effect of English proficiency is about linear.

References

Aitken, B., Hanson, G. H., and Harrison, A. E. (1997). "Spillovers, foreign investment, and export behavior." *Journal of International Economics*, *43*(1-2), 103–132.

Alfaro, L. (2016). "Multinational firms, value chains, and vertical integration." NBER Reporter, no. 3 (Research Summary).

Alfaro, L., Chanda, A., Kalemli-Ozcan, S., and Sayek, S. (2004). "Fdi and economic growth: the role of local financial markets." *Journal of International Economics*, *64*(1), 89–112.

Alfaro, L., and Charlton, A. (2009). "Intra-industry foreign direct investment." *American Economic Review*, *99*(5), 2096–2119.

Alfaro, L., and Chen, M. X. (2012). "Surviving the global financial crisis: foreign ownership and establishment performance." *American Economic Journal: Economic Policy*, *4*(3), 30–55.

Alfaro, L., and Rodríguez-Clare, A. (2004). "Multinationals and linkages: An empirical investigation." *Economía*, *4*(2), 113–156.

Alfaro-Ureña, A., Manelici, I., and Vasquez, J. P. (2022). "The effects of joining multinational supply chains: New evidence from firm-to-firm linkages." *The Quarterly Journal of Economics*, *137*(3), 1495–1552.

Allen, T. (2014). "Information frictions in trade." *Econometrica*, *82*, 2041–2083.

Anderson, J. E., and van Wincoop, E. (2004). "Trade costs." *Journal of Economic Literature*, *42*(3), 691–751.

Andrenelli, A., Cadestin, C., De Backer, K., Miroudot, S., Rigo, D., and Ye, M. (2018). "Multinational production and trade in services." Tech. Rep. OECD Trade policy papers No. 12, OECD Publishing, Paris.

Bahar, D. (2020). "The hardships of long distance relationships: time zone proximity and the location of mnc's knowledge-intensive activities." *Journal of International Economics*, *125*, 103311.

Balsvik, R. (2011). "Is labor mobility a channel for spillovers from multinationals? evidence from norwegian manufacturing." *The Review of Economics and Statistics*, *93*(1), 285–297.

Blalock, G., and Gertler, P. (2008). "Welfare gains from foreign direct investment through technology transfer to local suppliers." *Journal of International Economics*, *74*, 402–421.

Blinder, A. S. (2006). "Offshoring: the next industrial revolution?" *Foreign Affairs*, *85*, 113–128.

Blomstroem, M., and Kokko, A. (1998). "Multinational corporations and spillovers." *Journal of Economic Surveys*, *12*, 247–277.

Blomstroem, M., and Kokko, A. (2002). "Fdi and human capital: a research agenda." Tech. Rep. No. 195, OECD Development Centre Working Papers.

Borensztein, E., De Gregorio, J., and Lee, J.-W. (1998). "How does foreign direct investment affect economic growth?" *Journal of International Economics*, *45*, 115–135.

Bénétrix, A., Pallan, H., and Panizza, U. (2022). "The elusive link between fdi and economic growth." Tech. Rep. DP17692, CEPR Press Discussion Paper.

Carballo, J., Marra de Artiñano, I., Ottaviano, G., and Volpe Martincus, C. (forthcoming). "Linkages with multinationals: What are the effects on domestic firms' exports?" Tech. rep., IDB Working Paper.

Cristea, A. (2015). "The effect of communication costs on trade in headquarter services." *Review of World Economics*, *151*, 255–289.

Defever, F. (2012). "The spatial organization of multinational firms." *Canadian Journal of Economics/Revue canadienne d'economique*, *45*, 672–697.

Egger, P. H., and Lassmann, A. (2015). "The causal impact of common native language on international trade: evidence from a spatial regression discontinuity design." *The Economic Journal*, *116*(584), 699–745.

Egger, P. H., and Toubal, F. (2016). "Common spoken language and international trade." In V. Ginsburgh, and S. Weber (Eds.), *The Palgrave Handbook of Economics and Language*, vol. 1, chap. 9.

Fearon, J. D. (2003). "Ethnic and cultural diversity by country." *Journal of Economic Growth*, *8*(2), 195–222.

Feng, X., Lin, F., and Sim, N. C. S. (2019). "The effect of language on foreign direct investment." *Oxford Economic Papers*, *71*(1), 269–291.

Fidrmuc, J., and Fidrmuc, J. (2016). "Foreign languages and trade: evidence from a natural experiment." *Empirical Economics*, *50*(7), 31–49.

Garetto, S., Oldenski, L., and Ramondo, N. (2019). "Multinational expansion in time and space." Tech. rep., NBER Working Papers 25804, National Bureau of Economic Research, Inc.

Ginsburgh, V., and Weber, S. (2020). "The economics of language." *Journal of Economic Literature*, *58*(2), 348–404.

Guillouet, L., Khandelwal, A. K., Macchiavello, R., Malhotra, M., and Teachout, M. (2024). "Language barriers in multinationals and knowledge transfers." *The Review of Economics and Statistics*, 1–57.

Gumpert, A. (2018). "The organization of knowledge in multinational firms." *Journal of the European Economics Association*, *16*(6), 1929–1976.

Hanson, G. H. (2001). "U.s.-mexico integration and regional economies: Evidence from border-city pairs." *Journal of Urban Economics*, *50*(2), 259–287.

Harding, T., and Javorcik, B. S. (2012). "Foreign direct investment and export upgrading." *The Review of Economics and Statistics*, *94*(4), 964–980.

Hirano, K., and Imbens, G. (2004). "The propensity score with continuous treatments." In A. Gelman, and X.-L. Meng (Eds.), *Applied bayesian modeling and causal inference from incomplete-data perspectives*, 73–84, Chichester, UK: Wiley.

Hutchinson, W. K. (2002). "Does ease of communication increase trade? commonality of language and bilateral trade." *Scottish Journal of Political Economy*, *49*(5), 544–556.

Imai, K., and van Dyk, D. A. (2004). "Causal inference with general treatment regimes: generalizing the propensity score." *Journal of the American Statistical Association*, *99*, 854–866.

Isphording, I. E., and Otten, S. (2013). "The costs of babylon - linguistic distance in applied economics." *Review of International Economics*, *21*, 2.

Javorcik, B. S. (2004). "Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages." *American Economic Review*, *94*(3), 605–627.

Kim, M.-H., and Lee, H.-H. (2010). "Linguistic and nonlinguistic factors determining proficiency of english as a foreign language: a cross-country analysis." *Applied Economics*, *42*(18), 2347–2364.

Konara, P., and Wei, Y. (2019). "The complementarity of human capital and language capital in foreign direct investment." *International Business Review*, *28*(2), 391–404.

Ku, H., and Zussman, A. (2010). "Lingua franca: The role of english in international trade." *Journal of Economic Behavior & Organization*, *75*(2), 250–260.

Laitin, D. D. (2000). "Language conflict and violence: the straw that strengthens the camel's back." *European Journal of Sociology*, *41*(1), 97–137.

Melitz, J. (2008). "Language and foreign trade." *European Economic Review*, *52*(4), 667–699.

Melitz, J., and Toubal, F. (2014). "Native language, spoken language, translation and trade." *Journal of International Economics*, *93*(2), 351–363.

Muendler, M., Rauch, J., and Tocoian, O. (2012). "Employee spinoffs and other entrants: Stylized facts from brazil." *International Journal of Industrial Organization*, *30*(5), 447–458.

Neeley, T. (2012). "Global business speaks english." Harvard Business Review, https://hbr.org/2012/05/global-business-speaks-english.

Oldenski, L. (2012). "Export versus fdi and the communication of complex information." *Journal of International Economics*, *87*(2), 312–322.

Peltokorpi (2015). "Corporate language proficiency and reverse knowledge transfer in multinational corporations: interactive effects of communication media richness and commitment to headquarters." *Journal of International Management*, *41*, 49–62.

Poole, J. P. (2013). "Knowledge transfers from multinational to domestic firms: Evidence from worker mobility." *The Review of Economics and Statistics*, *95*(2), 393–406.

Ramondo, N., Rodríguez-Clare, A., and Tintelnot, F. (2015). "Multinational production: data and stylized facts." *American Economic Review: Papers & Proceedings*, *105*(5), 530–536.

Rauch, J. E. (1999). "Networks versus markets in international trade." *Journal of International Economics*, *48*(1), 7–35.

Rodrik, D. (2007). *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*. Princeton University Press.

Rodríguez-Clare, A. (1996). "Multinationals, linkages and economic development." *American Economic Review*, *86*(4), 852–873.

Ryder, B. (2014). "The english empire." The Economist, https://www.economist.com/business/2014/02/15/the-english-empire, last accessed 16 September 2023.

Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2015). "Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to ulrich and miller." *Journal of Experimental Psychology: General*, *144*(6), 1146–1152.

Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2020). "Specification curve analysis." *Nature Human Behaviour*, *4*(11), 1208–1214.

Tenzer, H., Pudelko, M., and Harzing, A.-W. (2013). "The impact of language barriers on trust formation in multinational teams." *Journal of International Business Studies*, *45*, 508–535.

Volpe Martincus, C., Marra de Artiñano, I., Sztajerowska, M., and Carballo, J. (2021). "Making the invisible visible: investment promotion and multinational production in latin america and the caribbean." Tech. rep., Inter-American Development Bank.

Volpe Martincus, C., and Sztajerowska, M. (2019). "How to solve the investment promotion puzzle: a mapping of investment promotion agencies in latin america and the caribbean and oecd countries." Tech. rep., OECD and Inter-American Development Bank.

Volpe Martincus, C., and Sztajerowska, M. (2025). "Fast forward in investment promotion: new digital tools for new policy challenges." Tech. rep., Inter-American Development Bank, mimeo.

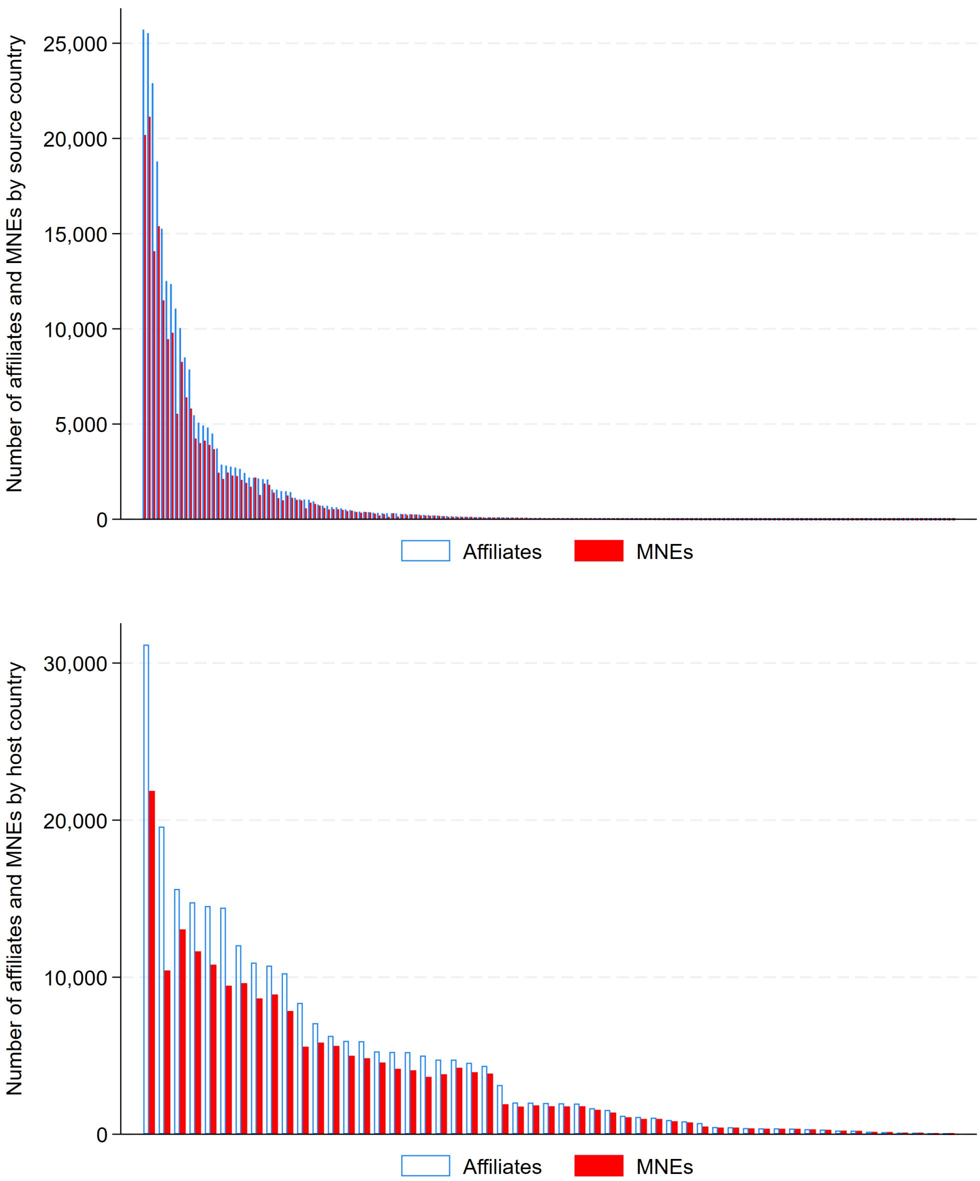
Wang, J.-Y., and Blomstrom, M. (1992). "Foreign investment and technology transfer : A simple model." *European Economic Review*, *36*(1), 137–155.

Welch, D. E., and Welch, L. S. (2008). "The importance of language in international knowledge transfer." *Management International Review*, *48*(3), 339–360.

Appendix

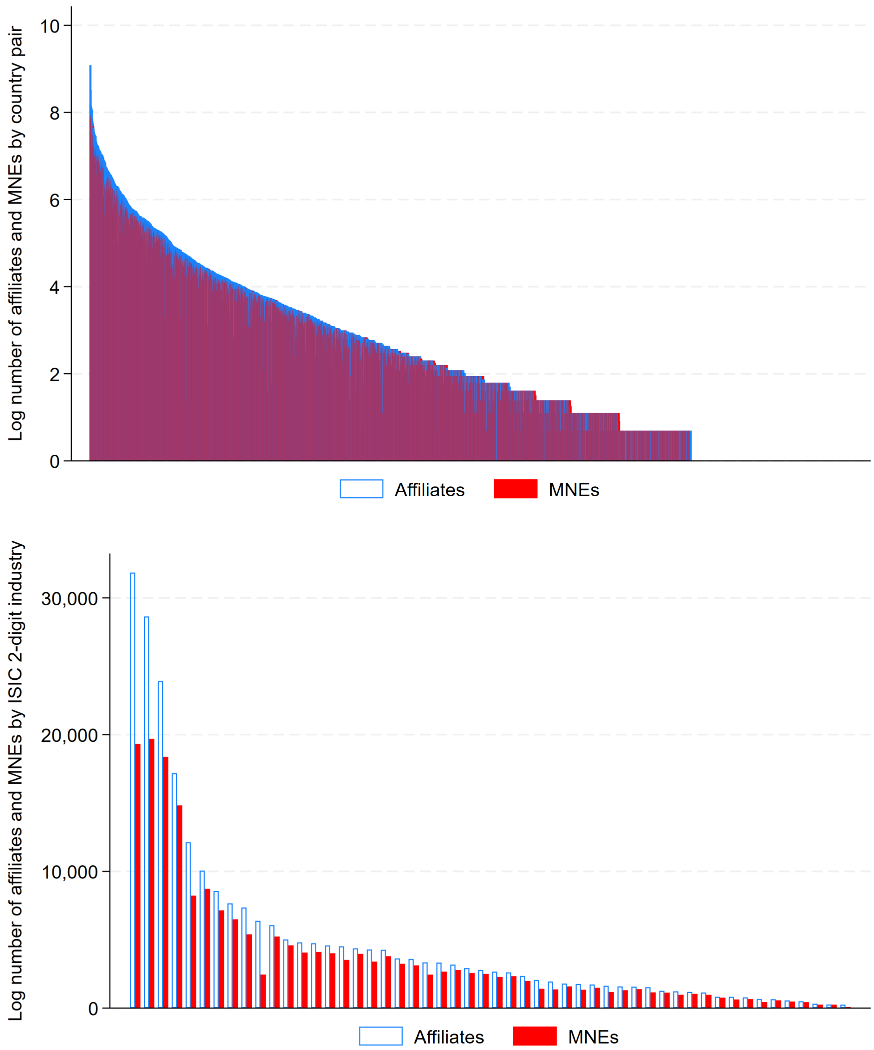
A. Supplementary Figures and Tables

Figure A1: Distribution of the number of MNEs and affiliates by home and host country



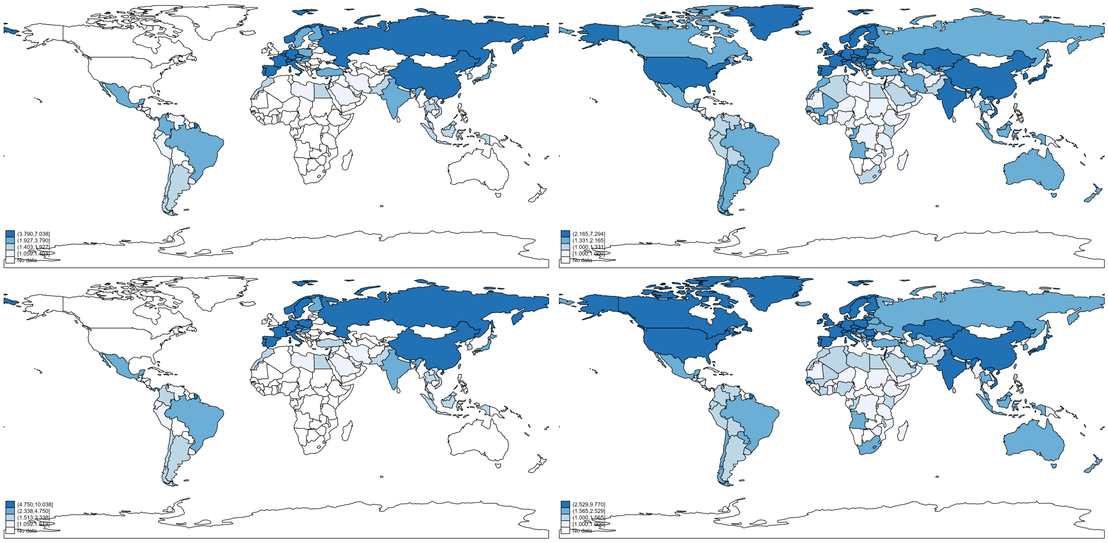
Source: Authors' calculations based on Dun&Bradstreet WorldBase, 2019. Notes: Number of MNEs and number of affiliates, by home country in upper and by host country in lower panel.

Figure A2: Distribution of the number of MNEs and affiliates by country-pair and industry



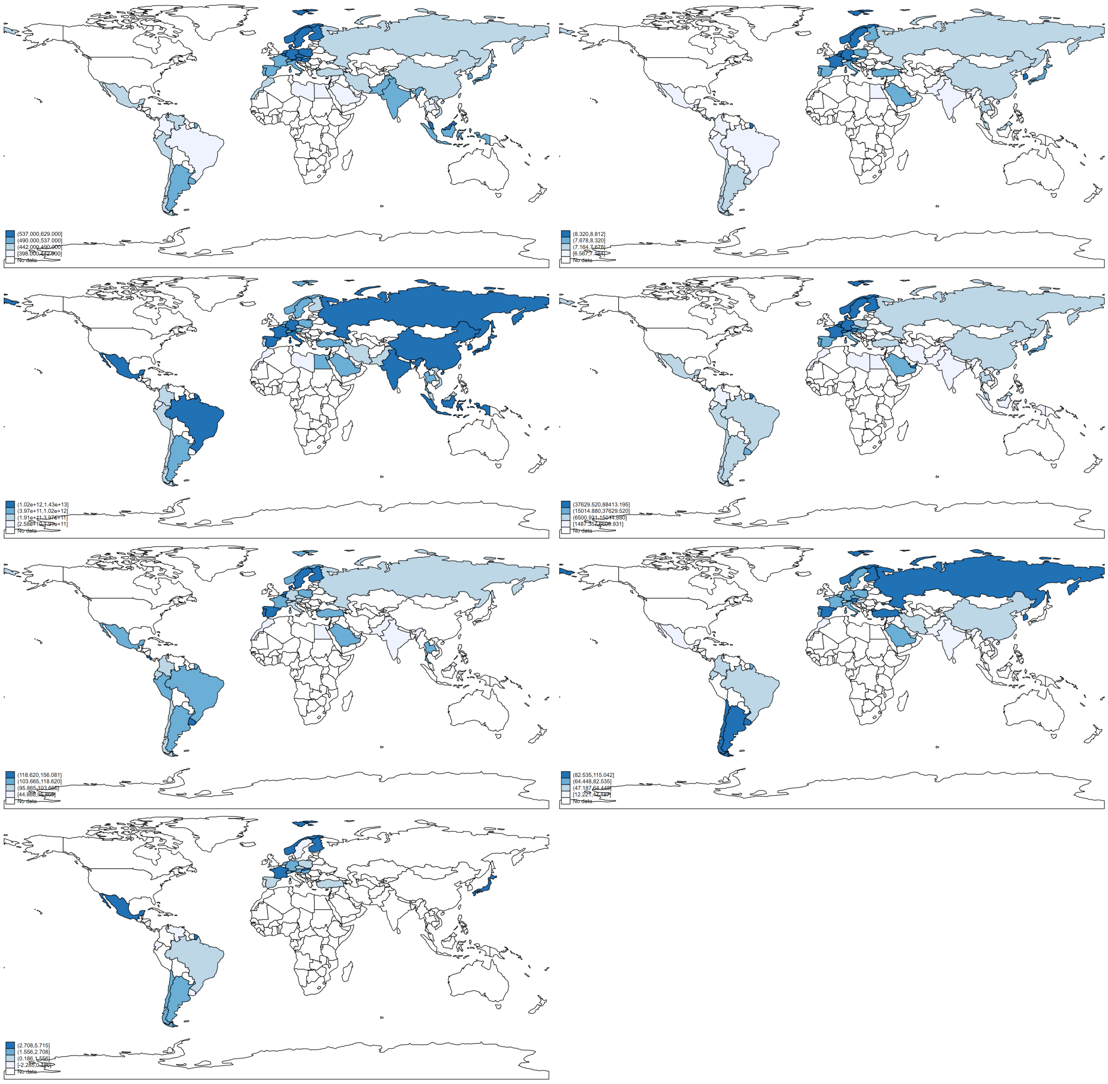
Source: Dun&Bradstreet WorldBase, 2019. Notes: Log number of MNEs and log number of affiliates, by country-pair in upper and industry in lower panel.

Figure A3: COVERAGE I (FOR ILLUSTRATION)



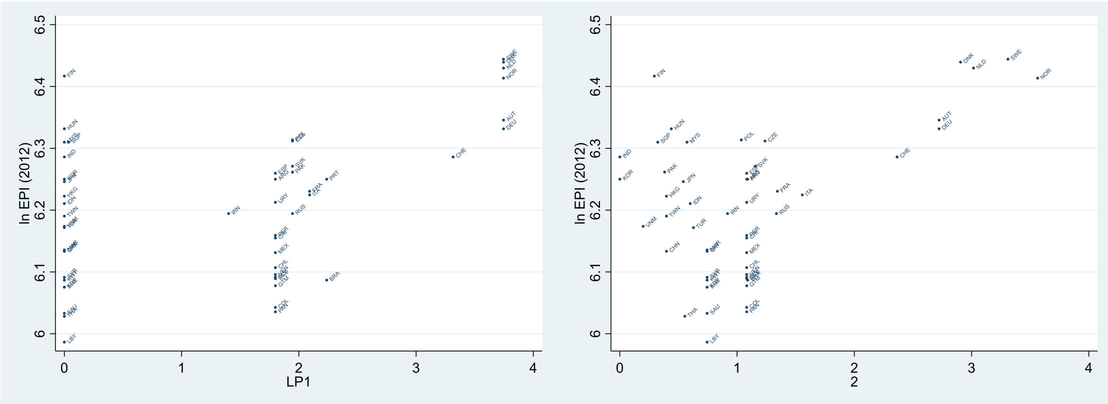
Notes: Upper panel: Number of MNEs by host (left) and source (right) country; lower panel: Number of affiliates by host (left) and source (right) country, Dun&Bradstreet WorldBase, 2019, over source countries and ISIC 4-digit industries. Countries for which the EPI is not available have been excluded because they are not used for the analysis.

Figure A4: Coverage II (for illustration)



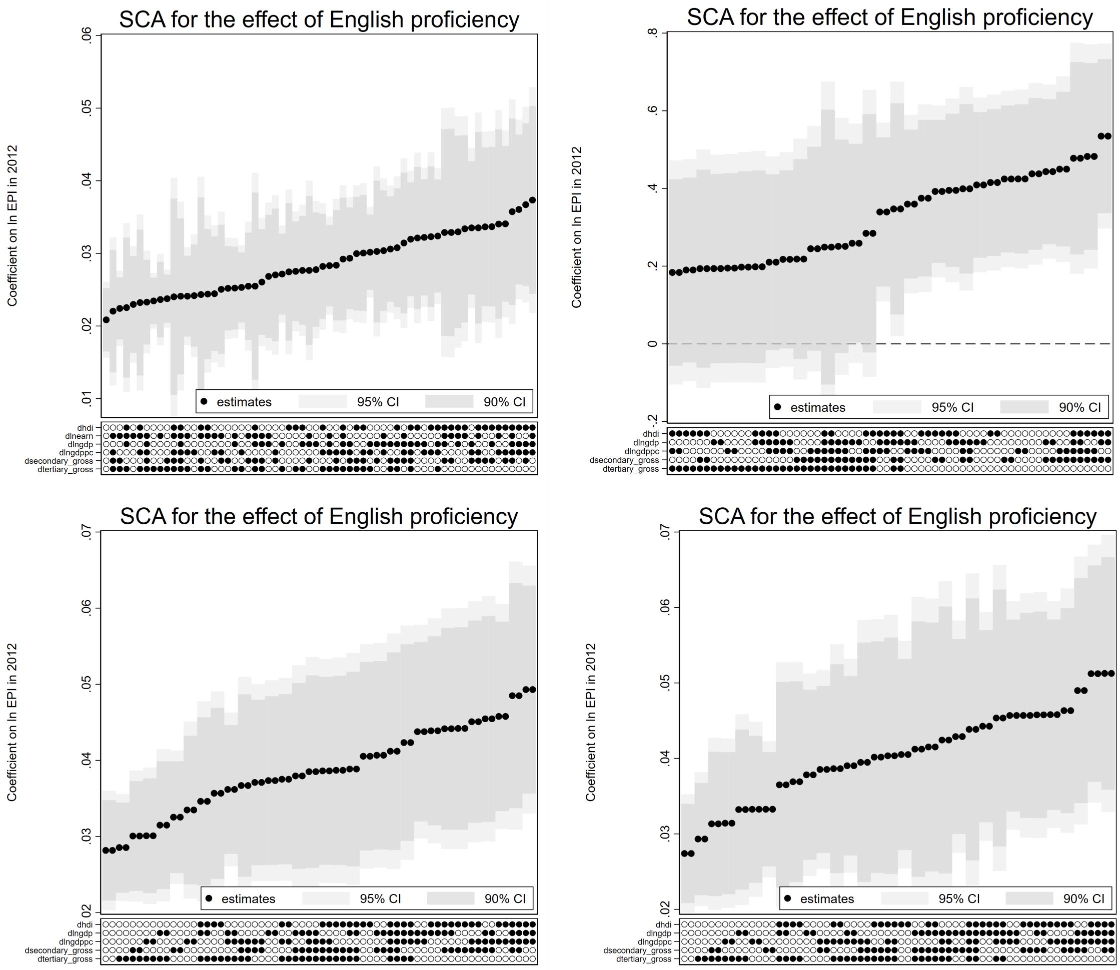
Notes: Left (top to bottom): English Proficiency Index (EPI) in 2012, EF (Education First); GDP in constant 2015 USD, World Bank, 2019; secondary school enrollment, World Bank, 2019; Investment Promotion Budget, Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025). Right (top to bottom): log earnings, ILOSTAT, 2019; GDP per capita in constant 2015 USD, World Bank, 2019; tertiary school enrollment, World Bank, 2019. Variables in host countries.

Figure A5: Validity of IV



Notes: The left panel shows the unconditional correlation between ln EPI (in 2012) and LP1 (LP2 in the right panel). Source: EPI, Education First (EF); LP1 and LP2, CEPII.

Figure A6: Specification curve analysis



Notes: This plot uses the methodology and different combinations of independent variables used in Tables 2 (upper left without and upper right with home country-industry fixed effects) and **??** (with LP1 as the IV in the lower left and LP2 as the IV in the lower left panel). Dependent variable: log number of MNEs, Dun&Bradstreet WorldBase.

Table A1: THE EFFECT OF ENGLISH PROFICIENCY ON THE NUMBER OF MNES (IV: LP1)

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			Δlog number of MNEs		
A. Without home country-sector fixed effects					
$\ln\text{EPI}_{it_0}$	0.028***	0.037***	0.044***	0.039***	0.052***
	(0.004)	(0.008)	(0.007)	(0.008)	(0.016)
Δln(earnings)		0.030			
		(0.104)			
Δln(gdp)		-0.448*	-1.682***	-1.332***	-1.866*
		(0.182)	(0.352)	(0.331)	(0.942)
Δln(gdp pc)			1.330***	1.034*	1.367*
			(0.303)	(0.467)	(0.679)
Δsecondary ed.				0.000	
				(0.001)	
Δtertiary ed.				-0.003	
				(0.002)	
lnIPA Budget					-0.012
					(0.015)
IPA Country Prioritization					-0.058+
					(0.031)
IPA Sector Prioritization					0.043
					(0.028)
N	52023	35202	50959	34668	32212
F-stat.	183.398	57.474	84.673	32.306	32.588
Home-industry FE	No	No	No	No	No
B. With home country-sector fixed effects					
$\ln\text{EPI}_{it_0}$	1.132*	1.282**	1.017*	0.967+	1.030*
	(0.434)	(0.450)	(0.394)	(0.521)	(0.452)
Δln(earnings)		0.118			
		(0.075)			
Δln(gdp)		-0.034	-0.342	-1.337**	-0.171
		(0.174)	(0.454)	(0.428)	(0.785)
Δln(gdp pc)			0.274	1.151*	0.124
			(0.454)	(0.558)	(0.644)
Δsecondary ed.				-0.002	
				(0.002)	
Δtertiary ed.				-0.005	
				(0.003)	
lnIPA Budget					0.007
					(0.014)
IPA Country Prioritization					0.014
					(0.049)
IPA Sector Prioritization					0.066**
					(0.023)
N	50926	34136	49866	33610	31240
F-stat.	11.965	7.734	11.887	2.861	10.970
Home-industry FE	Yes	Yes	Yes	Yes	Yes

Notes: Linear IV regressions in first differences with standard errors clustered at the host country level in parentheses, $^{+}(p < 0.1)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$. Language proximity (LP1) as IV. Differences refer to the difference of variables for years 2019 and 2012. *Sources*: Number of MNEs, Dun&Bradstreet WorldBase; English Proficiency Index (ln EPI) in the year 2012, EF (Education First); Earnings, ILOSTAT; GDP, GDP per capita, secondary and tertiary school enrollment, World Bank; IPA (Investment Promotion Activity) variables 2016, Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025). LP1, CEPII. The F-statistic reports the Kleibergen-Paap F statistic for weak identification. Source-industry fixed effects at the 2-digit ISIC industry level.

Table A2: First stage of IV regressions (IV: LP1)

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			ln EPI_{it_0}		
A. Without home country-sector fixed effects					
LP1	2.243***	1.648***	1.389***	1.444***	0.966***
	(0.165)	(0.217)	(0.151)	(0.254)	(0.169)
Δln(earnings)		0.128			
		(2.105)			
Δln(gdp)		12.151***	27.617***	-0.471	19.875**
		(2.683)	(5.066)	(8.616)	(6.949)
Δln(gdp pc)			-12.772*	19.589*	-3.837
			(6.219)	(8.616)	(8.150)
Δsecondary ed.				-0.017	
				(0.045)	
Δtertiary ed.				0.175***	
				(0.037)	
lnIPA Budget					0.544***
					(0.162)
IPA Country Prioritization					0.857
					(0.573)
IPA Sector Prioritization					0.377
					(0.294)
N	52023	35202	50959	34668	32212
Adj. R-sq.	0.603	0.901	0.908	0.953	
Home-industry FE	No	No	No	No	No
B. With home country-sector fixed effects					
LP1	0.35***	0.048**	0.044**	0.034+	0.052***
	(0.010)	(0.017)	(0.013)	(0.020)	(0.016)
Δln(earnings)		-0.057			
		(0.064)			
Δln(gdp)		0.138	-0.394	-0.026	-0.460
		(0.189)	(0.296)	(0.400)	(0.435)
Δln(gdp pc)			0.614*	0.404	0.884*
			(0.304)	(0.527)	(0.373)
Δsecondary ed.				0.002	
				(0.002)	
Δtertiary ed.				-0.003	
				(0.003)	
lnIPA Budget					0.012
					(0.010)
IPA Country Prioritization					-0.029
					(0.027)
IPA Sector Prioritization					-0.015
					(0.012)
N	50926	34136	49866	33610	31240
Adj. R-sq.	0.275	0.335	0.350	0.467	
Home-industry FE	Yes	Yes	Yes	Yes	Yes

Notes: First stage of linear IV regressions in first differences with standard errors clustered at the host country level in parentheses, $^{+}(p < 0.1)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$. Differences refer to the difference of variables for years 2019 and 2012. *Sources*: English Proficiency Index (ln EPI) in the year 2012, EF (Education First); LP1, CEPII; Earnings, ILOSTAT; GDP, GDP per capita, secondary and tertiary school enrollment, World Bank; IPA (Investment Promotion Activity) variables 2016, Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025). Source-industry fixed effects at the 2-digit ISIC industry level.

Table A3: The effect of English proficiency on the number of affiliates (IV: LP2)

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Δlog number of MNEs foreign affiliates				
A. Without home country-sector fixed effects					
$\ln\text{EPI}_{it_0}$	0.031***	0.044***	0.051***	0.047***	0.075**
	(0.005)	(0.010)	(0.009)	(0.012)	(0.026)
Δln(earnings)		0.020			
		(0.130)			
Δln(gdp)		-0.517*	-2.004***	-1.485***	-2.824^{+}
		(0.213)	(0.459)	(0.380)	(1.580)
Δln(gdp pc)			1.584***	1.060*	1.930^{+}
			(0.390)	(0.533)	(1.103)
Δsecondary ed.				0.000	
				(0.001)	
Δtertiary ed.				-0.004	
				(0.003)	
lnIPA Budget					-0.026
					(0.020)
IPA Country Prioritization					-0.072^{+}
					(0.042)
IPA Sector Prioritization					0.033
					(0.036)
N	52023	35202	50959	34668	32212
F-stat.	86.541	33.838	55.174	22.095	16.403
Home-industry FE	No	No	No	No	No
B. With home country-sector fixed effects					
$\ln\text{EPI}_{it_0}$	1.072*	1.320**	1.011**	0.730*	0.904*
	(0.416)	(0.395)	(0.366)	(0.305)	(0.347)
Δln(earnings)		0.130*			
		(0.076)			
Δln(gdp)		-0.031	-0.434	-1.388**	-0.194
		(0.184)	(0.445)	(0.395)	(0.731)
Δln(gdp pc)			0.370	1.238*	0.198
			(0.443)	(0.505)	(0.606)
Δsecondary ed.				-0.001	
				(0.001)	
Δtertiary ed.				0.003	
				(0.002)	
lnIPA Budget					0.007
					(0.014)
IPA Country Prioritization					0.0001
					(0.041)
IPA Sector Prioritization					0.068**
					(0.024)
N	50926	34136	49866	33610	31240
F-stat.	23.519	15.998	25.635	11.105	28.446
Home-industry FE	Yes	Yes	Yes	Yes	Yes

Notes: Linear IV regressions in first differences with standard errors clustered at the destination level in parentheses, $^{+}(p < 0.1)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$. Language proximity (LP2) as IV. in in columns 1-4 and columns 5-8, respectively. Differences refer to the difference of variables for years 2019 and 2012. *Sources*: Number of affiliates, Dun&Bradstreet WorldBase; English Proficiency Index (ln EPI) in the year 2012, EF (Education First); GDP, GDP per capita, secondary and tertiary school enrollment, World Bank; IPA (Investment Promotion Activity) variables 2016, Volpe Martincus et al. (2021), Volpe Martincus and Sztajerowska (2019, 2025); LP2, CEPII. The F-statistic reports the Kleibergen-Paap F statistic for weak identification. Source-industry fixed effects at the 2-digit ISIC industry level.

Table A4: The effect of English proficiency on the probability of multinational activity

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			ΔI(Number of MNEs $>$ 0)		
			A. OLS		
$\ln\text{EPI}_{it_0}$	0.0003***	0.0005**	0.0005***	0.001***	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.000)
Δln(earnings)		0.001			
		(0.001)			
Δln(gdp)		-0.006+	-0.017***	-0.022***	-0.010
		(0.003)	(0.005)	(0.004)	(0.009)
Δln(gdp pc)			0.016***	0.021***	0.016*
			(0.004)	(0.005)	(0.007)
Δsecondary ed.				-0.0001	
				(0.00001)	
Δtertiary ed.				-0.000	
				(0.0001)	
lnIPA Budget					0.000**
					(0.000)
IPA Country Prioritization					0.009**
					(0.003)
IPA Sector Prioritization					(0.002)+
					(0.001)
N	6,446,072	4,021,840	6,078,807	3,520,057	3,155,834
R-sq.	0.002	0.002	0.002	0.003	0.004
			B. IV		
$\ln\text{EPI}_{it_0}$	0.001***	0.001***	0.001***	0.002***	0.002*
	(0.0001)	(0.0003)	(0.0003)	(0.0004)	(0.001)
Δln(earnings)		-0.002			
		(0.002)			
Δln(gdp)		-0.020*	-0.048***	-0.039**	-0.100*
		(0.008)	(0.011)	(0.012)	(0.049)
Δln(gdp pc)			0.035***	0.015*	0.064+
			(0.009)	(0.014)	(0.033)
Δsecondary ed.				-0.0001	
				(0.00004)	
Δtertiary ed.				-0.0002*	
				(0.0001)	
lnIPA Budget					-0.001
					(0.001)
IPA Country Prioritization					0.008*
					(0.004)
IPA Sector Prioritization					0.001
					(0.001)
N	5,077,821	3,178,507	4,810,852	2,833,223	2,598,410
F-stat.	418.504	43.500	101.345	29.350	13.504

Notes: Linear probability model estimated by IV in first differences with standard errors clustered at the host country level in parentheses, $^{+}(p<0.1)$, $^{*}(p<0.05)$, $^{**}(p<0.01)$, $^{***}(p<0.001)$. The dependent variable is 0 if there is no multinational activity in a country-industry pair, and 1 if there is activity, whereby the difference between years 2019 and 2012 is used. *Sources*: English Proficiency Index (ln EPI) in the year 2012, EF (Education First); LP2, CEPII;. LP2 as IV. The p-value refers to the p-value of the Kleibergen-Paap LM statistic of underidentification. The F-statistic reports the Kleibergen-Paap F statistic for weak identification.

B. English language and ICT skills

In this appendix, we explore the joint effects of English language proficiency and two digital skills, H1D (Individuals who have used software for electronic presentations) and H1K (individuals who have written computer code) on multinational production. To do so, we use the generalized propensity score (GPS) framework (Hirano and Imbens, 2004, Imai and van Dyk, 2004). This allows for a flexible joint impact of English language proficiency and digital skills on the number of MNEs, whereby we assume selection of the two continuous treatment variables on observable joint determinants–the set of control variables included in our main specifications.

The first step is to estimate EPI_{it} (referring to $\ln EPI$) and DS_{it} (referring to the respective digital skill variable) as functions of the nonstochastic regressor vector X_{it}, which is also used in (1):[30]

$$EPI_{it} = f_{EPI}(X_{it}, \delta_{EPI}) + \epsilon_{EPI_{it}}, \qquad DS_{it} = f_{DS}(X_{it}, \delta_{DS}) + \epsilon_{DS_{it}}, \tag{B.1}$$

where $f_{EPI}(\cdot)$ and $f_{DS}(\cdot)$ are unknown, treatment-specific functions, $(\delta_{EPI}, \delta_{DS})$ are unknown parameter vectors, and $(\epsilon_{EPI_{it}}, \epsilon_{DS_{it}})$ are random terms uncorrelated with X_{it} and y_{it}. Assuming weak unconfoundedness for the continuous endogenous treatments implies that the potential number of multinational firms is conditionally independent of the actual levels of treatment, EPI_{it} and DS_{it}. This is obtained by constructing the scalar GPS, G_{it}, which is a function of $(\epsilon_{EPI_{it}}, \epsilon_{DS_{it}})$, as:

$$\hat{G}_{it} = \frac{1}{\sqrt{(2\pi)^2 det(\hat{\Sigma}_{EPI,DS})}} \exp\left(-0.5(\hat{\epsilon}_{EPI_{it}}, \hat{\epsilon}_{DS_{it}})\hat{\Sigma}^{-1}_{EPI,DS}(\hat{\epsilon}_{EPI_{it}}, \hat{\epsilon}_{DS_{it}})'\right), \tag{B.2}$$

where $det(\cdot)$ refers to the determinant, $\Sigma_{EPI,DS} = N^{-1}(\epsilon_{EPI_{it}}, \epsilon_{DS_{it}})'(\epsilon_{EPI_{it}}, \epsilon_{DS_{it}})$, and the term in the exponent is (minus-one-half times) the Mahalanobis distance of unit *it* to the average.

Finally, defining $Z_{it} = (EPI_{it}, DS_{it}, \hat{G}_{it})$, we specify a flexible function of y_{it} on Z_{it} as:

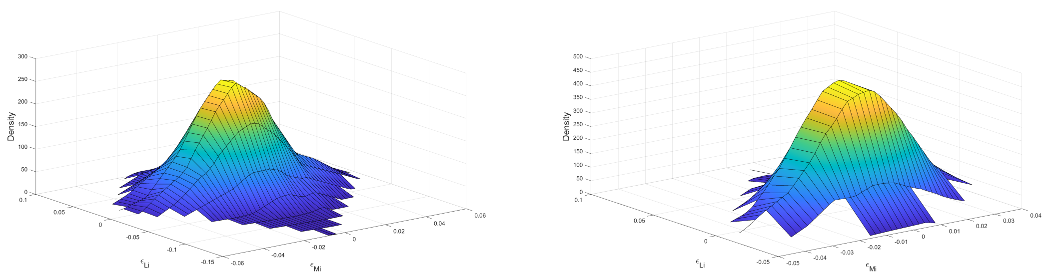
$$y_{it} = h(Z_{it}) + v_{it}. \tag{B.3}$$

In the last step we obtain the average hypothetical EPI, DS-specific level of multinational activity, $\hat{y}(EPI, DS) = N^{-1} \sum_{it=0}^{1} \hat{y}_{it}(EPI, DS)$, whereby the relationship of $\hat{y}(EPI, DS)$ to EPI, DS is the average dose–response function of the number of MNEs as a function of EPI, DS.

[30] The regressor vector is restricted to including the HDI and a squared term of the HDI.

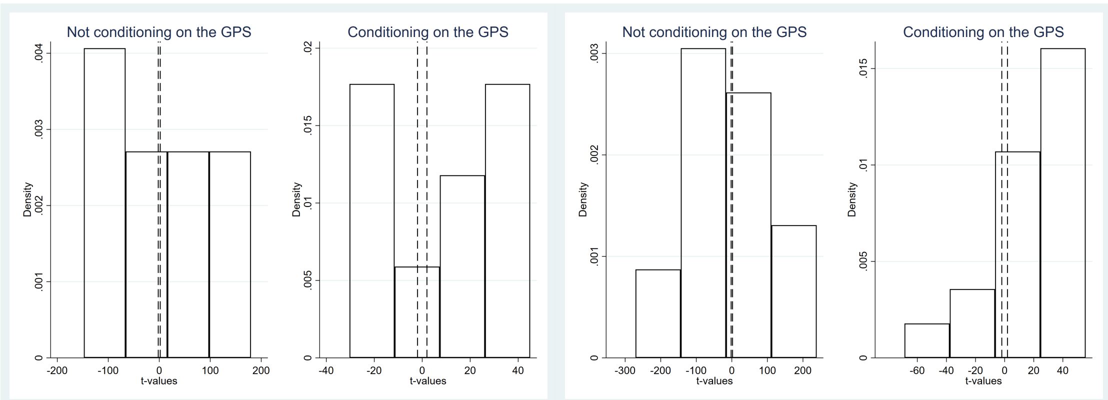
Figure B1 plots a bivariate-normal approximation of the bivariate densities of $(\epsilon_{EPI_{it}}, \epsilon_{DS_{it}})$. Figures B3 and B2 evaluate common support and covariate balancing by the GPS. Figure B3 suggests that the overlap in the bivariate densities $\hat{G}_{it}$ for observations within and outside of a constructed set of 9 groups (consisting of 3×3 size classes in EPI_{it}-DS_{it}-space) is good. Figure B2, which compares t-statistics from a test against the equality of means between the 9 treatment-level groups in EPI_{it}-DS_{it}-space without and with conditioning on the GPS, furthermore indicates that conditioning on the GPS substantially improves covariate balancing. Note that we condition on unbalanced covariates in Z_{it} in the estimation of the dose-response function.

Figure B1: Bivariate density of the residual EPI and ICT treatments



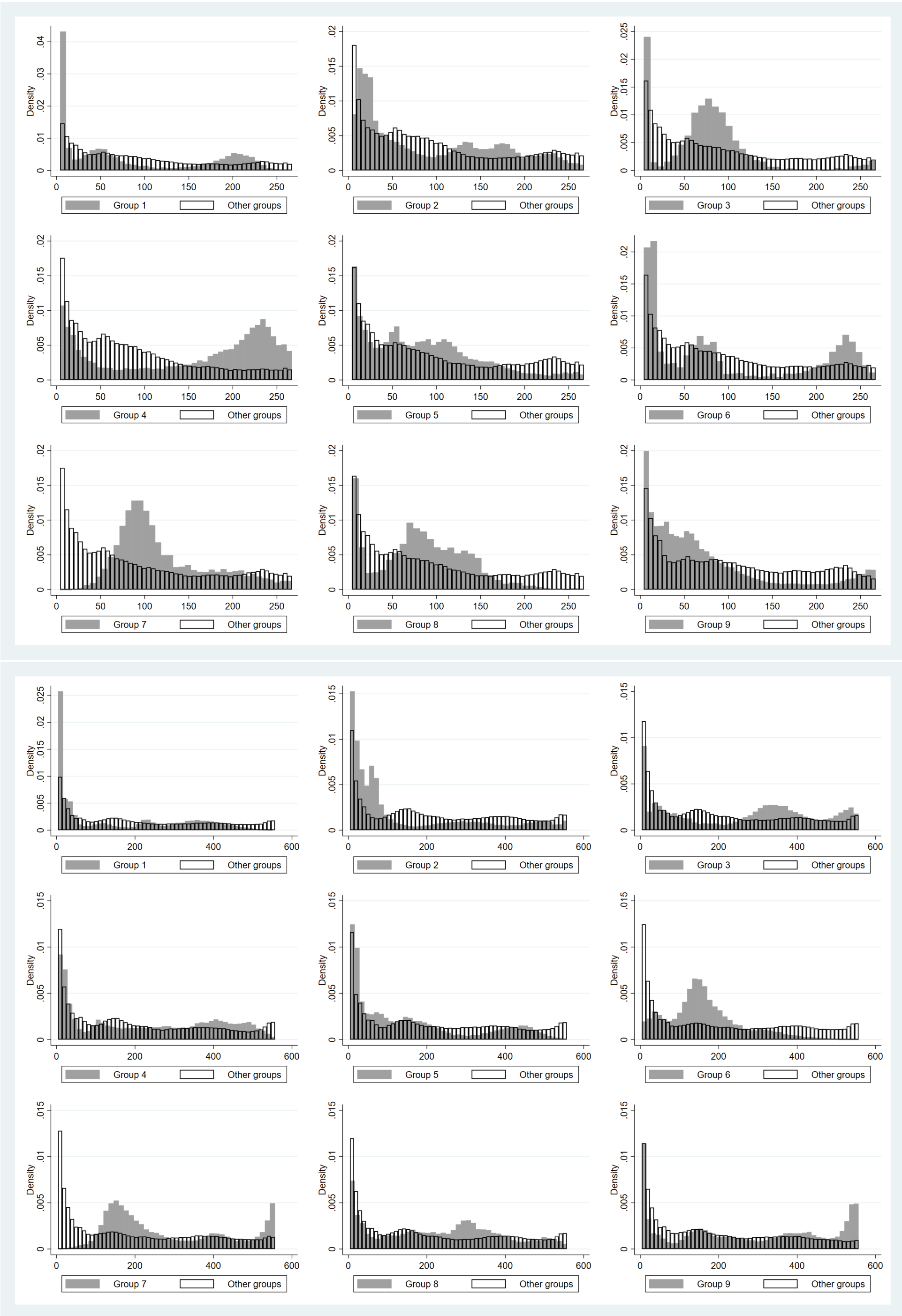
Notes: The figures plots bivariate estimates $\hat{G}_i$ of the GPS according to Equation (3). The digital skill variable refers to H1D (Individuals who have used software for electronic presentations) in the left panel and H1K (individuals who have written computer code) in right panel. Sources: EF (Education First) and OECD ICT Access and Usage by Households and Individuals Database.

Figure B2: Distribution of t-values from testing against mean equality for all covariates



Notes: The left panel reports t-statistics based on simple tests against the equality of means for all covariates (with the exception of fixed effects) involved in the estimation of equation (3). The right panel illustrates t-statistics based on 9 groups (3×3 in EPI_{it}-DS_{it}-space) according to the distribution of ln EPI and digital skills of approximately the same size, and 20-40 blocks of $\hat{G}_{it}$. The digital skill variable refers to H1D (Individuals who have used software for electronic presentations) in the left panel and H1K (individuals who have written computer code) in right panel. Sources: EF (Education First) and OECD ICT Access and Usage by Households and Individuals Database.

Figure B3: Common support of the GPS in 9 treatment groups



Notes: The figures is based on 9 groups (3×3 in EPI_{it}-DS_{it}-space) based on the distribution of ln EPI and digital skills of approximately the same size. The GPS is evaluated at the median of each of the two treatments within each group. The digital skill variable refers to H1D (Individuals who have used software for electronic presentations) in the upper panel and H1K (individuals who have written computer code) in the lower panel. Sources: EF (Education First) and OECD ICT Access and Usage by Households and Individuals Database.

Table A5 reports the estimated coefficients of interest. Estimates indicate that better English proficiency together with advanced digital skills as measured by the share of individuals who have written computer code exhibit a positive and significant impact on the number of MNEs. Graphical analysis that allows for nonlinearities in the joint effect of English proficiency and digital skills specifically suggests that high shares of digital skills in the population boost multinational production, whereas the effect of English proficiency is more or less linear. In contrast, basic digital skills as proxied by the share of individuals that have used software for electronic presentations do not have a significant average effect on the extensive margin of multinational production.

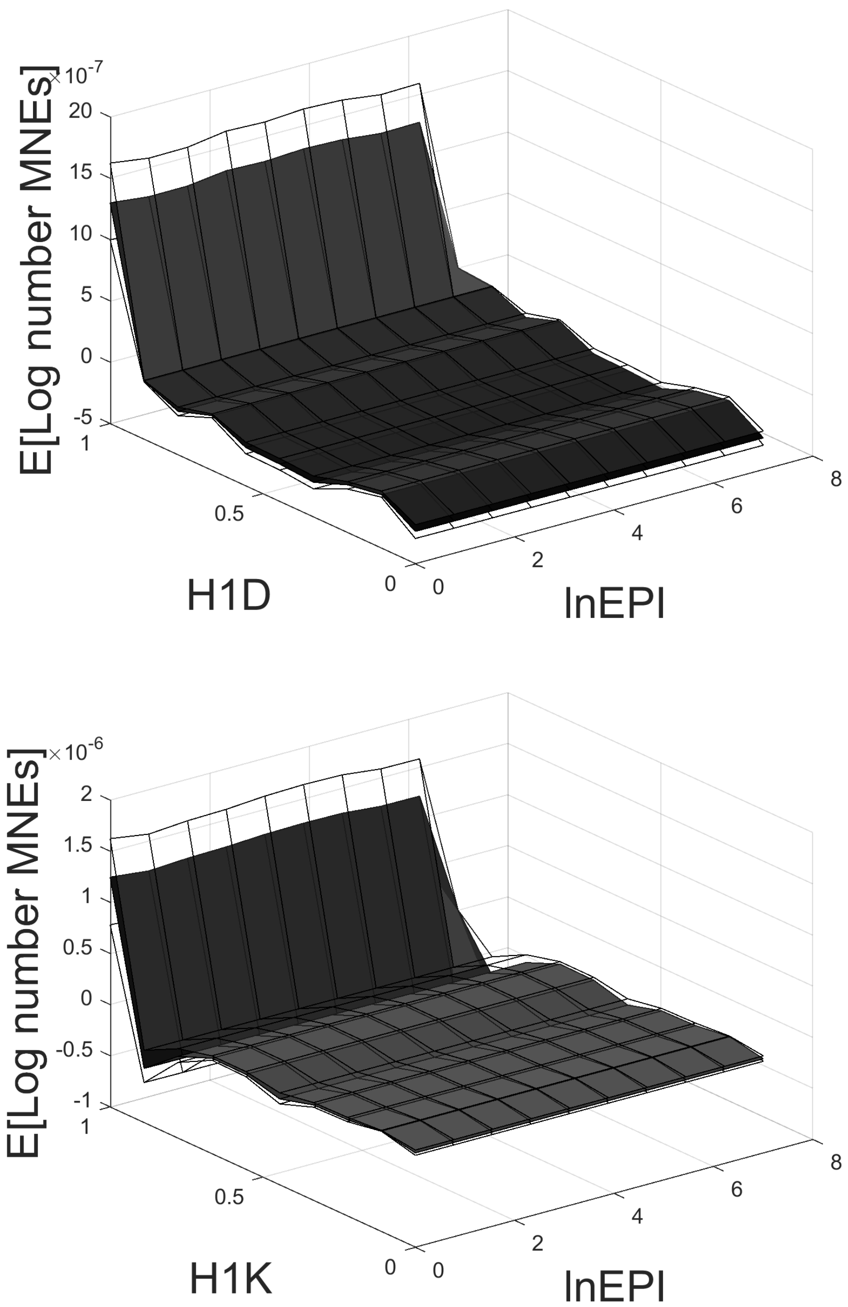
Figure B4 illustrates the estimates of the function after integrating out the differences between observational units in the average dose–response function. It plots the coefficient estimates using a 10×10 grid for all values (EPI, DS) in the support of the data, together with lower and upper bounds of the 95% confidence interval. The shape of this dose-response surface suggests that allowing for a flexible identification strategy when analyzing the joint impact of these skills is informative. The treatment functions are nonlinear in the sense that they are positively sloped over only parts of the support in DS-space and strongly increasing towards very high skill levels. This holds for all of the support in EPI-space. These findings indicate that high shares of digital skills in the population boost multinational production, whereas the effect of English proficiency is more or less linear.

Table A5: Parameter estimates of the unit-level dose–response function

Dep. var.: Log number of MNEs	H1D (1)	H1K (2)
EPI_{it}	-0.072	0.326
	(0.025)***	(0.013)***
DS_{it}	-0.154	6.245
	(0.462)	(0.867)***
$\hat{G}_{it}$	-0.005	-0.00009
	(0.0002)***	(0.00006)
$EPI_{it} \times \hat{G}_{it}$	0.001	0.00002
	(0.00004)***	(8.82e-06)**
$DS_{it} \times \hat{G}_{it}$	-0.001	0.00005
	(0.00003)***	(0.00002)**
$EPI_{it} \times DS_{it}$	0.019	-0.996
	(0.073)	(0.137)***

Notes: The dependent variable is the log number of MNEs. Explanatory variables are $\ln EPI_{it}$, the share of individuals with digital skills, the estimated GPS, and all variables included in its estimation. Standard errors are bootstrapped based on 100 independent draws. $^{+}(p < 0.1)$, $^{*}(p < 0.05)$, $^{**}(p < 0.01)$, $^{***}(p < 0.001)$. The digital skill variable refers to H1D (Individuals who have used software for electronic presentations) in the left panel and H1K (individuals who have written computer code) in right panel. Sources: Number of MNEs, Dun&Bradstreet WorldBase; English Proficiency Index (ln EPI), EF (Education First); and OECD ICT Access and Usage by Households and Individuals Database.

Figure B4: Average dose–response function of ln EPI and ICT skills with 95% confidence intervals for the number of MNEs



Notes: The 95% confidence intervals are bootstrapped based on 100 independent draws. The digital skill variable refers to H1D (Individuals who have used software for electronic presentations) in the upper panel and H1K (individuals who have written computer code) in the lower panel. Sources: Number of MNEs, Dun&Bradstreet WorldBase; English Proficiency Index (ln EPI), EF (Education First); and ICT skills, OECD ICT Access and Usage by Households and Individuals Database.