Appendix For Online Publication
Talent, Geography, and Offshore R&D

Jingting Fan
Pennsylvania State University

Contents

A Data and Facts **1**
A.1 Data Extraction, Cleaning, and Coverage . . . 1
A.2 Validating Patent Match Quality and Patent as a Measure of Offshore R&D . . . 5
A.3 Sample Restriction and Descriptive Statistics . . . 12
A.4 Regression Evidence for Facts 1 to 3 . . . 13

B Theory **17**
B.1 An Isomorphic Model with Vertical R&D at Affiliates . . . 17
B.2 Aggregation . . . 18
B.3 Definition of Equilibrium . . . 20

C Quantification **22**
C.1 Additional Data for Quantification . . . 22
C.2 Parameterization . . . 22
C.3 The Connection with Bilir and Morales (2020) and a Case without 'Bridge' R&D . . . 26

A Data and Facts

A.1 Data Extraction, Cleaning, and Coverage

The main source of the data for financial variables, ownership, and other firm characteristics is the Orbis Historic Disk Product (Bureau van Dijk, 2017), which links multiple vintages of Orbis products through firm identifiers and avoids some issues frequently arising in combining data vintages, as explained in Kalemli-Ozcan et al. (2019). The following describes my procedures for preparing the data, which largely follow those in Cravino and Levchenko (2017).

Financial data. I extract the financial data over 1996-2016 from the Historic Disk. The data are at firm identifier-year level. Each firm identifier represents a unique legal identity, possibly owned by another firm, an individual, or a family. I use sales (turnover) as the main measure of production and use value added for robustness. The initial extraction contains all firms with non-missing sales information for at least one year over the entire period. In a given year, firms might have multiple values of reported sales from different sources (local registry, annual report, or others), for consolidated or unconsolidated accounts. When different sources co-exist, I take local registry over other sources as it is likely more accurate. Consolidated account might include sales of other firms in the same conglomerate. For all analysis involving intensive margin measures of activities (i.e., sales and value added), I use only values reported in unconsolidated accounts and drop firms whose reporting is done solely in consolidated accounts; for analysis focusing on the extensive margin (i.e., whether an MNC operates in a host), I keep the latter group of firms. The extracted data include 58 million unique firm identifiers and 196 million identifier-year observations, among which 168 millions have non-missing sales data from unconsolidated accounts.

Due to expanding coverage, the representation of the sample varies over time. Table A.1 reports aggregate statistics of the raw data for 2013.[1] Column 1 reports the ratio between the total sales of firms and the GDP of the country, which is well above 0.8 for most countries.[2] For empirical analysis, relatively low coverage in some countries does not pose a threat because systematic variations in sample representation will be absorbed by fixed effects; in quantification, however, I will need to calculate the overall foreign shares in R&D and production, in which case low coverage could lead to biases. I explain how foreign shares are calculated for these countries in Section C.1 of this appendix.

For a subset of countries, Eurostat provides total sales for 'total business economy; repairs of computers, personal and household goods; except for financial and insurance activities,' and for manufacturing. I aggregate total sales of firms in these industries. Column 2 shows that the sample representation is reasonably good in these countries. Column 3 shows similar levels of representation for manufacturing. In both columns, the ratio is above 1 for some countries. One reason for this is that only the total sales of firms are reported, so I treat all sales of a firm as from its reported core industry. To the extent that some manufacturing firms also generate revenue from finance, I will not be able to exclude such revenue in this calculation.

Columns 4 and 5 reproduce Columns 2 and 3, restricting to firms after the match with the patent data. The matching process will be explained below, but in short, a firm is in the patenting data if one of its affiliates, its parent, or its sibling affiliates within the same MNCs has filed a patent in any country. Given that most firms do not own patents, the post-match sample is much smaller, but as Columns 4 and 5 show, it still accounts for a substantial share of the economy.

Ownership data. I extract a snapshot of shareholder information from the Historic Disk.[3] For each firm ID, I identify its global ultimate owner (GUO), the entity holding the majority control over the firm ID. This definition requires the owner to either directly hold more than 50% of the shares of the affiliate or—if the control is through other firms—hold more than 50% shares for every intermediate step along the ownership chain. For firms not linked to a GUO, I assume that their GUO are themselves, which practically means they are all domestic firms. To the extent that firms are more likely to have unreported links to foreign firms than to domestic firms, my treatment underestimates the importance of MNCs.

Time-invariant firm characteristics. I define the home country and industry of an MNC to be the country and industry of the GUO, respectively. When the GUO is an individual or a family, in which case industry classification and country information are unavailable, I use instead the industry and location of the largest affiliate (by sales) within the MNC. Note that because Table A.1 reports the aggregate statistics by affiliate country, it is not impacted by this choice.

[1] Except for Canada, which have a large number of missing value in 2013. I calculate statistics for Canada based on the 2014 data. In empirical analysis I average the data over each five-year interval, missing values in one year does not matter materially.

[2] Aggregate sales could be higher than GDP because they count value added multiple times. The U.S. has a low coverage because American firms report primarily their consolidated accounts and are thus excluded from this calculation.

[3] I measure the ownership information in 2016 and assume that it does not change throughout the sample period. For regressions exploiting over-time variations, this measurement error, if any, likely attenuates results. This choice is motivated by the limited coverage of earlier vintage of ownership data. However, all the empirical facts are robust if time-varying ownership information is used. Details are available upon request.

Table A.1: Coverage of the Firm-Level Data

	Full sample			Sample with patents	
	(1)	(2)	(3)	(4)	(5)
ISO	$\frac{\text{total sales}}{\text{GDP}}$	$\frac{\text{total sales exc. finance}}{\text{Eurostat total}}$	$\frac{\text{mfg. sales}}{\text{Eurostat mfg.}}$	$\frac{\text{total sales exc. finance}}{\text{Eurostat total}}$	$\frac{\text{mfg. sales}}{\text{Eurostat mfg.}}$
AUS	2.97	-	-	-	-
AUT	4.67	0.80	0.73	0.45	0.55
BEL	2.50	0.72	0.68	0.38	0.48
BGR	1.51	1.07	0.81	0.23	0.23
BRA	0.23	-	-	-	-
CAN	1.49	-	-	-	-
CHE	2.93	0.39	0.44	0.03	0.08
CHN	0.99	-	-	-	-
CZE	2.19	0.96	0.90	0.42	0.59
DEU	1.95	0.71	0.63	0.36	0.45
DNK	2.00	0.61	0.59	0.31	0.49
ESP	1.53	0.92	1.03	0.43	0.64
EST	2.27	0.96	0.88	0.20	0.33
FIN	2.85	-	-	-	-
FRA	1.88	0.81	0.80	0.42	0.56
GBR	0.83	0.26	0.31	0.11	0.21
GRC	0.96	0.64	0.78	0.08	0.08
HRV	1.13	0.93	1.13	0.14	0.20
HUN	2.20	1.10	0.99	0.40	0.51
IRL	4.33	0.83	0.56	0.61	0.34
ITA	2.02	0.87	0.97	0.31	0.46
JPN	1.91	-	-	-	-
KOR	2.10	-	-	-	-
LTU	1.34	0.68	0.48	0.08	0.14
LVA	1.82	0.99	1.04	0.10	0.18
MEX	0.31	-	-	-	-
NLD	0.92	0.26	0.15	0.11	0.12
NOR	2.68	0.99	0.82	0.40	0.47
POL	1.04	0.74	0.93	0.30	0.55
PRT	1.60	0.92	0.79	0.30	0.27
ROU	0.90	0.98	0.95	0.32	0.44
RUS	1.62	-	-	-	-
SVK	2.59	1.17	0.95	0.46	0.63
SVN	2.00	0.82	0.77	0.25	0.36
SWE	3.05	0.89	0.76	0.42	0.59
TUR	0.32	-	0.41	-	0.13
USA	0.02	-	-	-	-
Average	1.83	0.81	0.75	0.29	0.37

Notes: This table reports aggregate statistics constructed from the firm-level data, divided by the corresponding official statistics. Columns 1 through 3 are for the full sample; Columns 4 and 5 are for firms in the patenting sample. 'Eurostat Total' refers to sales reported by Eurostat in 'total business economy; repairs of computers, personal and household goods; except financial and insurance activities.' This definition includes NACE sectors 05-63, 68-82, and 95. I calculate the sample counterpart of this statistics in Columns 2 and 4 by aggregating over firms whose core industry is in these sectors. 'Eurostat mfg.' refers to total manufacturing sales from Eurostat. I calculate the sample counterpart of this statistics in Columns 3 and 5 using only manufacturing firms. In Columns 2 and 3, a couple of countries have ratios above 1. This is likely due to my treatment of the sales of multi-sector firms: only the total sales of a firm are reported, so I assume all sales is from the core industry.

Patent data and match to firms. I use patent-level data from PATSTAT Global (European Patent Office, 2018) to construct a measure of R&D. The database contains bibliographical data related to more than 100 million patent documents from 90 patent issuing authorities, including all major national, regional (e.g., the EPO), and global (e.g., the Patent Cooperative Treaty) patent offices.

I match individual patents from this database to their assignees (their owners) using a crosswalk from the Orbis Intellectual Property Database. This crosswalk links patent applications to firms using a string matching algorithm based on the (historic and current) name, address, and industry classification of the firm. A key step in the string matching underlying this crosswalk is the standardization of firm names—due to typo or abbreviations, firms often appear in the database under different names, which need to be harmonized before any match can be performed.[4] In the setting of global patents, this standardization is more challenging because it needs to take into account the variations in spelling patterns and common abbreviations across countries.[5] The Orbis Intellectual Property Database builds on harmonization efforts of several teams of researchers, respectively, the EPO, ECOOM at KU Lueven, and the OECD (HAN database). In Section A.2 of this appendix, I perform a few validation exercises on the quality of the match.

Having matched patents to firms, I further prepare the data as below:

1. **De-duplication.** Firms can, and often do, apply for multiple patents from different patent authorities for protection of the same underlying invention. As all such patents need to establish a common priority, i.e., the first applied patent on the invention, they can be identified as belonging to the same patent family in the PATSTAT database. I keep a family as long as one of its many patents is linked to a firm ID and, within each family, keep only one patent—the one with the most complete inventor location information. This de-duplication process reduces the number of unique patents to around 17 million, about two-thirds of the original number.

2. **Excluding design patents.** I exclude design patents and patents with unidentified types. Together, the excluded patents account for about 2% of the sample. The resulting sample contains patents from 90 patent offices, with the top 10 biggest patent offices accounting for 90% of the observations. The USPTO patents account for about 20% of this sample.

3. **Excluding patents without inventor location information.** PATSTAT does not receive inventor location information from the Japanese Patent office (JPO). Patents from other offices sometimes also have missing inventor locations. I exclude patents from the JPO or otherwise have missing inventor location information.[6]

   Table A.2 summarizes the contribution of each patent office to the sample and the fraction of these patents with non-missing inventor location information. Columns 1 and 2 are for all patents that can be matched to a firm in the Orbis data (after steps 1 and 2 described above), dating back to the early 20th century. U.S. and China are two biggest patent offices in this period, followed by Germany, Korea, and the EPO. Column 2 reports the fraction of observations from each patent office with non-missing inventor location. For six out of ten top patent authorities, inventor location is available for more than 70% of patents.

   Columns 3 and 4 reproduce Columns 1 and 2 for the period of my empirical analysis, 1996-2016. With increasing patenting in China, the top 10 offices now account for 94% of the sample. Aside from China, Australia, and Canada, all other major patent offices have close to universal availability of inventor location information. The increase in the availability of location information for

[4]For example, '3M Company' could appear in an application as any of the following: 'Minnesota Mining and Manufacturing Company', 'Minesota Mining and Manufacturing' (with misspelled 'Minnesota'), 'Minnesota Mining and Mfg.', 'Minn. Mining and Mfg.', or 'MMM Co.'

[5]For example, in Portugal, part of Spain, Poland, and most French speaking countries, the legal form that best resembles limited-liability corporations is 'Société anonyme,' abbreviated as S.A. or SA. Standardization should take into account such country-specific abbreviations of legal forms. For example, the string 'SA' in SAS Institute, an American company, should be treated as part of the name but not an abbreviation for legal form; on the other hand, 'SA' in the Spanish bank CaixaBank SA should be treated as legal form.

[6]Note that such exclusion does not necessarily mean that the R&D underlying the patent is excluded from my database. As long as *one* patent within a family has inventor location information, it will be preserved. This is an advantage of using the full PATSTAT Global data—I am able to piece together information about an invention from its multiple patents. For example, if a Japanese firm filed a patent in the U.S. and Japan at the same time, yet only the U.S. application reports inventor locations and only the Japanese application is linked to a firm ID, this patent will still be in my sample.

Table A.2: Sample Size and Availability of Inventor Location by Patenting Authority

	All historic patents		Patents filed in 1996-2016	
Patent office	% of obs. (1)	% with location (2)	% of obs (3)	% with location (4)
USA	22.35	73.37	18.00	100.00
CHN	21.11	21.33	30.84	21.22
GER	8.69	61.37	4.62	99.94
KOR	7.91	93.83	10.01	93.71
EPO	7.20	99.72	9.85	99.75
CAN	6.00	32.48	1.40	99.67
PCT	5.74	94.62	8.70	94.99
AUS	4.51	1.17	3.54	0.88
AUT	3.63	74.14	2.52	96.32
TWN	2.77	99.99	4.09	99.99
Others	10.08	57.26	6.43	75.74
Total	100.00	59.51	100.00	69.44

Notes: Columns 1 and 2 report information on all sample patents after Step 1 and 2 of the cleaning process and excluding JPO. Column 1 tabulates the fraction of patents in this sample from different patent offices. Column 2 reports, among all patents from an office, the fraction with inventor location available. Columns 3 and 4 reproduce Columns 1 and 2 for patents filed first between 1996 and 2016, the period of my empirical analysis.

most countries from Column 2 to Column 4 is likely due to changing reporting requirements at the patent application stage. That the missing information is concentrated in a small number of countries also reassures that these missing values are due to country-specific requirements, rather than MNCs' self selection into reporting.

4. **Aggregating by firm ID-inventor country-year.** I define the invention time of a patent as the earliest filing year among all patents within the patent family. I then sum across all patents assigned to a firm identifier in a given year to arrive at patent counts by firm identifier-inventor country-year. Note that because MNCs can assign a patent to any of its affiliates regardless of where the invention is performed and which patent office is involved, the result of this aggregation is not necessarily accurate for locations of R&D at the affiliate (firm identifier) level.[7] But after the final step below, it will be accurate for location of R&D at the parent level.

5. **Aggregating to parent firm-inventor country-year level.** I aggregate the R&D output from the previous step to parent firm-inventor country- year. I interpret inventor countries as the location of R&D. For example, if an American firm has 30 patents with inventors located in Japan, I interpret this as output of the American R&D center in Japan. I wish to emphasize that this assignment has nothing to do with whether the patents are from the USPTO or JPO, or whether the assignee on the patent is a U.S. affiliate in Japan or the headquarters in the U.S. The inferred location of R&D depends solely on the reported addresses of the inventors.

A.2 Validating Patent Match Quality and Patent as a Measure of Offshore R&D

To validate the match quality of the patent-firm crosswalk, I perform two validation exercises.

Manual inspection of the match quality. In the first exercise, I manually inspect the owner of 100 patents selected randomly from the merged patent-firm data (at the end of Step 2 described above). Figure A.1 shows the result from this check. These 100 patents have a total of 133 owners. 4 out of these 133 owners, colored gray, are not matched to a firm ID, which is likely due to the incomplete coverage of

[7] For example, Apple can apply for a patent invented entirely in California through its affiliate in China. It would be wrong to infer from this assignment that Apple China performs R&D in California.

the firm-level data. 2 firms, colored orange, are matched to a firm ID that cannot be matched to names.[8]

Among the remaining 127 owners, 126 are matched to correct firms, colored in either green or blue. In the figure, green means the names from the Orbis database are sufficiently close to the names from PATSTAT that it is clear that they are the same entity. Blue means that the names of the firms differ between the two databases, but I was able to verify that the difference are entirely due to simple name changes/re-organization/M&A. For example, the owner of the 6th patent in the figure changed its name from DB Cargo to Railion Deutschland in 2003 and back to DB Cargo AG to reflect its shifting business focuses. Such instances showcase the advantage of the Orbis database—with a database of historic firm names and firm ownership network, it is able to identify cases where simple string matching would miss.

In only 1 out of the 133 cases, colored red, the match is wrong. A patent by Huawei (the telecomm equipment manufacturer) is assigned to Dinglong Culture. I examined the source of this mistake. It turns out that Dinglong Culture, specialized in mining, was once named 'Huawei ltd.' At the time of patent filing, it was located in Huawei Industrial Park (named after the telecomm manufacturer).[9] This mistake can only be avoided with manual inspection of every patent.

Correlation with R&D expenditures. In the second exercise, I compare patent-based measure of invention to firm-level R&D expenditures in a small set of mostly listed global firms for which R&D expenditures are available from 2013. To allow for lags between expenditure and patenting, I use the average patent count between 2013 and 2016.

Table A.3 below reports the results. The first two columns use raw patent counts weighting patents equally. Columns 3 and 4 weight patents by their influence, measured as the number of citations *to* them (forward citation). All specifications include fixed effects for country and firms' main industry (3-digit). The results of the regressions suggest that patent invention is strongly correlated with R&D expenditures and that the correlation is not driven by firm size. Despite the challenges discussed above in merging global patents to global firms, the elasticities in our sample is in the same ballpark as estimates based on the match between USPTO data and Compustat firms (see Griliches, 1990 and the references thereto). This offers further assurance on the quality of the match.

Table A.3: Correlation between Patent Invention and R&D Expenditures

	(1)	(2)	(3)	(4)
	raw count		influence wgted	
log (R&D expenditures)	0.775^{***}	0.516^{***}	0.904^{***}	0.785^{***}
	(0.027)	(0.042)	(0.034)	(0.073)
log (employment)		0.382^{***}		0.184^{**}
		(0.053)		(0.081)
Observations	5551	4010	5547	4008
Country and industry FE	Yes	Yes	Yes	Yes
Psudo R^2	0.750	0.773	0.775	0.770

Notes: Specifications in this table regresses the logarithm of patent-based measures of firm invention on the logarithm of R&D expenditures and of employment. The R&D and employment data is from 2013. Patenting measure is averaged over 2013 to 2016. All specifications are in PPML and control for country and industry fixed effects. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[8]This is due to the change in firm ID between different vintages of the Orbis database. I verified manually that in the new vintage of the Orbis crosswalk, these two patents are linked to the correct owner.

[9]The company changes its name to Dinglong after it went public through the acquisition of a shell company named 'Dinglong'. IPO is heavily regulated in China and going public through acquisition to circumvent red tape is not uncommon.

id	application ID	year	person name	person name-matched	bvdidnumber	core sector
1	AT19720004352	1971	SIEMENS	SIEMENS AG	DE2010000581	2811
2	AT19750005745	1974	ITT INDUSTRIES	ITT INC.	US811197930	2813
3	AT19850305340T	1984	LABSYSTEMS	THERMO FISHER SCIENTIFIC OY	FI09215470	3250
4	AT19920000544	1992	CHEMIE LINZ		AT9070058281	
5	AT19960901173T	1995	ENTRE-MATIC	ENTREMATIC SWEDEN AB	SE*B000085591	
6	AT20040253492T	2003	RAILION DEUTSCHLAND	DB CARGO AKTIENGESELLSCHAFT	DE6190180117	4900
7	AT20090001457	2009	KOENIG MASCHINEN	KOENIG MASCHINEN GESELLSCHAF	AT9030027350	2893
8	AU20140294864	2013	TECHNIP FRANCE	TECHNIP ENERGIES FRANCE	FR391637865	7112
9	CA19770281917	1976	BAYER	BAYER AG	DE5330000056	2120
			GEA GROUP	METALLGESELLSCHAFT AG	DE3130292213	7010
10	CA19870541008	1986	UNION CARBIDE CORPORATION	UNION CARBIDE CORP	US131421730	2014
11	CA19952191440	1994	CSL	CSL LIMITED	AU051588348	2120
			UNIVERSITY OF NEW SOUTH WALES	UNIVERSITY OF NEW SOUTH WALE:	AU57195873179	5510
12	CA19962231199	1995	TROPIX	TROPIX INC	US135848339L	
13	CH19680012422	1968	SIEMENS	SIEMENS AG	DE2010000581	2811
14	CH19740016038	1973	SECRETARY OF STATE FOR DEFENCE IN HER BRITANNI(	QINETIQ LIMITED	GB03796233	7120
15	CHD353337	1953	BRITISH FILTERS	BRITISH FILTERS LIMITED	GB03839960	3299
16	CN2000121153	2000	CHINA PETROLEUM & CHEMICAL CORPORATION	CHINA PETROCHEMICAL GROUP TA	CN9360595881	2711
17	CN2002154077	2001	SAMSUNG ELECTRONICS COMPANY	SAMSUNG ELECTRONICS CO.,LTD.	KR1301110006246	2611
18	CN2007121657	2007	NANJING UNIVERSITY	NANJING UNIVERSITY	CN9363059219	8541
19	CN2007804290	2006	NEC CORPORATION	NEC CORPORATION	JP7010401022916	
20	CN2007822923	2006	KENNAMETAL	KENNAMETAL INC	US250900168	2849
21	CN2008182213	2007	FORD GLOBAL TECHNOLOGIES	FORD GLOBAL TECHNOLOGIES INC	US125737620L	2910
22	CN20091222791	2008	BROADCOM CORPORATION	BROADCOM CORP	US330480482	2611
23	CN20092112413U	2009	ZHEJIANG WANMA TIANYI TELECOM CABLE COMPAN\	ZHEJIANG WANMA CO., LTD.	CN31995PC	2733
24	CN20092297251U	2009	HENAN SHAOLIN AUTO COMPANY	HENAN SHAOLIN AUTO CO., LTD.	CN9360144275	3099
25	CN20101294175	2009	SILTRONIC	SILTRONIC AG	DE8170662128	2611
26	CN20111317246	2011	JIANGSU SANMU GROUP COMPANY	JIANGSU SANMU GROUP CO., LTD.	CN9360210541	2014
27	CN20121176930	2011	FUJIFILM	FUJIFILM CORPORATION	JP2010401064789	
28	CN2013124626	2013	JIMEI UNIVERSITY	JIMEI UNIVERSITY	CN9361069482	8541
			XIAMEN USUN YACHTS TECHNOLOGY COMPANY			
29	CN20132304550U	2013	ANHUI UNIVERSITY OF SCIENCE & TECHNOLOGY	ANHUI UNIVERSITY OF SCIENCE & 1	CN*HX000681393	
30	CN20132335132U	2013	YANTAI LONGYUAN POWER TECHNOLOGY COMPANY	YANTAI LONGYUAN POWER TECHN	CN40109PC	2899
31	CN20141602140	2014	HENAN PINGGAO ELECTRIC COMPANY	HENAN PINGGAO ELECTRIC COMP/	CN30866PC	2712
			PINGGAO GROUP COMPANY	PINGGAO GROUP CO.,LTD	CN9360360544	2790
			SGCC(STATE GRID CORPORATION OF CHINA)	STATE GRID CORPORATION OF CHI	CN9366719775	6619
32	CN20151161646	2015	CHINA UNIVERSITY OF PETROLEUM, BEIJING	CHINA PETROLEUM UNIVERSITY (B	CN9360836439	8541
			CNOOC (CHINA NATIONAL OFFSHORE OIL CORPORAT	CHINA NATIONAL OFFSHORE OIL C	CN9364410981	610
33	CN20162716110U	2016	CRRC MEISHAN COMPANY	CRRC MEISHAN CO., LTD.	CN9360325376	3020
34	CN201721369251U	2017	CHINA ACADEMY OF BUILDING RESEARCH	CHINESE ACADEMY OF BUILDING R	CN9366719001	
			CHINA ROAD & BRIDGE CORPORATION	CHINA ROAD & BRIDGE CORPORAT	CN9360658248	4299
35	CN20172487408U	2017	BAOTOU RESEARCH INSTITUTE OF RARE EARTHS	BAOTOU RARE EARTH RESEARCH IN	CN9360666021	7219
			BAOTOU REWIN RARE EARTH METAL MATERIALS COMPANY			
			BEIJING SCIAMPLE TECHNOLOGY COMPANY			
36	DE19742451940	1973	ELECTRIC POWER DEVELOPMENT CO., LTD., TOKIO/TO	ELECTRIC POWER DEVELOPMENT C	JP6010001050764	
			HITACHI	HITACHI LTD	JP7010001008844	
			TOSHIBA CORPORATION	TOSHIBA CORPORATION	JP2010401044997	
37	DE19883885217T	1987	ASAHI MALLEABLE IRON CO., LTD., SHIZUOKA	ASAHI TEC CORPORATION	JP4080401014785	
			OSAKA YUKI KAGAKU KOGYO K.K., OSAKA	OSAKA KAGAKU KOGYO KK	JP3120101025063	
38	DE1994628974T	1993	DIGISONIX	DIGISONIX LLC	US138443247L	
39	DE1995629199T	1994	TERUMO CARDIOVASCULAR SYSTEMS CORP., SOMERS	TERUMO CARDIOVASCULAR SYSTEI	US126296663L	3250
			THE CLEVELAND CLINIC FOUNDATION, CLEVELAND	CLEVELAND CLINIC FOUNDATION	US128551454L	8610
40	DE1998144443	1997	M. WATANABE & COMPANY	M.WATANABE & CO.,LTD.	JP3010001060923	
			MITSUBISHI ELECTRIC CORPORATION	MITSUBISHI ELECTRIC CORPORATIC	JP4010001008772	
			ORGANO	ORGANO CORPORATION	JP3010601025111	
41	DE2003615899T	2003	NORTEL NETWORKS	NORTEL NETWORKS LIMITED	CA32094NC	6190
42	DE20071005540	2006	FRIEDR. ISCHEBECK	FRIEDR ISCHEBECK GMBH	DE4310000217	2511
			MINOVA INTERNATIONAL	MINOVA INT LTD	GB04373275	7010
43	DK19990121089T	1995	KUHN	KUHN SAS	FR675580542	2830
44	DK20040292916T	2003	NORINCO	EJ EMEA	FR592029425	7010
45	EP19970950288	1996	DOMINO PRINTING SCIENCES	DOMINO PRINTING SCIENCES PUBI	GB01363137	2899
46	EP20040726049	2003	TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)	TELEFONAKTIEBOLAGET LM ERICSS	SE5560160680	2630
47	EP20090732515	2008	HONDA MOTOR COMPANY	HONDA MOTOR CO., LTD.	JP6010401027577	
			MAKINO J	MAKINO J CO.,LTD.	JP2021001020859	
48	EP20110858917	2011	HUAWEI TECHNOLOGIES COMPANY	DINGLONG CULTURE CO.,LTD.	CN40162PC	4690
49	FI19830002268	1982	BETHLEHEM STEEL COMPANY	BETHLEHEM STEEL CORP	US131219783L	
50	FR20040009725	2004	STMICROELECTRONICS ROUSSET SAS	STMICROELECTRONICS ROUSSET S/	FR414969584	2611
			UNIVERSITE DE PROVENCE AIX MARSEILLE 1	UNIVERSITE AIX -MARSEILLE 1 PRO	FR191318427	
51	GB19590041337	1959	BURNDEPT	AA EIGHT LIMITED	GB00928035	
52	GB19880029688	1988	NISSHINBO INDUSTRIES	NISSHINBO HOLDINGS INC	JP6010001034882	
			SUZUKI JIDOSHA KOGYO	SUZUKI JIDOSHA KOGYO KK	JP9030001061765	
			SUZUKI MOTOR CORPORATION	SUZUKI MOTOR CORPORATION	JP8080401002431	

53	GR20010400207T	1995	MEYHALL	DANISCO SWITZERLAND AG	CHCHE102615526	4675
54	IT19750050330	1974	UNION CARBIDE CORPORATION	UNION CARBIDE CORP	US131421730	2014
55	IT19760028287	1975	THE GOODYEAR TIRE & RUBBER COMPANY	GOODYEAR TIRE & RUBBER CO	US340253240	2211
56	KR19880071787	1987	DIGITAL EQUIPMENT CORPORATION	DIGITAL EQUIPMENT CORP	US042226590	2620
57	KR19950035944	1995	LG ELECTRONICS	LG ELECTRONICS INC.	KR1101112487050	2640
58	KR20000084850	2000	HYUNDAI MOTOR COMPANY	HYUNDAI MOTOR COMPANY	KR1101110085450	2910
59	KR20090112399	2009	INNOCEM KOREA. COMPANY	INNOCEM KOREA CO.,LTD	KR1101111697444	2740
			PLANET INT COMPANY	PLANET INT CO.,LTD	KR1101112621666	2611
60	KR20110096289	2010	USHIO	USHIO INC	JP1010001008651	
61	KR20140161879	2014	SAMHWA CAPACITOR COMPANY	SAMWHA CAPACITOR CO.,LTD.	KR1345110003696	2611
62	KR20167010392	2014	TOSHIBA CORPORATION	TOSHIBA CORPORATION	JP2010401044997	
			TOSHIBA LIFESTYLE PRODUCTS & SERVICES CORPORAT	TOSHIBA LIFESTYLE PRODUCTS & SI	JP4010001116880	
63	MA20070030009	2007	DRISCOLL STRAWBERRY ASSOCIATES	DRISCOLL STRAWBERRY ASSOCIATI	US123488847L	125
64	NL19740005891	1973	HERMANN HEYE, KOMMANDITGESELLSCHAFT TE OBEF	HERMANN HEYE KG	DE*D00303916	
65	PT20060829164T	2005	HECKLER & KOCH	HECKLER & KOCH GMBH	DE7230255706	2540
66	PT20120735174T	2011	GOJO INDUSTRIES	GOJO INDUSTRIES INC	US149107003L	2041
67	TW20110123682	2011	KEYSTONE ELECTRONICS CORPORATION	KEYSTONE ELECTRONICS CORPORA	TW12783703	
68	TW20130108537	2012	HTC CORPORATION	HTC CORPORATION	TW16003518	2620
69	TW20140105922	2013	RENESAS ELECTRONICS CORPORATION	RENESAS ELECTRONICS CORPORAT	JP8020001075701	
70	US19330687011	1933	DUPONT	E. I. DU PONT DE NEMOURS AND C	US510014090	2016
71	US19400371203	1940	GOODRICH CORPORATION	GOODRICH CORPORATION	US340252680	2651
72	US19430496931	1943	BROWN AND SHARPE MANUFACTURING COMPANY	BROWN & SHARPE MANUFACTURI	US120477224L	3299
73	US19620181327	1962	THE RIDGE TOOL COMPANY	RIDGE TOOL CO ELYRIA DOM	US163613088L	2790
74	US19620215539	1962	DOW CHEMICAL COMPANY	DOW CHEMICAL COMPANY (THE)	US381285128	2016
75	US19740470966	1974	TOWER PRODUCTS	TOWER PRODUCTS INC	US134231372L	2630
76	US19800163253	1979	BOEHRINGER MANNHEIM	BOEHRINGER MANNHEIM GMBH	DE90013ED	2120
77	US19860866603	1985	HONDA MOTOR COMPANY	HONDA MOTOR CO., LTD.	JP6010401027577	
78	US19960607704	1996	MICRON TECHNOLOGY	MICRON TECHNOLOGY INC	US751618004	2611
79	US19960609541	1992	IFF (INTERNATIONAL FLAVORS & FRAGRANCES)	INTERNATIONAL FLAVORS & FRAG	US131432060	2014
			UNIVERSITY OF FLORIDA		US163440498L	8541
80	US19970848726	1997	MOORE WALLACE	MOORE & WALLACE INC	US174582003L	4221
81	US19970990748	1997	AEC OIL SANDS	A.E.C. OIL SANDS, L.P.		
			AEC OIL SANDS LIMITED PARTNERSHIP		CA*H00089269070	
			ATHABASCA OIL SANDS INVESTMENTS	ATHABASCA OIL CORPORATION	CA35202NC	610
			CANADIAN OCCIDENTAL PETROLEUM	NEXEN INC	CA30003NC	610
			CANADIAN OIL SANDS INVESTMENTS	SUNCOR ENERGY VENTURES HOLDI	CA30776NC	910
			GULF CANADA RESOURCES	CONOCO CANADA RESOURCES LTD	CA90148NC	610
			IMPERIAL OIL RESOURCES			
			MOCAL ENERGY	MOCAL ENERGY LTD	CA*H00085772101	
			MURPHY OIL COMPANY	MURPHY OIL COMPANY LTD.	CA275568353L	610
			PETRO-CANADA	PETRO-CANADA	CA260826101L	
82	US20000614920	1996	PANASONIC CORPORATION	PANASONIC HOLDINGS CORPORAT	JP5120001158218	
83	US20010779730	2000	NEC ELECTRONICS CORPORATION	RENESAS ELECTRONICS CORPORAT	JP8020001075701	
84	US20010877147	1998	TOSHIBA CORPORATION	TOSHIBA CORPORATION	JP2010401044997	
85	US20020328329	2002	MEDTRONIC	MEDTRONIC INC	US126031475L	2660
86	US20030437083	2003	CHUNG-SHAN INSTITUTE OF SCIENCE AND TECHNOLO	NATIONAL CHUNG SHAN INSTITUT	TW45002910	8411
87	US20030439611	2000	UNIVERSITY OF CALIFORNIA	REGENTS OF UNIVERSITY OF CALIF(	US123847211L	8411
88	US20060640870	2005	HITACHI	HITACHI LTD	JP7010001008844	
89	US20080031610	2007	ETHICON ENDO-SURGERY	ETHICON ENDO-SURGERY INC	US149169268L	2120
90	US20080069225	2008	DICKE TOOL COMPANY	DICKE TOOL CO	US130390634L	2573
91	US20090472783	2008	SONY CORPORATION	SONY GROUP CORPORATION	JP5010401067252	
92	US200913139652	2008	LA SEDA DE BARCELONA	LA SEDA DE BARCELONA SA	ESA08010571	1310
93	US20100785884	2009	AZBIL CORPORATION	AZBIL CORPORATION	JP9010001096367	
94	US201113230128	2010	DNP (DAINIPPON PRINTING COMPANY)	DAI NIPPON PRINTING CO LTD	JP5011101012069	
95	US201213653934	2012	GOOGLE	GOOGLE LLC	US134090555L	6201
96	US201213658438	2012	LIEBHERR-WERK NENZING	LIEBHERR-WERK NENZING GMBH	AT9010000282	2899
97	US201213987306	2012	ARUZE GAMING AMERICA	ARUZE USA INC	US166550265L	3299
			UNIVERSAL ENTERTAINMENT CORPORATION	UNIVERSAL ENTERTAINMENT CORF	JP6010601025422	
98	US201514678981	2014	INNOCHIPS TECHNOLOGY	MODA-INNOCHIPS CO., LTD.	KR1355110106027	2611
99	US201815870539	2013	HUTCHINSON TECHNOLOGY	HUTCHINSON TECHNOLOGY INC	US410901840	2611
100	USD3764726	1971	SIEMENS	SIEMENS AG	DE2010000581	2811

Figure A.1: Match Quality Check on 100 Random Applications

Note: This figure illustrates the match quality for the 133 owners of 100 randomly selected patents. The column titled *id* is the index for these patents; *application ID* is these patents' ID in the PATSTAT database; *year* refers to the year in which the patent application was made; *person_name* refers to the name of the applicant from the PATSTAT database; *person name-matched* and *bvdidnumber* refers to the name and the firm ID from the Orbis database a patent is matched to; *core sector* is the NACIS code for the main sector of a firm. Among these patents, only one (the 48th patent) is matched to a wrong firm, indicated by red.

Patents as a measure for offshore R&D. As discussed in Section 2.1, using patent data to measure offshore R&D has three advantages: first, the universe of data are readily available at the firm level; second, it is less subject to different definitions of 'research and development' between countries; third, compared to affiliate R&D expenditures, the addresses of inventors are less likely to be manipulated by MNCs for tax avoidance. I now provide direct evidence for this last advantage.

Panel A of Table A.4 below reports the share of activities by U.S. multinationals in tax havens, defined according to Hines and Rice (1994). The first row is based on the public-use data from the BEA. It shows that these countries account for 5.3% of employment, 8.7% of employee compensation, 17.1% of R&D expenditures, and 48% of net income of all overseas affiliates of U.S. MNCs. In particular, the share of employment and employee compensation are vastly overshadowed by both R&D expenditure and net income, which is often interpreted as evidence of tax-avoidance activities.

The second row reports the share of patent invention by U.S. MNCs in these tax havens. When total invention is measured by inventor locations, depending on whether patents are weighted by influence (forward citation), tax haven accounts for about 6.7% or 8.25%, which fall between the share of employment and that of employee compensation in these countries. When measured by firm owners' location, however, the shares increase drastically, with the share of influence-weighted patents in these location more than doubled.[10] This is consistent with firms assigning intellectual property rights, especially high-value ones, to affiliates in tax heavens to facilitate tax avoidance.

Table A.4: Activities in Tax Haven Countries

Panel A. Tax haven countries' share of the activities of U.S. multinationals			
Employment	Employee compensation	R&D	Net income
5.26%	8.70%	17.09%	48.27
Patenting by inventor location		Patenting by firm location	
count	influence weighted	count	influence weighted
6.71%	8.25%	11.51%	18.06%
Panel B. Tax haven countries' share of world economy			
GDP	Population		
0.2%	0.08%		
Patenting by inventor location		Patenting by firm locations	
count	influence weighted	count	influence weighted
1.50%	2.04%	2.25%	3.20%

Note: Panel A reports the share of U.S. multinational activities taking place in tax haven countries, defined according to Hines and Rice (1994). Employment, employee compensation R&D, net income are from BEA public data. Patent based measures are authors' calculation. Panel B reports the share of tax haven countries in world GDP and patent invention (based on the location of the assignees versus inventors, respectively).

Panel B report the share of world GDP, population, and patents in these tax haven countries. These countries account for 0.2% of GDP and 0.08% of population. Their share in *world* patent invention measured using inventor locations is 1.5% (2.0% if patents are weighted). Importantly, switching from an inventor-location based measure to a firm-location based measure increases both the weighted and the unweighted share by 50%, consistent with the patterns documented in Panel A for U.S. firms

Table A.5 uses regressions to control for other country characteristics. The first two columns are at the country level. The dependent variable is the difference between the logarithm of patent counts based on firm locations and the logarithm of patent counts based on inventor locations. The regressions show that, controlling for country size (population), TFP, and human capital index, firm location-based measures inflate patent inventions in tax haven countries by 83% log point (75% log point if patents are weighted

[10] All of the patents ultimately belong to a U.S. multinational firms through a firm ownership network. Here I merely count patent based on the location of the *direct* owners of patents reported in patent applications.

Table A.5: Invention in Tax Heaven Countries: Firm- versus Inventor-based Measures

	(1)	(2)	(3)	(4)
dependent var: firm-based log patent counts minus inventor-based				
	country-level		firm-level	
	unweighted.	weighted	unweighted	weighted
tax heaven indicator	0.835***	0.753***	0.302***	0.376***
	(0.208)	(0.252)	(0.042)	(0.054)
Observations	111	111	76956	76956
R^2	0.328	0.217	0.303	0.294
Within R^2			0.019	0.021
Country controls	yes	yes	yes	yes
Firm FE			yes	yes

Notes: Columns 1 and 2 are at the country level; the dependent variable is the logarithm of the number of patents granted to firms in a country minus the logarithm of the number of patents with inventors located in that country. Columns 3 and 4 are at the firm level, with each observation being a firm-host country combination; the dependent variable is the logarithm of the number of the patent granted to the affiliates of a firm in a country minus the logarithm of the number of the firm's patents with inventors being in that country. All specifications control for country population, TFP, and human capital index; Columns 3 and 4 further control for firm fixed effects. Data include all patents from 1996-2016. Standard errors are clustered by country. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

by influence). Columns 3 and 4 are at the firm level, with the dependent variable being the difference between the logarithm of patent counts based on firm location and the logarithm of patent counts based inventor locations. The regressions show that a similar pattern holds even within individual firms, whether patents are weighted by citation or not.

Tables A.4 and A.5 demonstrate that an inventor-location-based measure of R&D can help alleviate concerns about multinational firms strategically allocating R&D expenditures for tax avoidance.

Additional concerns on patent as a measure of R&D. A well-known drawback of using patent data to measure R&D is that firms self select into patenting, which can introduces biases in the measure. I now explain why this concern seems unlikely to bias either my reduced-form or quantitative results.

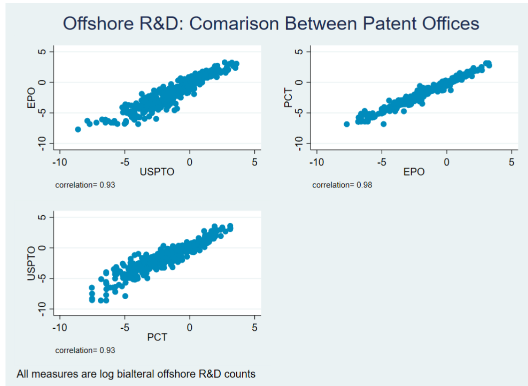
There are at least two types of selection, both of which are well recognized by the study of R&D using patent data in closed-economy settings (Griliches, 1998): some R&D efforts might not result in patentable outcomes; firms might choose not to patent a patentable R&D outcome. In the multinational setting considered in this paper, the threat is that such selection might be correlated with the characteristics of the firm or those of the host. For example, if more innovative firms are both more likely to engage in offshore R&D and have a higher propensity to patent their inventions, my measure of offshore R&D would be biased towards these firms.

An advantage of the multinational setting is that I will be able to flexibly control for firm and host characteristics that likely determine the decision to patent an invention in ways that are infeasible in closed-economy settings. Specifically, I control for firm-period fixed effects in all specifications, so any selection at the firm level is absorbed and the coefficients are only identified off within-firm variations. For Facts 1 and 2, where the selection concern is more relevant, I further control for affiliate fixed effects. This will purge out the influence of any time-invariant host-specific factors.[11]

In addition to these two types of selection, in my setting there is another type of selection: firms choose to patent inventions only in hosts in which they either have a manufacturing presence or the intention to launch a product. This selection makes the interpretation of Fact 2 potentially problematic: the observed colocation between production and patenting might be driven by such selection and therefore has nothing to do with the friction in separating the two. Because my measure of R&D is based *not* on which host country a patent is issued in, but on where the inventor of a patent is located, this concern does not in itself lead to a bias. To the extent that such selection can affect the measured locations of in-

[11] Admittedly, there could be time-varying factors. In Fact 1, I control for two usual suspects, the IPR protection index and R&D subsidies and show that they do not affect the main coefficient of interest. In Fact 2, since the variation exploited is at affiliate level, I am able to control for host-industry-period fixed effects, which absorb all time-varying characteristics of a country that might affect patenting in an industry.

Figure A.2: Comparison Among Patents From Individual Patent Offices



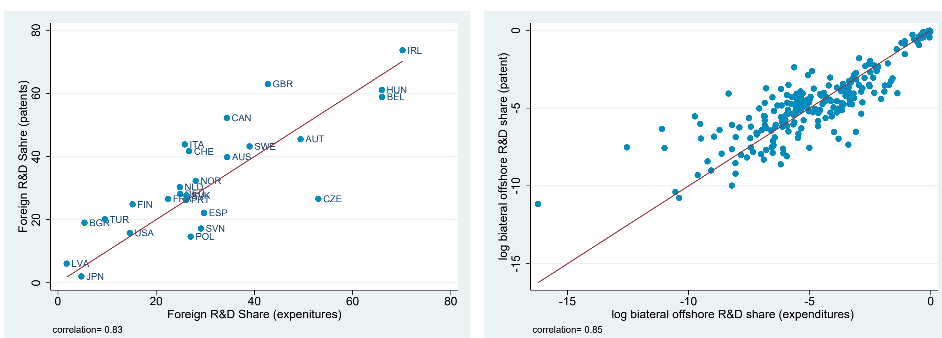
Notes: The figure shows log bilateral offshore R&D measured using three major patent offices (EPO, USPTO, PCT) is closely correlated.

ventors, I address this concern in two ways. First, through the rich control of fixed effects, which absorb all confounding factors that are specific to a host, an affiliate, or a firm. Second, I provide direct evidence that such selection does not affect the measured offshore R&D below.

Specifically, if the selection of patent offices affect the measured offshore R&D, then we should see that offshore R&D calculated based on data from different patent authorities differ significantly from one another. Figure A.2 shows the exact opposite: bilateral offshore R&D shares measured using data from the USPTO and two other major international patent offices, the PCT and the EPO, are highly correlated. I conclude that differential selection into patenting in specific hosts do not lead to significant biases in the measured offshore R&D.

Figure A.3: Aggregate and Bilateral Offshore R&D Measures: Patents v.s. Expenditures

(a) Aggregate Inward Offshore R&D Share (b) Log Bilateral Offshore R&D Share



Notes: The left panel is the fraction of R&D in a country carried out by foreign firms, measured using two different sources: business enterprise R&D expenditures (horizontal axis) and patents (vertical axis). The right panel plots the log of bilateral offshore R&D shares, in which each dot represents a country pair.

In quantitative analysis, I will use the share of inventions in a host by foreign firms as an additional input. For such aggregate shares, I will not be able to address the above concern through controls. Instead, I show directly that my measure is closely correlated with the one based on R&D expenditures in Figure A.3. In the left panel, the horizontal axis is calculated using business enterprise R&D expenditures from the OECD; the vertical axis is the shares calculated based on the patent data, described above. The figure shows a close mapping between the two measures, with a correlation of 0.83, despite that they are from two independent sources.[12] The right panel plots log *bilateral* offshore R&D shares measured

[12]The expenditures statistics are aggregated from firm-level surveys and other administrative data. It is possible that the differences between the two measures are mostly driven by sampling differences. In fact, the correlation between my data and

using R&D expenditures and patents. Again, the two measures are highly correlated. The takeaway from Figure A.3 is that even if one preferred to use R&D expenditures to construct foreign R&D shares, my measure is a good proxy.

A.3 Sample Restriction and Descriptive Statistics

I describe in this subsection the sample of my analysis and the descriptive statistics.

Sample period. For best coverage of the financial data, I focus on 1996-2016. To reduce measurement errors associated with patent counts (for example, firms might be continuously doing R&D but the patent application might be discrete), I aggregate the sample into four five-year periods. Within each period, I take the average values of patent and citation counts, and financial statistics.

Countries and their characteristics. I focus on a sample of 37 host countries (but include MNCs whose parent are from other countries in empirical analysis). This sample restriction is made to be consistent with the subsequent quantitative analysis. Specifically, for quantification I will use data on manufacturing output and trade from the 2016 release of the World Intput Output Database (Timmer et al., 2016). Among the 43 countries in this database, I exclude three countries with population below one million, Cypress, Luxembourg, Malta; I exclude Taiwan, as World Bank and Penn World Table does not report its economic statistics; finally, I exclude India and Indonesia due to their poor representation in the Orbis financial database. This results in the 37 countries reported in Table A.1. All empirical patterns remain virtually unchanged if I simply use all countries in the Orbis database.

I combine the firm-level data described previously with time-varying country characteristics. Concretely, I obtain GDP, GDP per capita, and the human capital index from the PWT 9.0. (Feenstra et al., 2016; see Feenstra et al., 2015 for descriptions); an updated version (Park, 2015) of the intellectual property right index created in Park (2008); R&D subsidies and the number of researchers from the OECD (OECD, 2018). All these variables are also averaged over each four-year period. I obtain bilateral distance measures from Mayer and Zignago (2011). Table A.6 summarizes these country characteristics for the last period, 2011-2016.

Table A.6: Host Characteristics: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
ln (GDP)	37	13.30	1.55	10.37	16.59
ln (GDP per capita)	37	13.30	1.55	10.37	16.59
Human capital idnex	37	3.24	0.38	2.29	3.72
ln (number of researchers)	32	11.38	1.48	8.61	15.07
R&D subsidies	36	0.13	0.12	-0.02	0.44
Intellectual property right index	33	4.31	0.33	3.59	4.88

Notes: This table reports characteristics of the 37 host countries in the sample, averaged over 2011-2016.

Structure of financial and R&D samples separately. In empirical analysis, I use the merged sample between financial and R&D datasets. Section 2.2 presents descriptive data for the matched sample. Tables A.7 gives an overview of these two datasets separately.

The left panel of Table A.7 summarizes the structure of the financial data. Columns 1 and 2 are the number of unique firms and unique production affiliates in each period, respectively. The numbers gradually increase as coverage of the database broadens, but Columns 1 and 2 track each other closely, reflecting that the overwhelming majority of firms have only one production affiliate.

The right panel of Table A.7 is the structure of the R&D data. This sample is larger than reported in Table 1 as it includes firms granted a patent but with no available financial information.[13] The R&D

these statistics for offshore production, measured using sales in both sources, is also around 0.8.

[13]The statistics reported is after excluding firms classified as education institutions and governments, or firms with unknown

Table A.7: Structures of Production and R&D Samples

	Financial data		R&D Data		
Period	# of unique firms	# of aff. with positive sales	# of unique firms	# of R&D centers (baseline)	# of R&D centers (Liberal)
1	3,615,341	3,643,392	118,953	112,217	140,344
2	7,992,947	8,050,202	138,792	133,432	164,173
3	13,814,682	13,906,270	161,114	158,297	191,442
4	49,264,872	49,389,390	136,323	136,694	162,139
Total	74,687,842	74,989,254	555,182	540,640	658,098
Unique firms	54,535,654		378,859		

Notes: This table summarizes separately the coverage of financial and R&D data over time (before the two are merged).

sample does not registered as dramatic an expansion as the financial sample. This is unsurprising, as PATSTAT covers close to the universe of world patents from the very beginning. In the fourth period, on average, a firm has 1.189 R&D centers according to the liberal definition, and only 1.003 according to the baseline definition.

A.4 Regression Evidence for Facts 1 to 3

In this subsection, I describe the regressions underlying Facts 1-3, and additional specifications that demonstrate the robustness of the results.

Fact 1. The first fact investigates how affiliate invention intensity varies with host talent quality. My specification is

$$y_{foh,t} = FE + \overrightarrow{\gamma_x} \cdot \overrightarrow{X}_{oh,t} + \epsilon_{ohf,t}.$$

The outcome variable is the log ratio of patent invention over sales, for an affiliate in host h of firm f from country o in period t. The explanatory variables $\overrightarrow{X}_{oh,t}$ include the measure of host talent, the human capital index from the Penn World Table, along with other controls. *FE* is fixed effects.

A concern in interpreting $y_{foht,t}$ as the invention intensity of an affiliate is that it might pick up that firms apply for more patents in more attractive markets. Note that in constructing $y_{foh,t}$, it is the number of patents *invented* in country h, which could be granted by any authorities, rather than the number of patents granted by the authority of country h, that is being counted. Selective patenting in more attractive hosts will thus not necessarily bias the measure. I include country fixed effects and time-varying country characteristics to soak up remaining variations in the propensity of patenting across hosts. Relatedly, some firms patent their inventions more frequently than others. This source of heterogeneity will be absorbed by firm-level fixed effects.

Table A.8 reports the results. The first column controls for firm-period fixed effects and four measures of bilateral distance between the home and the host: geographic distance, and indicators for whether countries o and h share an official language, are contiguous to each other, or have a colonial tie. This specification exploits within-firm, cross-host, variation, and finds significant positive correlation between the invention intensity of an affiliate and the human capital index. The size and income of the host, on the other hand, do not seems to be important.

Columns 2 adds affiliate fixed effects that absorb all invariant country characteristics. I further control for the protection of intellectual property rights (IPR, Park, 2008) and R&D subsidies (OECD) of the host, two policy measures likely correlated with R&D and patenting.[14] The specification shows that an improvement in host human capital over time is correlated with an increase in the patent sales ratio. In terms of magnitude, a one standard deviation increase in the human capital index ($\approx$ 0.38) more than doubles the outcome variable. Column 3 further includes a narrower measure of talent, the number of researchers in a country. An increase in this measure is positively correlated with the invention intensity

home countries—most likely individuals or families.

[14]Because changes in GDP and income are highly correlated, I include only one of them.

Table A.8: Human Capital and Affiliate Invention Intensity

	(1)	(2)	(3)	(4)
Dependent variable:	ln (patent/sales)			R&D Indicator
human capital index	0.95***	3.013**	3.431**	0.181**
	(0.257)	(1.365)	(1.334)	(0.076)
ln(GDP per capita)	-0.273	-0.622	-0.710*	0.081***
	(0.258)	(0.439)	(0.350)	(0.023)
IPR protection		0.563***	0.404**	0.020
		(0.205)	(0.176)	(0.015)
R&D subsidies		0.508	0.572	0.011
		(0.384)	(0.403)	(0.029)
ln (researchers)			0.421**	0.067***
			(0.172)	(0.016)
ln(GDP)	0.077			
	(0.097)			
tax haven indicator	0.069			
	(0.163)			
log (sales)				0.004***
				(0.001)
Observations	20893	11803	11464	80253
R^2	0.253	0.677	0.675	0.637
Within R^2	0.029	0.010	0.015	0.005
Distance measures	Y	-	-	-
Firm-period FE	Y	Y	Y	Y
Affiliate FE	-	Y	Y	Y

Note: The outcome variable is log of the ratio between patent counts and affiliate sales. The explanatory variables are country characteristics. Column 1 is a cross-sectional regression that controls for firm-period fixed effects and bilateral distance measures, including geographic distance and a set of dummies (see the text). Columns 2 through 4 control for time-invariant host characteristics through affiliate fixed effects. Standard errors (in parenthesis) are clustered by host country and by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results from Columns 3 and 4 are visualized in Figure 2 of the main text.

of affiliates, but the coefficient for the human capital index barely changes, suggesting that a broader interpretation of talent is warranted. Finally, Column 4 examines whether an improvement in human capital is associated with the entry of foreign R&D centers through the extensive margin, controlling for the sales of the affiliate. I find that both measures of talent are associated with entry.

Results from Columns 3 and 4 of this table are visualized in Figure 2 of the main text.

Fact 2. The second fact is on the colocation of invention and affiliate sales. The specification is:

$$y_{fh,t} = FE + \gamma_{R\&D}\, x_{fh,t} + \overrightarrow{\gamma}_{dist} \cdot \overrightarrow{dist}_{fh,t} + \epsilon_{fh,t},$$

where f, h, t, indicates firm, host, and period, respectively. Variables $y_{fh,t}$ and $x_{fh,t}$ measure affiliate sales and invention of firm f in host h. $\overrightarrow{dist}_{fh,t}$ is a vector consisting of four measures of the average distance between host h and all other countries in which firm f has an R&D center. $\overrightarrow{dist}_{fh,t}$ is firm specific because firms differ in their geographic presence. Coefficients $\gamma_{R\&D}$ and $\overrightarrow{\gamma}_{dist}$ capture the colocation patterns.[15]

The OLS estimate of $\gamma_{R\&D}$ and $\overrightarrow{\gamma}_{dist}$ might suffer from an omitted variable bias. For example, firms might conduct both invention and production in hosts with a comparative advantage in their industry; both activities might also grow in response to an expansion in the economy or market access of a host. To rule out these factors, I include host-period, home-host, and host-industry fixed effects. I further control for firm-period fixed effects. As discussed previously, these controls also help purge out systematic variations in patenting propensities across firms and hosts, so $x_{fh,t}$ can be interpreted as R&D.

[15]Since the quantitative analysis will be based on a static model, I focus on the contemporaneous effect in this specification. Including leads and lags of offshore R&D measures as additional explanatory variables result in positive and statistically significant coefficients for contemporary and lagged offshore R&D, and small and insignificant coefficient for future offshore R&D. Such dynamic results, available upon request, are also consistent with a colocation effect.

Table A.9: Co-location of Invention and Affiliate Sales

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent var.	aff. sales indicator	ln (sales)					
R&D Indicator$_{fh,t}$	0.281***	1.164***		1.042***			
	(0.003)	(0.024)		(0.026)			
ln(patent)$_{fh,t}$			0.331***		0.329***	0.205***	0.181***
			(0.012)		(0.012)	(0.044)	(0.042)
ln (distance)$_{fh,t}$				-0.024	-0.328**		
				(0.025)	(0.144)		
common language$_{fh,t}$				0.220***	0.408		
				(0.051)	(0.267)		
contiguity$_{fh,t}$				0.143***	0.224		
				(0.049)	(0.235)		
colonial tie$_{fh,t}$				0.090**	-0.563*		
				(0.046)	(0.306)		
Observations	7494979	119659	19519	119503	19519	14090	8839
R^2	0.704	0.495	0.572	0.496	0.572	0.969	0.963
Within R^2	0.042	0.045	0.093	0.047	0.094	0.022	0.020
Firm-period FE	Y	Y	Y	Y	Y	Y	Y
Host-period FE	Y	Y	Y	Y	Y	Y	-
Home-host FE	Y	Y	Y	Y	Y	-	-
Host-industry FE	Y	Y	Y	Y	Y	-	-
Affiliate FE	-	-	-	-	-	Y	Y
Host-industry-period FE	-	-	-	-	-	-	Y

Note: Column 1 estimates the relationship between having an R&D center in a host and the probability of having an affiliate with sales in the same host. Columns 2 through 7 estimate the relationship between having an R&D center (and the size of the R&D center) and affiliate sales in the same host. Columns 4 and 5 also control for the average distance of a production facility to other countries in which the firm has an R&D center. Industry effects are at two-digit level. Standard errors (in parenthesis) are clustered by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results from the first three columns are visualized in Figure 4 of the main text.

Table A.9 reports the results. The first column focuses on the extensive margin and shows that having an R&D center in a host increases the probability of having an affiliate with sales in the same country by 0.28, or about ten times the mean value of the outcome variable (2.7%). Column 2 focuses on the intensive margin of sales and finds that MNCs' affiliate sales are on average 116% larger in hosts where they have an R&D center. Column 3 uses log patent count as the explanatory variable. The sample size is substantially smaller, but the result is qualitatively similar: sales are correlated with the number of inventions at the affiliate level.

To explore whether R&D of the firm at other locations is correlated with the sales of an affiliate, Columns 4 (5, respectively) include the average value of the four distance measures between h and all countries in which firm f has an R&D center in the specification of Columns 2 (3, respectively). The distance coefficients, although not always statistically significant, are generally supportive of proximity to sibling R&D centers having an effect. Importantly, the coefficients for R&D Indicator$_{fh,t}$ and ln(patent)$_{fh,t}$ do not change much, underscoring the importance of co-location.

To ensure that the correlation is not due to idiosyncratic match quality between a firm and a host in both R&D and production, Column 6 controls for affiliate fixed effects; to further rule out the effect of *changes* in host economy that affects the entire industry, Column 7 adds host-industry-period fixed effects. Both specifications find that as affiliates increase invention, their production also increase.[16]

A remaining concern is that the correlation can be driven by over-time *changes* in the idiosyncratic match quality between a host and particular affiliates. Given that the estimate changes little with the inclusion of host-industry-period fixed effects, it appears that the scope for such shocks to affect the

[16] Two remarks on the within-affiliate specification: First, when exploiting over-time variations, firms growing from a small number of patents can have an extraordinary percentage growth rate. To avoid these firms having an outsized impact on the estimate, I weight firms by the square root of their patent counts. Not weighting would result in an estimate of 0.06 (s/e= 0.014) for Column 6 and an estimate of 0.09 (s/e= 0.02) for Column 7. Second, I do not include the average distance metrics as there are not enough over-time changes to estimate them precisely. Including these variables will not impact the coefficient for log(patent) materially.

estimate is limited. Nevertheless, I adopt an alternative IV strategy to further address this concern, under the following identifying assumption: controlling for host fixed effects and other time-varying host, industry, and firm characteristics, changes in the R&D environment of a host—e.g., R&D subsidies and the number of researchers—affect affiliate sales through affiliate R&D. The IV results, which is consistent with findings from Table A.9, is reported in the Supplementary Appendix (SA.A.3).

The robust finding in Table A.9 lends support to the existence of frictions impeding the separation of invention and production. **Results from the first three columns are visualized in Figure 4 of the text.**

Fact 3. The third fact concerns the headquarter effects on affiliate invention and sales. The specification is the following:

$$y_{foh,t} = FE + \overrightarrow{\gamma}_{dist} \cdot \overrightarrow{dist}_{oh} + \epsilon_{ohf,t},$$

in which the outcome variable $y_{foh,t}$ is measures of invention or sales in host h of firm f from country o in period t. $\overrightarrow{dist}_{oh}$, the distance measures between headquarters o and host h, is the focus of this specification. I exclude headquarters from the sample, so the comparison is among affiliates of the same firm in different countries.

Columns 1 and 3 of Table A.10 report the results of a linear probability model in which the outcome variable is an indicator for having an R&D center or affiliate sales in h. They show that geographic frictions play important but heterogeneous roles. Sharing a language is important for both invention and production, whereas distance and colonial ties matter more for production. These estimates are economically sizable compared to the mean value of the dependent variables (0.018 and 0.027, respectively). Columns 2 and 4 estimate the intensive margin effect of distance to the headquarters on affiliate activities. Sharing a common language is more important for invention, but other types of geographic frictions are in general more important for sales. Columns 5 and 6 the proximity to the headquarters matter for affiliate sales, along both extensive and intensive margins, even after the co-location indicator is controlled for.

These results indicate another possible role of geography: by affecting knowhow transfer, it can limit the reach of the activities by multinational firms. **Figure 4 of the main text visualize variants of Columns 2 and 4 that do not control for common language, contiguity, and colonial tie indicators.**

Table A.10: The Headquarter Effect on Invention and Sales

	(1)	(2)	(3)	(4)	(5)	(6)
	Affiliate Invention		Affiliate Sales			
Dependent var.	indicator	ln(patent)	indicator	ln (sales)	indicator	ln (sales)
$\ln(\text{distance})_{oh}$	-0.002**	-0.129***	-0.005***	-0.282***	-0.005***	-0.253***
	(0.001)	(0.034)	(0.002)	(0.028)	(0.001)	(0.027)
common language_{oh}	0.020***	0.258***	0.022***	0.162**	0.015**	0.094
	(0.004)	(0.072)	(0.009)	(0.064)	(0.007)	(0.061)
contiguity_{oh}	0.002	0.106	0.004	0.185***	0.003	0.174***
	(0.002)	(0.072)	(0.004)	(0.061)	(0.004)	(0.058)
colonial tie_{oh}	0.002	0.029	0.024***	0.153**	0.023***	0.129*
	(0.004)	(0.067)	(0.008)	(0.075)	(0.007)	(0.068)
R&D $\text{indicator}_{fh,t}$					0.375***	1.198***
					(0.019)	(0.031)
Observations	7295102	45364	7295102	103131	7295102	103131
R^2	0.124	0.336	0.302	0.420	0.339	0.446
Within R^2	0.004	0.012	0.006	0.012	0.058	0.056
Firm-period FE	Y	Y	Y	Y	Y	Y
Host-industry FE	Y	Y	Y	Y	Y	Y
Host-period FE	Y	Y	Y	Y	Y	Y

Note: Columns 1 and 2 estimate the relationship between affiliate invention and the proximity of the host to the headquarters. Columns 3 to 6 estimate the relationship between the proximity to the headquarters and affiliate production, among which Columns 5 and 6 also include the R&D center indicator. Headquarters are excluded from all regressions. Standard errors (in parenthesis) are clustered by country pair. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Theory

B.1 An Isomorphic Model with Vertical R&D at Affiliates

The model presented in the text assumes that varieties developed by the same R&D center are differentiated from each other, which can be a strong assumption. In this subsection, I show that the model can be alternatively micro-founded using vertical quality improvement at affiliate R&D centers.

Suppose by paying the fixed cost for a laboratory (lab), a firm obtains only *one* new differentiated variety. Firm can improve the quality of the product (or the efficiency with which the product can be manufactured) by hiring researchers to work in the lab. Assume product quality q is given by firms' knowhow and R&D investment as follows:

$$q = \underbrace{\tilde{z}^P \cdot \tilde{z}^R}_{\text{knowhow}} \cdot \underbrace{h^{\tilde{\gamma}}}_{\text{R\&D input}} \iff \log(q) = \log\left(\tilde{z}^P \cdot \tilde{z}^R\right) + \tilde{\gamma} \cdot \log(h). \tag{B.1}$$

Slightly abusing notation, I use tilde to denote the parameters in this isomorphic model. Here $\tilde{\gamma}$ determines how much R&D improves product quality.

Firms can produce this variety in any country, but offshore production is costly. Let the effective quality if the product is to be manufactured in m is given by $\phi^P_{oim} \cdot \eta_m$, in which η_m is an idiosyncratic efficiency draw from the following distribution:

$$H(\boldsymbol{x}|q) \equiv \text{Prob}(\eta_1 \leq x_1, ..., \eta_N \leq x_N) = \begin{cases} 1 - (\sum\limits_{m=1}^{N} (q)^{\tilde{\theta}} / N x_m^{-\tilde{\theta}}), \forall m \in \{1, ...N\},\ x_m \geq q \\ 0,\ \exists m \in \{1, ..., N\},\ x_m < q \end{cases}$$

Compared to the benchmark model, the difference is that exogenous manufacturing efficiency z^P is now replaced with endogenous product quality q.

Assume all the remaining aspects of the setup (e.g., CES final demand, iceberg trade cost) are the same as in the benchmark model, the *expected* total production of the firm in country m is simply the sum of its production for varieties developed in different affiliates:

$$r_{om}(\tilde{\boldsymbol{z}}^R, \tilde{\boldsymbol{z}}^P) \propto \frac{1}{N} \cdot \left[(\frac{T_m}{W^l_m})^{\theta} \right] \times \left[\sum_d \frac{X_d^{\frac{\theta}{\sigma-1}} P_d^{\theta} (W^h_d f^M_d)^{\frac{\theta+1-\sigma}{1-\sigma}}}{(\tau_{md})^{\theta}} \right] \times \left[\sum_{i \in \boldsymbol{R}} [\tilde{z}^P_i \cdot \tilde{z}^R_i \cdot h_{oi}(\tilde{z}^R_i, \tilde{z}^P_i)^{\tilde{\gamma}}]^{\tilde{\theta}} (\phi^P_{oim})^{\tilde{\theta}} \right]. \tag{B.2}$$

Setting $\tilde{\theta} = \theta$, $\tilde{\gamma} = \frac{\gamma}{\theta}$, $\tilde{z}^R_i = (z^R_i)^{\frac{1}{\theta}}$, $\tilde{z}^P_i = z^P_i$, then equation (B.2) is isomorphic to equation (8) in the text, and they imply the same first-order condition for the intensive margin R&D decisions.[17]

The above discussion takes the extensive margin offshore R&D decision as given. Because each lab invents one differentiated variety and make vertical improvement only on that variety, adding R&D centers affects production in any host country m additively. Therefore, in the absence of fixed cost for production, this alternative setup also implies that a firm's decision to establish offshore R&D centers in different countries are independent—exactly as in the benchmark model.

This analysis shows that the benchmark model can be interpreted alternatively as follows: the fixed cost for a lab represents the cost of developing a new variety and the intensive margin R&D expenditures represent vertical improvements for that variety. This alternative micro-foundation clarifies that it is not essential to assume that varieties within an R&D center are substitutable—tractability of the model stems from the assumption that varieties from different R&D centers are substitutable.

[17] Different from the benchmark model, in which each lab develops a continuum of varieties, here, each lab develops only one variety. As a result, the law of large numbers does not apply, and equation (B.2) corresponds to the expected revenue of a firm. However, as firms are risk-neutral, their extensive and intensive margin R&D decisions will be based on the expected revenue and expected profit, which coincide with those in the benchmark model.

B.2 Aggregation

This subsection derives a few results under Assumption 1 for aggregation. For convenience I first introduce the following lemma, which has been proved in Arkolakis et al. (2018) and is only included for this appendix to be self-contained. Proofs are provided in the Supplementary Appendix.

Lemma B.1. *Suppose* $\boldsymbol{\eta} = (\eta)_{h=1}^{N}$ *is a random variable with the following CDF:*

$$H(\boldsymbol{x}|z) \equiv Prob(\eta_1 \leq x_1, ..., \eta_N \leq x_N|z) = \begin{cases} 1 - (\sum_{m=1}^{N} \frac{z^\theta}{N} x_m^{-\theta}), \forall m \in \{1, ...N\},\ x_m \geq z \\ 0,\ \exists m \in \{1, ..., N\},\ x_m < z \end{cases}$$

Define $\zeta \equiv \max_m A_m \eta_m$*, where* $A_m,\ m = 1, ..., N$ *are positive constants. Then the following holds:*

1. *The CDF for* ζ *is*

$$Prob(\zeta \leq x|z) = \begin{cases} 1 - \tilde{A}^\theta x^{-\theta}, & \text{if } x \geq \bar{A} \\ 0, & \text{if } x < \bar{A}. \end{cases} \tag{B.3}$$

   where $\tilde{A} = z \cdot \left(\frac{1}{N}\sum_m A_m^\theta\right)^{\frac{1}{\theta}}$ *and* $\bar{A} = z \cdot \max_m A_m$

2. *The conditional expectation of* ζ *above* x *is:*

$$\mathbb{E}[\zeta|\zeta \geq x] = \frac{\theta}{\theta - 1} x,\ \forall x \geq \bar{A}.$$

3. *The conditional probability of the maximum value of* $\sum_{m'} A_{m'}\eta_{m'}$ *realizing at* m *is*

$$Prob(m = \arg\max_{m'} A_{m'}\eta_{m'} \Big| \zeta \geq x) = \frac{A_m^\theta}{\sum_m A_m^\theta},\ \forall x > \bar{A},\ \forall m = 1, 2, ..., N. \tag{B.4}$$

   Moreover, the distribution of ζ *conditional on the maximum value realizing at* m *is:*

$$Prob(\zeta \geq x'|m = \arg\max_{m'} A_{m'}\eta_{m'} \wedge \zeta \geq x\Big) = (\frac{x'}{x})^{-\theta},\ \forall x' \geq x > \bar{A}, \tag{B.5}$$

   which is independent of m*.*

Deriving equations (5), (6), (11) and (12). With Lemma B.3, I derive expressions for a few aggregate objects. For convenience, define a new random variable $\zeta_{oid} \equiv \max_m \frac{T_m \phi_{oim}^P}{W_m^m \tau_{md}} \cdot \eta_m$, then from the first part of the Lemma, the distribution of ζ_{oid} is given by:

$$H_{oid}(x|z^P) \equiv Prob(\zeta_{oid} \leq x|z^P) = \begin{cases} 1 - (\frac{\tilde{\zeta}_{oid}}{x} \cdot z^P)^\theta,\ x \geq \overline{\zeta_{oid}} \cdot z^P \\ 0,\ \ x < \overline{\zeta_{oid}} \cdot z^P \end{cases}$$

where $\tilde{\zeta}_{oid} \equiv \left(\sum_m \frac{1}{N}(\frac{T_m \phi_{oim}^P}{W_m^l \tau_{md}})^\theta\right)^{\frac{1}{\theta}}$ and $\overline{\zeta_{oid}} \equiv \max_m \frac{T_m \phi_{oim}^P}{W_m^l \tau_{md}}$. Here $\tilde{\zeta}_{oid} \cdot z^P$ and $\overline{\zeta_{oid}} \cdot z^P$ correspond to $\tilde{A}$ and $\overline{A}$ defined in equation (B.3), respectively.

Cast $p_{oid}(\omega)$ in the space of ζ_{oid}, then the price for a particular realization can be written as $p_{oid}(\zeta) = \frac{\sigma}{\sigma-1} \cdot \frac{1}{\zeta_{oid}}$. The probability that the product is manufactured in m is simply the probability that the best realization of ζ_{oid} realizes in m. From the third part of the lemma, this probability is independent of z^P, and given by

$$\psi_{oimd} \equiv Prob(m = \max_m \frac{T_m \phi^P_{oim}}{W^l_m \tau_{md}} \cdot \eta_m | \zeta_{oid} > x, z^P) = \frac{\frac{1}{N}(\frac{T_m \phi^P_{oim}}{W^l_m \tau_{md}})^\theta}{\frac{1}{N}\sum_{m'}(\frac{T_m \phi^P_{oim'}}{W^l_{m'} \tau_{m'd}})^\theta}, \forall x \geq \bar{\zeta}_{oid} \cdot z^P \tag{B.6}$$

Because the conditional distribution of ζ_{oid} is the same regardless of which country ends up with the maximum value for $\frac{T_m \phi^P_{oim}}{W^l_m \tau_{md}} \cdot \eta_m$ (equation (B.5)), the above choice probability is also equal to the share of sales produced in m.

For later use, I calculate the following:

$$\int_0^\infty \mathbb{1}(p_{oid}(\zeta) < \hat{p}_d) \cdot p_{oid}(\zeta)^{1-\sigma} dH_{oid}(\zeta | z^P) = \frac{\theta}{\theta - (\sigma - 1)} (\frac{\sigma - 1}{\sigma})^\theta P_d^{\theta+1-\sigma} (\frac{\sigma W^h_d f^M_d}{X_d})^{\frac{\theta+1-\sigma}{1-\sigma}} (\tilde{\zeta}_{oid} z^P)^\theta \tag{B.7}$$

I define $\bar{r}_{oid}(z^P)$ to be the expected revenue. Combine equation (B.7) with the definition of $\bar{r}_{oid}(z^P)$ to obtain:

$$\begin{aligned}
\bar{r}_{oid}(z^P) &= \frac{X_d}{P_d^{1-\sigma}} \int_0^\infty \mathbb{1}(p_{oid}(\zeta) < \hat{p}_d) \cdot p_{oid}(\zeta)^{1-\sigma} dH_{oid}(\zeta | z^P) \\
&= \frac{\theta}{\theta - (\sigma - 1)} (\sigma - 1)^\theta \sigma^{1 - \frac{\theta\sigma}{\sigma - 1}} X_d^{\frac{\theta}{\sigma - 1}} P_d^\theta (W^h_d f^M_d)^{\frac{\theta+1-\sigma}{1-\sigma}} (\tilde{\zeta}_{oid} z^P)^\theta
\end{aligned} \tag{B.8}$$

Define the expected marketing cost incurred for a variety as $\bar{f}^M_{oid}(z^P)$, then similar steps give:

$$\bar{f}^M_{oid}(z^P) = f^M_d W^h_d \int_0^\infty \mathbb{1}(p_{oid}(\zeta) < \hat{p}_d) dH_{oid}(\zeta | z^P) = (\frac{\sigma - 1}{\sigma})^\theta \sigma^{\frac{\theta}{1-\sigma}} (f^M_d W^h_d)^{\frac{\sigma-1-\theta}{\sigma-1}} X_d^{\frac{\theta}{\sigma-1}} P_d^\theta (\tilde{\zeta}_{oid} z^P)^\theta. \tag{B.9}$$

The operational profit is simply the difference between markup and marketing cost, given by:

$$\begin{aligned}
\bar{\pi}_{oid}(z^P) &= \frac{1}{\sigma} \frac{X_d}{P_d^{1-\sigma}} \int_0^\infty \mathbb{1}(p_{oid}(\zeta) < \hat{p}_d) \cdot p_{oid}(\zeta)^{1-\sigma} dH_{oid}(\zeta | z^P) - f^M_d W^h_d \int_0^\infty \mathbb{1}(p_{oid}(\zeta) < \hat{p}_d) dH_{oid}(\zeta | z^P) \\
&= \frac{(\sigma - 1)^{1+\theta}}{\theta - (\sigma - 1)} \sigma^{\frac{\sigma\theta}{1-\sigma}} (f^M_d W^h_d)^{\frac{\sigma-1-\theta}{\sigma-1}} X_d^{\frac{\theta}{\sigma-1}} P_d^\theta (\tilde{\zeta}_{oid} z^P)^\theta.
\end{aligned} \tag{B.10}$$

Equations (B.8), (B.9), (B.10) immediately imply:

$$\frac{\bar{f}^M_{oid}(z^P)}{\bar{r}_{oid}(z^P)} = \frac{\theta - (\sigma - 1)}{\theta\sigma}, \qquad \frac{\bar{\pi}_{oid}(z^P)}{\bar{r}_{oid}(z^P)} = \frac{\sigma - 1}{\theta\sigma}.$$

I now derive the aggregate price index P_d and trade flows X_{oid}

$$\begin{aligned}
P_d^{1-\sigma} &= \sum_o \sum_i R_{oi} \int_0^\infty \int_0^\infty v_{oi}(z^P, z^R) [\int_0^\infty \mathbb{1}(p_{oid}(\zeta) < \hat{p}_d) \cdot p_{oid}(\zeta)^{1-\sigma} dH_{oid}(\zeta | z^P)] \cdot g_{oi}(z^P, z^R) dz^P dz^R \\
&= \frac{\theta}{\theta - (\sigma - 1)} (\frac{\sigma - 1}{\sigma})^\theta P_d^{\theta+1-\sigma} (\frac{\sigma W^h_d f^M_d}{X_d})^{\frac{\theta+1-\sigma}{1-\sigma}} \sum_o \sum_i \tilde{\zeta}^\theta_{oid} \int_0^\infty (z^P)^\theta V_{oi}(z^P) dz^P.
\end{aligned}$$

$$X_{oid} = P_d^{\sigma-1} X_d \cdot R_{oi} \int_0^\infty \int_0^\infty v_{oi}(z^P, z^R)[\int_0^\infty \mathbb{1}(p_{oid}(\zeta) < \hat{p}_d) \cdot p_{oid}(\zeta)^{1-\sigma} dH_{oid}(\zeta|z^P)] \cdot g_{oi}(z^P, z^R) dz^P dz^R$$
$$= \frac{\theta}{\theta-(\sigma-1)} (\frac{\sigma-1}{\sigma})^\theta (\frac{X_d}{P_d^{1-\sigma}})^{\frac{\theta}{\sigma-1}} (\sigma W_d^h f_d^M)^{\frac{\theta+1-\sigma}{1-\sigma}} \tilde{\zeta}_{oid}^\theta \int_0^\infty (z^P)^\theta V_{oi}(z^P) dz^P.$$

B.3 Definition of Equilibrium

Definition 1. *Given the fundamentals, a competitive equilibrium of the model is characterized by a set of decision rules, prices, and allocations, such that $\forall o, i, d = 1, .., N$ the following holds:*

1. *Firms' production, market entry, and pricing decisions for each individual variety are optimal, which implies that the following holds $\forall z^P \in \mathbb{Z}^P$:*

$$\bar{r}_{oid}(z^P) = \frac{\theta(\sigma-1)^\theta \sigma^{1-\frac{\theta\sigma}{\sigma-1}}}{\theta-(\sigma-1)} X_d^{\frac{\theta}{\sigma-1}} P_d^\theta (W_d^h f_d^M)^{\frac{\theta+1-\sigma}{1-\sigma}} (\tilde{\zeta}_{oid} z^P)^\theta, \quad \bar{f}_{oid}^M(z^P) = \frac{\theta-(\sigma-1)}{\theta\sigma} \bar{r}_{oid}(z^P) \quad \text{(B.11)}$$
$$\bar{\pi}_{oid}(z^P) = \frac{1}{\sigma} \bar{r}_{oid}(z^P) - \bar{f}_{oid}^M(z^P) = \frac{\sigma-1}{\theta\sigma} \bar{r}_{oid}(z^P), \qquad \bar{\pi}_{oi}(z^P) = \sum_d \bar{\pi}_{oid}(z^P),$$
$$\textit{where } \tilde{\zeta}_{oid} \equiv [\sum_m \frac{1}{N} (\frac{T_m \phi_{oim}^p}{W_m^l \tau_{md}})^\theta]^{\frac{1}{\theta}}.$$

2. *Firms' R&D and offshore R&D decisions satisfy the following:*

$$v_{oi}(z^P, z^R) = z^{R\frac{1}{1-\gamma}} \Big(\frac{\gamma \bar{\pi}_{oi}(z^P)}{W_i^h}\Big)^{\frac{\gamma}{1-\gamma}}, \qquad \pi_{oi}^R(z^P, z^R) = (\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}) \Big(\frac{1}{W_i^h}\Big)^{\frac{\gamma}{1-\gamma}} \big(\bar{\pi}_{oi}(z^P) z^R\big)^{\frac{1}{1-\gamma}} \quad \text{(B.12)}$$
$$\pi_{oi}^R(z^R) = \int_0^\infty \pi_{oi}^R(z^P, z^R) g^P(z^P|z^R) dz^P, \quad \pi_{oi}^R(\hat{z}_{oi}^R \phi_{oi}^R) = f_{oi}^R W_i^h$$

3. *The distribution of R&D center innovation efficiency in each host is consistent with firms' offshore R&D decisions and the endowment distribution of the origin countries:*

$$R_{oi} = E_o \cdot \Big(1 - G_o^E(\hat{z}_{oi}^R)\Big) \quad \text{(B.13)}$$
$$g_{oi}^R(z^R) = \frac{1}{R_{oi}} \mathbb{1}(z^R > \hat{z}_{oi}^R \phi_{oi}^R) \cdot E_o \cdot g_o^E(\frac{z^R}{\phi_{oi}^R}) \cdot \frac{1}{\phi_{oi}^R}$$
$$g_{oi}(z^P, z^R) = g^P(z^P|z^R) g_{oi}^R(z^R).$$

4. *Firm decisions are consistent with aggregate trade flows*

$$
\begin{aligned}
V_{oi}(z^P) &= R_{oi} \int_0^\infty v_{oi}(z^P, z^R) \cdot g_{oi}(z^P, z^R) dz^R && \text{(B.14)} \\
X_{oid} &= \theta(\frac{\sigma}{\sigma - 1})^{-\theta} \frac{1}{\theta - (\sigma - 1)} (\sigma W_d^h f_d^M)^{\frac{\theta-(\sigma-1)}{1-\sigma}} \left(\frac{X_d}{P_d^{1-\sigma}}\right)^{\frac{\theta}{\sigma-1}} \tilde{\zeta}_{oid}^\theta \int_0^\infty (z^P)^\theta V_{oi}(z^P) dz^P \\
X_{oimd} &= \psi_{oimd} X_{oid} \\
Y_{om} &= \frac{\sigma - 1}{\sigma} \sum_{i,d} X_{oimd} \\
\Pi_{oi} &= \frac{1-\gamma}{\sigma} \left(\frac{\sigma - 1}{\theta}\right) \sum_{m,d} X_{oimd} \\
F_{od}^M &= \frac{1}{\sigma} \left(\frac{\theta - (\sigma - 1)}{\theta}\right) \sum_{i,m} X_{oimd} \\
I_{oi} &= \frac{\gamma}{1-\gamma} \Pi_{oi} \\
F_{oi}^R &= \mathbb{1}(o \neq d) \cdot E_o \cdot [1 - G_o^E(\hat{z}_{od}^R)] \cdot f_{oi}^R W_i^h,
\end{aligned}
$$

and the aggregate price indices

$$
P_d^{1-\sigma} = \theta(\frac{\sigma}{\sigma - 1})^{-\theta} \frac{1}{\theta - (\sigma - 1)} \left(\frac{\sigma W_d^h f_d^M}{X_d}\right)^{\frac{\theta-(\sigma-1)}{1-\sigma}} P_d{}^{\theta-(\sigma-1)} \sum_o \sum_i \tilde{\zeta}_{oid}^\theta \int_0^\infty (z^P)^\theta V_{oi}(z^P) dz^P \tag{B.15}
$$

5. *Workers' occupation choices are optimal:*

$$
\begin{aligned}
W_d^h \hat{\alpha}_d &= W_d^l && \text{(B.16)} \\
L_d^h &= L_d \cdot \int_{\alpha > \hat{\alpha}_d} \alpha \, dA_d(\alpha) \\
L_d^l &= L_d \cdot A_d(\hat{\alpha}_d).
\end{aligned}
$$

6. *Labor markets clear*

$$
\begin{aligned}
W_d^h L_d^h &= \sum_o I_{od} + \sum_o F_{od}^M + \sum_o F_{od}^R && \text{(B.17)} \\
W_d^l L_d^l &= \sum_o Y_{od}
\end{aligned}
$$

7. *Total income equals total expenditures*

$$
X_d = W_d^h L_d^h + W_d^l L_d^l + \sum_i (\Pi_{di} - F_{di}^R) \tag{B.18}
$$

Note that once $\{\hat{\alpha}_d, W_d^l, W_d^h : d \in 1, ..., N\}$, $\{X_d : d \in 1, ..., N\}$, and $\{P_d : d \in 1, ..., N\}$ are known, all other endogenous variables appearing in the above definition can be calculated sequentially with equations (B.11), (B.12), (B.13), and (B.14). The competitive equilibrium can therefore be viewed as a fixed point in $\{\hat{\alpha}_d, W_d^l, W_d^h : d \in 1, ..., N\}$, $\{X_d : d \in 1, ..., N\}$, and $\{P_d : d \in 1, ..., N\}$, so that equations (B.15), (B.16) , (B.17), and (B.18) hold, with endogenous objects in these four set of equations defined implicitly as functions of wages, prices and expenditures by equations (B.11), (B.12), (B.13), and (B.14).

C Quantification

This section provides additional details on the quantification procedure and counterfactual results.

C.1 Additional Data for Quantification

Sample countries. The model economy consists of the same 37 countries as in the empirical section. See Section A.3 of this appendix for a discussion on the sample selection. Among these countries, Ireland is frequently dubbed as a 'tax-haven' country, in which reported financial statistics might be unreliable. I include Ireland in the sample because this allows me to not take a stand on how to redirect the foreign linkages that run through it. That said, if these links are simply assumed to be non-existence, measured openness for other countries will be similar.

Country-specific openness measures. In disciplining host specific barriers to inward MNC activities, I use three targets: the share of production by foreign firms, the share of innovation by foreign R&D centers, and the share of foreign R&D centers in total R&D center counts. These targets are calculated from the firm-level data described in Section A.3, with the following modifications. First, given my interpretation of the model as for manufacturing, I focus only on manufacturing firms in calculating these ratios. Second, all firms in the model carry out some R&D, but not all firms in the data are granted patents. Instead of using the joint sample to construct the ratios, I use the full financial sample to calculate inward offshore production measure, and the full R&D sample to calculate the two (extensive and intensive margins) inward offshore R&D measures. Third, for the three countries for which the financial data have relatively low coverage (Mexico, Turkey, U.S.), instead of aggregating the firm-level data, I use the aggregate shares in Ramondo et al. (2019), described by Ramondo et al. (2015). Finally, recent research has found that patents granted by the Chinese patent office to local firms are systematically less likely to be global patents than the patents they grant to foreign firms, suggesting differential treatments based on where firms are from (Holmes et al., 2015). To avoid biases arising from potential discriminatory treatments, I exclude all patents issued by the Chinese patent authority when calculating the inward offshore R&D ratio for China. Table C.1 reports the three openness measures.

Endowment distributions. The calibration uses the World Management Survey (Bloom et al., 2012a) and an internationally comparable cognitive ability score database (Hanushek and Woessmann, 2012a). I take the exponent of the innovation management score so that its distribution has a right tail that resembles the firm size distribution. I compute the mean, standard deviation, and skewness of the exponent of scores in each country, which then serve as an input to the calibration. The distribution statistics for cognitive test scores are from Hanushek and Woessmann (2012b). These statistics include the average cognitive score for high school students in a country, the share of students achieving 'top' performance, and the share of student achieving 'basic' performance. Thresholds for 'top' and 'basic' performance are defined based on a common set of standards so these shares are comparable internationally.

A few countries in the sample are not included in the World Management Survey. I impute their management distribution statistics by regressing each statistics on income, R&D share, and geographic-region fixed effects, where geographic regions are at sub-continent level. The inclusion of income is motivated by the finding in Bloom et al. (2012b) that management knowhow explains a substantial share of cross-country income differences; the inclusion of geographic-region fixed effects is meant to capture management practice differences driven by culture. The R^2 of these regressions are all above 0.85.

Table C.1 reports these statistics for all countries.

C.2 Parameterization

Bilateral trade costs. I assume that the iceberg trade costs are symmetric and that trade cost with own country is 1, i.e., $\tau_{mm} = 1, \forall m$ and $\tau_{md} = \tau_{dm}, \forall m \neq d$. Under these assumptions, the approach in Head and Ries (2001) generalizes to my setting.

Table C.1: Calibration Targets: Country Characteristics

	Income and Openness					Innovation Mgt. Dist.			Talent Dist.		
ISO	$\frac{X_m}{P_m}$	$\frac{\sum_i V_{oi}}{\sum_{o,i} V_{oi}}$	$\frac{\sum_{o\neq m} Y_{om}}{\sum_o Y_{om}}$	$\frac{\sum_{o\neq i} V_{oi}}{\sum_o V_{oi}}$	$\frac{\sum_{o\neq i} R_{oi}}{\sum_o R_{oi}}$	mean	std	skew.	mean	% basic	% top
AUS	0.81	1.03	60.87	39.78	33.38	6.43	3.64	1.88	5.09	93.84	11.24
AUT	0.76	0.59	63.09	45.49	41.70	6.91	4.14	2.15	5.09	93.11	9.74
BEL	0.81	0.45	82.31	58.85	54.32	7.14	4.48	2.27	5.04	93.13	9.38
BGR	0.29	0.05	32.02	19.02	17.04	6.80	3.93	1.50	4.79	76.53	8.30
BRA	0.26	1.10	19.17	57.58	46.73	5.26	3.33	2.34	3.64	33.85	1.09
CAN	0.73	1.20	35.35	52.21	47.55	8.40	6.09	2.01	5.04	94.84	8.33
CHE	0.90	1.51	36.21	41.68	44.57	7.40	4.88	2.41	5.14	91.85	13.36
CHN	0.17	19.84	23.93	41.52	21.93	5.94	2.74	1.85	4.94	93.48	8.34
CZE	0.52	0.28	62.14	26.59	22.89	6.47	3.44	1.33	5.11	93.07	12.22
DEU	0.77	6.85	46.54	28.16	29.94	8.21	5.25	2.20	4.96	90.60	10.52
DNK	0.80	0.40	63.62	38.37	38.26	7.28	5.46	2.75	4.96	88.78	8.75
ESP	0.74	1.15	49.89	22.09	16.93	5.29	3.46	2.21	4.83	85.88	7.93
EST	0.46	0.03	53.49	28.48	34.27	6.11	3.71	2.14	5.19	97.32	9.46
FIN	0.73	0.80	28.05	24.86	18.43	6.84	4.79	2.51	5.13	95.78	12.39
FRA	0.82	3.89	42.68	26.58	40.34	6.43	4.25	2.52	5.04	92.62	8.49
GBR	0.70	1.85	87.73	62.95	53.21	7.36	5.04	2.24	4.95	92.88	8.79
GRC	0.59	0.09	24.51	50.08	80.00	5.63	3.70	1.89	4.61	79.77	4.24
HRV	0.51	0.03	28.34	53.37	60.87	5.45	3.21	1.93	4.70	83.35	4.76
HUN	0.48	0.08	47.70	61.09	38.01	6.56	3.57	1.37	5.05	94.11	10.28
IRL	1.13	0.69	77.34	73.67	68.78	7.14	6.73	3.86	4.99	91.37	9.40
ITA	0.79	1.23	33.85	43.84	29.03	6.47	4.15	2.17	4.76	87.54	5.45
JPN	0.63	11.76	6.65	1.99	17.42	7.83	5.57	1.82	5.31	96.67	16.76
KOR	0.60	4.87	9.71	6.72	14.33	7.06	4.42	1.92	5.34	96.16	17.84
LTU	0.55	0.04	46.93	20.30	16.76	6.70	4.59	2.44	4.78	89.07	2.97
LVA	0.45	0.02	38.21	6.07	31.25	5.94	3.45	2.05	4.80	86.95	4.99
MEX	0.34	0.33	17.80	55.01	64.30	6.90	4.43	1.66	4.00	48.93	0.88
NLD	0.80	2.01	93.21	30.27	29.58	7.34	4.79	2.37	5.11	96.54	9.16
NOR	1.34	0.39	46.67	32.27	22.60	8.81	7.72	3.54	4.83	89.44	5.61
POL	0.51	0.47	49.96	14.61	24.63	7.25	4.60	1.73	4.85	83.76	9.86
PRT	0.55	0.21	46.40	26.20	40.78	5.38	2.99	1.94	4.56	80.27	3.16
ROU	0.41	0.06	64.47	52.75	47.65	6.63	3.68	1.41	4.56	78.05	4.56
RUS	0.42	2.32	35.17	11.63	13.63	6.82	3.96	1.51	4.92	88.35	8.05
SVK	0.53	0.06	67.86	27.71	44.79	6.89	4.06	1.54	5.05	90.55	11.16
SVN	0.51	0.08	36.61	17.17	19.96	5.40	3.14	1.90	4.99	93.89	6.12
SWE	0.79	1.13	48.39	43.20	34.43	7.06	4.17	1.99	5.01	93.94	8.76
TUR	0.54	0.73	6.34	20.12	23.86	5.86	2.58	2.09	4.13	58.23	3.92
USA	1.00	32.38	15.29	15.77	13.62	10.94	8.15	2.15	4.90	91.82	7.33

Note: This table reports the country-level statistics used as targets in parameterization. The first set of targets are on the income and openness of countries, corresponding to Panel B of Table 5. These columns are: real income (U.S. normalized to 1), the contribution (%) of a country to the world world R&D (based on the origin of firms), the share (%) of domestic production by foreign firms, the share (%) of domestic R&D by foreign R&D centers, and the share (%) of foreign R&D centers among all active R&D centers in a host. '*Innovation Mgt. Dist.*' refers to the sample distribution statistics constructed from the World Management Survey as described in Section C.1. '*Talent Dist.*' refers to the talent distribution statistics from Hanushek and Woessmann (2012b), in which '*Mean*' is the mean score for a country, and '*% basic*' and '*% topc*' are shares of students achieving 'basic' and 'top' performance, respectively. The performance standards are common across countries.

To this end, using equation (B.14):

$$X_{oimd} = X_{oid}\psi_{oimd} \equiv B_d \frac{1}{N}\left(\frac{T_m \phi^P_{oim}}{W^l_m \tau_{md}}\right)^\theta \cdot B_{o,i},$$

where B_d and $B_{o,i}$ are functions of the equilibrium objects in d and in (o,i) introduced to shorten notations. Using this expression, we have:

$$\frac{\sum_{o,i} X_{oimd}}{\sum_{o,i} X_{oimm}} \cdot \frac{\sum_{o,i} X_{oidm}}{\sum_{o,i} X_{oidd}} = \frac{B_d(\frac{1}{W_m^I \tau_{md}})^\theta \sum_{o,i} \phi_{oim}^P B_{o,i}}{B_m(\frac{1}{W_m^I \tau_{mm}})^\theta \sum_{o,i} \phi_{oim}^P B_{o,i}} \times \frac{B_m(\frac{1}{W_d^I \tau_{dm}})^\theta \sum_{o,i} \phi_{oid}^P B_{o,i}}{B_d(\frac{1}{W_d^I \tau_{dd}})^\theta \sum_{o,i} \phi_{oid}^P B_{o,i}} = (\tau_{md})^{-2\theta}.$$

Notice that although the flow items such as X_{oimd} are not observable, $\sum_{o,i} X_{oidm}$ is simply the total sales from d to m, which is observable.

Slightly abusing notations, I write $\tau_{md} = (\frac{\sum_{o,i} X_{oimd}}{\sum_{o,i} X_{oimm}} \cdot \frac{\sum_{o,i} X_{oidm}}{\sum_{o,i} X_{oidd}})^{-\frac{1}{2\theta}} = (\frac{X_{md}}{X_{mm}} \times \frac{X_{dm}}{X_{dd}})^{-\frac{1}{2\theta}}$, where X_{md} denotes the sales from m to d as in the gravity literature. I obtain these sales for the aggregated manufacturing sector from the World Input Output Database.

Relating production efficiency to innovation efficiency. To discipline the relationship between firms' innovation and production management efficiency, I use micro data from the World Management Survey to estimate the following equation:

$$\text{Prob}(z^P \in H|z^R) = \frac{\exp(\delta_0 + \delta_1 \times z^R)}{1 + \exp(\delta_0 + \delta_1 \times z^R)}. \tag{C.1}$$

This dataset covers around 11338 firms from 34 countries. I classify a firm as being a H type, if its production management scores falls in the top 5% in the sample. Table C.2 presents summary statistics on this score and the indicator for H type. Table C.3 presents results from a logit regression of equation (C.1). Column 1 uses no fixed effects whereas Column 2 includes country fixed effects. Both specifications find positive and statistically significant coefficient, consistent with strong correlation between innovation efficiency and production efficiency. Based on the estimates, I set $\delta_1 = 0.21$ and $\delta_0 = -5$.

For illustration, Figure C.1 plots the parameterized distribution for the U.S. About 12% of the American firms end up being a H type.

Table C.2: Firm Management Score Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
z^R	11338	6.68	4.92	1	54.6
$\mathbb{1}(z^P \in G_H^P)$	11340	.051	0.22	0	1

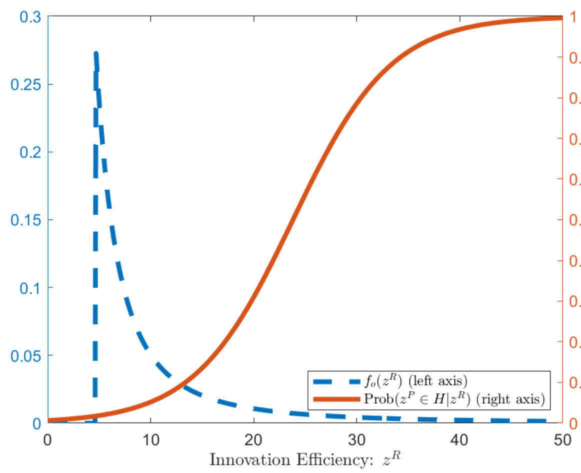
Notes: This table presents summary statistics for firm-level innovation management scores and the indicator for whether a firm is in the top 5% production efficiency.

Table C.3: Estimates for δ_0 and δ_1

	(1)	(2)
	$\mathbb{1}(z^P \in G_H^P)$	
z^R	0.213***	0.210***
	(0.00719)	(0.00797)
cons	-4.921***	-
	(0.0950)	-
N	11338	10637
pseudo R^2	0.251	0.281
country FE		yes

Notes: This table presents results from a Logit regression of the high production efficiency indicator $\mathbb{1}(z^P \in G_H^P)$, on firms' innovation efficiency, z^R. The high production efficiency indicator takes a value of 1 if the production management score of a firm is in the top 5% in the world. The second column controls for country fixed effects. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure C.1: The U.S. Firm Knowhow Distribution



Notes: The horizontal axis is the support of the innovation efficiency for firms from the U.S. The dashed line (left axis) is the probability density function of the innovation efficiency. The solid line (right axis) is the probability that a firm with a given z^R obtains a draw from $G^P_H(z^P)$, i.e., the value of equation (C.1).

Geographic friction parameters. Table C.4 reports the 22 targets used to pin down 17 geographic parameters. The left panel is the data, corresponding to regressions discussed in Section 4.3 of the text. The brackets are the 95% confidence intervals of these coefficients. The right panel reports the corresponding regression coefficients estimated with the exact same specifications using model-simulated data. The coefficients in general match the data counterparts closely. I highlight using underline the coefficient that is outside the 95% confidence interval of the original coefficient. Even this coefficient is not far off the target.

Table C.4: Calibration Targets: Geographic Friction Parameters

	A. Data (Table 4 of the text)					**B. Model**				
	Headquarter Effect			Colocation		Headquarter Effect			Colocation	
Dependent var.	R&D indicator	log (R&D)	log(sales)	log(sales)	log(sales)	R&D indicator	log (patents)	log(sales)	log(sales)	log(sales)
log(dist)$_{oh}$	-0.002	-0.129	-0.282	-0.253		-0.0003	-0.124	-0.274	-0.273	
	[-0.003,-0.000]	[-0.197,-0.062]	[-0.337,-0.227]	[-0.291,-0.214]						
Common language$_{oh}$	0.020	0.258	0.162	0.094		0.0185	0.238	0.133	0.113	
	[0.011,0.029]	[0.117,0.399]	[0.030,0.294]	[0.021,0.168]						
Contiguity$_{oh}$	0.002	0.106	0.185	0.174		0.0028	0.078	0.182	0.179	
	[-0.001,0.006]	[-0.036,0.248]	[0.059,0.311]	[0.103,0.245]						
Colonial tie$_{oh}$	0.002	0.029	0.153	0.129		0.0089	0.026	0.109	0.100	
	[-0.006,0.009]	[-0.102,0.160]	[-0.002,0.308]	[0.057,0.201]						
R&D center indicator				1.198	1.042				<u>**1.069**</u>	1.086
				[1.147,1.259]	[0.991,1.092]					
log(dist$_{fh,t}$)					-0.0235					-0.014
					[-0.073,0.025]					
Common language$_{fh,t}$					0.220					0.212
					[0.120,0.319]					
Contiguity$_{fh,t}$					0.143					0.140
					[0.046,0.239]					
Colonial tie$_{fh,t}$					0.090					0.176
					[0.001,0.179]					
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Host FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Home-host FE	N	N	N	N	Y	N	N	N	N	Y

Notes: This table reports moments that pin down geographic parameters. Panel A is a collection of reduced-form regressions, reported in Section 4.3 of the text. The bracket under each coefficient is its 95% confidence interval. The right panel reports results from regressions using model-simulated data. Coefficients highlighted by the underline are outside the 95% confidence interval.

C.3 The Connection with Bilir and Morales (2020) and a Case without 'Bridge' R&D

Empirical patterns. Figure 1c in Section 2.2 of the text shows that affiliate sales per invention is positively correlated with the firm's invention at the headquarters. I interpret the latter as a proxy for firms' innovation knowhow and use the above correlation to motivate the assumption that firms can transfer innovation knowhow—which is correlated with production knowhow—to affiliates.

Besides knowhow transfer, the correlation between affiliate sales and headquarters invention (and the correlation between headquarters sales and affiliate invention) can also reflect how R&D in one location benefits production in other locations, which connects with the findings of Bilir and Morales (2020). To zoom into this connection, Table C.5 below produces the correlation between sales and R&D in different locations.

Table C.5: Correlation between Sales and R&D at Headquarters and Affiliates

	(1)	(2)	(3)	(4)	(5)
Outcome var.		log parent sales		log affiliate sales	
average aff. R&D	0.218***	0.115***	0.015		
	(0.047)	(0.034)	(0.034)		
parent R&D		0.687***	0.558***	0.353***	0.243***
		(0.024)	(0.029)	(0.014)	(0.019)
# of R&D aff.			0.152***		
			(0.016)		
aff. R&D					0.382***
					(0.017)
Observations	3364	2330	2330	34881	6827
R^2	0.363	0.524	0.540	0.213	0.317
Within R^2	0.009	0.301	0.325	0.079	0.167

Note: Columns 1 to 3 are at the firm level and control for home country and industry fixed effects; Columns 4 and 5 are at affiliate level and control for home country, host country, and industry fixed effects. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

The first three columns focus on the correlation between affiliate R&D and parent sales. The regressions are at the firm level and control for home country and industry fixed effects. The outcome variable is the logarithm of parent sales. The main explanatory variable is the average log affiliate patents. Column 1 shows a statistically significant coefficient for this variable. This alone, however, does not mean parent's sales benefit significantly from affiliate R&D. Indeed, the coefficient can also be driven by the transfer of firm knowhow. The second column controls for the logarithm of patents at the HQ. This control serves two purposes. First, it captures the direct effect of parent R&D on parent sales. Second, it is a proxy for innovation knowhow. Not surprisingly, the coefficient for parent R&D is sizable and significant, and the within R^2 more than triples with this control. On the other hand, the coefficient of average affiliate R&D decrease by half. The third column controls for the number of R&D affiliates, which can be positively correlated with parent sales for two reasons. First, to the extent that parent R&D is only a noisy proxy for firm knowhow, the number of R&D affiliates can absorb the remaining knowhow variation unexplained by parent R&D. Second, conditional on the average affiliate R&D, the total spillovers from affiliate R&D to parent sales should increase in the number of affiliates. The regression suggests that having one more R&D affiliate is correlated with a 15% increase in parent sales. Importantly, the average affiliate R&D becomes statistically insignificant—once the proxies for correlation via knowhow is controlled for, affiliate R&D does not increase parent sales.

I contrast these findings with results on the relationship between *affiliate sales* and *parent R&D*, reported in Columns 4 and 5. The regression are at the affiliate level and control for home country, host country, and industry fixed effects. Column 4 shows that parent R&D is strongly correlated with affiliate sales; Column 5 shows that after controlling for the R&D at the same affiliate, which both has a direct effect on sales and serves as a proxy for the knowhow from the parent, the coefficient of parent R&D

Table C.6: Gains from Bridge R&D and the Role of Bridge R&D for the Gains from Openness

	The Gains from Bridge R&D (1)	**Gains from Openness in Re-Calibrated Model w/o Bridge R&D** (2)
BRA	0.18	2.28
CHN	0.14	2.80
POL	1.33	9.83
RUS	0.71	6.02
BEL	4.67	34.77
FRA	1.58	15.75
JPN	0.90	7.62
USA	2.90	22.03
Mean (all)	1.96	15.86
Std (all)	1.89	16.55

Notes: All numbers are in percent. The first column reports the gains from bridge offshore R&D, defined as the percentage decrease in real income as bridge offshore R&D is shut down from the baseline equilibrium; the second column reports the gains from openness in a re-calibrated model without bridge offshore R&D. See text in Appendix C.3 for detailed descriptions.

diminishes but remains sizable.

If we interpret the *residual correlation* between affiliate R&D and parent sales (in Columns 2 and 3); and between affiliate sales and parent R&D (in Column 5), as evidence for R&D spillovers, these results suggest that parent R&D has a larger impact on affiliate sales than affiliate R&D on parent sales. This is broadly consistent with Bilir and Morales (2020) and my model. See Section 3.3 for a discussion.

Quantitative exploration of a case without 'Bridge' offshore R&D. As previewed in footnote 18, in the baseline calibration, only 15% of all affiliate production is devoted to varieties invented by the sibling R&D centers of the affiliate. This implies a positive, although quantitatively small role for 'bridge' offshore R&D—R&D at one affiliate influencing production at the firm's other affiliates. In this appendix, I explore the role of bridge offshore R&D through two exercises.

In the first exercise, I shut down bridge offshore R&D from the baseline equilibrium by setting $\phi^P_{oim} \to 0, \forall o, m, i$, that satisfy $\mathbb{I}(o \neq i)\mathbb{I}(o \neq m)\mathbb{I}(i \neq m) = 1$ and solving for the counterfactual equilibrium. This change implies that the varieties developed in one affiliate cannot be produced at other affiliates. I then calculate the gains from bridge offshore R&D by comparing the real income between the baseline equilibrium and this counterfactual equilibrium. Table C.6 Column 1 reports the results. On average, countries gain 1.96% from bridge offshore R&D. As a special form of offshore R&D, it accounts for slightly more than half of the overall gains from offshore R&D that is reported in Column 1 of Table 8. This result might appear surprising, given that bridge R&D account for only 15% of affiliate sales. Note, however, that affiliate R&D requires fixed cost investment, so without the profit from bridge R&D, firms might not recoup the fixed R&D cost.

In the second exercise I re-calibrate a model imposing $\phi^P_{oim} \to 0, \forall o, m, i$, that satisfy $\mathbb{I}(o \neq i)\mathbb{I}(o \neq m)\mathbb{I}(i \neq m) = 1$.[18] I then calculate the gains from openness in this re-calibrated model. Table C.6 Column 2 reports the results. The average gains from openness are around 15.86, similar to the gains from openness in the baseline model that is reported in Column 4 of Table 8.

To see how incorporating offshore R&D this way—i.e., without allowing for bridge offshore R&D—affects the inferred gains from openness, we can compare Column 2 of Table C.6 to Column 5 of Table 8.[19] This comparison shows that, on average, incorporating offshore R&D increases the gains from openness

[18] I keep other components of geographic parameters as in the baseline model and re-calibrate parameters in Panel B of Table 5 to match the corresponding targets.

[19] Column 5 of Table 8 remains a relevant benchmark—if we shut down all offshore R&D in this alternative model and then re-calibrate it to calculate the gains from openness, we would arrive exactly at the numbers in Column 5 of Table 8.

by a factor of 1.3 (15.86/12.15), same as in the case of the baseline model.

Therefore, if we have imposed that affiliate R&D cannot benefit production at sibling affiliates, we would have reached a similar conclusion on the impact of incorporating offshore R&D for the inferred gains from openness.[20] This reflects that in the baseline calibration, only a moderate share of offshore R&D is for production at other offshore locations. It is instructive to compare this case to the re-calibrated model with $s = 0$, reported in Column 7 of Table 8. There, we find substantially larger gains from openness. The reason is that a substantial part of offshore R&D is for local production and by setting $s = 0$, we divert most of these to the headquarters. Thus, to match the same inward MP ratios, we need much more integration among countries (measured using the ratios in Proposition 4) than in the baseline model, leading to larger gains from openness.

References

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple, "Innovation and Production in the Global Economy," *American Economic Review*, 2018, *108* (8), 2128–73.

Bilir, L Kamran and Eduardo Morales, "Innovation in the Global Firm," *Journal of Political Economy*, 2020, *128* (4), 1566–1625.

Bloom, Nicholas, Christos Genakos, Raffaella Sadun, and John Van Reenen, "Data For 'Management Practices Across Firms And Countries'," https://worldmanagementsurvey.org/wp-content/uploads/2012/07/AMP1.zip, 2012. Accessed: 2016-03-22.

_ , _ , _ , and _ , "Management Practices Across Firms And Countries," *The Academy Of Management Perspectives*, 2012, *26* (1), 12–33.

Bureau van Dijk, "Orbis Historic Disk (1996-2016)," Distributed by Hard Disk, 2017. Accessed: 2017-04-20.

Cravino, Javier and Andrei A Levchenko, "Multinational Firms and International Business Cycle Transmission," *The Quarterly Journal of Economics*, 2017, *132* (2), 921–962.

European Patent Office, "PATSTAT Global 2018 Fall Version," European Patent Office FTP, 2018. Accessed: 2019-03-28.

Feenstra, Robert C, Robert Inklaar, and Marcel P Timmer, "The next generation of the Penn World Table," *American economic review*, 2015, *105* (10), 3150–3182.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer, "Penn World Table 9.0," Groningen Growth and Development Centre, https://doi.org/10.15141/S5J01T, 2016.

Griliches, Z, "Patent statistics as economic indicators: A survey.," *Journal of Economic Literature*, 1990, *28* (4), 1661–1707.

Griliches, Zvi, "Patent Statistics as Economic Indicators: a Survey," in "R&D and productivity: the econometric evidence," University of Chicago Press, 1998, pp. 287–343.

Hanushek, Eric A. and Ludger Woessmann, "Dataset For 'Do Better Schools Lead To More Growth? Cognitive Skills, Economic Outcomes, And Causation'," https://drive.google.com/file/d/1no5LgTLH9dIzL0sRICb0d_RIHK-RqMNw/view?usp=sharing, 2012. Accessed: 2016-03-21.

Hanushek, Eric A and Ludger Woessmann, "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation," *Journal of Economic Growth*, 2012, *17* (4), 267–321.

Head, Keith and John Ries, "Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade," *American Economic Review*, 2001, *91* (4), 858–876.

Hines, James R and Eric M Rice, "Fiscal paradise: Foreign tax havens and American business," *The Quarterly Journal of Economics*, 1994, *109* (1), 149–182.

Holmes, Thomas J, Ellen R McGrattan, and Edward C Prescott, "Quid pro quo: Technology capital transfers for market access in China," *The Review of Economic Studies*, 2015, *82* (3), 1154–1193.

[20] As Column 1 of Table C.6 alludes to, to ensure this alternative model matches the same inward offshore R&D, we would need slightly lower ϕ^P_{om} than in the baseline model.

Kalemli-Ozcan, Sebnem, Bent Sorensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yesiltas, "How to Construct Nationally Representative Firm Level Data from the Orbis Global Database: New Facts and Aggregate Implications," December 2019.

Mayer, Thierry and Soledad Zignago, "The GeoDist Database," CEPII, http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=6, 2011. Accessed: 01/05/2015.

OECD, *Multinational enterprises in the global economy: Heavily debated but hardly measured*, Retrieved at: https://www.oecd.org/industry/ind/MNEs-in-the-global-economy-policy-note.pdf, 2018.

Park, Walter G, "International Patent Protection: 1960–2005," *Research Policy*, 2008, *37* (4), 761–766.

_ , "Updated Data To 'International Patent Protection: 1960–2005'," http://fs2.american.edu/wgp/www/Patent%20index1960%20-%202015.xlsx, 2015. Accessed: 2019-07-25.

Ramondo, Natalia, Andrés Rodríguez-Clare, and Felix Tintelnot, "Multinational Production: Data and Stylized Facts," *The American Economic Review*, 2015, *105* (5), 530–536.

_ **,** _ **, and** _ , "Data For 'Multinational Production: Data And Stylized Facts'," ISPCR, https://doi.org/10.3886/E113384V1, 2019.

Timmer, Marcel P, Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J De Vries, "World Input-Output Database," Groningen Growth and Development Centre, https://doi.ORG/10.34894/PJ2M1C., 2016. Accessed: 2020-03-01.